Wiley Encyclopedia of FORENSIC SCIENCE

Editors-in-Chief
Allan Jamieson
Andre Moenssens

WILEY

Wiley Encyclopedia of
FORENSIC SCIENCE

Wiley Encyclopedia of
FORENSIC SCIENCE

Volume 1
A–B

Editors-in-Chief

Allan Jamieson
The Forensic Institute, Glasgow, UK

Andre Moenssens
Forensics and Law Center, Columbia City, IN, USA

WILEY

This edition first published 2009
© 2009 John Wiley & Sons Ltd

Registered office

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

Bomb-Pulse Dating, pp. 418–422; Radiocarbon Dating, pp. 2231–2233; Dissociative Disorders, pp. 784–792; Footwear and Foot Impressions: Foot Impressions and Linking Foot to Shoe, pp. 1244–1248; Footwear and Foot Impressions: Overview, pp. 1252–1255; Forged and Counterfeit Documents, pp. 1255–1276, are all US Government works in the public domain and not subject to copyright.

Accreditation: Laboratory, pp. 1–10, is copyright of The American Society of Crime Laboratory Directors/Laboratory Accreditation Board (ASCLD/LAB) and is used here with their consent.

Juvenile Justice: Adolescent Development, pp. 1608–1612, and Juvenile Justice: Transfer to Adult, pp. 1612–1618, are copyright of the John D. and Catherine T. MacArthur Foundation and are used here with their consent.

Bloodstain Pattern Interpretation, pp. 97–134, is copyright of the author and is used here with his consent.

The right of the authors to be identified as the authors of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data

Wiley encyclopedia of forensic science / editors in chief, Allan Jamieson, Andre Moenssens.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-0-470-01826-2 (set : cloth)
1. Forensic sciences–Encyclopedias. I. Jamieson, Allan. II. Moenssens, Andre A. III. Title: Encyclopedia of forensic science.
[DNLM: 1. Forensic Medicine–Encyclopedias–English. 2. Forensic Sciences–Encyclopedias–English. W 613 W714 2009]
HV8073.W55 2009
363.2503–dc22

2009001881

A catalogue record for this book is available from the British Library.

Set in 9½/11½ pt Times by Laserwords Private Limited, Chennai, India
Printed and bound by Grafos S.A., Barcelona, Spain

Editorial Board

Editors-in-Chief

Allan Jamieson
The Forensic Institute
Glasgow
UK

Andre Moenssens
Forensics and Law Center
Columbia City, IN
USA

Editors

Behavioral Sciences
Carl Edwards
Four Oaks Institute
Dover, MA
USA

Biological Sciences
Allan Jamieson
The Forensic Institute
Glasgow
UK

Criminalistics
Claude Roux
University of Technology (UTS)
Sydney, New South Wales
Australia

Digital Evidence, Multimedia and Engineering
Zeno Geradts
Netherlands Forensic Institute
Den Haag
The Netherlands

DNA Analysis of Biological Fluids
Allan Jamieson
The Forensic Institute
Glasgow
UK

Fire and Explosives
John J. Lentini
Scientific Fire Analysis, LLC
Big Pine Key, FL
USA

Investigation & Law Enforcement
Allan M. Scott
University of Central Lancashire (UCLAN)
Preston
UK

Law
Andre Moenssens
Forensics and Law Center
Columbia City, IN
USA

MEDICINE

Pekka J. Saukko
University of Turku
Turku
Finland

STATISTICS AND THE EVALUATION OF EVIDENCE

Christophe Champod
Institut de Police Scientifique
University of Lausanne
Lausanne
Switzerland

Tacha N. Hicks
Institut de Police Scientifique
University of Lausanne
Lausanne
Switzerland

TOXICOLOGY AND DRUG ANALYSIS

Olaf H. Drummer
Monash University
Southbank, Victoria
Australia

TRACE

Sheila Willis
Forensic Science Laboratory,
Department of Justice, Equality and
Law Reform
Garda Headquarters
Dublin
Ireland

Contents

VOLUME 1

Accreditation: Laboratory 1
Accreditation: Organizational 10
Acid Phosphatase 17
Addictions 18
Adversary Systems of Justice 23
Aggression 24
Aggression: Gender Differences in 36
Airbags 51
Alcohol 58
Alcohol: Analysis 81
Alcohol: Behavioral and Medical Effects 99
Alcohol: Interaction with Other Drugs 108
Alcohol: Use, Abuse, Tolerance, and Dependency 120
Allelic Designation 126
Alterations: Erasures and Obliterations of Documents 128
Amphetamine 134
Amplified Fragment Length Polymorphism 140
Analysis: Computer Network 141
Analysis: Neutron Activation 150
Anthropology 152
Anthropology: Age Determination of Remains 179
Anthropology: Aging the Living 188
Anthropology: Ancestry and Stature Determination 191
Archaeology 199
Arson Investigation: Misconceptions and Mythology 207
Asphyxia 224
Assault: Sexually Motivated 234
Autoerotic Deaths 243
Automated Fingerprint Identification System 249
Automatism as a Defense to Crime 253
Autopsy 256
Autoradiograph 262

Battered Child Syndrome 263
Battered Spouse Syndrome 270
Battered Woman's Reality 272
Bayesian Networks 276
Behavioral Science Evidence 281
Behavioral Toxicology 290
Benzodiazepines 293
Best Evidence Rule 298
Biological Agents 299
Biological Stains 314
Biological Swabs 320
Biometric Devices 321
Blood Grouping 338
Bloodstain Pattern Interpretation 359
Blunt Force Trauma 396
Bomb Scene Management 411
Bomb-Pulse Dating 418
Botany 422

VOLUME 2

Cannabis 431
Capacity Assessment 437
Capacity for Independent Living 444
Capacity to Consent to Medical Treatment 450
Capacity to Stand Trial 456
Capacity to Waive Miranda Rights 463
Cardiac and Natural Causes of Sudden Death 468
Case Assessment and Interpretation 483
Ceiling Principle: DNA 497
Chain of Possession of Tangible Evidence 498

Chemical, Biological, Radiological, and Nuclear Investigations 500
Chemical Warfare Agents 507
Child Sexual Abuse 529
Child Sexual Abuse Accommodation 537
Children: as Defendants 542
Children: as Witnesses 549
Children: Suggestibility of 553
Civil Commitment 556
Civil Law Systems of Justice 561
Cocaine 562
Cofiler/CofilerPlus 569
Compulsion 571
Compulsion Defense 576
Computer Animation and Simulation Evidence 579
Computers 584
Confessions: Evidentiary Reliability of 588
Confirmation Testing: Toxicology 595
Crime Scene Documentation 602
Crime Scene Investigation 614
Crime Scene Management 619
Crime Scene Photography: US Perspective 625
Crime Victims' Decision to Report Crime 643
Criminalization of the Mentally Ill 649
Cross-Examination: Impact on Testimony 656
Cross-Examination of Experts 662

Dangerousness: Risk of 667
Databases 677
Dating: Document 684
Daubert v. Merrell Dow Pharmaceuticals 692
Death: Time of 697
Death Penalty and Age 717
Deception: Detection of 720
Deception: Detection of and Brain Imaging 724
Deception: Truth Serum 728
Delusions 741
Demonstrative Evidence 745
Diatoms 748
Differential Extraction 757
Direct Examination of Experts 758
Disaster Mental Health 760
Disaster Victim Identification 764
Discovery: Depositions 772
Discovery: Discovery Motions 774
Discovery in the United States: Civil Cases 775
Discovery in the United States: Criminal Cases 778
Discovery of Expert Findings 781
Dissociative Disorders 784
DNA 792
DNA: an Overview 800
DNA: Degraded Samples 816
DNA: Sources of 821
DNA Databases and Evidentiary Issues 831
Documents: Authentication of 840
DQα 842
Drug Analysis 844
Drug Profiling 851
Drug Testing: Urine 860
Drug-Facilitated Sexual Assault 868
Drug-Impaired Driving 877
Duty to Warn 885

Earprints: Interpretation of 891
Education and Accreditation in Forensic Science 897
Elder Abuse: Policy 902
Elder Abuse: Risk 912
Elderly in Court 916
Electrical Engineering 920
Entomology 934
Environmental Science 946
Enzymes 954
Error Rates in Forensic Methods 955
Ethics: Codes of Conduct for Expert Witnesses 957
Evidence: Rules of 963
Evidence Collection and Preservation: Casting 963
Evidence Interpretation: a Logical Approach 968
Evil: Illusion of 977
Examination of Fibers and Textiles 985
Expert Opinion: Appeal v. Trial 998
Expert Opinion: United States 1001
Expert Opinion: United Kingdom, Canada, and Australia 1003
Expert Opinion in Court: a Comparison of Approaches 1004

Expert Opinion in Court: Civil Law Jurisdictions (France, Germany, Italy, and Spain) 1007
Expert Witness: Who Is? 1012
Expert Witnesses: Selection and Investigation of Credentials 1013
Explosions: Scene Investigation 1019
Explosion Debris: Laboratory Analysis of 1028
Extraction 1060
Eyewitness: Suggestibility of 1065
Eyewitness Lineups: Identification from 1072
Eyewitness Testimony 1075

VOLUME 3

Facial Comparison 1081
Facial Reconstruction 1086
Falsifiability Theory 1093
Federal Rule of Evidence 702 1095
Fibers 1095
Fire: Chemistry of 1103
Fire: Dynamics and Pattern Production 1112
Fire: Scene Investigation 1122
Fire and Explosion Investigations: Overview 1136
Fire Debris: Laboratory Analysis of 1137
Fire Investigator: Standardization, Accreditation, and Certification 1171
Fire Modeling and Its Application in Fire Investigation 1175
Firearm Discharge Residue: Analysis of 1189
Firearm Examination: Ballistics 1200
Firearms: Bullet and Cartridge Case Identification 1204
Firearms: Identification of Handling of Firearms/Trace Metal Detection 1211
Firearms: Overview 1216
Firearms: Scene Investigation 1219
Firesetting 1225
Footwear and Foot Impressions: Comparison and Identification 1230
Footwear and Foot Impressions: Databases 1240
Footwear and Foot Impressions: Foot Impressions and Linking Foot to Shoe 1244
Footwear and Foot Impressions: Intelligence 1248
Footwear and Foot Impressions: Overview 1252
Forged and Counterfeit Documents 1255
Foundation Testimony 1276
Friction Ridge Examination (Fingerprints): Interpretation of 1277
Friction Ridge Skin: Comparison and Identification 1282
Friction Ridge Skin: Fingerprint Detection and Recovery Techniques 1292
Friction Ridge Skin: Interaction between Fingerprint Detection and DNA/Biological Material 1318
Friction Ridge Skin: Morphogenesis and Overview 1322
Frye v. United States 1331

General Acceptance Test for Novel Expert Evidence 1333
General Electric v. Joiner 1334
Genomics and Behavioral Evidence 1335
Geographical Identification by Viral Genotyping 1342
Glass 1348
Glass Evidence: Bayesian Approach to 1351
GSM Analysis and PDAs 1360
Guardianships of Adults 1371
Gunshot Wounds 1380

Hair: Animal 1403
Hair: Microscopic Analysis 1415
Hair: Toxicology 1427
Hallucinations 1432
Handwriting and Signatures, Comparison of 1436
Handwriting and Signatures, Interpretation of Comparison Results 1451
Hardy-Weinberg Equilibrium 1458
Head Injury: Neuropsychological Assessment 1459
Hearsay Evidence 1465
Histology 1468

Homicide: Multiple (Behavior) 1474
Homosexual Panic 1480
Human Factors: Industrial Incidents 1483
Human Remains and Identity 1495
Hypnosis and Memory 1500
Hypothetical Question 1505

Identification and Individualization 1508
Identification of Human Remains 1511
Identifiler 1518
Image Processing and Analysis 1520
In Limine Motions and Hearings 1528
Injury: Burns, Scalds, and Chemical 1529
Ink Analysis 1541
Ink Comparison and Interpretation 1546
Insanity: Defense 1552
Interpretation: Document Evidence 1557
Interpretation: Legal Perspective 1561
Interpretation: Low Template DNA 1566
Interpretation: Observer Effects 1575
Interpreting Expert Opinions: History of 1579
Interrogation 1586
Interrogative Suggestibility 1590
Intersecting Lines: Documents 1594
Ipse Dixit Testimony 1599

Judicial Notice of Scientific Principles and Facts 1601
Jury Dynamics 1602
Jury Instructions on Expert Testimony 1607
Juvenile Justice: Adolescent Development 1608
Juvenile Justice: Transfer to Adult 1612

Kumho Tire v. Carmichael 1619

Learned Treatises as Evidence 1623
Length Measurement 1624
Light Bulbs and Filaments: Examination of 1632
Low Copy Number DNA 1639
Luminol 1645

VOLUME 4

Malingering: Forensic Evaluations 1657
Malpractice Actions against Experts 1663
Marks or Impressions of Manufactured Items 1668
Mass Grave Investigation 1674
Materials Science 1680
Matrix: DNA 1688
Medical Malpractice 1689
Memory: Reconstructive 1709
Memory: Repressed 1712
Mental Health Courts 1717
Mental Retardation 1724
Mental Retardation: Death Penalty 1730
Mental Status: Examination 1737
Microchemistry 1743
Microsatellites 1749
Microscopy: FTIR 1750
Microscopy: High Power 1758
Microscopy: Light Microscopes 1762
Microscopy: Low Power 1791
Microscopy: Scanning Electron Microscopy 1793
Mini-STRs 1804
Missing Persons and Paternity: DNA 1810
Mitigation Testimony 1818
Mitochondrial DNA: Interpretation 1823
Mitochondrial DNA: Profiling 1833
Mixture Interpretation: DNA 1838

Natural Causes of Sudden Death: Noncardiac 1843
Neuropsychological Assessment 1862
Neuropsychological Assessment: Child 1869
Northwest Juvenile Project 1877
Nuclear Forensics 1883

Odontology 1889
Opioids 1895
Oral Fluid Toxicology 1903

Packaging and Transport 1927
Paint 1931
Paint: Interpretation 1943
Palynology 1954
Paper Analysis 1968
Parental Alienation 1981
Parental Rights and Prerogatives 1984
Parenting: Assessment of Capacity 1989
Particles: Form 2001
Peak Height: DNA 2007

Peer Review as Affecting Opinion Evidence 2009
Pharmacogenomics 2012
Phenotype 2021
Photography: Marks, Impressions, and Documents 2036
Poisons: Detection of Naturally Occurring Poisons 2057
Police Use of Force 2068
Policing and Critical Incident Teams 2071
Polymorphism: Genetic 2075
Postmortem Biochemical Examinations 2076
Postmortem Interval: Anthropology 2089
Postmortem Toxicology: Artifacts 2093
Postmortem Toxicology: Interpretation 2115
Postmortem Toxicology: Laboratory Analysis 2119
Postpartum Psychosis 2136
Posttraumatic Stress Disorder 2141
PowerPlex 2149
Premenstrual Syndrome 2149
Profiles: Psychological and Behavioral 2156
Psychological Autopsy 2161
Psychological Testing 2173
Psychopathology: Terms and Trends 2186
Psychopathy 2193
Psychopathy Checklists 2197
Psychopharmacology 2201
Psychopharmacology: Child and Adolescent 2210

QiaAmp 2219
Quality Systems: Toxicology 2219

VOLUME 5

Radiocarbon Dating 2231
Radiology 2233
Rape Trauma Syndrome 2241
Recollective Accuracy of Traumatic Memories 2243
Reconstruction: Accident 2250
Reconstruction: Three Dimensional 2257
Report Writing for Courts 2268
Risk Assessment 2271
Risk Assessment: Patient and Detainee 2272

Sampling and Estimation of Quantities 2281
Sampling Trace Evidence 2291
Scientific Method Compared to Legal Method 2296
Seizures: Behavioral 2298
Sentencing: Demographic Factors in 2306
Serial Homicide 2311
Serial Number Restoration: Firearm 2324
Sex Determination of Remains 2328
Sex Offenders: Treatment of 2332
Shaken Baby Syndrome 2339
Shooting Distance: Estimation of 2351
Short Tandem Repeats 2354
Short Tandem Repeats: Interpretation 2365
Soil: Forensic Analysis 2377
Speaker Recognition 2389
Species Determination of Osseous Remains 2393
Stalking 2397
Statistical Evidence in Court 2401
Stockholm Syndrome 2409
Substance Abuse 2413
Suicide (Behavior) 2418
Sweat: Toxicology 2420
Syndromes: Psychological 2431

Temporary Insanity 2443
Therapeutic Jurisprudence 2449
Threat Assessment: School 2454
Threat Assessment: Workplace 2460
Time of Death Determinations 2466
Tire Impressions 2480
Toolmarks 2485
Toxicology: Analysis 2495
Toxicology: Forensic Applications of 2503
Toxicology: Initial Testing 2509
Trace Evidence: Transfer, Persistence, and Value 2534
Traffic Fatalities 2541
Training and Certification (in Criminalistics) 2545
Transfer: DNA 2555
Trauma Analysis of Skeletal Remains 2557

Trauma Causation: Analysis of Automotive 2565
Treatment, Mandated: Mental Health 2576
Treatment, Right to: Mental Health 2580
Treatment, Right to Refuse: Mental Health 2584

Ultimate Issue Evidence by Experts 2589
Use of Knowledge-Based Systems in Forensic Science 2590

Variable Number Tandem Repeats 2595
Violence Risk Assessment for Mental Health Professionals 2597
Visitation Rights 2602
Visual Recognition Systems in Identification 2611
Web Resources 2619
Weisgram v. Marley 2627
Whole Genome Amplification 2628
Wildlife 2635
Wood 2640
Wounds: Sharp Injury 2646
Writing Instruments and Printing Devices 2660

Y-Chromosome Short Tandem Repeats 2677

Glossary 2683
Author Index 2701
Subject Index 2707

Contributors

Abbondante, Serena F. *Australian Federal Police, Weston and Australian Chemical, Biological Radiological and Nuclear Data Centre, Canberra, ACT, Australia*

Abdel-Monem, Tarik *University of Nebraska Public Policy Center, Lincoln, NE, USA*

Abou-Khalil, Bassel *Vanderbilt University School of Medicine, Nashville, TN, USA*

Abram, Karen M. *Northwestern University Feinberg School of Medicine, Chicago, IL, USA*

Adams, Holly A. *Automotive Data Consultants, Centreville, VA, USA*

Aitken, Colin G. G. *University of Edinburgh, Edinburgh, UK*

Alberink, Ivo *Netherlands Forensic Institute, Den Haag, The Netherlands*

Aleksander, Adam K. *Aleksander & Associates P.A., Boise, ID, USA*

Allen, Rebecca S. *Center for Mental Health and Aging, Tuscaloosa, AL, USA*

Almog, Joseph *The Hebrew University of Jerusalem, Jerusalem, Israel*

Anderson, Robert N. *RNA Consulting, Inc., Losaltos Hills, CA, USA*

Andrews, Paul *Tyler, TX, USA*

Anetzberger, Georgia J. *Cleveland State University, Cleveland, OH, USA*

Aumeer-Donovan, Shaheen *University of Technology, Sydney, New South Wales, Australia*

Baden, Michael M. *New York State Police, Albany, NY, USA*

Bader, Scott *The Forensic Institute, Glasgow, UK*

Baker, David W. *The MITRE Corporation, McLean, VA, USA*

Balding, David *Imperial College, London, UK*

Baldwin, David *London Laboratory, London, UK*

Ballantyne, Jack *University of Central Florida and National Center for Forensic Science, Orlando, FL, USA*

Barnes, Sean *Binghamton University, Binghamton, NY, USA*

Barni, Filippo *Carabinieri Scientific Investigation Department of Rome, Rome, Italy*

Benbow, M. Eric *Michigan State University, East Lansing, MI, USA*

Berkowitz, Shari R. *University of California, Irvine, CA, USA*

Bernet, William *Vanderbilt University School of Medicine, Nashville, TN, USA*

Beyer, Jochen *Monash University, Southbank and Victorian Institute of Forensic Medicine, Melbourne, Victoria, Australia*

Bicknell, Danna E. *United States Secret Service, Washington, DC, USA*

Black, Sue *University of Dundee, Dundee, Scotland, UK*

Block, Stephanie *University of California, Davis, CA*

Bohnert, Michael *University of Freiburg, Freiburg, Germany*

Botluk, Diana *Stetson University College of Law, Gulfport, FL, USA*

Bottoms, Bette L. *University of Illinois at Chicago, Chicago, IL, USA*

BOWMAN-FOWLER, NICCI *University of California, Irvine, CA, USA*
BRAUN, MICHELLE *Wheaton Franciscan Healthcare, Racine, WI, USA*
BRESLER, SCOTT A. *University of Cincinnati, Cincinnati, OH, USA*
BRICK, JOHN *Intoxikon International, Yardley, PA, USA*
BRIGHT, JO-ANNE *Institute of Environmental Science and Research Limited, Auckland, New Zealand*
BRYANT, VAUGHN M. *Texas A&M University, College Station, TX, USA*
BUCHHOLZ, BRUCE A. *Lawrence Livermore National Laboratory, Livermore, CA, USA*
BUCKLETON, JOHN S. *Institute of Environmental Science and Research, Ltd., Auckland, New Zealand*
BULLING, DENISE *University of Nebraska Public Policy Center, Lincoln, NE, USA*
BURGESS, ANN W. *Boston College, Chestnut Hill, MA, USA*

CANTU, ANTONIO A. *Seven Oaks Place, Falls Church, VA, USA*
CARPENTER, DOUGLAS J. *Combustion Science & Engineering, Inc., Columbia, MD, USA*
CATTANEO, CRISTINA *Universitá degli Studi, Milan, Italy*
CHAMPOD, CHRISTOPHE *Institut de Police Scientifique, University of Lausanne, Lausanne, Switzerland*
CHENG, WING-CHI *Government Laboratory, Hong Kong Special Administrative Region, China*
CHOI, HYEYOUNG *National Institute of Scientific Investigation, Seoul, South Korea*
CHOI, SANGKIL *National Institute of Scientific Investigation, Seoul, South Korea*
CHRISTENSEN, THOMAS C. *San Ramon, CA, USA*
CHUNG, HEESUN *National Institute of Scientific Investigation, Seoul, South Korea*
CLEGG, CARL *West Virginia University, Morgantown, WV, USA*
COBLE, MICHAEL D. *The Armed Forces DNA Identification Laboratory, Rockville, MD, USA*
CONNOR, MELISSA *Nebraska Wesleyan University, Lincoln, NE, USA*
CORNELL, DEWEY G. *University of Virginia, Charlottesville, VA, USA*
COSTANZO, MARK *Claremont McKenna College, Claremont, CA, USA*
COSTELLO, JAN *Loyola Law School, Los Angeles, CA, USA*
COURT, DENISE S. *Barts and The London School of Medicine and Dentistry, London, UK*
COWELL, ANTHONY M. *University of Lincoln, Lincoln, UK*
CROSS, DOUGLAS W. *Lowick Bridge, Ulverston, UK*
CURRAN, JAMES M. *University of Auckland, Auckland, New Zealand*

DAVIS, MALCOLM *Vashaw Scientific, Inc., Norcross, GA, USA*
DAY, STEPHEN P. *Huntingdon Forensic Science Laboratory, Cambridgeshire, UK*
DE ANGELIS, DANILO *Università degli Studi, Milan, Italy*
DE BOECK, GERT *National Institute of Criminalistics and Criminology, Brussels, Belgium*
DE LA TORRE, RAFAEL *Neuropsychopharmacology Program IMIM-Hospital del Mar PRBB, Barcelona, Spain*
DEN DUNNEN, M. *Amsterdam-Amstelland Police, Amsterdam, The Netherlands*
DICKSON, STUART *Institute of Environmental Science and Research Limited, Porirua, New Zealand*
DIETZ, PARK *Threat Assessment Group, Inc., and Park Dietz & Associates, Inc., Newport Beach and University of California, Los Angeles, CA, USA*
DOUGLAS, KEVIN S. *Simon Fraser University, Burnaby, British Columbia, Canada*

DRUMMER, OLAF H. *Monash University, Southbank, Victoria, Australia*
DUTTON, GERARD *Tasmania Police, Hobart, Tasmania, Australia*
DUVINAGE, NICOLAS *Gendarmerie National Forensic Sciences Institute (IRCGN), Rosny-sous-Bios, France*
DU PREEZ, CHARL *University of Technology, Sydney, New South Wales, Australia*

EASTEAL, PATRICIA *University of Canberra, Canberra, ACT, Australia*
EDELMAN, GERDA *Netherlands Forensic Institute, Den Haag, The Netherlands*
EDELSTEIN, BARRY A. *West Virginia University, Morgantown, WV, USA*
EDWARDS, CARL N. *Four Oaks Institute, Dover, MA, USA*
ELKINGTON, KATE S. *Columbia University and New York State Psychiatric Institute, New York, NY, USA*
ERICKSON, STEVEN K. *University of Pennsylvania Law School, Philadelphia, PA, USA*
ERIKSSON, ANDERS F. *Umeå University, Umeå, Sweden*
ERVIN, THOMAS *The MITRE Corporation, McLean, VA, USA*
ESSEIVA, PIERRE *University of Lausanne, Lausanne, Switzerland*
EVETT, IAN W. *London Laboratory, London, UK*

FAGAN, JEFFREY *Columbia Law School, New York, NY, USA*
FELGATE, PETER *Forensic Science South Australia, Adelaide, South Australia, Australia*
FIDDIAN, SUSAN *Victoria Police Forensic Services Department, McLeod, Victoria, Australia*
FINCH, INDRA A. *Center for Forensic Services – Western State Hospital, Tacoma, WA, USA*
FINESCHI, VITTORIO *University of Foggia, Foggia, Italy*
FINKENBINE, RYAN *West Virginia University, Morgantown, WV, USA*
FITTERMAN, ELIZABETH *Stetson University College of Law, Gulfport, FL, USA*
FITZPATRICK, ROBERT W. *Centre for Australian Forensic Soil Science/CSIRO Land and Water, Adelaide, South Australia, Australia*
FLANAGAN, ROBERT J. *King's College Hospital NHS Foundation Trust, London, UK*
FOUND, BRYAN *Latrobe University, Bundoora and Victoria Police Forensic Services Department, Macleod, Victoria, Australia*
FOWLER, NICCI B. *University of California, Irvine, CA, USA*
FRAZIER, LEEANNE *Stetson University College of Law, Gulfport, FL, USA*
FREMOUW, WILLIAM J. *West Virginia University, Morgantown, WV, USA*
FRUDAKIS, TONY *DNAPrint Genomics, Inc., Sarasota, FL, USA*
FRUMKIN, BRUCE *Forensic and Clinical Psychology Associates, South Miami, FL, USA*

GALLO, FRANK J. *Western New England College, Springfield, MA, USA*
GANIS, GIORGIO *Harvard Medical School, Boston, and Atinoulas Martinos Center, Charlestown, and Harvard University, Cambridge, MA, USA*
GARDNER, ROSS M. *Bevel, Gardner and Associates, Inc., Lake City, GA, USA*
GELLER, JEFFREY L. *University of Massachusetts Medical School, Worcester, MA, USA*
GERADTS, ZENO *Netherlands Forensic Institute, Den Haag, The Netherlands*
GERAERTS, ELKE *Harvard University, Cambridge, MA, USA and Maastricht University, Maastricht, The Netherlands*
GEROSTAMOULOS, DIMITRI *Monash University, Southbank, Victoria, Australia*
GIBELLI, DANIELE *Universitá degli Studi, Milan, Italy*
GIBLIN, MARY *Garda Headquarters, Dublin, Ireland*

GILDER, JASON R. *Forensic Bioinformatics, Fairborn, OH, USA*
GILLMAN, VICTORIA C. *Australian Federal Police, Weston and Australian Chemical, Biological Radiological and Nuclear Data Centre, Canberra, ACT, Australia*
GITLOW, STUART *Mount Sinai School of Medicine, New York, NY, USA*
GODDARD, KEN *National Fish and Wildlife Forensics Laboratory, Ashland, OR, USA*
GOODMAN, GAIL S. *University of California, Davis, CA, USA*
GOODWIN, KERRI A. *Towson University, Towson, MD, USA*
GOULD, CHRISTINE E. *West Virginia University, Morgantown, WV, USA*
GRAHAM, ELEANOR A.M. *University of Leicester, Leicester, UK*
GREAVES, CAROLINE *BC Mental Health & Addiction Services, Port Coquitlam and Simon Fraser University, Burnaby, British Columbia, Canada*
GREENBERG, MARTIN S. *University of Pittsburgh, Pittsburgh, PA, USA*
GREENE, EDIE *University of Colorado at Colorado Spring, Colorado Spring, CO, USA*

HACKMAN, LUCINA *University of Dundee, Dundee, UK*
HAMILTON, WARREN D. *Rapidtox Pty. Ltd., Brisbane, Queensland, Australia*
HAMMER, LESLEY *State of Alaska Crime Laboratory, Anchorage, AK, USA*
HANSON, ERIN K. *University of Central Florida and National Center for Forensic Science, Orlando, FL, USA*
HAN, EUNYOUNG *National Institute of Scientific Investigation, Seoul, South Korea*
HARBISON, SALLY-ANN *Institute of Environmental Science and Research Ltd., Auckland, New Zealand*
HART, STEPHEN D. *Simon Fraser University, Burnaby, British Columbia, Canada*
HASEL, LISA E. *Iowa State University, Ames, IA, USA*
HATTERS-FRIEDMAN, SUSAN *University Hospital – Case Medical Center, Cleveland, OH, USA*
HAYNE, HARLENE *University of Otago, Dunedin, New Zealand*
HAZELWOOD, ROBERT R. *Academy Group, Inc., Manassas, VA, USA*
HENDERSON, CAROL *Stetson University College of Law, Gulfport, FL, USA*
HICKS, TACHA N. *Institut de Police Scientifique, University of Lausanne, Lausanne, Switzerland*
HILL, CHERYL A. *West Virginia University, Morgantown, WV, USA*
HONTS, CHARLES ROBERT *Boise State University, Boise, ID, USA*
HOPEN, THOMAS J. *The Bureau of Alcohol, Tobacco, Firearms and Explosives, Atlanta, GA, USA*
HUNTER, JOHN *University of Birmingham, Birmingham, UK*

IKEGAYA, HIROSHI *Kyoto Prefectural University of Medicine, Kyoto, Japan*
ISENSCHMID, DANIEL S. *Wayne County Medical Examiner's Office, Detroit, MI, USA*

JACKSON, GRAHAM *Advance Forensic Science, and University of Abertay Dundee, Dundee, UK*
JACKSON, MICHAEL *New South Wales Police Force, Sydney, New South Wales, Australia*
JAMIESON, ALLAN *The Forensic Institute, Glasgow, UK*
JONES, ALAN W. *National Board of Forensic Medicine, Linköping, Sweden*
JONES, GRAHAM R. *Office of the Chief Medical Examiner, Edmonton, Alberta, Canada*
JONES, PHILIP J. *York, North Yorkshire, UK*
JUST, REBECCA S. *The Armed Forces DNA Identification Laboratory, Rockville, MD, USA*

KASSIN, SAUL *John Jay College of Criminal Justice, New York, NY, USA*

KATSUMATA, YOSHINAO *National Research Institute of Police Science, Chiba, Japan*
KATTERWE, HORST *Bundeskriminalamt, Wiesbaden, Germany*
KAYE, DAVID H. *Arizona State University, Tempe, AZ, USA*
KAYE, NEIL S. *Widener University School of Law, Wilmington, DE, USA*
KEATON, RALPH *American Society of Crime Laboratory Directors/ Laboratory Accreditation Board (ASCLD/LAB), Garner, NC, USA*
KEEREWEER, ISAÄC *Netherlands Forensic Institute, The Hague, The Netherlands*
KEHN, ANDRE *University of Wyoming, Laramie, WY, USA*
KENAN, JOSEPH *University of California Los Angeles School of Medicine, Beverly Hills, CA, USA*
KENNEDY, ROBERT *Royal Canadian Mounted Police (Retired), Ottawa, Ontario, Canada*
KERNBACH-WIGHTON, GERHARD *University of Edinburgh, Edinburgh, UK*
KHANMY-VITAL, AITA *University of Lausanne, Lausanne, Switzerland*
KIM, EUNMI *National Institute of Scientific Investigation, Seoul, South Korea*
KINTZ, PASCAL *Laboratoire ChemTox, Illkirch, France*
KIRKBRIDE, K. PAUL *Australian Federal Police, Canberra, ACT, Australia*
KNOLL, IV, JAMES L. *SUNY Upstate Medical University, Syracuse, NY, USA*
KOEHLER, JONATHAN J. *Arizona State University, Tempe, AZ, USA*
KOPERSKI, J. GEORGE *Australian Federal Police, Weston and Australian Chemical, Biological Radiological and Nuclear Data Centre, Canberra, ACT, Australia*
KOSSLYN, STEPHEN M. *Harvard University, Cambridge and Massachusetts General Hospital, Boston, MA, USA*
KRANE, DAN E. *Wright State University, Dayton, OH, USA*
LAMENDOLA, GRETCHEN M. *Nova Southeastern University, Fort Lauderdale, FL, USA*
LANCASTER, SARAH L. *Defence Science and Technology Laboratories, Sevenoaks, UK*
LANGENBURG, GLENN *Minnesota Bureau of Criminal Apprehension, St. Paul, MN, USA*
LAPORTE, GERALD M. *United States Secret Service, Washington, DC, USA*
LAUX, DALE L. *Attorney General's Office, Richfield, OH, USA*
LEBEAU, MARC A. *FBI Laboratory, Quantico, VA, USA*
LEE, JUSEON *National Institute of Scientific Investigation, Seoul, South Korea*
LEE, LI-WEN G. *New York University School of Medicine, New York, NY, USA*
LEE, SOOYEUN *National Institute of Scientific Investigation, Seoul, South Korea*
LENNARD, CHRIS *University of Canberra, Canberra, ACT, Australia*
LENTINI, JOHN J. *Scientific Fire Analysis, LLC, Big Pine Key, FL, USA*
LENZ, KURT W. *Stetson University College of Law, Gulfport, FL, USA*
LEONG, GREGORY B. *University of Washington, Seattle, WA, USA*
LEO, RICHARD A. *University of San Francisco, San Francisco, CA, USA*
LEWIS, SIMON W. *Curtin University of Technology, Perth, Western Australia, Australia*
LIM, MIAE *National Institute of Scientific Investigation, Seoul, South Korea*
LIPTAI, LAURA L. *Biomedical Forensics, Moraga, CA and Orlando, FL, USA*
LOFTUS, ELIZABETH F. *University of California, Irvine, CA, USA*
LOVELL, ROBERT W. *Mercer Island, WA, USA*
LOVELOCK, TINA J. *LGC Forensics, Abingdon, UK*

LUONG, SUSAN *University of Technology, Sydney, New South Wales, Australia*
LYNN, STEVEN JAY *Binghamton University, Binghamton, NY, USA*

MAAT, G.J.R. *Netherlands Forensic Institute, The Hague and Leiden University Medical Center, Leiden, The Netherlands*
MACDONELL, HERBERT L. *Bloodstain Evidence Institute, Corning, NY, USA*
MACEO, ALICE V. *Las Vegas Metropolitan Police Department Forensic Laboratory, Las Vegas, NV, USA*
MACVAUGH III, GILBERT S. *Mississippi State Hospital, Whitfield, MS, USA*
MADEA, BURKHARD *University of Bonn, Bonn, Germany*
MAIDEN, NICHOLAS R. *South Australia Police, Adelaide, South Australiaa, Australia*
MALLETT, XANTHÉ *University of Dundee, Dundee, UK*
MARGOT, PIERRE *University of Lausanne, Lausanne, Switzerland*
MARLEEN, LALOUP *National Institute of Criminalistics and Criminology, Brussels, Belgium*
MARSHALL, MAURICE *Defence Science and Technology Laboratories, Sevenoaks, UK*
MARTELL, DANIEL A. *UCLA, Los Angeles, CA, USA*
MASSONNET, GENEVIÈVE *University of Lausanne, Lausanne, Switzerland*
MASTRUKO, VOJIN *Court Expert Witness, Zagreb, Croatia*
MATTHEWS, ABIGAIL *Binghamton University, Binghamton, NY, USA*
MAZZELLA, W.D. *University of Lausanne, Lausanne, Switzerland*
MCCOY, KATRINA *West Virginia University, Morgantown, WV, USA*
MCCULLOUGH, JOHN *Garda HQ, Dublin, Ireland*
MCDERMOTT, SEAN D. *Forensic Science Laboratory, Dublin, Ireland*
MCKENNA, LOUISE *Garda Headquarters, Dublin, Ireland*
MCNALLY, RICHARD J. *Harvard University, Cambridge, MA, USA*
MEIJERMAN, L. *Netherlands Forensic Institute, The Hague, and Leiden University Medical Center, Leiden, The Netherlands*
MELOY, J. REID *University of California, San Diego, California, USA*
MELSON, KENNETH E. *American Society of Crime Laboratory Directors/ Laboratory Accreditation Board (ASCLD/LAB), Garner, NC, USA*
MELTON, TERRY *Mitotyping Technologies, LLC, State College, LA, USA*
MERRITT, RICHARD W. *Michigan State University, East Lansing, MI, USA*
MEUWLY, DIDIER *Netherlands Forensic Institute, The Hague, The Netherlands*
MICHEALS, ANASTASIA D. *San Jose State University, San Jose, CA, USA*
MILES, SAMUEL I. *Geffen School of Medicine UCLA, Los Angeles and Cedars-Sinai Medical Center, Los Angeles, CA, USA*
MOENSSENS, ANDRE *Forensics and Law Center, Columbia City, IN, USA*
MOHAMMED, LINTON A. *San Diego Sheriff's Regional Crime Laboratory, San Diego, CA, USA*
MONNARD, FLORENCE *University of Lausanne, Lausanne, Switzerland*
MORETTI, MARLENE M. *Simon Fraser University, Burnaby, British Columbia, Canada*
MORRISH, BRONWYN C. *Australian Federal Police, Weston and Australian Chemical, Biological Radiological and Nuclear Data Centre, Canberra, ACT, Australia*
MUELLER-JOHNSON, KATRIN *University of Cambridge, Cambridge, UK*
MURPHY, JOHN P. *CSIRO Forest Biosciences, Clayton, Victoria, Australia*

NEHSE, KORNELIA *Forensic Science Institute, Berlin, Germany*
NELE, SAMYN *National Institute of Criminalistics and Criminology, Brussels, Belgium*

NELSON, KALLY J. *University of California, Irvine, CA, USA*
NERENBERG, LISA *Private Consultant, Redwood City, CA, USA*
NEUMANN, CEDRIC *The Forensic Science Service Ltd, Birmingham, UK and University of Lausanne, Lausanne, Switzerland*
NEUNER, JOHN K. *American Society of Crime Laboratory Directors/ Laboratory Accreditation Board (ASCLD/LAB), Garner, NC, USA*
NEUSCHATZ, JEFFREY S. *University of Alabama in Huntsville, Huntsville, AL, USA*
NICHOLLS, TONIA L. *BC Mental Health & Addiction Services, Port Coquitlam and University of British Columbia, Vancouver, British Columbia, Canada*
NIKOLOVA, NATALIA L. *Simon Fraser University, Burnaby, British Columbia, Canada*
NORMAN, KEITH W. *Australian Federal Police, Weston, ACT, Australia*
NUNEZ, NARINA L. *University of Wyoming, Laramie, WY, USA*

OJANPERÄ, ILKKA *University of Helsinki, Helsinki, Finland*
OLLEY, J. GREGORY *University of North Carolina, Chapel Hill, NC, USA*
ÖSTRÖM, MATS G. *Umeå University, Umeå, Sweden*
OXLEY, JIMMIE C. *University of Rhode Island, Kingston, RI, USA*

PARK, YONGHOON *National Institute of Scientific Investigation, Seoul, South Korea*
PAYNE-JAMES, JASON *Forensic Healthcare Services, Leigh-on-Sea and Royal College of Physicians and Barts & the Royal London Hospitals, London, UK*
PERRY, SYLVIA *University of Illinois at Chicago, Chicago, IL, USA*
PETERSON, TIAMOYO *University of California, Irvine, CA, USA*
PICHINI, SIMONA *Istituto Superiore di Sanità, Rome, Italy*
PINALS, DEBRA A. *University of California, Sacramento, CA, USA*
PIPER, AUGUST *Seattle, WA, USA*
POLLAK, STEFAN *University of Freiburg, Freiburg, Germany*
POLLANEN, MICHAEL S. *University of Toronto, Toronto, Ontario, Canada*
PORTA, DAVIDE *Università degli Studi, Milan, Italy*
PORTER, GLENN *University of Western Sydney, Penrith South DC, New South Wales, Australia*
POULSEN, HELEN *Institute of Environmental Science and Research Limited, Porirua, New Zealand*

QUINN, MARY J. *San Francisco Probate Court, San Francisco, CA, USA*

RAES, ELKE *Ghent University, Ghent, Belgium*
RAFF, ADAM N. *New York University Medical Center, New York, NY, USA*
RANDOLPH-QUINNEY, PATRICK *University of Dundee, Dundee, UK*
RAYMOND, JENNIFER J. *NSW Police Force, Pemulwuy, New South Wales, Australia*
RAY, NEELANJAN *New York University School of Medicine, New York, NY, USA*
REED, TOM *Widener University School of Law, Wilmington, DE, USA*
RESNICK, PHILLIP J. *Case Western Reserve University Medical School, Cleveland, OH, USA*
RESOR, MICHELLE R. *University of North Carolina at Charlotte, Charlotte, NC, USA*
RESSLER, ROBERT K. *Forensic Behavioral Services, Fredericksburg, VA, USA*
RIEZZO, IRENE *University of Foggia, Foggia, Italy*
ROBERTSON, JAMES *Australian Federal Police, Canberra, ACT, Australia*
ROFFEY, PAUL E. *Australian Federal Police, Weston and University of Canberra, Canberra, ACT, Australia*
ROMERO, ERIN G. *Northwestern University, Chicago, IL, USA*

ROUX, CLAUDE *University of Technology (UTS), Sydney, New South Wales, Australia*
ROYDS, DAVID *Australian Federal Police, Canberra, Australia*
RUIFROK, ARNOUT C. C. *Netherlands Forensic Institute, The Hague, The Netherlands*

SALEKIN, KAREN L. *University of Alabama, Tuscaloosa, AL, USA*
SALSAROLA, DOMINIC *Universitá degli Studi, Milan, Italy*
SAUKKO, PEKKA J. *University of Turku, Kiinamyllynkatu, Turku, Finland*
SAUVAGNAT, FRANÇOIS *Université de Rennes-II, Rennes, France*
SCHIFFER, BEATRICE *University of Lausanne, Lausanne, Switzerland*
SCHNECK, WILLIAM M. *Microvision Northwest-Forensic Consulting, Inc. Spokane, WA, USA*
SCHNEIDER, RICHARD D. *The Ontario Court of Justice and University of Toronto, Toronto, Ontario, Canada*
SCOTT, ALLAN MATHIESON *University of Central Lancashire (UCLAN), Preston, UK*
SCOTT, CHARLES L. *University of California, Sacramento, CA, USA*
SEGOVIA, DAISY A. *University of California, Davis, CA, USA*
SHARFE, GORDON A.I. *Wellington Central Police Station, Wellington, New Zealand*
SHEFCHICK, THOMAS P. *Shefchick Engineering, Sunnyvale, CA, USA*
SHIVER, FARRELL C. *Shiver & Nelson Document Investigation Laboratory, Inc., Woodstock, GA, USA*
SIEGEL, JAY A. *Indiana University Purdue University Indianapolis, Indianapolis, IN, USA*
SILVA, J. ARTURO *Private Practice of Forensic Psychiatry, San Jose, CA, USA*
SINGH, RAJVINDER *Punjabi University, Patiala, India*
SKOPP, GISELA *Ruprecht-Karls University, Heidelberg, Germany*
SMARTY, SYLVESTER *Case Western University/University Hospitals of Cleveland, Cleveland, OH, USA*
SMITH, ANN C. *Columbia City, Indiana, IN, USA*
SMITH, DELANEY M. *Twin Valley Behavioral Healthcare, Columbus, OH, USA*
SMYTH, LARRY D. *Red Toad Road Company, Havre de Grace, MD, USA*
SORRENTINO, RENEE *Institute for Sexual Wellness, Quincy, MA, USA*
SPIEGEL, DAVID *Stanford University School of Medicine, Stanford, CA, USA*
SQUIER, WANEY *John Radcliffe Hospital, Oxford, UK*
STANKOWSKI, JOY E. *Case Western Reserve University, Cleveland, OH, USA*
STAUFFER, ERIC *University of Lausanne, Lausanne, Switzerland*
STEINBERG, LAURENCE *Temple University, Philadelphia, PA, USA*
STRUB, DIANE S. *Simon Fraser University, Burnaby, British Columbia, Canada*
STUDEBAKER, CHRISTINA *ThemeVision LLC, Indianapolis, IN, USA*

TAKATORI, TAKEHIKO *National Research Institute of Police Science, Chiba, Japan*
TARONI, FRANCO *The University of Lausanne, Lausanne, Switzerland*
TEPLIN, LINDA A. *Northwestern University Feinberg School of Medicine, Chicago, IL, USA*
THAKAR, MUKESH KUMAR *Punjabi University, Patiala, India*
THEAN, A. *Netherlands Organisation for Applied Scientific Research (TNO), Delft, The Netherlands*
THOMAS, TRACY A. *West Virginia University, Morgantown, WV, USA*
THOMPSON, CHRISTOPHER *University of California Los Angeles School of Medicine, Los Angeles, CA, USA*
THOMPSON, WILLIAM C. *University of California, Irvine, CA, USA*
THURMAN, JAMES T. *Eastern Kentucky University, Richmond, KY, USA*

TOGLIA, MICHAEL P. *University of North Florida, Jacksonville, FL, USA*
TRAMONTANA, MICHAEL G. *Vanderbilt University Medical Center, Nashville, TN, USA*
TRIDICO, SILVANA R. *Australian Federal Police, Canberra, ACT, Australia*
TULLY, GILLIAN *Forensic Science Service, Birmingham, UK*
TURILLAZZI, EMANUELA *University of Foggia, Foggia, Italy*
TURNER, BARRY *University of Lincoln, Lincoln, UK*

VAN WAALWIJK VAN DOORN, K. *Amsterdam-Amstelland Police, Amsterdam, The Netherlands*
VERSTRAETE, ALAIN G. *Ghent University Hospital, Ghent, Belgium*
VINER, MARK D. *Cranfield Forensic Institute, Defence Academy of the United Kingdom, Shrivenham and St Bartholomew's and the Royal London Hospitals, London, UK*
VITACCO, MICHAEL J. *Mendota Mental Health Institute, Madison, WI, USA*
VUORI, ERKKI *University of Helsinki, Helsinki, Finland*

WALKER, JAMES S. *Vanderbilt University School of Medicine, Nashville, TN, USA*
WALSH, SIMON J. *Australian Federal Police, Canberra, ACT, Australia*
WASHBURN, JASON J. *Northwestern University Feinberg School of Medicine, Chicago, IL, USA*
WEINSTOCK, ROBERT *University of California, Los Angeles, CA, USA*
WEIR, BRUCE S. *University of Washington, Seattle, WA, USA*
WELLS, GARY L. *Iowa State University, Ames, IA, USA*
WENGER, ERIC *Australian Federal Police, Weston and Australian Chemical, Biological Radiological and Nuclear Data Centre, Canberra, ACT, Australia*
WEST, SARA G. *University Hospital Case Medical Center, Cleveland, OH, USA*
WETTON, JON *Forensic Science Service, Birmingham, UK*
WEYERMANN, CÉLINE *University of Lausanne, Lausanne, Switzerland*
WHEATE, RHONDA M. *The Forensic Institute, Glasgow, UK*
WHELPTON, ROBIN *University of London, London, UK*
WILSON, CATHERINE M. *Simon Fraser University, Burnaby, British Columbia, Canada*
WONG, STEVEN H.Y. *Medical College of Wisconsin and Milwaukee County Medical Examiner's Office, Milwaukee, WI, USA*

YANG, SUZANNE *University of Pittsburgh School of Medicine, Pittsburgh, PA, USA*
YORK, CATHERINE *The University of Illinois at Chicago, Chicago, IL, USA*

ZAJAC, RACHEL *University of Otago, Dunedin, New Zealand*
ZEICHNER, ARIE *Hebrew University of Jerusalem, Jerusalem, Israel*
ZOUN, RIKKERT *Netherlands Forensic Institute, The Hague, The Netherlands*

Foreword

Forensic scientists who are attempting to provide guidance that assists investigators and solves crimes are well aware of the capabilities of the various disciplines in which they toil. Even more importantly, they are also cognizant of what their chosen specialty cannot (yet) accomplish. With a carefully nurtured and expanded knowledge base of the strengths and weaknesses of science, experts are often key to solving perplexing and high-profile cases. Their accomplishments on those occasions draw widespread attention from the media but, even more importantly, these experts also labor quietly on a daily basis to solve the more common cases that represent the bulk of their laboratory efforts and analyses. The consumers of forensic science – be they courts, legislative bodies, or regulatory agencies – rely on the solutions and ideas which these dedicated researchers strive to supply.

In the past few decades, there have been many new discoveries and advances in scientific disciplines, enabling forensic science specialists to apply an ever increasing depth of expertise in carrying out their tasks. Cases are being solved today using techniques that were unheard of twenty, ten, five, or even one year ago. It seems as if every issue of a scientific or technical journal that is published reports new innovations that are progressing from the experimental to the practical. In that environment, the *Wiley Encyclopedia of Forensic Science* is no doubt destined to be recognized as the premiere compendium of knowledge. After all, the entries in its volumes have been compiled by recognized, internationally known, and respected experts in every field.

There is no denying the popularity of television programs such as *CSI*, *Cold Case*, *Forensic Files*, and others. Indeed, there is perhaps no topic of greater interest in today's cyber world than the use of forensics to solve crimes. The small (and not-so-small) screen of the home theater also allows people to view a plethora of movies dealing with scientific ways to demystify complex scenarios. Prominent actors are cast in the white coats of laboratory examiners and interact with other participants in an effort to convince us of their prowess.

Through television, the movies, and an ever growing number of Internet sites, the public receives an abundance of information about how crime laboratory science assists in solving mysteries. The so-called "CSI effect" influences jurors, and even some judges, who have come to expect real life to mirror what they see on television, the movies, or what they read in other media. Unfortunately, not all of the information purveyed via these mediums comports with reality. Expectations are inflated beyond the possible. Even the most accomplished expert cannot solve cases in one hour or less as their television counterparts suggest *they* can!

Expectations in the legal profession, the courts, those who serve on juries, and the public at large demand that those who deal with the judicial system be familiar and are knowledgeable about what is possible and what is perhaps coming in the future. Experts, and anyone else who seeks knowledge about forensics, must not only be familiar with the well-established foundations of scientific proof, but must also stay reliably informed about new developments in the various forensic science disciplines. Here again, the *Wiley Encyclopedia of Forensic Science* serves as a complete, accurate, realistic, and up-to-date resource that will sate the knowledge thirst of experts as well as that of consumers of forensic science.

The framework for this five-volume Wiley encyclopedia was designed and compiled by its two Editors-in-Chief, Allan Jamieson of the United Kingdom and Andre Moenssens of the United States, both pre-eminent authorities in forensic science and its practical and legal applications. They are intimately familiar with the formidable strengths of forensic evidence as well as its perhaps less-known weaknesses. In their daunting task to seek a balance reporting on topics from A to Z, all of which straddle different levels of achievement, the Editors-in-Chief were ably assisted by a selection of knowledgeable subject-specialist topic editors and hundreds of contributors from all over the world. The cooperation and efforts of these many leaders in the scientific enterprise ensures that information dispensed in each discipline is reliable, relevant to real-life problems, and useful to a broad audience.

There are over 370 articles written by more than 330 contributors. These articles reference the myriad subjects of forensic science and embrace the physical, biological, behavioral, as well as comparative sciences. Laboratory management and case investigative techniques, laboratory support mechanisms, quality control programs, and discussions on the modern trends in interpreting the confidence level accorded test results by reference to statistical likelihood ratios are also dealt with.

Concerns about the desired exactness of science, and the inexactness of the law, also required that precedent-setting court decisions and discussions of legal principles that impact on a forensic expert's performance be explained and analyzed in these volumes. Issues of examiner education, training, and accreditation in various disciplines, as well as the ethical requirements governing scientists' professional behavior and the problem of expert witness malpractice have not been neglected.

For the layperson, the Encyclopedia provides authoritative answers to most questions about specific forensic problems. For the practitioners in specific fields, it reminds them of the basic fundamentals in their own discipline while it also informs them of the totality of knowledge in other fields so as to better understand the system as a whole and its interrelated complexities. References to published data and source materials appended to most articles allow interested persons to study areas of special concern in greater depth. Truly a collection of useful information on most of what is known as "forensic science", the Encyclopedia provides ready answers for everyone who is either involved or simply interested in forensic science.

In assessing how well John Wiley & Sons and all those associated with the production and publication of these volumes have achieved their objectives, there is no doubt the verdict and judgment will be favorable.

The Honorable Haskell M. Pitluck
Past President, American Academy of Forensic Science
Retired Circuit Judge, State of Illinois
April 2009

Preface

Forensic science is a broad label. It covers the entire complement of human industry and experience at the point where they interface with legal and legislative processes. This interaction may involve civil as well as criminal concerns and often takes the form of expert testimony offered to answer questions posed by a variety of human institutions. In the exercise of that function, the opinions of experts may vary widely in probative value, weight, and persuasiveness. But forensic science also serves therapeutic and human aspects apart from problem solving, and is often a crucial component in the formulation of policy and the passage of regulatory endeavors destined to insure the health and safety of populations.

The advent of DNA profiling has heralded a revolution in forensic science that goes well beyond forensic biology. The techniques and principles used in evaluating forensic scientific evidence are being closely examined at this moment in the light of the work in DNA as an individualization process – considered somewhat as a Holy Grail for forensic science. The ability to develop population databases, based on the easy numerical nature of DNA profiles, enabled the introduction of other approaches to the evaluation of evidence. The scientific rigor of some aspects of this process has posed difficult-to-answer questions for those forensic disciplines that emerged primarily from experience-based endeavors but which, through use, had become accepted in criminal justice systems. The debate as to the response within many of these specialties is ongoing.

This state of events has brought us to an era of unprecedented development and debate, as well as a measure of uncertainty, in some forensic science functions. Several disciplines are now broadly divided into "old school" and "new school" practitioners. It is not easy to predict which school will emerge as the dominant thinking in these disciplines, but it is almost certain that the future will see more scientifically robust techniques of evidence evaluation gaining widespread acceptance.

Many of the entries in this work may represent the seeds of that future. The existence of these competing schools inevitably creates differences of opinion regarding the science. The constant and varied evolution of science means that the state of the art in one discipline (in instrumentation, analytical accuracy, or evaluation) may be very different from the state of the art in another.

Although science may be regarded as international, law is not. Jurisdictional differences also inevitably create differences in the practice of forensic science. We have encouraged authors to be candid, but to represent those differences in their writing, and occasionally different authors will discuss the same or similar topics while offering a slightly different approach. We regard this as healthy and a natural part of the scientific discourse. As such, the debate is encouraged. However, the consequence is that the views expressed within articles are not necessarily the definitive or final word on any topic, nor do they necessarily represent the view of any other author or of the Editors-in-Chief.

In bringing together such a large and varied selection of experts in one major work, we were cognizant that imposing a "house style" would be close to impossible. Our limit seemed to be the adoption of American-English spelling and a single reference style. Inevitably, our

workload and a variety of deeply ingrained localized or national approaches will have allowed some departures from the standard to slip through the editorial net. We hope that this does not detract from the content. We did not enforce any particular writing style and the varied entries reflect that.

One thing that all of our authors had to be aware of was that this is a very unusual major technical reference work in that we aimed to inform an audience that includes people who are not forensic science professionals and who may simply be interested in selected topics discussed in these volumes, such as writers, reporters, educators, but who may have only a sketchy knowledge of the core principles or language of science. The work also purports to inform the legal profession, the judiciary, and paralegals. With these different audiences in mind, we have attempted to maintain sufficient depth to provide a valuable reference source for practitioners and academics as well. Only use will establish the degree to which we have been successful in achieving those disparate aims.

The complexity of providing information to a wide potential readership, composed of such a variety of interested parties, placed important choices on the editors in terms of its coverage. No one source can serve all of humanity, and thus the editors sometimes had to make painful judgments on what to include and what to pass over. Sometimes the choices on what to include were dictated in part by the availability of experts willing to share their professional experience with our readers. At other times, choices related simply to the fact that the dividing line between science and pseudoscience required a decision as to whether conclusions reached in a particular field provided sufficient guarantees of trustworthiness and reliability. This in no way can be taken to mean that the lack of appearance here invalidates any particular discipline, nor for that matter that inclusion validates it! Whatever the choice, the editors acted in their best judgment and will remain vigilant so as to select, for later inclusions, those emerging fields currently perhaps on the fringes of forensic science that increase their underlying knowledge-based data and gain a modicum of acceptance in the broader forensic science profession.

No work of this nature is perfect. This will not prevent us from striving to improve it online and in subsequent editions. We welcome feedback on any aspect of this work, including topics that, almost certainly, we have missed in our attempts to be all-inclusive.

We are only part of a very large and skilled team including our Editorial Board and the team at Wiley. Our heartfelt thanks go to all of them and in advance to you, the reader, for your feedback which will assist us to provide better resources for your future work.

Allan Jamieson and Andre Moenssens
April 2009

Abbreviations and Acronyms

1,4-BD	1,4-butanediol
16 PF	16 Personality Factor
2,4-DNT	2,4-dinitrotoluene
2,6-DNT	2,6-dinitrotoluene
5-HT	5-Hydroxytryptamine
5HIAA	5-Hydroxyindole Acetic Acid
5HTOL	5-hydroxytryptophol
6-AM	Δ^9-Tetrahydrocannabinol
6-AM	6-Acetylmorphine
AA	Greatest Angular Apertures
AACC	American Association of Clinical Chemists
AAFS	American Academy of Forensic Sciences
AAIDD	American Association on Intellectual and Developmental Disabilities
AAMR	American Association on Mental Retardation
AAPA	American Association of Physical Anthropology
AAPI	Adult–Adolescent Parenting Inventory
AAS	Atomic Absorption Spectrophotometry
AAS	Atomic Absorption Spectroscopy
ABAS-II	Adaptive Behavior Assessment System – Second Edition
ABC	American Board of Criminalistics
ABFDE	American Board of Forensic Document Examiners
ABFE	American Board of Forensic Entomology
ABFM	American Board of Forensic Medicine
ABFO	American Board of Forensic Odontology
ABFP	American Board of Forensic Psychology
ABFT	American Board of Forensic Toxicology
ABI	Applied Biosystems
ABO	ABO Blood Groups
ABPN	American Board of Psychiatry and Neurology
ABPP	American Board of Professional Psychology
ABTS	2, 2′-azino-di-(3-Ethyl-Benzthiazolinesulfonate)
ACE-V	Analysis, Comparison, Evaluation and Verification
ACFE	American College of Forensic Examiners
ACh	Acetylcholine
ACP1	Acid Phosphatase 1
ACPO	Association of Chief Police Officers
ADA	Adenosine Deaminase
ADD	Accumulated Degree-Days

ADH	Accumulated Degree-Hours
ADH	Alcohol Dehydrogenase
ADHD	Attention Deficit Hyperactive Disorder
ADM	Alcohol, Drug, and Mental
ADP	Adenosine Diphosphate
AEDs	Antiepileptic Drugs
AEME	Anhydroecgonine Methylester
AES	Auger Electron Spectroscopy
AF	Acid Fuchsin
AFE	Amniotic Fluid Embolism
AFIS	Automated Fingerprint Identification System
AFM	Atomic Force Microscopy
AFR	Association of Forensic Radiographers
AFTE	Association of Firearm and Tool Mark Examiners
AHG	Antihuman Globulin
AIDS	Acquired Immunodeficiency Syndrome
AIMs	Ancestry Informative Markers
AIP	Acute Interstitial Pneumonitis
AK	Adenylate Kinase
AKA	Alcoholic Ketoacidosis
AKD	Alkyl Ketene Dimer
ALDH	Aldehyde Dehydrogenase
ALFPs	Amplified Fragment Length Polymorphisms
ALI	American Law Institute
ALS	Alternate Light Sources
ALT	Alanine Aminotransaminases
ALTEs	Acute Life-Threatening Events
AM	Antemortem
AMDIS	Automated Mass Spectral Deconvolution and Identification System
AMI	Acute Myocardial Infarction
AMP	Adenosine Monophosphate
AmpFLPs	Amplified Fragment Length Polymorphisms
AMPS	Advanced Mobile Phone System
AMS	Accelerator Mass Spectrometry
AN	Ammonium Nitrate
ANFO	Ammonium Nitrate and Fuel Oil
ANN	Artificial Neural Network
ANSI	American National Standards Institute
AP	Acid Phosphatase
AP	Ammonium Perchlorate
APA	American Psychological Association
APA	American Psychiatric Association
APCI	Atmospheric Pressure Chemical Ionization
APD	Antisocial Personality Disorder
APDS	Autonomous Pathogen Detection System
APHL	Association of Public Health Laboratories
API	Application Programming Interface
API	Atmospheric Pressure Ionization
AP-LS	American Psychology-Law Society

APPITA	Australian Pulp and Paper Industry Technical Association
APS	Adult Protective Services
ARDS	Acute Respiratory Distress Syndrome
ARIs	Actuarial Risk Assessment Instruments
ARVD	Arrhythmogenic Right Ventricular Disease
AS	Autonomous System
ASA	Acetylsalicylic Acid
ASA	Alkenyl Succinyl Anhydride
ASCH	American Society of Clinical Hypnosis
ASCLD	American Society of Crime Laboratory Directors
ASCLD/LAB	American Society of Crime Laboratory Directors/Laboratory Accreditation Board
ASD	Acute Stress Disorder
ASM	American Society of Metals
ASME	American Society of Mechanical Engineer
ASQDE	American Society of Questioned Document Examiners
ASRIS	Australian Soil Resources Information System
AST	Aspartate Aminotransaminases
ASTM	American Society for Testing and Materials
ATF	Alcohol, Tobacco, and Firearms
ATP	Adenosine Triphosphate
ATR	Attenuated Total Reflection
ATR-FTIR	Attenuated Total Reflectance-Fourier Transform Infrared
ATSA	Association for the Treatment of Sexual Abusers
ATV	Atmospheric Pressure Chemical Ionization
AUC	Area Under Curve
BAC	Blood Alcohol Concentration
BAPP	Beta Amyloid Precursor Protein
BARTS	Biological Agent Real-Time Sensor
BAWS	Biological Agent Warning Sensor
BCCH	Broadcast Control Channel
BCTMP	Bleached Chemi-Thermomechanical Hardwood Pulps
BE	Benzoylecgonine
BEECS	Benign Enlargement of the Extracerebral Spaces
BFRL	Building Fire and Research Laboratory
BGA	Ball Grid Array
BHT	Butylated Hydroxytoluene
BISFA	The International Bureau for the Standardization of Man-Made Fibres
BKV	BK Virus
BLEVE	Boiling Liquid and Expanding Vapor Explosion
BMI	Body Mass Index
BNs	Bayesian Networks
BP	Bandpass
bp	Base Pair
BPIF	Bandpass Interference Filters
BPS	Bricklin Perceptual Scales
BrAC	Breath-Alcohol Concentration
BS	Beam Splitter

BSA	Bovine Sera Albumin
BSDL	Boundary-Scan Description Language
BSE	Back-Scattered Electrons
BSE	Black Sheep Effect
BTB	Sickle Cell Anemia
BTEX	Benzene, Toluene, Ethylbenzene, and Xylene
BTU	British Thermal Unit
BWC	Biological and Toxins Weapons Convention
BWS	Battered Woman Syndrome
CA	Carbonic Anhydrase
CA	Cytosine-Adenine
CABs	Conformity Assessment Bodies
CABL	Compositional Analysis of Bullet Lead
CAC	California Association of Criminalistics
CAC	Child Advocacy Center
CAD	Computer-Aided Design
CAD	Coronary Artery Disease
CAF	Cyanoacrylate Fuming
CAFSS	Centre for Australian Forensic Soil Science
CAGE	Computer Aided Glass Evaluation
CAI	Case Assessment and Interpretation
CAI	Competence Assessment Instrument for Standing Trial
CAM	Computer-Aided Modeling
CAN	Cardiovascular Autonomic Neuropathy
CAP	College of American Pathologists
CAPI	Child Abuse Potential Inventory
CAPTA	Child Abuse Prevention and Treatment Act
carboxy-THC	11-Nor-9-carboxytetrahydrocannabinol
CAST/MR	Competence Assessment for Standing Trial for Defendants with Mental Retardation
CAT	Computed Axial Tomography
CBD	Cannabidiol
CBN	Cannabinol
CBRN	Chemical, Biological, Radiological, and Nuclear
CCBEMWG	Common Criteria Biometric Evaluation Methodology Working Group
CCD	Charge Coupled Device
CCSA	Certified Crime Scene Analyst
CCSI	Certified Crime Scene Investigator
CCTV	Closed Circuit Television
CD	Conduct Disorder
CD	Cyclodextrin
CDC	Centers for Disease Control
CDER	Center for Drug Evaluation and Research
CDMA	Frequency Division Multiple Access
CDR	Cartridge Discharge Residues
CDR	Crash Data Recorders
CDT	Carbohydrate Deficient Transferrin
CE	Capillary Electrophoresis

CEDIA	Cloned Enzyme Donor Immunoassay
CEIR	Central Equipment Identify Register
CESB	Council of Engineering and Scientific Specialty Boards
CF	Compact Flash
CF	Corrective Factors
CFA	Confirmatory Factor Analysis
CFAST	Consolidated Model of Fire Growth and Smoke Transport
CFC	Chlorofluorocarbon
CFD	Computational Fluid Dynamics
CGS	Crow–Glassman Scale
CHD	Coronary Heart Disease
CHF	Congestive Heart Failure
CHINS	Children in Need of Supervision
CI	Chemical Ionization
CI	Cognitive Interview
CI-MS	Chemical Ionization Mass Spectrometry
CIFA	Centre for International Forensic Assistance
CIL	Central Identification Laboratory
CIP	Commission Internationale Permanente
CIS	Canadian Information Society
CITES	The Convention on International Trade in Endangered Species Fauna and Flora
CIT	Concealed Information Test
CJA	Criminal Justice Act
CJP	Capital Jury Project
CL	Cathodoluminescence
CMOS	Complementary Metal-Oxide Semiconductor
CMR-R	Comprehension of Miranda Rights-Recognition
CMS	Consecutive Matching Striae
CMYK	Cyan, Magenta, Yellow, and Black
CN	Cyanide
CNS	Central Nervous System
CO	Carbon Monoxide
CO-Hb	Carbon Monoxide Hemoglobin
CODIS	Combined Offender DNA Index System
COHb	Carboxyhemoglobin
COPFS	Crown Office Procurator Fiscal Service
COVR	Classification of Violence Risk
CPA	Cyproterone Acetate
CPD	Continuing Professional Development
CPE	Combined Power of Exclusion
CPF	Cardiac Fibroelastoma
CPGR	Chlorophenolred-β-Galactoside
CPI	Combined Probability of Inclusion
CPIA	Criminal Procedure and Investigations Act
CPK	Creatine Phosphokinase
CPLE	Certified Latent-Print Examiner
CPR	Cardiopulmonary Rescuscitation
CPR	Chlorophenolred
CPR	Civil Procedure Rules

CPS	Child Protective Services
CPSC	Consumer Product Safety Commission
CPVT	Catecholaminergic Polymorphic Ventricular Tachycardia
CQT	Comparison Question Test
CR	Conditioned Response
CRFP	Council for Registration of Forensic Practitioners
CRT	Cathode-Ray Tube
CS	Conditioned Stimuli
CSA	Child Sexual Abuse
CSA	Crime Scene Analyst
CSAAS	Child Sexual Abuse Accommodation Syndrome
CSC	Crime Scene Coordinator
CSCSA	Certified Senior Crime Scene Analyst
CSD	Circuit Switched Data
CSE	Crime Scene Examiner
CSF	Cerebrospinal Fluid
CSFS	Canadian Society of Forensic Science
CSI	Consensual Sexual Intercourse
CSI	Crime Scene Investigators
CSM	Crime Scene Manager
CT	Computed Tomography
CTAB	Hexadecyltrimethlyammonium Bromide
CTS	Collaborative Testing Services
CV	Coefficient of Variation
CV	Curriculum Vita
CVD	Cardiovascular Disease
CVFI	Candidate for the Vehicle Fire Investigator
CW	Chemical Warfare
CWAs	Chemical Warfare Agents
CWC	Chemical Weapons Convention
CYP	Cytochrome P
CYP 2B6	Cytochrome P450 2B6
CYP 2D6	Cytochrome P450 2D6
CYP 3A5	Cytochrome P450 3A5
CYP3A	Cytochrome P450 3A
CZE	Capillary Zone Electrophoresis
D1T2	Direct Thermal Transfer
D2T2	Dye Diffusion Thermal Transfer
D-AMPS	Digital Advanced Mobile Phone System
DA	Dopamine
DAB	Diaminobenzidine
DAB	DNA Advisory Board
DAD	Diffuse Alveolar Damage
DAD	Diode-Array Detector
DAD	Drowning-Associated Diatoms
DAD	Photodiode Array Detection
DADP	Diacetone Diperoxide
DAG	Directed Acyclic Graph
DAI	Diffuse Axonal Injury

DART	Direct Analysis in Real Time
DEA	Drug Enforcement Agency
DECT	Digital Enhanced Cordless Telecommunication
DEM	Digital Elevation Model
DESNOS	Disorder of Extreme Stress not Otherwise Specified
DF	Dedicated File
DFC	Drug-Facilitated Crime
DFO	1,8-diaza-9-fluorenone
DFSA	Drug-Facilitated Sexual Assault
DHCP	Dynamic Host Configuration Protocol
DIC	Differential Interference Contrast
DID	Dissociative Identity Disorder
DIN	Deutsches Institut Für Normung
DIS-IV	Diagnostic Interview Schedule, Version IV
DLC	Diagnostic Link Connector
DM	Diabetes Mellitus
DM	Dichroic Mirror
DMA	Dimethoxyamphetamine
DNA	Deoxyribonucleic Acid
DNS	Domain Name System
DO	Dangerous Offender
DOB	4-Bromo-2,5-dimethoxyamphetamine
DOET	Dimethoxyethylamphetamine
DOM	4-Methyl-2,5-dimethoxyamphetamine
DOP	Degenerate Oligonucleotide Primed
DOS	Denial-of-Service
DOT	Department of Transportation
DPA	Diphenylamine
DPI	Dots Per Inch
DPS	Department of Public Safety
DRE	Drug Recognition Evaluation
DRIFT	Diffuse Reflectance Infrared Fourier Transform
DRM	Deese–Roediger–McDermott
DSC	Differential Scanning Calorimeter
DSM	Diagnostic and Statistical Manual of Mental Disorders
DSM-III	Diagnostic and Statistical Manual of Mental Disorders, Third Edition
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition
DSM-IV-TR	Diagnostic and Statistical Manual, Fourth Edition, Text Revision
DSPD	Dangerous Severe Personality Disorder
DTA	Differential Thermal Analysis
DTCs	Diagnostic Trouble Codes
DTGS	Deuterated Triglycine Sulfate
DTO	Dithiooxamide
DTT	Dithiothreitol
DUID	Driving Under the Influence of Drugs
DVI	Disaster Victim Identification
DVT	Deep Venous/Vein Thrombosis
DWI	Driving While Intoxicated

EA	Enzyme Acceptor
EA	European Co-Operation for Accreditation
EAAF	Equipo Argentino De Antropología Forense
EAFE	European Association for Forensic Entomology
EAP	Erythrocyte Acid Phosphatase
EBV	Epstein Barr Virus
EC	Ethyl Centralite
ECA	Epidemiologic Catchment Area
ECD	Electron Capture Detector
ECDS	Empirical Criteria for the Determination of Suicide
ECF	Elemental Chlorine Free
ECHR	European Court of Human Rights
ECLM	European Council of Legal Medicine
ECT	Electroconvulsive Therapy
ED	Enzyme Donor
EDD	Electrostatic Detection Device
EDDP	2-Ethylidene-1,5-Dimethyl-3,3-Diphenylpyrrolidine
EDM	Electrical Discharge Machining
EDMI	Electromyography
EDNAP	European DNA Profiling Group
EDS	Energy-Dispersive Spectrometer
EDS or EDX	Energy-Dispersive X-Ray
EDTA	Ethylene Diamine Tetraacetic Acid
EDX	Energy Dispersive X-Ray
EDXA	Energy Dispersive X-Ray Analysis
EEG	Electro-Encephalogram
EER	Equal Error Rate
EF	Elementary File
EFA	Exploratory Factor Analysis
EFG	European Fibres Group
EGA	Estimated Gestational Age
EGDN	Ethylene Glycol Dinitrate
E-HMM	Ergodic Hidden Markov Models
EI	Electron Impact
EI-MS	Electron Impact Positive Ion Mass Spectrometry
EIA	Environmental Impact Assessment
EIA	Enzyme Immunoassay
EIC	Extracted Ion Chromatogram
EIP	Extracted Ion Profiles
ELISA	Enzyme Linked Immunosorbent Assay
EME	Ecgonine Methyl Ester
EMG	Electromyography
EMIT	Enzyme Multiplied Immunoassay Technique
EMPOP	European DNA Profiling Group MtDNA Population Database
EMS	Enhanced Messaging Service
ENAA	Epithermal Neutron Activation Analysis
ENFSI	European Network of Forensic Science Institutes
EPA	Environmental Protection Agency
EPG	Electropherogram
EPI	Enhanced Product Ion

EPO	Erythropoetin
EPS	Extra-Pyramidal Symptoms
EQA	External Quality Assessment
ERPs	Event-Related Potentials
ESD	Environmental Secondary Detector
EsD	Esterase D
ESDA	Electrostatic Detection Apparatus
ESEM	Environmental SEM
ESI	Electronically Stored Information
ESI	Electrospray Ionization
ESLA	Electrostatic Lifting Apparatus
ESR	Environmental Science Research Limited
EtG	Ethyl Glucuronide
ETK	Explosive Test Kit
EtS	Ethyl Sulfate
EU	European Union
EUCAP	European Collection of Automotive Paint
EWG	Expert Working Group
EX	Exciter Filter
FAAS	Flameless Atomic Absorption Spectroscopy
FABMS	Fast Atom Bombardment Mass Spectrometry
FAEE	Fatty Acid Ethyl Esters
FAME	Fatty Acid Methyl Esters
FAR	False Acceptance Rate
FASE	Forensic Anthropology Society of Europe
FBI	Federal Bureau of Investigation
FDA	Food and Drug Administration
FDE	Forensic Document Examiner
FDR	Firearm Discharge Residue
FDS	Fire Dynamics Simulator
FDS	Fragment Data System
FE	Field Emission
FEC	Forensic Engineering Curriculum
FEPAC	Forensic Science Education Programs Accreditation Commission
FFT	Fast Fourier Transform
FGH	S-Formylglutathione Hydrolase
FHE	Forensic Handwriting Examiner
FIB	Focused Ion Beam
FID	Flame Ionization Detector
FINEX	Forensic International Network of Explosive Examiners
FIRS	Forensic Information Retrieval System
FISH	Fluorescent In Situ Hybridization
FLE	Filtered Light Examination
FLIM	Fluorescence Lifetime Imaging
FLO	Family Liaison Officer
FLS	Forensic Light Source
FMEA	Failure Modes and Effects Analysis
FMJ	Full Metal Jacket

fMRI	Functional Magnetic Resonance Imaging
FMSF	False Memory Syndrome Foundation
FNAA	Fast Neutron Activation Analysis
FOMA	Freedom of Mobile Multimedia Access
FPAC	Forensic Pathology Advisory Commitee
fpc	Finite Population Correction
FPD	Flame Photometric Detection
FPIA	Fluorescence Polarization Immunoassay
FPM	First-Pass Metabolism
FQS-I	Forensic Quality Services-International
FRE	Federal Rules of Evidence
FRI	Function of Rights in Interrogation
FRR	False Rejection Rate
FSAB	Forensic Specialties Accreditation Board
FSF	Forensic Sciences Foundation
FSH	Follicle-Stimulating Hormone
FSS	Forensic Science Society
FSS	Forensic Science Service Ltd
FSSoc	Forensic Science Society (UK)
FTA	Fault Tree Analysis
FTD	Flame Thermoionic Detector
FTIR	Frustrated Total Internal Reflection
FTL	Flash Translation Layer
FTP	File Transfer Protocol
FUT	Fucosyltransferase
FWA	Fluorescent Whitening Agent
G6PDH	Glucose-6-phosphate Dehydrogenase
GABA	Gamma Amino-Butyric Acid
GBL	Gamma Butyrolactone
GC	Gas Chromatography
GC-ECD	Gas Chromatography Electron Capture Detector
GC-FID	Gas Chromatography Flame Ionization Detector
GC-MS	Gas Chromatography-Mass Spectrometry
GC-MS/DFPD	Gas Chromatography-Mass Spectroscopy/Dual Flame Photometric Detection
GCS	Glascow Coma Scale
GD–MS	Glow Discharge–Mass Spectrometry
GERD	Gastroesophageal Reflux Disease
GFAAS	Graphite Furnace Atomic Absorption Spectrophotometry
GGT	Gamma-Glutamyltransferase
GH	Growth Hormone
GHB	γ-Hydroxybutyrate
GIS	Geographic Information System
GKT	Guilty Knowledge Test
GLO	Glyoxylase I
GMM	Gaussian Mixture Models
GMSC	Gateway Mobile Switching Center
GNDS	General Neuropsychological Deficit Scale
GnRH	Gonadotropin-Releasing Hormone

GPA	Grade Point Average
GPB	Glycophorin B
GPR	Ground-Penetrating Radar
GPRS	General Packet Radio Service
GPS	Global Positioning System
GPT	Glutamate-Pyruvate Transaminase
GRC	General Rifling Characteristics
GSM	Global System for Mobile
gsm	Grams Per Square Meter
GSR	Gunshot Residue
GSS	Gudjonsson Suggestibility Scales
GUI	Graphical User Interface
GuSCN	Guanidinium Thiocyanate
GUS	General Unknown Screening
Hb	Heterozygosity Balance
HBFP	Haematoxylin-Basic Fuchsin-Picric Acid
HCM	Hypertrophic Cardiomyopathy
HCN	Hydrogen Cyanide
HCR	Historical Clinical Risk
HELIN	Higher Education Library Information Network
HERG	Human Ether-a-go-go-Related Gene
HF	Human Factor
HFC	Hydrofluorocarbon
HFE	Human Factors Engineering
HFE	Hydrofluoroether
HGN	Horizontal Gaze Nystagmus
HHV-1	Human Herpes Virus Type 1
HIC	Head Injury Criterion
HIV AIDS	Human Immunodeficiency Virus Acquired Immune Deficiency Syndrome
HLA	Human Leucocyte Antigen
HLoQ	Higher Limit of Quantitation
HMTD	Hexamethylene Triperoxide Diamine
HMW	High-Molecular-Weight
HMX	Octogen
HPD	Heavy Petroleum Distillates
HPD	Highest Posterior Density
HPLC	High Performance Liquid Chromatography
HPLC-DAD	High Performance Liquid Chromatography Diode-Array Detector
HPLC-MS	High Performance Liquid Chromatography Mass Spectrometry
HPLC/PMDE	High Performance Liquid Chromatography with a Pendant Mercury Drop Electrode Detector
HPV	Human Papillomavirus
HRGS	High-Resolution Gamma Spectrometry
HRNB	Halstead–Reitan Neuropsychological Battery
HRNTB	Halstead–Reitan Neuropsychological Test Battery
HRR	Heat Release Rate
HSE	Health and Safety Executive

HTTP	Hypertext Transfer Protocol
HWE	Hardy–Weinberg Equilibrium
i.d.	Internal Diameter
I/IS	Insulin or Insulin Secretagogs
IAAC	Inter-American Accreditation Cooperation
IAAI	International Association of Arson Investigators
IABPA	International Association of Bloodstain Pattern Analysts
IACI	International Association for Craniofacial Identification
IAEA	International Atomic Energy Agency
IAFIS	Integrated Automated Fingerprint Identification System
IAFS	International Association of Forensic Scientists
IAI	International Association for Identification
ibd	Identical by Descent
IBG	International Biometrics Group
IBIS	Integrated Ballistics Identification System
IBS	Identical by State
IC	Ion Chromatography
ICC	International Criminal Court
ICC1	Intraclass Correlation Coefficient
ICCID	Integrated Circuit Card Identifier
ICD	International Classification of Diseases
ICDD	International Centre for Diffraction Data
ICF/DIC	Intravascular Coagulation and Fibrinolysis/Disseminated Intravascular Coagulation
ICN	Increased Cycle Number
ICP	Inductively Coupled Plasma
ICP-MS	Inductively Coupled Plasma Mass Spectrometry
ICP-OES	Inductively Coupled Plasma-Optical Emission Spectrometry
ICP/AES	Inductively Coupled Plasma Atomic Emission Spectroscopy
ICP/MS	Inductively Coupled Plasma Mass Spectrometry
ICTR	International Criminal Tribunal for Rwanda
ICTY	International Criminal Tribunal for the Former Yugoslavia
IDA	International Development Agencies
IDEIA	Individuals with Disabilities Education Improvement Act
iDEN	Integrated Digital-Enhanced Network
IDS	Intrusion Detection System
IEC	International Electrotechnical Commission
IED	Improvised Explosive Device
IEEE	Institute of Electrical and Electronic Engineers
IEEGFI	Interpol European Expert Group on Fingerprint Identification
IEF	Isoelectric Focusing
IFSTA	International Fire Service Training Association
IHL	International Humanitarian Law
IIFES	International Institute of Forensic Engineering Sciences
ILAC	International Laboratory Accreditation Cooperation
ILAE	International League against Epilepsy
ILR	Ignitable Liquid Residues
IMED	Indentation Materializer

IMEI/IMSI	International Mobile Equipment Identity/International Mobile Equipment Identity
IMMORTAL	Impaired Motorists, Methods of Roadside Testing and Assessment for Licensing
IMS	Ion Mobility Spectrometry
IMSI	International Mobile Equipment Identity
IMT-2000	International Mobile Telecommunications-2000
INAA	Instrumental Nuclear Activation Analysis
IND	1,2-Indanedione
IND	Improvised Nuclear Device
IND-Zn	Indanedione-Zinc
INFL	International Nuclear Forensic Laboratories
INH	Isoniazid
INS	International Neuropsychological Society
IOFOS	International Organisation for Forensic Odontostomotology
IP	Internet Protocol
IPEP	Improved Primer Extension Preamplification
IQ	Intelligence Quotient
IQC	Internal Quality Control
IR	Infrared
IRA	Irish Republican Army
IRC	Internet Relay Chat
IrDA	Infrared Data Association
IRL	Infrared Luminescence
IRMS	Isotope Ratio Mass Spectrometry
IRR	Infrared Reflectance
IS	Internal Standard
ISFG	International Society of Forensic Genetics
ISO	International Standards Organization
ISP	Internet Service Provider
ITU	International Telecommunication Union
ITWG	International Technical Working Group
JCAH	Joint Commission on Accreditation of Hospitals
JFFS2	Journalized Flash File System
JINS	Juveniles in Need of Supervision
JPAC	Joint Pow/MIA Accounting Command
JPEG2000	Joint Photographic Expert Group 2000
JTAG	Joint Test Action Group
KAAIT	Kaufman Adolescent and Adult Intelligence Test
KBS	Knowledge-Based System
KEBQ	Knowledge of Eyewitness Behavior Questionnaire
KIMS	Kinetic Interaction of Microparticles in Solution
KIPS	Keys to Interactive Parenting Scale
KM	Kastle Meyer test
KSA	Knowledge, Skills, and Abilities
LAB	Laboratory Accreditation Board
LAC	Location Area Code

LAMMA	Laser Microprobe Mass Analysis
LAMPA	Lysergic Acid Methyl Propyl Amide
LAN	Local Area Network
LBA	Logical Block Addressing
LC	Liquid Chromatography
LC/ESI/MS	Liquid Chromatography Electrospray Ionization Mass Spectrometry
LC-MS/MS	Liquid Chromatography Tandem Mass Spectrometry
LC/MS	Liquid Chromatography Mass Spectrometry
LCN	Low Copy Number
LCP	Life-Course Persistent
LCV	Leucocrystal Violet
LD	Lethal Dose
LDA	Linear Discriminant Analysis
LEAA	Law Enforcement Assistance Administration
LED	Light-Emitting Diode
LEL	Lower Explosive Limit
LH	Luteinizing Hormone
LHRH	Luteinizing Hormone-Releasing Hormone
LIBS	Laser-Induced Breakdown Spectroscopy
LLE	Liquid–Liquid Extraction
LLoQ	Lower Limit of Quantification
LMG	Leucomalachite Green
LNNB-CR	Luria–Nebraska Neuropsychological Battery-Children's Revision
LOC	Loss of Consciousness
LOD	Limit of Detection
LOQ	Limit of Quantification
LP	Liquefied Petroleum
LPC	Linear Predictive Coding
LPCC	Linear Prediction Cepstrum Coefficients
LPDs	Light Petroleum Distillates
LQTS	Long QT Syndrome
LR	Likelihood Ratio
LRs	Long Rifles
LSD	Lysergic Acid Diethylamide
LTDNA	Low Template DNA
LTO	Long-Term Offender
LVH	Left Ventricular Hypertrophy
M-FAST	Miller Forensic Assessment of Symptoms Test
M3G	Morphine-3-glucuronide
M6G	Morphine-6-glucuronide
MAC	Media Access Control
MacCAT-CA	MacArthur Competence Assessment Tool-Criminal Adjudication
MacCAT-T	MacArthur Competence and Assessment Tool-Treatment
MacSAC-CD	MacArthur Structured Assessment of the Competencies of Criminal Defendants
MALDI/TOF	Matrix-Assisted Laser Desorption/Ionization Time-of-Flight
MAM	Mechanical, Analytical, and Medical

MAM	Mono acetyl Morphine
MAO	Monoamine Oxidase
MB	Myocardial Bridging
MBD	4-(4-Methoxybenzylamino-7-Nitrobenzofurazan)
MBDB	*N*-Methyl-Benzodioxazoylbutanamine
MC-ICP/MS	Multicollector Inductively Coupled Plasma Mass Spectrometry
MC1R	Melanocortin 1 Receptor
MCMC	Monte Carlo Markov Chain
MCMI	Millon Clinical Multiaxial Inventory
MCQ	Multiple-Choice Questions
MCT	Mercury Cadmium Telluride
MCV	Mean Corpuscular Volume
MDA	3,4-Methylenedioxyamphetamine
MDA	Multiple Displacement Amplification
MDCT	Multislice Computed Tomography
MDE	3,4-Methylenedioxyamphetamine
MDEA	3,4-Methylenedioxy-*N*-Amphetamine
MDMA	3,4 Methylenedioxymethamphetamine
MDMA	Liquid Chromatography Coupled to Mass Spectrometry
MDMA	Methylenedioxymethamphetamine
MDMA	Methylenedioxymeth(yl)amphetamine
MDT	Multidisciplinary Team
ME	Medical Examiner
MECA	Methodology for Epidemiology of Mental Disorders in Children and Adolescents
MECC	Micellar Electrokinetic Chromatography
MECE	Micellar Electrokinetic Capillary Electrophoresis
MEOS	Microsomal Oxidizing System
MERMER	Memory and Encoding Related Multifaceted Electroencephalographic Response
MERS	Medical Error Reporting System
met-Hb	Methemoglobin
MFCC	Mel Frequency Cepstral Coefficients
MGT	Modified Griess Test
MHC	Major Histocompatibility Complex
MHL	Minimal Haplotype Loci
MI	Medullary Index
mIPEP	Modified Improved Primer Extension Preamplification
MIR	Mid-Infrared
MLE	Most Likely Estimate
MLP	Multilocus Profiling
MMC	Multi Media Card
MMDA	Methoxymethylenedioxyamphetamine
MMD	Multimetal Deposition
MMPI-2	Minnesota Multiphasic Personality Inventory 2
MMSD	Mass Memory Storage Device
MMSE	Mini Mental State Examination
MND	Malingering Neurocognitive Dysfunction
MO	Modus Operandi
MPA	Medroxyprogesterone

MPD	Medium Petroleum Distillate
MPD	Modified Physical Developer
MPD	Multiple Personality Disorder
MPS	Metropolitan Police Service
MP	Match Probability
MR	Metabolic Ratios
MRI	Magnetic Resonance Imagery
MRM	Multiple Reaction Monitoring
mRNA	Messenger Ribonucleic Acid
MRS	Magnetic Resonance Spectroscopy
MS	Mass Selective
MS	Mass Spectrometer
MSC	Mobile Switching Center
MSCT	Multislice Computed Tomography
MSD	Mass Selective Detector
MSE	Mental Status Examination
MSP	Microspectrophotometers
mtDNA	Mitochondrial Deoxyribonucleic Acid
MTPA	α-Methoxy-α-(Trifluoromethyl)Phenylacetic Acid
(*S*)-(+)-MTPACl	(*S*)-(+)-α-Methoxy-(Trifluoromethyl)Phenylacetyl Chloride
MTT	Dimethylthiazol 2 yl Diphenyltetrazolium Bromide
MVD	Multiwavelength Detection
MVN	Multivariate Normality
MW	Molecular Weight
NA	Numerical Aperture
NAA	Neutron Activation Analysis
NACB	National Academy of Clinical Biochemistry
NAD	Nicotine Adenine Dinucleotide
NADP	Nicontinamide Adenine Dinucleotide Phosphate
NADPH	Reduced Nicontinamide Adenine Dinucleotide Phosphate
NAE	Negligent Adverse Event
NAFE	National Academy of Forensic Engineers
NAFEA	North American Forensic Entomological Association
NAFI	National Association of Fire Investigators
NAGPRA	Native American Graves Protection and Repatriation Act
NAHI	Non-accidental Head Injury
NAME	National Association of Medical Examiners
NAPQI	*N*-Acetyl-*p*-benzoquinone Imine
NAS	National Academy of Sciences
NAS	Network Attached Storage
NATA	National Association of Testing Authorities
Native PAGE	Native Gel Electrophoresis
NBS	National Bureau of Standards
NC	Nitrocellulose
NCA	No Cause Apparent
NCANDS	National Child Abuse and Neglect Data System
NCIDD	National Criminal Identification DNA Database
NCIGC-MS	Negative Chemical Ionisation Gas Chromatograph with Mass Spectral Analyser

NCIS	National Coroners Information Service
NCJRS	National Criminal Justice Reference Service
NCNM	Noncorrosive and Nonmercuric
NCS	National Comorbidity Survey
NCSI	Nonconsensual Sexual Intercourse
NCSTL	National Clearinghouse for Science, Technology and the Law
NCVS	National Crime Victimization Survey
NDIS	National DNA Identification System
NDNAD	National DNA Database
nDNA	Nuclear DNA
NE	Norepinephrine
NEISS	National Electronic Injury Surveillance System
NEO-PI	Neo-Personality Inventory
NEO-PI-R	Neo-Personality Inventory-Revised
NFA	National Forensics Association
NFI	Netherlands Forensic Institute
NFIRS	National Fire Incident Reporting System
NFPA	National Fire Protection Association
NFSTC	National Forensic Science Technology Center
NFWFL	National Fish and Wildlife Forensics Laboratory
NG	Nitroglycerine
NGOs	Nongovernmental Organizations
NGRI	Not-Guilty-by-Reason-of-Insanity
NHTSA	National Highway Traffic Safety Administration
NIBIN	National Integrated Ballistic Information Network
NICHD	National Institute of Child Health and Human Development
NIFS	National Institute of Forensic Science
NIJ	National Institute of Justice
NIR	Near Infrared
NIST	National Institute of Standards and Technology
NLQ	Near Letter Quality
NMDA	*N*-Methyl-D-Aspartic Acid
NMR	Nuclear Magnetic Resonance
NMT	Nordic Mobile Telephones
NP	Neuropsychological
NPD	Nitrogen Phosphorus Detector
NPIA	National Policing Improvements Agency
NRC	National Research Council
NRY	Nonrecombining Region of the Y Chromosomes
NSAID	Nonsteroidal Anti-Inflammatory Drug
NSTC	National Science and Technology Council
NTT	Nippon Telegraph and Telephones
nuDNA	Nuclear Deoxyribonucleic Acid
NVFS	Nonvolatile File System
NY-OCME	New York Office of the Chief Medical Examiner
OBAs	Optical Brightening Agents
OCD	Obsessive Compulsive Disorder
OCDS	Operational Criteria for the Determination of Suicide
OD	Overdose

OEM	Original Equipment Manufacturer
OHS&W	Occupational Health Safety and Wellbeing
OLA	Oligonucleotide Ligation Assay
OOBNs	Object-Oriented Bayesian Networks
OPC	Organic Photoconductor
OPCW	Organization for the Prohibition of Chemical Weapons
OPGs	Orthopantomographs
ORO	Oil Red O
OSHA	Occupational Safety and Health Administration
OS	Operating Systems
OTA	Over the Air
OTC	Over-the-Counter
OVD	Optical Variable Device
P2P	Peer-to-Peer
PAC	Plasma Alcohol Concentration
PACE	Police and Evidence Act
PAE	Preventable Adverse Event
PAI	Personality Assessment Inventory
PAP	Prostatic Acid Phosphatase
PAR	Pseudoautosomal Region
PAS	Preliminary Alcohol Screening
PCA	Principal Component Analysis
PCB	Polychlorinated Biphenyl
PCB	Printed Circuit Board
PCC	Pyridinium Chlorochromate
PCDF	Polychlorinated Dibenzofurans
PCL	Psychopathy Checklist
PCL-R	Psychopathy Checklist-Revised
PCL:SV	Screening Version of the Hare Psychopathy Checklist-Revised
PCP	Phencyclidine
PCP	Primary Care Physician
PCPP	Phenyl Cyclopentyl Piperidine
PCR	Polymerase Chain Reaction
PCRI	Parent–Child Relationship Inventory
PCT	Procalcitonin Concentration
PD	Physical Developer
PDA	Personal Digital Assistant
PDC	Personal Digital Cellular
PDD	Psychophysiological Detection of Deception
PDM	Psychodynamic Diagnostic Manual
PDQ	Paint Data Query
PE	Pulmonary Embolism
Pep A	Peptidase A
PEP	Primer Extension Preamplification
PEP-PCR	Primer Extension Preamplification Polymerase Chain Reaction
PET	Positron Emission Tomography
PETN	Pentaerythritol Tetranitrate
PFA	Psychological First Aid
PGC–MS	Pyrolysis Gas Chromatography–Mass Spectrometry

PGD	6-Phosphogluconate Dehydrogenase
PGM	Phosphoglucomutase
PHA	Preliminary Hazard Analysis
PHA	Public Health Agency
PHP	Phenyl Cyclohexylpyrrolidine
PHT	Pulmonary Hypertension
PID	Photoionization Detector
PIN	Personal Identifying Number
PINS	Persons in Need of Supervision
PLM	Polarized Light Microscope
PLMN	Public Land Mobile Network
PLS	Partial Least-Squares
PM	Postmortem
PMA	*p*-Methoxy-Amphetamine
PMDD	Premenstrual Dysphoric Disorder
PMI	Postmortem Interval
PML	Progressive Multifocal Leukoencephalopathy
PMR	Postmortem Redistribution
PMS	Phenazine Methosulphate
PMS	Premenstrual Syndrome
PMT	Photo Multiplier Tube
PMT	Premenstrual Tension
PNE	Pediatric Neurological Exam
PNES	Psychogenic Nonepileptic Seizures
POCT	Point-of-Care Testing
PORT	Perception-of-Relationships Test
POW	Prisoners of War
ppb	Parts Per Billion
PPD	Postpartum Depression
PPE	Personal Protective Equipment
ppi	Pixels Per Inch
PPI	Proton Pump Inhibitor
PPP	Postpartum Psychosis
pRIA	Protein Radioimmunoassays
PRNU	Photo Response Nonuniformity
PSA	Prostate-Specific Antigen
PSI	Parenting Stress Inventory
PSTN	Public Switched Telephone Network
PTAH	Phosphotungstic Acid-hematoxylin
PTE	Pulmonary ThromboEmbolism
PTFE	Polytetrafluoroethylene
PTSD	PostTraumatic Stress Disorder
PT	Proficiency Testing
PUK	Pin Unlocking Key
PVC	Polyvinyl Chloride
Py-GC-MS	Pyrolysis Gas Chromatography Mass Spectrometry
QA	Quality Assurance
QC	Quality Control
QDE	Questioned Document Examiner

QM	Quality Management
QPN	Qualitative Probabilistic Network
RAID	Redundant Array of Independent Disks
RAM	Random Access Memory
RAPID	Ruggedized Advanced Pathogen Identification Device
RCMP	Royal Canadian Mounted Police
RC	Restructured Clinical
rCRS	Revised Cambridge Reference Sequence
RDC	Research Diagnostic Criteria
RDCT	Rey Dot Counting Test
RDD	Radiological Dispersion Device
RDX	Hexogen
RED	Radiological Emission Device
RF	Renal Failure
RFC	Request for Comment
RFID	Radio Frequency Identification Device
RFLP	Restriction Fragment Length Polymorphism
RFS	Robust File System
RFU	Relative Fluorescence Intensity
RFUs	Relative Fluorescent Units
RGB	Red, Green, and Blue
RH	Retinal Hemorrhage
RI	Refractive Index
RIA	Radio-Immunoassay
RIM	Research in Motion
RMNE	Random Man Not Excluded
RMP	Random Match Probability
ROC	Receiver Operating Characteristic
ROSITA	Roadside Testing Assessment
RP	Readiness Potential
RRT	Relative Retention Time
RSD	Relative Standard Deviation
RT	Retention Time
RT-PCR	Real-Time Polymerase Chain Reaction
RTS	Rape Trauma Syndrome
RUVIS	Reflected Ultraviolet Imaging System
Ry	Ryanodyine Receptor
S/N	Signal-to-Noise Ratio
S/P	Saliva-to-Plasma
SAAMI	Sporting Arms and Ammunition Manufacturers' Institute
SAC	Serum Alcohol Concentration
SADS-C	Schedule of Affective Disorders and Schizophrenia-Change
SAMHSA	Substance Abuse and Mental Health Service Administration
SARS	Severe Acute Respiratory Syndrome
SB-5	Stanford-Binet Intelligence Scale – Fifth Edition
SBP	Sellier Bellot, Prague
SBS	Shaken Baby Syndrome
SCAN	Scandinavian Pulp and Paper Association

SCC	Standards Council of Canada
SCID	Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, 4th ed.-Text Revision
SDH	Subdural Hemorrhage
SDIS	State DNA Identification System
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SD	Secure Digital
SD	Standard Deviation
SE	Secondary Electron
SEA	Strategic Environmental Assessment
SEIR	Surface-Enhanced Irregular Reflection
SEM/EDS	Scanning Electron Microanalysis with Energy Dispersive Sensor
SEM/EDX	Scanning Electron Microscopy/Energy Dispersive X-Ray
SEM/WDS	Scanning Electron Microanalysis with Wavelength Dispersive Sensor
SEM/WDX	Scanning Electron Microscopy/Wavelength Dispersive X-Ray Spectroscopy
SEM	Scanning Election Microscope
SF-ICP/MS	Sector Field-Inductively Coupled Plasma Mass Spectrometry
SFPE	Society of Fire Protection Engineers
SFST	Standardized Field Sobriety Test
SGAs	Second Generation Antipsychotics
SGM	Second-Generation Multiplex
SIB-R	Scales of Independent Behavior – Revised
SIDS	Sudden Infant Death Syndrome
SIM	Selected Ion Monitoring
SIM	Senior Identification Manager
SIM	Subscriber Identity Module
SIMCA	Soft Independent Modeling of Class Analogy
SIMS	Secondary Ion Mass Spectrometry
SIMS	Structured Inventory of Malingered Symptomatology
SIO	Senior Investigating Officer
SIPRI	Stockholm International Peace Research Institute
SIRS	Structured Interview of Reported Symptoms
SLA	Symbionese Liberation Army
SLP	Single Locus Probes
SLP	Single Locus Profiling
SLR	Single Lens Reflex
SMANZFL	Senior Managers of Australian and New Zealand Forensic Laboratories
SMD	Single-Metal Deposition
SMI	Severe Mental Illness
SMM	Stepwise Mutation Model
SMS	Short Message Service
SMTP	Simple Mail Transfer Protocol
SNAP-IV	Swanson, Nolan, and Pelham
SNP	Single Nucleotide Polymorphism
SNRI	Serotonin and Norepinephrine Reuptake Inhibitor
SOCO	Scenes of Crime Officer
SOFT	Society of Forensic Toxicologists

SOHO	Small and Home Office
SoHT	Society of Hair Testing
SOP	Standard Operating Procedure
SP	Shortpass
SPECT	Single Proton Emission Computed Tomography
SPE	Solid-Phase Extraction
SPIN	Service Planning Instrument
SPJ	Structured Professional Judgment
SPM	Scanning Probe Microscope
SPME	Solid Phase Micro Extraction
SPR	Small Particle Reagent
SRT	Sodium Rhodizonate Test
SSD	Scientific Support Department
SSM	Scientific Support Manager
SSM	Slipped Strand Mispairing
SSO	Sequence Specific Oligonucleotide
SSRI	Selective Serotonin Reuptake Inhibitor
SSSQ	Street Survival Skills Questionnaire
STA	Systematic Toxicology Analysis
STAG	Statistical Analysis of Glass
START	Short-Term Assessment of Risk and Treatability
STD	Sexually Transmitted Diseases
STEM	Scanning Transmission Microscope
STI	Sexually Transmitted Infections
StPO	Criminal Procedure Code
STR	Short Tandem Repeat
SUDEP	Sudden Unexplained Death in Epilepsy
SUDNIC	Sudden Unexpected Death Due to Neoplastic Disease in Infancy and Childhood
SVM	Support-Vector Machines
SVP	Sexually Violent Predator
SWAT	Special Weapons and Tactics
SWGDAM	Scientific Working Group on DNA Analysis Methods
SWEDOC	Science Working Group on Documents
SWGDRUG	Scientific Working Group for the Analysis of Drugs
SWGFAST	Scientific Working Group for Friction Ridge Analysis, Study, and Technology
SWGFEX	Scientific Working Group for Fire and Explosives
SWGGUN	Science Working Group on Guns
SWGhair	Scientific Working Group for Hairs
SWGIT	Scientific Working Group IT
SWGMAT	Scientific Working Group on Materials Analysis
TACS	Total Access Communication System
TAPPI	Technical Association of the Pulp and Paper Industry
TAT	Thematic Apperception Test
TATP	Triacetone Triperoxide
TBI	Traumatic Brain Injury
TBW	Total Body Water
TCD	Thermal Conductivity Detector

TCF	Totally Chlorine Free
TCP	Thienyl Cyclohexyl Piperidine
TCP	Transmission Control Protocol
TDMA	Time Division Multiple Access
TDx	Fluorescent Polarization Assay
TETRA	Terrestrial Trunked Radio
TDM	Target Disk Mode
TDM	Therapeutic Drug Monitoring
TdP	Torsades Des Pointes
TDx	Fluorescent Polarization Assay
TEA	Thermal Energy Analyzer
TEM	Transmission Electron Microscope
TEMED	Tetramethylethylenediamine
TETRA	Terrestrial Trunked Radio
TFPCl	*N*-(Trifluoroacetyl) Prolyl Chloride
TFS4	Transactional File System
TGA	Thermogravimetric Analysis
THC	Δ^9-Tetrahydrocannabinol
THC-COOH	11-Nor-9-Carboxy-Δ^9-Tetrahydrocannabinol
TIAs	Transient Ischemic Attacks
TIC	Toxic Industrial Chemical
TIC	Total Ion Chromatogram
TIM	Toxic Industrial Material
TIMS	Thermal Ionization Mass Spectrometry
TLC	Thin Layer Chromatography
TMA	Trimethoxyamphetamine
TMOT	Trace Metal Detection Test
TMJ	Total Metal Jacket
TMP	Thermomechanical Pulps
TMS	Trimethylsilyl
TNAZ	1,3,3-Trinitroazetidine
TNT	Trinitrotoluene
TOF	Time-of-Flight
TOF-SIMS	Time-of-Flight Secondary Ion Mass Spectrometry
TOMM	Test of Memory Malingering
TOR	The Onion Routing
TPMT	Thiopurine *S*-Methyltransferase
TPP	Thermal Protective Performance
TSH	Thyroid-Stimulating Hormone
TSOP	Thin Small-Outline Packages
TSWG	Technical Support Working Group
TTI	Transmit Terminal Identifier
TWGDAM	Technical Working Group on DNA Analysis and Methods
TWGED	Technical Working Group on Education and Training in Forensic Science
TWGFEX	Technical Working Group for Fire and Explosive Analysis
UAC	Urine–Alcohol Concentration
UCS	Unconditioned Stimuli
UDP	User Datagram Protocol

UEL	Upper Explosive Limit
UGPPA	Uniform Guardianship and Protective Proceedings Act
UHP	Ultra High Purity
UKAS/NAMAS	United Kingdom Accreditation Service/National Accreditation of Measurements and Sampling
UKNEQAS	United Kingdom National External Quality Assessment Scheme
UL	Underwriters Laboratories
UM	Ultrarapid Metabolizers
UMTS	Universal Mobile Telecommunication System
UN	United Nations
UNODC	United Nations Office on Drug Control
URN	Unique Reference Number
USDA	United States Department of Agriculture
USSC	United States Supreme Court
UV	Ultraviolet
UV-MSP	Microspectrophotometry Using Visible Light and Ultraviolet Light
UV/Vis	Ultraviolet/Visible
VA	Veterans Administration
VABS II	Vineland Adaptive Behavior Scales II
VCA	Vacuum Cyanoacrylate
VDAG	Vitamin D Binding Alpha Globulin
VIP	Validity Indicator Profile
VMD	Vacuum Metal Deposition
VNTR	Variable Number of Tandem Repeat
VoIP	Voice over Internet Protocol
VoIP/ToIP	Voice over Internet Protocol/Telephone over Internet Protocol
VQ	Vector Quantization
VRAG	Violence Risk Appraisal Guide
VRML	Virtual Reality Modeling Language
VSA	Video Spectral Analysis
VSA	Voice Stress Analyzers
VSA	Volatile Substance Abuse
VSC	Video Spectral Comparator
WADA	World Anti-Doping Agency
WAIS III	Wechsler Adult Intelligence Scale, Third Edition
WAIS-R	Wechsler Adult Intelligence Scale-Revised
WAN	Wide Area Network
WAP	Wireless Access Point
WAP	Wireless Application Protocol
WDX	Wavelength Dispersive X-Ray
WGA	Whole Genome Amplification
WHO	World Health Organization
WIRA	Wool Industries Research Association
WMD	Weapons of Mass Destruction
WMH-CIDI	World Mental Health – Composite International Diagnostic Interview
WMS	Wechsler Memory Scale

WRB	World Reference Base
WSQ	Wavelet Scalar Quantization
WTC	World Trade Center
WWI	World War I
WWII	World War II
XBO	Xenon Short Arc Lamp
XRD	X-Ray Diffraction
XRF	X-Ray Fluorescence Spectroscopy
XSR	eXtended Sector Remapper
Y-STRs	Y-Chromosome Short Tandem Repeats
YAFFS	Yet Another Flash File System
ZD	Z Direction
ZPO	Zivilprozessordnung(Civil Procedure Code)

The International System of Units (SI)

There are many different units of measure for most physical parameters such as length, mass, and temperature. The scientific community have agreed on a single system called the SI (Systeme Internationale) as the accepted international system of such units. SI units are either base or derived. Base units are fundamental and not reducible. Table 1 lists the main base units of interest in the discipline of materials science and engineering.

Derived units are expressed in terms of the base units, using mathematical signs for multiplication and division. For example, the SI units for density are kilogram per cubic meter (kg/m^3). For some derived units, special names and symbols exist; for example, N is used to denote the Newton, the unit of force, which is equivalent to 1 kg-m/s^2. Although many use the '/' notation (e.g. m/s) the system uses a superscript or exponent (to the power of) notation where a negative symbol replaces the '/'. For example, m/s becomes ms^{-1}, and m/s^2 becomes ms^{-2}. Table 2 contains a number of important derived units.

It is sometimes necessary, or convenient, to form names and symbols that are decimal multiples of SI units. Only one prefix is used when a multiple of an SI unit is formed, which should be in the numerator. These prefixes and their approved symbols are given in Table 3. Symbols for the main units are used in this book, SI or otherwise. Some disciplines retain other nomenclatures by practice or convention, although these are departures from the main scientific recommendations.

Table 1 The SI base units

Quantity	SI unit	Symbol
Length	meter	m
Mass	kilogram	kg
Time	second	s
Electric current	ampere	A
Thermodynamic temperature	kelvin	K
Amount of substance	mole	mol

Table 2 Some of the SI derived units (including examples of the superscript notation)

Quantity	Name	Formula	Special symbol
Area	square meter	m^2	–
Volume	cubic meter	m^3	–
Velocity	meter per second	m/s (ms^{-1})	–
Density	kilogram per cubic meter	kg/m^3	–
Concentration	moles per cubic meter	mol/m^3	–
Force	newton	$kg{\cdot}m/s^2$ ($kg{\cdot}ms^{-2}$)	N
Energy	joule	$kg{\cdot}m^2/s^2$, Nm	J
Pressure/Stress	pascal	kg/ms^2, N/m^2	Pa
Strain	–	m/m	–
Power, radiant flux	watt	$kg{\cdot}m^2/s^3$, J/s	W
Viscosity	pascal-second	kg/ms	Pa-s
Frequency (a periodic phenomenon)	hertz	1/s (s^{-1})	Hz
Electric charge	coulomb	A·s	C
Electric potential	volt	$kg{\cdot}m^2/s^2C$	V
Capacitance	farad	$s^2C/kg{\cdot}m^2$	F
Electric resistance	ohm	$kg{\cdot}m^2/sC^2$	Ω
Magnetic flux	weber	$kg{\cdot}m^2/sC$	Wb
Magnetic flux density	tesla	kg/sC, Wb/m^2	(T)

Table 3 SI multiple and submultiple prefixes

Factors by which multiplied	Prefix	Symbol
10^{12}	tera	T
10^9	giga	G
10^6	mega	M
10^3	kilo	k
10^{-2}	centi[(a)]	c
10^{-3}	milli	m
10^{-6}	micro	μ
10^{-9}	nano	n
10^{-12}	pico	p

[(a)] Avoided when possible.

Guide to Legal Citations

What is the Correct Form?

There are many systems for citing sources

When a publication such as this Encyclopedia is to accommodate technical contributions from authorities across the world, there is benefit in uniformity in the form that source references are printed. Thus, for scientific publications, we decided that the Vancouver style would be adopted. For the most part, that format is followed throughout these volumes.

It was more difficult to decide on the format for legal authorities and sources, because there is no single system of citing court cases, statutes, and other law publications. What exists is a conglomeration of conventions that are not always followed by various courts even within the same country. Local custom of reporting legal sources often supersedes what was prescribed as the "official" system in a jurisdiction.

Of course, that is only of limited help. In the United Kingdom, for instance, with no less than a dozen different courts that, at one time or another, publish or have published case reports (e.g., Appeal Cases (second and third series), Chancery Division, Criminal Appeal Reports, Queen's Bench, House of Lords, and other specialty courts), whatever convention exists is not always enlightening when one is confronted with a citation of a court. It is not easy for individuals who are not solicitors, barristers, or trained in law to locate where a particular case report may be found.

The confusion is even greater in the United States where there are 50 different states, each printing their case reports in official and/or unofficial reporter systems. In addition, the federal court system publishes cases decided by the United States Supreme Court (in three different reporter systems), by eleven Circuit Courts of Appeal and the District of Columbia (in two different publication forms), and by numerous United States District Court opinions. Then there is also a multitude of reports for specialized courts or federal agencies.

To seek to impart uniformity in citation form, editors from the Universities of Harvard, Columbia, Yale, and Pennsylvania Law Reviews have devised what is known as *The Bluebook – A Uniform System of Citation*. It is referred to as the *Bluebook*. The latest edition of the tome is the Eighteenth Edition, published in 2007. It comprises 415 pages. Unfortunately, the *Bluebook* has not imparted uniformity in citation form.

First, each of the *Bluebook's* editions advocated using conventions that were later partially modified in successive editions. Furthermore, the *Bluebook* is used essentially by law review journals published by law schools, but courts and other legal writers as well as publishers pay only scant attention to it and report cases in varying ways that they deem rational.

Clearly, then, some leeway must be accorded to authors in these volumes in the way they report case law, statutes, and other legal collections, since there is no one accepted way that might be said to be generally accepted as the "correct" format. In the face of these considerations we decided, as a matter of convenience, that legal sources would be cited in the form that was customary in the jurisdiction where the source originated.

It is more important, then, to understand the purpose that references serve. The purpose of a citation is so that the complete opinion of the court or other legal authority can be readily retrieved. Sadly, the convention whereby decisions of courts are cited makes sense only to legally trained individuals or to persons who work within a relatively closed environment.

Basic requirement of an informative citation form

An adequate citation of a reported opinion of a court is supposed to contain the following: (i) the name (or "style") of the case, (ii) the reporter system wherein the opinion has been published, (iii) the court that decided the case, and (iv) the year of decision. This is the preferred order in which court cases are cited in the United States and in many other countries, though the four elements of an informative citation are not always listed in the same order.

In the United Kingdom, for instance, the preferred form is to start with (i) the case name, (ii) the year of decision in parentheses, and thereafter (iii) the volume number (if available), (iv) the reporting system, (v) the page, (vi) the deciding court if that information is not otherwise clearly indicated by the reporting system, and (vii) the jurisdiction where the case was decided, if that information is not evident from the context or the citation. It is also preferred that cases be cited from official Law Reports rather than from ancillary sources. The Official Reports consist of the Appeal Courts (A.C.), Queen's Bench (Q.B.), Chancery (Ch.), Family (Fam.), and Probate (P) Reports. When citing decisions of the House of Lords, the Privy Council, or other sources that may report appeals from more than one jurisdiction, the jurisdiction from which the appeal was taken should be indicated parenthetically.

Most forensic scientists may be familiar with the way court opinions are published in their own country, but may not understand where and how foreign legal sources may be found. If, for research purposes, it becomes necessary for them to retrieve a complete court opinion or a statute for which the researcher possesses a citation, help in locating the full text referenced may be obtained from law-trained individuals. Major court systems also maintain libraries staffed by knowledgeable librarians. Legal libraries staffed with helpful librarians are further maintained in most law colleges and by many large legal firms.

Because a single uniform system for referencing legal authorities is lacking, some variations may be noted in the citation form used by authors of contributions. With the guidance offered here, the individual who is not legally trained will hopefully be able to decipher the meaning of legal references encountered throughout this Encyclopedia.

Andre Moenssens



Abuse of the Elderly ***see*** Elder Abuse: Policy

Accreditation: Criminalistics ***see*** Training and Certification (in Criminalistics)

Accreditation: Education ***see*** Education and Accreditation in Forensic Science

Accreditation: Laboratory

Crime laboratory accreditation has been one of the most powerful forces in improving the efficacy of forensic science services in the twenty-first century. Effective quality assurance programs are the linchpins for good forensic science, reliable techniques to apply the science, and trustworthy expert opinion testimony [1]. Encouraged by judicial opinions [2], mandated by four state legislatures [3], and implemented by crime laboratory directors [4], accreditation programs have brought needed regulation to a critical segment of our criminal justice system.

Introduction

The quality of scientific analyses conducted in crime laboratories varied widely throughout the United States before a program for crime laboratory accreditation was available. There were no generally accepted, published standards and no governmental regulations governing quality assurance in American crime laboratories [5]. A national proficiency testing research project during the mid-1970s [6], conducted by the Forensic Sciences Foundation (FSF) and funded with a grant from the National Institute of Law Enforcement and Criminal Justice Law Enforcement Assistance Administration (LEAA), brought public attention to proficiency problems and other inadequacies in some of America's crime laboratories [7]. An analysis of the results of the LEAA-funded project revealed "that some crime laboratories were experiencing serious problems in the examination and interpretation of several types of test specimens" [8] (*see* **Error Rates in Forensic Methods**; **Interpretation: Observer Effects**).

Some of the concerns cited by critics and commentators about unregulated crime laboratories included the following: (i) lack of written protocols for testing; (ii) lack of proficiency testing; (iii) level of education

and competency of examiners; (iv) uninformative nature, and in some cases the nonexistence, of laboratory reports and case files; (v) lack of review of examination results; (vi) experts testifying beyond the scope of their expertise or the science; (vii) questions about the integrity of evidence in the laboratories; (viii) questionable validity and reliability of scientific theories and methodologies used in the laboratories; and (ix) lack of oversight to remediate problems identified in laboratories [9, 10].

Crime laboratory directors and professional organizations [11], as well as legal commentators [12–14], recognized the need for the regulation of crime laboratories and the implementation of universally recognized quality assurance programs [15]. Crime laboratory accreditation programs help implement and oversee quality assurance programs that resolve the types of deficiencies mentioned above and improve the quality of the laboratory and the science performed in the laboratory [16].

That is not to say, however, that the absence of accreditation means a particular crime laboratory does not perform quality work. There are many unaccredited laboratories that have rigorous quality assurance programs ensuring valid and reliable scientific results [17]. Nor does accreditation guarantee that every examiner from an accredited laboratory will always perform up to appropriate standards. Accreditation is a safeguard that helps ensure that the laboratory provides accurate and valid results [18].

Since the LEAA-funded proficiency testing project, several forensic science laboratory accreditation programs have been developed. The American Board of Forensic Toxicology (ABFT) accredits toxicology laboratories, having 24 laboratories accredited under its program [19]. The National Association of Medical Examiners (NAME) established a voluntary accreditation program and has accredited 54 medicolegal death investigative offices [20]. Forensic Quality Services-International (FQS-I) provides accreditation to forensic science testing laboratories in the United States and has accredited 53 crime laboratories and other forensic science laboratories since 2001. It also provides accreditation to Identification Sections and Crime Scene Units within police departments [21]. The National Forensic Science Technology Center (NFSTC) conducts external DNA audits of publicly funded laboratories desiring to receive federal funding and entry into the National DNA Identification System (NDIS). The NFSTC is funded by a Cooperative Agreement with the National Institute of Justice, Office of Science and Technology [22] (*see* **Databases**).

By far the oldest, most prolific accreditation program, and the most widely recognized by courts, legislatures and the crime laboratory community, is the American Society of Crime Laboratory Directors/Laboratory Accreditation Board (ASCLD/LAB®) [23]. ASCLD/LAB came into existence through the work of the "Committee on Evaluation and Standards," which was one of the first committees appointed by the ASCLD. The committee was appointed largely because of deficiencies in crime laboratories, which were revealed by the LEAA proficiency testing program. The committee was charged to develop a program which sets standards of operation and a process for evaluation of crime laboratories. ASCLD/LAB received its first accreditation applications in early 1982 and accredited seven Illinois State Police laboratories.

By the fall of 1984, ASCLD/LAB had accredited 30 crime laboratories and had achieved a predetermined minimum size to become an independent not-for-profit organization, which was then incorporated in the state of Missouri. Since its inception, ASCLD/LAB has accredited more than 342 public and private crime laboratories at every governmental level, including international laboratories and the major federal law enforcement and military crime laboratories in the United States [24]. As of September 2008, the 342 laboratories holding current ASCLD/LAB accreditation included 178 state laboratories representing 48 states, 109 local government laboratories, 22 federal laboratories, 22 private laboratories, and 11 international (non-United States) laboratories. ASCLD/LAB maintains two programs, ASCLD/LAB-Legacy® and ASCLD/LAB-*International*®. The former is the original program under which accreditation is based upon requirements created, adopted and maintained by the ASCLD/LAB Delegate Assembly. The Legacy program is being phased out. The International program, implemented in April 2004, is based on the requirements of the International Organization for Standardization (ISO) and supplemental requirements such as the ASCLD/LAB-Legacy program requirements. In 2007, a separate accreditation program was implemented under the ASCLD/LAB-*International* program for breath alcohol calibration laboratories. As of September 2008, 69 testing laboratories and one

breath alcohol calibration laboratory were accredited under the International program and the remainder under the Legacy program.

With more than three-quarters of the publicly funded crime laboratories in the United States accredited by ASCLD/LAB [25], it is significant to the integrity of the criminal justice system that ASCLD/LAB's accreditation program provides professional regulation and review of a laboratory's quality assurance program. Accreditation is a part of a laboratory's quality assurance program which, under ASCLD/LAB's program, also includes proficiency testing, continuing education, and other programs to help the laboratory achieve excellence and give better overall service to the criminal justice system (*see* **Fire Investigator: Standardization, Accreditation, and Certification**; **Accreditation: Organizational**; **Quality Systems: Toxicology**). A quality assurance program which meets the ASCLD/LAB standards addresses the criticisms leveled at unregulated laboratories.

While the ASCLD/LAB accreditation program is voluntary, some states have mandated accreditation and have chosen ASCLD/LAB as their accrediting authority for multidisciplinary crime laboratories. New York established a Commission on Forensic Science which developed minimum standards and a mandatory program of accreditation for all forensic laboratories within the state [26]. All of the state's multidisciplinary crime laboratories are now accredited by ASCLD/LAB, while some of the toxicology laboratories are accredited by the ABFT. Likewise, Oklahoma requires that all crime laboratories within the state be accredited by ASCLD/LAB and its toxicology laboratories by ABFT [27]. Legislation passed in the state of Texas in 2003 requires the Texas Department of Public Safety (DPS) to establish a program of accreditation for "forensic examinations" [28]. Missouri requires that after December 31, 2012, any crime laboratory providing reports or testimony to a state court pertaining to a result of the forensic analysis of evidence shall be accredited or provisionally accredited by a laboratory accrediting organization approved by the DPS [29].

The ASCLD/LAB-International Accreditation Program

As discussed earlier, the ASCLD/LAB-Legacy program is being phased out of existence, and the ASCLD/LAB-*International* program is replacing it, using the ISO framework. This transition reflects the future of accreditation programs in the United States by which laboratories and accrediting bodies will be embracing internationally accepted accreditation standards.

The ISO has published at least two sets of standards having direct application to the accreditation of laboratories of all types [30]. The first, *ISO/IEC 17025: General Requirements for the Competence of Testing and Calibration Laboratories* [31] is not specific to crime laboratories, but the standards are, as the name implies, applicable to any organization performing tests and/or calibrations. The second ISO standard of interest [32], *ISO/IEC 17011:2004 Conformity assessment – General requirements for accreditation bodies accrediting conformity assessment bodies* [33], establishes a system of internationally recognized standards for the accreditation of laboratories [34].

Only in this century has the American forensic science community looked seriously at adopting ISO/IEC 17025 as the foundation of crime laboratory accreditation programs. NFSTC started conducting ISO/IEC 17025 compliant accreditation inspections of forensic laboratories in 2001 [35]. ASCLD/LAB-*International* was promulgated in December 2003 by the Delegate Assembly, and ISO/IEC 17025 compliant accreditation inspections began in the summer of 2004.

Engaging in ISO/IEC 17025 compliant inspections is but one requirement of ISO/IEC 17011 requirements. Compliance with all elements of ISO/IEC 17011 brings an added dimension of quality to crime laboratory accreditation efforts in America. The ultimate purpose of ISO/IEC 17011 is to establish standards for the recognition of competent accrediting bodies. An ISO/IEC 17025 accrediting body may voluntarily comply with the provisions of ISO/IEC 17011 and self-declare compliance, or apply for formal recognition to a body operating a recognition program for accrediting bodies, which in the Americas is the Inter-American Accreditation Cooperation (IAAC) [36]. Regional cooperations like the IAAC exist in most developed regions of the world for the primary purpose of offering recognition to accreditation bodies and facilitating the cooperation of such bodies in those regions.

In September 2008, ASCLD/LAB was granted formal recognition by the IAAC for meeting all

the applicable ISO and IAAC standards for the operation of a competent accrediting body. After a thorough evaluation by an IAAC international team of evaluators, ASCLD/LAB, and specifically the ASCLD/LAB-*International* accreditation program for testing laboratories, became the first forensic science accreditation body in the United States to gain international recognition for meeting all applicable ISO and IAAC standards for the operation of a competent accrediting body.

ISO standards and concepts in the accreditation of crime laboratories address criticism that previous crime laboratory accreditation programs were designed, adopted, implemented, and overseen solely by the users of each program. Embracing internationally adopted ISO/IEC 17025 standards and opening the accreditation programs to third-party ISO/IEC 17011 scrutiny was the next logical step in the evolution of crime laboratory accreditation in the United States.

Perhaps, the greatest strength in ISO compliant programs of accreditation is the increased physical presence in the accredited laboratories. FQS-I operates their program of ISO accreditation on a 2- to 5-year cycle. Depending on the period of accreditation granted, FQS-I will either conduct a full on-site inspection of the laboratory every second year or conduct periodic surveillance visits if the cycle of accreditation is longer than 2 years. The ASCLD/LAB-*International* program operates on a 5-year cycle of accreditation, but conducts annual on-site surveillance visits in each laboratory. Surveillance visits are abbreviated versions of full assessments.

However, there are weaknesses associated with ISO/IEC 17025. The most critical area where the international requirements for accreditation fall woefully short of American programs like ASCLD/LAB-Legacy is proficiency testing requirements. Undoubtedly as a direct result of the LEAA study in the mid-1970s, proficiency testing has become entrenched as a major part of the compliance monitoring phase of crime laboratory accreditation programs in the United States. Even many nonaccredited laboratories participate in internal and external proficiency tests each year. Interestingly, ISO/IEC 17025 is virtually silent on the subject of proficiency testing. To fill this critical gap, ASCLD/LAB-*International* includes supplemental requirements – like proficiency testing and specific educational requirements for laboratory analysts (*see* **Expert Witnesses: Selection and Investigation of Credentials**).

Embracing ISO standards and concepts should not be viewed, however, as a panacea for curing all the current quality issues in America's crime laboratories. Implementing ISO compliant accreditation programs is but another step in the ongoing process of assuring quality in the forensic sciences and integrity in the criminal justice system.

To that end, ASCLD/LAB-*International* has four objectives which define the purposes and nature of the program. The objectives are

- to improve the quality of laboratory services;
- to develop and maintain criteria that may be used by a laboratory to assess its level of performance and to strengthen its operation;
- to provide an independent, impartial, and objective system by which laboratories can benefit from a total operational review; and
- to offer to the general public and to users of laboratory services a means of identifying those laboratories, which have demonstrated that they meet established standards.

The forensic disciplines for which ASCLD/LAB-*International* affords accreditation include controlled substances, toxicology, trace evidence, biology (including DNA), firearms/toolmarks, questioned documents, latent prints, crime scene, and digital & multimedia evidence. ASCLD/LAB-*International* also offers accreditation to laboratories certifying the calibration of breath alcohol measurement devices. In the area of DNA, the Federal Bureau of Investigation (FBI's) Quality Assurance Standards for Forensic DNA Testing Laboratories and Convicted Offender DNA Databasing Laboratories are applied during the inspection [37] (*see* **Databases**). These quality assurance standards and programs were established pursuant to the DNA Identification Act of 1994 [38]. To receive federal funding for DNA analysis, laboratories must meet these standards. The FBI and ASCLD/LAB have agreed to cooperate in the implementation of these standards in laboratories inspected by ASCLD/LAB.

When a laboratory applies for accreditation, ASCLD/LAB requires an assessment of all the recognized disciplines in the laboratory, with the exception of the crime scene discipline. The applicant laboratory director has the option of including crime

scene and/or breath alcohol measurement device calibration in the accreditation process. Other than exempting those two disciplines, ASCLD/LAB will not do a partial accreditation of any laboratory.

The International program has adopted many accreditation requirements from the well-established standards provided by the Legacy program to supplement ISO/IEC 17025, making the program more relevant and applicable to forensic science testing laboratories. The written format of the supplemental requirements corresponds to the ISO format and contains the basic principles of quality assurance as well as notes to provide clarification or examples. The supplemental requirements are found in the *ASCLD/LAB-International Supplemental Requirements for the Accreditation of Forensic Science Testing Laboratories* and the *ASCLD/LAB-International Supplemental Requirements for the Accreditation of Forensic Science Testing Calibration Laboratories.*

The ASCLD/LAB-International Accreditation Process

As a part of the accreditation process, an assessment team, composed of a Lead Assessor and Technical Assessors, visits the laboratory or laboratory system and assesses the laboratory against all the applicable accreditation criteria in ISO/IEC 17025 and the Supplemental Requirements document. Assessments are conducted primarily by volunteer assessors from accredited laboratories who have successfully completed the ASCLD/LAB assessor training program and who are trained in the assessment criteria [39]. As a result of the growth in the program, there is now a paid staff employed by ASCLD/LAB, who act as lead assessors, program directors, quality manager, training manager, executive director, and support staff [40].

The assessment team determines whether the laboratory has met each of the accreditation requirements. If it has not, the laboratory receives a "no" as to every nonconforming criterion. "nonconformities" are classified as Level 1 or Level 2. The more serious nonconformity is Level 1, the nature of which directly affects and has a fundamental impact on the work product of the laboratory or the integrity of the evidence. All the other nonconformities are Level 2.

Prior to the completion of the on-site assessment, a Summary Assessment Report is prepared by the Lead Assessor and reviewed by an ASCLD/LAB Quality Review Panel to ensure consistency in the application of accreditation requirements and in grading the laboratory. The report is then given to the laboratory director at the summation conference that concludes the assessment visit. In addition, the laboratory director is given a Corrective Action Request for each criterion in which the laboratory was found to be nonconforming. Level 1 nonconformities must be corrected within 180 days and no laboratory will be accredited with an outstanding Level 1 nonconformity. Before the next annual on-site surveillance visit, all Level 2 nonconformities must be corrected or they will be treated as a Level 1 nonconformity.

After the summation conference, the Lead Assessor prepares a full assessment report that is again subject to a quality review, which is then given to the laboratory. This is a more formal report that includes all the nonconformities reported during the summation conference.

When the Lead Assessor determines that all Level 1 corrective actions have been completed, he or she will prepare a final assessment report to the Board of Directors [41]. The Board then grants or denies accreditation, or may require completion of additional corrective actions [42, 43]. When a laboratory system applies for accreditation, each laboratory within the system will be judged and accredited independently of each other.

Accreditation is granted for a period of 5 years, during which the laboratory is expected to remain compliant with the accreditation standards. During the accreditation cycle, compliance with accreditation criteria is monitored on an annual basis by three methods. The first method requires the submission of an Annual Accreditation Audit Report to ASCLD/LAB by each laboratory. If discrepancies with accreditation criteria are revealed, the laboratory is expected to remediate them under the supervision of ASCLD/LAB. If failures appear during the year, the standards require the laboratory to have a procedure in place that must be followed to correct the problems. If there are significant changes in the laboratory during the year, they must also be reported so that ASCLD/LAB can determine whether a special interim assessment is required to assure compliance with the standards.

The second method of determining compliance is on-site surveillance visits conducted annually in each accredited laboratory. These miniassessments

allow for first-hand, personal observations of the laboratory's quality assurance program. Core accreditation requirements are assessed, as well as a number of additional requirements selected for the on-site assessment. The annual surveillance visits are conducted much like the full assessment, with a summation conference, and a full and then final surveillance report, with intervening quality reviews.

Proficiency testing is the third means of monitoring compliance. ASCLD/LAB has adopted a Proficiency Review Program that outlines the review process for external proficiency test results from approved test providers [44]. For each forensic discipline accredited by ASCLD/LAB, there is a Proficiency Review Committee. These committees review external proficiency test results provided by approved test providers in accordance with the ASCLD/LAB Proficiency Review Program. If a discrepancy is indicated in the report, the committee requests a response from the laboratory as to its investigation of the cause of the discrepancy and the corrective measures that have been taken to correct it. There are three classes of discrepancies. Class I discrepancies are the most serious, the nature and cause of which raise immediate concern regarding the quality of the laboratory's work product. The Board of Directors has the authority to impose sanctions for discrepancies and the failure to correct them, including probation, suspension, or revocation of the accreditation status.

If a failure to comply with the accreditation standards is reported from within the laboratory, from outside the laboratory, or discovered by a special interim assessment or surveillance visit, an investigation will ensue in appropriate cases. The same sanctions described above may be imposed by the Board of Directors on a laboratory that fails to remain compliant. The Board has the authority to send an assessment team into the laboratory at any time to determine compliance with accreditation criteria. In these ways, ASCLD/LAB provides oversight in the remediation of problems in accredited laboratories.

For nonaccredited laboratories that experience quality-related problems, ASCLD/LAB has developed an assistance program. At the invitation of the laboratory or the laboratory's supervising authority, ASCLD/LAB will provide a team of trained assessors to conduct an on-site review of the laboratory's processes and casework to determine the cause of the deficiencies and errors in examination.

ASCLD/LAB Accreditation Criteria

ASCLD/LAB-*International* accreditation criteria for testing and calibration laboratories cover the following broad categories: laboratory management and operations, personnel qualifications, and physical plant.

Laboratory Management and Operations

The laboratory management and operations criteria contain the heart of the accreditation program and include the principles mentioned later.

Evidence Control. This series of criteria requires a comprehensive chain of custody to maintain the integrity of the evidence while in the laboratory. Evidence control requires marking evidence for identification purposes, storage under proper seal, and protection from loss, cross-transfer, contamination, and/or deleterious change (*see* **Chain of Possession of Tangible Evidence**).

Management System. A fully documented management system [45] is required, including a quality manual, in which the laboratory documents the policies and procedures that pertain to quality assurance, including a requirement of compliance by the management and staff. A quality manager must also be appointed who is responsible for the quality system in the laboratory. Internal audits are one of the primary tools used to evaluate, confirm, or verify activities related to quality. Their purpose is to assess compliance with the operational requirements of the entire management system. Periodic audits, along with day-to-day review of scientific reports, provide an effective means for ensuring that quality control activities are being implemented and that each forensic examiner performs in a manner consistent with the quality management system.

Procedures that are used in the laboratory must be documented to be demonstrably capable of producing valid results. They must be generally accepted in the scientific field or supported by data gathered and recorded according to the scientific method. New procedures must be scientifically validated before being used. These safeguards, reviewed by the laboratory's quality manager and by ASCLD/LAB assessors, help assure that valid theories and methodologies are in place for the examination of evidence.

The quality aspects of the management system also require that written technical procedures are in place for sample preparation methods, controls, standards, and equipment calibration procedures. These practices are crucial for the reliability of the methodology involved in many scientific tests. These protocols and the protocols established for individual tests in the different forensic disciplines assure that tests are conducted properly and under generally accepted procedures. These protocols not only help to assure validity and reliability but also help to meet the admissibility requirement of the "existence and maintenance of standards controlling the technique's operation" set forth by the Supreme Court [46] (*see* **Expert Opinion in Court: a Comparison of Approaches**; **Expert Opinion: United States**); ***Daubert v. Merrell Dow Pharmaceuticals***; and **Federal Rule of Evidence 702**).

Laboratory reports must be supported by a comprehensive case record that includes all notes, worksheets, photographs, spectra, printouts, charts, and other data or records used by the examiners to support their conclusions (*see* **Report Writing for Courts**). Documentation to support conclusions must be such that in the absence of the examiner, another competent examiner or supervisor could evaluate what was done and interpret the data. Acceptable ways to document the basis for conclusions derived from evidence examination may include a narrative description of the examination process and observations made, photographs, photocopies, diagrams, drawings, and worksheets. This requirement is essential because it allows a technical review. With these requirements in place, it will be much harder for an unscrupulous examiner to hide his or her incompetency, shortcuts, or fraud. Moreover, the complete case record required by ASCLD/LAB allows a more thorough and comprehensive review by defense experts of the tests conducted (*see* **Discovery of Expert Findings**).

As an additional quality assurance procedure to ensure compliance with testing protocols, technical review by another competent examiner of a percentage of laboratory reports is required (*see* **Peer Review as Affecting Opinion Evidence**). In addition, there must be a course of action mandated when discrepancies or nonconforming testing are found. The oversight of the quality of the testing and the reporting is an essential criterion that affords the public confidence that the conclusions issued in reports are accurate and reliable.

The courtroom testimony of each examiner is also reviewed annually for several purposes, including determinations that testimony is scientifically consistent with the work documented in the case file. Periodic, random review of testimony discourages examiners from testifying beyond the scope of his or her report and helps discover those that do [47].

Proficiency Testing. Proficiency testing is an integral part of a quality assurance program. It measures the capability of the examiner and the reliability of the analytical results. Such tests help determine where more training is required or where more stringent quality control may benefit. It also demonstrates, in part, the current competence of the laboratory. One test must be administered annually for each discipline in the laboratory and must be provided by an external proficiency test provider. ASCLD/LAB approves providers meeting its requirements. The external review of proficiency test results is a crucial element of the compliance monitoring process. In addition to the external tests required for each discipline, the program requires that each individual in the laboratory participate in at least one proficiency test annually. It should also be noted that the program encourages laboratories to give proficiency tests to each examiner annually in each subdiscipline in which casework is performed. In addition to participating in external proficiency testing, a laboratory should conduct proficiency testing using blind tests prepared internally or externally and submitted as normal casework evidence or by reexamination by another examiner of evidence on which casework was previously completed.

Personnel Qualifications

The qualifications of a laboratory's examiners are of great importance. These criteria establish minimum education, experience requirements, and training. They also require, for most of the forensic disciplines, a minimum of a baccalaureate degree in a natural science or a related field relevant to the scientific discipline in which the examiner is practicing. These criteria have raised the education levels required in many disciplines [48].

While certification of examiners by an appropriate certification organization is not required by ASCLD/LAB, the accreditation standards do require that examiners have the education, experience, and

training commensurate with the examinations and testimony they provide [49]. When certification is obtained, it should be by a recognized organization [50].

Initial competency and continued proficiency of examiners is also crucial. Each examiner must pass a competency test prior to assuming casework responsibilities and must successfully pass a proficiency test at least annually.

Physical Plant

A well-designed and outfitted laboratory can greatly add to its productivity and reliability. These criteria require that appropriate security measures be practiced and address the adequacy and appropriateness of laboratory space. The laboratory's physical plant should reflect due consideration of space, design, security, health, and safety. A reference guide used by laboratories is the "Forensic Laboratories: Handbook for Facility Planning, Design, Construction, and Moving," publication NCJ168106, prepared by the US Department of Justice, Office of Justice Programs, National Institute of Justice.

Conclusion

While crime laboratory accreditation programs have improved substantially over the last two decades, there is still a lot of work to be done. Report writing, proficiency testing, and ethics training are some of the areas that still need improvement (*see* **Ethics: Codes of Conduct for Expert Witnesses**; **Report Writing for Courts**). Most important, there remain a considerable number of laboratories, both public and private, that have not been accredited by any organization. Forensic sciences now play such an integral role in the criminal justice system [51] that the public and the criminal justice system as a whole demand the implementation of quality assurance programs monitored by outside entities. Accreditation by an internationally recognized program is a safeguard that helps ensure laboratories provide accurate and valid results.

End Notes

[a.] Kenneth E. Melson is a board member on the American Society of Crime Laboratory Directors/Laboratory Accreditation Board (ASCLD/LAB) and a prosecutor. Ralph Keaton, ASCLD/LAB Executive Director, and John K. Neuner, ASCLD/LAB International Program Manager, contributed to this article, an early version of which appeared in American Bar Association, Criminal Justice Section, *Crime Laboratory Accreditation: The Assurance of Quality in the Forensic Sciences*, The State of Criminal Justice (August 2003).

References

[1] The elements which make up a comprehensive quality assurance program are described in National Research Council (1992). *Communications on DNA Techniques in Forensic Science. DNA Technology in Forensic Science*, p. 98.

[2] In Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579, 594 (1993), the Supreme Court noted that a court ordinarily should consider the existence and maintenance of standards controlling the technique's operation when determining admissibility of scientific evidence, (citing United States v. Williams, 583 F.2d 1194, 1198 (2nd Cir., 1978) (noting professional organization's standards governing the technique). Judges are citing the accreditation standards in decisions on admissibility of scientific evidence. *See, e.g.*, Smith v. State, 702 N.E.2d 668, 673 (Ind. 1998).

[3] New York, Oklahoma, Texas and Missouri are now the only jurisdictions that require forensic laboratories to be accredited. *See* N.Y.EXEC. § 995-a (McKinney 1996); OKLA. STAT. tit. 74, § 150.37 (2002); TEX. CODE CRIM. PROC. art. 38.35 (1991), *amended by* H.B. 2703, 2003 Leg. (Tex. 2003); TEX.GOV.CODE ANN. § 411 (Vernon 1998) *amended by* H.B. 2703, 2003 Leg. (Tex. 2003); Mo.Rev.Stat. § 650.060 (2008).

[4] The American Society of Crime Laboratory Directors voted to begin a voluntary accreditation program for their laboratories in (1981).

[5] For a history of accreditation programs in the fields of health care, legal education, and law enforcement and corrections, see Dodd, V.J. (1989). Toward a national system of state court accreditation, *Judicature* **73**, 141.

[6] The results of the study were reported in 1978 in Peterson, J.L., Fabricant, E.L., Field, K.S. Thornton, J.I. (1978). *Crime Laboratory Proficiency Testing Research Program*.

[7] Peterson, J.L. & Markham, P.N. (1995). Crime laboratory proficiency testing results, 1978–1991, I: identification and classification of physical evidence, *Journal of Forensic Sciences* **40**, 994–1008.

[8] *Id.* at 994.

[9] For examples of problems in crime laboratories, see Giannelli, P.C. (1997). The abuse of scientific evidence in criminal cases: the need for independent crime laboratories, *Virginia Journal of Social Policy & the Law* **4**, 439.

[10] Maier, T.W. (2003). *Inside the DNA Labs, Insight on the News*, May 26, (accessed May 2003). Available at www.insightmag.com/news/436794.html.
[11] Hansen, L.B. (1992). Stemming the DNA tide: a case for quality control guidelines, *Hamline Law Review* **211**(165), 16. (stating that organizations calling for quality assurance standards include the National Association of Attorneys General, the National District Attorneys Association, and the American Society of Crime Laboratory Directors).
[12] Lander, E. (1989). DNA fingerprinting on trial, *Nature* **339**, 501–505.
[13] Jonakait, R. (1991). Forensic science: the need for regulation, *Harvard Journal of Law & Technology* **4**, 109–191.
[14] Weinstein, J.B. (1998). Science, and the challenge of expert testimony in the courtroom, *Oregon Law Review* **77**, 1005–1011.
[15] Bales, S. (2000). Turning the microscope back on forensic scientists, *Litigation* **26**(2), 51–54, (explaining why crime laboratory accreditation is important).
[16] A primary recommendation made in the 1997 Office of Inspector General's report on its investigation of allegations concerning the FBI laboratory was that it obtain accreditation by the American Society of Crime Laboratory Directors/Laboratory Accreditation Board (ASCLD/LAB) as soon as possible. *Id.* at 54. *See* U.S. Department of Justice, Office of Inspector General, The FBI Laboratory: An Investigation into Laboratory Practices and Alleged Misconduct in Explosives-Related and Other Cases (April 1997).
[17] National Research Council (1992). *Cf. Communications on DNA Techniques in Forensic Science, DNA Technology in Forensic Science*, p. 107, (suggesting that courts should view the absence of appropriate accreditation as constituting a prima facie case that the laboratory has not complied with generally accepted standards).
[18] Bales, S. (2000). Turning the microscope back on forensic scientists, *Litigation* **26**(2), 51–54.
[19] As of September (2008). The website for the American Board of Forensic Toxicologists is www.abft.org.
[20] As of July (2008). The website for the National Association of Medical Examiners is www.thename.org.
[21] As of September (2008). The FQS-I website is www.forquality.org.
[22] As of November (2008). The website for the National Forensic Science Technology Center is www.nfstc.org.
[23] As of November (2008). The website for ASCLD/LAB is www.ascld-lab.org.
[24] As of November (2008). A list of accredited laboratories can be viewed at www.ascld-lab.org.
[25] In 2005, 78% of 293 laboratories responding to a Bureau of Justice Statistics' survey of the 389 publicly funded crime laboratories in the United States were accredited by ASCLD/LAB, with 3% accredited by some other organization. Durose, Census of Publicly Funded Forensic Crime Laboratories, 2005, Bureau of Justice Statistics Bulletin, July 2008. Since 2005, ASCLD/LAB has increased the number of publicly funded crime laboratories accredited by one of its two accreditation programs.
[26] *See* N.Y. Exec. § 995-b (McKinney 1996).
[27] *See* Okla. Stat.tit. 74, § 150.37 (2002).
[28] *See* Tex. Code Crim. Pro. art. 38.35 (1991), *amended by* H.B. 2703, 2003 Leg. (Tex. 2003); Tex.Gov.Code Ann. § 411 (Vernon 1998), *amended by* H.B. 2703, 2003 Leg. (Tex. 2003).
[29] Mo.Rev.Stat. § 650.060 (2008).
[30] As of November (2008). The website for ISO is www.iso.org.
[31] As of November (2008). *ISO/IEC 17025: General Requirements for the Competence of Testing and Calibration Laboratories*, is available for purchase from the American National Standards Institute (ANSI) at www.ansi.org. IEC stands for the International Electrotechnical Commission.
[32] There are more than two ISO standards related to the work and accreditation of crime laboratories. For example, ISO Guide 43 (Parts I & II) addresses, respectively, standards of proficiency testing by interlaboratory comparisons, and the selection and use of proficiency testing schemes by laboratory accreditation bodies.
[33] In this context, the term "conformity assessment body" means the laboratory being evaluated for accreditation.
[34] As of November (2008). *ISO/IEC 17011:2004 Conformity assessment – General requirements for accreditation bodies accrediting conformity assessment bodies* is available for purchase at www.ansi.org.
[35] NFSTC was the predecessor to FQS-I, and no longer offers full laboratory accreditations.
[36] As of November (2008). Additional information about the Inter-American Accreditation Cooperation (IAAC) is available at http://www.iaac.org.mx/.
[37] As of November (2008). *See* www.fbi.gov/hq/lab/fsc/backissu/july2000/codispre. htm, for the standards.
[38] 42 U.S.C. § 14131.
[39] Using volunteer assessors from accredited laboratories has two distinct advantages. First, these individuals are experts in their field, doing work in the subject matter expertise in the area they are inspecting. Second, participation in the accreditation program gives additional training in quality assurance standards and practices to examiners who work in crime laboratories.
[40] Paid staff assessors were added to the program to assure consistency in application of accreditation criteria in the inspection process and the review of inspection reports. The quality manager is responsible for the Proficiency Review Program and the quality of the accreditation program.
[41] The Board of Directors is composed of nine voting members, a non-voting ex-officio member representing ASCLD (a separate non-profit organization) and a non-voting Executive Director. Seven of the voting members are elected from among the Delegate Assembly. One elected Board member is a representative of law enforcement and prosecuting attorneys and one is

a public member. The Board is elected by the Delegate Assembly, which is composed of the directors of all ASCLD/LAB accredited laboratories and laboratory systems.

[42] Contrary to outside criticism suggesting the accreditation program is a trade association administered by a group of golfing buddies, less than ten percent of applicant laboratories under the Legacy program achieved accreditation following the initial inspection. It is expected that the same will apply under the International program. *See* McDonald, R. (1998). Juries and crime labs: correcting the weak links in the DNA chain, *American Journal of Law & Medicine* **24**, 345–356.

[43] Arvizu, J. (2000). *Shattering the Myth: Forensic Laboratories*, CHAMPION **18**, 20–21, 24-May. Despite these early characterizations of ASCLD/LAB, two of the states requiring mandatory accreditation and the United States Department of Justice Office of the Inspector General have all recognized the ASCLD/LAB as an appropriate accreditation body, despite its peer-reviewer structure. The National Research Council in its second study of DNA analysis recommended that laboratories should make every effort to be accredited for DNA work by such organizations as ASCLD/LAB. COMM. ON DNA FORENSIC SCI., NAT'L RESEARCH COUNCIL, THE EVALUATION OF FORENSIC DNA EVIDENCE 4 (1996). The long-standing practice of accrediting bodies like ASCLD/LAB to use senior forensic scientists, with actual experience in a crime laboratory, and who receive specialized inspection training, meets both the letter and intent of peer-reviewer philosophy shared worldwide. *See* The International Organization of Standardization's (ISO) *ISO/IEC 17011:2004 Conformity assessment – General requirements for accreditation bodies accrediting conformity assessment bodies* (assessors (inspectors) must be selected and used based upon their having appropriate technical knowledge and practical experience directly related to the work of the laboratory being inspected); *KPI 3: Accrediting Body Staff, Assessors and Experts*, in the International Laboratory Accreditation Cooperation's (ILAC) publication *ILAC P7:2003 / Key Performance Indicators (KPIs)* (reiterating the necessity for accrediting bodies to use only inspectors who possess the appropriate "technical qualifications and practical experience.").

[44] As of November (2008). The ASCLD/LAB Proficiency Review Program document may be viewed at www.ascld-lab.org.

[45] The term "management system" is defined in *ISO/IEC 17025:2005 General requirements for the competence of testing and calibration laboratories* as "the quality, administrative and technical systems that govern the operations of a laboratory."

[46] Daubert v. Merrell Dow Pharmaceuticals, Inc, 509 U.S. 579, 594 (1993).

[47] For an example of a capital murder case where the trial testimony of an expert went beyond the conclusions in the lab report, see Troedel v. Wainwright, 667 F. Supp. 1456 (S.D. Fla. 1986), *aff'd* 828 F.2d 670 (11th Cir., 1987).

[48] There are many outstanding forensic science programs being offered in colleges and universities. The American Academy of Forensic Sciences formed an accrediting body, called the Forensic Science Education Programs Accreditation Commission (FEPAC), which has as its mission the accreditation of college-level academic programs that lead to a baccalaureate or graduate degree.

[49] An indication of the growing recognition of certification is the 2003 amendment to the Oklahoma accreditation statute which exempts latent print identifications performed by latent print examiners certified by the International Association for Identification (IAI) (www.theiai.org) from the requirements of a technical peer review system, a proficiency testing program and accreditation. *See* OKLA. STAT. tit 74, § 150.37, *amended by* H.B.1802, 2003 Leg. (OKLA. 2003). For a listing of certification organizations, see Carol Henderson Garcia, *Expert Witness Malpractice: A Solution to the Problem of the Negligent Expert Witness,* 12 MISS. C.L. REV. 39, 62 (1991). Many of the certification organizations may be found by reviewing the website for the American Academy of Forensic Sciences (AAFS) at www.aafs.org.

[50] The American Academy of Forensic Sciences has established an accrediting body, the Forensic Specialties Accreditation Board, to accredit certification organizations.

[51] Evidence from an estimated 2.7 million criminal investigations was submitted to the nation's forensic crime laboratories in 2005. Durose, Census of Publicly Funded Forensic Crime Laboratories, 2005, Bureau of Justice Statistics Bulletin, July 2008.

KENNETH E. MELSON, RALPH KEATON AND JOHN K. NEUNER

Accreditation: Organizational

Introduction

Accreditation is becoming (*see* **Accreditation: Laboratory**; **Fire Investigator: Standardization, Accreditation, and Certification**) more widely present in criminalistics. According to the most important forensic organizations, such as the European Network of Forensic Science Institutes

(ENFSI), the American Society of Crime Laboratory Directors (ASCLD), and the Senior Managers of Australian and New Zealand Forensic Laboratories (SMANZFL), the promulgation of accreditation programs is a visible sign of the increasing importance of accreditation in forensic sciences [1, 2]. As an illustration, the ENFSI 2005–2008 strategic plan states that [3] "All ENFSI laboratories need support in complying with best practices and raising the level of quality and science. To achieve this, ENFSI continuously encourages its member institutes to move toward formal quality management and accreditation. Scientific cooperation within expert working groups needs more momentum and a competence assurance system for ENFSI laboratories will be developed." SMANZFL includes the promotion and enhancement of management and quality practices in their strategic plan (Davey A. *Personal communication 2008*).

There are several reasons for this growing phenomenon. One of them lies in the general tendency for all scientific laboratories to seek accreditation. Forensic DNA laboratories were among the first to follow this development. Simultaneously, the increasing number of private laboratories and customer demand for more accountability also has a great influence. Management systems introduced in the public sector are also undergoing similar changes. Finally, admissibility of scientific evidence is subjected to a constantly increasing scrutiny in the court. Accreditation has been accepted as a means of helping courts to make decisions about the reliability of the evidence provided by a crime laboratory [4].

Contrary to certification (*see* **Training and Certification (in Criminalistics)**), which deals with the evaluation of the competency of a forensic scientist, accreditation is concerned with the overall operation of the organization. It is an overview of the managerial and technical requirements of a laboratory in order to comply with a given standard.

Accreditation

Accreditation, in its broad meaning, can be defined as [5] "the formal recognition of the technical and organizational competence of an authority to execute a specific service as described in the scope of accreditation. Competence is the key to transparency, confidence, and comparability." More specifically, laboratory accreditation can be defined as the [4] "formal recognition that a testing laboratory is competent to carry out specific tests or specific types of tests". It may also include the [4] "recognition of both the technical competence and impartiality of a testing laboratory".

In other words, the entity that wants to be accredited agrees to structure its aims/goals, management system, and technical facilities (i.e., infrastructure, personnel, and methods) so that they meet preestablished agreed criteria (described in a standard). The accrediting body then verifies whether the laboratory seeking accreditation (also referred to as the *applicant*) fulfills the criteria and complies with the standard.

The partnership between the ISO and the IEC has led to the creation of the standard *ISO/IEC 17025 general requirements for the competence of testing and calibration laboratories* [6]. This standard has been chosen by many forensic/criminalistics laboratories around the world for their accreditation purposes. It describes in a very general but exhaustive manner what is necessary to comply with it. ISO/IEC 17025 includes five chapters covering the entire laboratory operation: scope, normative references, terms and definitions, management requirements, and technical requirements. The first three chapters explain cases in which the standard is applicable and references on which it is based. The management requirements chapter deals primarily with the operation and effectiveness of the quality management system in place in the laboratory. It explains how to describe the system in place, how to detail all stages of the working process, and how to deal with more specific situations. The technical requirements part addresses methodology, test/calibration equipment, and competency of personnel. The combination of management and technical requirements covers the entirety of crime laboratory operations.

Because ISO/IEC 17025 is designed for all kinds of testing and calibration laboratories, it does not necessarily consider all of the needs and specificities of a criminalistics laboratory. The application of standards to specific areas may require further explanations through additional documentation called *applications* or *supplementary requirements*. These documents are an elaboration of the more general criteria stated in the standards. Supplementary requirements provide an interpretation of the intent of parent standard (ISO/IEC 17025), when applied to specific disciplines

within forensic sciences. They are usually established with the help of specialists in the field, in the present case criminalists.

In the United States, in addition to ISO/IEC 17025 accreditation available through the ASCLD/LAB International program, the ASCLD/LAB Legacy program also accredits crime laboratories. As a matter of fact, more than 90% of US laboratories do not rely on ISO/IEC 17025, but on the ASCLD/LAB Legacy program [7].

In Australia, the National Association of Testing Authorities (NATA) provides a tailored forensic science accreditation program, which comprises ISO 17025 and supplementary requirements for accreditation in the field of forensic science for organizations whose core business is forensic science service provision. In addition, the "forensic module" is available to organizations that hold ISO 17025 accreditation in another field of testing but conduct a limited amount of forensic testing (Davey A. *Personal communication*, Parsell M. *Personal communication 2008*).

Two other international standards are of interest with regard to accreditation in criminalistics: *ISO/IEC 17020 general criteria for the operation of various types of bodies performing inspections* and *ISO 9001 quality management systems – requirements* [8, 9]. ISO 9001 describes the general requirements of a quality management system. Its application is, thus, much broader than is defined by 17025, as it applies not only to laboratories but also to any entity providing services or products. The managerial criteria defined in 17025 are based on the ISO 9001 requirements; however, the main difference between the two standards is that the former also includes criteria on technical requirements and competency. More recently, there have been many discussions regarding the application of ISO/IEC 17020 to criminalistics, more specifically to crime scene investigation [10], although this is not the case in all countries. However, although the application of 17025 for laboratories and 9001 for management systems is not contested, the use of 17020 for crime scene investigation units is still under discussion and has not been widely accepted yet. ENFSI is presently working on a joint project with the European Co-operation for Accreditation (EA) to define which standard, ISO/IEC 17020 or ISO/IEC 17025, is to be used for the different forensic fields [11]. The latest version of the ENFSI's *standards for accreditation* document states that [12] "For activities other than the testing part of the forensic process e.g. work at the scene of crime, ISO/IEC 17020 can be implemented as the standard used to achieve accreditation".

The main goals and requirements for all three international standards (ISO 9001, ISO/IEC 17020, and ISO/IEC 17025) are the same. One of the most important goals is oriented toward customer service. It is crucial to provide a satisfactory service or product, which is recognized as a basic condition by the customer. Another important requirement is documentation. First, documentation of the goals and procedures allows the accredited laboratory to establish a repository of knowledge existing within their laboratory. When training new employees, this comprehensive body of documentation is available and knowledge can be easily passed on. It should, however, be emphasized that this documentation does not replace in-house training but facilitates the whole process. Second, written documentation assists the laboratory to work in a transparent and traceable manner. Ultimately, this helps to establish customers' confidence. This accountability also renders the customers' task much easier when choosing between various laboratories to meet their requirements. Third, documentation helps to have an efficient quality management system by keeping a documented chain of custody, a particularly important requirement in criminalistics, and by keeping track of errors. Although a complete management system is more than just the chain of custody and corrective actions, several goals are reached by simply applying a comprehensive documentation policy.

Accreditation Process

There are several institutions that are competent to provide accreditations, most of which are national bodies. In the United States, ASCLD/LAB is the most popular. It is also the body chosen by 10 laboratories outside the United States, notably in New Zealand [13]. In Canada, the Standards Council of Canada (SCC) is competent. In Australia, it is the NATA and in the United Kingdom the United Kingdom Accreditation Service/National Accreditation of Measurements and Sampling (UKAS/NAMAS) can provide accreditation.

There are three main stages in the accreditation process: preparing for, obtaining, and maintaining

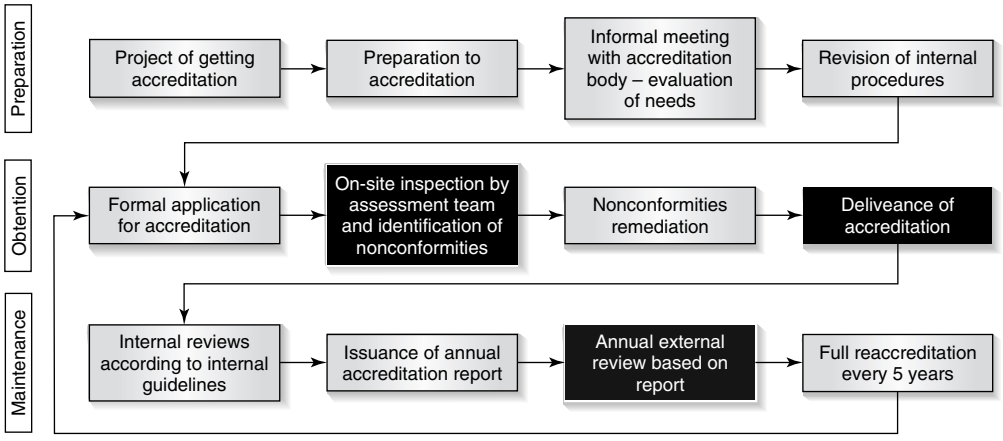


Figure 1 The different steps in the accreditation process. Light gray indicates the steps carried out by the body getting accredited. Dark gray indicates the steps carried out by the accrediting body. The blend of gray indicates the steps involving both the parties

accreditation.[a] Figure 1 shows the detailed process of accreditation as described hereafter.

One of the most important steps in preparing for accreditation is to devise a documented portfolio comprised of the aims/goals, management system, and technical facilities of the laboratory seeking accreditation. This includes the presentation of the laboratory's internal quality manual, which may already exist in one form or another, but which may need to be adapted to the exact needs of the process. The required documentation can usually be classified into three different levels. The most general level is the quality manual that describes, in general terms, the policies and goals of the work done by the laboratory seeking accreditation. It also describes the structure of the entity and its processes (in a general way) and the basic concepts of the (quality) management system. The second level includes guidelines and standard operating procedures (SOPs). These describe the exact processes used to achieve each task carried out at the laboratory and include information on organizational and technical know–how. Finally, the third level comprises detailed documentation, which may include checklists (or pro formas) to perform specific tasks and/or to maintain instruments as well as other specific documents.

Then, the accrediting institution will assign a technical officer who will conduct a preassessment of the laboratory. During this preliminary visit, discussions take place to review the existing measures and procedures in place. These measures and procedures are then evaluated as to their compliance with the requirements of the standard in question, such as ISO/IEC 17025 for a crime laboratory. Next, the laboratory will review its processes, documentation, and facilities in order to adapt to the standard and meet the accreditation requirements. The time required to carry out this step highly depends on what was already in place at the laboratory. If extensive adaptations must be incorporated, this step may be quite long and tedious. Depending on the accrediting body, an advisory visit may be necessary to address any noncompliances identified during the on-site visit.

To obtain accreditation, a formal inspection or assessment of the laboratory by a team of inspectors or assessors must take place. The previously prepared documentation is reviewed in light of the specifications described in the standard. The laboratory is also visited, and the accreditation inspectors scrutinize the operations of the laboratory. All nonconformities are noted and the appropriate corrective actions are discussed. Depending on the accrediting body, there may be two levels of nonconformities [14]. Level 1 nonconformities have a significant impact on the quality of the laboratory product and must be resolved before accreditation is granted. Level 2 nonconformities have a minimal impact and must be corrected prior to the next yearly visit. Some accrediting bodies, such as NATA, require that all conditions identified during the assessment must be satisfactorily addressed before accreditation can be considered.

Once obtained, accreditation must be maintained. To this effect, internal and external evaluation measures are put into place. Internal measures include the constant update of documentation on techniques, methods, and administration. External measures include annual surveillance visits, the submission of an annual accreditation report to the accrediting body, and results of proficiency tests. Depending on the accrediting body, a full reinspection of the laboratory may take place every five years [15]. NATA's surveillance program is a little different and includes 18-month surveillance visits and 36-month reassessment visits, with no annual reports required (Parsell M. *Personal communication*).

Accreditation Applied to Criminalistics

Criminalistics laboratories are not the usual testing and calibration laboratories (as covered by ISO/IEC 17025). Therefore, the propositions made in the international standards have to be interpreted according to the specificities of forensic sciences, in general, and criminalistics, in particular. As mentioned earlier, such a supplemental document is an interpretative document of 17025 and is called *an application document*. One such example is the International Laboratory Accreditation Cooperation (ILAC) *Guidelines for forensic science laboratories* (ILAC-G19:2002), which provides guidance for laboratories involved in forensic analysis and examination to comply with ISO/IEC 17025 [16] or the NATA supplementary requirements. These documents are structured in the same way as 17025. Its scope is illustrated by providing a specific list of activities that might be undertaken by a forensic laboratory, without being exhaustive. A succinct overview of the different techniques used is also provided. Reference is made to various relevant ISO/IEC guides. Some terms and definitions are also explained: objective test, reference collection, and court statement. Following this, the management requirements are specified. This includes not only the control of records (the aforementioned importance of documentation) but also the mention of negative results. The technical requirements include personnel (competency); accommodation and environmental conditions (e.g., contamination issues); test and calibration methods and method validation (how to carry out validation); equipment (maintenance and calibration); measurement traceability (documentation); sampling (also depends on competency of staff); handling of test and calibration items (chain of custody and evidence management); assuring the quality of test and calibration results (quality control measures); and reporting the results (preparation of the report and court testimony).

Although it is relatively straightforward to adopt ISO 17025 and its application (such as ILAC-G19:2002 or the NATA supplementary requirements) to fields such as toxicology, DNA analysis, or drug testing, it is much more challenging with disciplines such as handwriting, fingerprints, and other identification areas. One reason for this is the high similarity of activities between analytical forensic sections (DNA, drugs, toxicology, etc.) and regular analytical/pharmaceutical laboratories, since most processes can be described and handled in a standardized manner. Both fields are also largely numerical or quantitative in nature. Although extensive literature and practices exist describing the best way of handling exhibits for fingerprint enhancement, each incoming item has its particularities. Thus, the methodology cannot be applied in a completely standardized manner, although maintaining the integrity of the evidence does not change with the discipline involved – these procedures remain reasonably standard. This means that the description of the work done by the unit or laboratory is a more difficult task.

A further consequence of the particular and sometimes random nature of forensic samples is that it might be possible that the technique or method needed is not very commonly used. In turn, this renders it difficult to create or maintain standards for them. Likewise, it might be that a certain technique is only used once a year, which will not be sufficient to demonstrate a regular use and guarantee a routine use by laboratory personnel. This specific issue is covered by ILAC-G19:2002 and the NATA supplementary requirements.

The difficulty for the forensic community to adapt to external standards, such as 17025, is not only due to the "degree of subjectivity" of the techniques used but also due to the nature of the work performed, mainly because of the unique and varied character of each item of evidence. As the title of the ILAC application mentions, the guidelines are for forensic science laboratories. The question remains whether or not a crime scene unit falls under the definition of ISO/IEC 17025 (testing and calibrating laboratory). Until now the question has been raised

but no satisfactory answer has been found. In many countries, the tendency has been to accredit laboratories with ISO/IEC 17025 without considering the crime scene units. This is not true in some countries such as Australia where there has been considerable discussion on this issue and the mindset that "the laboratory" is four walls in a fixed building is seen as narrow and incorrect. As a result, in Australia, Crime Scene Laboratories are accredited to ISO 17025. In Switzerland, the question that has been asked is whether ISO 17020 (designed for inspecting bodies) would be appropriate for crime scene units [10]. Another way is to certify according to ISO 9001, which means that the managerial part, including the documentation, is covered, whereas the technical requirements are still not accredited, but will have to be covered by another form of quality control (internal or other entity than an accreditation body). Creating specific standards for crime scene accreditation remains a possibility [10].

Because of the existence of private or public laboratories and the differences of evidence admissibility in court, accreditation has not undergone the same development in different countries. At the same time, as previously mentioned, the development has also not been the same for the different stages of the criminalistic process. In Europe, the development is reflected by the guidelines established by ENFSI, which state that all their member laboratories [12] "should have achieved or should be taking steps toward ISO/IEC 17025 compliant accreditation for their laboratory testing activities". According to a recent survey of ENFSI members, one-third of the laboratories are accredited, one-third plan to become accredited by 2009, and one-third do not have a definite timeline or have not undertaken any action yet [17]. Nevertheless, 94% of the laboratories have a quality assurance system in place or are developing one. Furthermore, accreditation as perceived by a laboratory may not agree with the general understanding of accreditation as granted by an international institution or an official national body. Thus, only 20% of the laboratories surveyed are actually accredited according to the above-mentioned international standard ISO/IEC 17025 [17]. In the United States, 294 laboratories are accredited through the ASCLD/LAB Legacy system and 26 through the ASCLD/LAB International system that corresponds to ISO/IEC 17025 [13]. Only 18 of these laboratories are private. In Australia, approximately 85% of the government forensic science service providers (including crimes scene) are accredited to ISO 17025 and the majority of relevant private laboratories either have forensic science accreditation or are accredited with the "Forensic Module" or are moving in that direction (Davey A. *Personal communication*).

Advantages and Disadvantages

Accreditation offers several advantages: transparency, traceability, and accountability. This profits both the provider of a service and its clients. The laboratory will gain a much clearer understanding of what work is done and how it is done. This information is documented so that it is much easier to transmit to new employees who undertake on-the-job training. The client can have a much better view on how his product or service has been performed. It may also be easier for the client to choose which provider has the best offer to meet his needs. Along with this, accreditation leads by its structure to a certain harmonization between laboratories, as the conditions to fulfill are the same and ISO/IEC 17025 gives a common backbone to all laboratories. By rendering comparisons between laboratories possible, the quality of each of them can more easily be evaluated. An expansion on an international market, rather than a national one, is facilitated, as well. Furthermore, international collaboration based on criminalistics may be simplified by the common approach [17].

Among the first disadvantages, most commonly named are the initial cost of implementing all the measures needed and the permanent cost of maintaining accreditation. Sometimes it is also perceived that the extra paperwork and administrative effort needed is disproportionate to the gain resulting in implementing all the measures. A certain loss of flexibility in applying nonvalidated testing or implementing new techniques or methods has also been advanced as a negative point, as each and every change has to be documented according to procedures and accepted for court purposes. The quality of work may also become limited, as a tendency may prevail to comply only with the standards imposed by regulations, instead of using previously higher internal standards. Accreditation can also be used to limit competition with other laboratories by fixing high standards so that not all (private) laboratories can follow them. However, some of these criticisms may disappear when

it is understood that in reality accreditation sets the "minimum standard only" and not the top standard.

Conclusion

Accreditation, as part of a quality management system, is gaining more and more importance in criminalistics and has been fixed by professional bodies such as ENFSI as one of the goals to be achieved by their members. For criminalistics laboratories, the international standard that seems most appropriate is ISO/IEC 17025 for testing and calibrating laboratories. The specifications of how this standard can be applied to forensic/criminalistics laboratories are found in ILAC-G19:2002 and the NATA supplementary requirements. In the United States, most laboratories are still accredited according to ASCLD/LAB Legacy program. However, ASCLD/LAB also offers an ISO/IEC 17025 accreditation program, which will eventually replace the Legacy program. Future developments are turned toward discussions about how criminalistics/forensic units operating in the field (crime scene units) could comply with these international standards and become accredited. In this regard, the standard ISO/IEC 17020, designed for bodies performing inspections, is being considered by some people. However, one must be cautious about the fact that investigation and inspection are not synonymous.

Acknowledgments

The authors would like to thank Dr Sarah D. Brown for her editorial review of this article.

End Notes

[a.] A more detailed description of the whole process as performed by ASCLD/LAB International is available at http://www.ascld-lab.org/international/pdf/alpd3013.pdf, last visited 23 November 2007.

References

[1] European Network of Forensic Science Institutes (2007). *Welcome to the ENFSI Portal*, available at http://www.enfsi.eu.

[2] American Society of Crime Laboratory Directors (2007). *About ASCLD*, available at http://www.ascld.org/about/.

[3] ENFSI (2005). Strategic plan 2005–2008–Vision of ENFSI, *European Network of Forensic Science Institutes*.

[4] Smith, F.P. & Kidwell, D.A. (2000). Accreditation of Forensic Science Laboratories, in *Encyclopedia of Forensic Sciences*, J.A. Siegel, P.J. Saukko & G.C. Knupfer, eds, Academic Press, London, England, pp. 58–64.

[5] Swiss Accreditation Service (2007). *What is Accreditation*? available at, http://www.seco.admin.ch/sas/00026/index.html?lang=en:.

[6] ISO (2005). *ISO/IEC 17025:2005 General Requirements for the Competence of Testing and Calibration Laboratories*, International Organization for Standardization, Geneva, Switzerland.

[7] Stauffer, E., Dolan, J.A. & Newman, R. (2000). *Fire Debris Analysis*, Elsevier Academic Press, Burlington.

[8] ISO (1998). *ISO/IEC 17020:1998 General Criteria for the Operation of Various Types of Bodies Performing Inspections*, International Organization for Standardization, Geneva, Switzerland.

[9] ISO (2000). *ISO 9001:2000 Quality Management Systems – Requirements*, International Organization for Standardization, Geneva, Switzerland.

[10] Swiss Accreditation Service (2005). *Guide for the Assessment in the Field of Forensic Evidence Recovery ISO/IEC 17020:1998*, State Secretariat for Economic Affairs–Swiss Confederation, Bern, Switzerland.

[11] European Co-operation for Accreditation (2007). *Standards for Accreditation*, available at http://www.europeanaccreditation.org/content/news/cooperation.htm.

[12] ENFSI (2007). *Standards for Accreditation*, European Network of Forensic Science Institutes.

[13] American Society of Crime Laboratory Directors/Laboratory Accreditation Board (2007). *Laboratories accredited by ASCLD/LAB*, available at http://www.ascld-lab.org/legacy/aslablegacylaboratories.html.

[14] ASCLD/LAB (2004). *ASCLD/LAB-International Accreditation Program*, American Society of Crime Laboratory Directors/Laboratory Accreditation Board, Garner.

[15] Caddy, B. & Cobb, P. (2004). Forensic Science, in *Crime Scene to Court: The Essentials of Forensic Science*, P. White, ed, Royal Society of Chemistry, Cambridge, United Kingdom, pp. 1–20.

[16] ILAC (2002). *ILAC-G19:2002 Guidelines for Forensic Science Laboratories*, International Laboratory Accreditation Cooperation, Rhodes, NSW, Australia.

[17] Malkoc, E. & Neuteboom, W. (2006). The Current Status of Forensic Science Laboratory Accreditation in Europe, *Forensic Science International* **167**(2–3), 121–126.

BEATRICE SCHIFFER AND ERIC STAUFFER

Acid Phosphatase

Prostatic acid phosphatase (PAP) is an enzyme secreted by the prostate gland and secreted in seminal fluid. Acid phosphatase (AP) is a generic term for a group of isoenzymes, of which PAP is one. AP is not unique to the prostate (although PAP is), being found in other biological fluids such as vaginal secretions. However, AP activity is 50–1000 times greater in human semen than in any other body fluid, and so the test is regarded as a presumptive test for the presence of semen. The only approach to differentiating semen from vaginal secretion is by quantitative analysis or the microscopic inspection for the presence of sperm. In many circumstances, a strong positive reaction for AP indicates that semen is present and that further testing is warranted (e.g., for DNA).

There are a number of ways to test for AP. Most, especially those used to identify semen-stained areas, involve a color change. Clothing and swabs are probably the most frequently analyzed items.

PAP is found in the vagina following intercourse. Its activity was found to correlate better with the time since intercourse than the presence or absence of sperm.

According to Davies and Wilson, and using their test method, a reaction time of less than 30 s was regarded as a "very good indication" of the presence of semen as no vaginal AP reacted within this time.

In a series of postmortem examinations using ELISA methods, a level of PAP greater than $100\,\mathrm{ng\,ml^{-1}}$ was considered positive for sexual intercourse.

Importantly, a number of other materials can produce "false positive" reactions with AP (i.e., positive, but not semen). Examples include cauliflower, clover, bindweed, turnips, raisins, and ginger.

Further Reading

Collins, K.A. & Bennett, A.T. (2001). Persistence of spermatozoa and prostatic acid phosphatase in specimens from deceased individuals during varied postmortem intervals, *The American Journal of Forensic Medicine and Pathology* **22**, 228–232.

Davies, A. & Wilson, E. (1974). The persistence of seminal constituents in the human vagina, *Forensic Science* **3**, 45–55.

Moenssens, A.A., Henderson C.E. & Portwood, S.G. (2007). *Scientific Evidence in Criminal Cases*, 5th Edition, Thomson West, p. 997.

Ricci, L.R. & Hoffman, S.A. (1982). Prostatic acid phosphatase and sperm in the post-coital vagina, *Annals of Emergency Medicine* **11**, 530–534.

ALLAN JAMIESON

Action ***see*** Compulsion

Actuarial Risk Assessment ***see*** Risk Assessment: Patient and Detainee

Actus Reus ***see*** Automatism as a Defense to Crime

Acute Stress Disorder ***see*** Posttraumatic Stress Disorder

Adaptive Behavior ***see*** Mental Retardation

Adaptive Functioning ***see*** Mental Retardation: Death Penalty

Addictions

Definition

Addictive diseases are complex primary disease states caused by a combination of genetic, psychosocial, and environmental factors and generally resulting in both physical and behavioral manifestations [1] (*see* **Substance Abuse**). Their essential elements are often defined as continued use of a substance or similar maladaptive behavior, such as gambling, despite significant problems resulting from that use or behavior [2]. Extensive scientific research has explored the genetic underpinnings and the environmental contributors [3], the long-term physical findings generally resulting from substance use, and the behavioral differences observed in the addict [4]. For the purposes of this article, psychoactive substances are being used as a prototype addictive agent, but addictive disease also may pertain to other entities, gambling being one which has received diagnostic recognition [2]. Contributions to the literature have also explored other potential addictive agents, including sex, pornography, the Internet [5], videogame [6], and chocolate [7], though whether these represent medical disease states remains a controversial topic.

From a physiologic perspective, addiction takes place when a psychoactive drug is taken for a sufficiently long period of time and in sufficient doses to result in either tolerance or withdrawal phenomena. Tolerance is represented by the need for an individual to take more of a given substance to achieve a given effect, or by the decrease in perceived effect over time when a substance is taken in an unchanging dose. Withdrawal is represented by the presence of measurable physiologic change upon cessation of drug intake. Physiologic addiction need not be present for addictive disease to exist. Similarly, the presence of physiologic addiction does not alone imply the presence of addictive disease. Many psychoactive drugs cause measurable tolerance and withdrawal phenomena after only one dose and within a short period of time. A hangover, for example, after a single night of alcohol use, represents a withdrawal state. There are also longer-term effects of even brief exposure; for example, rats fed alcohol for 4 weeks have abnormal responses 6 months later when reexposed to alcohol [8]. The term *addiction*, as used by the general medical community, or *dependence*, as used by the psychiatric community, are rough equivalents that do not apply simply to physiologic addiction but to the well-characterized disease state involving use of substances that cause physiologic addiction. To wit, alcohol is an addictive drug. It causes physiologic addiction in all humans, and indeed in all mammals, who consume the drug. As a result, any individual who drinks alcohol may experience intoxication, and will experience tolerance and withdrawal to an extent consistent with the quantity and duration of alcohol use. Such experiences are neither indicative of the presence of addictive disease, nor are they necessary for such a diagnosis. Gambling is not a physiologically addictive drug, yet it has been accepted within the diagnostic rubric and leads to comparable behavioral consequences in some of those who participate in the activity.

Substance use itself can be categorized into several groups, using any number of classification schemes. Terminology has been inconsistent, with substance misuse, abuse, overuse, harmful use, addiction, and dependence all holding different meanings depending upon the individuals using the terms. The fourth edition of The Diagnostic and Statistical Manual of Mental Disorders uses the terms abuse and dependence as referring to condition states with respect to addictive substances. The background as to how these terms were developed and why they differ from generally used terminology has been addressed in the literature [9]. There is agreement that substance use refers to any use of addictive substances, whether such use is by prescription or within constraints of legal codes, or if illegal and of such quantity as to lead to intoxication.

Substance use, then, is a quantity-oriented term as it contains all those activities in which substances are used in any quantity other than zero. Addictive disease is not a quantity-oriented term; its presence involves some use during the illness, but neither the quantity nor the frequency of such use is germane to the diagnosis [1, 2].

Addiction within the psychiatric construct [2] can involve alcohol and other sedative agents, amphetamines, cannabis, cocaine, hallucinogens, inhalants, nicotine, opioids, phencyclidine, and steroids.

Once an individual has been diagnosed with addictive disease, the diagnosis remains permanently, just as the disease does, so long as the diagnosis was

retrospectively accurate. Addictive substance use can end as a result of remission, recovery, agonist therapy such as methadone, or a controlled environment such as prison. Where pertinent, such qualifications are added to the diagnosis.

Disease Course

Addictive illnesses are primary illnesses in that they are not symptoms of other disease states. It is natural that people try to determine causality as to why they have developed a disease, and many therefore choose to blame their addictive disease upon depression, anxiety, or specific life stressors. Addictive disease, however, has not been shown to arise as a symptom of any other illness.

Addictive diseases are involuntary in that afflicted individuals did not choose to have their illness, but this does not imply that those with addictive illness are relieved of the responsibility necessary to recover from their illness [1]. The diseases, without intervention, are often progressive and fatal. Patients often focus their attention upon the addictive substance or behavior, have adverse consequences resulting from the addiction, and deny the relationship between their addictive focus and the negative consequences.

In 1990, Block and Shedler published a key longitudinal study which revealed that addictive disease may be present, but for the use of drugs, at a very early stage of life [10]. Psychological differences between frequent drug users and others were noted as being traced to the earliest years of childhood and related to the quality of parenting received. The findings indicated that problem drug use is a symptom, not a cause, of personal and social maladjustment, and that the meaning of drug use can be understood only in the context of personality structure and developmental history. Disease course must therefore examine a complete individual history, not simply document the less important substance use history. More recent studies add the possibility that the problem drug use is not only a symptom of these environmental issues but of specific genetic abnormalities as well; no longitudinal study has yet been conducted, however, to confirm these theories.

At some point, the individual with addictive disease will use an addictive substance for the first time. This is intuitively more likely to take place within a society or subculture that encourages such use. Positive subjective experiences during the first use are more likely in those who have addiction [11]. This may be interpreted in one of two ways: either those with addictive illness are more likely to have the disease because they had a positive initial experience, or those with addictive disease have a physiologic abnormality that causes them to have a positive initial experience. Use of the substance increasingly takes on greater importance for the individual afflicted. Efforts at control may be instituted in which the addict attempts to limit his own use, but nevertheless as time passes, important social, occupational, or recreational activities become less important than the addictive behavior itself [2].

Whether through intervention or through self-awareness, addicts often pass through stages of change related to their ongoing behavior. Such stages include precontemplation, contemplation, and preparation, all of which precede any significant alteration of substance use, and are followed by action and maintenance [12, 13], both efforts to obtain and maintain abstinence. Discontinuation of the addictive behavior is a requirement during the action and maintenance phases. Stages may be, and often are, repeated, with relapses common after an initial attempt at abstinence. Cutting back, sometimes referred to as *controlled use*, is an ineffective treatment approach [14]. Continuation of any addictive behavior places the individual with addictive disease at risk; tobacco use, for example, not only increases the risks of morbidity and mortality due to smoking, but increases the risk of relapse of other substance use as well [15, 16]. Successful long-term recovery may accompany simultaneous long-term abstinence [17] during which no diagnostic signs of the illness remain.

Epidemiology

With respect to illness, the term *prevalence* refers to the percentage of a given population with a given diagnosis at a specific cross-section in time. Widely varying prevalences of addictive disease have been reported, in part because of differing application of definitions. Whether an individual in recovery was counted as one with the diagnosis or not depended upon the study. Some studies separated alcoholism from other drug addictions, again resulting in a

difficulty in determining even approximate figures. The Epidemiologic Catchment Area (ECA) study indicates a lifetime prevalence of alcohol disorders to be 23.8%, with a one-year prevalence of alcohol disorders to be 11.9% [18]. The National Comorbidity Survey (NCS) showed a lifetime prevalence of alcohol dependence to be 20.1% in men, with a one-year prevalence of 10.7% [19]. In women, the parallel figures were 8.2 and 3.7%. The differential between one-year and lifetime prevalence figures suggests that in these cases, the diagnosis was applied only to cases in which an individual is actively using substances. Clearly, though, these studies indicate how much more prevalent alcoholism is than other substance addiction; the ECA revealed an overall lifetime prevalence of drug abuse and dependence, not including alcohol, at 6.2%. Similarly, the NCS indicated a lifetime prevalence of drug dependence to be 9.2% among men and 5.9% among women.

Incidence refers to the risk that any one member of the study population will become afflicted with the disease under study in a specified period of time. In the case of addictive disease, incidence can be difficult to determine, as subjects often have used substances for some time prior to their precisely meeting the definition of illness. Retrospective reviews are therefore problematic. The most dependable findings are those referring to that population at greatest risk: men between 18 and 29 years of age [18].

Treatment

Addictive disease treatment is generally quite successful. The Federal Aviation Administration has demonstrated an 85% rate of success for alcoholic pilots attaining and maintaining recovery while being followed closely by physicians during a nine-year study period [20]. Military studies also indicate high rates of recovery over several years of treatment [21]. Similar research of healthcare professionals with substance use disorders has found similarly strong recovery rates [22]. Each of these studies involved ongoing treatment provided primarily by physicians. There have been many studies [23] indicating a far lower rate of success in which subjects generally have the same occupational or financial issues but in which patients are not provided with physician-provided medical care. No studies have yet been performed to directly compare and contrast outcomes among health care professionals and other clinicians.

Initial treatment focuses on the physiologic addiction to achieve a safe withdrawal from substance use. Cessation of the substance use is merely the first step in the treatment of addiction, however, and an ongoing maintenance plan is critical to a successful outcome [24]. Successful medical care generally involves a combination of therapeutic approaches, treatment of any physical and behavioral comorbidities, and the active involvement by the patient in 12-step recovery groups. Pharmacotherapy has proven useful for treatment of opioid [25] and nicotine [26] dependence. Disulfiram (Antabuse) remains the only medication demonstrated to have long-term value in the treatment of alcoholism [27]. Treatment is entirely ambulatory in many cases, but in more severe cases involves inpatient stays at various levels of care, including hospitalization, rehabilitation, partial day programs, and halfway houses.

Relapse Prevention

Once patients with addictive disease have stopped their addictive behavior, they now begin to feel the discomfort representative of the disease state – the very discomfort that they gained relief from through the use of substances. Abstinence merely solves the physiologic addiction. Prevention of relapse can be obtained only by addressing the discomfort that is present for a sober addict. There are four common myths regarding relapse [28]:

- **The relapse takes place because the patient wants to relapse:**

It is common for treatment professionals to accuse the patient of a moral lapse leading to relapse, as if the relapse is the desired state for the patient. The patient wants only to feel better. A relapse is the best possible alternative the patient can conceive of if the relapse has taken place. One role for the treating physician is to provide alternative options. Punitive policies such as those held by some programs that administratively discharge a patient following a relapse are inappropriate. Patients who relapse require more care, not less.

- **The patient needs to hit bottom before recovery is possible:**

No empirical data exist to support this.

- **Relapse means the patient has begun using substances again:**

Relapse begins long before the patient starts using again. The path to relapse is progressive, with the actual use coming as the final step on the path. Self-help groups often refer to the term *dry drunk* as meaning that an individual is thinking in the same manner as one who is actively using.

- **Once a relapse has taken place, complete loss of control will result:**

Relapses often consist of a brief episode of use after which the patient suddenly and quickly returns to sobriety. These relapses often follow a period in which the patient has dropped out of treatment or out of self-help groups. The patient then denies any further difficulties until the relapse takes place. Then the individual recognizes that indeed the disease has continued despite the past time of recovery; a return to therapy follows. Such a course is not at all atypical.

An individual's risk of relapse may fall within any of three groups [29]:

- **Recovery prone**

Forty percent of addicts attempting recovery fall into this group. Some attain sobriety with no clinical intervention and no attendance at self-help groups. Others seek such assistance and remain sober following this initial intervention.

- **Transitionally relapse prone**

Twenty percent of all patients periodically relapse, generally within treatment, but as time passes their relapse episodes become less severe, with shorter durations and greater time periods separating episodes. These patients often enter a long-term sobriety within 3–5 years.

- **Relapse prone**

This group of 40% are thought to develop progressive patterns of more severe episodes. Levels of functioning decrease during periods of abstinence. These patients often die of their illness within the first two decades of treatment. This group can be subdivided into those with motivation and those without. The group without is unlikely to present for ongoing treatment. The group with motivation will dutifully try, participating in treatment and self-help, but eventually failing to succeed.

Conclusion

As with many other diseases, addictive illnesses have a wide spectrum of severities with an accompanying range of related morbidities. Quantity and frequency of substance use, however, are independent of severity [30] and are therefore neither predictive nor useful measures of disease course. Terminology has caused significant difficulties in the field. Alcoholism, the most common addictive disease, has long been a reference to an illness that involves use of any sedative agent that exhibits cross-tolerance with alcohol. Alcohol dependence, the syndrome originally defined in 1976 [31], has come to focus specifically upon alcohol use; other sedatives are included within the "Sedative Dependence" section of DSM-IV. As a result, studies of alcoholism and alcohol dependence differ substantively, resulting in findings for one group that may not apply to the other. There are also many studies involving heavy use of alcohol in which readers might presume incorrectly that findings apply to those with alcoholism. The literature must be analyzed closely and critically to determine whether findings in any one study or group of studies are applicable to any specific individual.

References

[1] Morse, R.M. & Flavin, D.K. (1992). The definition of alcoholism, *JAMA* **268**, 1012–1014.

[2] American Psychiatric Association (2000). *The Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition, Text Revision (DSM IV-R), American Psychiatric Association, Washington, DC.

[3] Sher, K.J., Grekin, E.R. & Williams, N.A. (2005). The development of alcohol use disorders, *Annual Reviews of Clinical Psychology* **1**, 493–523.

[4] Oscar-Berman, M. & Marinkovic, K. (2007). Alcohol: effects on neurobehavioral functions and the brain, *Neuropsychology Review* **17**(3), 239–257.

[5] Meerkerk, G.J., Van Den Eijnden, R.J. & Garretsen, H.F. (2006). Predicting compulsive Internet use: it's all about sex!, *Cyberpsychology and Behavior* **9**(1), 95–103.

[6] Grüsser, S.M., Thalemann, R. & Griffiths, M.D. (2007). Excessive computer game playing: evidence for addiction and aggression? *Cyberpsychology and Behavior* **10**(2), 290–292.

[7] Bruinsma, K. & Taren, D.L. (1999). Chocolate: food or drug? *Journal of the American Dietetic Association* **99**(10), 1249–1256.

[8] Gitlow, S.E., Bentkover, S.H., Dziedzic, S.W. & Khazan, N. (1973). Persistence of abnormal REM sleep

response to ethanol as a result of previous ethanol ingestion, *Psychopharmacologia* **33**, 135–140.
[9] Saundersa, John.B. & Linda, B. (2007). Cottlerc the development of the diagnostic and statistical manual of mental disorders version V substance use disorders section: establishing the research framework, *Current Opinion in Psychiatry* **20**(3), 208–212.
[10] Shedler, J. & Block, J. (1990). Adolescent drug use and psychological health: A longitudinal inquiry, *The American Psychologist* **45**(5), 612–630.
[11] Acosta, M.C., Eissenberg, T., Nichter, M., Nichter, M. & Balster, R.L. (2007). Characterizing early cigarette use episodes in novice smokers, *Addictive Behaviors*. DOI: 10.1016/j.addbeh.2007.09.005.
[12] Prochaska, J.O., Norcross, J.C. & DiClemente, C.C. (1994). *Changing for Good: a Revolutionary Six-stage Program for Overcoming Bad Habits and Moving Your Life Positively Forward*, William Morrow and Co, New York.
[13] Ramos, D. & Perkins, D.F. (2006). Goodness of fit assessment of an alcohol intervention program and the underlying theories of change, *Journal of the American College Health* **55**(1), 57–64.
[14] Pendery, M.L., Maltzman, I.M. & West, L.J. (1982). Controlled drinking by alcoholics? New findings and a reevaluation of a major affirmative study, *Science* **217**(4555), 169–175.
[15] Hser, Y.I., McCarthy, W.J. & Anglin, M.D. (1994). Tobacco use as a distal predictor of mortality among long-term narcotics addicts, *Preventive Medicine* **23**(1), 61–69.
[16] Kelly, M., Chick, J., Gribble, R., Gleeson, M., Holton, M., Winstanley, J., McCaughan, G.W. & Haber, P.S. (2006). Predictors of relapse to harmful alcohol after orthotopic liver transplantation, *Alcohol Alcohol* **41**(3), 278–283.
[17] Gitlow, S. (2007). Recovery and research: a better paradigm, *Journal of Substance Abuse Treatment* **33**(3), 277–278.
[18] Helzer, J.E., Burnam, A. and McEvoy, L.T. (1991). Alcohol abuse and dependence, In *Psychiatric Disorders in America: The ECA Study*, L.N. Robins & D.A. Regier eds, The Free Press/MacMillan Inc, New York, 81–115.
[19] Kessler, R.C., McGonagle, K.A., Zhao, S., Nelson, C.B., Hughes, M., Eshleman, S., Wittchen, H.U. & Kendler, K.S. (1994). Lifetime and 12 month prevalence of DSM-IIIR psychiatric disorders in the US, *Archives of General Psychiatry* **51**, 8–19.
[20] Russell, J.C. and Davis A.W. (1985). *Alcohol Rehabilitation of Airline Pilots*. NTIS Technical Report AD-A163076. Federal Aviation Administration, Office of Aviation Medicine, Washington, DC.
[21] Wright, C., Grodin, D.M. & Harig, P.T. (1990). Occupational outcome after military treatment for alcoholism, *Journal of Occupational Medicine* **32**(1), 24–32.
[22] Domino, K.B., Hornbein, T.F., Polissar, N.L., Renner, G., Johnson, J., Alberti, S. & Hankes, L. (2005). Risk factors for relapse in health care professionals with substance use disorders, *JAMA* **293**, 1453–1460.
[23] Anton, R.F., O'Malley, S.S., Ciraulo, D.A., Cisler, R.A., Couper, D., Donovan, D.M., Gastfriend, D.R., Hosking, J.D., Johnson, B.A., LoCastro, J.S., Longabaugh, R., Mason, B.J., Mattson, M.E., Miller, W.R., Pettinati, H.M., Randall, C.L., Swift, R., Weiss, R.D., Williams, L.D. & Zweben, A. (2006). Combined pharmacotherapies and behavioral interventions for alcohol dependence, *JAMA* **295**, 2003–2017.
[24] Mackay, P.W. & Marlatt, G.A. (1990–1991). Maintaining sobriety: stopping is starting, *The International Journal of the Addictions* **25**(9A-10A), 1257–1276.
[25] Collins, G.B. & McAllister, M.S. (2007). Buprenorphine maintenance: a new treatment for opioid dependence, *Cleveland Clinic Journal of Medicine* **74**(7), 514–520.
[26] Glover, E.D. & Rath, J.M. (2007). Varenicline: progress in smoking cessation treatment, *Expert Opinion on Pharmacotherapy* **8**(11), 1757–1767.
[27] Gitlow, S. & Gold, M.S. (2007). The inadequacies of the evidence, *Addiction Professional* **5**, 17–25.
[28] Daley, D.C. (1987). Relapse prevention with substance abusers: clinical issues and myths, *Social Work* **32**(2), 138–142.
[29] Gorski, T.T. (1986). Relapse prevention planning, a new recovery tool, *Alcohol Health and Research World (Fall 1986)* 6–11, 63.
[30] Gitlow, S.E. (1979). The disease of alcoholism, *Cancer Research* **39**, 2836–2839.
[31] Edwards, G. & Gross, M.M. (1976). Alcohol dependence: provisional description of a clinical syndrome, *BMJ* **1**, 1058–1061.

STUART GITLOW

Admissibility: Expert Opinion, USA *see* Expert Opinion: United States

Admissibility of Expert Opinion Evidence: a Comparison of Approaches *see* Expert Opinion in Court: a Comparison of Approaches

Admissibility of Expert Opinions Evidence: Civil Law Jurisdictions in France, Germany, Italy, and Spain ***see*** Expert Opinion in Court: Civil Law Jurisdictions (France, Germany, Italy, and Spain)

Admissibility of Expert Opinion Evidence in the UK, Canada, and Australia ***see*** Expert Opinion: United Kingdom, Canada, and Australia

Admissibility of Expert Opinion Evidence in the United States ***see*** Expert Opinion: United States

Admissions of Guilt ***see*** Confessions: Evidentiary Reliability of

Adults: Suggestibility of ***see*** Eyewitness: Suggestibility of

Advance Directives ***see*** Therapeutic Jurisprudence

Adversary Systems of Justice

Legal systems through the world may be classified as belonging in one of two philosophical camps: those that are called *adversary systems of jurisprudence* and those that are typically referred to as *civil law systems* (*see* **Civil Law Systems of Justice**).

Adversary systems of jurisprudence, also referred to as *common law systems*, are in use in that part of the world where English-speaking rule exists or where the English legal system was the predominant jurisprudential system at one point in time. This includes the United Kingdom, the United States of America, Canada, Australia, New Zealand, and some other countries whose juridical systems were created at a time when the Anglo–Saxon tradition of the common law inspired their development.

The adversary system is one that focuses on a fact-finding trial, which is a contest between opposing sides in litigation. In this contest, each side seeks to convince the fact finder that their contentions have merit by offering supporting evidence. The fact finder in such a system is, alternatively, the presiding judge if it is a bench trial (nonjury trial), or a jury of the defendant's peers if a jury trial is permitted under local or constitutional law. In a jury trial, the judge's role is to preside over the trial, rule on matters of law and on objections interposed by counsel for the litigants, and instruct the jurors on the legal principles they should apply to the facts that they have determined, from the evidence presented, to have occurred.

Though there exist many variants in the rules of various jurisdictions, presentation of proof in adversary or common law system countries is subject to an often complex system of rules of evidence. These rules are designed to limit the types of proof that courts will admit. The rules of evidence are designed to assure that the fact finder will base the

outcome of the litigation on mostly appropriately reliable evidence. They also exclude, at times, reliable evidence that, for policy reasons, courts or legislatures have determined to be worthy of special protection and are therefore shielded from disclosure. This would include privileged communications between well-defined classes of individuals, such as communications between doctors, lawyers, clergy, spouses, and parties to civil litigation or defendants in criminal prosecutions.

Related Articles

Chain of Possession of Tangible Evidence

Civil Law Systems of Justice

Expert Opinion in Court: a Comparison of Approaches

Hearsay Evidence

***In Limine* Motions and Hearings**

ANDRE MOENSSENS

Affirmative Defense *see* Automatism as a Defense to Crime

AFIS Storage and Retrieval of Fingerprints *see* Automated Fingerprint Identification System

Age Determination *see* Anthropology: Age Determination of Remains, Anthropology: Aging the Living

Age Determination of Documents *see* Dating: Document

Age Determination of Skeletal Remains *see* Anthropology: Age Determination of Remains

Age Determination of the Living *see* Anthropology: Aging the Living

Aggression

Introduction and Definitions

> *The extreme form of power is All against One, the extreme form of violence is One against All* [1], p. 42.
>
> *The preeminence of aggressiveness in our civilization would already be sufficiently demonstrated by the fact that it is usually confused in everyday morality with the virtue of strength* [2], p. 98.

In ordinary language, the concepts of aggression and violence have taken on broad, often metaphorical, meanings. Diseases may be treated aggressively by physicians. A storm may be violent. An aggressive attitude is necessary in training for competitive sports. These common usages suggest the need to define aggression and violence precisely in the setting of legal proceedings and the scientific study of behavior in its social context.

Aggression refers to a state of destructive intent, with or without harm. It has been studied in insects and animals [3], with an emphasis on biological and evolutionary factors in behavior [4, 5]. Research on human aggression, in contrast, takes into account the

additional specificity of social and cultural processes. The impact of civilization on instinctual drives may largely override biological determinants, making human aggression a phenomenon qualitatively distinct from aggression in animals. In its mildest forms, such as irritation or brief moments of anger, human aggression remains within the range of universal, normative experience. According to Bronfenbrenner and Ricciutti, aggression includes "any action, thought or impulse the presumed aim of which is physical or psychological injury either real or symbolic to an individual or his surrogate" (1960, cited in [6]). In this definition, aggression thus reflects an internal state that manifests in a wide spectrum of behaviors. It is characterized by harmful intent, with destructive feelings toward others as well as objects. Aggression may include "[i]nsults and spreading harmful rumors" [7], p. S7. Anger, fear, and irritation are emotional states often associated with aggression, but are not synonymous with it. Frustration may result in anger that does not involve destructive intent of any kind, and may be resolved through verbal expression. As Rothenberg stated in 1971, "Violence and revenge are destructive direct discharges, but they are not expressions of anger *per se*; they are in part expressions of failed or unattempted communication" [8].

Violence involves physical injury or imminent threat of physical injury to another human being. The Centers for Disease Control (CDC) in the United States conceptualizes violence under the larger public health problem of physical injury, differentiating unintentional injury (accidents) from intentional injury toward self and others [9]. The World Health Organization (WHO) defines violence broadly as "The intentional use of physical force or power, threatened or actual, against oneself, another person, or against a group or community, that either results in or has a high likelihood of resulting in injury, death, psychological harm, maldevelopment or deprivation" [10], p. 5. Although this definition includes violence to oneself and collective violence, researchers have operationalized violence more narrowly as intentional interpersonal acts resulting in injury of another, in order to focus the object of empirical study. The categorization used by the WHO divides interpersonal violence into family and intimate partner violence on the one hand and community violence (between unrelated persons) on the other hand [10], p. 6.

The authors of the MacArthur Study of Mental Disorder and Violence defined two types of violent incidents as outcome variables: violence and other aggressive acts, distinguishing violence as "the infliction of injury or the threat of considerable, credible harm" [11], p. 18. The authors of the Historical Clinical Risk-20 (HCR-20), one of the most extensively validated rating instruments for structured clinical judgment of violence risk, define violence as actual or attempted harm to a person, or a threat of harm that is unambiguous [12], p. 24. Acts that would induce fear in most people but that do not actually result in injury, such as nonconsensual attempted physical sexual contact or stalking (*see* **Stalking**), are included in this definition. A general guideline according to the HCR-20 is to include as violence any behavior that would be "serious enough to result in criminal or civil sanctions" [12], p. 25 and to exclude from the definition those acts that would not. Verbal insults without overt threat of harm, destruction of property, and harm to animals without intended harm to persons, appropriate self-defense, and sporting injuries are excluded from their definition [12], pp. 25–26.

From a historical and sociological perspective, violence broadly encompasses phenomena such as war, genocide, and revolution. Writing from the point of view of a psychiatrist with extensive experience in the correctional system, Gilligan elaborates on the idea that violence serves either to "achieve and maintain justice, or to undo or prevent injustice" [13], p. 12, and that acts of violence can be construed as a reaction to injustice, either real or perceived [13], pp. 18–19. In his formulation of violence as a medical and public health problem, he goes on to extend its definition beyond interpersonal violence to include accidents, which do not involve the specific intent to harm. This latter definition of violence stems from a conceptualization of *structural violence* as a form of harm that is built into a social system, and for which there is collective responsibility [14]. In this sense, socioeconomic disadvantages such as poverty and large-scale prejudice such as racism are manifestations of structural violence contributing to direct violence, in the sense of preventable deaths [13], p. 192+. The associated problems of war, genocide and torture may become relevant to legal proceedings in international courts of law or in immigration proceedings and merit further discussion, which is beyond the scope of this article.

In addition to structural and direct forms of violence, Galtung defined a third major type in 1990, *cultural violence*, which refers to a system of values that

reinforces or legitimizes both structural and direct violence. In his causal model, structural, cultural, or direct violence can each generate either of the other two types of violence [14]. Although the conceptualization of structural violence is valuable in emphasizing that, with regards to public health outcomes, many events resulting in physical injury are not the result of direct action by individuals with intent to harm others, the empirical study of violence requires differentiation of the phenomena in order to examine them in a meaningful manner. As noted above, for the purposes of research, interpersonal violence is generally distinguished from collective and self-inflicted violence. Different methodologies have been necessary, with reference to the specificities of each type of violence, particularly with attention to the role of human agency and the intent to cause direct harm to another person.

This heuristic approach, differentiating multiple types of violence in order to study them separately, has created problems for the generalizability of findings. Even within the restricted field of interpersonal violence, studies have been organized according to the characteristics of specific kinds of violence, resulting in groupings that provide useful information regarding each subtype of violence but which are lacking in the consistency needed for a general or unified theory. An unresolved question is whether interpersonal violence indeed constitutes a coherent whole or whether the word "violence" is used to describe disparate phenomena that have distinct causation and risk, even if the outcome, physical injury, is the same [15]. Recent attempts at synthesis suggest that child abuse, adolescent violence, and domestic violence share common risk factors and that the phenomena are interrelated [16].

Studies of assaultive behavior have proceeded by distinguishing *different classes of victims* in order to constitute the phenomenon to be characterized, such as child abuse [17], intimate partner abuse [18, 19], violence toward strangers [20], violence among gang members [21], abuse of the elderly [22], victimization of the mentally ill [23–28], targeted violence toward persons in positions of authority (police or politicians [29]), and sexual offenses, where the victims are children [30, 31], adult women [32, 33], or adult men in the context of prison [34, 35]. Another approach has involved the study of violence according to the *location where it occurs*, for example, at home [36] (domestic or intimate partner violence), at school [37, 38], in prison [39, 40], in the psychiatric hospital on an inpatient unit [41, 42], or in the workplace [19, 43–45]. Other studies have focused on the *identity of the perpetrator* by studying violence committed by specialized populations: severely mentally ill patients [11, 46–49], individuals not mentally ill but who meet criteria for psychopathy (*see* **Psychopathy**), women (*see* **Aggression: Gender Differences in**), children [50], adolescents [51–53], individuals with organic brain disease [54, 55] or low IQ [56, 57], military veterans [58], and substance abusers [59–61].

In addition to the above, features that distinguish types of interpersonal violence include characteristics of the act itself. Violent acts have been distinguished according to whether they are *impulsive* (driven by an immediate emotional reaction) *versus premeditated* (planned through thinking in advance) [62, 63]. Studies have also focused on the *magnitude and frequency* of violent behavior, distinguishing between offenders who engage in acts of serious violence (e.g., characteristics of homicide offenders [64, 65]) and those who repeatedly commit low-magnitude assaults.

Homicide refers to the most extreme and irreversible consequence of violence: the loss of life as a result of physical injury by another (see [66]). Although some studies have characterized differences between spousal homicide offenders and perpetrators of lower-magnitude domestic violence [67], the extent to which homicide should be considered a phenomenon fundamentally distinct from assault remains unclear at this time. Further research is needed to determine the frequency with which homicides are the unintended consequence of assaultive behavior.

Behavioral Manifestations

Aggression can be viewed in terms of gradations of severity, ranging from the inner state of irritation, hostility, or anger, to physical injury of another person, to death, and can be examined through outward manifestations in behavior. In its mildest forms, aggression may be expressed through passive-aggressive behavior (avoidance, subtly hostile comments, and neglectful behavior that interferes with task completion), which is rarely the basis for legal proceedings, unless poor work performance has led to a fitness for duty evaluation. An individual who is experiencing aggressive feelings may be aware of

violent fantasies or intense hatred, or he may not have conscious awareness of hostility at all, expressing these feelings instead through maladaptive forgetting of important meetings, complaining, procrastination, or other indirect means.

Escalation

Overt behavior indicative of escalating aggression includes verbal expression of anger, which may involve threats that are stated in either explicit or implicit form (veiled threats). Physical gestures, such as finger pointing, shouting, pacing, or intrusiveness are relevant indicators of increasing aggression and hostile intent, along with confusion and irritability [41]. Other behavioral manifestations of escalation include obtaining a weapon, handling or moving a weapon [68], and aggressive gestures toward objects, such as pounding a fist on the table or the destruction of property [41]. Preparatory behaviors indicative of aggressive intent for sexual crimes may not appear overtly hostile, since these crimes are often organized by the perpetrator in a surreptitious fashion, involving seduction and "grooming" of the victim, or secretive observation of the victim's daily patterns of movement. Acts of violence may also be sudden, impulsive, and without premonitory signs that the individual is moving towards action with harmful intent.

The Act

Gilligan [13] and Junginger [69] each argue that analysis of the specific characteristics of the act may reveal a dimension of violence as "symbolic language" that serves as a form of communication. In addition, some acts of violence may serve as an escape from meaning or a failure in symbolization that later on, afterwards, leads others to interpret the act, even though at the time, it was an act that was "senseless" or beyond any intent to communicate a message (see [70, 71]). As previously noted, acts of violence can be analyzed according to whether they are predominantly impulsive (driven by an immediate emotional reaction) or premeditated (involving a purpose known in advance). Many instances of violence may remain unknown to public authorities, particularly in the context of domestic disputes.

Responses to the Act

After the immediate management of the aftermath of violent behavior, which includes medical stabilization of vital functions and treatment of injuries, a variety of reactions can be anticipated concerning each of the involved parties. The *perpetrator* of violence may or may not experience extreme distress at his own act. If the victim was a loved one who has suffered serious injury or death, or in the context of apprehension by law enforcement, the perpetrator may experience an elevated risk of suicide in the time following the act (*see* **Suicide (Behavior)**). The perpetrator may also have memory loss for the act itself [72], either due to failure to register the memory [73], in the context of an altered state of consciousness during and after the act ("red out") [73, 74], or he may claim amnesia due to malingering (*see* **Malingering: Forensic Evaluations**). Perpetrators of violence may experience flashbacks and other symptoms of posttraumatic stress [75]. The presence of posttraumatic stress disorder symptoms (*see* **Posttraumatic Stress Disorder**) in perpetrators of serious violence is increasingly recognized among psychiatrists and may be the focus of treatment [76].

Posttraumatic stress disorder has been more extensively recognized in the *victims* of violence. Ongoing, repeated victimization may result in chronic syndromes described in legal proceedings as battered spouse syndrome (*see* **Battered Spouse Syndrome**) or sexual abuse accommodation syndrome (*see* **Child Sexual Abuse Accommodation**). Mental health professionals should be alert to the possibility that participation in legal proceedings may itself induce distress in the victim, beyond that arising from the act of violence. Bystanders or others who were not actually present at the time of the violent act may also experience emotional distress in relation to the act.

Sequelae

In cases that are prosecuted, one major consequence is punishment of the offender, who may serve time in prison, in jail, or be constrained by home arrest or electronic monitoring. Whether or not the offender is prosecuted, victims and close relations of the victim or the perpetrator may experience anticipatory fear of future violence. The violent incident may have a profound impact on the relationships between the victim and perpetrator and with other family

members, resulting in social isolation. On the other hand, in some communities, participation in an act of violence may be valorized by the group, such as in rites of initiation in gangs or organized crime. In these contexts, violence may serve the symbolic purpose of establishing the perpetrator's social identity and his or her claim on valued relationships or property. For the victim, enduring physical disability may cause significant impairment and serve as a constant reminder of past trauma.

Sources of Information on Prevalence

The most comprehensive prevalence data regarding violent behavior is collected by governmental agencies, though each of these has limitations and they use different definitions of violence. The subjective experience of aggression, hostility, and anger is nearly universal, however, and its prevalence is therefore difficult, if not impossible, to measure.

At the international level, the most comprehensive study to date was released by the WHO in 2002 [10]. Although this report demonstrated that the United States has a much higher level of homicide and physical assault when compared with other developed countries, the homicide rate in the United States is substantially lower than in underdeveloped regions, even when deaths due to war are excluded [77].

The CDC in the United States conducted a survey in 1996, the National Violence Against Women Survey, in collaboration with the National Institutes of Justice and the National Center for Injury Prevention and Control, in which 8000 women and 8005 men were interviewed with regards to their fear of and experience of violence [78]. The CDC is currently piloting a new study, the National Intimate Partner and Sexual Violence Survey, to provide national and state estimates of the incidence and prevalence of all forms of intimate partner abuse, sexual violence of all types, and stalking [79]. In addition, the CDC maintains continuous collection of data from 17 states in its National Violent Death Reporting System, which began in 2002 with the aim of combining data from multiple sources in order to provide a comprehensive and detailed picture of the circumstances in which a violent death has occurred, and the relationship of the perpetrator and victim [80].

The Federal Bureau of Investigation (FBI) has maintained crime statistics on homicide, forcible rape, robbery, and aggravated assault since the introduction of its Uniform Crime Report program in 1929 [81]. From these data, the FBI generates a yearly report called *Crime in the United States*, compiled from data from over 94% of US jurisdictions, based on monthly reports from state agencies. The US Bureau of Justice Statistics analyses the FBI data and issues yearly and multi-year reports regarding violent crime [82]. Although most homicides and attempted murders do come to the attention of the legal system, crime statistics based on prosecution and convictions are biased in that numerous acts of lower-magnitude violence are reported neither to the criminal justice system nor to medical facilities. The National Crime Victimization Survey (NCVS) is an ongoing data collection project that addresses this bias in law enforcement statistics [83]. Begun in 1973, it underwent a methodological revision that was completed in 1993 in order to improve the wording of questions particularly regarding sexual assault. In its current form, the NCVS collects data on a yearly basis in interviews of a nationally representative sample of approximately 77 200 households, including approximately 134 000 individuals [84].

Many other countries maintain similar databases. Comprehensive statistics regarding mental illness in the perpetrators and victims of violent crime are more difficult for government sources to obtain in a systematic manner. Prevalence information can also be found in articles referenced above under different categories of victims and offenders, though care should be taken to critically examine differences in the sampling methods and definitions of violence.

Causes

> *What a potent obstacle to civilization aggressiveness must be, if the defence against it can cause as much unhappiness as aggressiveness itself!* [85] p. 143.

Human aggression and its behavioral expression as violence comprise a complex interweaving of biological, individual, interpersonal, and social or environmental factors, which cannot easily be summarized here (see [86]). The person who engages in an act of violence may have some awareness of aggressive thoughts and feelings, but the determinants of a decision to take violent action, rather than to use other

strategies for handling aggression, are challenging to identify with precision.

Elias describes a societal process of decreased acceptability of aggressiveness and identifies the end of the Middle Ages as a time when new mores and social practices resulted in the "repression" of aggressive behavior through education and internalized values [87]. The incursion of this "civilizing process" on biological drives was the source, according to Freud, of many modern neurotic symptoms [85]. Scientific and policy debate in the 1960s and 1970s led scientists to issue the Seville statement [88], which counters the claim that violence is biologically inherent in the human organism.

The empirical study of causation of violence is methodologically complex. Relevant variables associated with violence are numerous and interact with one another [89]. Furthermore, the association of variables with an outcome of violence in a study population does not clearly lead to explanatory power in individual cases: the presence of known risk factors does not prove, in a causal manner, that these risk factors led up to a specific act of violence.

Violence is commonly thought of in association with men, though gender differences in aggression have been overestimated and misunderstood (*see* **Aggression: Gender Differences in**). Although all age groups are affected by violent victimization and offending, individuals between the ages of 20 and 24 have the highest rates of homicide victimization and offending [66]. The impact of exposure to forced institutional settings, such as present or past imprisonment, is not directly assessed in currently available rating instruments. The biological study of aggression at a molecular level in animals and humans has improved knowledge regarding neurotransmitter systems and the genetic contribution to aggressive behavior [90], though the integration and application of this knowledge to treatment in humans is in a rudimentary but promising state at this time [91, 92].

Risk factors are commonly categorized as *static* (related to historical elements of the life story that are unlikely to change) and *dynamic* (modifiable) factors. The HCR-20 [12] includes items such as substance use and major mental illness, which may be assessed according to both static and dynamic components. The use of biographical information and the clinical interview to identify static and dynamic risk factors is an approach that may be called *subjective* in the sense that the risk assessment integrates objectively verifiable data and the individual's own report of his reaction to the environment in a manner that is specific to the person being evaluated. Subjectivity, however, in regards to the individual's emotional experience of and attribution of meaning to his past or current life events and situations is not generally measured within validated assessment instruments, even though it is taken into account in the clinical risk assessment interview and plays a significant role in treatment.

A recent emphasis on dynamic risk rather than on risk status (based on unchanging past features of the person's history) has led to the conceptualization and study of *risk state* as a time-dependent convergence of transient and modifiable risk factors [89]. In their 2005 review, Douglas and Skeem called for theory development and recommended that a theoretical model of the individual's risk state over time include a model of "relational complexities" of risk factors leading up to violent acts. They also recommended that theory development include an account of strategies for reducing violence in persons at elevated risk [89], pp. 367–368. The formulation of an explanatory theory would require the integration of known risk factors into a model of their interrelation, which may vary widely from individual to individual. Case-specific risk factors and configurations of risk based on the person's life history are currently not discernible in statistical models using actuarial methods, as these elements emerge only through detailed study of particular cases and may not rise to a level of generalizable patterns within large populations. Qualitative methods have been useful in addressing this gap, by identifying variables of potential relevance to the person's internal reasoning [93, 94].

Studies outside the field of risk assessment *per se* point to additional factors that may lead to more accurate prediction as well as to a general theory of causation. Barratt and Slaughter [63] propose a "discipline neutral" model which integrates cognitive, biological, behavioral, and "environmental (physical and social)" concepts of the person. This conception includes the idea that the person is defined in part by his social relations, and a sense of belonging or not belonging to a group. One study has indicated that, under experimental conditions, social rejection or exclusion results in decreased "prosocial" behavior, defined as behavior that privileges

collective values over immediate self-interest [95]. Reminders of "connectedness" or social ties may restore prosocial tendencies and reduce aggressiveness after experimentally induced social exclusion [96].

From several different theoretical perspectives, dehumanization has been studied in relation to violence. In particular, the finding that animal cruelty in childhood may be a predictor of future violence [97], especially when motivated by amusement [98], suggests that treating animals as objects of destruction is an index of the person's reduced ability to imagine human qualities in other living beings. Bestiality (sexual assault of animals) may represent an inability to socially relate to human beings and a means of obtaining sexual gratification through aggression, coercion, and manipulation of animals as objects [99]. A lack of empathy or reduced capacity to appreciate the mind of another human being [100] suggests a willingness to harm without appreciating the consequences for others. These latter traits are characteristic of individuals with psychopathy (*see* **Psychopathy**), which has been strongly associated with an increased risk for violence.

Desensitization to and minimization of aggressive behavior, either due to environmental exposure to violence [101] or due to exposure to cultural representations, may be a developmental and dynamic risk factor. Violence in the media includes visual representations of aggression in video games and fictional works as well as in the news and may have an impact analogous to "an environment filled with real violence" [7]. Controversy exists regarding the correlation of violent video games or representations in the media with acts of serious violence [102]. Some studies suggest that there is a process of identification with violent perpetrators when individuals view these scenarios [103]. Although there is solid evidence that viewing violent media produces short-term increases in aggressive behavior (verbal and physical) in children, adolescents, and young adults [7, 104–106], the long-term effects of violent media are difficult to demonstrate, due to intervening factors in the individual's life story and environment [102, 106].

On the basis of extensive clinical observation, Gilligan theorizes that interpersonal violence is largely driven by feelings of shame or of being disrespected and he brings known risk factors to converge upon this common final pathway. He writes, "... it is not poverty, racism, sexism, or age-discrimination, as such, that cause violence. It is, rather, that each correlates with violence because each increases the statistical probability that individuals exposed to these social forces will be subjected to intolerable and potentially self-destroying intensities of shame, from which they do not perceive themselves as having any means of rescuing themselves except by violence..." [107], p. 66. A relationship between pathologies of narcissism (threatened egotism) to violence has begun to be studied empirically. Whereas previous authors had posited that low self-esteem was correlated with violence, Baumeister *et al.* [108] reanalyzed the literature in light of the alternative explanatory hypothesis that self-protective behavior in the face of perceived threat is guided by elevated self-esteem. In a subsequent study, Bushman and Baumeister found that the construct of egotism could be differentiated into stable self-esteem and narcissism, the latter condition being unstable and particularly hostile in response to perceived threat [109]. Perceived threat to egotism has also been associated with aggression in psychopathic individuals [110].

Implicit in this theorization of the stimulus for and mechanism of violent action is that past life events, through their effects on the individual's inner experience, set the stage for interpretation of present experiences, particularly of perceived threat to the individual's sense of self. Lacan wrote in 1948 that there may be a "persistence in the subject of the shadow of 'bad internal objects,' related to some accidental 'association' ... [R]e-evoking certain imaginary *personae* and reproducing certain situational inferiorities may *disconcert* the adult's voluntary functions in the most rigorously predictable way – namely, by their fragmenting impact on the imago involved in the earliest identification" [2], p. 94. Although Lacan's focus is on persons who remind the perpetrator of past experiences of structural inferiority during infancy, the internal mental mechanism he describes is in principle relevant to past situations, such as the recollection of environments where potential violence was the norm. In this regards, given that perpetrators may be sentenced to prison, it is possible that, beyond learned behavior, exposure to incarceration may contribute, by these means, to an increased future risk for violence.

Assessment and Treatment

In clinical settings as well as in sentencing or postsentence release hearings, assessment of risk for violence (*see* e.g., **Dangerousness: Risk of**) is an essential step in individualized treatment planning (*see also* **Psychopathy**; **Psychopathy Checklists**; **Addictions**; **Substance Abuse** and [111]). When violence risk assessment is specifically requested by the court, the examiner often performs both a clinical interview and a standardized assessment using a validated rating scale, since these approaches are complementary. Assessment of risk is an appreciation of likelihood and does not rise to the level of accuracy for prediction *per se*. It allows clinicians and decision makers in the judicial system to determine what interventions are needed to reduce an individual's risk for future violence, on a case-by-case basis [112], whether that risk arises from psychiatric or criminological factors (*see* e.g., **Insanity: Defense**; **Therapeutic Jurisprudence**).

Prevention

> *Power and violence are opposites; where the one rules absolutely, the other is absent. Violence appears where power is in jeopardy, but left to its own course it ends in power's disappearance . . . Violence can destroy power; it is utterly incapable of creating it* [1], p. 56.

Viewed from a public health perspective, prevention may be primary (general measures to prevent the condition from developing), secondary (efforts focused on individuals who are at elevated risk of developing the condition), or tertiary (measures introduced to prevent recurrence in individuals known to have the condition) [107], pp. 20–22+. In the setting of legal proceedings, the latter two types of prevention are of greatest relevance.

In 1969, Ilfeld categorized the prevention of violence into treatments of three main types focusing either on the individual or the environment: (i) redirection of aggression through alternate means of expression; (ii) modification of social learning; and (iii) modification of sources of frustration, such as socioeconomic factors [113]. Significant advances have been made in the past decades in understanding the effective basis for treatment at individual and organizational levels. Institutional and community responses [114], including public information campaigns, have resulted in a decrease in the rate of domestic violence, due to increased reporting and help seeking [115]. One study indicates that employers' awareness of intimate partner violence and the availability of support in the workplace are predictors of a victim's continued employment [43].

Although institutional responses such as mandatory arrest may also have a deterrent effect by indicating to the public that violence is not acceptable, prosecution and intervention programs have not consistently demonstrated a reduction in recidivism of offenders [36]. Data collection and the compilation of accurate statistics using combined emergency room and local police sources allow for improved surveillance and planning for prevention strategies [116]. Further study of risk factors and protective factors that serve as a basis for effective prevention programs in widely different but overlapping types of violence may provide indices for a common theorization of aggressive behavior [101].

Conclusions

Aggression and its behavioral manifestations encompass a wide range of phenomena that cannot at this time be meaningfully studied as a whole. Reductionist approaches have been useful in defining different types of violence in specialized populations of victims, offenders, and locations. At the present time, a unified theory of causation of violent acts is lacking, even when the field of study is limited to interpersonal violence, at the exclusion of war, torture, genocide, poverty, racism, and other forms of collective or structural violence. In principle, the widely varied forms of violence have in common a choice in favor of destruction, whether this destruction is directly intended or intrinsic to the structure of social institutions. The violent individual's subjective point of view as a factor in his choice of behavior remains a relevant topic to explore from a clinical or ethnographic perspective, and has implications for a theory of causation. Although the risk for future violence appears to be strongly associated with environmental factors and static features of the individual's past life experience that are not susceptible to change, the issue of the person's capacity to choose how to express aggressive

feelings and his willingness to change past patterns is an important element to consider in future studies of the possible treatment and prevention of violent behavior.

References

[1] Arendt, H. (1970). *On Violence*, Harcourt Brace & Company, San Diego.

[2] Lacan, J. (1966/2006). Aggressiveness in psychoanalysis (1948), in *Ecrits: The First Complete Edition in English*, W.W. Norton & Company, New York, pp. 82–101.

[3] Miczek, K.A., de Almeida, R.M. & Kravitz, E.A., Rissman, E.F., de Boer, S.F. & Raine, A. (2007). Neurobiology of escalated aggression and violence, *Journal of Neuroscience* **27**, 11803–11806.

[4] Gilula, M.F. & Daniels, D.N. (1969). Violence and man's struggle to adapt, *Science* **164**, 396–405.

[5] Spitz, R.A. (1969). Aggression and adaptation, *The Journal of Nervous and Mental Disease* **149**, 81–90.

[6] Zegans, L.S. (1971). Towards a unified theory of human aggression, *The British Journal of Medical Psychology* **44**, 355–365.

[7] Huesmann, L.R. (2007). The impact of electronic media violence: scientific theory and research, *The Journal of Adolescent Health* **41**, S6–S13.

[8] Rothenberg, A. (1971). On anger, *The American Journal of Psychiatry* **128**, 454–460.

[9] National Center for Injury Prevention and Control (2006). *CDC Injury Fact Book*, Centers for Disease Control and Prevention, Atlanta.

[10] Krug, E.G., Dahlberg, L.L., Mercy, J.A., Zwi, A.B. & Lozano, R. (2002). *World Report on Violence and Health*, World Health Organization, Geneva.

[11] Monahan, J., Steadman, H.J., Silver, E., Appelbaum, P.S., Robbins, P.C., Mulvey, E.P., Roth, L.H., Grisso, T. & Banks, S. (2001). *Rethinking Risk Assessment: The MacArthur Study of Mental Disorder and Violence*, Oxford University Press, New York.

[12] Webster, C.D., Douglas, K.S., Eaves, D. & Hart, S.D. (1997). *HCR-20: Assessing Risk for Violence*, Mental Health, Law, and Policy Institute, Burnaby.

[13] Gilligan, J. (1996). *Violence: Reflections on a National Epidemic*, Vintage Books, New York.

[14] Galtung, J. (1990). Cultural violence, *Journal of Peace Research* **27**, 291–305; See also: Galtung, J. (1969). Violence, peace and peace research, *Journal of Peace Research* **6**, 167–191.

[15] Stalans, L.J., Yarnold, P.R. & Seng, M., Olson, D.E. & Repp, M. (2004). Identifying three types of violent offenders and predicting violent recidivism while on probation: a classification tree analysis, *Law and Human Behavior* **28**, 253–271.

[16] Daro, D., Edleson, J.L. & Pinderhughes, H. (2004). Finding common ground in the study of child maltreatment, youth violence, and adult domestic violence, *Journal of Interpersonal Violence* **19**, 282–298.

[17] Waterston, T. & Mok, J. (2008). Violence against children: the UN report, *Archives of Disease in Childhood* **93**, 85–88.

[18] Scott, K. & Straus, M. (2007). Denial, minimization, partner blaming, and intimate aggression in dating partners, *Journal of Interpersonal Violence* **22**, 851–871.

[19] Swanberg, J.E., Logan, T. & Macke, C. (2005). Intimate partner violence, employment, and the workplace: consequences and future directions, *Trauma Violence Abuse* **6**, 286–312.

[20] Shaw, J., Amos, T. & Hunt, I.M., Flynn, S., Turnbull, P., Kapur, N. & Appleby, L. (2004). Mental illness in people who kill strangers: longitudinal study and national clinical survey, *British Medical Journal* **328**, 734–737.

[21] Taylor, T.J., Freng, A., Esbensen, F.A. & Peterson, D. (2008). Youth gang membership and serious violent victimization: the importance of lifestyles and routine activities, *Journal of Interpersonal Violence* **23**, 1441–1464.

[22] Lachs, M.S. & Pillemer, K. (2004). Elder abuse, *Lancet* **364**, 1263–1272.

[23] Teplin, L.A., McClelland, G.M., Abram, K.M. & Weiner, D.A. (2005). Crime victimization in adults with severe mental illness: comparison with the National Crime Victimization Survey, *Archives of General Psychiatry* **62**, 911–921.

[24] Hiday, V.A., Swartz, M.S., Swanson, J.W., Borum, R. & Wagner, H.R. (1999). Criminal victimization of persons with severe mental illness, *Psychiatric Services* **50**, 62–68.

[25] Honkonen, T., Henriksson, M., Koivisto, A.M., Stengard, E. & Salokangas, R.K. (2004). Violent victimization in schizophrenia, *Social Psychiatry and Psychiatric Epidemiology* **39**, 606–612.

[26] Dean, K., Moran, P. Fahy, T., Tyrer, P., Leese, M., Creed, F., Burns, T., Murray, R. & Walsh, E. (2007). Predictors of violent victimization amongst those with psychosis, *Acta Psychiatrica Scandinavica* **116**, 345–353.

[27] Fitzgerald, P.B., de Castella, A.R., Filia, K.M., Filia, S.L., Benitez, J. & Kulkarni, J. (2005). Victimization of patients with schizophrenia and related disorders, *The Australian and New Zealand Journal of Psychiatry* **39**, 169–174.

[28] Choe, J.Y., Teplin, L.A. & Abram, K.M. (2008). Perpetration of violence, violent victimization, and severe mental illness: balancing public health concerns, *Psychiatric Services* **59**, 153–164.

[29] Borum, R., Fein, R., Vossekuil, B. & Berglund, J. (1999). Threat assessment: defining an approach for evaluating risk of targeted violence, *Behavioral Sciences and the Law* **17**, 323–337.

[30] Fagan, P.J., Wise, T.N., Schmidt Jr, C.W. & Berlin, F.S. (2002). Pedophilia, *The Journal of the American Medical Association* **288**, 2458–2465.

[31] Johnson, C.F. (2004). Child sexual abuse, *Lancet* **364**, 462–470.

[32] Bergen, R.K. & Bukovec, P. (2006). Men and intimate partner rape: characteristics of men who sexually abuse their partner, *Journal of Interpersonal Violence* **21**, 1375–1384.

[33] Feldhaus, K.M., Houry, D. & Kaminsky, R. (2000). Lifetime sexual assault prevalence rates and reporting practices in an emergency department population, *Annals of Emergency Medicine* **36**, 23–27.

[34] Wolff, N., Blitz, C.L. & Shi, J. (2007). Rates of sexual victimization in prison for inmates with and without mental disorders, *Psychiatric Services* **58**, 1087–1094.

[35] Wolff, N., Blitz, C.L., Shi, J., Bachman, R. & Siegel, J.A. (2006). Sexual violence inside prisons: rates of victimization, *Journal of Urban Health* **83**, 835–848.

[36] Stover, C.S. (2005). Domestic violence research: what have we learned and where do we go from here? *Journal of Interpersonal Violence* **20**, 448–454.

[37] Anderson, M., Kaufman, J., Simon, T.R., Barrios, L., Paulozzi, L., Ryan, G., Hammond, R., Modzeleski, W., Feucht, T. & Potter, L. (2001). School-associated violent deaths in the United States, 1994–1999, *The Journal of the American Medical Association* **286**, 2695–2702.

[38] Eisenbraun, K.D. (2007). Violence in schools: prevalence, prediction and prevention, *Aggression and Violent Behavior* **12**, 459–469.

[39] Walters, G.D. (1998). Time series and correlational analyses of inmate-initiated assaultive incidents in a large correctional system, *International Journal of Offender Therapy and Comparative Criminology* **42**, 124–132.

[40] Loza, W. & Dhaliwal, G.K. (2005). Predicting violence among forensic-correctional populations: the past 2 decades of advancements and future endeavors, *Journal of Interpersonal Violence* **20**, 188–194.

[41] Abderhalden, C., Needham, I., Miserez, B., Almvik, R., Dassen, T., Haug, H.J. & Fischer, J.E. (2004). Predicting inpatient violence in acute psychiatric wards using the Broset-Violence-Checklist: a multicentre prospective cohort study, *Journal of Psychiatric and Mental Health Nursing* **11**, 422–427.

[42] Abderhalden, C., Needham, I., Dassen, T., Halfens, R., Fischer, J.E. & Haug, H.J. (2007). Frequency and severity of aggressive incidents in acute psychiatric wards in Switzerland, *Clinical Practice and Epidemiology in Mental Health* **3**, 30.

[43] Swanberg, J., Macke, C. & Logan, T.K. (2007). Working women making it work: intimate partner violence, employment, and workplace support, *Journal of Interpersonal Violence* **22**, 292–311.

[44] Beck, J.C. & Schouten, R. (2000). Workplace violence and psychiatric practice, *Bulletin of the Menninger Clinic* **64**, 36–48.

[45] Fletcher, T.A., Brakel, S.J. & Cavanaugh, J.L. (2000). Violence in the workplace: new perspectives in forensic mental health services in the USA, *The British Journal of Psychiatry* **176**, 339–344.

[46] Swanson, J., Swartz, M., Estroff, S., Borum, R., Wagner, R. & Hiday, V. (1998). Psychiatric impairment, social contact, and violent behavior: evidence from a study of outpatient-committed persons with severe mental disorder, *Social Psychiatry and Psychiatric Epidemiology* **33**,(Suppl 1), S86–S94.

[47] Steadman, H.J., Mulvey, E.P., Monahan, J., Robbins, P.C., Appelbaum, P.S., Grisso, T., Roth, L.H. & Silver, E. (1998). Violence by people discharged from acute psychiatric inpatient facilities and by others in the same neighborhoods, *Archives of General Psychiatry* **55**, 393–401.

[48] Swanson, J.W., Van Dorn, R.A., Swartz, M.S., Smith, A., Elbogen, E.B. & Monahan, J. (2007). Alternative pathways to violence in persons with schizophrenia: the role of childhood antisocial behavior problems, *Law and Human Behavior* **32**(3), 228–240.

[49] Taylor, P.J. (1998). When symptoms of psychosis drive serious violence, *Social Psychiatry and Psychiatric Epidemiology* **33**,(Suppl 1), S47–S54.

[50] Tremblay, R.E., Nagin, D.S., Seguin, J.R., Zoccolillo, M., Zelazo, P.D., Boivin, M., Perusse, D. & Japel, C. (2004). Physical aggression during early childhood: trajectories and predictors, *Pediatrics* **114**, e43–e50.

[51] Herrenkohl, T.I., Maguin, E., Hill, K.G., Hawkins, J.D., Abbott, R.D. & Catalano, R.F. (2000). Developmental risk factors for youth violence, *The Journal of Adolescent Health* **26**, 176–186.

[52] Herrenkohl, T.I., McMorris, B.J., Catalano, R.F., Abbott, R.D., Hemphill, S.A. & Toumbourou, J.W. (2007). Risk factors for violence and relational aggression in adolescence, *Journal of Interpersonal Violence* **22**, 386–405.

[53] Barker, E.D., Seguin, J.R., White, H.R., Bates, M.E., Lacourse, E., Carbonneau, R. & Tremblay, R.E. (2007). Developmental trajectories of male physical violence and theft: relations to neurocognitive performance, *Archives of General Psychiatry* **64**, 592–599.

[54] Brower, M.C. & Price, B.H. (2001). Neuropsychiatry of frontal lobe dysfunction in violent and criminal behaviour: a critical review, *Journal of Neurology, Neurosurgery, and Psychiatry* **71**, 720–726.

[55] Santos, C.O., Caeiro, L., Ferro, J.M., Albuquerque, R. & Luisa Figueira, M. (2006). Anger, hostility and aggression in the first days of acute stroke, *European Journal of Neurology* **13**, 351–358.

[56] Sigafoos, J., Elkins, J., Kerr, M. & Attwood, T. (1994). A survey of aggressive behaviour among a population of persons with intellectual disability in Queensland, *Journal of Intellectual Disability Research* **38**(Pt 4), 369–381.

[57] Crocker, A.G., Mercier, C., Lachapelle, Y., Brunet, A., Morin, D. & Roy, M.E. (2006). Prevalence and types of aggressive behaviour among adults with intellectual

disabilities, *Journal of Intellectual Disability Research* **50**, 652–661.

[58] Taft, C.T., Pless, A.P., Stalans, L.J., Koenen, K.C., King, L.A. & King, D.W. (2005). Risk factors for partner violence among a national sample of combat veterans, *Journal of Consulting and Clinical Psychology* **73**, 151–159.

[59] Mulvey, E.P., Odgers, C., Skeem, J., Gardner, W., Schubert, C. & Lidz, C. (2006). Substance use and community violence: a test of the relation at the daily level, *Journal of Consulting and Clinical Psychology* **74**, 743–754.

[60] Thompson, M.P. & Kingree, J.B. (2006). The roles of victim and perpetrator alcohol use in intimate partner violence outcomes, *Journal of Interpersonal Violence* **21**, 163–177.

[61] Shaw, J., Hunt, I.M., Flynn, S., Amos, T., Meehan, J., Robinson, J., Bickley, H, Parsons, R., McCann, K. & Burns, J. (2006). The role of alcohol and drugs in homicides in England and Wales, *Addiction* **101**, 1117–1124.

[62] Barratt, E.S., Stanford, M.S., Dowdy, L., Liebman, M.J. & Kent, T.A. (1999). Impulsive and premeditated aggression: a factor analysis of self-reported acts, *Psychiatry Research* **86**, 163–173.

[63] Barratt, E.S. & Slaughter, L. (1998). Defining, measuring, and predicting impulsive aggression: a heuristic model, *Behavioral Sciences and The Law* **16**, 285–302.

[64] Shaw, J., Appleby, L., Amos, T., McDonnell, R., Harris, C., McCann, K., Kiernan, K., Davies, S., Bickley, H. & Parsons, R. (1999). Mental disorder and clinical care in people convicted of homicide: national clinical survey, *British Medical Journal* **318**, 1240–1244.

[65] Shaw, J., Hunt, I.M., Flynn, S., Meehan, J., Robinson, J., Bickley, H., Parsons, R., McCann, K., Burns, J. & Amos, T. (2006). Rates of mental disorder in people convicted of homicide. National clinical survey, *The British Journal of Psychiatry* **188**, 143–147.

[66] Malmquist, C.P. (2006). *Homicide: A Psychiatric Perspective*, American Psychiatric Publishing, Arlington.

[67] Dobash, R.E., Dobash, R.P., Cavanagh, K. & Medina-Ariza, J. (2007). Lethal and nonlethal violence against an intimate female partner: comparing male murderers to nonlethal abusers, *Violence Against Women* **13**, 329–353.

[68] Kausch, O. & Resnick, P.J. (1999). Psychiatric assessment of the violent offender, in *Handbook of Psychological Approaches with Violent Offenders: Contemporary Strategies and Issues*, V. Hasselt & M. Hersen, eds, Kluwer Academic/Plenum Publishers, pp. 439–457.

[69] Junginger, J. (1996). Psychosis and violence: the case for a content analysis of psychotic experience, *Schizophrenia Bulletin* **22**, 91–103.

[70] Lacan, J. (2004). *Le Séminaire de Jacques Lacan. Livre X: L'Angoisse (1962–1963)*, J.-A. Miller, ed, Seuil, Paris, pp. 135–153.

[71] Miller, J.A. (1988). Jacques Lacan: remarque sur son concept de passage à l'acte, *Actualités Psychiatriques* **1**, 5–14.

[72] Evans, C. (2006). What violent offenders remember of their crime: empirical explorations, *The Australian and New Zealand Journal of Psychiatry* **40**, 508–518.

[73] Hervé, H., Cooper, B.S. & Yuille, J.C. (2007). Memory formation in offenders: perspectives from a biopsychosocial model of eyewitness memory, in *Offenders' Memories of Violent Crimes*, S.A. Christianson, ed, John Wiley & Sons, West Sussex, pp. 37–74.

[74] Merckelbach, H. & Christianson, S.A. (2007). Amnesia for homicide as a form of malingering, in *Offenders' Memories of Violent Crimes*, S.A. Christianson, ed, John Wiley & Sons, West Sussex, pp. 165–190.

[75] Moskowitz, A. (2004). Dissociation and violence: a review of the literature, *Trauma Violence Abuse* **5**, 21–46.

[76] Evans, C., Ehlers, A., Mezey, G. & Clark, D.M. (2007). Intrusive memories in perpetrators of violent crime: emotions and cognitions, *Journal of Consulting and Clinical Psychology* **75**, 134–144.

[77] Mercy, J.A., Krug, E.G., Dahlberg, L.L. & Zwi, A.B. (2003). Violence and health: the United States in a global perspective, *American Journal of Public Health* **93**, 256–261.

[78] Tjaden, P. & Thoennes, N. (2000). Full report on the prevalence, incidence, and consequences of violence against women, in *Findings from the National Violence Against Women Survey*, U.S. Department of Justice, Office of Justice Programs, Washington, DC.

[79] U.S. Department of Health and Human Services, Centers for Disease Control (2008). *Violence Prevention: Intimate Partner Violence Prevention, CDC Data Sources*, http://www.cdc.gov/ncipc/dvp/IPV/ipv-data.htm (accessed Mar 2008).

[80] U.S. Department of Health and Human Services, Centers for Disease Control (2008). *National Violent Death Reporting System*, http://www.cdc.gov/ncipc/profiles/nvdrs/default.htm (accessed Mar 2008).

[81] Federal Bureau of Investigation U.S. (2008). *Uniform Crime Reports*, http://www.fbi.gov/ucr/ucr.htm (accessed Mar 2008).

[82] U.S. Department of Justice (2008). *Bureau of Justice Statistics*, http://www.ojp.usdoj.gov/bjs/ (accessed Mar 2008).

[83] U.S. Department of Justice, Bureau of Justice Statistics (2004). *Comparing UCR and NCVS*, at: http://www.ojp.usdoj.gov/bjs/pub/html/ntcm.htm (accessed Mar 2008).

[84] U.S. Department of Justice, Bureau of Justice Statistics (2007). *Crime and Victim Statistics*, http://www.ojp.usdoj.gov/bjs/cvict.htm (accessed Apr 2008).

[85] Freud, S. (1953). *The standard edition of the complete psychological works of Sigmund Freud*, Civilization and its Discontents (1930), in vol. 21, The Future of an Illusion, Civilization and its Discontents and Other

Works, J. Strachey, ed, Hogarth Press and the Institute of Psycho-Analysis, London.

[86] Flannery, D.J., Vazsonyi, A.T. & Waldman, I.D. (eds) (2007). *The Cambridge Handbook of Violent Behavior and Aggression*, Cambridge University Press, New York.

[87] Elias, N. (2000). *The Civilizing Process: Sociogenetic and Psychogenetic Investigations (1939, 1968)*, Blackwell Publishing, Malden.

[88] United Nations Educational, Scientific and Cultural Organization (UNESCO) (1986). *Seville Statement on Violence, Adopted 1989*, http://portal.unesco.org/education/en/ev.php-URL_ID=3247&URL_DO=DO_TOPIC&URL_SECTION=201.html (accessed Apr 2008).

[89] Douglas, K.S. & Skeem, J.L. (2005). Violence risk assessment: getting specific about being dynamic, *Psychology, Public Policy, and Law* **11**, 347–383.

[90] Nelson, R.J. & Trainor, B.C. (2007). Neural mechanisms of aggression, *Nature Reviews Neuroscience* **8**, 536–546.

[91] Siever, L.J. (2008). Neurobiology of aggression and violence, *The American Journal of Psychiatry* **165**, 429–442.

[92] Bufkin, J.L. & Luttrell, V.R. (2005). Neuroimaging studies of aggressive and violent behavior: current findings and implications for criminology and criminal justice, *Trauma Violence Abuse* **6**, 176–191.

[93] Beech, A., Fisher, D. & Ward, T. (2005). Sexual murderers' implicit theories, *Journal of Interpersonal Violence* **20**, 1366–1389.

[94] Tilley, D.S. & Brackley, M. (2005). Men who batter intimate partners: a grounded theory study of the development of male violence in intimate partner relationships, *Issues in Mental Health Nursing* **26**, 281–297.

[95] Twenge, J.M., Baumeister, R.F., DeWall, C.N. (2007). Social exclusion decreases prosocial behavior, *Journal of Personality and Social Psychology* **92**, 56–66.

[96] Twenge, J.M., Zhang, L., Catanese, K.R., Ciarocco, N.J. & Bartels, J.M. (2007). Replenishing connectedness: reminders of social activity reduce aggression after social exclusion, *The British Journal of Social Psychology* **46**, 205–224.

[97] Miller, C. (2001). Childhood animal cruelty and interpersonal violence, *Clinical Psychology Review* **21**, 735–749.

[98] Hensley, C. & Tallichet, S.E. (2007). The effect of inmates' self-reported childhood and adolescent animal cruelty: motivations on the number of convictions for adult violent interpersonal crimes, *International Journal of Offender Therapy and Comparative Criminology* **52**(2), 175–184.

[99] Hensley, C., Tallichet, S.E. & Singer, S.D. (2006). Exploring the possible link between childhood and adolescent bestiality and interpersonal violence, *Journal of Interpersonal Violence* **21**, 910–923.

[100] Goldberg, B.R., Serper, M.R., Sheets, M., Beech, D., Dill, C. & Duffy, K.G. (2007). Predictors of aggression on the psychiatric inpatient service: self-esteem, narcissism, and theory of mind deficits, *The Journal of Nervous and Mental Disease* **195**, 436–442.

[101] Guterman, N.B. (2004). Advancing prevention research on child abuse, youth violence, and domestic violence: emerging strategies and issues, *Journal of Interpersonal Violence* **19**, 299–321.

[102] Olson, C.K. (2004). Media violence research and youth violence data: why do they conflict? *Academic Psychiatry* **28**, 144–150.

[103] Konijn, E.A., Bijvank, M.N. & Bushman, B.J. (2007). I wish I were a warrior: the role of wishful identification in the effects of violent video games on aggression in adolescent boys, *Developmental Psychology* **43**, 1038–1044.

[104] Anderson, C.A. & Bushman, B.J. (2001). Effects of violent video games on aggressive behavior, aggressive cognition, aggressive affect, physiological arousal, and prosocial behavior: a meta-analytic review of the scientific literature, *Psychological Science* **12**, 353–359.

[105] Anderson, C.A. (2004). An update on the effects of playing violent video games, *Journal of Adolescence* **27**, 113–122.

[106] Huesmann, L.R. & Taylor, L.D. (2006). The role of media violence in violent behavior, *Annual Review of Public Health* **27**, 393–415.

[107] Gilligan, J. (2001). *Preventing Violence*, Thames & Hudson, New York.

[108] Baumeister, R.F., Smart, L. & Boden, J.M. (1996). Relation of threatened egotism to violence and aggression: the dark side of high self-esteem, *Psychological Review* **103**, 5–33.

[109] Bushman, B.J. & Baumeister, R.F. (1998). Threatened egotism, narcissism, self-esteem, and direct and displaced aggression: does self-love or self-hate lead to violence? *Journal of Personality and Social Psychology* **75**, 219–229.

[110] Cale, E.M. & Lilienfeld, S.O. (2006). Psychopathy factors and risk for aggressive behavior: a test of the "threatened egotism" hypothesis, *Law and Human Behavior* **30**, 51–74.

[111] Simon, R.I. & Tardiff, K. (eds) (2008). *Textbook of Violence Assessment and Management*, American Psychiatric Publishing, Washington, DC.

[112] Webster, C.D., Hucker, S.J. (2007). *Violence Risk: Assessment and Management*, UK: John Wiley & Sons, West Sussex.

[113] Ilfeld Jr, F.W. (1969). Overview of the causes and prevention of violence, *Archives of General Psychiatry* **20**, 675–689.

[114] Sabol, W.J., Coulton, C.J. & Korbin, J.E. (2004). Building community capacity for violence prevention, *Journal of Interpersonal Violence* **19**, 322–340.

[115] Shepard, M. (2005). Twenty years of progress in addressing domestic violence: an agenda for the next 10, *Journal of Interpersonal Violence* **20**, 436–441.

[116] Shepherd, J.P. (2001). Emergency medicine and police collaboration to prevent community violence, *Annals of Emergency Medicine* **38**, 430–437.

SUZANNE YANG

Aggression: Gender Differences in

Female aggression traditionally has received little attention in the research literature and has been regarded with a certain degree of skepticism by the public and professionals alike. In the past three decades, however, the precipitants and purpose of female aggression and the profiles of girls and women who commit aggressive, violent, and criminal acts have increasingly attracted the attention of researchers, criminal justice, mental health professionals, advocates, and policy makers. The field is beginning to flourish with a small number of longitudinal and large-scale examinations of sex differences and similarities in antisocial behavior and aggression [1–5].

In addition, the importance of examining female aggression separately from male aggression and using that data to inform policies and practice is increasing, and now appears to be well-recognized in developed nations worldwide [6]. This relatively abrupt turnaround reflects increasing numbers of girls and women being charged, arrested, and incarcerated for violent crimes. In addition, research demonstrates that in certain contexts (e.g., romantic relationships and parent–child relationships) and in highly specific populations and settings (e.g., inpatient psychiatric patients) the gender gap in the perpetration of aggressive actions is largely reduced or entirely absent. These circumstances require scholars and practitioners to evaluate the extent to which our understanding of female aggression reflects substantiated research findings *versus* unsubstantiated generalizations and stereotypes.

This article operationalizes aggression and violence (*see also* **Aggression**) and reports the prevalence and incidence of aggression among females; documents sex differences and similarities in aggressive and violent behavior; critically examines the assertion that risk and protective factors are sex-specific, briefly explores clinical implications for prevention and intervention, and closes with reflections on gaps in knowledge and directions for future research.

Prevalence and Incidence of Aggression among Females

There is virtually universal agreement that males are more aggressive than females. Across age categories, regardless of the data source (i.e., self-report, family/collateral reports, official records, and victimization surveys), males outnumber females in the perpetration of physical and sexual aggression, violence, and crime. However, methodological advances (e.g., data collection strategies and sources) to studying aggression and more inclusive definitions of aggression have revealed comparable rates across the two genders for some types of aggression and/or in discrete settings and populations. Given the importance of developmental milestones in understanding human behavior, in terms of diagnostic categories (e.g., diagnostic and statistical manual of mental disorders (DSM-IV)-TR, [7]) and with regard to prevention and intervention in the criminal justice system, we present the prevalence and incidence of aggression among females separately for female youth and adult women.

Operationalizing Aggression

Although there remains little agreement in the literature regarding how best to operationalize aggressive and violent actions, increasing clarity has been achieved as a result of important studies such as the MacArthur violence risk assessment study [4]. Building on prior definitions [8] and psychometrically advanced measures (Conflict Tactics Scale, [9]), the MacArthur study operationalized "physical aggression" as laying one's hands on another with the intention to cause physical harm. In contrast, "violence" was defined as actions that result in injury, sexual assaults, or verbal threats of physical aggression with a weapon in hand [4]. "Relational aggression" refers to interpersonal interactions and verbal exchanges intended to harm others through social exclusion and public humiliation [10]. Finally, "verbal aggression" includes threats, ridiculing, name-calling, and

shouting. Although imperfect, with these definitions in mind, we can explore the question: how common is female aggression?

Female Youth

Boys perpetrate approximately three times as many violent acts as girls. Girls are also less likely to report carrying a weapon and tend to engage in violent acts at a lower frequency than their male counterparts [3, 11, 12]. Nonetheless, multiple sources of data consistently show increased rates of violence among girls. In the United States, between 1988 and 1998, person-related offenses increased at more than twice the rate among adolescent females (157%) than among adolescent males (71%; [13]) and between 1993 and 2002, arrests for aggravated assault decreased 29% for boys but increased 7% for girls. Canadian statistics show comparable trends: between 1988 and 1998, the violent crime rate more than doubled for girls (+127%) compared to a smaller increase for boys (+65%; [14]). Furthermore, between 1996 and 2002, when a small decrease was noted in the rate of violent crime committed by boys, a modest increase was observed for girls [15].

Outside North America, the picture is much the same: in the United Kingdom, between 1981 and 1999, there was a 23% decrease in juvenile male offenders and an 8% increase in female offenders, although in 1999 males still outnumbered females by 3 : 1–4 : 1 [16]. Similar trends are evident in epidemiological studies. According to the US Surgeon General's report [17], between 1993 and 1998, the gap between adolescent girls' and boys' self-reported engagement in violent acts shrunk by approximately 50%. Turning to relational aggression, research shows that girls engage in at least equal or higher levels of relational aggression than do boys [18].

Relational aggression can be reliably detected as early as preschool and children who engage in it are more likely to suffer rejection from their peers and are more likely to affiliate with deviant peers who also engage in relational aggression [19, 20]. Rates of relational aggression increase during elementary school for girls but not boys [21], possibly reflecting sex differences in the complexity and psychological relevance of relational contexts. The fact that social aggression can have social "payoffs" for some girls has garnered support from recent studies. For example, Cillessen and Mayeux [22] found that young adolescents who were relationally aggressive to others held high social prominence, although they were not well-liked by their peers. This was particularly true for girls. Girls also experience relational aggression as more distressing and harmful than do boys [20].

Adult Women

Official criminal data and incarceration rates confirm that within the general population, men are considerably more aggressive than women and come into conflict with the law with much greater frequency. In North America, men vastly outnumber women in correctional settings. Early in 2006, there were 408 women federally incarcerated in Canada [23] compared with 12 263 men. Although the actual number of women admitted to federal institutions increased from 238 to 276 between 2004–2005 and 2005–2006, women constitute a very small proportion (5.8% in 2005–2006) of all federal admissions. The gender disparity in incarceration is most apparent for violent offenses (homicides, sexual offenses, and other violent crimes (men = 96.5%; women = 3.5%).

In the United Kingdom, the total prison population is comprised of 6.1% women, 17% of whom were incarcerated for violent offenses [24]. In the United States, a country with one of the highest incarceration rates in the world, females comprised just 7% of the total prison population in 2005 [25]. In other words, males were 14 times more likely than females to be incarcerated, relative to the general population. The gender breakdown for violent offenders, in particular, is slightly higher in the United States relative to Canada and the United Kingdom (4.4% female, 95.6% male). Thus, the picture that emerges from a consideration of official criminal justice data sources clearly demonstrates a vast disparity in criminal offending by gender of perpetrator, that is even greater for when one considers violent offenses.

We find that although women are markedly underrepresented in criminal courts and criminal justice settings, female aggression is *not* uncommon. That is, the gender disparity in aggressive behavior noted in the general population does not appear to be evident among individuals with mental illness [2] nor when one examines aggression that occurs within romantic and familial relationships [26, 27]. Among civil psychiatric patients [28] and forensic psychiatric patients (Nicholls, Brink, Greaves, Lussier, and Verdun-Jones,

in press) women match, and sometimes exceed, men in the prevalence and incidence of aggression. Similarly, it has long been recognized that women perpetrate as much as 50% of aggression within familial and intimate relationships (i.e., child abuse, elder abuse, and partner abuse) [29, 30].

Is Female Aggression Increasing?

Between 1980 and 2001 the proportion of females incarcerated in the United States nearly doubled (4–7%, respectively) [31]. The US Department of Justice, Bureau of Justice Statistics reported that the number of women under the jurisdiction of State or Federal prison authorities increased 4.5% from year end 2005, reaching 112 498, and the number of men rose 2.7%, totaling 1 458 363 [32]. The increase in female contact with the US criminal justice system is particularly remarkable given that the number of serious violent offenses committed by persons ages 12–17 declined 61% from 1993 to 2005, while those committed by persons older than 17 fell 58% (see [32]). In Canada, the rate of "serious violent crime" committed by adult women increased from 25 to 46 per 100 000 between 1986 and 2005 [6]. Similarly, as noted above, the rate has more than doubled among female youth since 1986, growing from 60 per 100 000 to 132 by 2005 [6].

Despite claims that female offending is rapidly escalating, any increase should be understood in relation to the low base rate of violent crime among women. It is also noteworthy that the rate at which female youth have been charged with serious violent crimes has been on a slow downward trend since 2001. Comparisons between females and males for the same offenses and over the same period of time provide a clearer picture of trends. Between 1986 and 2005, with the exception of a few downturns over the years, the rate at which women were charged with assault level 1 (simple assault) more than doubled (44–93 per 100 000 population). In comparison, the rates among male adults have shifted downward since the early 1990s. Between 1991 and 2005, the charge rate for male adults for serious violent crime dropped 30% (412–290 per 100 000). From 1993 to 2005, the charge rate for assault level one for male adults fell 25% (from 606 to 455 per 100 000 population). These data confirm the narrowing gap between adult females and males charged with violent crime: in 1986, nine men were charged with a violent offence for every one woman charged. In 2005, this ratio had decreased substantially from five to one. These statistics raise two important questions: what accounts for decreasing male aggression and what contributes to female aggression increasing in both the United States [32] and Canada [6] despite the fact that the rate of violent crime is dropping?

It is clear from the accumulation of knowledge to date that the "myth of female passivity" [33] is not borne out of the extant empirical data. Knowledge of these base rates is essential, particularly for informing violence risk assessments and public funding decisions; for instance, but an appreciation of the nature of female aggression, what motivates females to aggress against others, and the consequences of those actions is likely to drive treatment and intervention policy, research, and clinical efforts.

Exploring the Topography of Female Aggression: Sex Differences and Similarities in Aggressive and Violent Behavior

Recent research shows a slow but progressive trend in developing research objectives and methodologies to better understand the contexts, functions, targets, and implications of women's aggression. Moving beyond the question of how often females are aggressive relative to males, research now compares and contrasts the topography in which male and female aggression occurs [34]. An appreciation of the contexts in which aggression occurs and the purposes it serves for females helps us understand, prevent, and reduce the risk of female aggression.

Form and Function

Topographical similarities in aggressive acts may serve to mask gender differences, thereby obscuring the underlying motivations and mechanisms involved in females' use of aggression and exaggerating similarities in the potential risk posed to victims. Trends have been evidenced in research, which reveal both gendered and nongendered forms and functions of aggression. Emerging findings continue to indicate similarities (e.g., perpetration of any assaultive act, instigation, and injury incurred, [35]; nature and

location, Nicholls *et al.*, under review), in addition to a substantial amount of divergence (e.g., use of very severe forms of violence, [35]), between male and female aggression.

Form – What is the Nature of Female Aggression?. Self-report surveys and victimization reports generally confirm what we see in the official criminal justice and corrections databases reviewed above; violent crimes remain disproportionately low in women and, that is particularly true of certain forms of interpersonal offending. Data from the United States indicates that, as reported by victims, females account for only 1 out of 7 violent offenders. Further, 1 in 50 offenders committing a violent sex offense (including rape and sexual assault) were women; one in 14 robbers were women; 1 in 9 aggravated assault perpetrators were women; as were approximately 1 in 6 offenders who committed a simple assault [32]. Corresponding Canadian figures for those convicted in 2003–2004 indicate that women accounted for approximately 1 in 99 sexual offenders, 1 in 12 robbers, 1 in 6 offenders committing a major assault, and 1 in 7 who committed a common assault [6]. Charges for murder or manslaughter are rare, regardless of gender (in 1991, 48 charges were laid against women, and 486 against men).

Overall, women's violent criminal charges are primarily for common assault [6]. Fewer are brought about by the commission of robbery, and sexual forms of aggression by women are an exceedingly rare occurrence [36], a pattern that is similar in both Canada and the United States [37, 38]. As with sexual offenders, females engaging in stalking behaviors are relatively rare as compared to their male counterparts (accounting for 15–20% of those who perpetrate stalking offenses, with one in five ultimately attacking the victim; [39]), their occurrence may be a function of stalking as a variant of domestic violence (see [40]). It is important to note that when women are aggressive their assaults occupy the entire continuum of aggression. To clarify, women commit both minor (e.g., verbal aggression) and severe forms of aggression (e.g., kicking, beating, choking, and using weapons), particularly within intimate relationships and against family members (see [41, 42]). In fact, when there is reciprocal aggression (i.e., both partners are abusive) or when only one partner is abusive it is most likely to be the woman who uses severe aggression (see [41, 42]; see [26, 29] for reviews).

Setting and Targets – Where Does Female Aggression Occur and Who Do Females Aggress Against?. As compared with men's aggressive acts, women's expressions of aggression are more likely to occur in the private (e.g., the perpetrator's home) *versus* the public (e.g., bars) domain; a finding that holds in data gathered from psychiatric patient samples [35, 43] and is highly consistent with the disproportionate amount of women's aggression that involves domestic violence, child abuse, and elder abuse. Similar to data from other sources on female aggression, female-perpetrated homicide data suggests that women are less likely than males to aggress against strangers (3% of victims killed by women *vs*. 14% killed by men) or casual acquaintances (13% *vs*. 21%, respectively) than men. The relational aspect of female aggression, however, is perhaps most evident across forms of aggression involving abuse of others who are intimate (or perceived intimate).

Differences across gender are readily apparent when considering perpetrator-victim relationship in instances of homicide. In 1994, Statistics Canada reported that 71% of Canadian women charged with homicide were related "domestically" to their victim, whereas this was true among only 24% of their male counterparts. Indeed, in general, the most common target of women's aggressive acts is the current or previous spouse or common-law partner (30%). Further, in the context of domestic violence, women often report themselves to be either the primary [44] or sole aggressor against their nonviolent partners [42, 45]. Another frequent target is the woman's child (28% of violent convictions in the United States [6]; 10.4% of females convicted of murder in Canada killed their child/stepchild, [46]). Between 1976 and 1997 in the United States, parents and stepparents killed nearly 11 000 children. Mothers and stepmothers perpetrated about half those child killings [46].

According to the Canadian Centre for Justice Statistics, females (3%) and males (15%) are both unlikely to have more than one victim of a homicide [6]. Child maltreatment studies have long identified mothers perpetrating abuse to a comparable extent as fathers [41, 47]. Sexual abuse as a form of women's aggression has only recently been examined as perpetrated against adults [48] and in its more prevalent form, against children [49] (for a review see [36]). Consistent with other forms of female aggression, it is also the case that women more often

than men sexually offend against those to whom they provide care (their own offspring or other related children, children they baby-sit or educate [50]).

Function – What Motivates Female Aggression?. An appreciation of why women use aggression and to what extent women's aggression has similar or unique roots to men's aggression is an essential means of developing theoretical explanations and informing prevention and intervention strategies. Felson [51] asserted that there are principally three reasons that people perpetrate aggression: (i) to obtain compliance or control the target; (ii) to attain retribution or justice; and (iii) to promote or defend their self-image.[a] A fourth motive, self-defense is a predominant theme in much of the debate about women's involvement in intimate abuse. In a comprehensive review of the literature, Graham-Kevan [52] found that, in the context of intimate relationships, there are no consistent sex differences in the use of controlling behaviors. Even in samples selected for high rates of physical aggression, she noted that women sometimes use controlling behaviors with similar frequency men. A consideration of the empirical data suggests that men and women do not differ in their desire to control their partners though they may use different methods to achieve control (for reviews, see [52, 53]).

Both men and women experience jealousy, frustration, and disappointment in relationships; thus, efforts to save face are not unique to men and it is not surprising that women are motivated to use aggression for retribution just as are men. Empirical work in the area details both similar and divergent motivations cited by domestically aggressive women themselves. Women often report using aggression against their partners for purposes that are similar to those of men who perpetrate partner violence, that is, a desire to control or punish their partners, to get their attention, as a response to partner emotional abuse, and to express anger [54]. Yet a number of other studies do cite women's additional motivations of self-defense or retaliation [55, 56], which some feminist scholars argue are not common motivations among males but in fact are frequently cited by men and women [57, 58]. For instance, Follingstad *et al.* [54] found that women were significantly more likely to report using physical force in retaliation for emotional hurt (55.9% *vs.* 25.0%; $\chi^2 = 13.11$, $p < 0.0001$) and men were more likely to report using physical force in response to being hit first (29.2% *vs.* 13.6%; $\chi^2 = 5.61$, $p < 0.05$) whereas men were more likely to report jealousy (41.7%) as a motivator than women (8.5%) ($\chi^2 = 29.62$, $p < 0.0001$).

It is also noteworthy that there was no difference in the likelihood that men and women used aggression to "punish the person for wrong behavior" (12.5 *vs.* 16.9%). Contrary to widely held conceptions [59], the notion that women are aggressive against partners primarily in self-defense has not withstood empirical scrutiny [54]. Briefly, women are known to initiate physical assaults, aggress against nonabusive partners, and as few as 10–20% of women report using physical aggression to defend themselves (for a discussion, see [26]). In fact, the varied motivational influences reported to account for female aggression against intimate partners and stalking victims tend to relate to dysfunctional expressions of anger, loneliness, and frustration; and to that of power and anger, very similar to their male counterparts [60, 61]. In the particular case of female stalkers, rage at abandonment and perceived betrayal are most often driven by the desire to establish intimacy with their targets [39].

Among Canadian female homicide offenders, the most common motives cited were escalation of an argument (39%) and frustration (22%) [6]. In only 11% of the cases did women cite revenge, jealousy, or resolving accounts as their motive, compared to 27% cited by male homicide offenders [6]. Laboratory research has also offered considerable insight into female aggression, demonstrating the circumstances under which females aggress [33, 62] and challenging traditional conceptions of females. Briefly, studies suggest that many of the same contexts and circumstances that promote male aggression are also found to increase the likelihood that females will aggress (e.g., emotional arousal, rumination).

Summary. In sum, women are unlikely to commit certain forms of aggression (e.g., robbery, sexual assault, and physical attacks against strangers) but they are equally represented among perpetrators of physical violence in North American family homes and there is increasing evidence that this finding is consistent in other developed nations [42]. Overall, the general trends concerning the form (i.e., interpersonal/familial based), location (i.e., outside the public domain), and target (e.g., family members) of female aggression are likely contributors to underreporting, lower arrest rates for females perpetrating violence,

and a persistent perception of lower severity. It is important to note here that both men and women view female aggression as less severe than that perpetrated by males. Although it is essential not to lose sight of the fact that women suffer more harm as a result of domestic violence, it may be the expectation of less harm and injury that is fueling increasing rates of female aggression, and thus allows females to justify and minimize the impact of their aggressive behaviors. Moreover, we need to always be mindful of the implications of female aggression (and male aggression) for child witnesses to partner abuse and direct child abuse [63, 64].

Outcomes and Implications – How Serious Is Female Aggression?

Aggression is known to have deleterious implications for victims and witnesses that include physical, psychological, emotional, and financial harm (e.g., physical injuries, fear, shame, and posttraumatic stress disorder). Now that there is considerable agreement that the frequency of aggression in women may be on par with men in certain populations (e.g., hospitalized mentally ill patients) or contexts (i.e., within romantic and familial relationships), the debate has shifted to a consideration of whether the consequences of female aggression are on par with the consequences suffered as a result of male aggression. Most scholars agree that female aggression is at least somewhat less likely to result in injury than male aggression. As we have done throughout the manuscript, we focus here on family violence and inpatient aggression, because those are areas of our expertise and there is an abundance of data on the topic. Finally, we end this section by comparing and contrasting the severity of male and female offending as demonstrated by criminal justice statistics on injuries and weapon use.

Several studies have found that, compared to men, women sustain more severe injuries as a result of partner abuse [58, 65]. In a meta-analysis of domestic violence research, Archer [66] reported that women were more likely than men to be injured by a partner and men were more likely than women to inflict an injury. Perhaps given physical differences in size and strength it seems to reason that men would be more likely to injure their targets; however, women often even the playing field with weapons or attack their partners when they are defenseless (e.g., sleeping; [26]). On the basis of their review of the literature, Noller and Robillard [67] concluded that several studies show that women commit a larger share of severe violence. Nonetheless, some studies find sex differences in the reaction of victims confronted with male *versus* female perpetrators [68]: some male victims of female aggression reportedly find the abuse "humorous"[b] whereas female victims of male aggression do not report such a response [69]. Ultimately, however, it is essential to evaluate harm on a case-by-case basis; it is the extent of exposure to trauma, not gender that predicts the long-term emotional implications of aggression [70, 71].

When evidence of the correspondence between inpatient aggression among male and female psychiatric patients became available, critics argued that the findings failed to address the severity of women's aggression and the likelihood that women would cause serious injury [35]. Increasingly, however, research is demonstrating that while it is certainly the case that the victim of a male perpetrator faces a greater risk of harm than the victim of a female perpetrator, that discrepancy is small (Nicholls *et al.*, under review). Conversely, in some circumstances female aggression may have more severe consequences than male aggression, at least in part, as a function of the relationship with the victim. For instance, maternal aggression results in severe consequences to children that are unique from the implications of paternal aggression. Maternal aggression may have particularly salient influences because mothers are the primary attachment figure throughout the lifespan and mothers spend more time interacting with their children [64]. It should be noted, however, that there is considerable between-victim variability in this regard, and furthermore, the majority of children from violent homes (35%–45%) do not experience clinically significant outcomes [63]. Finally, there is a likelihood of early parenthood, and a negative implication of maladaptive parenting. These features are not completely unique to females, but are of greater consideration given that most single parent households continue to be headed by women [72].

Criminal justice statistics evaluate the level of injury sustained by victims and the use of weapons as indicators of the seriousness of violent crimes. Statistics show that compared to males, females rarely commit violent crimes, but when they do they are just as likely to injure their victim and to use weapons [6]. Just over half of victims sustained no injury from either female or male perpetrated violence (51 and

54%, respectively), minor injury requiring no professional medical treatment was less frequent (43 and 38%) and a small minority resulted in major injury requiring professional treatment or death (2 and 4%; [6]).

In closing, it is important to be mindful that there is considerable evidence that the gap between the suffering experienced by male and female victims of aggression has historically been exaggerated because of methodological limitations and political agendas [26, 27, 29, 33, 73] further research in this area is required. The field is still relatively new and emerging methodological sophistication (e.g., prospective longitudinal examinations including mental health and physical injury outcomes) enhances our full understanding of the different consequences of male and female aggression [69].

Understanding Female Aggression

Are Risk Factors for Aggression Sex-Specific?

Considerable discussion in the literature has revolved around the extent to which common or unique risk factors underlie aggression in males and females. Despite sound rationale for the importance of gender-sensitivity, few efforts have been made to study empirically whether the predictors and moderators of aggression are sex-specific. Similarly, in the development of risk assessment instruments and other forensic assessment measures (e.g., psychopathy measures) it has been rare for them to incorporate theoretical evidence of sex differences and similarities in the variables of relevance. In addition to an insufficient amount of research, an examination of the extant literature suggests that the findings to date have been equivocal; thus, we examine (i) the extent to which risk factors operate in a similar way, increasing the likelihood of aggression in both males and females; (ii) the differential influence of the same risk factors as a function of gender (i.e., being more influential in one gender than the other); and (iii) gender-specific risk factors (i.e., increasing risk in males or females, but not in both; having the opposite effect, increasing risk for aggression in one and decreasing risk in the other).

Many Commonalities. Not surprisingly, many of the factors that leave males vulnerable to committing aggressive behavior also increase the risk of aggression among females. There is considerable evidence that static risk factors (i.e., unchangeable variables such as a diagnosed serious mental illness, a history of child abuse, etc.) as well as dynamic risk factors (i.e., changeable predictors that are potentially influenced by treatment and interventions, such as substance abuse, anger, impulsivity, and poor social support) are relevant across populations and settings (e.g., mentally disordered and nondisordered offenders; correctional inmates, civil psychiatric patients, and forensic psychiatric patients). Thus, it stands to reason that risk factors known to be relevant to male aggression are potentially relevant to female aggression (see [28, 74–76]).

On the rare occasions that scholars have made efforts to develop gender-informed assessment measures from the ground up (e.g., [77]; service planning instrument SPIN) the result has been a remarkable degree of overlap in the variables found in measures previously developed for males. For instance, Blanchette and Taylor [77] examined 176 variables identified as theoretically, empirically, or operationally relevant to security classifications of women in correctional settings. The result was a measure composed of nine variables – all but one of which was common to classification measures developed for men. The authors concluded that despite the attempt to develop a gender-informed measure their results suggest there is little evidence for gender-specific variables; however, they caution that "the order of relevance and weighting of predictive items might differ by gender" (p. 376). Blanchette and Taylor went on to note that evidence of considerable overlap in risk factors for male and female offending is highly consistent with well-established theory [78] and prior research in the field of corrections [79, 80].

Another approach to addressing the question of the extent to which risk factors have a common influence over male and female aggression has been to study the psychometric properties of existing measures constructed based on research with males and test their applicability to populations of females. These efforts have yielded revealing, though generally inconsistent, results. Some studies have found predictive accuracy of existing measures result in similar or better predictive capacity with females [28, 81–84] while other studies have found small or moderate and often insignificant associations with women's aggression [85–87] (for a review of violence risk assessment with women see [88].

In their longitudinal study of a birth cohort (ages 3–21) Moffitt *et al.* [3] concluded that the same risk factors predict antisocial behavior in both males and females (also see [1]). Although they did *not* find any evidence of "replicable sex-specific risk factors" the authors did note that family adversity, compromised intelligence, difficult temperament, and hyperactivity had somewhat stronger effects on males than females. They caution, however, that the sex differences are small and "at best, offer only weak support to the hypothesis that males are more vulnerable than females to risk factors for antisocial behavior" [[3], p. 108]. Given these findings it is important to consider to what extent the same risk factors have a unique bearing on the expression of aggression in males *versus* females.

The Differential Influence of Similar Variables. While there appears to be considerable symmetry in male and female risk markers for violence, evidence also exists for potential differential influences by a number of those shared variables. One particularly noteworthy domain is exposure to elements often present in dysfunctional families of origin. Findings suggest that child abuse and witnessing domestic violence may be more influential in the development of aggression among girls than it is among boys. Differential outcomes are evidenced when the nature of the abuse and the perpetrator gender are taken into consideration as a function of the gender of the victim. For instance, some evidence suggests that childhood maltreatment in the form of sexual abuse may be a risk factor that is especially important in the emergence of girls' antisocial behavior [89, 90]. Further, experiencing childhood abuse at the hands of ones' mother has been found to be a powerful predictor of relationship violence [91], as has childhood abuse perpetrated by ones' father, which appears to explain more of the statistical variance among females than males [64, 92].

There is still reason to expect a greater likelihood of previously victimized women employing violent strategies in their own intimate relationships, as witnessing parental aggression has predicted women's subsequent use of verbal and physical aggression toward their partner [90]. These findings have been replicated and further refined through more recent research, albeit utilizing a similar sample. Explaining an astounding 51% of the variance in violence, women reported perpetrating more violent acts toward their partners if they had seen their mothers aggress against their fathers [91]. That various gender differences have been evidenced with regard to the impact of witnessing domestic violence, for instance, suggest refinement of the current blanket conceptualization as applied across gender.

Evidence from epidemiological studies [93], population-based research [2, 5, 94] and many patient-based studies (see for citations and brief discussions, [28, 93]) call into question the extent to which the gender gap in aggression witnessed in the general population is reflected among individuals with serious mental illness. Mental disorder, a robust predictor of violence, may have a differential association with the likelihood of aggression among females, although the findings to date are equivocal [93]. Swanson *et al.* [5] assessed the prevalence of self-reported violence over one year in community participants. The authors found that among persons with no mental disorder, violence was much more common among males. The gender difference was substantially reduced among mentally ill individuals. Hodgins [2] reported similar findings through an examination of mental disorder and intellectual disabilities in a Swedish birth cohort. She found that these risk factors had a substantial and differential impact on the risk of crime and violence among females. Specifically, Hodgins [2] reported that women who had a serious mental illness or intellectual handicap were five times more likely to commit a criminal offence than women without those characteristics. Particularly notable for the present discussion, that was twice the increased risk reported for men for the same variables.

When specific aspects of mental disorder have been considered, differences are again evident in the expression of violence; for instance, the presence of positive psychotic symptoms has been found to be more prominently associated with physical violence in women than in men [95], although the reason for these differences remains unspecified. Substance abuse as a contributing factor in violence perpetration is more prevalent, and relates more strongly with physical assault in men than in women [95]. Among homicide perpetrators, 71% of males and 65% of females were reportedly under the influence of drugs and/or alcohol during the commission of the offence [6]. Gender differences in substance use-related aggression are further evident when the type of drug is separated in further analysis [96]. Together,

these findings suggest the differential impact of certain clinical and psychosocial factors on subsequent aggression, and emphasize the need to "drill down" past only the first level of factors under investigation.

Socialization and the influence of societal norms and values may have a unique influence on the likelihood of aggression in females as well, protecting females by promoting prosocial behavior, the development of empathy, and an appreciation of caregiving. Conversely, our gender-role socialization of males may actually promote subsequent displays of aggression through rewarding competitiveness and machismo. In general, the influence of protective factors has been overlooked in much of the literature [97, 98] this seems to be no less true of females [82].

Gender-Specific Variables. Reflecting the limited body of evidence to inform this field, it remains unknown to what extent there may be risk variables or protective variables that are unique to the risk of aggression for one gender or the other. For instance, precocious pubertal development and having a mature and/or sexualized physical appearance may present a sex-specific risk for girls entering into aggressive and antisocial behavior [3]. There is also some evidence from the general population and particularly from the sexual offending literature, [36] that antisocial male partners may play an important role in some female aggression. Official criminal statistics offer some support to this hypothesis, demonstrating that women committing violent offenses are more likely to have done so in partnership with a male than are male violent offenders to have committed violent offenses in partnership with a female.

Logan's [36] review of the sexual offending literature reminds us that women do commit serious offenses independently of males and a small but robust proportion of women's involvement in aggression may reflect the influence of antisocial men (the reverse is likely also true). Though the impact of antisocial peers and negative relationships is hardly unique to females, the role of antisocial male partners, (perhaps particularly older men) may be a unique predictor of adolescent girls' involvement in aggression. There is some evidence for this in the expression of antisocial and aggressive behaviors emerging later in girls, typically in adolescence, around the time that same-sex socialization gives way to increasing mixed gender socialization and sexuality. Moreover, criminologists have suggested that being in a romantic relationship with a woman generally inhibits antisocial behaviors in males [99]. Not surprisingly, though, antisocial males and females selectively mate (assortative mating), likely escalating the risk of aggression in both partners, as well as the risk of an intergenerational transmission of aggression to offspring.

In what appears to be the most comprehensive examination of empirical data on the etiology of physical violence by male and female dating partners, Medeiros and Straus [57] asserted that most of what has been written about the causes and motives for women's aggression has been based on writers' assumptions, in the absence of empirical evidence. Based on their extensive review of the literature Medeiros and Straus [57] found four types of studies they categorized by the type of data reported: (i) seven studies evaluated 25 variables and the statistical relationships between motive and gender; 72% of the relationships analyzed demonstrated no significant gender difference. (ii) The second type of study compared violent men and women on 56 characteristics (e.g., educational attainment, measures of anger, etc.) and demonstrated that in 73% of comparisons no significant difference was found. (iii) The third category also examined violent men and women but did not test significance; therefore, they categorized gender differences that were 20% or more as a gender difference. According to that threshold, they concluded that 43% of variables were similar for men and women (28 variables in six studies). (iv) Finally, the fourth category included studies that examined risk factors for partner assault separately for men and women but did not test for significant differences.

In 23 studies, reporting results in relation to 147 risk factors, 60% of variables showed the same relationships for men and women, 39% showed the direction of the relationship to the risk factor was the same for men and women but was significant in one but not the other, and 1% of variables showed opposite relationships in men and women (one positive and significant, the other negative and significant). The authors interpreted these findings to mean that of the risk factors considered there was a similar etiological pattern for men and women for 60% of the risk factors examined, or that 99% of the studies showed the effect of the variable was in the same direction for men and women [57].

Much work remains to parse out the underlying mechanisms by which female aggression emerges. These disparate findings highlight the need for gender-sensitivity when considering female aggression. Although further study is necessary, the research to date seems to suggest that many of the same clinical, psychosocial, and environmental risk factors pertain to males and females, however, there is some evidence that the clinical and psychosocial factors that are associated with increased risk for aggression have a different impact on males and females [3, 95]. As Crick [100] speculated, nonnormative forms of aggression (i.e., overt aggression and physical violence) may reflect maladjustment more than gender normative forms of aggression (i.e., relational aggression and verbal aggression). The as yet unresolved question remains whether gender-specific models of aggression are necessary to explain female, as separate from male, perpetrated aggression. Further methodologically sound exploration into disparities and overlap concerning the roots of female aggression may definitively direct us to adopting extant male models, or alternatively, to considering female aggression as a separate phenomenon.

The Developmental Trajectory

Until recently, research on aggression and violence in childhood and adolescence was based on the assumption that aggressive behavior increased from childhood to adolescence, often as a result of exposure to various risk factors. New work on developmental trajectories reveals a different picture: first, aggressive acts such as hitting and biting are *sometimes* present in over 40% of two-year old boys and almost 35% of girls, and are *frequently* present in 5% of boys and 1% of girls [101]. After age 2, aggressive behavior tapers off quickly, and by age 11 only 10–15% of boys and girls *sometimes* engage in aggressive acts and fewer than 5% do so *frequently* [102]. From age 6 to 16, further desistance is noted for the vast majority of children; however, 4% of boys continue to be aggressive [103]. While girls also show drops in physical aggression with age, girls but not boys increase their use of relational aggression during primary school [21]. Together these findings show that most children learn to inhibit aggression very early in childhood, however a small proportion of boys and girls are not taught or do not learn how to desist, and girls in general begin to use greater amounts of relational aggression.

The failure to desist in childhood aggression, and to acquire new aggressive behavior early in childhood, is a clear marker for future pathology. Approximately 95% of boys who show severe aggressive behavior early in development (i.e., prior to age 10) continue to show antisocial and aggressive behavior into adolescence and adulthood, thereby lending credence to the distinction between early-onset life-course persistent (LCP) *versus* adolescent limited (AL) conduct disorder [104–106]. Moffitt *et al.* [3] asserted that the LCP *versus* AL taxonomy applies equally well to males and females, however the rate of early-onset *versus* adolescent-onset cases is extremely low among females. For example, only 6 of the approximately 450 females (1.3%) from the Dunedin Longitudinal Study were identified as life-course offenders, whereas 78 (17%) were identified as adolescent-onset. Consistent with this finding, the gender gap in rates of conduct disorder is greater in childhood than in adolescence (for reviews see [3, 107, 108]).

The fact that adolescent-onset aggression is more common in girls than is childhood-onset has led some to question the validity of this distinction in girls [104, 109]. Silverthorn and Frick [109] proposed that the delayed-onset pattern in girls is comparable to the early-onset pattern in boys in terms of risk markers, stability, and persistence to adulthood. They present findings that show adolescent-onset girls resemble early-onset boys on a range of risk factors [109], and they are more likely to suffer from a multitude of mental health problems in adulthood, including substance dependence, poor physical health, involvement in abusive relationships, antisocial personality disorder, and social welfare dependence [3, 109–111]. However, Moffitt and Caspi [112] proposed that the same model applies to the development of antisocial behavior in girls and boys, and that the delayed onset in girls simply reflects the slower rate of accumulated risk factors for girls. In particular, the higher prevalence of neurocognitive and temperamental risk markers in boys than girls exerts a significant impact on early development and results in boys reaching a threshold of risk for antisocial behavior more quickly than do girls [112]. Yet, whether or not risk factors operate similarly for girls and boys is unclear. Moffitt and Caspi [112] assume a linear and additive model of risk; however, consideration of other

models is warranted. Some risk factors may have gender-specific impacts, or may interact with other risk factors in a gender-specific way [113]. There is simply too little research to conclude that the risk models developed primarily on boys are accurate in predicting onset and developmental course in girls.

Conclusions: Implications, Gaps in Knowledge, and Future Directions

Female aggression remains a topic fraught with controversy and heated debate. Despite widespread consensus of large gender disparities in the amount and consequence of aggression committed by males *versus* females, we recommend exercising caution in the wholesale acceptance of "well-established" knowledge for which there is little or equivocal empirical evidence. Aggression is complex [114] and no one variable, including gender is a sufficient explanation for why one person is aggressive and another is not [115]. The lens through which society has traditionally examined female aggression has been colored largely by our knowledge of women's engagement in criminal violence that comes to the attention of the criminal justice system. As we have reviewed here, the picture that emerges from a consideration of arrest and incarceration data clearly confirms a vast disparity in violent criminal offending by males *versus* females.

For this reason, it is not surprising that research calling into question the accuracy of our comfortable classification of females as gentle, nurturing, empathic, caregivers is often received both by the research community and the general population with disbelief and caution and at other times with outrage (see [73]). People who study female aggression have been ostracized and vilified, their efforts to solicit funding and to communicate their research findings have been blocked [73, 116]. Similarly, advocates who work to provide support to victims of female aggression have encountered severe criticism and empty pockets [117]. Readers should be cognizant that sex differences have been exaggerated in the literature as a result of ideology and stereotypes and assumptions have been maintained often due to a lack of empirical evidence to contradict our socially sanctioned assumptions about females [33]. In fact, the extent to which this wisdom holds depends very much on the population and setting, as well as the type of aggression being examined.

As we have suggested, gaining a more thorough understanding of aggression requires a consideration of the multiple manifestations it takes, discriminating between verbal and physical assaults, physical aggression and severe violence, for instance. The extant literature offers compelling evidence that violent females are vastly outnumbered by violent males in the general population but that women contribute to nearly half of the aggression that occurs in inpatient psychiatric settings, intimate and familial relationships; though, they remain somewhat less likely than males to commit harm that results in injury.

Violence by females has not been recognized as a public health concern, there is little public education along those lines, funding for research has been purposely directed away from examining the issue (Straus) and the study of relevant variables has been intentionally blocked (e.g., psychopathy, [118]; for a discussion see [84, 119]). As we have demonstrated here, avoiding the difficult questions is not an effective means of achieving increased health and safety. While male aggression is on the decline, female aggression is increasing and that aggression is now known to have widespread, lasting, and substantial implications for victims, perhaps most importantly, children. To move ahead, to effectively reduce aggression in society (i.e., not only among females because female aggression has implications for male aggression and the intergenerational transmission of violence) we must be willing to challenge our most firmly held beliefs about gender, patriarchy, and sexism.

As Murray Straus, one of the pioneers in this field has admonished, we have to ask ourselves if we are more committed to maintaining our political perspectives or are we committed to reducing aggression? While always remaining cognizant of many important gender differences, an increasing recognition that female aggression is not uncommon means that we can now can turn our attention away from attempting to credit or discredit research showing gender equity and begin to uncover what contributes to, or conversely prevents, female aggression [29] and why it might be that in certain settings and populations female aggression is uniquely common. Continued contributions to the female aggression knowledge base carry the potential for directly informing development of proactive aggression prevention programs, as well as clinical treatment options to curtail further expressions of violence amongst those most at risk.

Acknowledgments

Michael Smith Foundation for Health Research; Social Sciences and Humanities Research Council; Canadian Institutes of Health Research.

End Notes

[a] There are other potential motives of which readers should be mindful, but they are beyond the scope of this article (e.g., excitement, [51]).

[b] While this finding may offer important insight into gender differences with respect to the fear experienced by female *versus* male victims of intimate partner abuse several related issues require further study and careful consideration. For instance, it is unknown to what degree reporting reflects male socialization and sex-role expectations (e.g., we do not teach male children to fear their female peers). Further, although men may report less fear than women that offers little evidence that they are actually at less risk. As we noted above, when women commit violent offenses generally and when they are violent to their partners, specifically, there is relatively little difference in the risk of injury to the victim.

References

[1] Fergusson, D.M. & Horwood, L.J. (2002). Male and female offending trajectories, *Development and Psychopathology* **14**, 159–177.

[2] Hodgins, S. (1992). Mental disorder, intellectual deficiency and crime: evidence from a birth cohort, *Archives of General Psychiatry* **49**, 476–483.

[3] Moffitt, T.E., Caspi, A., Rutter, M. & Silva, P.A. (2001). *Sex Differences in Antisocial Behaviour: Conduct Disorder, Delinquency, and Violence in the Dunedin Longitudinal Study*, Cambridge University Press, Cambridge.

[4] Monahan, J., Steadman, H., Silver, E., Appelbaum, P., Robbins, P., Mulvey, E., Roth, L., Grisso, T. & Banks, S. (2001). *Rethinking Risk Assessment: The MacArthur Study of Mental Disorder and Violence*, Oxford University Press, New York.

[5] Swanson, J., Holzer, C. & Ganju, V. (1990). Violence and psychiatric disorder in the community: evidence from the Epidemiological Catchment Area Survey, *Hospital and Community Psychiatry* **41**, 761–770.

[6] Juristat (2008). *Female Offenders in Canada*, (Catalogue No. 85-002-XIE, 28[1]). Retrieved January 28, 2008, from http://www.statcan.ca/english/freepub/85 002-XIE/85-002-XIE2008001.pdf.

[7] American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition, Author, Washington, DC.

[8] Lidz, C.W., Mulvey, E.P. & Gardner, W. (1993). The accuracy of predictions of violence to others, *The Journal of the American Medical Association* **269**(8), 1007–1011.

[9] Straus, M.A., Hamby, S.L., Boney-McCoy, S. & Sugarman, D.B. (1996). The revised Conflict Tactics Scale (CTS2): Development and preliminary psychometric data, *Journal of Family Issues* **17**, 283–316.

[10] Moretti, M.M., Obsuth, I., Odgers, C.L. & Penney, S.R. (2008). Female perpetrators of violence, in *The Encyclopaedia of Interpersonal Violence*, C.M. Renzetti & J.L. Edleson, eds, Sage, Thousand Oaks, CA.

[11] Snyder, H.N. & Sickmund, M. (1999). *Juvenile Offenders and Victims: 1999 National Report*, (NCJ 178257), U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention, Washington, DC.

[12] Webster, P.S., Gainer, P.S. & Champion, H.R. (1993). Weapon carrying among inner-city junior high school students: defensive behavior vs aggressive delinquency, *American Journal of Public Health* **83**(11), 1604–1608.

[13] Puzzanchera, C., Stahl, A.L., Finnegan, T.A., Tierney, N. & Snyder, H.N. (2003). *Juvenile Court Statistics 1998*, National Center for Juvenile Justice, Office of Juvenile Justice and Delinquency Prevention, Washington, DC.

[14] Savoie, J. (2000). *Youth Violent Crime*, (Statistics Canada-Catalogue number 85-002-XPE. Vol. 19 no. 13), Canadian Centre for Justice Statistics, Ottawa.

[15] Juristat (2004). *Criminal Victimization in Canada*, (Catalogue No. 85-002-XPE, 25[7]). Retrieved January 28, 2008, from http://dsp-psd.pwgsc.gc.ca/Collection-R/Statcan/85-002-XIE/0070585-002-XIE.pdf.

[16] East, K. & Campbell, S. (1999). *Aspects of Crime: Young Offenders*, Home Office, London.

[17] U.S. Department of Health and Human Services (2001). *Youth Violence: A Report of the Surgeon General*, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Injury Prevention and Control; Substance Abuse and Mental Health Services Administration, Center for Mental Health Services; and National Institutes of Health, National Institute of Mental Health, Rockville.

[18] Crick, N.R. & Grotpeter, J.K. (1995). Relational aggression, gender, and social-psychological adjustment, *Child Development* **66**, 710–722.

[19] Werner, N.E. & Crick, N.R. (2004). Maladaptive peer relationships and the development of relational and physical aggression during middle childhood, *Social Development* **13**(4), 495–514.

[20] Coyne, S.M., Archer, J. & Eslea, M. (2006). 'We're Not Friends Anymore! Unless . . .': the frequency and harmfulness of indirect, relational, and social aggression, *Aggressive Behavior* **32**(4), 294–307.

[21] Murray-Close, D., Ostrov, J.M. & Crick, N.R. (2007). A short-term longitudinal study of growth of relational aggression during middle childhood: associations with gender, friendship intimacy, and internalizing

problems, *Development and Psychopathology* **19**(1), 187–203.

[22] Cillessen, A.H.N. & Mayeux, L. (2004). From censure to reinforcement: developmental changes in the association between aggression and social status, *Child Development* **75**(1), 147–163.

[23] Public Safety and Emergency Preparedness Portfolio Corrections Statistics (2006). *Corrections and Conditional Release Statistical Overview*, Public Works and Government Services Canada, Ottawa Cat. No. PS1-3/2006E.

[24] Home Office (2003). *Statistics on Women and the Criminal Justice System*, Home Office Press Office, London.

[25] Bureau of Justice Statistics Bulletin (2006). *Prison and jail inmates at midyear 2005*, (NCJ 213133). Retrieved January 28, 2008 from http://www.ojp.gov/bjs/pub/pdf/pjim05.pdf.

[26] Dutton, D.G. & Nicholls, T.L. (2005). The gender paradigm in domestic violence research and theory: part 1–the conflict of theory and data, *Aggression and Violent Behavior* **10**(6), 680–714.

[27] Hamel, J. & Nicholls, T.L. (2007). *Family Interventions in Domestic Violence: A Handbook of Gender-Inclusive Theory and Treatment*, Springer, London.

[28] Nicholls, T.L., Ogloff, J.R.P. & Douglas, K.S. (2004). Assessing risk for violence among male and female civil psychiatric patients: the HCR-20, PCL:SV, and VSC, *Behavioral Sciences and the Law* **22**(1), 127–158.

[29] Straus, M.A. (2006). Future research on gender symmetry in physical assaults on partners, *Violence Against Women* **12**, 1086–1097.

[30] Straus, M.A. & Gelles, R.J. (1986). Societal change and change in family violence from 1975 to 1985 as revealed by two national surveys, *Journal of Marriage and the Family* **48**, 465–479.

[31] Kruttschnitt, C. & Gartner, R. (2003). Women's imprisonment, in *Crime and Justice: An Annual Review of Research*, M. Tonry, ed, University of Chicago Press, Chicago, Vol. 30, pp. 1–81.

[32] Bureau of Justice Statistics (2008). *Crime and Victim Statistics*, Retrieved January 28, 2008 from http://www.ojp.usdoj.gov/bjs/cvict.htm.

[33] Richardson, D.S. (2005). The myth of female passivity: thirty years of revelations about female aggression, *Psychology of Women Quarterly* **29**, 238–247.

[34] Putallaz, M. & Bierman, K.L. (2005). *Aggression, Antisocial Behavior, and Violence Among Girls*, Guilford, New York.

[35] Hiday, V.A., Swartz, M.S., Swanson, J.W., Borum, R. & Wagner, H.R. (1998). Male-female differences in the setting and construction of violence among people with severe mental illness, *Social Psychiatry and Psychiatric Epidemiology* **33**(1), S68–S74.

[36] Logan, C. (2008). Sexual deviance in females: Psychopathology and theory, in *Sexual deviance: Theory, Assessment, and Treatment*, 2nd Edition, D.R. Laws & W. O'Donohue, eds, Guilford Press, New York.

[37] Reiss Jr, A.J. & Roth, J.A. (1993). *Understanding and Preventing Violence*, National Research Council, Commission on Behavioral & Social Sciences & Education, Committee on Law & Justice, Panel on the Understanding & Control of Violent Behavior, National Academy Press, Washington, DC, Vol. 1.

[38] Steffensmeier, D. (1995). Trends in female crime: it's still a man's world, in *The Criminal Justice System and Women: Offenders, Victims, and Workers*, B.R. Price & N. Sokoloff, eds, McGraw-Hill, New York.

[39] Meloy, J.R. & Boyd, C. (2003). Female stalkers and their victims, *The Journal of the American Academy of Psychiatry and the Law* **31**(2), 211–219.

[40] Langhinrichsen-Rohling, J. (2006). An examination of sheltered battered women's perpetration of stalking and other unwanted pursuit behaviors, *Violence and Victims* **21**(5), 579–595.

[41] Straus, M.A. & Gelles, R.J. (1999). *Physical Violence in American Families: Risk Factors and Adaptations to Violence in 8, 145 Families*, Transaction publishers, New Brunswick.

[42] Straus, M.A. (2007). Dominance and symmetry in partner violence by male and female university students in 32 nations, *Children and Youth Services Review*. DOI:10.1016/j.childyouth.2007.10.004.

[43] Robbins, P.C., Monahan, J. & Silver, E. (2003). Mental disorder, violence, and gender, *Law and Human Behavior* **27**(6), 561–571.

[44] Fiebert, M.S. & Gonzalez, D.M. (1997). College women who initiate assaults on their male partners and the reasons offered for such behavior, *Psychological Reports* **80**(2), 583–590.

[45] Goldenson, J., Geffner, R., Foster, S. & Clipson, C.R. (2007). Female domestic violence offenders: their attachment security, trauma symptoms, and personality organization, *Violence and Victims* **22**(5), 530–543.

[46] Bureau of Justice Statistics (1999). *Women Offenders*, (NCJ 175688). Retrieved January 28, 2008 from http://www.ojp.gov/bjs/pub/pdf/wo.pdf.

[47] Schaeffer, S.M., Alexander, P.C., Bethke, K. & Kretz, L.S. (2005). Predictors of child abuse potential among military parents: comparing mothers and fathers, *Journal of Family Violence* **20**(2), 123–129.

[48] Krahé, B., Waizenhöfer, E. & Möller, I. (2003). Women's sexual aggression against men: prevalence and predictors, *Sex Roles* **49**, 219–232.

[49] Grayston, A.D. & de Luca, R.V. (1999). Female perpetrators of child sexual abuse: a review of the clinical and empirical literature, *Aggression and Violent Behavior* **4**, 93–106.

[50] Saradjian, J. & Hanks, H. (1996). *Women Who Sexually Abuse Children: From Research to Clinical Practice*, John Wiley & Sons, Chichester.

[51] Felson, R.B. (2002). *Violence & Gender Re-examined*, American Psychological Association, Washington, DC.

[52] Graham-Kevan, N. (2007). Power and control in relationship aggression, in *Family Interventions in Domestic Violence: A Handbook of Gender-Inclusive Theory and Treatment.*, J. Hamel & T. Nicholls, eds, Springer, London, pp. 87–107.

[53] Felson, R.B. & Outlaw, M.C. (2007). The control motive and marital violence, *Violence and Victims* **22**(4), 387–407.

[54] Follingstad, D.R., Wright, S., Lloyd, S. & Sebastian, J.A. (1991). Sex differences in motivations and effects in dating violence, *Family Relations* **40**(1), 51–57.

[55] Kernsmith, P. (2005). Exerting power or striking back: a gendered comparison of motivations for domestic violence perpetration, *Violence and Victims* **20**(2), 173–185.

[56] Swan, S.C. & Snow, D.L. (2003). Behavioral and psychological differences among abused women who use violence in intimate relationships, *Violence Against Women* **9**, 75–109.

[57] Medeiros, R. & Straus, M. (2007). Risk factors for physical violence between dating partners: implications for gender-inclusive prevention and treatment of family violence, in *Family Interventions in Domestic Violence: A Handbook of Gender-Inclusive Theory and Treatment*, J. Hamel & T. Nicholls, eds, Springer, London, pp. 87–107.

[58] Whitaker, D.J., Haileyesus, T., Swahn, M. & Saltzman, L.S. (2007). Differences in frequency of violence and reported injury between relationships with reciprocal and nonreciprocal intimate partner violence, *American Journal of Public Health* **97**, 941–947.

[59] Swan, S.C. & Snow, D.L. (2002). A typology of women's use of violence in intimate relationships, *Violence Against Women* **8**(3), 286–319.

[60] Nathan, P. & Ward, T. (2001). Females who sexually abuse children: assessment and treatment issues, *Psychiatry, Psychology and Law* **8**, 44–55.

[61] Nathan, P. & Ward, T. (2002). Female sex offenders: clinical and demographic features, *The Journal of Sexual Aggression* **8**, 5–21.

[62] Knight, G.P., Guthrie, I.K., Page, M.C. & Fabes, R.A. (2002). Emotional arousal and gender differences in aggression: a meta-analysis, *Aggressive Behavior* **28**, 366–393.

[63] Davies, P.T. & Sturge-Apple, M.L. (2007). The impact of domestic violence on children's development, in *Family Interventions in Domestic Violence: A Handbook of Gender-Inclusive Theory and Treatment*, J. Hamel & T. Nicholls, eds, Springer, London, pp. 165–189.

[64] Moretti, M.M., Penney, S.R., Obsuth, I. & Odgers, C. (2007). Family lessons in attachment and aggression: the impact of interparental violence on adolescent adjustment, in *Family Interventions in Domestic Violence: A Handbook of Gender-Inclusive Theory and Treatment*, J. Hamel & T. Nicholls, eds, Springer, London, pp. 165–189.

[65] Ehrensaft, M.K., Moffitt, T.E. & Caspi, A. (2006). Is domestic violence followed by an increased risk of psychiatric disorders among women but not among men? A longitudinal cohort study, *The American Journal Psychiatry* **163**(5), 885–892.

[66] Archer, J. (2000). Sex differences in aggression between heterosexual partners: a meta-analytic review, *Psychological Bulletin* **126**(5), 651–680.

[67] Noller, P. & Roillard, L. (2007). Couple violence: a new look at some old fallacies, in *Family Interventions in Domestic Violence: A Handbook of Gender-Inclusive Theory and Treatment*, J. Hamel & T. Nicholls, eds, Springer, London, pp. 125–144.

[68] Holtzworth-Munroe, A., Smutzler, N. & Bates, L. (1997). A brief review of the research on husband violence. Part III: sociodemographic factors, relationship factors, and differing consequences of husband and wife violence, *Violent Behavior* **2**, 285–307.

[69] Holtzworth-Munroe, A. (2005). Female perpetration of physical aggression against an intimate partner: a controversial new topic of study, *Violence and Victims* **20**, 253–261.

[70] Dutton, D.G. (2007). *Rethinking Domestic Violence*, UBC press, Vancouver.

[71] Pimlott-Kubiak, S. & Cortina, L.M. (2003). Gender, victimization, and outcomes: reconceptualizing risk, *Journal of Consulting and Clinical Psychology* **71**, 528–539.

[72] Serbin, L.A., Stack, D.M., De Genna, N., Grunzeweig, N., Temcheff, C.E., Schwartzman, A.E. & Ledingham, J. (2005). When aggressive girls become mothers: problems in parenting, health, and development across two generations, in *Aggression, Antisocial Behavior and Violence Among Girls*, M. Putallaz & K.L. Bierman, eds, Guilford Press, New York, pp. 262–285.

[73] Straus, M.A. (2006). The national family violence surveys, in *Physical Violence in American Families: Risk Factors and Adaptations to Violence in 8,145 Families*, M.A. Straus & R.J. Gelles, eds, Transaction Publishers, New Brunswick, NJ, pp. 3–16.

[74] Moretti, M.M., Catchpole, R. & Odgers, C. (2005). The dark side of girlhood: recent trends, risk factors and trajectories to aggression and violence, *Canadian Child and Adolescent Psychiatry Review* **14**, 32–38.

[75] Odgers, C., Burnette, M.L., Chauhan, P., Moretti, M.M. & Repucci, D. (2005a). Misdiagnosing the problem: mental health profiles of incarcerated juveniles, *Canadian Child and Adolescent Psychiatry Review* **14**, 39–44.

[76] Webster, C.D. (1999. December). *Risk Assessment and Risk Management with Women Offenders*. Report to the National Parole Board, Ottawa.

[77] Blanchette, K. & Taylor, K.N. (2007). Development and field test of a gender-informed security reclassification scale for female offenders, *Criminal Justice and Behavior* **34**, 362–379.

[78] Andrews, D.A. & Bonta, J. (1998). *The Psychology of Criminal Conduct*, 2nd Edition, Anderson, Cincinnati.
[79] Dowden, C. & Andrews, D.A. (1999). What works for female offenders: a meta-analytic review, *Crime and Delinquency* **45**, 438–452.
[80] Loucks, A.D. & Zamble, E. (2000). Predictors of criminal behavior and prison misconduct in serious female offenders, *Empirical and Applied Criminal Justice Review* **1**, 1–47.
[81] Douglas, K.S., Strand, S., Belfrage, H., Fransson, G. & Levander, S. (2005). Reliability and validity evaluation, *Assessment* **12**, 145–161.
[82] Holtfreter, K. & Cupp, R. (2007). Gender and risk assessment: the empirical status of the LSI-R for women, *Journal of Contemporary Criminal Justice* **23**, 363–382.
[83] Lovins, L.B., Lowenkamp, C.T. & Latessa, E.J. (2007). Application of the risk principle to female offenders, *Journal of Contemporary Criminal Justice* **23**, 383–398.
[84] Nicholls, T.L., Ogloff, J., Brink, J. & Spidel, A. (2005). Psychopathy in women: a review of its clinical usefulness for assessing risk for aggression and crime, *Behavioral Sciences and the Law* **23**(6), 779–802.
[85] de Vogel, V. & de Ruiter, C. (2005). The HCR-20 in personality disordered female offenders: a comparison with a matched sample of males, *Clinical Psychology and Psychotherapy* **12**, 226–240.
[86] Odgers, C., Reppucci, N.D. & Moretti, M.M. (2005b). Nipping psychopathy in the bud: an examination of the convergent, predictive and theoretical utility of the PCL-YV among adolescent girls, *Behavioral Sciences and the Law* **23**, 1–21.
[87] Warren, J., South, S., Burnette, M., Rogers, A., Friend, R., Bale, R. & Van Patten, I. (2005). Understanding the risk factors for violence and criminality in women: the predictive validity of the PCL-R and HCR-20, *International Journal of Law and Psychiatry* **28**(3), 269–289.
[88] Garcia-Mansilla, A., Rosenfeld, B. & Nicholls, T.L. (in press). Risk assessment: Are current methods applicable to women? *International Journal of Forensic Mental Health*.
[89] Trickett, P.K. & Gordis, E.B. (2005). Aggression and antisocial behavior in sexually abused girls, in *Aggression, Antisocial Behavior, and Violence among Girls: A Developmental Perspective*, M. Putallaz & K.L. Bierman, eds, Guilford Press, New York, pp. 162–185.
[90] White, J.W. & Humphrey, J.A. (1994). Women's aggression in heterosexual conflicts, *Aggressive Behavior* **20**, 195–202.
[91] Hendy, H.M., Weiner, K., Bakerofskie, J., Eggen, D., Gustitus, C. & McLeod, K. (2003). Comparison of six models for violent romantic relationships in college men and women, *Journal of Interpersonal Violence* **18**(6), 645–665.
[92] Langhinrichsen-Rohling, J., Neidig, P. & Thorn, G. (1995). Violent marriages: gender differences in levels of current violence and past abuse, *Journal of Family Violence* **10**(2), 159–176.
[93] Stueve, A. & Link, B.G. (1998). Gender differences in the relationship between mental illness and violence: evidence from a community-based epidemiological study in Israel, *Social Psychiatry and Psychiatric Epidemiology* **33**(Suppl. 1), S61–S67.
[94] Hodgins, S., Mednick, S.A., Brennan, P.A., Schulsinger, F. & Engberg, M. (1996). Mental disorder and crime: evidence from a Danish birth cohort, *Archives of General Psychiatry* **53**, 489–496.
[95] Krakowski, M. & Czobor, P. (2004). Gender differences in violent behaviors: relationship to clinical symptoms and psychosocial factors, *The American Journal Psychiatry* **161**(3), 459–465.
[96] Tardiff, K., Marzuk, P.M., Leon, A.C., Portera, L. & Weiner, C. (1997). Violence by patients admitted to a private psychiatric hospital, *The American Journal Psychiatry* **154**, 88–93.
[97] Rogers, R. (2000). The uncritical acceptance of risk assessment in forensic practice, *Law and Human Behavior* **24**, 595–605.
[98] Webster, C.D., Nicholls, T.L., Martin, M.L., Desmarais, S.L. & Brink, J. (2006). Short-Term Assessment of Risk and Treatability (START): the case for a new violence risk structured professional judgment scheme, *Behavioral Sciences and the Law* **24**, 747–766.
[99] Laub, J.H., Nagin, D.S. & Sampson, R.J. (1998). Trajectories of change in criminal offending: good marriages and the desistance process, *American Sociological Review* **63**(2), 225–238.
[100] Crick, N.R. (1997). Engagement in gender normative versus nonnormative forms of aggression: links to social-psychological adjustment, *Developmental Psychology* **33**, 610–617.
[101] Baillargeon, R.H., Zoccolillo, M., Keenan, K., Côté, S., Pérusse, D., Wu, H.X., Boivin, M. & Tremblay, R.E. (2007). Gender differences in physical aggression: a prospective population-based survey of children before and after 2 years of age, *Developmental Psychology* **43**(1), 13–26.
[102] Tremblay, R.E., Boulerice, B., Harden, P.W., McDuff, P., Perusse, D., Pihl, R.O. & Zoccolillo, M. (1996). Do children in Canada become more aggressive as they approach adolescence? in *Growing Up in Canada: National Longitudinal Survey of Children and Youth*, HRDC, Ottawa, pp. 127–137.
[103] Nagin, D. & Tremblay, R.E. (1999). Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and non violent juvenile delinquency, *Child Development* **70**(5), 1181–1196.
[104] Loeber, R. & Stouthamer-Loeber, M. (1998). Development of juvenile aggression and violence: some common misconceptions and controversies, *The American Psychologist* **53**(2), 242–259.

[105] Moffitt, T.E. (1993). AddedAdolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy, *Psychological Review* **100**(4), 674–701.
[106] Tremblay, R.E. (2000). The development of aggressive behaviour during childhood: what have we learned in the past century? *International Journal of Behavioral Development* **24**(2), 129–141.
[107] Lahey, B.B., Schwab-Stone, M., Goodman, S.H., Waldman, I.D., Canino, G., Rathouz, P.J., Miller, T.L., Dennis, K.D., Bird, H. & Jensen, P.S. (2000). Age and gender differences in oppositional behavior and conduct problems: a cross-sectional household study of middle childhood and adolescence, *Journal of Abnormal Psychology* **109**(3), 488–503.
[108] Zoccolillo, M. (1993). Gender and the development of conduct disorder, *Development and Psychopathology* **5**(1–2), 65–78. Special issue: Toward a developmental perspective on conduct disorder.
[109] Silverthorn, P. & Frick, P.J. (1999). Developmental pathways to antisocial behavior: the delayed-onset pathway in girls, *Development and Psychopathology* **11**(1), 101–126.
[110] Bardone, A.M., Moffitt, T.E., Caspi, A., Dickson, N. & Silva, P.A. (1996). Adult mental health and social outcomes of adolescent girls with depression and conduct disorder, *Development and Psychopathology* **8**, 811–829.
[111] Robins, L.N. (1986). Deviant children grown up, *European Child and Adolescent Psychiatry* **5**(Suppl 1), 44–46.
[112] Moffitt, T.E. & Capsi, A. (2001). Childhod predictors differentiate life-course persistent and adolescence-limited antisocial pathways amoung males and females, *Development and Psychology* **13**, 355–375.
[113] Moretti, M.M., Odgers, C. & Jackson, M. (2004). *Girls and Aggression: Contributing Factors and Intervention Principles*, Kluwer-Plenum, New York.
[114] Hart, S.D. (1998). The role of psychopathy in assessing risk for violence: conceptual and methodological issues, *Legal and Criminological Psychology* **3**, 121–137.
[115] World Health Organization (2002). *World Report on Violence and Health: Summary*, WHO, Geneva.
[116] Straus, M.A. (2008,. February) *30 years of research: Denials and distortions of the evidence and what to do about it*. *Paper presented at the meeting From Ideology to Inclusion: Evidence-based Policy and Intervention in Domestic Violence Conference*, National Family Violence Legislative Resource Centre, Sacramento, California.
[117] Pizzey, E. (2008,. February) *A history of the domestic violence in the Western world*. *Paper presented at the meeting From Ideology to Inclusion: Evidence-based Policy and Intervention in Domestic Violence Conference*, National Family Violence Legislative Resource Centre, Sacramento, California.
[118] Laishes, J. (2002). The 2002 mental health strategy for women offenders, Correctional Service of Canada. Retrieved Feb 23, 2008 from http://www.csc-scc.gc.ca/text/prgrm/fsw/mhealth/8-eng.shtml.
[119] Nicholls, T.L. & Petrila, J. [Guest Editor](2005). Gender and psychopathy: an overview of important issues and introduction to the special issue. Guest editor for this special issue of, *Behavioral Sciences and the Law* **23**(6), 729–741.

Further Reading

Denov, M.S. (2003). The myth of innocence: sexual scripts and the recognition of child sexual abuse by female perpetrators, *Journal of Sex Research* **40**, 303–314.
Laishes, J. (2002). The 2002 mental health strategy for women offenders. Correctional Service of Canada. Retrieved Feb 23, 2008 from http://www.csc-scc.gc.ca/text/prgrm/fsw/mhealth/8-eng.shtml.
Monahan, J. (2001a). Major mental disorder and violence: epidemiology and risk assessment, in *Clinical Assessment of Dangerousness: Empirical Contributions*, G. Pinard & L. Pagani, eds, Cambridge University Press, New York, pp. 89–102.
Nicholls, T.L., Brink, J., Greaves, C., Lussier, P. & Verdun-Jones, S. (in press). *Female forensic psychiatric inpatients and aggression: incidence, prevalence, severity, and interventions*, *International Journal of Law and Psychiatry*.
Swan, S.C. & Snow, D.L. (2002). A typology of women's use of violence in intimate relationships, *Violence Against Women* **8**(3), 286–319.

TONIA L. NICHOLLS, CAROLINE GREAVES AND MARLENE M. MORETTI

Aging the Living *see*
Anthropology: Aging the Living

Airbags

As with any electronic system, air bag system technology is evolving and being used in more creative ways. Passenger vehicle air bag systems have three main components: the air bag module, the sensors, and the main computer. Within each of these groups,

there are several types of components. Additionally, items such as seatbelt pretensioners and steering column shear capsules factor into the design of the overall passenger safety system.

Air bags are passive safety devices. They require no action by the occupants, other than turning on the vehicle, to perform their protective role in the event of an accident. They are designed to work in conjunction with seatbelts. More advanced air bag systems not only work in conjunction with seatbelts, but also use the information regarding whether or not the seatbelt is actually buckled to aid in deployment decisions. This was one of the first types of occupant detection incorporated into air bag systems.

Occupant detection is an area of air bag technology that is rapidly evolving. From the simple seatbelt switch to the more complicated seat sensors and cameras, occupant sensing and position discrimination employ some of the newest technologies. But, before the newer systems are discussed, an understanding of the basics of air bag systems must be understood.

As stated before, there are three main components to air bag systems: the air bag module, the sensors and the computer. The air bag module is the actual fabric air bag. It is the most visible component. Before it is deployed, the fabric bag is folded in a specific manner and housed behind a molded plastic housing. This housing is designed with a deliberate weak area, called the tear seam, which allows the air bag to break through the housing when it is deployed. This housing is what the occupant sees in the steering wheel or instrument panel of their vehicle. On the back side, there is an inflator containing the electrical components and chemicals that when ignited by the squib create the gas needed to deploy the air bag. When a deployment decision is made, an electrical current is sent through the wires to the squib causing the ignition of the chemicals and the gas that inflates the air bag.

The decision to deploy the air bags is made by the system in the vehicle. Early air bag systems (late 80's –early 90's) were called distributed sensor systems. In a distributed system, there are at least two inertia-based switches located outside of the computer box and one inside the computer box. An inertia-based switch is a mechanical switch that closes due to a change in momentum of the vehicle. There is no accelerometer.[a] To deploy the air bags, one of the external switches and the internal switch had to be closed simultaneously to complete an electrical circuit, sending current to the squib and deploying the air bags simultaneously, as shown in Figure 1. Distributed systems do not contain the amount of crash data that is commonly found in today's vehicles. Since there is no accelerometer, no acceleration or Delta V[b] data can be recorded. The timing of the sensor closures can be recorded along with the status of the seatbelt switch (generally driver only).

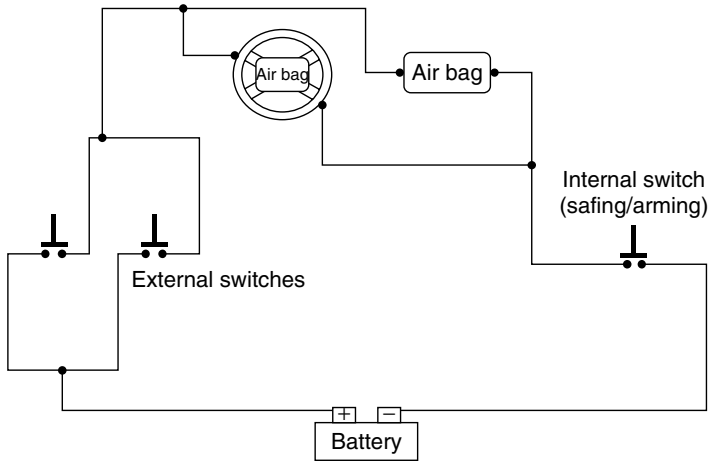


Figure 1 A distributed air bag system employs the use of electromechanical sensors. One is on the battery side of the air bag circuit, and one is on the ground side of the air bag circuit. Both must be closed simultaneously to deploy the air bags

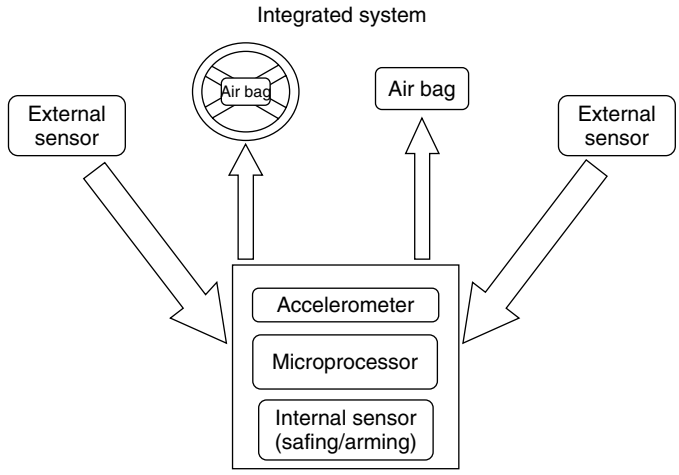


Figure 2 An integrated air bag system uses external sensors connected to the air bag computer in conjunction with the internal accelerometer and internal sensor to make a deployment decision

The next generation of air bag systems began appearing in the early to mid 90's. These systems employed accelerometers located inside the computer box along with a sophisticated algorithm that analyzed the incoming data to make a deployment decision. The fundamental difference between this integrated system and the distributed system is that the simple completion of an electrical circuit does not deploy the air bags. The air bag computer algorithm must make a deployment decision. The integrated system consists of an accelerometer, an inertia-based switch, and a microprocessor with a deployment algorithm housed inside the computer box. Generally at least one external inertia-based switch is used; however, some integrated systems do not use any external sensors, as shown in Figure 2. With the addition of the accelerometer, more detailed crash data including acceleration and/or Delta V of the crash can be recorded.

Successive generations of air bag systems have built upon the basic integrated sensor design. The external sensors have become more sophisticated–some are accelerometers themselves. The algorithms have become more finely tuned. The air bag computer is part of a vehicle-wide network that shares information to help with deployment decisions. Some of the data from the other vehicle computers is saved in the air bag computer crash data files. Seatbelt switches are included in the deployment decision making process, especially when seatbelt pretensioners are present. Seatbelt pretensioners are pyrotechnic devices that remove the slack from the seatbelt webbing to minimize the forward excursion of the occupant. The air bags are not necessarily deployed simultaneously. And occupant detection systems can override a deployment decision if the air bag may do more harm than good. The amount and type of information recorded by each system is completely dependent on the manufacturer of the vehicle and the manufacturer of the air bag system.

Before a discussion on data, directional sensitivity must be addressed. All of the sensors involved in the air bag system are installed in a specific orientation to respond to acceleration pulses coming in certain directions. The sensors associated with frontal air bags are oriented longitudinally on the vehicle. They only respond to acceleration pulses coming from the front to the back of the vehicle. In the same manner, sensors oriented laterally are for side air bag systems. They only respond to acceleration pulse moving right-to-left (passenger side) or left-to-right (driver side). If the acceleration pulse does not occur exactly perpendicular to the front or side of the vehicle (as it rarely does in an actual crash), then the accelerometer is responding to either the lateral or longitudinal component of the crash pulse. As shown in Figure 3, this is called the angle of impact. The farther the angle of impact is from perpendicular, the greater the impact force required to close the sensor, as shown in Figure 4. Issues

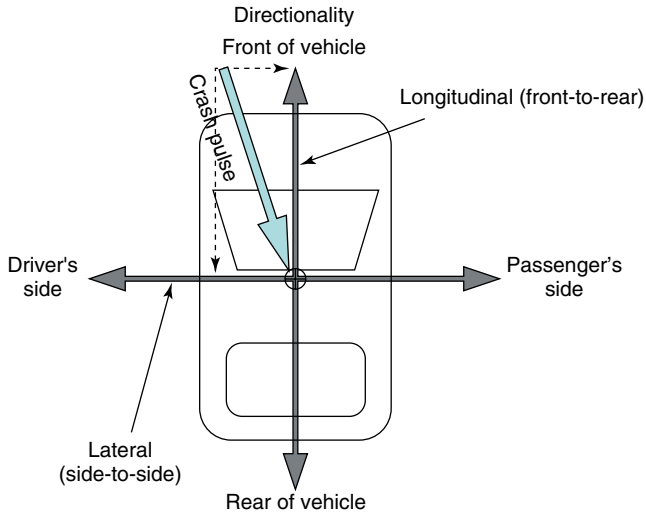


Figure 3 The direction of the crash pulse is important in determining whether or not the air bags will deploy. The front air bags are deployed based on the severity of the longitudinal component of the crash pulse. Side air bags are deployed based on the severity of the lateral component of the crash pulse

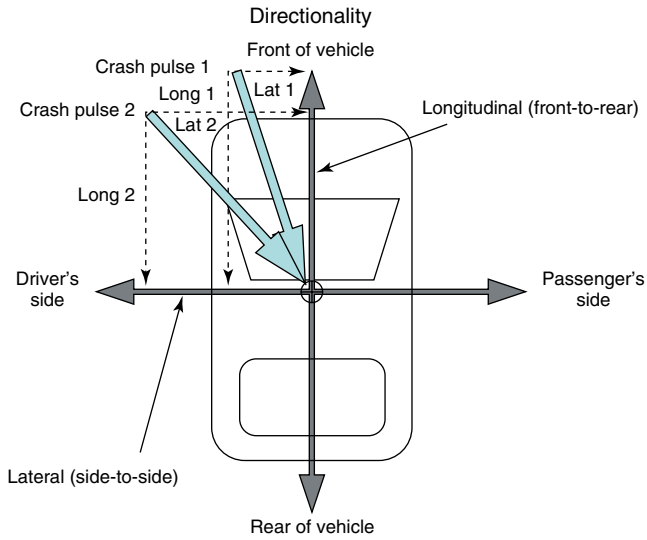


Figure 4 Crash Pulse 1 has a larger longitudinal (front-to-rear) component than lateral (side-to-side). Crash Pulse 2 has equal longitudinal and lateral components

regarding the angle of impact are best resolved by accident reconstruction.

The second part to making sensors an effective part of the air bag system is their placement. The sensors must be placed so that they correctly respond to different crash pulses. In distributed systems, there are two types of sensor placement: triangular and in-line. A triangular sensor placement used two sensors at the front of the vehicle, left and right. Generally, these sensors are found in the area between the headlight and the outside edge of the radiator. The third sensor is the safing/arming sensor located inside

the air bag computer case, which is located near the center of the occupant compartment.

The in-line sensor placement used one sensor at the front, center of the vehicle and one sensor behind the firewall in the occupant compartment. The firewall sensor is variously located. The third sensor is the safing/arming sensor, again located inside the air bag computer case.

The introduction of integrated systems consolidated all of the sensors into the same casing as the air bag computer through the use of an accelerometer. The use of a centrally located accelerometer only required the development of complex computer algorithms to analyze the acceleration data. Eventually, external sensors were brought back into the air bag system to give additional information to the algorithms and help discriminate whether or not to deploy in crashes that may be borderline or have a very narrow pulse transmission zone, such as poles or trees.

A very general rule of thumb for deployment parameters in terms of miles per hour is in a full frontal crash with an impact severity equivalent to hitting a rigid barrier at 8 mph or less the air bags should not deploy. In the same type of crash described above with the impact at 15 mph, the air bags should always deploy. The area between 8 mph and 15 mph is called the gray zone. As air bag systems have become more sophisticated, especially in the area of occupant detection, this general range of deployment parameters can vary widely, going up as high as 20 mph for a properly positioned and seatbelted occupant.

Now that all of the parts of the air bag system have been discussed, the basic types of decision tree structures can be presented. These are all meant to be general examples of the progression of air bag technology.

Starting with the distributed system, the only requirement for deployment is to have one crash sensor and the safing/arming sensor close simultaneously. This would deploy both frontal air bags simultaneously. There is no input from the air bag computer. This is illustrated in Figure 5.

Next, there is an integrated system. In the first generation of integrated systems, the computer algorithm analyzes the accelerometer input and the closure of any external sensors to make a deployment decision, as shown in Figure 6. As the integrated system gains more inputs, the decision process becomes more complicated. The status of seatbelts and the position of

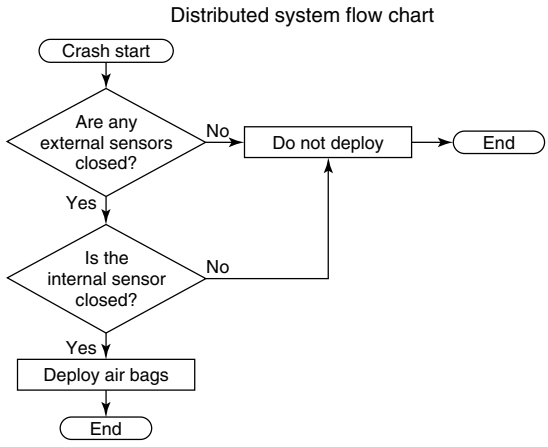


Figure 5 This flow chart outlines the base logic behind a distributed air bag system

the front seats can override the air bag deployment command. This is shown in Figure 7. Additionally, in many trucks and SUVs the passenger air bag can be manually turned off by a keyed switch.

As the air bag system has become more complicated, the amount of information saved by the air bag computer has increased. The first air bag computers were not intended to record information regarding the deployment event. They were there to monitor the status of the air bag system components and record any diagnostic trouble codes (DTCs or fault codes) that occurred. These DTCs would activate the air bag light in the instrument panel signaling the vehicle should be brought in for service. When being serviced, the technician can access the DTCs through a handheld scan tool to help identify the component causing the air bag light.

The first crash-relevant piece of information recorded in the air bag computer was the timing between sensor closures. This time was recorded in milliseconds (1/1000th of a second) and enabled some determination of how quickly the air bags were commanded to deploy. The next piece of data that began to show up consistently was whether or not the driver's seatbelt was buckled. As air bag systems migrated to integrated systems, the recorded timing data became the time from when the computer recognized the possibility of a non-normal event (wake-up or algorithm enable) to the time the air bags were commanded to deploy. Additionally, since these integrated systems employ an accelerometer, the crash

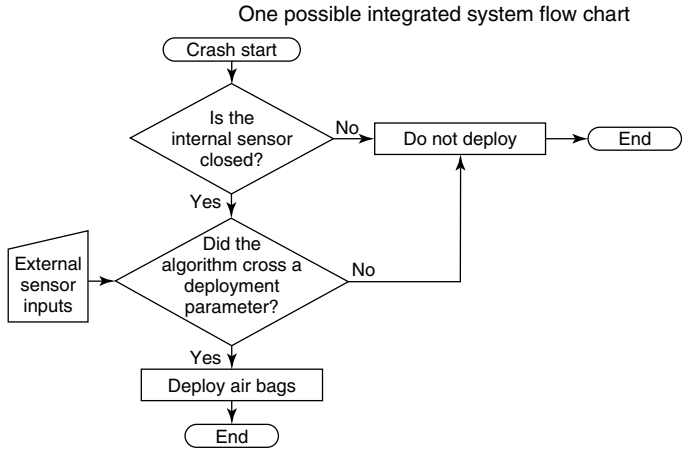


Figure 6 This flow chart outlines the base logic behind one possible configuration of an integrated air bag system

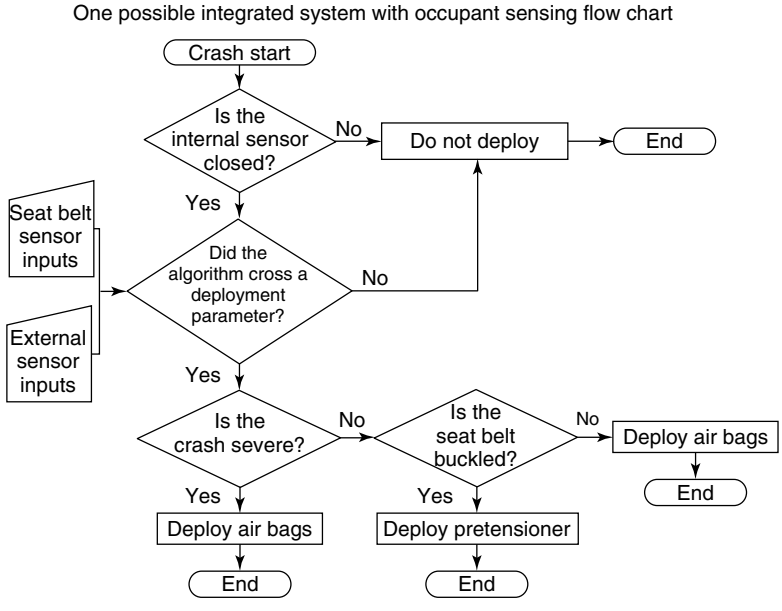


Figure 7 This flow chart outlines the base logic behind one possible configuration of an integrated air bag system that includes occupant sensing

data began to include longitudinal acceleration and/or Delta V. With the advent of side air bags, lateral acceleration and/or Delta V may also be recorded.

The final major piece of crash data that may be recorded is pre-crash data. As of the writing of this article, only two manufacturers, GM and Ford, have employed this type of data. In GM's pre-crash data, four parameters are recorded independently and continuously in a data storage buffer. When a non-normal event is detected, the last five data samples of these parameters are saved into the crash data. The four parameters that GM currently records are vehicle speed, engine RPM, throttle percent and brake switch status. In Ford's pre-crash data, there are many additional data parameters saved for a time interval varying between 26 seconds and over 6 minutes,

depending on the vehicle in question. However, Ford does not save this data in the actual air bag computer. It is stored in volatile memory in the powertrain control module. This data must be retrieved very carefully to prevent its loss. Additionally, power must not be applied to the vehicle in question or Ford's pre-crash data will be overwritten with information pertaining to the current, post-accident state of the vehicle.

If the vehicle in question has seatbelt pretensioners or any form of occupant position sensing, the status of these parameters may also be recorded, especially if they are part of the deployment decision tree. If one front seat occupant is wearing their seatbelt and that seatbelt has a pretensioners and the other front seat occupant is not wearing their seatbelt, the air bag computer algorithm may make the decision to deploy the pretensioners of the belted occupant and the air bag of the unbelted occupant. Belting and occupant position may also factor into the force with which an air bag is deployed. Some air bag systems have two inflators for their air bags which can be timed to change how fast the air bags are deployed or how long they stay inflated.

Once it has been determined that the vehicle in question has the ability to save crash-related data, it can frequently be retrieved. This retrieval process involves the subject vehicle, an interface box, and a laptop computer. There is currently one publicly available interface that allows the user to download the crash data from certain vehicles. If the vehicle in question is not on the list of this public interface system, then the data stored on the air bag computer must be retrieved by the vehicle manufacturer.

The amount and type of crash data saved in the air bag computer varies widely by manufacturer. It can also vary within the manufacturer by model and model year. There are also situations, such as power loss during the accident, which can cause the data to not be recorded or to be only partially recorded. The first priority of the air bag computer is to discriminate between a deployment and non-deployment event and deploy the air bags when necessary. The second priority is to record the crash-related data. The data is first stored in a volatile memory called RAM. This type of memory is erased if power is lost. However, the data must be stored in RAM first because, until a deployment decision is made, the computer would not know whether to place the data in the non-deployment or deployment section of the permanent memory. This permanent memory is called EEPROM. Once the data is in EEPROM it cannot be erased due to a power loss. The air bag computer only has a limited amount of backup power which would first be used to deploy the air bags and then record the crash data.

Now, manufacturers are just beginning to add crash prevention systems. One type of system employs different types of radar to determine how close the vehicle is getting to the one in front of it. The crash prevention system then predicts whether or not the distance between the vehicles is becoming too small. If it is, the vehicle's brakes are applied automatically through the cruise control system. This is called adaptive cruise control. The next level of adaptive cruise control will be able to be used at low speeds which will help to prevent minor accidents in bumper-to-bumper traffic.

Another facet of crash prevention systems are warnings for lane-departure or side alerts if another vehicle is getting too close. Some manufacturers offer night vision, a head-up display, and even a drowsy driver monitor. As of the writing of this article, not only cost is preventing the widespread use of these technologies, but also the integration of the multiple computers and cameras over the vehicle network needs to be developed that is fast and reliable. As this technology continues to evolve, it will help the air bag system by creating more detailed pre-crash inputs.

The following is a list of references for more information on air bag systems and their data:

auto.howstuffworks.com/airbag.htm

actsinc.org	Automotive Coalition for Traffic Safety
nhtsa.dot.gov	National Highway Traffic Safety Administration
iihs.org	Insurance Institute for Highway Safety
sae.org	Society of Automotive Engineers (papers, standards, books, etc.)

End Notes

[a.] An accelerometer is a device that changes voltage when subjected to a mechanical stress. This change in voltage is measured and calibrated to relate to acceleration values through the use of known pulses.

b. ΔV is the difference between the initial velocity and final velocity. It is not the barrier equivalent velocity or the initial velocity of the vehicle before impact.

HOLLY A. ADAMS

Alcohol

Introduction

Alcohol (ethanol or ethyl alcohol) is the world's favorite recreational drug, legally available to adults and widely used for pleasure and relaxation without consequence. Indeed, moderate drinking (one to two glasses of wine each day) has a number of beneficial effects on health and is often recommended as a prophylactic treatment to reduce the risk of dying from coronary artery disease and stroke [1]. However, ethanol is also a drug of abuse and chronic drinking eventually leads to dependence and craving for alcohol with serious consequences for the individual and society [2]. Binge drinking and drunkenness have emerged as major public health problems and overconsumption of alcohol is frequently a catalyst and contributing factor in violent crimes such as physical and sexual assault and homicide [3]. This means that alcohol ranks as the foremost psychoactive substance encountered in forensic investigations of unnatural deaths such as suicides, drowning, and especially road-traffic fatalities where 20–40% of crashes are caused by drunken drivers.

Forensic investigators need to understand what happens to alcohol in the body and how excessive drinking influences a person's behavior in a negative way [4]. The ability to perform skilled tasks, to comprehend and communicate and to form intent are questions that often arise in forensic casework and legal proceedings. The need to translate a person's blood-alcohol concentration (BAC) into the quantity of alcohol consumed is also important and, indeed, ethanol is one of the few drugs for which this calculation is feasible and defensible (see later in this section for details).

During prosecution of drunken drivers, a forensic investigator or expert witness might be asked to perform a backward extrapolation of a person's BAC from the time of blood sampling to an earlier time, such as the time of driving [5]. This procedure is known as *retrograde extrapolation, back-calculation*, or *back-tracking* and is a dubious practice, owing to the many assumptions and unknown factors involved.

Alcohol is the foremost psychoactive substance encountered in postmortem toxicology and heavy drinking and drunkenness, either directly or indirectly, are responsible for considerable morbidity and mortality. Whether or not acute alcohol poisoning was the cause of death is not always easy to determine because of wide variation in the sensitivity and reaction of different people to the same dose of alcohol.

Interpreting the concentration of a drug measured in blood or other body fluid in relation to the dose administered or the effects produced on the individual is fraught with difficulties. Much depends on the interplay between chemical, pharmaceutical, physiological, psychological, and genetic factors. In the case of ethanol, the situation is simplified thanks to its special physiochemical properties, namely, small molecular size, distribution into the total body water (TBW), and lack of binding to plasma proteins. Table 1 gives a summary of the main physicochemical properties of ethanol.

The BAC reached after drinking a known amount of alcohol depends on a host of factors, the most important of which are the quantity consumed and the speed of drinking. Higher doses and more rapid ingestion lead to higher peak BAC (C_{max}) and an earlier

Table 1 The major physicochemical properties of ethanol

Property	Ethanol
CAS number[a]	64–17–5
Molecular weight	46.07
Molecular formulae	C_2H_6O
Chemical formulae	CH_3CH_2OH
Structure	Primary aliphatic alcohol
Common name	Beverage or grain alcohol
Boiling point	78.5 °C
Melting point	−114.1 °C
Density	0.789 at 20 °C
Water solubility	Mixes completely with water

[a] Chemical abstract service registry number

occurring time to peak (t_{max}), and consequently more pronounced effects on the individual concerned. Among pharmacological agents, ethanol is classified as a depressant of the central nervous system even though drinking small quantities tends to elicit a state of excitement and euphoria. However, these feelings arise by suppression of inhibitions in the cerebral cortex and not as a direct stimulant action of the drug [6]. Ethanol is sometimes administered by intravenous infusion, such as 8–10% v/v solutions in physiological saline, as is common practice in some research applications and also in the emergency clinic as an antidote for treatment of methanol poisoning [7]. Absorption of ethanol is rapid when given rectally as an enema, but for all practical purposes alcohol is ingested orally by drinking an alcoholic beverage.

Trace amounts of ethanol are produced naturally in the body, either by fermentation of carbohydrates in the gut or during certain minor biochemical reactions of the intermediary metabolism [8]. The concentrations of the so-called endogenous alcohol in blood are normally so low ($<0.001\,g\,l^{-1}$) that they are without forensic significance, except under very exceptional circumstances. Ethanol is produced endogenously via reduction of acetaldehyde, which is generated from pyruvate during microbiological and fermentation processes. Many Asians (e.g., Japanese) lack an effective hepatic enzyme for the catabolism of acetaldehyde, which means that the concentrations of this precursor of ethanol can increase appreciably. Moreover, in people suffering from yeast infections, such as C*andida albicans* in the gut, concentrations of endogenous ethanol are likely to be higher after intake of carbohydrate-rich meals, such as rice (*see also* **Alcohol: Analysis**).

How much Alcohol was Consumed?

The strengths of different alcoholic beverages (beers, wines, and spirits) are usually written on the labels, and are almost always expressed in terms of percentage of alcohol by volume (% v/v) and sometimes as proof spirit. The use of the term proof spirit has a long history, and in North America a beverage declared as being 100% proof contains 50% v/v alcohol. In any type of blood-alcohol calculation, the concentration of alcohol in the drink consumed needs to be converted from % v/v to % weight/volume (w/v) to arrive at the number of grams of ethanol ingested [9].

Examples of commonly imbibed alcoholic beverages such as beers, table wines, desert, or fortified wines as well as distilled spirits and the concentrations of ethanol they contain are given in Table 2. The concentrations of ethanol are expressed as %

Table 2 Relationships between the alcoholic strength of different beverages (vol%) and quantity of ethanol (EtOH) they contain and the corresponding amounts of ethanol ingested in a typical serving

Beverage type	EtOH conc. (% v/v)	EtOH conc. (% w/v[a])	Volume of a typical drink (ml)	Ethanol (g)	Ethanol (g) in one bottle[c]
Beer[a]	3	2.4	500[b]	12.0	7.9
	4	3.2		16.0	10.6
	5	4.0		20.0	13.2
	6	4.7		23.5	15.6
Table wine	9	7.1	150	10.6	53.3
	10	7.9		11.9	59.2
	12	9.5		14.2	77.3
Fortified wines (sherry and port)	16	12.6	100	12.6	94.5
	18	14.2		14.2	106.5
	20	15.8		15.8	118.5
Spirits (whisky, gin, vodka, and brandy)	35	27.6	25	6.9	201
	40	31.6		7.9	231
	45	35.6		8.9	267

[a] % v/v × 0.79 (density of ethanol)
[b] In the United Kingdom, a pint of beer is equivalent to 568 ml
[c] Beer (330 ml), wines (750 ml), and spirits (750 ml)

v/v and % w/v and these are related by the density of ethanol (Table 1), which is 0.79 g ml^{-1} at room temperature. Also, Table 2 lists the volumes of the above-mentioned beverages dispensed in a typical serving along with the quantity (grams) of ethanol they contain. Lastly, the number of grams of ethanol contained in one bottle of each beverage, namely, beer (330 ml), wine (750 ml), or spirits (750 ml) is listed.

If a person drinks two bottles of beer (5 vol%) and one bottle of table wine (12 vol%), this corresponds to an intake of 2 × 13.2 g or 26.4 g ethanol from the beer and 77.3 g from the wine making a total consumption of 103.7 g of pure ethanol. This information is necessary when various types of blood-alcohol calculations are made for clinical, research, and legal purposes [9].

Alcohol in the Body

Knowledge about the disposition and fate of ethanol in the body and the factors influencing these processes is necessary to answer many questions arising in routine forensic investigations and medicolegal casework. The basic facts relating to absorption, distribution, and elimination of ethanol have been known for many decades and a vast scientific literature exists on this topic [10].

Absorption

Absorption is the process by which a drug passes from the site of administration into the bloodstream and gets transported to all body organs, fluids, and tissues. Alcohol (ethanol) is a small uncharged molecule totally miscible with water and absorption from the gut occurs by passive diffusion across mucous membranes according to the concentration gradient as predicted by Fick's law. Small amounts of alcohol can be absorbed through the mucous surfaces of the oral cavity if an alcoholic drink is kept in the mouth sufficiently long enough.

Alcohol is absorbed along the entire gastrointestinal tract and the speed of absorption depends on the absorption surface area and the concentration gradient of ethanol across the membrane. The absorption of alcohol begins in the stomach but is considerably faster from the upper part of the small intestine, owing to the much larger area of the absorption surface furnished by the microvilli in the duodenum and jejunum. The absorbed alcohol enters the portal venous blood and passes through the liver and then to the heart, and after picking up oxygen in the lungs, it returns back to the heart and distributes throughout the entire systemic circulation.

Absorption of alcohol occurs much faster when gastric emptying is rapid, facilitating passage through the pyloric sphincter into the small intestines. Under normal conditions, the pylorus remains nearly totally closed because of tonic contraction of the pyloric muscle. Factors that influence gastric emptying play an important role in the rate of uptake of alcohol from the gut, which in turn influences the maximum concentration in the blood (C_{max}) and the intensity of the effects of alcohol on the individual.

Consumption of alcohol together with or after a meal slows the rate of absorption because of a food-induced delayed stomach emptying, as reflected in a lower C_{max} and a later-occurring t_{max}. The amount of food ingested is seemingly more important than its composition in terms of fat, protein, or carbohydrate content [10].

Table 3 lists the main factors that can impact on C_{max} and the time of its occurrence (t_{max}) after drinking a given dose of alcohol. Most of the variables listed either alter gastric emptying or are important for the distribution of ethanol between the blood and tissue water (e.g., gender and adiposity), owing to differences in TBW.

Distribution

Ethanol distributes into the aqueous compartment of the body and the concentrations reached in all body fluids and tissues at equilibrium depend primarily on the water content of these fluids and tissues. Specimens such as sweat, saliva, and urine, which are almost 100% water, will contain a higher concentration of ethanol than whole blood, which is ~80% w/w water. Likewise, the concentrations of ethanol in plasma and serum (~93% water) are higher than the concentration in whole blood. Studies have shown that the plasma/blood distribution ratio of ethanol on average is 1.16 : 1 [11]. Note that plasma and serum are the body fluids mostly analyzed at hospital clinical chemistry laboratories, whereas specimens of whole blood are analyzed at toxicology laboratories.

The speed with which ethanol equilibrates between blood water and the extracellular fluid and

Table 3 Factors influencing the peak blood-alcohol concentration or C_{max} reached after drinking the same quantity of alcohol under various conditions

Variable/conditions	Possible mechanism and explanation	Effect on peak BAC (C_{max})
Low body weight	Less body water	Higher
Appreciable adiposity	Less body water	Higher
Female gender	Less body water	Higher
Rapid drinking	Faster absorption	Higher
Drinking on empty stomach	Faster absorption	Higher
Drinking after a meal	Delayed absorption	Lower
More concentrated alcoholic drinks	Faster absorption	Higher
Obstruction in liver blood flow	Slower absorption	Lower
Smoking cigarettes	Delayed absorption	Lower
Drugs that alter gastric emptying	Slow or more rapid absorption	Lower or higher
Gastric bypass surgery	Faster absorption	Higher
Trauma and massive blood loss	Delayed absorption	Lower

tissue water depends on the cross-sectional area of the local capillary bed and the blood flow per gram of tissue. Studies have shown that the volume of distribution of ethanol corresponds very closely with the TBW. Indeed, a person's TBW can be determined by dilution experiments with ethanol as a marker and the results agree well with isotope dilution when ^2H_2O and ^3H_2O are used as tracers.

Between 50 and 60% of a person's bodyweight is water and the volume of the blood amounts to about 5 l in an adult. This makes it clear that large amounts of ethanol must be consumed to increase a person's BAC to a level that causes inebriation and drunkenness. Once absorbed into the blood, the alcohol easily passes through capillary walls into the various tissues and fluids and equilibration is achieved in about 60 min after the end of drinking. Heavier people tend to have a larger body water space to dilute the alcohol they consume resulting in a lower BAC (Table 3) compared with a person of lesser weight. During aging, especially in men, TBW per kg bodyweight decreases as it does in obesity. Elderly (>60 years) and those overweight reach higher BAC for the same dose of ethanol administered compared with a young nonobese individual.

Well-known gender differences exist in the distribution space for ethanol because women tend to have less body water per kilogram body weight than men. On the basis of experiments dating back to the 1930s, the average V_d for women was determined as $0.61\,kg^{-1}$ compared with $0.71\,kg^{-1}$ for men [12]. Note that these values represent the distribution ratio of alcohol between the body as a whole and the blood compartment. Since the experiments used to determine V_d for ethanol date back to the 1930s, they probably need updating considering the epidemic of obesity in modern society [13].

Table 4 shows the relationship between body mass index (BMI) and clinical manifestations or stages of obesity, which should be considered when choosing the most appropriate distribution factor for use in blood-alcohol calculations. Higher BMI (>30) is associated with lower values of V_d such as $0.4–0.51\,kg^{-1}$ rather than $0.6–0.71\,kg^{-1}$ [14].

Instead of using the population average Widmark *r*-factors for men and women, another approach is to calculate a person's TBW from information available about age, height, and weight (see later in this article). As a general recommendation, an intersubject variation of ±20% should be used in

Table 4 Clinical manifestations of obesity in relation to body mass index (BMI), which might impact on the distribution volume of ethanol (Widmark *r*-factor)

Classification	BMI ($kg\,m^{-2}$)
Underweight	<18.5
Normal	18.5–24.9
Overweight	25.0–29.9
Obesity class I	30.0–34.9
Obesity class II	35.0–39.9
Obesity class III	>40

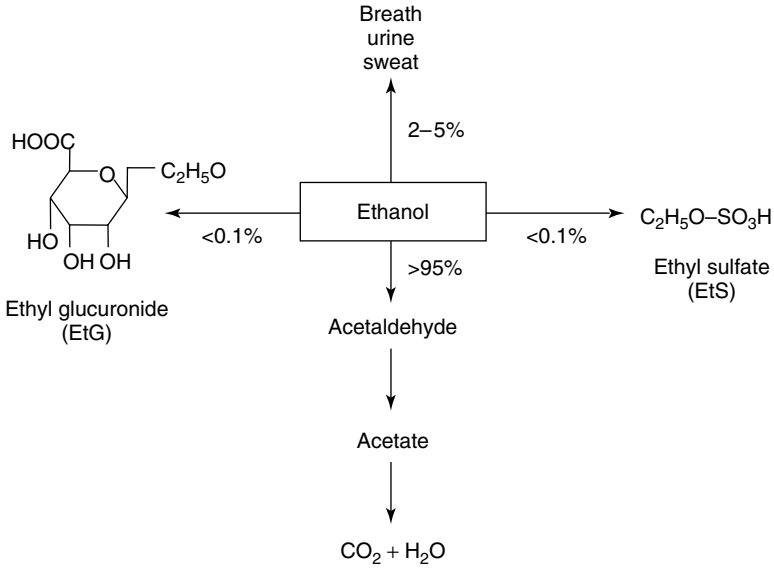


Figure 1 Scheme showing the metabolism of ethanol in the body, the relative amounts excreted unchanged in breath, urine and saliva, and the oxidative and nonoxidative pathways

blood-alcohol calculations to allow for variations in the proportions of fat-to-lean body mass. In obesity (Table 4) or emaciation, a larger variation in body water is expected to occur, which impacts on the BAC reached after a given dose of ethanol.

Metabolism and Excretion

Most of the ethanol a person drinks (95–98%) is removed from the body by oxidative metabolism and a minor fraction (<0.2%) undergoes conjugation reactions to produce ethyl glucuronide (EtG) and ethyl sulfate (EtS), which are the major nonoxidative metabolites [10]. The remainder of the dose of alcohol (2–5%) is eliminated unchanged via the kidney in the urine, through the skin by perspiration and also by pulmonary excretion via the lungs. Figure 1 schematically illustrates the fate of alcohol in the body, showing the relative amounts metabolized and excreted unchanged.

The metabolism of ethanol occurs primarily in the liver by oxidative reactions that liberate considerable energy, actually 7.1 kcal g^{-1} (29.7 kJ), which exceeds the energy content of both proteins and carbohydrates. The liver cells (hepatocytes) are equipped with enzymes that specialize in the degradation of ethanol, the most important of which is the class I alcohol dehydrogenase (ADH), which is located in the cytosol. Enzymatic oxidation of ethanol produces acetaldehyde and this toxic metabolite is converted to acetate by the action of low K_m aldehyde dehydrogenase (ALDH), located in the mitochondria. The acetate produced during the biotransformation of ethanol is transported away from the liver and enters the Krebs cycle where it is converted to CO_2 and H_2O in peripheral organs and tissues. Figure 2 summarizes the salient features of the ADH and ALDH pathways for oxidative metabolism of ethanol.

During the oxidation of ethanol to acetaldehyde and subsequent conversion of the latter to acetate, the coenzyme nicotinamide adenine dinucleotide (NAD^+) is involved. The NAD^+ is simultaneously reduced to NADH and the ratio NADH/NAD^+ increases appreciably, which affects other NAD-dependent metabolic reactions in the liver [15]. Among others, the reduction of pyruvate to lactate is favored, which results in hyperlactacidemia (lactic acidosis) and there is also a reduced rate of gluconeogenesis. The latter is important when the glycogen stores in the liver are depleted, such as when food intake is negligible, leading to a dangerous ethanol-induced hypoglycemia (low blood sugar). The increased NADH/NAD^+ ratio also diminishes

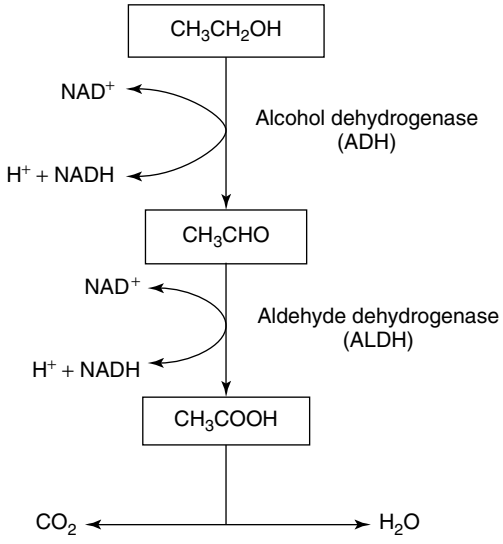


Figure 2 Scheme showing alcohol dehydrogenase (ADH) and aldehyde dehydrogenase (ALDH) pathways for oxidation of ethanol. The same coenzyme nicotinamide adenine dinucleotide (NAD^+) is involved in these redox reactions. The final metabolism of acetate into CO_2 and H_2O occurs mainly in extrahepatic tissues

the oxidation and utilization of fatty acids, which results in an accumulation of fat in the liver and hence the fatty liver (hepatic steatosis) seen in heavy drinkers and alcoholics at autopsy [15].

The class I ADH has a low K_m for oxidation of ethanol ($0.05-0.1\,g\,l^{-1}$), so the enzyme becomes saturated with substrate after just a couple of drinks [16]. Another ADH enzyme (class IV), which is mainly located in the gastric mucosa, has been said to account for presystemic metabolism of ethanol. Both gender and ethnic differences in the activity of gastric ADH exist, which has implications for first-pass metabolism (FPM) and bioavailability of the dose of alcohol administered [17]. Differences of opinion exist regarding the quantitative importance of FPM and whether this is primarily of gastric or hepatic origin.

Besides the ADH pathway, alcohol-metabolizing enzymes are also located within the smooth endoplasmic recticulum, particularly the microsomal fraction. The microsomal enzymes constitute a large family of proteins, known as the *cytochrome P450 monooxygenases*, being responsible for oxidation and reduction of a wide range of foreign chemicals (xenobiotics) that might enter the body, including atmospheric pollutants, industrial solvents, as well as diverse pharmaceutical agents [18]. The particular form of P450 enzyme that converts ethanol to acetaldehyde is denoted by CYP2E1 and has a Michaelis constant (K_m) of $0.6-0.8\,g\,l^{-1}$, being appreciably higher than that of class I ADH ($0.05-0.10\,g\,l^{-1}$). Accordingly, CYP2E1 plays a more important role in the oxidation process when relatively high BACs are reached, such as in heavy drinkers and alcohol-impaired drivers.

$$CH_3CH_2OH + H^+ + O_2 \xrightarrow{CYP2E1} CH_3CHO + NADP^+ + 2H_2O \quad (1)$$

Another feature of CYP2E1 is that its activity increases after sustained periods of heavy drinking lasting weeks or months. The enzyme becomes more effective in the oxidation of its substrate, which explains the old observation that alcoholics on a drinking spree acquire an enhanced capacity to eliminate alcohol from the bloodstream. Involvement of CYP2E1 in the metabolism of alcohol is also responsible for a number of undesirable drug-alcohol interactions [7]. For example, the widely used over-the-counter antipyretic drug paracetamol (acetaminophen or Tylenol®) is converted into a potentially toxic metabolite by CYP2E1. If this enzyme is activated, as it might be in an alcoholic, this can potentially cause hepatic necrosis, loss of liver function, and death. The older sedative–hypnotic drug chloral hydrate (Noctec®) is more potent if taken together with a large dose of ethanol and this adverse drug-alcohol interaction has sometimes led to serious life-threatening consequences [7].

A third liver enzyme that in theory can oxidize ethanol is catalase, which is located in the peroxisomes. This enzyme is capable of oxidizing ethanol under *in vitro* conditions although its role *in vivo* has been questioned because of the lack of hydrogen peroxide necessary for the reaction to proceed. For all practical purposes, ADH and CYP2E1 account for the oxidative metabolism of ethanol in humans.

The fate of ethanol in the body and the pathways of metabolism by the major oxidative enzymes ADH and ALDH are shown in Figures 1 and 2.

The Blood-Alcohol Curve

The time course of alcohol in the body is usually displayed as a graph on which the BAC is plotted on

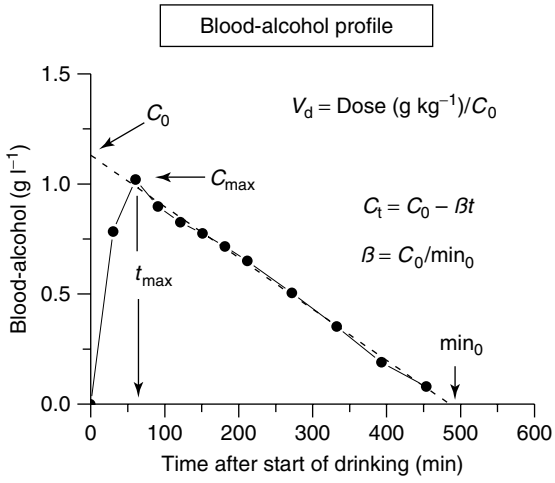


Figure 3 Concentration-time profile of ethanol in blood in one subject after consumption of 0.70 g ethanol per kilogram body weight as neat whisky on an empty stomach

the *y*-axis and the time of sampling blood, measured from the start of drinking, is plotted on the *x*-axis [12]. The shape of the resulting blood-alcohol curve furnishes important information about the absorption, distribution, and elimination processes. Hundreds or even thousands of BAC curves have been generated over the years by scientists from various countries by experiments with healthy volunteers under controlled conditions. The doses of ethanol generally ranges from 0.3 to 1.0 g kg^{-1} and blood samples are then taken for analysis at 15- to 30-min intervals for up to 3–10 h, depending on the dose. A plot of BAC *versus* time is a neat way to visualize the concentration-time profile of ethanol and absorption, distribution, and elimination stages of alcohol metabolism.

A typical blood-alcohol curve obtained after a male subject drank a moderate dose of alcohol as neat whisky on an empty stomach is shown in Figure 3. The quantitative evaluation of BAC profiles is fundamentally important in forensic pharmacology and the basic principles and concepts were developed and introduced already in the 1930s by a physician and scientist from Sweden, namely, Erik MP Widmark [19]. The results of this research remain valid today and have become the cornerstone of forensic blood-alcohol calculations and interpretation of BAC in relation to the quantity of alcohol a person has consumed.

The Widmark Equation

In 1922, a reliable method was presented for the quantitative determination of ethanol in small amounts of blood (~0.1 ml). The blood specimens were obtained without medical risk simply by pricking a fingertip or an earlobe [12]. The controlled experiments on human drinking showed that after the absorption of alcohol was complete and the peak concentration in blood (C_{max}) was passed the declining portion of the BAC curve followed more or less a straight line. This was interpreted to mean that the rate of metabolism of ethanol occurred at a constant rate independent of the prevailing BAC according to zero-order elimination kinetics [12, 19]. This contrasted with the metabolism and pharmacokinetics of most other drugs for which the rate of elimination from the blood decreases at lower concentrations as expected for a first-order process [20].

Extrapolating the linear declining portion of the BAC curve back to the time of starting to drink (the dotted line in Figure 3) gives the BAC expected if absorption and distribution had occurred instantaneously without any metabolism taking place. Widmark denoted the *y*-intercept on the blood-alcohol axis with the symbol C_0 that stands for concentration (BAC) at zero-time. This pharmacokinetic parameter has gained great importance in studies of clinical pharmacokinetics of ethanol and other drugs [21].

If the entire dose of ethanol had been absorbed and distributed evenly in all body fluids and tissues, one would have expected that the ratio of the dose (grams) to body weight (kilograms) would be identical (within experimental error) to the C_0 value in units of gram per kilogram. But this was not the case; the C_0 was always higher than the concentration of alcohol in the body as a whole (g kg^{-1}). Because ethanol dissolves in the body water compartment and solubility in lipids and bone is negligible, Widmark introduced a correction to the body weight (reduction factor), which he denoted as *r* or *rho* [12]. This *r*-factor was less than unity and meant that the concentration of alcohol in the body when calculated as g/(kg × r) was the same as C_0 given by the *y*-intercept of the BAC curve.

$$C_0 = \text{gram alcohol}/(\text{body weight} \times r) \quad (2)$$

$$\text{Grams alcohol } (A) = C_0 \times \text{body weight} \times r \quad (3)$$

The basic Widmark equation above expresses a quantitative relationship between the BAC (C_o) and

the amount of alcohol absorbed and distributed in all body fluids and tissues (A) at the time the blood was sampled. The Widmark equation can also be written as follows

$$A(\text{g}) = \text{BAC (mg g}^{-1}) \times \text{BWt (kg)} \times r \text{ (w/w)} \quad (4)$$

$$A(\text{g}) = \text{BAC (g l}^{-1}) \times \text{BWt (kg)} \times r \text{ (v/w)} \quad (5)$$

The decision as to which of the above two equations should be used depends on the concentration units used to report BAC, whether mass/mass or mass/volume. In Widmark's fundamental work, the BAC was reported as mass/mass (mg g^{-1} or g kg^{-1}) because the aliquots of blood analyzed were measured by weighing. The subject's body weight was in kilogram so the original r-factor was a dimensionless ratio. Today, it is more common to measure and report BAC in units of mass/volume (mg ml^{-1} or g l^{-1}) so the Widmark factor will have the units of liters per kilogram. The connection between mass/mass and mass/volume is the density of whole blood, which is 1.055 g ml^{-1} on average [11]. Accordingly, both the C_o value and the Widmark r-factor are 5.5% less if the BAC is measured and reported in mass/volume units.

Table 5 summarizes values of the distribution factor (r-factor) for men and women based on Widmark's work and the corresponding values if mass/volume (g l^{-1}) units are used to report BAC as is more common today. The gender-related difference (0.68 for men and 0.55 for women) depends primarily on differences in body composition and especially the lower TBW and the higher proportion of fat-to-lean tissue in women.

In any individual case, the r-factor depends on age, gender, adiposity, and BMI although population average values of 0.7 for men and 0.6 for women are generally used in BAC calculations [9]. An important assumption when using the Widmark equation is that the entire dose of alcohol has reached the systemic circulation at the time the blood sample was obtained. However, this is only feasible if alcohol was administered intravenously (100% bioavailability) although drinking a bolus dose on an empty stomach, so that gastric emptying is rapid, gives close to 100% availability.

During social drinking, some of the ingested alcohol might be metabolized either in the stomach or the liver or both organs before reaching the central compartment and this is known as *FPM*. This might be appreciable if the absorption of alcohol is slow and when drinking extends over several hours, such as when dining out and drinking with a meal or in small repetitive doses.

Widmark's equation is useful and widely used in forensic science practice to translate a person's BAC into the amount of alcohol absorbed and distributed in all body fluids. If alcohol remains unabsorbed in the stomach when the blood sample is taken, then it is not considered as being in the body; hence, the total quantity consumed is underestimated by use of this calculation.

Figure 4 shows the relative frequency distribution of BAC in drunken drivers apprehended in Sweden. The mean and median concentrations were about the same, being close to 1.8 g l^{-1} (180 mg/100 ml). Most drunken drivers (>85%) are men, so if one assumes an average body weight of 80 kg and a Widmark r-factor of 0.71 kg^{-1} a simple calculation using equation (4) or (5) shows that 100.8 g of ethanol are absorbed and distributed in the whole body at the time of sampling the blood. Using the information in Table 2 leads to the conclusion that there are 325 ml of whisky or vodka (40% v/v) in all body fluids.

Table 6 presents further examples of how the Widmark equation can be used to calculate the quantity of

Table 5 Mean, standard deviation, and range of Widmark r-factors depending on whether BAC is measured and reported as mass/mass (g kg^{-1}) or mass/volume (g l^{-1}) units

Gender	Widmark's r-factor, mean ± SD (range)[a]	Adjusted r-factor, mean ± SD (range)[b]
Men ($N = 20$)	0.68 ± 0.085 (0.51–0.90)	0.64 ± 0.081 (0.48–0.85)
Women ($N = 10$)	0.55 ± 0.055 (0.46–0.65)	0.52 ± 0.052 (0.44–0.62)

[a] Blood-alcohol reported as milligram per gram (gram per kilogram)
[b] Blood-alcohol reported as milligram per milliliter (gram per liter) where 1.0 mg g^{-1} = 1.055 mg ml^{-1}

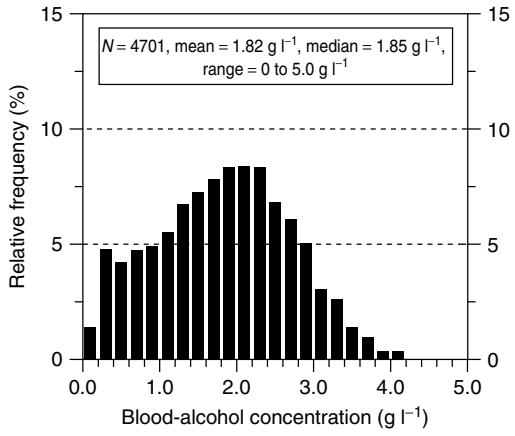


Figure 4 Relative frequency distribution of the concentrations of ethanol in blood from people apprehended by the police in Sweden for drunken driving

alcohol in the body at certain BAC spanning from 0.5 to 4.5 g l^{-1} (50–450 mg/100 ml) for men with body weights of 60, 80, or 100 kg. Besides the amount of ethanol absorbed and distributed in the body at the time of sampling blood (*r*-factor 0.7), some of the ingested alcohol is lost by hepatic metabolism during the 3-h drinking spree. A good rule of thumb for use in this calculation is that the rate of ethanol metabolism is 0.1 g kg^{-1} body weight per hour independent of gender.

Equations (4) and (5) can also be used to calculate the BAC expected at some future point in time after absorption and distribution are complete. Widmark insisted that such a calculation should not be attempted until at least 1.5–2 h had elapsed after the end of drinking. This amount of time was necessary to ensure that negligible amounts of alcohol

Table 6 Relationship between blood-alcohol concentration (BAC) and the quantity of alcohol absorbed and distributed in all body fluids of a healthy male subject. Also shown is the total quantity of ethanol consumed if metabolism had taken place over 3 h

BAC g l^{-1} (mg/100 ml)	Body weight kg[a]	Ethanol (g) in body	Amount metabolised during 3 h[b]	Ethanol (g) consumed[c]
0.5 (50)	60	21.1	6	27.1
	80	28.2	8	36.2
	100	35.2	10	45.2
0.8 (80)	60	33.8	6	39.8
	80	45.0	8	53.0
	100	56.3	10	63.3
1.0 (100)	60	42.2	6	48.2
	80	56.3	8	64.3
	100	70.4	10	80.4
1.5 (150)	60	63.4	6	69.4
	80	84.5	8	92.5
	100	105.6	10	115.6
2.0 (200)	60	84.5	6	90.5
	80	112.6	8	120.6
	100	140.8	10	150.8
2.5 (250)	60	105.6	6	111.6
	80	140.8	8	148.8
	100	176.0	10	186.0
3.5 (350)	60	126.2	6	132.2
	80	169.0	8	177.0
	100	246.4	10	256.4
4.5 (450)	60	190.1	6	196.1
	80	253.5	8	261.8
	100	316.9	10	326.9

[a] The calculations assume a healthy male subject with Widmark *r*-factor of 0.71 kg^{-1}
[b] Ethanol is metabolized at a rate of 0.1 g ethanol per kg body weight per hour from the start of drinking
[c] The sum of alcohol in body and the amount metabolized over 3 h

still remain unabsorbed in the gut when blood was sampled.

Total Amount of Alcohol Ingested

The metabolism and elimination of ethanol start to occur immediately after the start of drinking and continues for as long as alcohol is still measurable in the blood. When BAC drops below 0.1–0.2 g l^{-1}, the major metabolizing enzymes are no longer saturated with substrate and zero-order kinetics changes to first-order kinetics. The BAC time curve changes from a straight line to become curvilinear and the elimination rate thereafter decreases with diminishing BAC [22]. In reality, however, ethanol is a good example of a drug that displays dose-dependent saturation kinetics and the shape of the elimination phase spanning from high- to low-BAC looks more like the shape of a hockey stick rather than a straight line [20–22]. Mathematically, the concentration–time data in the postabsorptive part of the BAC *versus* time curve is best described by the Michaelis–Menten equation as

$$\text{Rate of metabolism } (-dC/dt) = (V_{max} \times \text{BAC})/K_m + \text{BAC} \quad (6)$$

This equation assumes the existence of a single enzymatic reaction with a Michaelis constant K_m and a maximum velocity of V_{max} and BAC represents the substrate concentration [20]. When the BAC is considerably higher than the value of K_m (0.05–0.1 g l^{-1} for ADH), the above equation simplifies to an equation resembling zero-order kinetics.

$$\text{Rate of metabolism } (-dC/dt) = V_{max} \quad (7)$$

When the BAC is much less than K_m, the Michaelis–Menten equation reduces to the equation for first-order kinetics with the elimination rate being directly proportional to the substrate concentration. The first-order rate constant (k_1) is given by the ratio of V_{max}/K_m.

$$\text{Rate of metabolism } (-dC/dt) = (V_{max}/K_m) \times \text{BAC} \quad (8)$$

$$\text{Rate of metabolism } (-dC/dt) = k_1 \times \text{BAC} \quad (9)$$

Since BAC greatly exceeds 0.2 g l^{-1} in most forensic science applications, the concepts of zero-order kinetics as developed by Widmark are still widely used and remain valid.

As mentioned above, the rate of ethanol metabolism from the whole body is close to 0.1 g kg^{-1} body weight per hour or 7 g h^{-1} for a 70 kg person. The Widmark equation, which was used above to calculate the quantity of alcohol in the body, can be expanded to allow for metabolism of ethanol since the time of starting to drink ($0.1 \times \text{BWt} \times t_h$), where t_h is the number of hours elapsed.

$$A = [(\text{BAC} \times \text{BWt} \times r) + (0.1 \times \text{BWt} \times t_h)] \quad (10)$$

The above formula is useful to evaluate statements made by drunken drivers as to how much alcohol had actually been consumed that evening. The quantity of alcohol obtained in grams can then be converted into the volumes of different kinds of beverages depending on their alcoholic strength (Table 2).

Making forward projections of BAC from information provided by a person suspected of alcohol-impaired driving is not recommended because of the obvious lack of truthfulness in such statements. Moreover, there is always uncertainty, as mentioned above, in the extent of FPM of orally ingested alcohol. The concentration of alcohol in a specimen of venous blood provides the most reliable information about the amount of alcohol a person has consumed. Use of the above equation tends to give a conservative estimate of the amount of alcohol actually ingested because bioavailability of the dose is rarely 100%.

The BAC expected to exist some time after the end of drinking when absorption and distribution are complete is given by the equation below, where t_h is the time in hours after drinking started, β is the factor developed and used by Widmark to represent the rate of alcohol elimination form the bloodstream (see later), A is dose (g) of ethanol ingested, kg is body weight, and r is the apparent volume of distribution of ethanol.

$$\text{BAC} = (A/\text{kg} \times r) - (\beta \times t) \quad (11)$$

Updating the Widmark Equation

Widmark's pioneer work on alcohol pharmacokinetics was done in the 1930s with 20 men and 10 women serving as the test subjects [12]. A number of attempts have been made to update the Widmark factors, which seemed to be motivated considering that body composition has changed considerably since the 1930s with more obese people in today's society [13].

The results from many controlled drinking experiments led to the conclusion that C_0 values as derived by Widmark's method were less than expected, which meant that the r-factor (ratio of dose/C_o) was correspondingly higher. This finding led to the notion of an alcohol absorption deficit amounting to about 10% of the dose of ethanol administered. The most likely explanation for this deficit is FPM of ethanol occurring in the stomach, the liver, or both organs. Accordingly, there is a general agreement that under social drinking conditions a part of the dose of alcohol ingested (~10%) never actually reaches the systemic circulation, which has consequences when blood-alcohol calculations are made for legal purposes.

Another modification of the Widmark formula entailed calculating a person's TBW based on readily available anthropometric data such as age (years), weight (kg) the height (cm) [23]. The TBW derived in this way and knowledge of the water content of whole blood (80% w/w or 84% w/v) allow calculation of the Widmark r-factor as the ratio between the percentages of water in the body to that in the blood. Two widely used formulae for computing TBW for men and women using anthropometric data are given below:

$$\text{TBW (liters for men)} = 2.447 - 0.09516\ \text{age (years)} + 0.1074\ \text{height (cm)} + 0.3362\ \text{weight (kg)} \quad (12)$$

Residual standard deviation = 3.78 l

$$\text{TBW (liters for women)} = -2.097 + 0.1069\ \text{height(cm)} + 0.2466\ \text{weight (kg)} \quad (13)$$

Residual standard deviation = 3.60 l

The above equations offer the advantage that they are tailored for the individual concerned and consider differences in body shape and size. Note that the water content of blood in this equation should have the same unit mass/volume or mass/mass as used for reporting the BAC. For a man aged 45 years, a body weight of 85 kg, and a height of 184 cm, use of equation 12 gives TBW as 46.5 l or 54% of body weight. If the water content of whole blood is taken as 84 g/100 ml (84% w/v), the ratio of TBW (%) to blood water (%) is 0.64 l kg^{-1}, which is another way to derive the Widmark r-factor for this person.

The residual standard deviation (SD) in the above equations gives an idea of the amount of uncertainty in the calculation of TBW using anthropometric data. Thus, 95% of individuals will have a TBW within $\pm 2 \times$ SD of mean, which corresponds to roughly $\pm$ 20% for a person with a TBW of 40 l.

Interindividual Variations in Elimination Rate of Alcohol from Blood

Two pharmacokinetic parameters of alcohol are important when BAC is interpreted in forensic science and legal medicine. The first is the distribution volume of ethanol, which corresponds to the Widmark r-factor discussed in detail above and the second is the rate of elimination of alcohol from the bloodstream, corresponding to the slope of the linear elimination phase (see Figure 3). On the basis of hundreds of controlled alcohol dosing studies in healthy volunteers as well as in alcoholics during detoxification and in drunken drivers, a wide range of values for the alcohol burn-off rate have been published [10, 22].

Figure 5 presents a cumulative frequency distribution of the elimination rates of alcohol from blood in drunken drivers. These values were derived by taking two blood samples about 60 min apart from each apprehended driver and using the equation β-slope = $(BAC_1 - BAC_2)/t_{diff}$ [24]. This calculation assumes

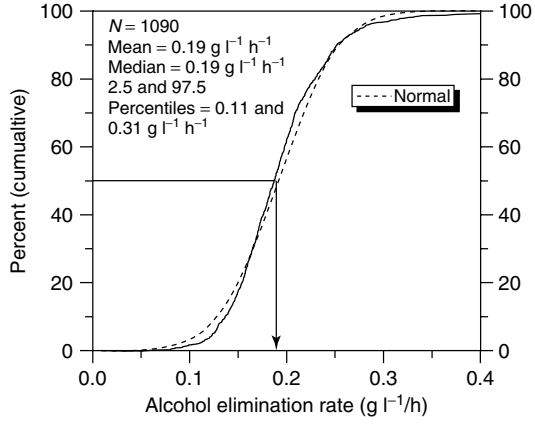


Figure 5 Cumulative frequency distribution (%) of elimination rates of ethanol from blood in drunken drivers derived from double blood samples (1 h apart), which assumes the existence of the postabsorptive stage of the blood-alcohol curve when the first blood sample was taken

that all individuals are already in the postabsorptive phase of the BAC curve when the first sample of blood was taken. The frequency distribution shows a good fit to a Gaussian curve with mean and median values 0.19 $g\,l^{-1}\,h^{-1}$ and a 95% range from 0.11 to 0.31 $g\,l^{-1}\,h^{-1}$. Values less than 0.1 $g\,l^{-1}\,h^{-1}$ should be considered suspect and probably represent those individuals still absorbing alcohol when the first-blood sample was taken.

On the basis of hundreds of controlled alcohol dosing studies with moderate drinkers as well as alcohol-dependent individuals, a realistic physiological range of ethanol elimination rates from blood in humans (Widmark's β-factor) is given in Table 7. The rate of elimination of ethanol in the vast majority of people can be expected to range form 0.1 to 0.25 $g\,l^{-1}\,h^{-1}$ and these extreme values are recommended for use if a retrograde extrapolation is made [5, 25]. The average elimination rate in drunken drivers is 0.19 $g\,l^{-1}\,h^{-1}$, which can be compared with 0.15 $g\,l^{-1}\,h^{-1}$ in moderate drinkers. The difference can be explained by a higher proportion of heavy drinkers and alcoholics among apprehended drivers having an enhanced capacity to metabolize alcohol via the CYP2E1 pathway.

There is some evidence of racial and ethnic differences in the rate of ethanol elimination from the bloodstream, because of polymorphism of the class I hepatic ADH enzyme. Studies have shown that metabolism of ethanol in East Asians, such as Japanese, occurs at a slightly faster rate than in Caucasians, judging by a steeper slope of the BAC curve in the postabsorptive state [10, 17]. Genetic polymorphism of the class I ADH enzyme means that Asians inherit a β_2 enzyme with higher V_{max} for ethanol as substrate [10]. Nevertheless, Asians show a considerable overlap with Caucasians and African Americans in the elimination rate of ethanol from blood, so the racial differences probably lack any forensic significance.

Forward and Backward Extrapolation of BAC

The BAC is a valid and reliable measure of the concentration of ethanol in the blood only at the time of sampling. In forensic casework, police authorities often want to know what the BAC was a few hours before or after the blood sample was taken [5, 25]. For example, after involvement in a traffic crash, a blood sample from the driver might not be obtained until several hours afterwards and the question naturally arises what was the BAC at the time of the crash.

Another situation arises in cases of alleged sexual assault when the victim reports the attack several hours afterward. The long delay between time of the

Table 7 Suggested physiological range of ethanol elimination rates from blood and the body as a whole under various conditions and circumstances

Elimination rate from blood ($g\,l^{-1}\,h^{-1}$)	Elimination rate form body ($g\,h^{-1}$)[(a)]	Subject conditions, treatment, or special circumstances
0.08–0.10	4–5	People with liver dysfunction (e.g., owing to cirrhosis or carcinoma) or who might be malnourished or on low-protein diets. Treatment with the drug fomepizole (4-methyl pyrazole), which blocks activity of alcohol dehydrogenase.
0.10–0.12	5–6	Consumption of moderate doses of alcohol by healthy individuals after an overnight (10 h) fast.
0.12–0.16	6–8	Consumption of moderate doses of ethanol under nonfasting conditions.
0.16–0.25	8–12	Drinking by healthy individuals to reach intoxicating BAC ($>1.2\,g\,l^{-1}$ or 120 mg/100 ml) such as in many drunken drivers.
0.25–0.35	12–17	Alcoholics during detoxification and people that drink continuously over several days or weeks to reach high BAC ($>3\,g\,l^{-1}$). People with a genetic predisposition or in a hypermetabolic state (e.g., after burn trauma or hyperthyroidism).

[(a)]This calculation assumes a normal nonobese person with a body weight of 70 kg and a Widmark r-factor (distribution volume) of 0.71 kg^{-1}

crime and obtaining a blood sample for toxicological analysis means that some drugs are no longer measurable. The police therefore might want to know what the BAC was at the time of the alleged assault and whether the victim was incapacitated because of too much drink or drugs [26].

In an actual case of drunken driving, a man was arrested by the police and a blood sample was taken for determination of alcohol content. The man was released but sometime afterward he was again observed driving, although on this occasion a blood sample was not available for analysis. Under these circumstances, the police might want to know whether the man was above the legal limit on the second occasion based on the concentration in the first blood sample and information about the rate of alcohol metabolism.

Since the metabolism of ethanol occurs at a constant rate per unit time in a large segment of the BAC curve, this makes forward or backward extrapolation feasible under certain circumstances [5, 10, 25]. The first condition necessary is that the postabsorptive phase was reached at the time of driving or when a crash occurred (e.g., 90 min in Figure 3). The second is that the rate of alcohol elimination from the bloodstream is known or a value that does not prejudice the suspect (e.g., $0.1\,g\,l^{-1}\,h^{-1}$) is assumed. Third, a deduction should be made from the average BAC to allow for uncertainty in the analysis and back-extrapolation should then start from this lower value and not the average BAC.

Interpreting Blood-Alcohol Concentrations in Samples Obtained at Autopsy

The use and abuse of alcohol in society has meant that drunkenness is a prime factor in many accidents on the roads, at home and in the workplace [3]. Alcohol intoxication is often an underlying factor in violent crime such as homicide, muggings, and sexual assault. Alcohol therefore tops the list of drugs identified in blood specimens taken during routine forensic autopsies and investigations of all out of hospital deaths [27, 28].

During investigations of road-traffic fatalities, a key question is whether the driver was under the influence of alcohol or drugs at the time of the crash. This becomes important when responsibility for the crash is investigated and insurance claims are made. Likewise, alcohol intoxication and drunkenness are prominent in many other traumatic events including fights, homicides, suicides, and drowning. Knowledge of the BAC in the deceased provides useful information when the cause of death is determined.

The methods used to measure alcohol (ethanol) in blood and other body fluids are the same regardless of whether specimens are obtained from the living or the dead [29]. However, great care is needed when the results of postmortem analysis are interpreted and a conclusion is reached about the state of inebriation or drunkenness at the time of death. Unlike drawing blood from a living person, there are a number of artifacts that should be considered in postmortem toxicology [27–29]. However, the situation is simplified because a much wider section of biological specimens are available from a corpse, thus making toxicological results easier to interpret.

Postmortem Aspects

The quality and composition of the blood samples obtained at autopsy can vary widely depending on the circumstances surrounding the death and the condition of the body, such as degree of trauma and whether decomposition and putrefaction had commenced [28–30].

The recommended sampling site for blood in postmortem toxicology is a femoral vein after cross-clamping. Sampling blood from the heart or pleural cavity is not recommended because of the risk that alcohol might have escaped from the stomach to contaminate the sampling site. Aspiration of stomach contents during the agonal period means that alcohol gets into plural cavity blood via the lungs. After death the body should be handled with care by the crime scene investigators or during transportation and storage at the mortuary to minimize spread of any alcohol. All such factors can lead to redistribution of alcohol from stomach contents if heavy drinking had occurred just prior to death [31].

In bodies autopsied within 24 h of death and when trauma is minimal, the concentration of alcohol in femoral venous blood is the closest one can come to knowing the antemortem concentration. Indeed, there is some evidence that a decrease in concentration occurs after death because metabolizing enzymes retain some activity for a few hours after death as the body cools. If the autopsy and sampling of blood

for toxicology is done the same day, then a chemical preservative, such as sodium fluoride, is probably not necessary. For longer delays and when specimens are sent by mail to another laboratory, it is imperative to include sodium or potassium fluoride as a preservative in blood and other biological specimens (1–2% w/v). The fluoride ion functions as an enzyme inhibitor and prevents the synthesis of ethanol by fermentation processes after sampling. However, this does not rule out that ethanol had already been produced before the autopsy was performed [30, 32] (*see also* **Toxicology: Analysis**.)

Alternative Body Fluids

The results of postmortem blood-alcohol analysis are considerably strengthened if other body fluids are submitted for analysis along with the blood sample (e.g., urine and vitreous humor (VH)). In postmortem work in addition to cardiac or femoral venous blood, urine, VH, and cerebrospinal fluid (CSF) are the most useful specimens for determination of ethanol [33, 34]. These other biofluids, which are ~100% water, are obtained from the urinary bladder, the eye, and the back of the neck (cisternal fluid), respectively. The concentrations of alcohol depend on water content and also on time after drinking and the status of alcohol absorption and distribution in the body when death occurred. The concentrations of alcohol measured in VH, urine, and CSF are displaced in time compared with the venous BAC. During the absorption stage, the concentrations of alcohol are lower in these alternative fluids compared to the blood, although 1–2 h later when equilibration of ethanol in all body fluids is complete, the VH, urine, and CSF contain a higher concentration, by ~20%, compared with that of the blood samples.

Vitreous fluid from the eye is an important biological specimen in postmortem toxicology for several reasons. First, VH is a relatively clean watery fluid easily obtained with syringe and needle even without conducting a complete autopsy. Second, the remote nature of the eyes compared with the gut means that the spread of bacteria to contaminate the VH is much less likely than the contamination of central or peripheral blood during autolysis [34]. When the corpse has been subjected to severe trauma, a suitable blood sample might not be available and with some reservations the BAC can be estimated indirectly by the analysis of ethanol in the VH or the urine, albeit with considerable uncertainty.

When the body has undergone putrefaction, the BAC result is obviously suspect, owing to the risk of microbial formation of ethanol from glucose present in the blood. Under these circumstances, VH gives a more reliable indication of whether the deceased had consumed alcohol and might have been drunk at the time of death. Figure 6 shows a high correlation between the concentrations of ethanol in VH and in femoral blood ($r = 0.98$) at autopsy. The average VH/BAC ratio of ethanol concentration was 1.19 : 1 (SD = 0.28) and the range was from 0.28 to 2.90 [34].

The distribution ratios for urine/blood, vitreous/blood, and CSF/blood change as a function of time after drinking. Much depends on the position of the alcohol curve and absorption or postabsorptive phase of metabolism at the moment of death [27]. On the ascending limb of the BAC profile, the concentrations of ethanol in urine, VH, and CSF are lower than or about the same as in venous blood. In the postpeak descending limb of the BAC profile, which corresponds to the postabsorptive phase, the concentration of ethanol in urine, VH, and CSR are always higher than that in the blood. Indeed, alcohol might still be measurable in these alternative specimens even though BAC is reported as negative. Table 8 compares the water contents of body fluids and tissues usually available for the analysis of ethanol at autopsy and gives expected ratios of the concentrations of ethanol relative to blood samples.

Traumatic injuries and death by blunt force with open wounds and massive blood loss are factors that increase the risk of bacteria entering the body. Under these circumstances fermentation processes might produce ethanol before an autopsy is performed. This risk is heightened at elevated environmental temperatures (summer months) and when a long time elapses before recovery of the body, e.g., after an air disaster or death at sea [28]. Although a fluoride preservative is routinely added to blood specimens taken at autopsy, some alcohol might have been synthesized in body cavities between the time of death and autopsy.

After skull trauma, blood is often sampled from a subdural hematoma or a clot in the brain, which furnishes useful information in postmortem toxicology. Because of reduced or nonexistent blood circulation to the clot, any alcohol it contains does not

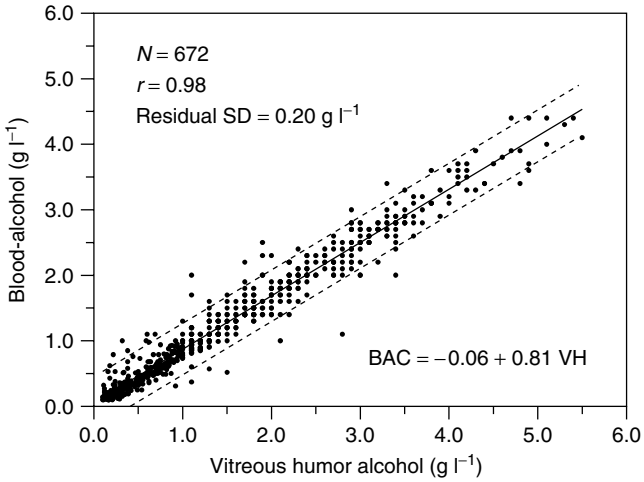


Figure 6 Correlation between the concentrations of ethanol determined in femoral venous blood and vitreous humor samples obtained at autopsy

Table 8 Average water contents of whole blood, body fluids, organs, and tissues and the concentration ratios of ethanol relative to whole blood

Specimen	Water content (% w/w)	Ratio of ethanol concentration relative to whole blood[a]
Whole blood[b]	78–81	1.0
Plasma/serum	91–93	1.1–1.2
Erythrocytes	68–71	0.8–0.9
Urine	98–99	1.2–1.4[c]
Vitreous humor	99–100	1.1–1.3
Cerebrospinal fluid	98–99	1.1–1.3[d]
Bile	87–97	0.8–1.0
Synovial fluid	92–96	1.1–1.2
Liver	80	0.6–0.8
Brain	75	0.8–1.0
Skeletal muscle	76	0.8–0.9
Kidney	79	0.6–0.7

[a] The values can differ depending on the concentration of alcohol in the samples and the time after drinking when death occurred
[b] Depends on water content and hematocrit of blood sample, which in turn depends on gender and the condition of the body
[c] The ratio depends on the stage of alcohol absorption and distribution at time of death
[d] Lumbar fluid taken during postabsorptive phase

undergo metabolism. The concentration of ethanol in the sequestered hematoma therefore gives an indication of the person's BAC some hours earlier, such as when the trauma occurred. For example, if a person suffers a blow to the head when drunk but survives for many hours before death, considerable amounts of alcohol are eliminated by metabolism ($0.15\,g\,l^{-1}\,h^{-1}$). If 10 h elapses before death, then a BAC of $1.5\,g\,l^{-1}$ is no longer measurable. However, the concentration measured in the blood clot can furnish useful information about the BAC at an earlier time, such as when the clot was formed. Contamination of the wound with bacteria and the rate of formation of the clot and other factors need to be considered.

Any emergency life-saving treatment administered at the crash site, such as intravenous fluids to counteract shock, and massive blood loss are also important to consider. Table 9 lists many of the considerations necessary when results of postmortem alcohol analysis are reported and interpreted.

Biochemical Markers of Alcohol Consumption

Considerable research has been done to develop a biochemical marker of acute and chronic alcohol consumption. In this connection, the formation of the nonoxidative metabolites of ethanol, namely, EtG and EtS, has attracted a lot of attention [35, 36]. The presence of these metabolites in body fluids taken at autopsy means that ethanol must have undergone metabolism during life, which supports

Table 9 Factors to consider when the results of blood-alcohol analysis in postmortem specimens are interpreted

1. Specificity of the analytical method and whether other volatile substances might have interfered with the determination of ethanol.
2. Time between death, recovery of the body, and the postmortem examination.
3. Condition and location of the body:
   (a) Indoors or outdoors
   (b) Time of year
   (c) Temperature and humidity of the environment
   (d) Extent of traumatic injuries
   (e) Abdominal trauma, e.g., ruptured stomach
   (f) Incinerated body
   (g) Body recovered from water
   (h) Extent of decomposition/putrefaction, bad smell, maggots, etc
4. Was a preservative (NaF ~2%) added to all specimens submitted for toxicology?
5. Risk of contamination of specimens with extraneous solvents during emergency service treatment at hospital, the mortuary, or the analytical laboratory.
6. Was the body embalmed and if so what embalming fluids were used and did they contain any alcohols?
7. Compare and contrast the concentrations of ethanol expected in different sampling sites after considering water content of the specimens (blood, urine, and vitreous humor)
8. Should the water content of blood be determined and results of alcohol analysis adjusted to blood-water content of 80% w/w?
9. Variations in alcohol concentration depending on absorption or postabsorptive stage of the blood-alcohol curve when death occurred.
10. How was the body handled at the scene of death and during transport and storage at the mortuary?
11. Any evidence of recent consumption of alcohol before death?
12. Concentration of ethanol in stomach contents, risk of postmortem diffusion, and aspiration of vomit during the agonal period.
13. Changes in ethanol concentration caused by evaporation, dilution, or degradation of ethanol after death.

antemortem consumption of alcohol. If ethanol was present in blood and urine taken at postmortem and EtG and EtS were shown to be negative, then this would raise a warning flag that the ethanol was probably produced after death as a result of microorganisms acting on carbohydrates or other substrates. Even other urinary markers of acute alcohol ingestion such as the metabolites of serotonin and the ratio of 5-hydroxytryptophol (5HTOL) to 5-hydroxyindole acetic acid (5HIAA) are available [36].

Biochemical markers are useful to detect damage to organ and tissue as a consequence of heavy drinking so that treatment can be given before liver cirrhosis and death occurs. This entails measuring the concentration of certain hepatic enzymes such as gamma-glutamyltransferase (GGT), aspartate aminotransaminases, and alanine aminotransaminases (AST and ALT). Concentrations are elevated in serum after a long period of heavy drinking because they leak into the bloodstream when the liver is damaged by the alcohol [37]. Enlarged size of the red blood cells, mean corpuscular volume (MCV), is another widely used marker of chronic drinking in combination with the enzyme markers [36].

Carbohydrate deficient transferrin (CDT) is another marker of heavy drinking used in clinical medicine but also with forensic applications thanks to good sensitivity and high specificity. Transferrin is a plasma protein synthesized and secreted by the liver that serves to transport iron in the body [36]. After a period of continuous heavy drinking (60–80 g/day over many weeks), glycosylation of the transferrin molecule is altered and this is reflected in the loss of one or more of the carbohydrate moieties, hence the name carbohydrate deficient.

The methods used to determine serum CDT are now much improved, e.g., by use of capillary electrophoresis and high performance liquid chromatography (HPLC), and interlaboratory reporting of the results of CDT analysis is now better standardized. CDT is the marker of choice to screen individuals for excessive drinking before the liver is too severely damaged and helps to corroborate results from the use of clinical interviews and questionnaires [35–37]. CDT is considered to be the most specific marker of

Table 10 Typical signs and symptoms of alcohol influence in relation to the person's blood-alcohol concentration (BAC)

BAC $g\,l^{-1}$ (mg/100 ml)	Signs and symptoms of alcohol influence[a]
0.0–0.2 (0–20)	Sobriety, no untoward effects, or outward signs
0.3–0.5 (30–50)	Less inhibited (euphoria), more talkative, impairment of certain cognitive tasks and skills that require divided attention
0.5–1.0 (50–100)	More rowdy and daring, sensory and motor disturbances, slowed reaction time especially in choice situations
1.0–1.5 (100–150)	Lack of coordination, unsteady gait, slurred speech, prolonged reaction to sights and sounds.
1.5–2.0 (150–200)	Obvious drunkenness, aggressive behavior, and ataxia significantly slowed reaction time even when relatively simple tasks are performed; nausea and vomiting in some individuals especially if the BAC increases rapidly to reach these levels
2.0–3.0 (200–300)	Not able to stand upright or walk unaided, incoherent speech, and motor areas of the brain are markedly influenced, thus distorted perception of time and judgment and near comatose state
3.0–4–0 (300–400)	Confusion, stupor, or coma with shallow breathing, risk of respiratory arrest, loss of gag reflex, and risk of inhalation of vomit
4.0–5.0 (400–500)	Profound risk of death from respiratory paralysis and cardiopulmonary arrest

[a]Both subjective and objective measures of alcohol influence are more pronounced on the rising part of the BAC curve close to the C_{max} compared with several hours later during the postabsorptive phase. This difference in the effects of ethanol is related to the phenomenon of acute tolerance or the Mellanby effect (see main text for details)

excessive drinking and false-positive test results are rare. CDT has found forensic applications to evaluate drinking habits of convicted drunken drivers before they are allowed to retake a drinking test. Biomarkers find use in many other situations when there is a need to verify that a person is no longer a problem drinker, such as when it comes to rehabilitation of alcoholics and custody of children, etc.

Development of Tolerance to Ethanol

It is common knowledge that the effects of drugs differ widely between different individuals despite intake of the same dose under otherwise identical conditions [16]. Regular drinkers can tolerate more alcohol and appear less influenced compared with moderate or nondrinkers with the same BAC. Among other things, this raises the question of whether BAC is really such a reliable indicator of the degree of alcohol influence in any individual case.

Table 10 lists various clinical signs and symptoms of alcohol influence as a function of increasing BAC. Much variation can be expected in the individual case, depending on the time after drinking when the observations are made. The degree of impairment after drinking alcohol depends on the speed of drinking, and elapsed time after the end of drinking when tests or observations were made. People feel more impairment on the rising limb of the BAC curve compared with the declining limb although previous experience with drinking alcohol and habituation are important considerations.

The development of tolerance to the effects of alcohol or other drugs and the underlying mechanisms involved are complex topics involving behavioral changes and learning effects [38, 39]. Tolerance implies the ability to lessen the untoward effects of drugs after repeated exposure. To the pharmacologist, tolerance is often illustrated by a shift in the sigmoid dose–response curve to the right so that larger doses of a drug are required to produce the same degree of effect on the individual. Several different kinds of tolerance need to be distinguished and defined in connection with interpretation of BAC in forensic casework.

The first and simplest kind of tolerance is referred to as *dispositional tolerance*, which has to do with altered absorption, distribution, metabolism, or excretion of ethanol during chronic treatment. The rate of absorption of ethanol is closely linked to gastric emptying, which in turn depends on factors such as dose and speed of drinking, fed or fasting state, concentration of ethanol in the beverage, and various hormonal influences [10]. Speed of distribution is determined by blood flow and tissue mass and the proportion of lean-to-fatty tissue in the body. The amounts of ethanol eliminated via breath, sweat, and urine are trivial (2–5%) and are dependent on simple physical

diffusion, which in turn is a function of the BAC. These diffusion processes are not markedly altered after a period of continuous heavy drinking.

The rate of metabolism of ethanol is accelerated after a period of continuous heavy drinking, although the amounts of ethanol that have to be consumed and the duration of drinking are not well defined. The mechanism seems to involve the CYP2EI enzyme, which acquires an enhanced capacity to oxidize ethanol after chronic administration of substrates, such as ethanol as well as other drugs (e.g., barbiturates). This faster elimination of certain drugs from the body after chronic exposure is referred to as *metabolic tolerance*. However, studies have shown that after a few days of abstinence from alcohol, the rate of metabolism returns to normal as CYP2E1 enzyme becomes resynthesized. Thus, heavy drinkers or alcoholics do not necessarily show a faster rate of ethanol metabolism after they have remained sober a few days. Some alcoholics might have an abnormally slow rate of metabolism owing to malnutrition and liver damage (cirrhosis) caused by the abuse of alcohol.

A well-studied and well-recognized form of tolerance is *acute tolerance*, which develops during a single exposure to ethanol [40]. The degree of intoxication, both subjective and objective, changes as a function of time after the end of drinking a known amount of ethanol. The signs and symptoms of alcohol influence are less pronounced on the descending limb of the BAC curve about 3–4 h after the end of drinking compared with the ascending limb before reaching C_{max}.

Acute tolerance to alcohol is therefore characterized by a diminished effect of the drug at the same blood or brain concentration on the descending phase of the BAC curve compared with the ascending phase [41]. This phenomenon seems to be related to an adaptation of the brain cells and membrane receptors to an alcohol environment within a few hours of exposure to a single intoxicating dose. The exact mechanisms involved are unknown but seem to involve both learning phenomenon and adaptation to the impairment effects of the drug [38, 39]. Acute tolerance to alcohol is sometimes referred to as the *Mellanby effect*, named after the British pharmacologist (Sir Edward Mellanby), who first observed the phenomenon and published results of his experiments of alcohol intoxication in dogs [41].

Chronic tolerance is a more gradual process resulting in a diminished intoxicating effect after a period of continuous heavy drinking lasting several days or weeks without a period of abstinence. Those with a developed chronic tolerance at first sight will not exhibit overt signs of drunkenness, apart from a strong smell of alcohol on the breath, with BAC of $2–3\,g\,l^{-1}$ (0.2–0.3 g%). Table 11 shows the results from a clinical examination and questionnaire administered to a large number of apprehended drunken drivers 2 h or more after they were arrested. The examining physicians were not aware of the person's BAC and the assessment was based on answers to questions and simple clinical tests of motor and cognitive functioning. The physicians were asked to grade their findings and to conclude whether the people examined were not influenced, slightly influenced, moderately influenced, or heavily under the influence of alcohol.

The data in the table demonstrate large discrepancies at one and the same BAC depending on, among other things, differences in training, experience, and enthusiasm on the part of the physician for

Table 11 Relationship between blood-alcohol concentration (BAC), number and percent of individuals judged not, slightly, moderately, or severely influenced by alcohol according to a clinical examination and questionnaire administered by a physician (police surgeon)[(a)]

BAC ($g\,l^{-1}$) (mg/100 ml)	N	Not under influence (%)	Slightly under influence (%)	Moderately under influence (%)	Severely under influence (%)
0.0–0.49 (0–49)	103	52 (50)	47 (46)	4 (4)	0 (0)
0.50–0.99 (50–99)	223	111 (49)	90 (40)	22 (10)	0 (0)
1.00–1.49 (100–149)	226	58 (26)	107 (47)	59 (26)	2 (2)
1.50–1.99 (150–199)	195	17 (9)	65 (33)	81 (41)	32 (16)
2.00–2.49 (200–249)	104	5 (5)	31 (30)	52 (50)	16 (15)
2.50–2.99 (250–299)	44	0 (0)	9 (20)	21 (48)	14 (32)

[(a)] Those examined were suspected for drunken driving in Sweden

this task, and inherent differences in tolerance (acute and chronic) in the people being examined.

The type of tolerance that develops after long-term use of alcohol is sometimes referred to as *behavioral or functional tolerance* because it relates to an altered sensitivity of the brain to the effects of the drug. The underlying mechanism of functional tolerance, which manifests in marked behavioral effects of the drug after repeated exposure, is complex involving environmental, psychological, and genetic factors. The drinking pattern and amounts of alcohol necessary to elicit a chronic tolerance also causes physical dependence. Evidence of this comes from the physiological disturbances after cessation of drinking and when the BAC drops toward a zero concentration. Indeed, the severity of physiological disturbances during alcohol withdrawal can prove life threatening. To alleviate the withdrawal symptoms, patients are treated with drugs, such as barbiturates or benzodiazepines, that exhibit cross-tolerance to ethanol. The corresponding withdrawal stage after acute ingestion of ethanol (a single exposure) corresponds to the well-known hangover syndrome.

How Dangerous is Ethanol?

In the field of forensic toxicology, drugs are classified as either licit or illicit and also in terms of their toxicity. Along with nicotine from use of tobacco products and caffeine from consumption of coffee or tea, ethanol is a socially accepted legal drug. All pharmaceutical products are legal drugs although many such as opiates and benzodiazepines are classified as narcotics and are subject to abuse. The classic illicit drugs are heroin, cannabis, cocaine, ecstasy, lysergic acid diethylamide (LSD), amphetamine, methamphetamine, and gamma-hydroxybutyrate (GHB). When assessing the dangerousness of drugs, there are many factors to consider besides the risk of acute toxicity, such as dependence liability and social and medical harm caused to the individual and the costs to society for treatment and rehabilitation of those addicted to drugs [42].

With regard to acute toxicity, ethanol should be considered a fairly dangerous drug. The BAC associated with mild euphoria ($0.4–0.5\,g\,l^{-1}$) is only 10 times less than the BAC that could cause death ($4.0–5.0\,g\,l^{-1}$), through respiratory paralysis. This gives a ratio of lethal dose to effective dose of 10 : 1, which is a narrow safety margin for a legal drug. Besides acute toxicity, it is also important to consider the potential of a drug to cause physical and psychological dependence and craving.

In a large case series of forensic autopsies ($N = 693$), the concentration of ethanol in femoral venous blood was evaluated when the cause of death was attributed to acute alcohol poisoning [43]. Figure 7 depicts the frequency distribution of BAC in these blood specimens and this shows a good fit to a bell-shaped or Gaussian curve. This means that ~95% of cases are within ±2 SD around the mean. The mean, SD, and median concentration of ethanol in femoral

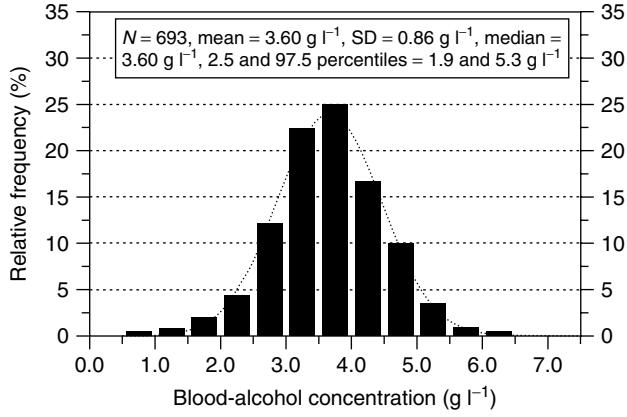


Figure 7 Frequency distribution of the concentration of ethanol in femoral blood in deaths attributed to acute alcohol poisoning (no other drugs present in blood) by the pathologist

blood were 3.6 , 0.86, and 3.6 g l^{-1}, respectively, and therefore 95% of cases are within 1.9 and 5.3 g l^{-1} (mean ±2SD).

Multiple sampling is a golden rule in forensic medicine and toxicology, so besides blood or plasma urine should be also obtained and sent for the analysis of ethanol and other drugs. Figure 8 shows the frequency distribution of urine–alcohol concentration (UAC) in deaths attributed to acute ethanol poisoning and when no other drugs were detected [44]. The mean and median concentrations are noticeably shifted toward higher values compared with BAC (Figure 7) as expected from the difference in water content between blood and urine. The frequency distribution of UAC is also a good fit to a Gaussian curve; mean, median, and 2.5 and 97.5 percentiles of 4.3, 4.3, and 2.4 and 6.1 g l^{-1}, respectively.

The UAC is higher than BAC not only because of the higher water content of ~100% in urine compared with that of ~80% in blood, but also because most people die after ethanol is fully equilibrated in all body fluids [44]. The value of the UAC/BAC ratio can give a clue to whether absorptive was ongoing (UAC/BAC values <1.1) or whether the postabsorptive phase was reached (UAC/BAC values >1.25) at the time of death [33, 44].

The BAC and UAC curves are shifted in time with UAC being less than BAC during the absorption phase and always higher than BAC in the postabsorptive period (Figure 9**).** This difference in the concentration–time profiles of ethanol in successive urine voids and in blood has implications for the pharmacokinetic parameters of ethanol in blood and urine. In the experiment depicted in Figure 9, each subject emptied the bladder of residual urine before the start of drinking. Alcohol was served as neat whisky (0.85 g kg^{-1}), which was finished in 25 min and consumed on an empty stomach after an overnight fast.

Lethal Dose of Ethanol

Deaths caused by uncomplicated alcohol poisoning, that is, through depression of the central nervous system and paralysis of respiration and collapse of circulation generally occur at a BAC between 4 and 5 g l^{-1} [40, 43]. After drinking massive amounts of alcohol, the individual progresses through various stages of intoxication (see Table 10) before death occurs. If a BAC of 5 g l^{-1} is considered sufficient to cause death, then the Widmark equation can be used to calculate the amount of ethanol that needs to be consumed to reach this level.

Table 12 gives the results of this calculation showing the number of grams of ethanol that are absorbed and distributed in all body fluids in a person with a BAC of 5 g l^{-1}. The results are illustrated for males and females with body weights ranging from 10 to 100 kg. Besides the mean quantity of ethanol in the body, the 95% range is included based on knowledge about intersubject variation in the distribution volume of ethanol (Widmark's *r*-factor). The lethal dose of ethanol is slightly higher for men 3.5 g kg^{-1} (range 2.8–4.2 g kg^{-1}) compared with that

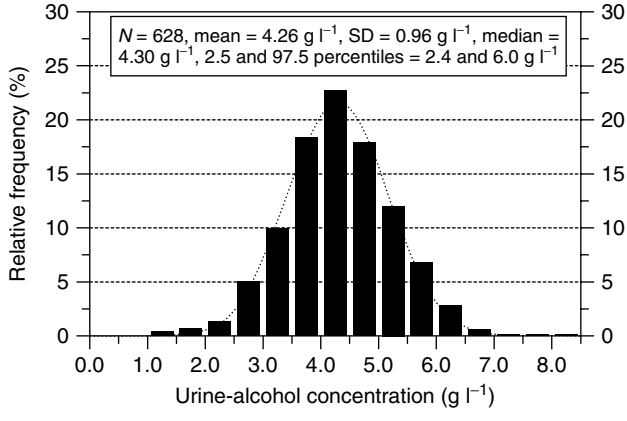


Figure 8 Frequency distribution of the concentrations of ethanol in bladder urine in deaths attributed to acute alcohol poisoning

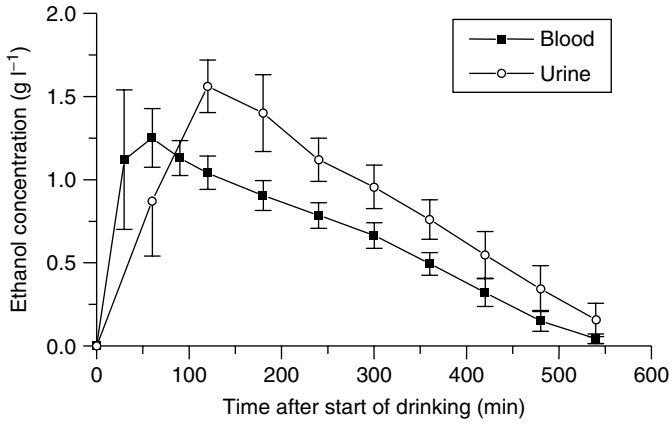


Figure 9 Concentration–time profiles of ethanol in successive urinary voids and in blood samples from healthy volunteers after they drank 0.85 g ethanol per kg body weight on an empty stomach

Table 12 The amounts of ethanol absorbed and distributed in all body fluids and tissues and the 95% range considered necessary to cause death by acute alcohol poisoning. Results are shown for both sexes as a function of body weight. In the calculation, a BAC of 5 g l^{-1} was considered sufficient to cause death by paralysis of respiratory centers in the brain

Body weight (kg)	Males[a] (g ethanol) Mean (95% range)	Females[b] (g ethanol) Mean (95% range)
10	35 (28–41)	30 (24–36)
20	70 (56–84)	60 (48–72)
30	105 (84–126)	90 (72–108)
40	140 (112–168)	120 (96–144)
50	175 (134–216)	150 (120–180)
60	210 (168–252)	180 (144–216)
70	245 (196–294)[c]	210 (168–252)
80	280 (224–336)	240 (192–288)[c]
90	315 (252–378)	270 (216–324)
100	350 (280–420)	300 (240–360)

[a] Average r-factor for males = 0.71 kg^{-1} with 95% range ±20%
[b] Average r-factor for females = 0.61 kg^{-1} with 95% range ±20%
[c] A bottle of spirits 40% v/v contains about 240 g ethanol

for women 3.0 g kg^{-1} (range 2.4–3.6 g kg^{-1}). For a man with a body weight of 70 kg, he would need to consume 240 g of ethanol or the amount contained in a whole bottle (750 ml) of liquor (whisky or vodka) to die from acute alcohol poisoning.

Heavily intoxicated people are overrepresented among patients admitted to hospital for treatment of head trauma and brain hemorrhage, which is a common cause of death in alcoholics [3]. Drinking on an empty stomach and neglecting to eat properly is dangerous because of ethanol-induced disturbances in carbohydrate metabolism and risk of developing life-threatening hypoglycemia (low blood sugar). The combined effect of alcohol and other drugs, such as barbiturates or benzodiazepines, has resulted in many adverse drug-related fatalities [7]. Exposure to the cold, such as if a person sleeps outdoors after a night of heavy drinking, has been responsible for many hypothermia-related fatalities because ethanol lowers core body temperature in addition to the low ambient temperature.

Inhalation of vomit (aspiration) and asphyxia has killed many drunken people, especially young inexperienced drinkers, who have a nonfunctional gag reflex inactivated owing to gross intoxication and lack of consciousness [29]. A semicomatose person should not be allowed to "sleep it off" unattended and instead should be placed in a semiprone position and inspected regularly until they regain consciousness. A partial obstruction of the trachea during the time a person is in a comatose state after heavy drinking is yet another circumstance that could lead to death by asphyxiation [27–29].

Drinking alcohol too quickly to reach a BAC of 1.5 g l^{-1} or more often causes nausea and vomiting, especially in novice drinkers. This vomit reflex has saved the lives of many young people by removing unabsorbed alcohol from the stomach contents. However, if a person vomits when already in a state

of gross intoxication with depressed gag reflex, there is a grave risk of vomit entering the airways and death being caused by asphyxiation [29]. Many teenagers with little or no previous experience with alcohol consumption have lost their lives in this way. The BAC determined in blood taken at autopsy is often considerably lower than the level that caused incapacitation at some earlier time because of decreases through metabolism at a rate of $0.15\,g\,l^{-1}$ per h until the time of death.

Concluding Remarks

Overconsumption of alcoholic beverages and drunkenness play a major role in many types of accidents, trauma deaths, suicides, muggings, and other crimes of violence, as evidenced by death certificates and reports from hospital accident and emergency departments worldwide [3, 45]. Moreover, heavy drinking and drunkenness are often the underlying factor in domestic violence, sexual assault, and road-traffic crashes. Accordingly, requests to determine the concentration of ethanol in blood and other body fluids and to interpret the results in relation to degree of impairment and ability to form intent are everyday services obtainable from forensic science and toxicology laboratories [26, 46].

Because a diagnosis of alcohol influence has deep-rooted social, medical, and legal ramifications, great care is needed when the analytical results are interpreted and conclusions reached for use in civil and/or criminal proceedings. Whether a person's BAC was above or below a threshold limit, such as the legal limit for driving, makes the difference between punishment (fines, loss of driving permit, or incarceration) and acquittal [47].

The qualitative and quantitative determination of ethanol in body fluids is a relatively simple analytical procedure and the use of gas liquid chromatography provides accurate, precise, and specific results [48]. However, correctly interpreting these results requires knowledge about the pharmacokinetics and pharmacodynamics of ethanol as well as intersubject and intrasubject variation, drug-alcohol interactions, and the development of tolerance [11].

When it comes to interpreting the results of ethanol determinations in autopsy specimens, a number of other considerations are necessary, owing to the risk of various postmortem artifacts [27–29]. These include the condition and type of specimen, the stability of ethanol in the body after death, diffusion from the gut to other tissue after death, and the possibility of microbial synthesis occurring if the body has undergone decomposition [26–31]. Police investigations, insurance claims, and financial compensation to surviving family members might be jeopardized or invalidated if the deceased was judged to be drunk and incapable at the time of death. An accurate and precise measurement of the BAC is indispensable to reach a conclusion about a person's state of inebriation at the time of death. (*see also* **Drug-Impaired Driving** and **Behavioral Toxicology**).

References

[1] Klatsky, A.L. (2003). Drink to your health, *Scientific American* **288**, 74–81.

[2] Room, R., Babor, T. & Rehm, J. (2005). Alcohol and public health, *Lancet* **365**, 519–530.

[3] Cherpitel, C.J. (2007). Alcohol and injuries: a review of international emergency room studies since 1995, *Drug and Alcohol Review* **26**, 201–214.

[4] Garriott, J.C. (ed) (2008). *Medicolegal Aspects of Alcohol*, 5th Edition, Lawyers and Judges Publishing Company, Tuscon, pp. 1–534.

[5] Anderson, R.A. (2005). Back-tracking calculations, in *Encyclopedia of Forensic and Legal Medicine*, J. Payne-James, R.W. Byard, T.S. Corey & C. Henderson, eds, Elsevier, Oxford, pp. 261–270.

[6] Meyer J.S. & Quenzer L.F. (2005). Chapter 9 Alcohol, in *Psychopharmacology; Drugs, The Brain and Behavior*, Sinauer Associates, Sunderland, pp. 215–243.

[7] Jones, A.W. (2003). Drug alcohol interactions, in *Handbook of Drug Interactions: A Clinical and Forensic Guide*, C. Miozayani & L. Raymon, eds, Humana Press, Totowa, pp. 395–462.

[8] Logan, B.K. & Jones, A.W. (2000). Endogenous ethanol "autobrewery syndrome" as a drunk driving defense challenge, *Medicine Science and the Law* **40**, 206–215.

[9] Brick, J. (2006). Standardization of alcohol calculations in research, *Alcoholism Clinical and Experimental Research* **30**, 1276–1287.

[10] Jones, A.W. (2008). Biochemical and physiological research on the disposition and fate of ethanol in the body, in *Medicolegal Aspects of Alcohol*, 5th Edition, J.C. Garriott, ed, Lawyers and Judges Publishing Company, Tuscon, pp. 47–155.

[11] Jones, A.W. & Pounder, D.J. (2007). Update on clinical and forensic analysis of alcohol, in *Drug Abuse Handbook*, 2nd Edition, S.B. Karch, ed, CRC Press, Boca Raton, pp. 333–376.

[12] Widmark, E.M.P. (1981). *Principles and Applications of Medicolegal Alcohol Determination*, Biomedical Publications, Foster City, pp. 1–163.

[13] Haslam, D.W. & Jones, W.P.T. (2005). Obesity, *Lancet* **366**, 1197–1209.
[14] Jones, A.W. (2007). Body mass index and blood-alcohol calculations, *Journal of Analytical Toxicology* **31**, 177–178.
[15] Lieber, C.S. (1982). *Medical Disorder of Alcoholism*, W.B. Saunders publishing Co, Philadelphia.
[16] Kalant, H. & Khanna, J.M. (2007). The alcohols, in *Principles of Medical Pharmacology*, 7th Edition, H. Kalant, D.M. Grant & J. Mitchell, eds, Elsevier, Toronto, pp. 275–288.
[17] Lieber, C.S. (2005). Metabolism of alcohol, *Clinics in Liver Disease* **9**, 1–35.
[18] Lieber, C.S. (2004). The discovery of the microsomal ethanol oxidizing system and its physiologic and pathologic roles, *Drug Metabolism Reviews* **36**, 511–529.
[19] Andreasson, R. & Jones, A.W. (1996). The life and work of Erik MP Widmark, *American Journal of Forensic Medicine and Pathology* **17**, 117–190.
[20] Wagner, J.G. (1991). *Pharmacokinetics for The Pharmaceutical Scientist*, Technomic Publishing Company, Basel.
[21] Rowland, M. & Tozer, T.N. (1995). *Clinical Pharmacokinetics; Concepts and Applications*, 3rd Edition, Williams & Wilkins, Philadelphia.
[22] Norberg, A., Jones, A.W., Hahn, R. & Gabrielsson, J. (2003). Role of variability in explaining ethanol kinetics – research and forensic applications, *Clinical Pharmacokinetics* **42**, 1–31.
[23] Watson, P.E., Watson, I.D. & Batt, R.D. (1981). Prediction of blood alcohol concentration – updating the Widmark equation, *Journal of Studies on Alcohol* **42**, 547–556.
[24] Jones, A.W. & Andersson, L. (1998). Influence of age, gender and blood alcohol concentration on the disappearance rate of alcohol from blood in drinking drivers, *Journal of Forensic Sciences* **41**, 922–928.
[25] Ferner, R.E. (1996). *Forensic Pharmacology – Medicines, Mayhem and Malpractice*, Oxford University Press, Oxford.
[26] Drummer, O.A. (2001). *The Forensic Pharmacology of Drugs of Abuse*, Arnold, London.
[27] Jones, A.W. (2000). Alcohol; post-mortem, in *Encyclopedia of Forensic Sciences*, J.A. Siegel, P.J. Saukko. & G.C. Knupfer, eds, Academic Press, London, pp. 112–126.
[28] Kugelberg, F.C. & Jones, A.W. (2007). Interpreting results of ethanol analysis in postmortem specimens: a review of the literature, *Forensic Science International* **165**, 10–29.
[29] Pounder, D.J. & Jones, A.W. (2007). Post-mortem alcohol aspects of interpretation, in *Drug Abuse Handbook*, 2nd Edition, S.B. Karch, ed, CRC Press, Boca Raton, pp. 376–401.
[30] Corry, J.E.L. (1978). Possible sources of ethanol ante - and post-mortem: its relationship to the biochemistry and microbiology of decomposition, *Journal of Applied Bacteriology* **44**, 1–56.
[31] Forrest, A.R.W. (1993). Obtaining samples at post-mortem examination for toxicological and biochemical analyses, *Journal of Clinical Pathology* **46**, 292–296.
[32] O'Neal, C.L. & Poklis, A. (1996). Postmortem production of ethanol and factors that influence interpretation: a critical review, *American Journal of Forensic Medicine and Pathology* **17**, 8–20.
[33] Jones, A.W. (2006). Urine as a biological specimen for forensic analysis of alcohol and variability in the urine-to-blood relationship, *Toxicological Reviews* **25**, 15–35.
[34] Jones, A.W. & Holmgren, P. (2001). Uncertainty in estimating blood alcohol concentration by analysis of vitreous humor, *Journal of Clinical Pathology* **54**, 699–702.
[35] Helander, A. & Jones, A.W. (2007). Recent advances in biochemical tests for acute and chronic alcohol consumption, in *Drug Abuse Handbook*, 2nd Edition, S.B. Karch, ed, CRC Press, Boca Raton, pp. 401–427.
[36] Jones, A.W. (2008). Biomarkers of acute and chronic alcohol ingestion, in *Medicolegal Aspects of Alcohol*, 5th Edition, J.C. Garriott, ed, Lawyers and Judges Publishing Company, Tuscon, pp. 157–203.
[37] Rainio, J., Giorgio, F.D., Bortolotti, F. & Tagliaro, F. (2008). Objective post-mortem diagnosis of chronic alcohol abuse – a review of studies on new markers, *Legal Medicine* **10**, 229–235.
[38] Kalant, H., LeBlanc, A.E. & Gibbins, R.J. (1971). Tolerance to and dependence on some non-opiate psychotropic drugs, *Pharmacological Reviews* **23**, 135–191.
[39] Rigter, H. & Crabbe, J.C. (eds) (1980). *Alcohol Tolerance and Dependence*, Elsevier Biomedical Press, Amsterdam.
[40] Jones, A.W. (2005). Alcohol, acute and chronic use, post-mortem findings, in *Encyclopedia of Forensic and Legal Medicine*, J. Payne-James, R.W. Byard, T.S. Corey & C. Henderson, eds, Elsevier, Oxford, pp. 39–58.
[41] Kalant, H. (1998). Research on tolerance; what can we learn from history, *Alcoholism Clinical and Experimental Research* **22**, 67–75.
[42] Jaffe, J.H. (ed) (1995). *Encyclopedia of Drugs and Alcohol*, Macmillan Library Reference, Simon & Schuster and Prentice Hall International, New York.
[43] Jones, A.W. & Holmgren, P. (2003). Comparison of blood-alcohol concentrations in deaths attributed to acute alcohol poisoning and chronic alcoholism, *Journal of Forensic Sciences* **48**, 874–879.
[44] Jones, A.W. & Holmgren, P. (2003). Urine/blood ratios of ethanol in deaths attributed to acute alcohol poisoning and chronic alcoholism, *Forensic Science International* **135**, 206–212.
[45] Gibbons, B. (1992). Alcohol – the legal drug, *National Geographic* **181**, 3–35.
[46] Levine, B.S. (ed) (1999). *Principles of Forensic Toxicology*, 2nd Edition, American Association of Clinical Chemistry, Washington, DC, pp. 1–394.
[47] Walls, H.J. & Brownlie, A.R. (1985). *Drink, Drugs and Driving*, 2nd Edition, Sweet and Maxwell, London.

[48] Jones, A.W. (2000). Medico-legal alcohol determination – Blood or breath alcohol concentration? *Forensic Science Review* **12**, 23–48.

Further Reading

Nutt, D., King, L.A., Saulsbury, W. & Blackmore, C. (2007). Development of a rational scale to assess the harm of drugs of potential misuse, *Lancet* **369**, 1047–1053.

ALAN W. JONES

Alcohol: Analysis

Introduction

In the field of forensic science and legal medicine, investigators are mainly concerned with just one type of alcohol, namely ethyl alcohol or ethanol, which is the pharmacologically active constituent in all alcoholic beverages. However, the word alcohol is a generic term used by chemists to denote a family of organic compounds, the simplest member of which is methanol or methyl alcohol (CH_3OH). Ethanol (CH_3CH_2OH) is the second member in this homologous series of aliphatic alcohols and is known as a *primary alcohol*, because the hydroxyl bearing carbon atom is attached to one alkyl group. Next comes *n*-propyl alcohol ($CH_3CH_2CH_2OH$) along with its structural isomer isopropyl alcohol $(CH_3)_2CHOH$, the latter is designated as a *secondary alcohol* because the hydroxyl carbon is bonded to two methyl groups. Structural formulae of the simplest alcohols are given in Table 1.

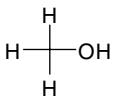
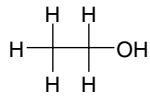
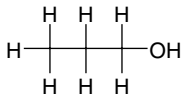
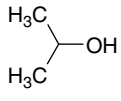
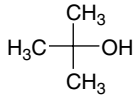
The common feature shared by the entire family of alcohols is that they contain one or more hydroxyl (–OH) functional groups. This property determines solubility in water and body fluids, ability to form hydrogen bonds with other biomolecules as well as many chemical and biochemical reactions, such as oxidation, reduction, and conjugation.

Ethanol is produced on a huge scale by the fermentation of carbohydrates (beverage alcohol) or by the catalytic addition of water to ethylene (nonbeverage usage). Ethanol is considered as a clean fuel and is widely used in industry as a solvent and also occurs in many household products, pharmaceuticals, cosmetics, toiletries, perfumes, etc. Even certain foodstuffs, such as fresh fruits and fruit juices, might contain trace quantities of ethanol or ethyl esters that can undergo hydrolysis in the stomach to form ethanol.

Alcoholic beverages are made by fermentation of the sugars contained in a wide variety of naturally occurring raw materials such as potato mashes, barley corn, fruit juices, beet, cane sugar, and molasses. The enzymes in yeast convert one molecule of glucose into two molecules of ethanol and two molecules of carbon dioxide. Depending on the fermentation conditions, such as the source of the yeast and the nature of the starting material, the endproduct might contain up to 15% v/v ethanol. Beers contain 3–5% v/v ethanol and wines generally contain 9–15% v/v. A drink containing higher concentrations of ethanol cannot be produced by fermentation alone, because the yeast enzymes become inactivated. Distillation or fortification of the primary fermentation product is necessary if more concentrated alcoholic drinks (distilled spirits), such as whisky, vodka, gin, and brandy (35–55%v/v) are required.

Alcohol is a dependence-producing drug and excessive drinking and drunkenness lead to many negative social, medical, psychological, and economic consequences for the individual and society [1]. Overconsumption of alcohol, binge drinking, and drunkenness constitute major public health issues in today's society. Intoxicated people are overrepresented at hospital emergency rooms to receive

Table 1 Chemical structures of the simplest monohydroxy aliphatic alcohols

				
Methanol CH_3OH	Ethanol CH_3CH_2OH	*n*-propanol $CH_3CH_2CH_2OH$	Isopropanol $(CH_3)_2CHOH$	*t*-butanol $(CH_3)_3COH$

treatment for injuries and in deaths caused by drowning, suicide, or road-traffic fatalities. Forensic science and toxicology laboratories worldwide have a long standing interest in the analysis and the interpretation of a person's blood-alcohol concentration (BAC), which requires considerable knowledge and expertise about methodological, physiological, and pharmacological aspects of ethanol [2, 3].

This article deals with the principles and practice of ethanol determination in biological specimens. The major emphasis is given to the analysis of ethanol in blood and breath, which are the specimens used in forensic casework as evidence for prosecuting drunken drivers [4]. However, the procedures described for handling blood samples can be applied to many other biological liquids such as urine, plasma, serum, and saliva [5, 6]. Indeed, any biofluid or tissue that contains water can serve as a specimen for the analysis of ethanol.

Punishable Alcohol Concentration Limits for Driving

Alcohol-related motor vehicle crashes are a major cause of premature death and disability. Many case-controlled studies have established that the risk of a crash appreciably increases as the driver's BAC increases above 0.5 $g\,l^{-1}$ (50 mg/100 ml or 0.05 g/100 ml). Impaired drivers are overrepresented in crash statistics and 30–40% of those killed on the roads test positive for alcohol and most victims have BACs above the legal limit for driving. To improve traffic safety and deter drunken driving, governments have enacted punishable limits of alcohol concentration in specimens of blood, breath, or urine above which it is an offense to drive a motor vehicle on the public roads and highways [7, 8]. However, these statutory alcohol limits differ among different countries, which has a lot to do with alcohol control policy and local politics rather than traffic safety research (Table 2).

In many publications dealing with forensic aspects of alcohol, confusion arises owing to the different concentration units used to report the results. Examples of the various blood- and breath-concentration units and the countries where they apply are given in Table 3. Also shown is the concentration unit used in clinical and laboratory medicine, namely millimoles per liter (21.7 $mmol\,l^{-1} = 1.0\,g\,l^{-1}$). Note also that hospital laboratories commonly determine ethanol in specimens of plasma or serum and not in whole blood. The differences in water content among the different biofluids have important consequences if and when results from hospital clinical laboratories are later used as evidence in legal proceedings, such as impaired driving trials [2].

Ethanol distributes into the water compartment of the body and unlike many other drugs, it does not bind to plasma proteins. The water content of whole blood is ~80% w/w on average compared with ~93% w/w for serum or plasma. Because the

Table 2 Threshold concentration limits of alcohol (statutory limits) in blood and breath for operating a motor vehicle in various countries and the blood/breath ratio assumed when setting the corresponding breath-alcohol concentration limit

Country/nation[a]	Blood-alcohol concentration	Breath-alcohol concentration	Assumed blood-to-breath ratio
Most European nations	0.50 $mg\,ml^{-1}$ ($g\,l^{-1}$)	0.25 $mg\,l^{-1}$	2000 : 1
The Netherlands	0.50 $mg\,ml^{-1}$ ($g\,l^{-1}$)	220 $\mu g\,l^{-1}$	2300 : 1
Norway and Sweden[b]	0.20 $mg\,g^{-1}$ ($g\,kg^{-1}$)	0.10 $mg\,l^{-1}$	2100 : 1
Finland	0.50 $mg\,g^{-1}$ ($g\,kg^{-1}$)	0.22 $mg\,l^{-1}$	2400 : 1
The United States	0.08 g/100 ml	0.08 g/210 l	2100 : 1
The United Kingdom and Ireland[c]	80 mg/100 ml	35 µg/100 ml	2300 : 1
Canada	0.08 g/100 ml	0.08 g/210 l	2100 : 1
Australia	0.05 g/100 ml	0.05 g/210 l	2100 : 1
New Zealand	80 mg/100 ml	400 $\mu g\,l^{-1}$	2300 : 1

[a] In several countries listed above a lower limit of alcohol concentration operates for novice or provisional drivers (e.g., 0.02 g% in the United States), for drivers under 21 years and 0.04 g% for operators of commercial vehicles

[b] Because a BAC of 0.20 $mg\,g^{-1}$ is equivalent to 0.21 $mg\,ml^{-1}$, the actual blood/breath ratio operating in Norway and Sweden is close to 2100 : 1

[c] If urine is the specimen donated the threshold concentration of alcohol is 107 mg/100 ml

Table 3 Concentration units used to report blood-alcohol concentration (BAC) and breath-alcohol concentration (BrAC) for clinical and forensic purposes in different countries

Concentration unit for blood alcohol	Countries where used	Concentration unit for breath-alcohol	Countries where used
$mg\,g^{-1}$ ($g\,kg^{-1}$)	Sweden, Denmark, Norway, Finland, Germany	$mg\,l^{-1}$	Sweden, Denmark, Norway, Finland, Germany, Spain, other EU countries
$mg\,ml^{-1}$ ($g\,l^{-1}$)	France, Holland, Spain, Belgium	$\mu g\,l^{-1}$	Holland
mg/100 ml ($mg\,dl^{-1}$)	The United Kingdom, Ireland, Canada, New Zealand	µg/100 ml	The United Kingdom, Ireland, New Zealand
g/100 ml (g%)	The United States, Australia	g/210 l	The United States
mmol/l[a]	Hospital clinical chemistry laboratories in most countries	ppm[b]	Environmental health applications

[a] Molecular weight of ethanol is 46.07; thus, a concentration of 21.7 mmol l^{-1} is equivalent to 1.0 g l^{-1} (100 mg/100 ml or 0.1 g/100 ml)
[b] ppm stands for parts per million where 200 ppm = 0.365 mg ethanol per liter of breath at 34 °C

distribution of alcohol between plasma and whole blood follows the distribution of water, the mean concentration ratio (plasma/blood) for ethanol is 1.16 : 1 and this can be used as a conversion factor [2]. The actual plasma/blood distribution ratio depends on the factors influencing the water content of the specimens analyzed such as hematocrit value and any medical conditions that influence the volume of red cells (e.g., anemia). Abnormal blood lipids tend to influence the concentration of alcohol in the blood sample. The gender-related differences in hematocrit (men 42–50 ml/100 ml and women 37–47 ml/100 ml) also have a small effect on the plasma/blood distribution ratio of ethanol.

If nothing is known about the individual concerned and a request is made to convert the plasma alcohol concentration (PAC) or the serum alcohol concentration (SAC) into the blood alcohol concentration (BAC) the factor used should be chosen to the person's advantage if the results are intended for use as evidence in a criminal prosecution.

$$BAC = (PAC \text{ or } SAC)/1.16 \quad \text{(gives the best average estimate)} \quad (1)$$

$$BAC = (PAC \text{ or } SAC)/1.20 \quad \text{(gives a lower limit for most people)} \quad (2)$$

Analytical Methods for the Determination of Ethanol

The first analytical methods for the identification and determination of ethanol in human body fluids and tissues were described approximately 150 years ago [9]. Because of the strong association between a person's BAC and the impairment of performance of skilled tasks, such as driving, it has become increasingly necessary to interpret the results of analysis in a legal context.

The BAC encountered in most forensic casework ranges from 0 to 5 $g\,l^{-1}$ (500 mg/100 ml), and a concentration below 0.1 $g\,l^{-1}$ is usually reported as negative and no further action or interpretation is necessary. Although the analytical methods used in many forensic toxicology laboratories are capable of measuring much lower concentrations of ethanol below 0.1 $g\,l^{-1}$, this level is appropriate as a practical cutoff concentration for reporting negative results.

The limit of quantitation (LOQ) of an analytical method is an important concept in clinical, forensic, and laboratory medicine and this is defined as the lowest concentration of analyte that gives acceptable precision and accuracy under the stated operating conditions of the method. In practice, LOQ is approximately 3.3 times the limit of detection (LOD), which is defined as the lowest concentration of analyte that can be detected. With instrumental methods of analysis, LOD corresponds to a signal to background noise ratio of about 3 : 1.

During the development of new analytical methods, a good starting point is to consider the physical, chemical, and structural properties of the substance of interest – the analyte. Ethanol is a low-molecular weight volatile substance (boiling point 79 °C) with a specific gravity of 0.789 and mixes with water in all proportions. The hydroxyl group is chemically

reactive and readily undergoes oxidation to an aldehyde group (–CHO) or a carboxylic acid group (–COOH), depending on the reaction conditions such as time and temperature.

Methods Based on Chemical Oxidation

The earliest methods suitable for the analysis of alcohol in blood and urine were based on chemical oxidation of the hydroxyl group (–OH), to yield the corresponding aldehyde (–CHO) and then carboxylic acid (–COOH). Various oxidizing agents, such as acidified potassium dichromate or potassium permanganate, were employed. Prior to adding the required chemicals to start the reaction, it was necessary to remove the ethanol from the biological matrix by distillation, desiccation, or aeration or also by the precipitation of blood proteins, e.g., with perchloric acid [10]. The oxidizing reagent was added to the aqueous distillate or filtrate containing ethanol and when necessary the mixture was heated (e.g., to 50 °C) to speed up the completion of the reaction. The amount of ethanol in the original sample could then be determined by titrimetric analysis or by colorimetry by comparison with known strength standard solutions of ethanol.

The basic equation for the oxidation of ethanol with a mixture of potassium dichromate and sulfuric acid is shown below.

$$2K_2Cr_2O_7 + 8H_2SO_4 + 3C_2H_5OH \longrightarrow$$

(yellow color)

$$2Cr_2(SO_4)_3 + 2K_2SO_4 + 11H_2O + 3CH_3COOH$$

(green color) (3)

Methods based on chemical oxidation were rather time consuming and required considerable training to ensure reliable results and not many samples could be completed each working day. Moreover, the oxidation reaction was not specific and other alcohols or acetone if present in the blood samples could not be distinguished from ethanol. Nevertheless, only wet-chemistry oxidation was used between 1900s and 1950s for the quantitative analysis of ethanol in body fluids for legal purposes [8, 9]. When specimens were taken during autopsy, various preliminary chemical tests were necessary to resolve whether other volatile substances might be present that could compromise the reliability of the analytical results.

Methods Based on Enzymatic Oxidation

Milder oxidation conditions and improved selectivity of analysis were achieved when enzymatic procedures were developed in the early 1950s. This coincided with the isolation and purification of the liver enzyme responsible for the oxidation of ethanol *in vivo*, namely alcohol dehydrogenase (ADH). Mammalian ADH was not so suitable for forensic analysis because other aliphatic alcohols (e.g., methanol and isopropanol) were also oxidized and under certain circumstances, these substances might be present in the blood samples. Accordingly, the enzyme derived from yeast was developed commercially and widely used in the enzymatic (ADH) determination of ethanol. Methanol reacted very slowly or not at all under optimum reaction conditions and did not interfere with the determination of ethanol [11].

A major advantage of enzymatic methods over chemical methods was that the procedures could be more easily automated. Indeed, this was becoming increasingly necessary considering the huge increase in forensic casework as drunken driving escalated in the 1950s and 1960s.

The oxidation of ethanol to acetaldehyde by ADH is a reversible reaction. Under *in vitro* conditions, the reaction is facilitated by adjusting pH to 8.8 and adding a reagent (semicarbazide hydrochloride) to trap the acetaldehyde produced. The semicarbazide forms a semicarbazone by reacting with the acetaldehyde to drive the reaction (see below) to the right. During the redox reaction, the coenzyme nicotinamide adenine dinucleotide (NAD^+) is converted to the reduced form NADH and its formation is monitored by UV spectrometry at a wavelength of 340 nm for quantitative analysis.

$$CH_3CH_2OH + NAD^+ \overset{ADH}{\longleftrightarrow} CH_3CHO + NADH + H^+$$

CH_3CHO + semicarbazide gives a semicarbazone driving the reaction to completion (4)

Enzymatic methods gradually replaced the chemical oxidation procedures and became widely used until the emergence of physicochemical methods in the early 1960s. The upsurge of immunoassay methods designed for analysis of drugs of abuse in urine has led to a renaissance for the NAD^+-NADH reaction. Highly automated and sophisticated

instruments are available for the analysis of abused drugs including ethanol by immunoassay. Indeed, some laboratories make a rapid preliminary screening analysis of all specimens by enzymatic methods so as to distinguish negative and positive specimens. Those specimens that tested positive for ethanol are then reanalyzed by more specific methods, such as gas chromatography [12].

Nowadays, the method of choice for blood-alcohol analysis in clinical, forensic, and research applications is gas-liquid chromatography (GC). For the determination of alcohol in samples of breath, dedicated instruments are available that utilize infrared spectrometry or electrochemical oxidation as the analytical principles (see later in this article).

Methods Based on Gas Chromatography

The development and implementation of gas liquid partition chromatography in the early 1960s and onwards revolutionized the practice of analytical chemistry at forensic science and toxicology laboratories. The instrument bay in a modern toxicology laboratory is dominated by advanced separation methods involving either gas- or liquid-chromatographic procedures with various detector systems depending on the particular analyte or drug of interest [13].

Basic Principles of Gas Chromatography

The principle components of GC are an inert carrier gas (nitrogen or helium), which represents the moving or mobile phase. This is made to flow through a long thin coiled tube (the chromatographic column) containing the liquid stationary phase. The column is kept in an oven held at a precisely controlled temperature, which is a key parameter for the effective separation of the components of a mixture. The liquid stationary phase is coated onto an inert support material to provide a large surface area and to facilitate an intimate contact and mixing of components between the moving gaseous phase and the stationary phase.

The components in the sample are vaporized before mixing with the gaseous mobile phase, which transports the substances to be separated through the column. During passage through the column, the components of the mixture partition between the moving gas phase and the liquid stationary phase. Different compounds partition differently depending on their physical and chemical properties and are retained for a shorter or longer time inside the column so that a partial or complete separation occurs.

The effluent from the column passes to a detector, which produces a signal proportional to the amount of substance it detects. Different detectors are available for different applications depending on the chemical structure and elemental composition of the target analytes. For compounds with C–C or C–H bonds in the molecules, a flame ionization detector (FID) is widely used. The carrier gas and the substances eluting from the column enter an air-hydrogen burning flame to produce ions and electrons that pass between anode and cathode. The current flowing at a particular voltage between the positive and negative poles is amplified and the signal is monitored as a trace on a recorder known as *the chromatogram*.

The choice of stationary phase is important and when water-soluble alcohols are analyzed hydrophilic phases, such as polyethylene glycols of different average molecular weights, are appropriate. The traditional packed columns, which worked very well for decades and still do, have gradually been replaced by capillary or wide bore columns, which are more expensive and less robust compared with packed columns.

The main advantage of gas chromatographic analysis over other methods is that the target analyte (ethanol) is separated from impurities or potential interfering compounds (e.g., acetone) that might be normally present in forensic blood samples. Moreover, both qualitative and quantitative analysis is possible in the same analytical run. The time that elapses after injecting the sample into the column to the appearance of the apex of the peak on the chart recorder or electronic integrator is called the *retention time* (*RT*) and is useful for qualitative analysis. The RT of the unknown substance can be compared with the RT of authentic known standards purchased from a reliable source. The integrated area under the detector response (peak) on the chromatogram is the parameter used for quantitative analysis and is directly proportional to the amount of substance in the sample.

Analytical Details

In forensic casework, the specimens of blood or other body fluids should always be analyzed in duplicate and whenever possible, these determinations

should be made by different technicians working independently with different sets of equipments [14]. Importantly, the chromatographic conditions, such as the nature of the stationary phase, should be such that different RTs are obtained for potential interfering substances. The risk that two closely related compounds will have the same RT when analyzed on two different stationary phases is very remote. Alternatively, an increased selectivity can be achieved by the use of two independent analytical methods, such as GC and an enzymatic oxidation procedure, to make the duplicate determinations.

Approximately 1–5 µl of the diluted blood specimen is injected directly onto the chromatographic column or into a heated injection port attached to the column where vaporization occurs. Any volatile substances in the sample mix with the mobile phase (carrier gas) and pass through the column, where separation occurs. The effluent from the column is then directed to the detector for quantitative analysis. Blood and other body fluids contain nonvolatile constituents (proteins, fats, etc.), which tend to clog the syringe and injection port of the GC and the column packing material deteriorates after many samples are analyzed. To avoid this problem, investigators developed the headspace sampling technique and this entails the analysis of the vapor phase above the diluted blood sample. The diluted blood specimen is held in a glass vial, which is kept airtight with a stopper and crimped on aluminum cap and is heated to 50 or 60 °C to achieve a liquid-vapor equilibration. The use of the headspace method meant that the GC column was not contaminated with nonvolatile substances and several thousands of blood samples could be analyzed without damaging the column and the overall analytical performance. A scheme that illustrates the basic features of headspace GC applied the determination of blood ethanol concentration is shown in Figure 1.

A short analysis time and adequate resolution of components in the sample are the key elements in any chromatographic analysis. Ethanol and other low molecular volatiles can be analyzed under isothermal temperature control of the chromatographic oven, which saves much time compared with a temperature program (heating and cooling) being used. For routine applications, the analysis takes about 2–3 min after the injection was made into the GC. By raising the oven temperature or operating with a higher carrier gas flow rate, the RTs can be shortened but often at the expense of an incomplete separation of the components in a mixture.

Examples of the RTs (min) for GC analysis of nine different volatile agents analyzed on four different

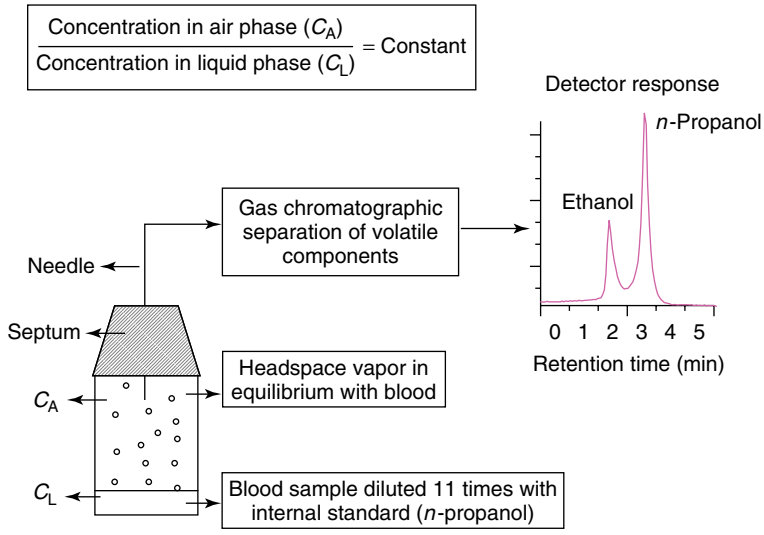


Figure 1 Diagram showing the main features of static headspace gas chromatography where an equilibration is achieved between volatiles in the diluted blood specimen (*n*-propanol internal standard) and the air phase above the blood. After equilibration is reached (20–30 min), a portion of the air-phase is removed and directed into the gas chromatograph where a separation of the components and a quantitative analysis is made with a flame-ionization detector

Table 4 Retention times (RT) of ethanol and other low molecular volatile substances analyzed by headspace gas chromatography on four different stationary phases commonly used in forensic toxicology laboratories. RT's relative to *n*-propanol as internal standard are shown in brackets

Volatile substance	Retention time (min) Carbopak C[(a)]	Retention time (min) Carbopak B[(b)]	Retention time (min) Rtx-BAC1[(c)]	Retention time (min) Rtx-BAC2[(d)]
Acetaldehyde	0.56 (0.38)	0.53 (0.29)	1.19 (0.53)	0.82 (0.34)
Acetone	1.00 (0.68)	0.86 (0.46)	2.05 (0.91)	1.36 (0.57)
n-Butanol	4.68 (3.16)	4.11 (2.22)	4.63 (2.06)	5.28 (2.20)
Ethanol	0.72 (0.49)	0.98 (0.53)	1.32 (0.59)	1.27 (0.53)
Methanol	0.49 (0.33)	0.67 (0.36)	1.06 (0.47)	0.96 (0.40)
Methyl ethyl ketone	2.45 (1.66)	1.49 (0.81)	3.08 (1.37)	2.49 (1.04)
Isopropanol	1.16 (0.78)	1.31 (0.71)	1.66 (0.74)	1.48 (0.62)
n-Propanol	1.05 (1.00)	1.85 (1.00)	2.25 (1.00)	2.40 (1.00)
t-Butanol	1.90 (1.28)	1.68 (0.91)	1.98 (0.88)	1.65 (0.69)

[(a)] Packed glass column (2 m × 0.5 mm i.d.) with 0.2% Carbowax (polyethylene glycol) 1500
[(b)] Packed glass column (2 m × 0.5 mm i.d.) with 5% Carbowax (polyetheylene glycol) 20 M
[(c)] Capillary column, 30 m × 0.53 mm i.d.
[(d)] Capillary column, 30 m × 0.53 mm i.d.

columns and stationary phases (two packed and two capillary) under isothermal oven temperature conditions are shown in Table 4. Also listed are the RTs of the *n*-propanol peak – the internal standard. The overall time of analyzing these nine compounds was only 4–5 min and the longest RT was for *n*-butanol.

Recommended Procedure for Routine Purposes

The basic steps in a well tested and recommended procedure for the quantitative determination of ethanol in blood or urine for legal purposes by GC are outlined below [15].

1. The tube with the material to be analyzed should be gently inverted for some time to ensure homogeneity of the specimen. This step is important when ethanol is determined in blood samples because the red cells and the plasma fraction tend to separate out on standing.
2. An aliquot (0.1 ml) of the specimen (whole blood, plasma, serum, or urine) is accurately removed from the tube and immediately diluted with an aqueous solution of an internal standard (usually another type of alcohol). This maneuver is best done by the use of specially constructed diluter-dispenser equipment although micropipettes can be also be used. The dilution of the biofluid with aqueous internal standard should be at least 1 : 5 (6-fold) or 1 : 10 (11-fold). Adequate dilution of the specimen eliminates matrix effects, which is an important consideration when headspace GC analysis is used and determines the choice of the calibration standards. With 6- to 11-fold dilution of blood meant that aqueous ethanol standards can be used to construct a calibrate plot for use in quantitative analysis. Suitable internal standards are aqueous solutions of *n*-propanol (specimens from living subjects) or *t*-butanol (in autopsy work) at concentrations of 0.05–0.10 g l^{-1}. It is important that the substance used as the internal standard is not likely to occur or be produced in the biofluid analyzed.
3. For headspace analysis, the diluted blood specimens are ejected into a glass vial, which is made airtight with a Teflon-coated stopper and crimped on aluminum cap. For analysis of liquid samples the diluted specimens are ejected into a clean dry glass vial and an aliquot (1–5 µl) is injected into the GC using a microsyringe.
4. The peak area ratio of the responses for ethanol and the internal standard are recorded and compared with the corresponding ratios after the analysis of a series of aqueous ethanol solutions with known concentrations as the calibration standards. Alternatively, the ratio of heights of the two peaks on the gas chromatograph can be measured and used for the quantitative analysis.

The plot of peak height or peak area ratio against concentration of ethanol in the standards is used to construct a calibration line for quantitative analysis. A linear relationship exists between the detector response (peak area ratio) and the concentration of ethanol in the sample analyzed over a wide range, such as 0–5 $g\,l^{-1}$ (0–500 mg/100 ml) that might be encountered in forensic blood samples.

5. Quality control standards should be included within each analytical run dispersed liberally between the unknown blood samples.

Besides the analysis of unknown blood specimens, each analytical run should include one or more blank specimens (blood without any ethanol) and control standards of known concentration positioned at intervals in the series. Some laboratories might also include a sample containing a mixture of potential interfering substances (e.g., acetaldehyde, methanol, acetone, and isopropanol). The ethanol control standards should be prepared independently from the solutions used as calibrators and the latter should have good traceability to a primary source.

A wide range of instruments are available to perform a GC analysis of ethanol in blood and some manufactures produce dedicated instruments for headspace sampling and analysis (e.g., Perkin-Elmer Corporation). Other types of GC instruments are more flexible and can handle liquid injection and are adaptable for headspace analysis [16].

The first GC methods for alcohol analysis involved the use of a packed column, which consisted of a coiled tube made of glass, copper, or stainless steel that was normally 2–3 m long with an inside diameter of a few millimeters. The column contained an inert solid support material impregnated with the stationary liquid phase to facilitate separations and analysis. Currently, most GC analysis in forensic toxicology is done with capillary columns made of fused silica of normally 20–30 m long and with an inside diameter of 0.3–0.5 mm.

Figure 2 shows examples of chromatograms obtained from analysis of blood containing methanol, acetone, isopropanol as well as ethanol and the *n*-propanol internal standard. The blood specimen was analyzed on two different columns and stationary phases thus furnishing different RTs for the volatile components in the mixture. The detector response was normalized to 100% for the most prominent component in the sample. Note that the chart speed has been purposely increased for clarity, which causes the peaks to be fairly broad compared with if a slower chart speed had been used.

Few substances of physiological and forensic interest can be determined with such high accuracy and precision as the BAC. The analytical principles involved and brief practical details of the procedures used for the forensic determination of ethanol in biological specimens spanning the past century are presented in Table 5 in chronological order (*see also* **Confirmation Testing: Toxicology**).

Blood Sampling from Living Subjects

The quality of the specimen sent for analysis is often a neglected aspect of the overall procedures, which can have negative consequences for the reliability and acceptance of the results for legal purposes. The preanalytical stages include sampling, labeling, transport, and storage prior to arrival at the laboratory [17]. Factors to consider include obtaining informed consent from the subject or suspect, the blood sampling site on the body, the use of a disinfectant to swab the skin, the type of blood-sampling equipment used, whether an evacuated tubes or a syringe and needle. The procedures used to draw blood, the volume and type of container and whether chemical preservatives were added, the mode of transport of specimens, and the storage after arrival are all key elements in the overall reliability of the final results. The chain of custody of the specimens also needs to be well documented to guarantee integrity of laboratory work that might be used as evidence in criminal prosecutions.

Table 6 outlines major issues that deserve consideration when blood samples are collected and sent for forensic analysis of alcohol from living subjects.

The sampling site for obtaining blood might be a vein, an artery, or a capillary but for all practical purposes the sample is taken from a cubital vein. The blood is conveniently taken using a sterile evacuated glass tube (gray stopper) containing sodium fluoride as a preservative and potassium oxalate as an anticoagulant. The evacuated tubes usually hold a volume of 9–10 ml when full and 100 mg NaF (1% w/v) and 25 mg of potassium oxalate are present as preservatives. Immediately after sampling, the tube of blood should be gently inverted for about 20 s

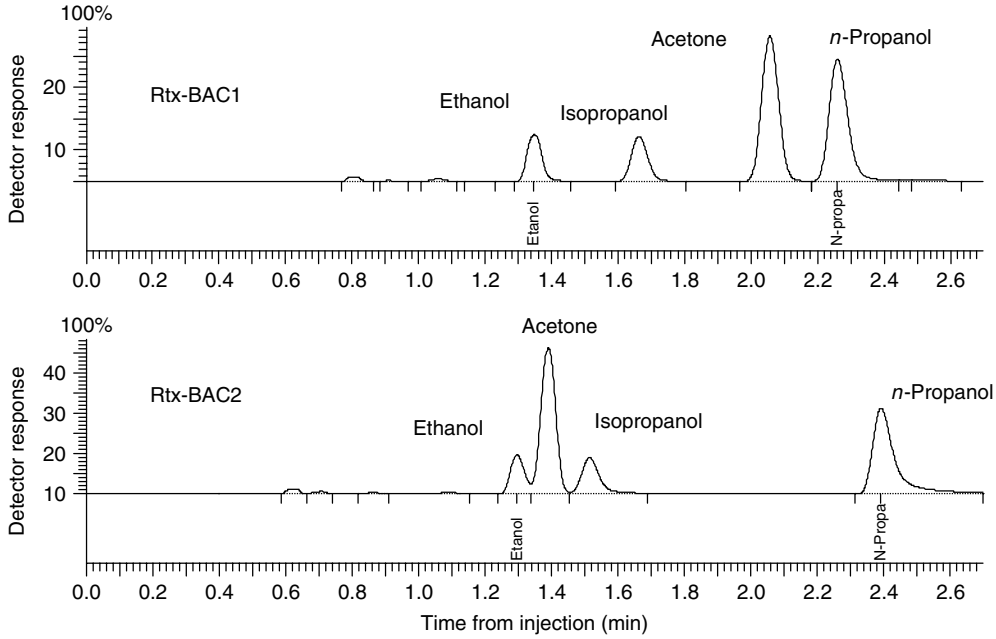


Figure 2 Chromatographic traces from analysis of ethanol and other volatiles in a blood sample also containing acetone and isopropanol. On two different stationary phases (Rtx-BAC1 and Rtx-BAC2) thus furnishing different retention times for identification of components

to ensure proper mixing with the preservatives to prevent coagulation. In practice, the volume of the blood in the evacuated tubes might vary from about 1 to 10 ml depending on the care taken with the sampling procedure. Such a variation in volume does not impact adversely on the accuracy or precision of the resulting BAC even when the volume of blood in the tube is abnormally low.

The evacuated tubes (5 ml) often used at hospital laboratories for collecting specimens for the analysis of blood-glucose are also suitable for blood-alcohol analysis. These 5 ml tubes normally contain less sodium fluoride and heparin instead of potassium oxalate is present as the anticoagulant. Because the tubes used to sample blood for alcohol analysis are sterile, one might wonder whether an enzyme inhibitor, such as sodium fluoride, is really necessary. However, a claim could be made that microorganisms had entered the blood when the needle penetrated the skin and ethanol, at least in part, was produced by the fermentation of blood-glucose after sampling. Swabbing the skin with a nonalcohol disinfectant or simply cleaning with soap and water is a standard practice before sampling blood. Studies have shown that even if an alcohol-containing swab had been used by mistake, the risk of contamination of the sample by carryover of ethanol from the skin is negligible.

Ethanol can be determined in any body fluid that contains water and methods have been described for the analysis of cerebrospinal fluid, saliva, sweat, and urine by the same GC methods described above for dealing with blood samples.

Specimens Taken at Autopsy

The methods described above for the determination of ethanol in blood from living subjects are the same as those used to analyze autopsy specimens [2]. During the postmortem examination a much wider selection of fluids or tissues are available for analytical toxicology. The composition of the autopsy blood samples in terms of fluidity, presence of clots, and degree of putrefaction might differ widely depending on the condition of the body and the circumstances surrounding the death [18]. When the analytical results are interpreted, the water content of the blood

Table 5 Historical developments in the methods used for determination of ethanol in blood and urine samples

Time period	Analytical procedure and brief details of the principles for ethanol determination
1900–1950s	The first analytical methods were based on chemical oxidation of the hydroxyl group in the ethanol molecule. After separation of ethanol from the biological matrix by diffusion, distillation or aeration or protein precipitation, an oxidizing agent such as a mixture of potassium dichromate ($K_2Cr_2O_7$) and sulfuric acid (H_2SO_4) was added. The ethanol is oxidized to acetic acid and the endpoint of the reaction is either determined by volumetric titration or by photometry
1950–1970s	Oxidation of ethanol by means of enzymes, such as alcohol dehydrogenase (ADH), offered milder conditions and gave a higher analytical selectivity when the enzyme derived from yeast was used. Ethanol was first separated from the biological matrix by precipitation of the blood-proteins with perchloric acid ($HClO_4$) and then centrifugation. After adjusting pH of the supernatant to alkaline conditions (e.g., pH 8.8) with buffers that contained the coenzyme (nicotinamide adenine dinucleotide, NAD^+), the ADH enzyme was added to start the reaction. This redox reaction, which is reversible, was driven to completion by adding semicarbazide hydrochloride, which reacts with the acetaldehyde produced from the oxidation of ethanol. During the reaction, the coenzyme (NAD^+) is reduced to NADH, which is monitored by absorption of ultra violet light at a wavelength of 340 nm. Enzymatic methods were better suited for automation, e.g., with various autoanalyzer instruments
1960s	Gas chromatography (GC) is today the method of choice and furnishes both a qualitative identification and a quantitative analysis of ethanol. An aliquot of the blood sample (~100 µl) is first diluted 1 : 5 or 1 : 10 with an aqueous solution of another alcohol (*n*-propanol or *t*-butanol) to serve as the internal standard. About 1–5 µl of the diluted blood is then injected into the heated injection port of the gas chromatograph and vaporized in a stream of nitrogen carrier gas (mobile phase). The volatiles pass through a long thin metal or glass tube (the GC column) containing the stationary liquid phase held on an inert support material. Depending on physicochemical properties and solubility in the liquid phase, the volatiles are held in the column for different times thus achieving separation. A detector, usually a flame ionization detector (FID), is used to measure the effluents when they emerge from the column. The time after injection to appearance of the apex of the peak (retention time) is a characteristic of the substance analyzed and can be used for identification. The area under the peak was related to the amount of analyte. Quantitative analysis required calibration of the detector response with known strength alcohol standards
1970s	Instead of injecting the diluted blood specimen a modification of the basic GC procedure involved the use of headspace analysis. This entailed sampling a portion of the vapor phase in equilibrium with the blood sample. This offers the advantage that the column packing material is not contaminated by nonvolatile constituents of the biological matrix. As with liquid injection, the blood samples and ethanol standards are first diluted (1 : 5 or 1 : 10) with internal standard and transferred to glass vials made airtight with rubber septa and crimped on aluminum caps. The vials are equilibrated at 50 or 60 °C and after about 30 min a portion of the vapor phase is removed for GC analysis. Sampling is done with the aid of a gas-tight syringe or by an automated procedure. The original packed GC columns have been gradually replaced by capillary or wide bore columns. The sensitivity of the headspace method can be enhanced and matrix effects can be eliminated by saturating the blood samples and aqueous ethanol standards with salt (NaCl or K_2CO_3), e.g., 0.5 ml blood + 1 g salt
1980s	Positive identification of ethanol is accomplished by mass spectrometry with an electron-impact detector producing characteristic ion fragments m/z 31 (base peak for primary alcohols), m/z 46 (molecular ion), and m/z 45 after loss of a proton. The ions m/z 31 and m/z 45 are sufficiently intense to allow quantitative analysis and deuterium-labeled ethanol can serve as internal standard

deserves consideration, such as when conclusions are reached regarding the antemortem BAC. Some investigators recommend that the water content of autopsy blood samples should be determined in conjunction with the analysis of ethanol. This is easily done by desiccation or freeze drying an aliquot of the specimen so that the BAC can then be adjusted to a blood-water content of 80% w/w, which applies to fresh blood from living subjects.

Table 7 gives a list of the biological specimens recommended for use in connection with forensic analysis of ethanol from living subjects and also in postmortem toxicology. (*see also* **Toxicology: Analysis**).

Table 6 Important considerations (preanalytical factors) in connection with sampling of blood from suspected drunken drivers intended for analysis of ethanol for legal purposes

Aspects of blood sampling	Considerations
Qualifications of the person commissioned to draw blood	Doctor, nurse, laboratory technician, or trained phlebotomist
Preparation of the patient or person	Informed consent about the purpose of sampling and information whether the person was sitting, standing, laying down, struggling, conscious/unconscious, or a victim of a traffic crash?
Blood source and sampling site	Source of the blood sample whether left or right arm cubital vein, artery, fingertip, or earlobe? Any medical treatment given to the person prior to sampling, such as intravenous fluids
Sampling technique	Type of skin disinfection used if any. Syringe and needle or an evacuated tube – was a tourniquet applied? Volume of blood specimen collected and the time needed for obtaining the sample
Identification and labeling of the specimen.	Record name of suspect, date of birth, time and date of blood draw, and name of the person who took the blood sample
Collection tubes for blood and preservatives	Volume of tubes and whether these were made of glass or plastic. What anticoagulant was present and whether an enzyme inhibitor such as sodium fluoride was included and at what concentration
Transport of sample to forensic laboratory	Tamperproof packaging, mode of transport to the laboratory (normal postal service or special delivery). Where the specimens were refrigerated during transport. How was chain of custody ensured
Inspection and registration at the laboratory	Date and time of arrival (date stamp). How many tubes of blood were received? What volume of blood in the tubes? Was there obvious hemolysis or clotting, any attempt at manipulation? Were security tapes or seals still intact? Had the tubes been opened
Storage prior to analysis	Time delay from sampling to analysis, storage of blood in a refrigerator or frozen and known stability of the analytes

Importance of Method Validation

Quality assurance and process control are crucial in all kinds of routine analytical laboratory work and especially when results are used in criminal investigations [19]. All analytical methods should undergo a rigorous validation and standardization before they can be considered fit for purpose. A careful documentation of the method characteristics including information on accuracy, precision, and the results of external proficiency tests are essential in today's laboratory environment [19]. The uncertainty

Table 7 Biological specimens suitable for analysis of ethanol in living subjects and also at autopsy

Living subjects	Dead bodies
Venous blood (10 ml)[a]	Femoral blood (20 ml)[b]
Capillary blood (fingertip)	Heart blood (if femoral unavailable)
Plasma/serum	Blood clot (subdural hematoma)
Urine fresh void (10 ml)[a]	Urine (50 ml)[b]
Tear fluid	Vitreous humor (all available both eyes)[b]
Cerebrospinal fluid (lumber fluid)[c]	Cerebrospinal fluid (cisternal fluid)[d]
Saliva	Bile (gall bladder)
Perspiration/sweat	Synovial fluid (knee joint)
Breath (end expired)[a]	Tissue (brain, skeletal muscle, liver)

[a] Recommended for analysis in people arrested for drunken drivers
[b] Recommended for analysis in connection with forensic autopsies
[c] From base of spine with the patient in a crouching position
[d] From back of neck (lumber sampling not practical)

in the analytical results also needs to be recognized, reported, and allowed for especially if these are compared with some threshold value, such as the legal alcohol limit for driving [15].

Method validation embraces the process and events by which a newly developed method or laboratory procedure is subjected to a strict testing protocol to establish key performance characteristics including accuracy, precision, specificity, bias, linearity, LOD, and LOQ.

Many of the terms used in connection with validation and accreditation of laboratory methods are defined in Table 8.

Over the past few decades, more and more attention has been given to laboratory accreditation, which is a formal process by which a laboratory is evaluated for its competence to perform designated analytical tasks or measurements [20]. Aspects of the accreditation process include participation in external proficiency schemes as well as regular onsite inspections by outside experts who examine and evaluate protocols, oversee laboratory facilities, and verify that the staff has appropriate qualifications and training for their task. In today's climate a forensic science or toxicology laboratory would not survive for long without being accredited for the services it provides including regular inspections and some form of peer recognition.

When the results of blood-alcohol analysis are reported to the police authorities, some laboratories consider the uncertainty in analytical results and allow for this in the report. This is done by making a deduction from the mean result of duplicate or replicate determinations. The amount deducted depends on the reliability of the analytical method and the magnitude of random and systematic errors. Because precision tends to decrease as the concentration of ethanol in the blood increases, the amount deducted is greater at high BAC to ensure the same degree of confidence (e.g., 99 or 99.9%) in the final result. After making the deduction, a statement can be made to the court to the effect that the person's BAC is not less than the value reported with a high degree of confidence such as 99 or 99.9%.

Principles and Application of Breath-Alcohol Analysis

It must be a very old observation that some portion of the alcohol a person consumes gets exhaled via the lungs and can be detected on the breath [9]. Indeed, the smell of alcohol on the breath, together with the person's general appearance and behavior, often constituted the first indications, albeit primitive, of overindulgence in alcohol. The first scientific studies attempting to measure accurately the concentration of alcohol in a person's breath were published over 150 years ago [9]. These showed that only a small fraction of the alcohol consumed (2–5%) could be recovered unchanged in the breath and urine collected for several hours afterward. The bulk of the dose of alcohol ingested underwent biochemical oxidation in the liver to provide a rich source of calories (7.1 kcal per gram ethanol).

These first studies on breath-alcohol concentration (BrAC) also identified the potential problem caused by the presence of alcohol dissolved in the mucous surfaces of the mouth from a recent drink. This led to a warning (cited below from FE Anstie, The Lancet, Sept. 28, 1867), which still holds true. When a breath-alcohol test is made for evidential purposes there is always a mandatory 15–20 min deprivation and observation period after the last drink to avoid the risk of contaminating the sample with alcohol in the mouth [21].

> Much caution is necessary, however, in applying this test. It must not be tried during at least the first quarter of an hour after a dose has been taken, for the mouth retains the characteristic smell, even of the most moderate dose, of any of the stronger smelling drinks, for fully this time.

Human breath consists of a mixture of gases being mainly oxygen, nitrogen, carbon dioxide and is saturated with water vapor at body temperature [22]. Additionally, expired air contains trace amounts of other organic volatile compounds (VOCs), which are produced either naturally in the body (endogenous volatiles) or inhaled together with the ambient air breathed. The most abundant VOCs in human breath are ethanol, methanol, acetone, carbon monoxide, methane, and isoprene. The concentrations of these substances are normally extremely low and have no practical relevance to challenge the results when these are used for legal purposes. However, under some circumstances, such as diabetes or other metabolic disorder, or after eating very low carbohydrate diets or if a person consumed denatured alcohol, the concentrations of acetone and isopropanol in breath samples might increase appreciably.

Table 8 Important characteristics to consider during validation of an analytical method intended for forensic purpose, such as the determination of ethanol in biofluids

Characteristic of the analytical method	Brief description and/or definition of the characteristic
Accuracy	Accuracy is a measure of the closeness of agreement of analytical results with the true or known concentration of substance in the sample. The true value might be the accepted reference value or a "target" concentration obtained by spiking with a known quantity of analyte
Bias	Bias is a systematic deviation either constant or proportional to the concentration of the analyte. Bias has to do with correctness of the analytical result and is derived as the difference together with its sign between the known quantity or target value and the result or average of several determinations by the method
Precision	Precision is the degree of mutual agreement between independent measurements under specified test conditions. This reflects the spread of analytical results when the method is applied repeatedly to aliquots of a homogenous sample. Random errors inherent in the method determine the analytical precision. In mathematical terms precision is computed as the standard deviation of repeated measurements
Repeatability	This term is used to define within-run analytical precision, same laboratory same operator same equipment – short-term precision
Intermediate precision	Precision of analysis between-runs in the same laboratory over longer time periods (weeks) by different operators, instruments, reagent batches, and calibrations
Reproducibility	This term is used to define precision of the analytical method when performed in different laboratories, with different instruments and analysts, such as the standard deviation in external proficiency testing
Linearity	The ability of an analytical method to give results directly proportional to the concentration of the analyte in the sample within a defined range of values
Range	The concentration interval between the upper and lower levels of analyte that can be determined with acceptable accuracy and precision. The range usually represents the difference between the lowest and the highest points on the calibration curve
Sensitivity	This is defined as the difference in analyte concentration corresponding to the small detectable difference in the detector response. It is represented by the slope of the calibration curve.
Specificity/selectivity	These terms are often used interchangeably and are related to the ability of a particular method to quantify the target analyte in the presence of interfering compounds
Recovery	Normally expressed as a percentage, recovery refers to the amount of drug removed from the original sample which reaches the end of the analytical procedure. Poor recovery can be compensated for by adding an internal standard to the biofluid before staring the analysis
Limit of detection (LOD)	LOD is the lowest concentration of analyte in a sample that can be detected but not necessarily quantified
Limit of quantitation (LOQ)	LOQ is the lowest concentration of an analyte in a sample that can be quantified with acceptable precision and accuracy. LOQ is approximately $3.3 \times$ LOD
Robustness	Robustness is the capacity of an analytical method to produce accurate and precise results despite small deliberate changes in test conditions and method parameters. In GC analysis of ethanol the addition of an internal standard helps to ensure the method is robust
Ruggedness	Ruggedness is a measure of the degree of reproducibility of the analytical results when performed under varying tests conditions, such as when work is done by different technicians, instruments, source of reagents, laboratories, or even in different countries
Traceability	Traceability means that a result of measurement can be traced back, through an unbroken chain of comparisons, to a national or international standard value. The traceability of the ethanol standards used to calibrate the gas chromatography needs to be well documented
Uncertainty	The word uncertainty means having doubt in something, such as the result of analysis. Analytical uncertainty is defined as a parameter associated with the result of a measurement, which characterizes the spread or dispersion of the results that could reasonably be attributed to the quantity being measured

The physiological principles that govern the excretion of alcohol in breath and the notion of indirectly estimating the BAC by breath analysis emerged in the 1930s when the concept of a blood/breath distribution ratio was formulated [22]. The first use of BrAC was to estimate the person's coexisting BAC, because this was accepted as the best objective evidence of drunkenness. The concentration of alcohol in the breath sample analyzed was multiplied by the blood/breath ratio of alcohol (2100 : 1), which allowed reporting the corresponding BAC. Some breath-test instruments, such as the Breathalyzer®, was designed and calibrated based on the assumption of a 2100 : 1 factor [22].

$$BAC = BrAC \times 2100 \quad (5)$$

Experience has shown that use of the above formulae tended to underestimate the true coexisting venous BAC by about 10% [23]. In some applications of breath-alcohol testing, the factor 2300 : 1 is used for calibration, which gives a more unbiased estimate of the venous BAC. However, the above equation is physiologically incorrect because BrAC runs closer to the concentration of ethanol in arterial (A) blood, whereas the sample analyzed for legal purposes is venous (V) blood. The arterial BAC (e.g., blood from a radial artery) is higher than the venous BAC (e.g., from a cubital vein) during the absorption and distribution stages of the BAC curve. The arterial BAC and venous BAC are the same at one time point (60–90 min post drinking), which marks complete equilibration of the ingested dose between the blood and tissue water. After this time, the arterial BAC is less than the venous BAC and remains less for the remainder of time that alcohol is being metabolized. The A–V differences are more pronounced on the rising or absorption phase of the BAC curve and the negative A–V differences on the descending postabsorptive phase is rather small or negligible and can be ignored [2]. In forensic practice it has become common to assume that most drunken drivers are apprehended 1–2 h after the end of drinking when the blood-alcohol curve has entered the postpeak descending phase.

Methods of Analyzing Alcohol in Breath

The first methods of breath-alcohol analysis employed chemical oxidation principles as exemplified by the classic Borkenstein Breathalyzer® device, which was widely used for law enforcement purposes, such as for testing drunken drivers in the United States, Canada, and Australia. The oxidizing reagent consisted of a mixture of sulfuric acid and potassium dichromate contained in a glass ampoule through which a known volume of the subject's breath was passed. Any ethanol present in the breath sample was oxidized to acetic acid with a concomitant reduction of chromium VI (yellow) in dichromate to chromium III (green) and this color change was monitored by photometry after exactly 90 s of reaction. Accordingly the reaction endpoint and the quantitative analysis of ethanol in breath were determined by optical photometry.

In the 1970s, physicochemical methods were developed for the analysis of alcohol in samples of breath particularly the use of infrared absorptiometry, which was incorporated in the Intoxilyzer range of instruments [24]. Figure 3 shows the infrared spectrum of ethanol in the gas phase and the major absorption bands corresponding to vibrational frequencies in the ethanol molecule. Those used for quantitative analysis at 3.4 μm (C–H stretch) and 9.5 μm (C–O stretch) are indicated.

The quantitative analysis of ethanol utilizes the Lambert–Beer law and monitors the absorption of radiation at 3.4 μm, which corresponds to the carbon–hydrogen stretching in the alkyl groups of ethanol. Under some conditions, acetone might be present in human breath at elevated concentrations and can also absorb IR radiation at 3.4 μm, which suggests a potential interference problem. Accordingly selectivity of the analysis of ethanol was enhanced in later versions of the Intoxilyzer instruments by monitoring IR radiation and more than one wavelength by use of narrow band-path filters set at 3.38 and 3.49 μm so that acetone could be distinguished from ethanol. The latest generation of infrared breath-alcohol instruments incorporate filters to monitor IR radiation also at a wavelength of 9.5 μm, which corresponds to the C–O stretching frequency in the alcohol molecule (see Figure 4). A few instruments (e.g., Intoxilyzer 9000) use both 3.4 and 9.5 μm to give an added analytical selectivity. By contrast, the Alcotest 7110 is a dual detector device with IR absorption at 9.5 μm and electrochemical oxidation (fuel cell) to ensure high selectivity for identification of ethanol. The advantage of IR analysis is that this is a nondestructive analytical method and

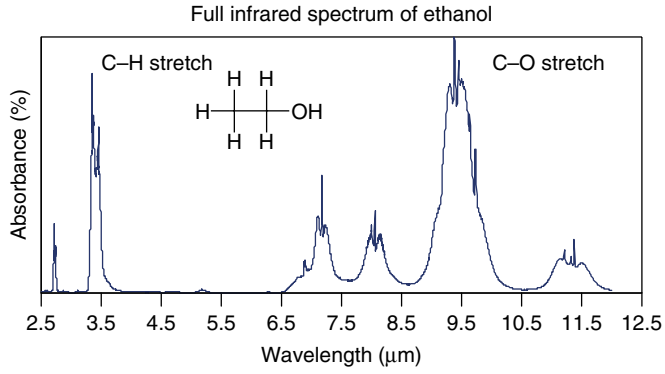


Figure 3 Infrared spectrum of ethanol in the vapor phase showing the absorption bands around 3.4 μm (C–H stretching) and 9.5 μm (C–O stretching) used for quantitative determination in modern evidential breath-ethanol analyzers

if necessary the sample analyzed can be preserved and used for other purposes.

The methodologies (analytical principles) incorporated into a wide selection of devices and instruments for breath-alcohol testing over many years are described in Table 9 [24, 25]. In this table, one should distinguish breath analyzers intended for roadside screening analysis with those designed for evidential purposes. The first screening devices were constructed from chemical tubes containing potassium dichromate and sulfuric acid impregnated on a silica gel support material. This mixture of chemicals changed in color from yellow to green if there was alcohol in the suspect's breath after blowing through the tube and inflating a balloon of fixed volume (~1 l).

The breath-alcohol screening devices used today by the police for roadside mass testing of motorists are small handheld instruments that incorporate an electrochemical sensor for the analysis of ethanol. Many such instruments are available such as Alcolmeter, Alcotest, AlcoSensor, and LifeLoc, which are sometimes referred to as *fuel cell instruments* (see Table 9).

The ethanol contained in a small portion of the end-exhaled breath (~1.0 ml) is directed via an inlet tube into a chamber fitted with an electrode (platinum black) and an acidic electrolyte solution, such as phosphoric acid. Oxidation of ethanol at the electrode surface produces acetaldehyde with a simultaneous liberation of electrons. These are captured, amplified, and displayed as a measurable current in direct

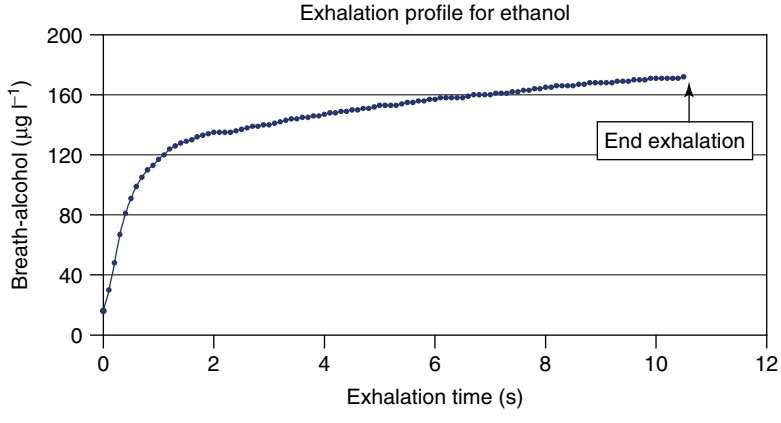


Figure 4 Increase in breath-alcohol concentration during a prolonged exhalation into a modern infrared breath-alcohol analyzer

Table 9 Analytical principles used for determination of ethanol with a wide selection of instruments and devices intended for breath-alcohol analysis

Analytical method	Example of breath instruments	Brief details of the operating procedure and principles
Chemical oxidation	Drunkometer Alcometer Breathalyzer Alcotest tubes Alcolyzer tubes	Ethanol in a known volume of breath is analyzed by oxidation with a mixture of potassium dichromate and sulfuric acid, or potassium permanganate or iodine pentoxide. The endpoint of the reaction was either a visual color change (Alcotest and Alcolyzer tubes) or by UV photometry (Breathalyzer)
Gas chromatography	GC Intoximeter AlcoAnalyzer GC	Compact instruments designed for analysis of ethanol in exhaled breath or vapor samples. The gas chromatograph (GC) Intoximeter incorporated a flame ionization detector and the AlcoAnalyzer used a thermal conductivity detector. The need for an external gas supply and frequent recalibration and maintenance meant they these instruments were not very practical for use at police stations
Infrared spectrometry	Intoximeter DataMaster Intoxilyzer Alcotest Evidenzer	Ethanol in the exhaled breath passes through a sample chamber ($\sim$50 ml) and analysis is done by measuring the absorption of infrared radiation according to the Lambert–Beer law. Ethanol molecules absorb radiation at wavelengths of $\sim$3.4 μm corresponding to the C–H stretching and $\sim$9.5 μm corresponding to the C–O stretching frequencies. To enhance selectivity for identifying ethanol, the absorption of infrared radiation is done at several different wavelengths
Electrochemical oxidation	AlcoSensor Alcolmeter Alcotest Lifeloc	Ethanol in a fixed volume of breath ($\sim$1.0 ml) undergoes electrochemical oxidation with the help of a platinum black catalyst and acid electrolyte mounted with electrical connections to form a fuel cell. The molecules of ethanol enter one side of the cell and are oxidized to acetaldehyde thus liberating electrons. The current produced is proportional to the concentration of ethanol in the sample of breath
Metal oxide semiconductor (Taguchi cell)	ALERT J4X	Many small handheld devices incorporate a Taguchi sensor and these are generally inexpensive and mainly intended as self-testers for use by the public. The ALERT J4X incorporated a Taguchi cell and this unit gained approval as a breath-alcohol screening instrument by police forces in Canada. The Taguchi sensor incorporates a tin-oxide (SnO_2) bead mounted on a ceramic cylinder that measures changes in surface conductivity. The heated bead has a high surface resistance in ambient air but when exposed to combustible gases the surface conductivity increases in proportion to the concentration of substance in the gas phase. The semiconductor is not specific for analysis of ethanol and constituents of cigarette smoke as well as acetone on the breath give a response

proportion to the number of molecules of ethanol reacting at the platinum electrode surface. The output from the fuel cell changes as a function of time after starting the reaction to reach a peak response after about 20 s. The response curve then exponentially falls back to zero. Quantitative analysis is achieved by measuring the peak response or the area under the entire response curve. To speed up the time between tests, the return to a zero signal is achieved by short circuiting the cell in readiness for the next subject test.

One important category of breath-alcohol instruments are those intended and used for evidential testing, that is to generate evidence of sufficient reliability that drunken drivers can be prosecuted. Most evidential breath-alcohol instruments incorporate infrared technology for the detection and analysis of ethanol as discussed above [25]. The test subject

exhales into the heated breath inlet tube and makes a continuous forced exhalation for as long as possible, usually for at least 6 s. The concentration of ethanol increases as a function of time during a prolonged exhalation, first rapidly and then more slowly as the volume of exhaled breath reaches a maximum forced exhalation (Figure 4).

Quality assurance of the breath sample is achieved by means of slope detectors that help to monitor the shape of the breath-alcohol exhalation profile to ensure that it conforms to that expected for normal human breath containing only ethanol [24, 25]. The BrAC at the end of exhalation gives a good indication of the concentration of alcohol residing in the deep lung alveolar air as it leaves the lungs. Instruments used for evidential testing should provide a print-out of the results in real time including identification of the person tested, proof of calibration control of the instrument and the date and time of testing as well as all analytical results. It is good forensic practice to make two independent tests of the person's breath-alcohol concentration about 6 min apart and take the average result [24].

An important aspect of any chemical analysis is the calibration and standardization of the measuring equipment used [19,21]. Calibration refers to the process by which the relationship between the output signal or response of the instrument and the value of the input quantity or concentration is determined. For the purpose of calibration, a series of ethanol standards with known strength are prepared to cover the range of interest. The calibration and control of breath-alcohol instruments is done either by means of a so-called wet-bath simulator device or by use of a compressed dry gas cylinder containing a known content of ethanol mixed with an inert gas such as argon. The wet-bath simulators are the traditional way of calibrating breath analyzers and have the advantage that they resemble the biological specimen (human breath) in composition, that is, being saturated with water vapor at the temperature of exhaled breath (34 °C). However, the compressed gas standards are more convenient to handle when tests are made away from the laboratory, such as at a police station or in a police vehicle [24]. It seems that compressed dry gas standards are gradually replacing the traditional wet-bath simulators for the purpose of calibration control checks of evidential breath-testing instruments.

Considerable interest exists in the marketing and sale of breath-testing instruments that are intended primarily for use by the general public as a means of self-testing. These devices are usually small, compact and battery operated, and fairly cheap to buy. However, the results obtained are not always reliable and trustworthy for several reasons. First, a proper control of the way the sample of breath is introduced into the instrument cannot be guaranteed. Second, the alcohol sensors incorporated are usually tin-oxide semiconductors that measure changes in conductivity, which is a nonspecific way to measure ethanol (Table 9). Third, a control of the calibration stability of the instrument is rarely done during long-term usage. Fourth, cigarette smoke, carbon monoxide, exhaust fumes, methane, and other combustible gases react and give a response indistinguishable from ethanol. Accordingly, the use of these self-testers for alcohol should not be encouraged. Another rapidly emerging public safety application of breath-alcohol testing is in connection with ignition interlock devices or so-called alcolocks. These devices are fitted to public transportation vehicles (e.g., buses and trains) and also into many private cars. The aim is to prevent a person who has consumed too much alcohol from being able to start the vehicle and reduce the risk of drunken driving. If alcohol is detected on the person's breath above a predefined threshold level, the ignition switch is locked and the vehicle will not start.

The concentration of ethanol in breath at which the vehicle fails to start is usually set to be fairly low, such as at a BAC of $0.2\,g\,l^{-1}$ (20 mg/100 ml) or less, even though the legal limit for driving might be higher (e.g., 0.5 or $0.8\,g\,l^{-1}$). The overall aim of interlock devices is to improve traffic safety by reducing the risk of alcohol-related crashes. Interlocks are particularly useful as a countermeasure to prevent a previously convicted drunken driver from reoffending. Indeed, some countries make it one of the requirements for relicensing to have an interlock fitted to the vehicle belonging to a previously convicted drunken driver. Most of the currently available interlock devices make use of the electrochemical oxidation reaction to analyze ethanol (Table 9).

Concluding Remarks

Knowledge about a person's BAC or BrAC provides compelling evidence when a conclusion is reached about the effects of alcohol, such as the degree of drunkenness, impairment of cognitive and

psychomotor functioning, and the ability to form intent. Discussion and debate about culpability is tightly linked to a person's BAC and this has gained paramount importance in criminal proceedings, such as investigation of alcohol-related crimes including murder, physical and sexual assaults, and drunken driving. This means that the quantitative analysis of alcohol in body fluids is undoubtedly the highest volume investigation performed at most forensic science and toxicology laboratories worldwide [26, 27].

Few substances can be determined in blood and other body fluids with such a high degree of accuracy, precision, and selectivity as the concentration of ethanol [15, 19]. Because the alcohol a person drinks becomes diluted with the total body water, which is ~50–60% of body weight, the amount of ethanol necessary to cause inebriation and impairment vastly exceed that of all other drugs of abuse. The BACs, typically encountered in legal medicine ($0.2–5.0\,g\,l^{-1}$), are about 1000 times higher than those for most other licit or illicit drugs.

The sampling and analysis of alcohol in breath has the advantage that it is noninvasive and the results of the test are obtained immediately afterward [4]. The on-the-spot results of a breath-alcohol test allow immediate decisions to be made, such as allowing a motorist to continue driving, or whether a casualty patient might require emergency surgery e.g., for head trauma or simply be allowed to recover from gross intoxication. Breath-alcohol instruments are being increasingly used in connection with workplace alcohol testing, at probation and rehabilitation centers where people must refrain from drinking and also at school parties and dance venues where alcohol use is forbidden.

This review hopefully made it clear that ethanol can be determined in body fluids with a high degree of accuracy, precision, and specificity by a variety of analytical methods [20, 23]. However, the gold standard procedure is the application of GC together with the headspace sampling technique. Such GC methods furnish both a qualitative analysis, based on the comparison of the RT of components in the sample with authentic known substances, and also a quantitative analysis based on detector response (peak height or peak area) displayed on the chromatogram [16]. The GC methods of analysis are easily automated and under optimum conditions, the response for ethanol is adequately resolved from that of other low molecular weight volatiles that might be present in forensic bloods samples, such as acetaldehyde, methanol, acetone, and isopropanol [15].

Although the methods are the same for dealing with specimens from the living and the dead, the analytical results from autopsy specimens are more difficult to interpret, owing to various postmortem artifacts. This limitation is offset to some extent by the fact that many different biological specimens are available for toxicological analysis during a postmortem examination. Aspects of postmortem alcohol analysis and interpretation have been dealt with in depth in several recent review articles [27, 28]. The recommended specimens for the determination of ethanol in corpses are femoral venous blood from the left or right leg (Table 7) and use of heart blood or from the chest cavity should be avoided. To simplify interpretation of analytical results, pathologists should strive to obtain additional specimens, such as urine from the intact bladder and vitreous fluid from the eyes, for the analysis of ethanol [29].

The consumption of alcoholic beverages is part and parcel of normal social life in most countries and for many individuals, especially among men, moderate drinking often escalates into overconsumption, abuse and dependence. Binge drinking and drunkenness have many negative consequences for the individual and society and results in premature death. Heavy drinking tends to trigger deviant and aggressive behavior, such as family violence and drunken driving. The continued use and abuse of alcohol in society ensure that requests to measure this drug in body fluids and to interpret the results in a legal context will remain the most commonly requested service from forensic science and toxicology laboratories in the distant future.

References

[1] Room, R., Babor, T. & Rehm, J. (2005). Alcohol and public health, *Lancet* **365**, 519–530.

[2] Jones, A.W. & Pounder, D.J. (2007). Update on clinical and forensic analysis of alcohol Chapter 5.2, in *Drug Abuse Handbook*, 2nd Edition, S.B. Karch, ed, CRC Press, Boca Raton, pp. 333–376.

[3] Hunsaker, D.M. & Hunsaker, J.C. (2005). Blood and body fluid analysis, in *Encyclopedia of Forensic and Legal Medicine*, J. Payne-James, R.W. Byard, T.S. Corey & C. Henderson, eds, Elsevier, Oxford, pp. 29–38.

[4] Jones, A.W. (1988). Enforcement of drink-driving laws by use of per-se legal alcohol limits: blood and/or breath

concentration as evidence of impairment, *Alcohol, Drugs and Driving* **3**, 99–112.
[5] Moriya, F. (2005). Forensic sciences – alcohol in body fluids, in *Encyclopedia of Analytical Sciences*, 2nd Edition, Elsevier Sciences, pp. 368–366.
[6] Dubowski, K.M. (1986). Recent developments in alcohol analysis, *Alcohol, Drugs and Driving* **2**, 13–46.
[7] Jones, A.W. (2000). Medico-legal alcohol determination – blood or breath alcohol concentration? *Forensic Science Review* **12**, 23–48.
[8] Jones, A.W. (1989). Measurement of alcohol in blood and breath for legal purposes, in *Human Metabolism of Alcohol*, *Vol 1, Pharmacokinetics, Medicolegal Aspects, and General Interest*, K.E. Crow & R.D. Batt, CRC Press, Boca Raton, Florida, pp. 71–99.
[9] Jones, A.W. (1996). Measuring alcohol in blood and breath for forensic purposes – a historical review, *Forensic Science Review* **8**, 13–44.
[10] Widmark, E.M.P. (1922). Eine Mikromethode zur Bestimmung von Äthylalkohol im Blut, *Biochemische Zeitskrift* **131**, 473–484.
[11] Jones, A.W. (1995). Forensic science – determination of alcohol in body fluids, in *Encyclopedia of Analytical Sciences*, Academic Press, London, Vol. 2, pp. 1585–1594.
[12] Kristoffersen, L. & Smith-Kielland, A. (2005). An automated alcohol dehydrogenase method for ethanol quantitation in urine and whole blood, *Journal of Analytical Toxicology* **29**, 387–389.
[13] Logan, B.K. (1992). Analysis of alcohol and other volatiles, in *Gas Chromatography in Forensic Sciences*, Ian Tebbett, ed, Elsevier, North Holland, pp. 87–108.
[14] Emerson, V. (2004). Alcohol analysis, in *Crime Scene to Court; the Essentials of Forensic Science*, 2nd Edition, P.C. White, ed, Royal Society of Chemistry, Cambridge, pp. 350–376.
[15] Jones, A.W. & Schuberth, J. (1989). Computer-aided headspace gas chromatography applied to blood-alcohol analysis: importance of online process control, *Journal of Forensic Sciences* **34**, 1116–1127.
[16] Tagliaro, F., Lubli, G., Ghielmi, S., Franchi, D. & Marigo, M. (1992). Chromatographic methods for blood alcohol determination, *Journal of Chromatography* **580**, 161–190.
[17] Scopp, G. (2004). Preanalytical aspects of postmortem toxicology, *Forensic Science International* **142**, 75–100.
[18] Pounder, D.J. & Jones, A.W. (2007). Postmortem alcohol - aspects of interpretation. Chapter 5.3, in *Drug Abuse Handbook*, 2nd Edition, S.B. Karch, ed, CRC Press, Boca Raton, pp. 376–401.
[19] Taylor, J.K. (1987). *Quality Assurance of Chemical Measurements*, Lewis Publishers, Chelsea, Michigan.
[20] Burnett, D. (1996). *Understanding Accreditation in Laboratory Medicine*, A C B Venture Publications, Association of Clinical Biochemists.
[21] Gullberg, R.G. (2000). Methodology and quality assurance in forensic breath alcohol testing, *Forensic Science Review* **12**, 49–68.
[22] Jones, A.W. (1990). Physiological aspects of breath alcohol measurement, *Alcohol, Drugs and Driving* **6**, 1–25.
[23] Gullberg, R.G. (2005). Breath alcohol analysis, in *Encyclopedia of Forensic and Legal Medicine*, J. Payne-James, R.W. Byard, T.S. Corey & C. Henderson, eds, Elsevier, Oxford, pp. 21–29.
[24] Gullberg, R.G. (2006). Estimating the measurement uncertainty in forensic breath-alcohol analysis, *Accreditation Quality Assurance* **11**, 562–568.
[25] Harding, P. & Zettl, J.R. (2008). Methods for breath analysis, in *Medical-Legal Aspects of Alcohol*, 5th Edition, J.C. Garriott, ed, Lawyers & Judges Publishing Company, Tuscon, pp. 229–253.
[26] Garriott, J.C. (ed) (2008). *Garriott's Medicolegal Aspects of Alcohol*, 5th Edition, Lawyers & Judges Publishing Company, Tuscon, pp. 1–534.
[27] Carey, K.B. & Hustad, J.T.P. (2005). Methods for determining blood alcohol concentration current and prospective, in *Comprehensive Handbook of Alcohol-Related Pathology, Vol 3*, V. Preedy & R. Watson, eds, Academic Press, pp. 1429–1444.
[28] Kugelberg, F.C. & Jones, A.W. (2007). Interpreting results of ethanol analysis in postmortem specimens, *Forensic Science International* **165**, 10–29.
[29] Jones, A.W. & Holmgren, P. (2001). Uncertainty in estimating blood-alcohol concentration by analysis of vitreous humor, *Journal of Clinical Pathology* **54**, 699–702.

Further Reading

Dubowski, K.M. (1991). *The Technology of Breath-Alcohol Analysis*, US Department of Health and Human Services, Public Health Service DHHS Publication No. (ADM), pp. 1–38.

ALAN W. JONES

Alcohol: Behavioral and Medical Effects

Introduction

Alcohol is one of the oldest drugs known and it affects virtually every organ system in the body. The medical consequences including the economics of medical treatment of alcohol-related disorders are staggering and can have important forensic implications. Alcohol damages the heart and can elevate

blood pressure, increasing the risk of heart failure and stroke and can impair biochemical regulation of a variety of cellular and metabolic functions that can increase the risk for certain forms of cancer, the risk for accidental injuries and impairs the recovery from those injuries and significantly contribute to the years of life lost. This article reviews the most significant and well-known medical consequences of alcohol in three basic areas of forensic interest: accidental injuries, skeletal fragility, and liver pathologies.

Alcohol and Accidental Injuries

Accidental injuries are a direct medical consequence of alcohol intoxication. Laboratory as well as epidemiological field studies conducted over the last few decades clearly establish the fact that alcohol-induced impairment of cognitive and psychomotor functioning while engaging in a variety of behavioral activities increases the risk for injury. Among these, the effects of alcohol on automobile, bicycle, motorcycle, boating, aquatic, and pedestrian injuries, as well as homicide, suicide, and death from fire have been examined.

Impaired Driving

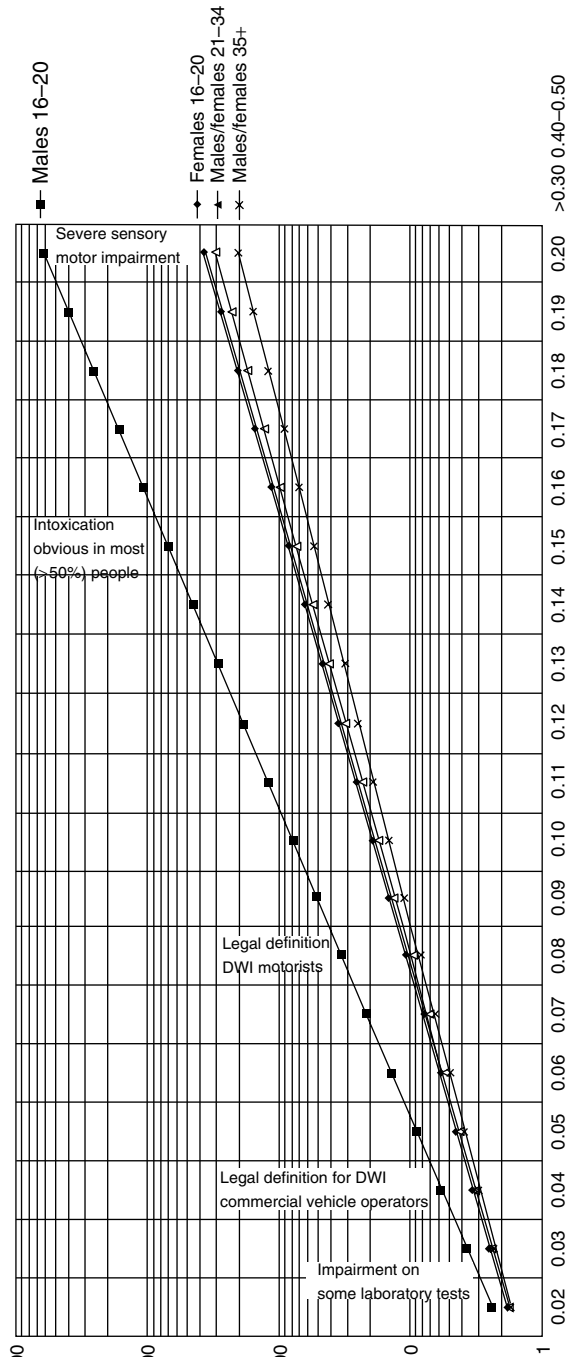
Driving while intoxicated is probably the well-studied injurious consequence of drinking of interest and is of importance to forensic examiners. The use of alcohol coupled with increased risk taking and impulsivity, at least among young males [1], and decreased seat belt use [2]. Even at very low blood–alcohol concentrations (BACs), the performance of complex divided attention, a critical factor in a variety of tasks both inside and outside the laboratory, is impaired and the most likely cause of motor vehicle collisions. At BAC above 50 mg/dl that impairment translates into actual highway statistics (in which the intoxicated driver is deemed to be the cause of the accident). At higher BACs (e.g., 150 mg/dl, or more), impairment in proprioception, visual perception, and lengthened simple reaction time are additional significant contributing factors that should be considered into forensic investigations of crashes involving intoxicated drivers, pedestrians, and others. Most people who present with obvious symptoms of intoxication are impaired drivers and at increased risk for a fatal crash. However, the lack of obvious intoxication does not mean lack of impairment. Most subjects do not appear visibly intoxicated, even though they are intoxicated according to law in regard to motor vehicle operation [3–6]. When most people appear obviously intoxicated, their BAC is probably well in excess of any legal definition of intoxication.

Regardless of which functions are affected by alcohol, impaired drivers clearly present a public health risk because of the increased number of accidental injuries due to intoxication. About 16 000 fatalities occur each year related to drunken driving [7] and about 10% of all personal injury accidents and at least 180 000–200 000 property and personal injury crashes, respectively, are caused by alcohol intoxication per year [8]. The risk of injury as well as the responsibility for causing a collision when driving while intoxicated is proportional to the BAC. With the current legal definition of driving while intoxicated in the United States (80 mg/dl), the relative risk for a crash is significantly greater compared with sober drivers. When the interaction among blood alcohol, gender, and single *versus* multiple vehicle collisions is considered, the relative risk is many times greater than previously believed. For example, the relative risk of a fatal crash in an 18-year-old male with a BAC of 80 mg/dl is about 34 times greater than a sober driver. In comparison, a 40-year-old male with the same BAC has a relative risk that is about 11 times greater than a sober control [9, 10]. A more specific breakdown of relative risk based on age, gender, and BAC is presented in Figure 1.

Pedestrian and Fall-down Injuries

Earlier studies estimated that about one-third of all fatally injured pedestrians had a BAC of 100 mg/dl or more at the time of their death. More than two-thirds of drivers, pedestrians, and bicyclists (see below) killed each year are intoxicated [12]. Driving and pedestrian activity rely on divided attention and visual motor processes. Therefore, it is reasonable to infer that these behaviors share similar alcohol-induced changes in relative risk. However, driving is obviously a more challenging task than walking, but both behaviors require divided attention, vigilance, and other cognitive skills that are sensitive to the impairment produced by alcohol, even at low BACs [13].

In the United States, injuries related to falls are the second leading cause of accidents and account for about 13 000 deaths per year. Most studies suggest



Age 16–20 females*	1.8	2.5	3.3	4.5	6	8	11	15	20	27	37	49	67	90	122	164	221	299	403	Comatose	~LD:50	
Age 16–20 males	2.4	3.7	5.8	9.0	14	22	34	53	82	127	196	305	473	735	1141	1772	2752	4273	6634			
Age 21–34 females/males	1.8	2.4	3.2	4.3	6	8	10	14	18	24	33	43	58	78	104	138	185	247	330			
Age 35+ females/males	1.7	2.3	3.0	3.9	5	7	9	11	15	20	26	33	44	57	75	99	129	169	221			

Figure 1 Alcohol intoxication, behavior, and relative risk for a fatal crash. Line graph shows relative risk (log) as a function of BAC, age, and gender. Data derived from stepwise logistic regression coefficients for relative risk based on single vehicle fatalities [9]. Note that for women age 16–20, a coefficient of 0.03 (range 0.044–0.014) was used. Coefficients are rounded to nearest 10th or whole number [Reproduced from Ref. 11. © Taylor & Francis Group, 2008.]

that alcohol increases the risk for injuries due to falls. Honkanen *et al.* [14] evaluated intoxicated emergency room patients involved in fall-down injuries and compared them with sober pedestrians who were at the same location of the accident one week later at the same time of the day. The comparison revealed the relative risk or a fall was three times greater for patients with blood–alcohol levels between 50 and 100 mg/dl, 10 times greater for patients with blood alcohol levels between 100 and 150 mg/dl, and 60 times greater for patients with BACs 160 mg/dl, or higher.

Bicycling

Also, there are about 200 fatalities and 7000 injuries from alcohol-related bicycle crashes each year. The relative risk of an alcohol-related bicycle crash found that alcohol was involved in 25% of the collision accidents and in 63% of the single accidents involving cyclists aged 15–64 years compared with intoxication in 4% of the nonaccident controls. A relative risk was of the order of 3 overall, and 58 for the collisions related to alcohol use [15]. Moreover, bicyclists who died at the scene were four times as likely as those who died at hospitals to be at or above the legal definition of intoxicated driving. This increased mortality may be due to the effects of alcohol on injury outcome (discussed later in this article). Other studies suggest that fatally injured bicyclists were about twice as likely to be intoxicated as cyclists treated for nonfatal injuries [16].

Fires and Burns

In the United States, alcohol intoxication plays a role in the estimated 5000 fatalities and about 1.4 million burn injuries each year [17]. In a review of studies on alcohol and burn injuries published between 1947 and 1986, Howland and Hingson [18] and reported on the percentage of the victims who were intoxicated with alcohol. In the overwhelming majority of the studies published in that period, alcohol exposure was found to be more likely among those who died in fires ignited by cigarettes than from other causes, suggesting that alcohol plays a role in the cause of fires, subsequent burn injuries, and is overly represented in burn victims. In fact, one-third to two-thirds of these victims had blood alcohol levels greater than 100 mg/dl. The authors concluded from these data that alcohol intoxication is a risk factor for fire deaths [18]. Later studies further revealed that alcohol was a factor in about 22% [19] to 26% [20] of burn injuries. Overall, intoxicated patients have a significantly higher fatality rate in severe burn cases. These data are reviewed elsewhere [21].

Although, impairment from alcohol is a risk factor in a substantial percentage of burn victims, it is not the only factor and may also interact with other factors. For example, Brezel and Kassenbrock [22] examined alcohol and drug abusers, psychiatric patients, and those with neurological dysfunction to determine whether this group had more medical complications, surgical procedures, and longer hospital stays than burn patients without these disorders. Alcohol abuse (defined as six or more cans of beer or the equivalent, per day) was the most common form of impairment, and the authors found that these patients had more complications and required a longer period of hospitalization, alcohol intoxication was only one of the several contributory factors.

Both acute and chronic alcohol abuse contribute to burn injuries. In a study of 1074 patients admitted to a medical center burn unit, 40% who were positive for alcohol were more likely to have a greater proportion of bodily burns and greater incidence of smoke inhalation than the controls [23]. Chronic alcoholics also seem to have a higher fatality rate than do patients without a history of chronic alcohol abuse [24].

In a recent review of the incidence and toxicological complications of burn cases, Brick [25] concluded that there is a clear relationship between alcohol or drug intoxication and the risk for thermal injury. It was suggested that the reason alcohol-intoxicated persons were at increased risk for accidental thermal injuries is impaired judgment or psychomotor coordination while engaging in normal fire-starting activities (e.g., cooking). We also noted that psychomotor impairment might be only one part of the problem. Other factors including neuropsychological status while intoxicated may impair various domains of cognitive functioning resulting in a decreased ability to anticipate problems, lowered inhibitions, and increased risk taking. For example, once a fire has started, mental confusion, and failure to recognize risk or danger may lead to an inability to anticipate or respond to danger, particularly at high levels of intoxication.

Water Sports

Alcohol is infinitely miscible in water, at least chemically, but with regard to water sports, alcohol and water do not mix. Nearly half of drowning victims in one study had consumed alcohol and 22% were intoxicated with BACs of 100 mg/dl, or more [26]. More recent studies suggest that alcohol consumption significantly increases the risk for boating fatalities was a factor in up to 21% of reported boating [27]. As alcohol deleteriously impairs balance, motor function, and judgment, intoxicated passengers, as well as vessel operators, are probably at risk for injury [28].

Alcohol intoxication also contributes to and aggravates spinal cord injuries following diving accidents. In this context, Perrine *et al.* [29] examined the effects of alcohol on the ability to perform shallow-water entry dives under experimental conditions. A progressive and significant impairment of specific aspects of diving performance was detected at blood alcohol levels as low as 40 mg/dl. This study also correlated diving performance with psychomotor performance using the Standardized Field Sobriety Tests (SFSTs). Impaired diving correlated well with impaired SFST performance criteria for the detection of drivers with a blood alcohol level of more than 100 mg/dl [29].

Aircraft Operation

Pilots who must divide attention and process information derived from an array of instrumentation and make perceptual and cognitive decisions based on a large amount of information in a multidimensional environment should not drink and fly. It is known that alcohol deleteriously influences the ability of pilots to evaluate their performance [30] and that low levels of alcohol (25–40 mg/dl) impair performance of trained pilots in flight simulators [31, 32]. Although it might appear that both motor vehicle drivers and aircraft pilots are impaired at similar low BACs (e.g., 30–40 mg/dl), there is evidence to suggest that piloting an aircraft is significantly more sensitive to the effects of alcohol. For example, there is some research suggesting that alcohol continues to impair performance on flight simulators many hours after blood alcohol levels have returned to zero [33]. Although some laboratory studies failed to find impairment 12 h after drinking [34], the forensic implications of impairment when alcohol is not in the blood are enormous, need to be evaluated further.

Suicide

Suicide is the 11th leading cause of death for young people (aged 15–34), and the third leading cause of death in the United States [35]. Suicide is also highly correlated with alcohol intoxication.

In a review of suicide attempts, intoxication was present in 40% (range 10–73%) for attempted suicide and 10–69% for completed suicides. The relationship between alcohol use and suicide is not difficult to understand. Acute intoxication reduces inhibitions, narrows attention, impairs the ability to appreciate the consequences of behavior, and may promote depressive thoughts and hopelessness. Chronic alcohol abuse is often complicated by mental illness, including depression [36–38]. Both acute and chronic intoxication, impair cognitive skill, may enhance aggression including self-aggression, or the combination with medications may precipitate pathways mediating this behavior [39]. Although, it is unlikely that people commit or attempt to commit suicide simply because they are intoxicated, premorbid suicidal ideology is more likely to be acted upon while intoxicated. Acute intoxication may be a greater risk factor for suicide than the previous drinking history [40] but the causal mechanism for the interrelationship among alcohol intoxication, alcohol dependence, and suicide is only partially understood [37, 41].

Miscellaneous Injuries

People die from many causes including accidental injuries, homicide, and suicide, in which alcohol intoxication is a significant risk factor. Relative to lifetime abstainers and infrequent drinkers, the risk of death from external causes increased logarithmically among infrequent binge drinkers [42] and no evidence of reduced risk of death among light or moderate drinkers. The group at highest risk of death from external causes was drinkers who consumed five or more drinks less than once a month. Within this group of binge drinkers, older subjects (defined as 65-plus years) were at the highest risk. High risk was also observed in younger drinkers (defined as 18–24 years of age), probably because of their lack of tolerance and experience, as middle-aged drinkers (25–64 years of age) who presumably have more experience both as drinkers and drivers, did not show the same increased mortality risk. In summary, infrequent binge drinking increases risk as a function of

age, possible tolerance, age-related experience, and other variables.

Intoxication and Injury Outcome

Alcohol intoxication not only contributes significantly to accidental injuries but may also affect injury outcome, particularly among patients with head injuries. This is highly significant as up to half of the traumatic brain-injured patients have BACs that exceed the legal definition for intoxicated driving at the time of injury [43]. For example, motorcycle riders with head injuries are about twice as likely to have fatal head injuries if they are intoxicated, compared to injury matched, sober riders [44]. Drunk drivers are more likely to be seriously or fatally injured than sober drivers [45] and alcohol-intoxicated accident victims with central nervous system injuries were more than twice as likely to die sooner than anatomically matched controls [46].

Such a belief is not supported by the majority of more current research on this topic [47]. The biochemical cause of the deleterious effect of intoxication in the injury outcome is not proven but there are some intriguing potential mechanisms. For example, it is now believed that the severity of hemorrhagic shock is greater when intoxicated and results in a higher mortality rate compared with the sober controls [48]. Hemorrhagic shock also induces acidosis with marked hypercarbia. In such cases, alcohol-induced acidosis would likely increase morbidity and mortality [48–50] possibly because of the effects of acidosis on ventilatory responses.

Alcohol and the Skeletal System

The relationship between alcohol abuse and increased risks for skeletal fractures was observed by the ancient Egyptians [51–53], but the epidemiology and mechanisms of this effect have only recently come under scientific scrutiny. Since alcohol is known to increase risk for injuries that may involve skeletal fractures, generalized skeletal fragility among alcoholics may be an important contributing factor to such injuries.

Alcohol-induced Fractures

Epidemiological research on the prevalence of fractures in alcoholic subjects suggests positive association between alcohol intake and fractures. For example, men hospitalized for alcohol-related problems are four times more likely to have rib fractures than nondrinking patients [54] and up to 14 times more likely to have spinal crush fractures [55, 56]. Two to six drinks per week also increases the risk for fractures in men compared with the same injuries in subjects who consumed less than two drinks per week, and for heavy drinkers, there was almost 10 times the risk of hip fractures as men in the same age group who drank lightly [57].

Weekly alcohol intake was associated with greater risks for osteoporotic fractures in postmenopausal women [58], and women who consumed more than eight drinks per week, the fractures were almost twice as likely as in nondrinkers. Other studies show that the equivalent of two standard drinks per day is associated with a significant increase in hip, wrist, and other fractures [59–61]. However, other investigators have not identified any significant association between alcohol intake and the risks for various fractures in women [62–66].

In addition to the risk of falls and related injuries, alcoholics may also suffer from a generalized skeletal fragility. Low bone density (osteoporosis) is a predictor of fractures [7], and a consequence of excess alcohol use reduced bone density has been confirmed in by many, but not all studies [67–73]. Dietary [68], hormonal [69], metabolic [73], and other factors contribute to this phenomenon (see [11] for a review).

Alcohol-induced Liver Injury

Underreporting of alcohol consumption makes the exact prevalence of alcoholic liver disease in the United States difficult to measure, but health statistics suggest that some form of alcoholic liver disease affects more than 2 million drinkers [74]. It is estimated that 900 000 Americans have cirrhosis, and of the 26 000 who die each year, 40–90% have a history of alcohol abuse [74]. It is clear that the development of alcoholic liver disease is due to a combination of factors, most notably, prolonged alcohol consumption. Alcohol abuse is the leading cause of liver-related mortality in the United States. Excessive alcohol consumption leads to three serious types of liver injuries: fatty liver, hepatic inflammation (alcoholic hepattits), and progressive liver scarring (fibrosis or cirrhosis). Chronic heavy drinking can alter the normal metabolism and lead to an accumulation of fat

in the liver. As a result, the liver cells become so infiltrated that the liver itself becomes enlarged and cell damage may occur. However, fatty liver is reversible with abstinence. Continued alcohol abuse may result in hepatitis, a more serious medical condition, characterized by prolific inflammation and tissue damage.

Hepatitis is life threatening but there can be significant recovery of function following abstinence. The most serious form of liver damage is cirrhosis, an irreversible disease characterized by scarring and cell death. Impaired liver functioning can also cause primary hepatic encephalopathy. Although this brain disorder precipitated by liver disease is rare, forensic examiners should be aware of it because it is characterized by altered psychomotor, intellectual, and behavioral functioning absent acute intoxication.

Chronic, heavy drinking may produce metabolic tolerance and unusually high rates of alcohol elimination. Hepatitis and fibrosis will ultimately impair liver function and produce a reverse metabolic tolerance and impaired oxidation of alcohol. When justified, pharmacokinetic analyses used in making estimates of alcohol consumption or intoxication should account for potential metabolic changes.

One question of interest to many people is, "How much alcohol does one need to drink before liver damage occurs"? Epidemiological studies suggest that reliable signs of injury begin after a "threshold" dose of about 600 kg for men and between 150 and 300 kg for women. To place this in perspective, at the high end (for men), this is roughly equivalent to the average consumption of 10–12 drinks a day for 10 years, and at the low end (for women), about three drinks per day. Below these doses, it is difficult but not impossible to reliably detect liver injury [75–78]. The differences in threshold doses between genders cannot be accounted for by anthropometrics or pharmacokinetics. For example, many individuals who consume large amounts of alcohol never develop liver disease and less than one-half of heavy drinkers develop alcoholic hepatitis or liver fibrosis [75]. This suggests that hereditary and/or environmental factors interact with alcohol to affect the natural history of liver injury [76]. Numerous possible mechanisms may affect the susceptibility of certain people to alcohol-induced liver damage. Nevertheless, alcoholic liver disease is a significant cause of death in the United States and significantly contributes overall ages to 24 (in men) to 28 (in women) years of potential life lost [79, 80].

Summary and Conclusions

As a pharmacological agent, alcohol is a relatively simple compound. The ubiquitous nature of this drug on most, if not all major organ systems is consistent with its simple molecular structure and its widespread use [11]. Of specific interest to forensic issues, alcohol directly affects hepatic and skeletal systems and indirectly affects health and well being when accidental injuries due to intoxication are considered. From the available alcohol research, several conclusions may be drawn regarding the medical consequences of alcohol use. Most notably and across physiological systems, the effects of alcohol are complex and vary as a function of gender, diet, environment, lifestyle, genetics, dose and frequency of alcohol use, use of other drugs, and age. Even so, the majority of studies suggest that, overall, higher doses of alcohol are deleterious to many physiological systems and precipitate a range of psychosocial and biobehavioral problems including shortened lifespan that would be of interest in forensic evaluations.

References

[1] Cherpitel, C.J. (1993). Alcohol, injury, and risk-taking behavior: data from a national sample, *Alcoholism, Clinical and Experimental Research* **7**(4), 762–766.

[2] Centers for Disease Control (1991). *Morbidity and Mortality Weekly Report*, June 21.

[3] Brick, J. & Carpenter, J.A. (2001). The identification of alcohol intoxication by police, *Alcoholism, Clinical and Experimental Research* **25**(6), 850–855.

[4] Wells, J., Greene, M., Foss, R., Ferguson, S. & Williams, A. (1997). Drinking drivers missed at sobriety checkpoints, *Journal of Studies on Alcohol* **58**, 513–517.

[5] Brick, J., Adler, J., Cocco, K. & Wesrick, E. (1992). Alcohol intoxication: pharmacokinetic prediction and behavior in humans, *Current Topics in Pharmacology* **1**, 57–67.

[6] Langenbucher, J. & Nathan, P. (1983). Psychology, public policy and the evidence for intoxication, *American Psychologist* **38**, 1070–1077.

[7] National Institute on Alcohol Abuse and Alcoholism (2000). *Tenth Special Report to the US Congress on Alcohol and Health*, US Department of Health and Human Services, Washington, D.C.

[8] Wieczorek, W.F. (1995). The role of treatment in reducing alcohol-related accidents involving DWI offenders, in *Alcohol, Cocaine and Accidents, Drug Alcohol Abuse Rev*, R.R. Watson, ed, Humana Press, Totowa, N.J., USA Vol. 7, pp. 105–129.

[9] Zador, P.L., Krawchuk, S.A. & Voas, R.B. (2000). Alcohol-related risk of driver fatalities and driver

involvement in fatal crashes in relation to driver age and gender: an update using 1996 data, *Journal of Studies on Alcohol* **61**, 387–395.

[10] Zador, P.L., Krawchuk, S.A. & Voas, R.B. (2000). *Relative Risk of Fatal Crash Involvement by BAC, Age and Gender*, DOT HS 809 050, US Department of Transportation National Highway Traffic Safety Administration.

[11] Brick, J. (2008). Medical consequences of alcohol abuse, in *Handbook of the Medical Consequences of Alcohol and Drug Abuse*, The Haworth Medical Press, Binghamton.

[12] National Highway Traffic Safety Administration (1994). Traffic safety facts 1993, in *Alcohol*, US Department of Transportation National Center for Statistics and Analysis, Washington, D.C.

[13] National Institute on Alcohol Abuse and Alcoholism; US Department of Health and Human Services (1988). *Effects of Alcohol on Driving Performance, Effects of Low Doses of Alcohol on Driving-related Skills: A Review of the Evidence*, DOT HS 807 280, US Department of Transportation/National Highway Traffic Safety Administration.

[14] Honkanen, R., Ertoma, L., Kuosmanen, P., Linnoina, M., Alah, A. & Visori, T. (1983). The role of alcohol in accident falls, *Journal of Studies on Alcohol* **44**, 231–245.

[15] Olkkonen, S. & Honkanen, R. (1994). The role of alcohol in non-fatal bicycle injuries, *Accident; Analysis and Prevention* **22**, 89–96.

[16] Li, G., Baker, S.P., Sterling, S., Smialek, J.E., Dischinger, P.C. & Soderstron, C. (1996). A comparative analysis of alcohol in fatal and non-fatal bicycling injuries, *Alcoholism, Clinical and Experimental Research* **20**, 1553–1559.

[17] Baker, S.P., O'Neill, B. & Karpf, R. (1992). *The Injury Fact Book Second Edition*, Oxford University Press, New York.

[18] Howland, J. & Hingson, R. (1987). Alcohol as a risk factor for injuries or death due to fires and burns: review of the literature, *Public Health Reports* **102**, 475–483.

[19] Cherpitel, C.J. (1989). Breath analysis and self reports as measures of alcohol-related emergency room admissions, *Journal of Studies on Alcohol* **50**(2), 155–161.

[20] Jones, J.D., Barber, B., Engrav, L. & Heimbach, D. (1991). Alcohol use and burn injury, *The Journal of Burn Care and Rehabilitation* **12**(2), 148–152.

[21] National Institute on Alcohol Abuse and Alcoholism (1997). *Ninth Special Report to the US Congress on Alcohol and Health*, NIH Publication No 97-4017, National Institute on Alcohol Abuse and Alcoholism, Bethesda.

[22] Brezel, B.S. & Kassenbrock, M. (1988). Burns in substance abusers and in neurologically and mentally impaired patients, *The Journal of Burn Care and Rehabilitation* **9**(2), 169–191.

[23] McGill, V., Kowal-Vern, A., Kahn, S. & Gamelli, R.L. (1995). The impact of substance use on mortality and morbidity from thermal injury, *The Journal of Trauma* **38**(6), 931–934.

[24] Haum, A., Perbix, W., Hack, H.J., Stark, G.B., Spilker, G. & Doehn, M. (1995). Alcohol and drug abuse in burn injuries, *Burns* **21**(3), 194–199.

[25] Brick, J. (2006a). Interaction between toxicology and burn victim physiology, in *Burns: The Medical and Forensic Model*, A. Clark, ed, Lawyers & Judges Publishing Company, pp. 221–256.

[26] Hoxie, P., Cardosi, K., Stearns, M. & Mengert, P. (1988). *Alcohol in Fatal Recreational Boating Accidents*, Pub No DOT CGD 0488, US Department of Transportation US Coast Guard, Washington, D.C.

[27] Molberg, P., Hopkins, R., Paulson, J. & Gunn, R. (1993). Fatal incident risk factors in recreational boating in Ohio, *Public Health Repots* **108**(3), 340–346.

[28] Hingson, R. & Howland, J. (1993). Alcohol and nontraffic unintended injuries, *Addiction* **88**(7), 877–883.

[29] Perrine, M.W., Mundt, J.C. & Winer, R.I. (1994). When alcohol and water don't mix: diving under the influence, *Journal of Studies on Alcohol* **55**, 517–524.

[30] Morrow, D., Leirer, V., Yesavage, J. & Tinklenberg, J. (1991). Alcohol, age, and piloting: Judgment, mood, and actual performance, *The International Journal of the Addictions* **26**(6), 669–683.

[31] Billings, C., Demosthesen, T., White, T. & O'Hara, D. (1991). Effects of alcohol on pilot performance in simulated flight, *Aviation, Space and Environmental Medicine* **62**(3), 2323–2335.

[32] Ross, L., Yeazel, L. & Chau, A. (1992). Pilot performance with blood alcohol concentrations below 0.04 percent, *Aviation, Space and Environmental Medicine* **63**(11), 951–956.

[33] Yesavage, J., Dolhert, N. & Taylor, J. (1994). Flight simulator performance of younger and older aircraft pilots. Effects of age and alcohol, *Journal of the American Geriatric Society* **42**(6), 577–582.

[34] Lemon, J., Chesher, G., Fox, A., Greeley, J. & Nabke, C. (1993). Investigation of the "Hangover" effects of an acute dose of alcohol on psychomotor performance, *Alcoholism, Clinical and Experimental Research* **17**(3), 665–668.

[35] US Department of Health and Human Services Washington DC (2003). *National Center for Health Statistics National Vital Statistics Records*, Vol. 52(3), available at http://www.cdc.gov/nchs/data/nvsr/nvsr54_03.pdf..

[36] Cheng, A.T.A. (1995). Mental illness and suicide, *British Journal of Psychiatry* **170**, 441–446.

[37] Conner, K.R., Beautrais, A.L. & Conwell, Y. (2003). Risk factors for suicide and medically serious suicide attempts among alcoholics: analyses of canterbury suicide project data, *Journal of Studies on Alcohol* **64**, 551–554.

[38] Wallen, M., Lorman, W. (2008). Special issues in patients with comorbid psychiatric and chemical dependency disorders, in *Handbook of the Medical Consequences of Alcohol and Drug Abuse*, J. Brick, ed, The Haworth Medical Press, Binghamton.

[39] Cherpitel, C.J., Borges, G.L.G. & Wilcox, H.C. (2004). Acute alcohol use and suicidal behavior: a review of the literature, *Alcoholism, Clinical and Experimental Research* **28**(s1), 18S–28S.

[40] Borges, G. & Rosovsky, H. (1996). Suicide attempts and alcohol consumption in an emergency room sample, *Journal of Studies on Alcohol* **57**, 543–548.

[41] Hufford, M.R. (2001). Alcohol and suicidal behavior, *Clinical Psychology Review* **21**, 797–811.

[42] Dawson, D. (2001). Alcohol and mortality from external cues, *Journal of Studies on Alcohol* **62**, 790–797.

[43] Zink, B.J., Schultz, C.H., Stern, S.A., Mertz, M., Wang, X., Johnston, P. & Keep, R.F. (2001). Effects of ethanol and naltrexone in a model of traumatic brain injury with hemorrhagic shock, *Alcoholism: Clinical and Experimental Research* **25**(6), 916–923.

[44] Luna, G.K., Maier, R.V., Sowder, L., Copass, M.K. & Oreskovich, M.R. (1984). The influence of ethanol intoxication on outcome of injured motorcyclists, *The Journal of Trauma* **24**(8), 695–700.

[45] Waller, P., Steward, J., Hansen, A., Stutts, J., Popkin, C. & Rodgman, E. (1986). The potentiating effects of alcohol on driver injury, *The Journal of the American Medical Association* **256**(11), 1461–1466.

[46] Zink, B., Maoi, R. & Chen, B. (1996). Alcohol, central nervous system injury, and time to death in fatal motor vehicle crashes, *Alcoholism Clinical and Experimental Research* **20**(9), 1518–1522.

[47] Fell, J. & Hertz, E. (1993). The effects of blood alcohol concentration on time of death for fatal crash victims, *Alcohol Drugs and Driving* **9**(2), 97–106.

[48] Molina, P.E., McClain, C., Valla, D., Guidot, D., Diehl, A.M., Lang, C.H. & Neuman, M. (2002). Molecular pathology and clinical aspects of alcohol-induced tissue injury, *Alcoholism Clinical and Experimental Research* **26**(1), 120–128.

[49] Chen, R.J., Fang, J.F., Lin, B.C., Hsu, Y.P., Kao, J.L. & Chen, M.F. (2000). Factors determining operative mortality of grade V blunt hepatic trauma, *The Journal of Trauma* **49**, 886–891.

[50] Kinkaid, E.H., Miller, P.R., Meredith, J.W., Rahman, N. & Change, M.C. (1998). Elevated arterial base deficit in trauma patients: a marker of impaired oxygen utilization, *Journal of the American College of Surgeons* **187**, 384–392.

[51] Conn, H.O. (1985). Natural history of complications of alcoholic liver disease, *Acta Medica Scandinavica* **703**, 127–134.

[52] Mathew, V.M. (1992). Alcoholism in biblical prophecy, *Alcohol and Alcoholism* **27**, 89–90.

[53] Seller, S.C. (1985). Alcohol abuse in the Old Testament, *Alcohol* **20**, 69–76.

[54] Lindsell, D.R., Wilson, A.G. & Maxwell, J.D. (1982). Fractures on the chest radiograph in detection of alcoholic liver disease, *BMJ (Clinical Research Ed.)* **285**, 597–599.

[55] Crilly, R.G., Anderson, C., Hogan, D. & Delaquerriére-Richardson, L. (1988). Bone istomorphometry, bone mass, and related parameters in alcoholic males, *Calcified Tissue International* **43**, 269–276.

[56] Israel, Y., Orrego, H., Holt, S., Macdonald, D.W. & Meema, H.E. (1980). Identification of alcohol abuse: thoracic fractures on routine chest x-rays as indicators of alcoholism, *Alcoholism* **4**, 420–422.

[57] Felson, D.T., Kiel, D.P., Anderson, J.J. & Kannel, W.B. (1988). Alcohol consumption and hip fractures: the framingham study, *American Journal of Epidemiology* **128**, 1102–1110.

[58] Paganini-Hill, A., Ross, R.K. & Gerkins, V.R. (1981). Menopausal estrogen therapy and hip fractures, *Annals of International Medicine* **95**, 28–31.

[59] Tuppurainen, M., Kroger, H., Honkanen, R., Puntial, E., Huopia, J., Saarikoski, S. & Alhave, E. (1995). Risks of perimenopausal fractures: a prospective population-based study, *Acta Obstetricia Gynecologica Scandinavica* **74**, 624–628.

[60] Hernandez-Avila, M., Colditz, G.A., Stampfer, M.J., Rosner, B., Speizer, F.E. & Willett, W.C. (1991). Caffeine, moderate alcohol intake, and risk of fractures of the hip and forearm in middle-aged women, *The American Jounal of Clinical Nutrition* **54**, 157–163.

[61] Fujiwara, S., Kasagi, F., Yamada, M. & Kodama, K. (1997). Risk factors for hip fracture in a Japanese cohort, *Journal of Bone and Mineral Research : The Official Journal of the American Society for Bone and Mineral Research* **12**, 998–1004.

[62] Cumming, R.G. & Klineberg, R.J. (1994). Case-control study of risk factors for hip fractures in the elderly, *American Journal of Epidemiology* **139**, 493–503.

[63] Diaz, M.N., O'Neill, T.W. & Silman, A.J. (1997). The influence of alcohol consumption on the risk of vertebral deformity, *Osteoporos International* **7**, 65–71.

[64] Huang, Y.S., Chan, C.Y., Wu, J.C., Pai, C.H., Chao, Y. & Lee, S.D. (1996). Serum levels of interleukin-8 in alcoholic liver disease: relationship with disease stage, biochemical parameters, and survival, *Journal of Hepatology* **24**(4), 377–384.

[65] Johnell, O., Gullberg, B., Kanis, J., Allander, E., Elffors, L., Dequeker, J., Dilsen, G., Gennari, C., Lopes, V., Lyritis, G., Mazzuoli, G. Miravet, L., Passeri, M., Perez, C., Rapado A. & Robot C. (1995). Risk factors for hip fracture in European women: the MEDOS study, *Journal of Bone and Mineral Research* **10**, 1802–1815.

[66] O'Neill, T.W., Marsden, D., Adams, J.E. & Silman, A.J. (1996). Risk factors, falls, and fracture of the distal forearm in Manchester, UK, *Journal of Epidemiology and Community Health* **50**, 288–292.

[67] Holbrook, T.L. & Barrett-Connor, E. (1993). A prospective study of alcohol consumption and bone mineral density, *British Medical Journal* **306**, 1506–1509.

[68] Lairinen, K., Karkkainen, M., Lalla, M., Lambergallardt, C., Tunninen, R., Tahtela, R. & Valimaki, M. (1993). Is alcohol an osteoporosis-inducing agent for young and middle-aged women? *Metabolism: Clinical and Experimental* **42**(7), 875–881.

[69] Lairinen, K., Valimaki, M. & Keto, P. (1991). Bone mineral density measured by dual-energy X-ray absorptiometry in healthy Finnish women, *Calcified Tissue International* **48**, 224–231.

[70] Orwoll, E.S., Bauer, D.C., Vogt, T.M. & Fox, K.M. (1996). Axial bone mass in older women: study of osteoporotic fracture research group, *Annals of Internal Medicine* **124**, 187–196.

[71] Peris, P., Guanabens, N., Parés, A., Pons, F., Del Rio, L., Monegal, A., Suris, X., Caballeria, J., Rodes, J. & Munoz-Gómez, J. (1995). Vertebral fractures and osteopenia in chronic alcoholic patients, *Calcified Tissue International* **57**, 111–114.

[72] Blaauw, R., Albertse, E.C., Beneke, T., Lombard, C.J., Laubscher, R. & Hough, F.S. (1994). Risk factors for the development of osteoporosis in a South African population: a prospective analysis, *South African Medical Journal. Suid-Afrikaanse Tydskrif Vir Geneeskunde* **84**, 328–332.

[73] Gonzalez-Calvin, J.L., Garcia-Sanchez, A., Bellot, V., Munoz-Torres, M., Raya-Alvarez, E. & Salvatierra-Rios, D. (1993). Mineral metabolism, osteoblastic function, and bone mass in chronic alcoholism, *Alcohol and Alcoholism* **28**, 571–579.

[74] Dufour, M.C., Stinson, F.S. & Caces, M.F. (1993). Trends in cirrhosis morbidity and mortality: United States 1979–1988, *Seminars in Liver Disease* **13**(2), 109–125.

[75] Lelbach, W.K. (1975). Cirrhosis in the alcoholic and its relation to the volume of alcohol abuse, *Annals of the New York Academy of Sciences* **252**, 85–105.

[76] Marbet, U.A., Bianchi, L., Meury, U. & Stalder, G.A. (1987). Long-term histological evaluation of the natural history and prognostic factors of alcoholic liver disease, *Journal of Hepatology* **4**(3), 364–372.

[77] Mezey, E., Kolman, C.I., Diehl, A.M., Mitchell, M.C. & Herlong, H.F. (1988). Alcohol and dietary intake in the development of chronic pancreatitis and liver disease in alcoholism, *The American Journal of Clinical Nutrition* **48**(1), 148–151.

[78] Tuyns, A. & Pequignot, G. (1984). Greater risk of ascitic cirrhosis in females in relation to alcohol consumption, *International Journal of Epidemiology* **13**(1), 53–57.

[79] Center for Disease Control (2001–2005). *Years of Potential Life Lost Report, Average for United States*, available at http:apps.nccd.cdc.gov (last accessed 3/18/08).

[80] Center for Disease Control (2001–2005). *Alcohol-attributed Death Report, Average for United States*, available at http:apps.nccd.cdc.gov (last accessed 3/18/08).

JOHN BRICK

Alcohol: Elimination *see* Alcohol: Analysis

Alcohol: Interaction with Other Drugs

Introduction

The interaction between alcohol (ethanol, unless otherwise specified) and therapeutic medications is often encountered in accident, homicide, and suicide investigations. This review will focus on alcohol–other drug interactions most applicable to forensic evaluations. Reviews of the interaction between alcohol and a broader range of medications on health are available elsewhere [1, 2]. The widespread use of alcohol is coupled with extensive medical consequences and overly represented in accidental or other injuries and criminal actions. For a variety of reasons, including the ability to routinely test for the presence of medications and illicit drugs, the use of alcohol with drugs in accidents and crimes is of increasing forensic interest.

The term *drug* is used here to refer to both licit and illicit drugs, whereas the term *medication* will refer specifically to therapeutic medications and the term *alcohol* refers to *ethanol*, unless otherwise indicated. In the controlled environment of the laboratory the biobehavioral effects of drug interactions are often complex. Outside of the laboratory alcohol–drug interactions are often complicated by real-world variations in dose and drug potency, duration and frequency of use, use of other drugs, and individual characteristics of the user, including physiological factors such as tolerance, metabolic state, diseases, and anthropometrics.

Understanding Alcohol

Alcohol is a central nervous system (CNS) depressant, although under some conditions, it is perceived as a stimulant because it increases locomotor activity, loquaciousness, and other behaviors. The biphasic effect of alcohol is probably more related to the

decrease in inhibitions produced by this drug rather than actual stimulant effects, at least in humans. Pharmacologically, alcohol acts much more like a CNS depressant or anxiolytic [3, 4]. Suffice it to say, the pharmacology of alcohol is complex and its biobehavioral effects quite broad, since alcohol is capable of altering receptors, ion transport, cell membranes, most cellular mechanisms critical to neurophysiology, and ultimately a range of behaviors. Alcohol also has the ability to alter the pharmacokinetics and pharmacodynamics of other drugs. Changes in bioavailability and efficacy of drugs in the presence of alcohol, or in patients with a history of alcohol dependence, have important implications for diagnoses, treatment, and outcome and should be part of a complete clinical evaluation.

Why Multiple Drug Use?

Persons trained in neuropharmacology, toxicology, or medicine are aware of the pernicious effects of some of the more commonly used drugs. The inherent dangers of combining drugs and public service announcements have saturated the media for decades. Yet multiple drug use is still relatively common. Although the biopsychosocial factors of drug use and drug interactions are complex and not fully identified, drug use is usually related to the reinforcing effects of the drug, psychosocial influences, and enhanced performance. Multiple drug use is may be motivated by many factors, including increasing the primary drug effect (e.g., increasing the intensity of the "high" or, when used clinically, increasing the effectiveness of treatment), self-medication to decrease the undesirable side effects of the primary drug (e.g., a depressant may be used to alleviate the edgy feeling after the desired effects of a stimulant wear off), or short supply of the primary or preferred drug (e.g., when the availability of the drug of choice is limited, another drug with similar properties will be substituted).

Heroin addicts, for example, may use or combine drugs for many reasons. Among these are sensation seeking (e.g., drug combinations without regard to safety and without any particular rationale other than the desire to become intoxicated), medical management (e.g., in the course of medical treatement, it is not uncommon to prescribe drugs that may interact with therapeutic or nontherapeutic drugs), or they may consume large quantities of alcohol or use other depressants to reduce or delay the opioid withdrawal syndrome until additional narcotics can be procured). Physicians must recognize the potential for these interactions (since in addition to a prescription medication, their patients may ingest other drugs, alcohol being the most common) and advise patients accordingly.

Pharmacological Basis of Alcohol–Drug Interactions

Drug interactions can produce alterations in physiology and ultimately behavior through changes in pharmacokinetics or changes in pharmacodynamics. Pharmacokinetic mechanisms account for drug interactions when the presence of one drug affects the bioavailability of another drug. Pharmacodynamic interactions account for drug interactions when drugs interact at the receptor level.

Pharmacokinetic Interactions

Pharmacokinetics can alter the bioavailability of the drug. If bioavailability increases or decreases the quantity of drug that interacts with receptors or other cellular components, the effect of the drug will in most instances, increase or decrease as well, depending on the dose. Pharmacokinetic interactions can occur through changes in drug absorption, distribution, metabolism, and excretion.

Pharmacodynamic Interactions

Pharmacodynamics is the study of the physiological and biochemical effects of drugs and their mechanism of action. Psychoactive drugs alter the functional activity of receptors or endogenous ligands (i.e., neurotransmitters and hormones) in the brain, whereas other drugs may exert similar effects outside the CNS (e.g., cardiac β-receptor antagonists). Changes in the pharmacodynamics or functional activity of neurons that modulate cognitive and psychomotor effects, for example, will alter physiology and behavior. Pharmacodynamic drug interactions can be classified as addictive, synergistic, potentiated, and antagonistic.

Additive Effects. Addictive effects occur when the combination of two drugs is equal to the sum of the effect of each drug (e.g., $2 + 2 = 4$). For example,

the CNS depressant effects of many benzodiazepines and alcohol are additive, as are the effects of many barbiturates with alcohol.

Synergistic Effects. Synergistic interactions occur when the combination of two drugs produces an effect that is greater than the sum of the effect of each drug (e.g., $2 + 2 = 6$). Synergistic interactions produce effects far greater than would be predicted from the sum of effect of either drug. For example, alcohol and carbon tetrachloride, a cleaning fluid, are toxic to the liver. However, the combination of the two produces much more liver damage than would be predicted from the sum of their individual effects. Similarly, in some cases, alcohol synergistically enhances the sedative effect of barbiturates and some effects of opiates.

Drug Potentiation. Potentiated drug effects are similar to synergistic effects, but usually describe an increase in the toxic effect of a drug when combined with a nontoxic drug (e.g., $0 + 1 = 2$). Histamine (H_2) antagonists such as cimetidine can be considered to potentiate the toxic effects of alcohol by increasing alcohol bioavailability.

Antagonism. An antagonist is a drug that blocks the effect of another drug (e.g., $2 + 2 = 1$ or $2 + 2 = 0$). Antagonists are very specific. For example, Naloxone has a much higher affinity for opiate receptors than heroin. Administering Naloxone to someone who has overdosed on heroin will produce a rapid reversal of the respiratory depression produced by heroin. Some drugs exhibit dispositional antagonism when the absorption, metabolism, distribution, or excretion of one drug is affected by another drug. For example, ethyl alcohol decreases the metabolism of methanol.

Specific Alcohol–Drug Interactions

With a basic understanding of the pharmacological mechanisms by which alcohol–drug interactions may occur, let us examine specific alcohol–drug interactions that may be of forensic interest.

Alcohol and Acetaminophen

The over-the-counter (OTC) medication acetaminophen (Tylenol) is one of the most commonly consumed medications in the United States because of its effective analgesic and antipyretic properties. Acetaminophen is metabolized by the CYP2E1 isozyme to a toxic intermediate, *N*-acetyl-*p*-benzoquinone imine (NAPQI). Normally, NAPQ1 is detoxified by the antioxidant glutathione [5], but chronic alcohol use reduces the amount of the glutathione produced in the liver. In other words, the combination of large amounts of acetaminophen reduced glutathione levels from chronic alcohol use increases the bioavailability, and the result is toxic. The resulting hepatotoxicity may progress to the point of fulminant hepatic failure and death. There is no agreement in the medical community as to the amount of alcohol consumed and/or the amount of acetaminophen necessary to cause this toxic effect to occur. Earlier studies indicated that liver damage can occur in alcoholics at therapeutic doses of acetaminophen [6–8], but a more recent study found no increase in liver toxicity among alcoholics given the maximal therapeutic dose (4 g/day) of acetaminophen [9]. They concluded that there was no clinical evidence of increased risk for these patients when acetaminophen is used within recommended doses.

Alcohol and Antibiotics

Most medications have some side effects, but when antibiotics used to treat infectious diseases are combined with acute alcohol consumption, some may cause nausea, vomiting, headache, and possibly convulsions. Among these antibiotics are furazolidone, griseofulvin, metronidazole, and the antimalarial quinacrine. Isoniazid and rifampin are used together to treat tuberculosis, a disease especially problematic among the elderly in nursing homes and among homeless alcoholics. In addition, acute alcohol consumption decreases the bioavailability of isoniazid in the bloodstream, whereas chronic alcohol use decreases the bioavailability of rifampin. The pharmacokinetic interaction between alcohol and some antibiotics may reduce their effectiveness [10]. Erythromycin accelerates gastric emptying and may reduce first-pass metabolism of alcohol in the stomach, resulting in increased absorption in the intestines and higher blood alcohol levels. Conversely, many aerobic bacteria in the colon are capable of metabolizing alcohol because they possess alcohol dehydrogenase (ADH) activity. In rats, treatment

with ciprofloxacin totally eliminated the colonic metabolism of alcohol, resulting in increased blood alcohol concentrations [11]. Although controversial, patients drinking alcohol while taking metronidazole or ketoconazole may suffer from symptoms similar to those found with disulfiram (Antabuse): abdominal distress, nausea, vomiting, and headache [12]. Similarly, cefoperazone (Cefobid), griseofulvin (Fulvicin, Grisactin), isoniazid (INH), metronidazole (Flagyl), nitrofurantoin (Furadantin, Macrodantin), and sulfamethoxazole (Bactrim, Septra) inhibit aldehyde dehydrogenase and may also produce a disulfiram-like reaction.

Alcohol and Anticoagulant Medications

Warfarin (Coumadin) reduces the ability of the blood to clot and is commonly used to treat patients with irregular heart rhythms, artificial heart valve, and following open-heart surgery. Warfarin is metabolized by the cytochrome P450 enzyme system in the liver, which is also part of the pathway of alcohol metabolism. Therefore, if a person consumes alcohol, the anticlotting effect of warfarin may be increased above the desired therapeutic effect. This increased bioavailability of warfarin is due to alcohol-related inhibition of warfarin metabolism by the cytochrome P450 enzyme system [13]. This increase could result in the emergence of potentially life-threatening hemorrhages. However, in people who drink alcohol regularly, and especially in some alcoholics, the chronic consumption of alcohol will result in induction of the cytochrome P450 enzyme system. The result of this will be an increased rate of metabolism of warfarin, thereby decreasing its bioavailability and interfering with its effectiveness in reducing blood clotting [13]. Such individuals will often need larger doses of warfarin to achieve the desired therapeutic effect, but there is no reported effect of warfarin on alcohol pharmacokinetics or pharmacodynamics.

Alcohol and Anticonvulsants

More currently available anticonvulsants in comparison with the classic agents have good absorption, linear kinetics, and minimal potential for interaction with other drugs. The newer anticonvulsants are eliminated through different combinations of liver metabolism and direct renal excretion, such as felbamate, gabapentin, lamotrigine, topiramate, tiagabine, levetiracetam, oxcarbazepine, and zonisamide. Depending on its chronicity of use, alcohol will have totally different pharmacokinetic effects with the older anticonvulsants such as phenobarbital, phenytoin, primidone, ethosuximide, carbamazepine, and valproate. Acute alcohol consumption increases the availability of phenytoin and the risk of drug-related side effects. Chronic drinking may decrease phenytoin bioavailability, significantly reducing the patient's protection against seizures, even during a period of abstinence [14, 15].

Alcohol and Antidepressants

The literature on the correlation between alcoholism and depression is well established [16] and particularly important in forensic evaluations because alcoholics are often involved in accidents and crimes. Many active alcoholics are prescribed antidepressants leading to a high potential for alcohol–antidepressant interactions. Several classes of antidepressants are available and are defined by how they affect brain neurochemistry. Some antidepressants cause varying degrees of sedating activity but these drugs should not be described as depressants or sedatives. Nevertheless, alcohol increases the sedative and other effects of tricyclic antidepressants such as amitriptyline (Elavil) [17]. In addition, alcohol-induced liver disease further impairs antidepressant metabolism and causes significantly increased levels of active medication in the body [18]. The disposition of imipramine in alcoholic and nonalcoholic patients with depression varied significantly. For example, oral imipramine clearance was more than 2 times greater in alcoholic patients than controls [19] and in recently detoxified men with alcohol dependence, and the elimination half-life for imipramine was more than doubled in alcoholics after intravenous infusion. Plasma concentrations of imipramine were also significantly lower in the alcoholics, suggesting that standard doses of some antidepressants may fail to produce adequate therapeutic changes in alcoholics. For imipramine, the doses may have to be doubled [19].

Popular serotonin reuptake inhibitors (SSRIs) such as fluoxetine (Prozac), sertraline (Zoloft), paroxetine (Paxil), and citalopram (Celexa) have the best safety profile of all antidepressants, and no serious interactions seem to occur when these agents are consumed

with moderate doses of alcohol [20]. In addition, neither fluoxetine nor alcohol alters the pharmacokinetics or psychomotor effects of the other, although alcohol impairs performance of most subjects on psychomotor tests [21, 22].

The use of the monoamine oxidase inhibitors (MAOIs) with alcohol will potentially precipitate a hypertensive crisis and enhance sedation. The mechanism for this reaction has been attributed to increased concentrations of the amino acid tyramine [23], which is a potent hypertensive agent present in many alcoholic beverages (e.g., wines) and foods (e.g., cheeses and bananas). Although most dietary tyramine is destroyed by MAO in the intestines and liver, tyramine will enter the circulation in patients treated with MAOIs and may produce hypertension and, rarely, sudden death [24]. Therefore, alcohol use should be avoided in patients prescribed MAOIs and considered by medical examiners in such cases.

Atypical antidepressants generally do not seem to have any problematic interactions with alcohol. However, mirtazapine (Remeron), when combined with alcohol, causes impaired cognition and decreased motor performance [25].

Alcohol and Antidiabetic Medications

Oral hypoglycemic agents are commonly prescribed for the treatment of diabetes mellitus in patients not requiring insulin. As previously noted, chlorpropamide (Diabinase), glyburide (Diabeta, Micronase), and tolbutamide (Orinase) inhibit aldehyde dehydrogenase and can cause disulfiram-like reactions following alcohol consumption. Metformin (Glucophage) may increase lactic acid levels in the blood following alcohol ingestion, which could result in acute lactic acidosis with potentially lethal results. Alcohol consumption by patients taking many of these medications could increase the risk of causing lower than normal blood sugar levels owing to impairment of gluconeogenesis during fasting when blood sugar is already low and the body depends on the production of new glucose [18]. Since low blood sugar can cause symptoms of impairment virtually indistinguishable from alcohol intoxication [26], behavioral observations of patients who are diabetic, or diabetics who have consumed alcohol, should be interpreted cautiously.

Alcohol and Antihistamines (H_1-Antagonists)

Antihistamines (H_1-antagonists), such as diphenhydramine, are commonly used to treat allergic disorders, and some antihistamines, such as hydroxyzine (Vistaril, Atarax), to treat anxiety. Many antihistamines cause drowsiness, which make them useful to treat insomnia, but are potentially dangerous because they can impair psychomotor and cognitive skills. Alcohol can substantially enhance the sedating effects of these agents and may further impair the ability to drive or operate other types of machinery [27]. Common sedative antihistamines, such as chlorpheniramine and diphenhydramine, significantly impair psychomotor performance and significantly increase the deleterious effects of alcohol on reaction time, coordination, and related psychophysical tests [28–30]. Newer antihistamines such as fexofenadine, loratadine, and cetirizine have been developed to minimize drowsiness and sedation while still providing effective therapeutic value. Nevertheless, these newer medications may still increase risk of hypotension and fall-down injuries among the elderly, particularly when combined with alcohol [18].

The effects of many of the antihistamines with alcohol have not been fully investigated, but it seems more probable than not that the combined use of these drugs will result in increased drowsiness and increased driving risks [31], possibly due to the observation that histamine receptor antagonism can affect alcohol metabolism and change the sensitivity to the hypnotic effects of alcohol [32].

Alcohol and Antipsychotics

Many antipsychotics produce sedation and psychomotor impairment. When combined with alcohol, they may increase sedation, impair coordination, and fatally depress respiration. This effect appears to be additive, but the mechanisms of this interaction are uncertain. Antipsychotics, such as chlorpromazine and thioridazine, are more sedating than the high-potency antipsychotics such as haloperidol and fluphenazine and tend to cause more significant CNS depression. For example, chlorpromazine coupled with very low blood alcohol concentrations ($\sim 40\,\text{mg dl}^{-1}$) impairs skills related to driving and produces subjective complaints of feeling sleepy, lethargic, dull, groggy, and poorly coordinated behavior [33]. Less psychomotor impairment was observed

when alcohol was combined with flupenthixol [34, 35] or thioridazine [34–36] but not with haloperidol [34, 35]. Changing doses, steady-state pharmacokinetics, and types of antipsychotic medication make predictions about the acute interactions of these drugs with alcohol difficult. Long-term use of some psychotropic medications may result in extrapyramidal motor system disorders that can be misinterpreted as some form of intoxication. Also, chronic alcohol consumption causes increased metabolism of the antipsychotic medications, resulting in lower blood levels and, ultimately, lesser efficacy of the medication. Since antipsychotic medications are typically used to treat mental illnesses such as schizophrenia, when combined with alcohol, the results may be particularly challenging and unpredictable.

Alcohol and Cannabinoids

It is well known by forensic examiners, police, and toxicologists that alcohol and tetrahydrocannabinol (THC) are two of the most widely used and commonly encountered drugs detected in motor vehicle collisions. Tetrahydrocannabinol is the primary psychoactive compound in marijuana and presumed to be the effect-producing drug in studies where marijuana is smoked. The effects of THC on drowsiness, memory, and distortion of space and time are particularly important because of the obvious need for such skills in motor vehicle operation, for example.

In one of the earlier interactive studies on the effects of THC and alcohol on driving, Casswell [37] found that (i) alcohol alone increased speed and impaired steering, (ii) THC alone reduced speed and slowed response to instruction, and (iii) alcohol and marijuana tended to increase speed, impair steering, and increase response times to instructions. More recent studies concluded that the effects of THC and alcohol appear to be additive. When THC is combined with low blood alcohol concentrations (BACs) (40 mg dl^{-1}), there is evidence of impaired visual search patterns while driving, and reaction times are greater than from either drug alone. The authors point out that the effects of alcohol are greater than those produced by smoking marijuana, noting that the combination is particularly dangerous with regard to motor vehicle operation [38, 39].

Although the majority of more recent studies support the conclusion that alcohol and THC impair driving [40–42], not all investigators found an interaction between these drugs. For example, Smiley, Ziedman, and Moskowitz [43] found that after smoking marijuana, drivers became more cautious in overtaking tasks. Alcohol (45 or 75 mg dl^{-1}) had a slight effect, but no interaction between alcohol and smoked marijuana was observed. Using a driving simulator, Stein *et al.* [44] found that after alcohol consumption (100 mg dl^{-1}), drivers had more "accidents" and "speeding tickets" and slower and less accurate responses to road signs. Smoking marijuana had "only an occasional effect" and no interaction was observed.

The nature of any pharmacodynamic interaction between alcohol and cannabinoids is difficult to explain since the neuropharmacological effects of these drugs are complex and diverse. There is some evidence of a pharmacokinetic interaction between these drugs. Lukas *et al.* [45] found that smoking marijuana decreases the bioavailability of alcohol, reducing the maximum serum alcohol concentration and delaying the time to peak concentration from 78 mg dl^{-1} (50 min after drinking) to about 55 mg dl^{-1} (105 min after drinking).

More research on this potential pharmacokinetic interaction is needed, and forensic examiners are cautioned that pharmacokinetic and pharmacodynamic interactions that are measurable at low drug concentrations may be obscured at the higher concentrations, which are more likely to be encountered in forensic evaluations.

Alcohol and Cardiovascular Medications

Acute alcohol consumption interacts with some cardiovascular medications (e.g., nitroglycerin, used to treat angina; reserpine, methyldopa (Aldomet), hydralazine (Apresoline), and guanethidine (Ismelin), used to treat hypertension) to cause dizziness or fainting upon standing (orthostatic hypotension). Chronic alcohol consumption decreases the availability of propranolol (Inderal), used to treat high blood pressure, potentially reducing its therapeutic effect. Since alcohol acts as an osmotic diuretic and also causes hypokalemia, patients taking loop diuretics (e.g., furosemide, ethacrynic acid, and bumetanide) or less potent thiazide and various sulfonamide derivatives, are at greater risk for dehydration, hypokalemia and, therefore, risk of seizure. Verapamil, which inhibits the metabolism of the alcohol by the liver, can increase blood alcohol content, whereas alcohol may

increase the bioavailability of nifedipine by inhibiting its metabolism [46].

The inhibition of aldehyde dehydrogenase by the coronary artery dilator isosorbide dinitrate (Isordil, Dilatrate, and Sorbitrate) and nitroglycerin (Nitro-Bid and Nitrostat) can produce a disulfiram-like reaction (described in the following section) when combined with relatively low amounts of alcohol. Such interactions may complicate drug recognition evaluations (DREs).

Alcohol and Disulfiram

Disulfiram (Antabuse) is used to control drinking through aversion therapy, but the combination of drugs has serious and even fatal consequences due to their interactive biochemistry. Alcohol is metabolized by ADH to acetaldehyde, which is metabolized by aldehyde dehydrogenase. Disulfiram inhibits the enzyme aldehyde dehydrogenase, resulting in the accumulation of acetaldehyde. Acetaldehyde is highly toxic and produces many aversive sympathomimetic effects including facial flushing, nausea, vomiting, breathing difficulties, and headache. In extreme cases, respiratory depression, cardiovascular collapse, cardiac arrhythmias, unconsciousness, and convulsions leading to death can occur.

In addition to disulfiram, other prescription medications can inhibit aldehyde dehydrogenase and can produce a disulfiram-like reaction in people who consume alcohol while taking them [47]. The following medications may produce severe sympathomimetic reactions when combined with alcohol: the antidiabetic oral medications chlorpropamide (Diabinase), glyburide (Micronase, Diabeta), tolazamide, and tolbutamide; a number of antibiotic medications including cefoperazone (Cefobid), griseofulvin (Fulvicin, Grisactin), INH, metronidazole (Flagyl), nitrofurantoin (Furadantin, Macrodantin), and sulfamethoxazole (Bactrim, Septra); the analgesics phenacetin and phenylbutazone; and the coronary artery dilator medications isosorbide dinitrate (Isordil, Dilatrate, and Sorbitrate) and nitroglycerin (Nitro-Bid and Nitrostat). In patients with certain medical conditions (i.e., those with coronary artery disease), the disulfiram-like reaction could be potentially fatal as a result of cardiovascular effects involved in the pathogenesis of the disulfiram-like reaction (dilation of blood vessels, hypotension, and tachycardia). As it is impossible to predict with certainty the severity of the disulfiram-like reaction, individuals taking any of these medications should be advised to avoid alcohol, and a thorough review of prescription medications should always be part of any forensic examination if unusual symptoms are present in intoxicated patients.

Alcohol and Histamines (H_2-Antagonists)

Alcohol abuse may contribute to gastrointestinal diseases, including gastritis, ulcers, and gastroesophageal reflux disorder (GERD). H_2-antagonists, such as cimetidine (Tagamet), ranitidine (Zantac), and nizatidine (Axid), are commonly used to treat these disorders.

As a result, these medications can increase the bioavailability of alcohol. Moreover, gastric ADH may account for a significant percentage of alcohol metabolism [48], at least at low blood alcohol concentrations. In addition, the first-pass metabolism of alcohol is also reduced by cimetidine because of the effect it may have on increasing the rate of gastric emptying, again resulting in increased blood alcohol levels [49]. In many studies, the effect was quite significant – increases of blood alcohol concentrations of about 17–33% by cimetidine and to a much lesser degree by rantidine, if at all [50, 51]. Although these investigators examined this effect through a series of detailed studies and identified dose, drug type, gender, and drinking history to be important variables, the clinical significance of this interaction has been questioned by some researchers [52], and this effect is not universal to all H_2 antagonists. For example, Famotidine (Pepcid) appears to have no effect on blood alcohol levels.

The use of the H_2-antagonists for the treatment of gastroesophageal reflux disease (GERD) has largely been supplanted by a newer class of agents that reduce gastric acid secretion, the proton pump inhibitors (PPIs) such as omeprazole (Prilosec), lansoprazole (Prevacid), esomeprazole (Nexium), and rabeprazole (Acidphex). The PPIs do not appear to interact significantly with alcohol. Therefore, broad generalizations about alcohol and H_2 antagonists should be avoided, and careful examinations of the specific medications and dose should be made in evaluating the role of these new medications in combination with alcohol. In other words, some, but not all, H_2 antagonists affect alcohol pharmacokinetics.

Alcohol and Lipid-Reducing Medications

Statins, medications used to reduce elevated lipids and cholesterol, are usually metabolized through the cytochrome P450 enzyme system. Atorvastatin (Lipitor), simvastatin (Zocor), and lovastatin (Mevacor, Altocor) are metabolized through the CYP3A4 isoenzyme, which is also involved in the metabolism of alcohol and a number of other medications that bind more strongly to the enzyme than statins. When alcohol or any of these substances block the statin from binding to the CYP3A4 enzyme, the metabolism of the statin is reduced, resulting in increased statin levels and the potential for statin-related toxicity [53–56]. The major toxic reactions include myotoxicity (myalgia, myopathy, and rhabdomyolysis) and hepatotoxicity. Milder cases are often reversible without serious sequelae with substance discontinuation, but severe cases, although rare, may be potentially fatal.

Alcohol and Methanol

Alcoholics may drink other forms of alcohol such as methanol when beverage alcohol (ethanol) is not available. Methanol (methyl alcohol) is highly toxic, and consumption of small quantities may result in metabolic acidosis, blindness, and death. Interestingly, the interaction between ethanol and methanol is critical in treating methanol poisoning. Methanol is metabolized by the enzyme ADH to formaldehyde and then to formic acid, a highly toxic compound [57]. Even when relatively small doses of methanol (several ounces) may cause metabolic acidosis, blindness, and cardiovascular instability and death. It is noteworthy that methanol poisoning can be prevented by the administration of ethyl alcohol because ethanol (alcohol) is preferentially metabolized by ADH, thereby decreasing the formation of toxic metabolites from methanol. The decrease in methanol metabolism allows methanol to be excreted unchanged and before toxic metabolites are formed. Patients admitted for acute alcohol intoxication or detoxification should be screened for methanol use so that appropriate prophylactic treatment (e.g., hemodialysis, ADH inhibitors, and ethanol administration) can be initiated.

Alcohol and Non-narcotic Pain Relievers (ASA and NSAIDS)

Aspirin (acetylsalicylic acid (ASA)) and nonsteroidal anti-inflammatory drugs (NSAIDs) frequently used to treat mild to moderate pain also decrease the activity of gastric ADH. This biochemical damage increases the bioavailability of ingested alcohol and heightens the effects of a given dose of alcohol [50] at least at low blood alcohol concentrations. There are a number of other non-narcotic pain relievers including phenacetin, an acetaminophen precursor (and often found in other drugs including acetaminophen, aspirin, caffeine, codeine, and propoxyphene), and phenylbutazone (Butazolidin), both of which inhibit aldehyde dehydrogenase. Therefore, the use of these drugs with alcohol may result in an aversive disulfiram-like reaction (see previous section on Alcohol–Disulfram interactions).

Alcohol and Opiates

The effects of opiate medications such as codeine, morphine preparations, propoxyphene (Darvon), oxycodone preparations (Percocet, Oxycontin), hydromorphone (Dilaudid), hydrocodone (Vicodin, Lortab), meperidine (Demerol), and fentanyl can be enhanced by the depressant effect of alcohol, resulting in decreased motor skills, respiratory problems, drowsiness, and sedation. For example, a single dose of alcohol can increase the bioavailability of propoxyphene, potentially increasing its sedative effect [58]. Opiate medications (i.e., codeine, propoxyphene, and oxycodone) are manufactured as combination products with the non-opiate analgesics (e.g., acetaminophen), which can result in an interaction between alcohol and the acetaminophen, as well as the opiate. The accumulation of toxic breakdown products forms an acetaminophen/alcohol interaction and can be potentially dangerous, resulting in liver damage or failure. Patients who are prescribed any of the opiate/acetaminophen combination preparations should be cautioned about consuming any additional amounts of acetaminophen.

Methadone, a synthetic opioid with a relatively long half-life, is used to reduce heroin use in opiate-dependent patients. New patients or patients receiving a significant increase in their daily oral dose of methadone may present with symptoms of mild psychomotor and cognitive impairment. However, unlike many other opioids, prolonged methadone use, even at relatively high doses, does not impair cognitive or psychomotor performance [59] including driving [60–62]. However, the combination of alcohol and opiates does increase respiratory depression. It is

generally advisable not to drink alcohol while taking methadone for a variety of reasons related to recovery, but the interaction between these drugs is limited.

Alcohol and Sedative Hypnotics

The depressant effects of barbiturates range from mild sedation to general anesthesia and are similar to alcohol. The interaction between these drugs is additive probably because of both pharmacokinetic and pharmacodynamic mechanisms. Alcohol may also inhibit the hepatic metabolism of barbiturates [63], which would increase their bioavailability and effectiveness. Gamma aminobutyric acid (GABA) interacts with receptors containing recognition sites for anxiolytics and sedative benzodiazepines and barbiturates, which regulate gating of chloride channels. Since alcohol also alters the gating properties of these receptors, a pharmacodynamic interaction between alcohol and these medications exists even though alcohol has little affinity to recognition sites for GABA, benzodiazepines, or barbiturates [3].

Both barbiturates and alcohol derive some pharmacodynamic properties through the $GABA_A$ receptor. Barbiturates enhance the binding of GABA to $GABA_A$ receptors. Alcohol also shares the ability to increase GABA-mediated synaptic inhibition and chloride ion flux, which explains in part interaction of these drugs. Although other mechanisms are involved in the psychoactive effects of both drugs, the combination of alcohol and barbiturates will result in greater psychomotor and other CNS impairments than either drug alone.

Many benzodiazepines and alcohol share similar pharmacokinetic and pharmacodynamic mechanisms, so it is not surprising that the combination of alcohol and benzodiazepines is associated with drug-induced deaths, drug overdoses, and traffic accidents or fatalities [64, 65].

A pharmacokinetic interaction between alcohol and benzodiazepines also exists. For example, there is a decline in the efficiency of the metabolism of the benzodiazepines as a result of increasing age or liver disease. In the elderly, there is a 50% decrease in clearance, with a four- to ninefold increase in half-life, and a two- to four fold increase in the volume of distribution [66].

Alprazolam (Xanax), a benzodiazepine analog, is used in the treatment of anxiety disorders and is currently the most prescribed medication in the United States. Although there is no synergistic action between this medication and alcohol, an additive interaction has been reported on certain psychomotor and cognitive tasks, such as information processing and memory [67]. The combination of alcohol and alprazolam produces increases in self-reported drowsiness [67], but no significant interaction between alprazolam and alcohol was reported [68] even though each drug produces several effects.

Some nonbenzodiazepine anxiolytics, such as buspirone, also do not appear to interact with alcohol to potentiate cognitive or motor performance impairment. Unlike some benzodiazepines, buspirone has been found to have no significant effect on body sway, coordination skills, tracking skills, or nystagmus even when combined with alcohol [69]. This lack of alcohol interactions with some benzodiazepine-like medications should be noted when interpreting DREs performed by police.

Alcohol and Stimulants

Although there is some evidence that stimulants decrease fatigue and reaction time while causing an increase in arousal, body temperature, heart rate, blood pressure, and other changes that may decrease some of the depressant effects of alcohol (e.g., sleepiness), the combination of a stimulant and a depressant is often erroneously assumed by laypersons to result in a neutralizing or balancing out of these two opposite effects. The scientific literature indicates the interactions between alcohol and stimulants to be inconsistent and often complex. For example, studies reported no antagonistic effect of dextroamphetamine on the mental and psychomotor impairment produced by alcohol, whereas others have found improved performance compared with controls [70, 71]. Perez-Reyes *et al.* [71] reported that alcohol ($100\,mg\,dl^{-1}$)) significantly increased the bioavailability of high, but not low, doses of dextroamphetamine ($25.5\,ng\,ml^{-1}$ vs. $15.7\,ng\,ml^{-1}$). Moreover, amphetamine did not significantly alter peak blood alcohol concentrations, but alcohol significantly increased the bioavailability of dextroamphetamine. Dextroamphetamine attenuated alcohol-induced increases in latency and accuracy while performing an eye–hand–foot reaction time task believed to be related to driving abilities. As the effect was greatest about 4 h after alcohol administration, when blood alcohol concentrations had dropped

from a peak of about 100 mg dl^{-1} to about 60 mg dl^{-1}, fatigue and dose were probably important factors in the latter finding [71]. The statistical significance of these results probably outweighs their actual value.

Although the interactions between alcohol and amphetamines have been studied, the only clear conclusion is that the interaction is complex and task specific, and that there is no simple antagonism between the drugs.

There does seem to be an interaction between alcohol and another stimulant, cocaine. One novel consequence of cocaine and alcohol use is that a third compound, cocaethylene, is formed. Cocaethylene (sometimes referred to as *ethyl cocaine* or *cocaine ethyl-ester*) is not a natural alkaloid of the coca plant and is not found in pharmaceutical or street cocaine. In fact, cocaine is metabolized to its ethyl configuration only in the presence of ethanol. Cocaine and cocaethylene have similar effects, but the latter extends the euphorogenic and reinforcing effects of the former. In humans, the combination of alcohol and cocaine is greater than the effect of either drug alone, and is associated with an enhanced subjective euphoria, increased heart rate, and increased plasma cocaine concentrations [72, 73]. In humans, insufflation of cocaine before alcohol ingestion does not appear to alter blood alcohol levels or subjective ratings of intoxication of alcohol intoxication. However, when alcohol is administered prior to cocaine insufflation, there is a significant increase in both cocaine plasma levels (possibly due to an inhibition of hepatic cocaine metabolism produced by alcohol) and an augmentation of cocaine's subjective and heart rate effects [74, 75]. In another study, combining cocaine and alcohol produced a nonsignificant decrease in subjective feelings of drunkenness, an increase in cocaine-induced euphoria, and a significant improvement in alcohol-related changes in psychomotor performance along with a marked increase in heart rate. Subjects who were administered cocaine and alcohol interpreted the effects as "more pleasant" than compared to alcohol alone [76]. Although many drugs are metabolized to form other psychoactive compounds, cocaethylene is the only known example of a third psychoactive drug being formed as a result of administering two other psychoactive drugs of abuse. As cocaine-induced deaths have been observed with a wide range of postmortem cocaine concentrations, often with presence of low blood alcohol levels, the possibility exists that cocaethylene may be partially responsible for these deaths [77]. Additional research is clearly required in this area.

Summary and Conclusions

With some exceptions, such as the interaction between drugs and specific antagonists to receptors for those drugs, or combinations of drugs of the same class, drug interactions are often difficult to predict with great precision. The interactions between alcohol and therapeutic medications or other drugs are very complex, highly variable events, and dependent upon a number of pharmacodynamic and pharmacokinetic factors. These interactions are further complicated by the potential for alcohol to cause liver pathologies that impact on the therapeutic effectiveness of some medications, even if the patient is sober at the time of the testing. In evaluating drug effects and drug interactions, there are a multitude of factors that can affect outcome including, but not limited to, the anthropometric characteristics of the patient (age, gender, body weight, and height), medication/drug dosage, and the amount and frequency of alcohol consumption. Patients in certain populations (e.g., geriatric) may use multiple medications or be susceptible to age-related physiological effects that could increase the potential for more severe drug interactions. When interpreting empirical research or clinical studies, forensic examinations must be cognizant of the fact that most of these studies involve relatively small alcohol doses for a variety of practical and ethical reasons, and the effects observed at relatively low blood alcohol concentrations may be very different than those in highly intoxicated patients. Similarly, the lack of a significant interaction with low doses of a drug does not guarantee the same results when higher doses are used. As a result, it is very difficult to stipulate a safe dose of alcohol that can be consumed when taking medications. Individuals with little or no tolerance to alcohol or other drugs may be more sensitive to these interactions. In conclusion, the CNS depressant effects of alcohol can be expected to increase the depressant effects of most medications and other drugs with similar CNS effects, but the significance of that interaction will be drug specific, even within some classes of medications. Generalizations about medication or drug effects in combination with alcohol should be avoided, as not alcohol–other drug

or medication combinations produce pharmacokinetic or pharmacodynamic interactions.

References

[1] Brick, J., Wallen, M.C. & Lorman, W.J. (2008). Interaction of alcohol with medications and other drugs, in *Handbook of the Medical Consequences of Alcohol and Drug Abuse*, J. Brick, ed, Haworth Press/Taylor Francis Group, USA, pp. 527–564.

[2] Adams, W.L. (1995). Interactions between alcohol and other drugs, *International Journal of Addictions* **30**, 1903–1923.

[3] Crews, F.T. (2008). Effects of alcohol abuse on Brain Neurochemistry, in *Handbook of the Medical Consequences of Alcohol and Drug Abuse*, J. Brick, ed, The Haworth Medical Press, Binghamton, pp. 123–176.

[4] Moser, R.S., France, C.E. & Brick, J. (2008). The neuropsychological consequences of alcohol and drug abuse, in *Handbook of the Medical Consequences of Alcohol and Drug Abuse*, J. Brick, ed, The Haworth Medical Press, Binghamton, pp. 57–96.

[5] Kuffner, E.K. (2001). New perspectives on acetaminophen and alcohol use, *The American Journal of Managed Care* **7**(19), S590–S591.

[6] Black, M. (1984). Acetaminophen hepatotoxicity, *Annual Review of Medicine* **35**, 577–593.

[7] Girre, C., Hispard, E., Palombo, S., N'Guyen, C. & Daily, S. (1993). Increased metabolism of acetaminophen in chronically alcoholic patients, *Alcoholism: Clinical and Experimental Research* **17**(1), 170–173.

[8] Seeff, L.B., Cuccherini, B.A., Zimmerman, H.J., Adler, E. & Benjamin, S.B. (1986). Acetaminophen hepatotoxicity in alcoholics: a therapeutic misadventure, *Annals of Internal Medicine* **104**(3), 399–404.

[9] Kuffner, E.K. & Dart, R.C. (2001). Acetaminophen use in patients who drink alcohol: current study evidence, *The American Journal of Managed Care* **7**(19), S592–S596.

[10] Jacobson, J.M. (1992). Alcoholism and tuberculosis, *Alcohol Health and Research World* **16**(1), 39–45.

[11] Nosova, T., Jokelainen, K., Kaihovaara, P., Vakevainen, S., Rautio, M., Jousimies-Somer, H. & Salaspuro, M. (1999). Ciprofloxacin administration decreases enhanced ethanol elimination in ethanol-fed rats, *Alcohol and Alcoholism* **34**(1), 48–54.

[12] Cina, S.J., Russell, R.A. & Conradi, S.E. (1996). Sudden death due to metronidazole/ethanol interaction, *American Journal of Forensic Medicine and Pathology* **17**(4), 343–346.

[13] Lieber, C.S. (1994). Alcohol and the liver: 1994 update, *Gastroenterology* **106**, 1085–1105.

[14] Ellenhorn, M. & Barceloux, D. (1988). *Medical Toxicology: Diagnosis and Treatment of Human Poisoning*, Elsevier Publishing Co., p. 785.

[15] Greenspan, K. & Smith, T.J. (1991). Perspectives on alcohol and medication interactions, *Journal of Alcohol and Drug Education* **36**(3), 103–107.

[16] Hesselbrock, M.H., Meyer, R.E. & Keener, J.J. (1985). Psychopathology in hospitalized alcoholics, *Archives of General Psychiatry* **42**, 1050–1055.

[17] Dorian, P., Sellers, E.M., Reed, K.L., Warsh, J.J., Hamilton, C., Kaplan, H.L. & Fan, T. (1983). Amitriptyline and ethanol: pharmacokinetic and pharmacodynamic interaction, *European Journal of Clinical Pharmacology* **25**, 325–331.

[18] Weathermon, R. & Crabb, D. (1999). Alcohol and medication interactions, *Alcohol Research and Health* **23**(1), 40–54.

[19] Ciraulo, D.A., Anderson, L.M., Chapron, D.J., Jaffe, J.H., Bollepalli, S. & Kramer, P.A. (1988). Clinical pharmacokinetics of imipramine and desipramine in alcoholics and normal volunteers, *Journal of Clinical Psychopharmacology* **43**, 509–518.

[20] Mattila, M.J. (1990). Alcohol and drug interactions, *Annals of Medicine* **22**, 363–369.

[21] Allen, D., Lader, M. & Curran, H.V. (1988). A comparative study of the interactions of alcohol with amitriptyline fluoxetine and placebo in normal subjects, *Progress in Neuro-Psychopharmacology and Biological Psychiatry* **12**, 63–80.

[22] Lemberger, L., Rowe, H., Vergstron, R.F., Farid, K.Z. & Enas, G.G. (1985). Effect of fluoxetine on psychomotor performance, physiological response and kinetics of ethanol, *Clinical Pharmacology and Therapeutics* **37**, 658–664.

[23] Simpson, G.M. & Gratz, S. (1992). Comparison of the pressor effect of tyramine after treating with phenelzine and moclobemide in healthy male volunteers, *Journal of Clinical Pharmacology and Therapeutics* **52**, 2896–2291.

[24] Blackwell, B., Barley, E., Price, J. & Taylor, D. (1967). Hypertensive interactions between monoamine oxidase inhibitors and food stuffs, *British Journal of Psychiatry* **113**, 349–365.

[25] Sitsen, J.M.A. & Zikov, M. (1995). Mirtazapine: clinical profile, *CNS Drugs* **4**(Suppl 1), 39–48.

[26] Brick, J. (1983). Diabetes breath acetone and breathalyzer accuracy. A case study, *Alcohol, Drugs and Driving* **9**(1), 27–28.

[27] Ridout, F., Shamsi, Z., Meadows, R., Johnson, S. & Hindmarch, I. (2003). A single center, randomized, double-blind, placebo-controlled, crossover investigation of the effects of fexofenadine hydrochloride 180 mg alone and with alcohol, with hydroxyzine hydrochloride 50 mg as a positive internal control, on aspects of cognitive and psychomotor function related to driving a car, *Clinical Therapeutics* **25**(5), 1518–1538.

[28] Burns, M. (1989). Alcohol and antihistamines in combination, effects on performance, *Alcoholism: Clinical and Experimental Research* **13**, 243.

[29] Burns, M. & Moskowitz, H. (1980). Effects of diphenhydramine and alcohol on skills performance, *European Journal of Clinical Pharmacology* **17**, 259.

[30] Franks, H.M., Hensley, V.R., Hensley, W.J., Starmer, G.A. & Teo, R.K.C. (1979). The interaction between ethanol and antihistamines, *The Medical Journal of Australia* **1**, 185–186.

[31] Roehrs, T., Zwyghuizen-Doorenbos, A. & Roth, T. (1993). Sedative effects and plasma concentrations following single doses of tiazolam, diphenhydramine, *Ethanol and Placebo Sleep* **16**(4), 301–305.

[32] Zimatkin, S.M. & Anichtchik, O.V. (1999). Alcohol-histamine interaction, *Alcohol and Alcoholism* **34**(2), 141–147.

[33] Zirkle, G.A., King, P.D., McAtee, O.B. & Van Dyke, R. (1959). Effects of chlorpromazine and alcohol on coordination and judgment, *Journal of the American Medical Association* **171**, 1496–1499.

[34] Linnoila, M. (1973). Effects of diazepam, chlordiazepoxide, thioridazine, haloperidol, flupenthixol and alcohol on psychomotor skills relating to driving, *Annales Medicinae Experimentalis et Biologiae Fenniae* **51**, 125.

[35] Linnoila, M., Sarrio, I., Olkonieme, J., Liljeqvist, R. & Maki, M. (1975). Effect of two weeks' treatment with chlordiazepoxide or flupenthixol alone or in combination with alcohol, on psychomotor skills related to driving, *Arzneim-Forschung Drug Research* **25**, 1088.

[36] Millner, G. & Landauer, A.A. (1971). Alcohol, thioridazine and chlorpromazine effects on skills related to driving behaviour, *The British Journal of Psychiatry* **118**, 351–352.

[37] Casswell, S. (1977). Cannabis and alcohol: effects on closed course driving behaviour, *Paper presented to Seventh International Conference on Alcohol, Drugs and Traffic Safety*, Melbourne.

[38] Lamers, C.T.J. & Ramaekers, J.G. (2000). *Visual Search and Urban City Driving Under the Influence of Marijuana and Alcohol*, Report DOT HS 809 020, National Highway Traffic Safety Administration US Department of Transportation, Washington, D.C.

[39] Robbe, H.W.J. & O'Hanlon, J.F. (1999). *Marijuana, Alcohol and Actual Driving Performance*, Report DOT HS 808 939, National Highway Traffic Safety Administration US Department of Transportation.

[40] Moskowitz, H. (1976). Marijuana and driving, *Accident Analysis and Prevention* **8**, 21–26.

[41] Moskowitz, H. (1985). Marijuana and driving, *Accident Analysis and Prevention* **17**, 323–346.

[42] National Institute on Alcohol Abuse and Alcoholism (2000). *Tenth Special Report to the US Congress on Alcohol and Health* US Department of Health and Human Services, Washington, D.C.

[43] Smiley, A., Ziedman, K. & Moskowitz, H. (1981). *Pharmacokinetics of Drug Effects on Driving Performance: Driving Simulator Tests of Marijuana Alone and in Combination with Alcohol*. Report prepared for NIDA and the National Highway Traffic Safety Administration Contract 271-76-3316, Southern California Research Institute, Los Angeles, CA.

[44] Stein, A.C., Allen, R.W., Cook, M.L. & Karl, R.L. (1983). *A Simulator Study of the Combined Effects of Alcohol and Marijuana on Driving Behavior – Phase II Report*, DOT HS-5-01257, National Highway Traffic Safety Administration, Washington, D.C.

[45] Lukas, S.E., Benedikt, R., Mendelson, J.H., Kouri, E., Sholar, M. & Amass, L. (1992). Marijuana attenuates the rise in plasma ethanol levels in human subjects, *Neuropsychopharmacology* **7**, 77–81.

[46] Bauer, L.A., Schumock, G., Horn, J. & Opheim, K. (1992). Verapamil inhibits ethanol elimination and prolongs the perception of intoxication, *Clinical Pharmacology and Therapeutics* **52**(1), 6–10.

[47] Lieber, C.S. (ed) (1992). Interactions of ethanol with other drugs, in *Medical and Nutritional Complications of Alcoholism: Mechanisms and Management*, Plenum Press, New York, pp. 165–183.

[48] Baraona, E., Abittan, C.S. & Lieber, C.S. (2002). Contribution of gastric oxidation to ethanol first pass metabolism in baboons, *Alcoholism: Clinical and Experimental Research* **24**(7), 946–951.

[49] Lieber, C.S. (1997). Gastric ethanol metabolism and gastritis: Interactions with other drugs, Helicobacter pylori, and antibiotic therapy, *Alcoholism: Clinical and Experimental Research* **21**, 1360–1366.

[50] Caballeria, J., Baraona, E., Rodamilans, M. & Lieber, C.S. (1989). Effects of cimetidine on gastric alcohol dehydrogenase activity and blood alcohol levels, *Gastroenterology* **96**, 388–392.

[51] Roine, R., Gentry, R.T., Hernandez-Munoz, R., Baraona, F. & Lieber, C.S. (1990). Aspirin increases blood alcohol concentrations in humans after ingestion of ethanol, *Journal of the American Medical Association* **264**(18), 2406–2408.

[52] Levitt, M.D. (1993). Review article: lack of clinical significance of the interaction between H_2-receptor antagonists and ethanol, *Alimentary Pharmacology & Therapeutics* **7**, 131–138.

[53] Farmer, J.A. & Torre-Amione, G. (2000). Comparative tolerability of the HMG-CoA reductase inhibitors, *Drug Safety* **23**, 197–213.

[54] Igel, M., Sudhop, T. & von Bergmann, K. (2001). Metabolism and drug interactions of 3-hydroxy-3-methyglutaryl coenzyme A-reductase inhibitors (statins), *European Journal of Clinical Pharmacology* **57**, 357–364.

[55] Paoletti, R., Corsini, A. & Bellosta, S. (2002). Pharmacological interactions of statins, *Atherosclerosis Supplements* **3**, 35–40.

[56] Worz, C.R. & Bottorff, M. (2001). The role of cytochrome P450-mediated drug-drug interactions in determining the safety of statins, *Expert Opinion on Pharmacotherapy* **2**, 1119–1127.

[57] Baselt, R. (2004). Disposition of toxic drugs and chemicals, in *Man*, 7th Edition, Biomedical Publications, Foster City California.
[58] Girre, C., Hirschhorn, M., Bertaux, L., Palombo, S., Dellatolas, F., Ngo, R., Moreno, M. & Fournier, P.E. (1991). Enhancement of propoxyphene bioavailability by ethanol: relation to psychomotor and cognitive function in healthy volunteers, *European Journal of Clinical Pharmacology* **41**(2), 147–152.
[59] Zacny, J.P. (1995). A review of the effects of opioids on psychomotor and cognitive functioning in humans, *Experimental and Clinical Psychopharmacology* **3**, 432–466.
[60] Byas-Smith, M.G., Chapman, S.L., Reed, B. & Cotsonis, G. (2005). The effects of opioids on driving and psychomotor performance in patients with chronic pain, *Comparative Study Journal Article Research Support Non-US Gov't Clinical Journal of Pain* **21**(4), 345–352.
[61] Galski, T., Williams, J.B. & Ehle, H.T. (2000). Effects of opioids on driving ability, *Journal of Pain and Symptom Management* **19**, 200–208.
[62] Zacny, J.P. (1996). Should people taking opioids for medical reasons be allowed to work and drive, *Addiction* **91**, 1581–1584.
[63] Rubin, E., Gang, H., Misra, P.S. & Lieber, C.S. (1970). Inhibition of drug metabolism by acute ethanol intoxication: a hepatic microsomal mechanism, *The American Journal of Medicine* **49**, 801.
[64] Girre, C. & Schuckit, M.A. (1988). Alcohol and drug interactions with antianxiety medications, *The American Journal of Medicine* **82**(S5A), 27–33.
[65] Girre, C., Facy, F., Lagier, G. & Dally, S. (1988). Detection of blood benzodiazepines in injured people. Relationship with alcoholism, *Drug and Alcohol Dependence* **21**(1), 61–65.
[66] Peppers, M.P. (1996). Benzodiazepines for alcohol withdrawal in the elderly and in patients with liver disease, *Pharmacotherapy* **16**, 49–58.
[67] Linnoila, M., Stapleton, J.M., Lister, R., Moss, H., Lane, E., Granger, A. & Eckardt, M.J. (1990). Effects of single doses of alprazolam and diazepam alone and in combination with ethanol, on psychomotor and cognitive performance and on autonomic nervous system reactivity in healthy volunteers, *Journal of Clinical Pharmacology* **39**, 21–38.
[68] Hindmarch, I. (1983). Measuring the side-effects of psychoactive drugs: a pharmacodynamic profile of alprazolam, *Alcohol and Alcoholism* **18**(4), 361–367.
[69] Mattila, M.J., Aranko, K. & Seppala, T. (1982). Acute effects of buspirone and alcohol on psychomotor skills, *Journal of Clinical Psychiatry* **43**, 56–61.
[70] Kaplan, H.L., Forney, R.D., Richards, A.B. & Hughes, F.W. (1966). Dextro-amphetamine, alcohol and dextro-amphetamine-alcohol combination and mental performance, in *Alcohol and Traffic Safety*, R.N. Harger, ed, Indiana University Press, USA, pp. 211–214.
[71] Perez-Reyes, M., White, W.R., McDonald, S.A. & Hicks, R.E. (1992). Interaction between ethanol and dextroamphetmaines: effects of psychomotor performance, *Alcoholism: Clinical and Experimental Research* **16**, 75–81.
[72] McCance, E.F., Price, L.H., Kosten, T.R. & Jatow, P.I. (1995). Cocaethylene: pharmacology, physiology and behavioral effects in humans, *Journal of Pharmacology and Experimental Therapeutics* **274**(1), 215–223.
[73] McCance-Katz, E.F., Price, L.H., McDougle, C.J., Kosten, T.R., Black, J.E. & Jetlow, P.I. (1993). Concurrent cocaine-ethanol ingestion in humans: pharmacology, physiology, behavior, and the role of cocaethylene, *Psychopharmacology (Berl)* **111**(1), 39–46.
[74] Perez-Reyes, M., Jeffcoat, A.R., Myers, M., Sihler, K. & Cook, C.E. (1994). Comparison in humans of the potency and pharmacokinetics of intravenously injected cocaethylene and cocaine, *Psychopharmacology (Berl)* **116**(4), 428–432.
[75] Perez-Reyes, M. & Jeffcoat, A.R. (1992). Ethanol/cocaine interaction: cocaine and cocaethylene plasma concentrations and their relationship to subjective and cardiovascular effects, *Life Sciences* **51**(8), 553–563.
[76] Farre, M., De LaTorre, R., Llorente, M., Lamas, X., Ugena, B., Segura, J. & Cami, J. (1993). Alcohol and cocaine interactions in humans, *Journal of Pharmacology and Experimental Therapeutics* **266**(3), 1364–1373.
[77] Bunn, W.H. & Ginnini, A.J. (1992). Cardio vascular complications of cocaine abuse, *American Family Physician* **46**(3), 769–773.

JOHN BRICK

Alcohol: Metabolism *see* Alcohol: Analysis

Alcohol: Use, Abuse, Tolerance, and Dependency

Alcohol use, abuse, tolerance, and chemical dependence are often factors in the forensic investigation of crimes, accidents, or deaths (accidental or otherwise). Although these terms apply to other drugs, this article

will focus on and review basic concepts in defining alcohol, its use, abuse, tolerance, and dependence.

What is Alcohol?

Most people know what the term *alcohol* is, but unless specified it can be misleading or inaccurate. In some instances, this may result in a fatal misunderstanding. To an alcohol research scientist, toxicologist, or chemist, three of the most common alcohols are ethyl alcohol (ethanol), methyl alcohol (methanol), and isopropyl alcohol (isopropanol). Each has a similar chemical structure; a hydroxyl group attached to a saturated carbon molecule, each causes intoxication, and each has been represented in forensic cases. The general intoxicating effects of ingesting different alcohols are somewhat similar, but the side effects are quite different.

Methanol, also known as *wood alcohol*, is the most toxic of the three examples given. Relatively small amounts (less than one ounce) may cause retinal damage. Larger quantities (one or more ounces) can be fatal, but toxicity varies greatly [1, 2]. Methanol's toxicity is the result of its metabolism to formaldehyde, which is metabolized to formic acid, a cellular toxin that is about 6 times more poisonous than methanol itself. Formic acid produces severe metabolic acidosis and tissue hypoxia, and more than 6 to 7 ounces of methanol is lethal for many adults [1, 2]. Methanol intoxication usually occurs by accident, because of lack of knowledge about its harmful effects. For example, mixing methanol in "jungle juice" (a concoction of many different types of liquors and juices), typically served to large groups and popular among college students in the United States, may place the uninformed drinker at significant medical risk.

Methanol is also consumed by some alcoholics who, in the absence of beverage alcohol, may consume products containing it, such as antifreeze or dry gas used in automobiles. Initial symptoms of methanol intoxication are somewhat similar to those of beverage alcohol. Unless specifically alerted to possible methanol intoxication, a patient who becomes ill from methanol may be misdiagnosed by untrained or unobservant medical personnel and not receive proper treatment for methanol poisoning. For example, emergency treatment for methanol poisoning, such as ethanol administration, is highly unlikely if the patient is believed to be "drunk".

As mentioned, methanol's toxicity is derived from its metabolism to formaldehyde and formic acid. Methanol is metabolized by the enzyme alcohol dehydrogenase (ADH), which also metabolizes beverage alcohol. Because the affinity of beverage alcohol (ethanol) for ADH is about 10–20 times greater than that of methanol, when beverage alcohol and methanol are present at the same time, the liver preferentially metabolizes beverage alcohol. This allows time for methanol to pass through the excretory system before it is metabolized to harmful metabolites. Therefore, a rapid emergency treatment for methanol poisoning is beverage alcohol. Beverage alcohol, such as whiskey, vodka, rum, etc., is not likely to be administered to an unknown patient who presents symptoms of intoxication in an emergency department.

Isopropanol, or common rubbing alcohol, is less toxic than methanol but about twice as toxic as beverage alcohol. Small amounts, as little as several ounces, can also cause permanent damage to the visual system, and 8 ounces is estimated to be lethal [1]. Dose estimates are often based upon postmortem findings. Obviously, it would be highly unethical and illegal to conduct a range of dose–response studies in humans to determine the lethal dose. Actual lethal doses may be lower or higher, depending upon many factors. Some alcoholics may also consume isopropanol intentionally (e.g., some brands of dry gas contain both methanol and isopropanol) or unknowingly, also with potentially harmful or even lethal consequences.

However, the alcohol that is the primary subject of this review, and the alcohol consumed as a beverage by most people is ethyl alcohol (ethanol). Ethanol is a relatively odorless chemical formed through a process of fermentation and a drug that affects most organs of the body, including the brain [3, 4]. Ethanol is also lethal but at much higher doses than methanol or isopropanol. The lethal dose (LD50) of acute ethanol is estimated to be a blood alcohol concentration (BAC) of about 400–500 mg/dl, although death may occur at higher or lower concentrations depending upon factors such as tolerance or the presence of other drugs. For example, compared to 6 or 8 ounces of methanol or isopropanol, it would require about 24 ounces of 80 proof liquor (about 16 standard drinks) in a 150-pound male consumed over an hour to reach a BAC that would probably be lethal for about half the population. In this article, unless

otherwise indicated, the term *alcohol* will be used to denote ethanol.

Alcohol Use

Use in the United States

Alcohol use predates recorded history, but for all practical purposes the medical literature documenting the consequences of alcohol abuse goes back less than 200 years. Alcohol consumption and related problems have been well documented. Alcohol-related medical problems account for a disproportionate number of hospital admissions and medical complications [5, 6]. Yet, despite general public awareness about alcohol-related problems, longitudinal studies suggest that nearly 9% of adults in the United States consume, on average, more than two drinks per day [7]. Among senior high school students (typically 17–18 years old), about 3% use alcohol daily and about half had used alcohol within 30 days of being surveyed. About two-thirds of the adult population in the United States consumes alcohol to varying degrees, and 6% of adolescents (age 12–17) and about 8% of adults meet diagnostic criteria for alcohol abuse or alcohol dependence [8].

Defining Alcohol Use, Abuse, and Dependence

Alcohol use can first be defined in terms of how much alcohol was consumed. For example, a standard alcoholic drink is defined as 1.5 ounces of 80 proof liquor (e.g., rum, whiskey, vodka), a 12-ounce beer (about 5.0% v/v), or a 5-ounce glass of wine (12% v/v). Each of these drinks contains the same amount of ethanol. However, drink size can vary considerably in restaurants, bars, or at parties [9–11]. Nevertheless, for reporting purposes, standardization of drinks is useful and recommended in research and in forensic analyses [11].

The term *social drinker* or *light drinker* is often used by laypersons and some professionals but these terms may be inaccurate or misleading if not defined correctly. For example, the social use of alcohol is now generally described as a cold beer after a ball game, a glass of wine with meals, or a glass of champagne at festive occasions [12]. Similarly, the terms *light, moderate,* and *heavy* drinking are also used to describe drinking habits, but these terms are relative and may have different meanings when other factors are considered. For example, the US Department of Agriculture Health and Human Services defines moderate drinking as one drink per day or less for women and two or fewer drinks per day for men [13]. However, if someone admits to drinking six or seven drinks per week, that might equate to moderate drinking since it averages about one drink per day, but have an entirely different implication if all six or seven drinks were consumed on Friday nights. Such a pattern of drinking constitutes binge drinking, now defined as four to five drinks within 2 h or a BAC of 80 mg/dl [8]. Whereas one to two drinks per day is considered moderate, heavy drinking is considered by many scientists as the consumption of more five or six drinks a day, but many other factors are considered in making a diagnosis of heavy, or other types of drinking [13, 14], including the impact of drinking on life events [15]. Therefore, the terms *light, moderate*, and *heavy* should be interpreted carefully on the basis of individual drinker characteristics or research design.

The most recent edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) [16] defines two types of drinkers: those with alcohol abuse, and those with alcohol dependence. Alcohol abuse is intentional, conscious, and voluntary. Abusers drink too much too often, and make poor choices and decisions regarding their drinking. Their subsequent intoxication often results in injuries to themselves or others, medical consequences and expenses, lost productivity, and family problems. An example of alcohol abuse is drinking and then driving a car, or possibly drinking when pregnant. Most people with alcohol "problems" are alcohol abusers.

In contrast to abuse, alcohol dependence is pathological and unintended. Alcoholics lack control over their use of alcohol in lifestyle situations in which abusers would ordinarily stop drinking, even if the desire to do so is strong. For example, a man who drinks a quart of whiskey a day cannot stay employed, has lost his family and friends because of constant drunkenness and is vomiting blood several times a day, and consults with his medical doctor. Even when he is told that alcohol is causing bleeding in his esophagus, the man still cannot stop drinking. Even though there are serious medical consequences to continued drinking, he cannot stop. This is alcohol dependence.

The new classifications of dependence are confusing to those who remember the old 1950 World Health Organization definition. This latter definition, no longer valid, stated that an addicting drug had three qualities: (i) psychological dependence (equating generally with habituation or "craving"), (ii) tolerance (reduced drug effect so that more drug is needed to produce the original desired effect), and (iii) physical dependence (adaptation of the body to the drug so that one could only function normally when the drug was present. When the drug is discontinued, the previous adaptation of the nervous system no longer functions in the drugfree cellular environment and physical withdrawal symptoms appear). While this definition properly describes dependence on central-nervous-system (CNS) drugs such as heroin and alcohol, the definition falls short of describing dependence on cocaine, a CNS stimulant that has profound emotional withdrawal symptoms, but no significant physical withdrawal.

Impaired control, which is at the center of our new definition of addiction, is a characteristic of all drugs that are "addicting". The new term *dependence* (addiction) still has both psychological and physical components, but they are different from the old WHO definition. For example, impaired control is an obsessive preoccupation with the use of the drug (psychological) caused primarily by a neurochemical (physical) dysfunction in the brain. Older concepts such as "physical addiction" and "psychological addiction" have no accurate meaning and terms such as *alcohol addict* or *drug addict* do not appear in the current edition of DSM-IV. Terms such as *addict* or *addiction* are vague and scientifically imprecise and should not be used. Erickson [12] points out that these are pejorative terms (e.g., addict, junkies, and drunks) that detract from the science of alcohol (or drug) dependence as a brain disease [12, 16].

Tolerance and Dependence

Tolerance and physical withdrawal are side effects usually associated with moderate to heavy drug consumption, but are not rigid criteria for diagnosing alcohol dependence. In fact, social drinkers and alcohol abusers can demonstrate tolerance.

Tolerance is a decrease in the response to a drug as a function of exposure to that drug. The several types of tolerance are discussed below. Although tolerance is still a criterion for alcohol dependence, it cannot be inferred that if someone is tolerant they are an alcoholic. Tolerance is merely one of seven maladaptive patterns of use leading to a possible diagnosis of alcohol dependence [17]. Similarly, while anyone who consumes alcohol acquires some form of tolerance (see below), evaluations of alcohol-dependent patients should avoid the error of assuming that all "alcoholics" have a degree of tolerance that precludes any interpretation of their behavior while intoxicated. Although tolerance varies within and between subjects, it rarely confers immunity from most of the intoxicating effects of alcohol that would be of interest in forensic examination (e.g., drinking and driving). However, tolerance (or lack thereof) may be a factor in dramshop or comparative liability cases, where the question of general appearance (not performance on tests) is of legal interest. There are five primary forms of tolerance: acute, chronic, dispositional/metabolic, cross, and nonpharmacological.

Acute tolerance was first noted by Mellanby, who observed that in dogs there is greater behavioral impairment when the BAC is rising compared to the same BAC when alcohol is post absorption and declining [18]. A similar effect was observed in humans [19–22], and has since been termed *acute tolerance.* This phenomenon, which can occur in a single drinking episode, is termed *acute tolerance*.

Chronic tolerance develops in response to chronic or repeated exposure to ethanol and is a resistance to the intoxicating effects of the drug. Chronic tolerance occurs as the nervous system responds to the presence of alcohol and adapts to maintain homeostasis. A history of heavy drinking is likely to confer tolerance and less impairment compared to those in people who are moderate drinkers or abstainers [19, 23]. All drinkers acquire one or more forms of tolerance, but some chronic heavy drinkers become of exceptional tolerance and are able to function (to varying degrees) at BACs that would render less-tolerance drinkers unconscious or dead [23, 24]. Contrary to popular belief, although chronic drinking may reduce obvious symptoms of alcohol intoxication in some drinkers (compared with drinkers having little or no tolerance), it does not confer the same immunity from the impairment of complex tasks, such as driving, particularly at the high BACs often achieved by such drinkers.

Dispositional and metabolic tolerance results from differences in ethanol absorption and distribution, whereas metabolic tolerance occurs from more efficient metabolism. The key factor here is tolerance

occurs because of a change in drug availability to some target tissue.

Alcohol is metabolized primarily by enzymes in the liver. For many drugs, including alcohol, repeated administration results in enzyme induction. Under such circumstances, the increased availability of enzyme results in more rapid metabolism of the drug and, therefore, less pharmacological effect. In human alcoholics, the magnitude of metabolic tolerance can vary significantly. For example, most men and women eliminate alcohol at a rate of about 15–20 mg/dl h^{-1}, but in some chronic heavy drinkers that rate can be significantly higher [25]. When a history of chronic heavy drinking is known or suspected, metabolic tolerance should be considered in pharmacokinetic analyses. In such cases, a range of elimination rates should be used and not a single "average [11]".

Cross-tolerance is present when the use of one drug results in a change in response to a different drug. In other words, the effect of the use of one drug may increase or decrease the response of another drug. Cross-tolerance may be due to changes in CNS sensitivity or metabolic tolerance that occurs when drugs share similar pathways of metabolism. For example, oxidative metabolism catalyzed by cytochrome P-450, reduced nicotinamine–adenine dinucleotide phosphate (NADP), and the microsomal oxidizing system (MEOS) pathways are shared by barbiturates and some minor tranquilizers. In the presence of alcohol, the removal of these drugs can be delayed, resulting in higher than expected drug concentrations [26, 27]. Interactions between ethanol and environmental factors as well as other drugs may also contribute to cross-tolerance [28].

Nonpharmacological or learning-based theories of tolerance demonstrate state-dependent learning, in which a task learned and practiced under a drug state is better performed in the same drug state. The work of Siegel and others has provided additional evidence that learning plays a significant role in the development of tolerance. This research suggests that environmental stimuli preceding drug intake elicit a conditioned compensatory response that attenuates the drug effect. In a classical Pavlovian conditioning paradigm, the conditioned stimuli (CS) consist of the various experimental procedures and environmental stimuli associated with the administration of the drug. The unconditioned stimuli (UCS) are the actual direct pharmacological effects of the drug. The conditioned response (CR), once formed by repeated administrations of the drug, may then be demonstrated with a placebo by presenting the usual drug administration cues (CS) in the absence of the pharmacological effects of the drug (UCS). Conditioned tolerance has been demonstrated for many different drugs, including ethanol, both in animals [29–31] and in humans [32–34].

Conclusions

Alcohol comes in many forms, including ethanol, the form contained in beverages. Alcohol is one of the most widely used and abused drugs in the United States and elsewhere, and is frequently found in toxicology results stemming from investigations of accidental and intentional injuries. Forensic evaluations of cases where alcohol is involved should take into account differences between social alcohol use and alcohol abuse when describing drinker characteristics. Similarly, the biobehavioral consequences of alcohol abuse and dependence should be interpreted in the context of current definitions of abuse and dependence. Multiple forms of tolerance can result from acute or chronic alcohol use. Tolerance is a relative term and should not be considered indicative of alcoholism. Similarly, a diagnosis of alcohol dependence (alcoholism) should not lead to speculation regarding exceptional tolerance unless supported by additional evidence. Finally, tolerance should be considered in pharmacokinetic estimates of alcohol use for example, as well as in describing or predicting behaviors following the consumption of alcohol. By using standardized definitions of what constitute use, abuse, dependence, and understanding the nature of tolerance, forensic examiners can communicate accurately and without bias information relevant to jurors and the court.

References

[1] Baselt, R. (2004). *Disposition of Toxic Drugs and Chemicals in Man*, 7th Edition, Biomedical Publications, Foster City, California.

[2] Brahmi, N., Blel, Y., Abidi, N., Kouraichi, N., Thabet, H., Hedhil, A. & Amamou, M. (2006). Methanol poisoning in Tunisia: report of 16 cases, *Clinical Toxicology* **45**, 717–720.

[3] Pohorecky, L.A. & Brick, J. (1990). Pharmacology of Ethanol, in *Psychotropic Drugs of Abuse*, D.J.K. Balfour, ed, Pergamon Press, Elmsford, pp. 189–282.

[4] Brick, J. (2008). Medical Consequences of Acute and Chronic Alcohol Abuse, in *Handbook of the Medical Consequences of Alcohol and Drug Abuse*, J. Brick, ed, The Haworth Medical Press, Binghamton, pp. 9–56.

[5] Dufour, M. (1999). What is moderate drinking? *Alcohol Research & Health* **23**, 15–14.

[6] National Institute on Alcohol Abuse and Alcoholism, *Tenth Special Report to the U.S. Congress on Alcohol and Health* (2000). U.S. Department of Health and Human Services, Washington, DC.

[7] Dawson, D.A., Grant, B.F., Chou, S.P. & Pickering, R.P. (1995). Subgroup variation in U.S. drinking patterns: results of the 1992 national longitudinal alcohol epidemiologic study, *Journal of Substance Abuse* (3), 331–344.

[8] National Institute on Alcohol Abuse and Alcoholism, US Department of Health and Human Services (2008). *Alcohol Alert*, Vol.74.

[9] Kerr, W.C., Greenfield, T.K., Tujague, J. & Brown, S.E. (2005). A drink is a drink? Variation in the amount of alcohol contained in beer, wine and spirits drinks in the US methodological sample, *Alcoholism, Clinical and Experimental Research* **29**(1), 2015–2021.

[10] Case, G.A., Destefano, S. & Logan, B.K. (2000). Tabulation of alcohol content of beer and malt beverages, *Journal of Analytical Toxicology* **24**, 202–210.

[11] Brick, J. (2006). Standardization of alcohol calculations in research, *Alcoholism Clinical and Experimental Research* **30**(8), 1276–1287.

[12] Erickson, C. (2007). *The Science of Addiction: From Neurobiology to Treatment*, Norton Press, New York.

[13] U.S. Department of Agriculture and U.S. Department of Health and Human Services (1995). *Home and Garden Bulletin, No.232*, 4th Edition, U.S. Department of Agriculture, Washington DC.

[14] Oates, J. & McCoy, O. (1973). *Laboratory evaluation of alcohol safety interlock systems Instrument performance at high BALS. Report to the Highway Research Institute*, PB224 702/1, National Highway Traffic Safety Administration, Department of Transportation, Washington DC.

[15] McLellan, A., Luborsky, L., O'Brien, C. & Woody, G. (1980). An improved diagnostic instrument for substance abuse patients: the addiction severity index, *The Journal of Nervous and Mental Disease* **168**, 26–33.

[16] Erickson, C. & Wilcox, R. (2001). Neurobiological causes of addiction, *Journal of Social Work Practice in the Addictions* **1**(3), 7–22.

[17] American Psychiatric Association (APA) (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition, American Psychiatric Association, Washington, DC.

[18] Mellanby, E. (1919). *Alcohol: Its absorption into and disappearance from the blood under different conditions*, Special Report Series No 31 HMSO, Medical Research Committee, London.

[19] Goldberg, L. (1963). Quantitative studies on alcohol tolerance in man, *Acta Physiologica Scandinavica*, **5**(Suppl 5), 1–126.

[20] Carpenter, J.A. (1963). Effects of alcohol on some psychological processes, *Quarterly Journal of Studies on Alcohol* **23**, 274–314.

[21] Vogel-Sprott, M. (1979). Acute recovery and tolerance to low does of alcohol: differences in cognitive and motor skill performance, *Psychopharmacology* **61**, 287–291.

[22] Niaura, R., Nathan, P.E., Frankenstein, W., Shapiro, A. & Brick, J. (1987). Gender differences in acute psychomotor & pharmacokinetic response to alcohol, *Addictive Behaviors* **12**, 345–356.

[23] Minion, G.E., Slovis, C.M. & Boutiette, L. (1989). Severe alcohol intoxication: a study of 204 consecutive patients, *Clinical Toxicology* **27**(6), 375–384.

[24] Perper, J.A., Twerski, A. & Weinand, J. (1986). Tolerance at High BACs: a study of 110 cases and review of the literature, *Journal of Forensic Sciences* **31**(1), 212–221.

[25] Jones, A.W. & Sternebring, B. (1992). Kinetics of ethanol and methanol in alcoholics during detoxification, *Alcohol and Alcoholism (Oxford, Oxfordshire)* **27**, 641–647.

[26] Chakraborty, J. (1980). Metabolic Basis of Ethanol-Drug Interactions, in *Psychopharmacology of Alcohol*, M. Sandler, ed, Raven Press, New York, pp. 191–198.

[27] Lieber, C.S., Pirola, R. (1982). Clinical Relevance of Alcohol-Drug Interactions, in *Recent Advances in the Biology of Alcoholism*, C.S. Lieber & B. Stimmel, eds, Haworth Press, New York, pp. 41–65.

[28] Kalant H., Khanna J.M. (1980). Environmental Neurochemical Interactions, in *Ethanol Tolerance in Psychopharmacology of Alcohol*, M. Sandler, ed, Raven Press. New York, pp. 107–112.

[29] Le, A.D., Poulos, C.X. & Cappell, H. (1979). Conditioned tolerance to the hypothermic effects of ethyl alcohol, *Science* **206**, 1109–1110.

[30] Hinson R.E., Siegel S. (1980). The Contribution of Pavlovian Conditioning to Ethanol Tolerance and Dependence, in *Alcohol Tolerance, Dependence and Addiction: A Research Handbook*, H. Rigter & J. Crabbge, eds, Elsevier/ North-Holland Biomedical Press, Amsterdam, pp. 181–199.

[31] Melchior, C. & Tabakoff, B. (1981). Modification of environmentally cued tolerance to ethanol in mice, *The Journal of Pharmacology and Experimental Therapeutics* **219**, 175–180.

[32] Dafters, R. & Anderson, G. (1982). Conditioned tolerance to the tachhcardia effect of ethanol in humans, *Psychopharmacology* **78**, 365–367.

[33] Newlin, D. (1985). Human conditioned compensatory response to alcohol cues: initial evidence, *Alcohol* **2**, 507–509.

[34] Tiffany, S.T. & Baker, T.B. (1986). Tolerance to alcohol: psychological models and their application to alcoholism, *Annals of Behavioral Medicine : A Publication of the Society of Behavioral Medicine* **8**, 7–12.

JOHN BRICK

Algor Mortis *see* Death: Time of

Allelic Attribution *see* Allelic Designation

Allelic Designation

Introduction

Short tandem repeat (STR)-based DNA profiling methodology is effectively at the theoretical limit of detection (LOD) in that typable results can be generated from as little starting material as a single cell [1, 2]. However, one of the most challenging aspects of forensic DNA analysis is the interpretation of low-level testing results where it is difficult to reliably distinguish between noise and signal from template DNA that is associated with an evidence sample [3, 4]. This difficulty with minimal samples is often compounded by the consumptive nature of PCR-based DNA testing [5, 6] when material is unavailable for replicate testing. Forensic DNA testing laboratories typically endeavor to minimize the effect of baseline noise and stochastic artifacts by relying upon very conservative minimum peak height thresholds (commonly fixed in the range of 50–200 relative fluorescent units (RFUs)) that are established during the course of their validation processes [7–10]. However, the conservative nature of these commonly employed thresholds can also arbitrarily remove from consideration legitimate signal from trace and secondary contributors to an evidentiary sample – matters of critical importance in many criminal investigations.

Any measurement made with a light-detecting instrument, such as a genetic analyzer is subject to at least some level of background noise [11] – defined here as signal not associated with amplified DNA. Instrument-related factors that may contribute to background noise in DNA testing experiments are typically run-specific and include (but are not necessarily limited to) the age and condition of the polymer and capillary being used, dirty capillary windows, and dirty pump blocks [12]. Background noise may also differ between instruments due to differences in charged couple device (CCD) detectors, laser effectiveness and alignment, and cleanliness and alignment of the optical components [10]. Many amplification-related factors that contribute to background noise (such as analyst skill and stocks of chemicals) are also run-specific and might be reasonably expected to have varying impacts over time.

Limits of Detection and Quantitation

Many analytical disciplines aside from forensic DNA profiling have needed to rigorously account for background noise mixed with low levels of signal [13, 14]. It is common for background noise, such as that associated with DNA testing results, not to be constant and for noise levels to be distributed in a Gaussian fashion that can be effectively characterized with a mean and a standard deviation [11, 13–15]. In such circumstances, the LOD is expressed as a statistical confidence limit of noise error, usually 99.7% (i.e., three standard deviations) or

$$\mathrm{LOD} = \mu_b + 3\sigma_b \tag{1}$$

where μ_b is the average amount of background noise and σ_b is the standard deviation associated with that value [11, 13–15]. A limit of quantitation (LOQ) represents the threshold beneath which measurements of signal strength cannot be reliably used to determine the relative quantity of detected analyte (e.g., because such measurements may include an appreciable amount of signal arising from background noise). LOQ is commonly expressed as the average background signal plus 10 standard deviations [11, 13–15] or

$$\mathrm{LOQ} = \mu_b + 10\sigma_b \tag{2}$$

Forensic DNA testing laboratories routinely test a positive control, negative control, and reagent blank with every DNA analysis run [7–9]. While these controls are utilized primarily as sentinels for gross failures of the DNA testing processes, such as cross contamination of samples, as well as contamination or inappropriate activity of reagents, they also contain an abundance of subtle but important information about

the running environment of the DNA testing system – particularly as it pertains to background noise. These ubiquitous controls can be used to establish objective run-specific electropherogram peak height thresholds that need to be exceeded to minimize the possibility that noise will be considered signal arising from a sample being tested [16].

LOD thresholds have been found to be consistently much lower (often by an order of magnitude) than what testing laboratories had previously established as minimum peak height thresholds [16]. Disregarding information associated with electropherogram peaks well above an analytical threshold of detection (and even above an analytical threshold of quantitation) might be considered abundantly conservative in some circumstances, given that DNA testing is a very sensitive process subject to a variety of technical artifacts such as pull-up, voltage spikes, and stutter. However, in this abundance of caution, valid information about the presence of real DNA peaks is being discarded or ignored – especially concerning given that means of reducing or at least identifying most technical artifacts are known or could be developed.

Technical Artifacts – Stutter

Among the most commonly observed technical artifacts in PCR-based DNA profiling are signals associated with DNA polymerase "stutter" [17]. Stutter artifacts can be particularly problematic in mixed DNA samples where small peaks situated in stutter positions may be interpreted as either being due to stutter or to amplification product corresponding to an allele from a secondary contributor. Stutter products are amplification products that differ in size from a primary peak (corresponding to the actual genomic template) by integral numbers of the underlying core repeat sequence. They result from slippage of *Taq* DNA polymerase on template molecules during amplification: either forward (resulting in $n-4$ stutter or in $n-8$ stutter if two core units are skipped) or backward (resulting in $n+4$ stutter) [18]. Numerous organizations including the Scientific Working Group on DNA Analysis Methods (SWGDAM) [9] and the DNA Advisory Board (DAB) [7] have recommended that forensic DNA testing laboratories conduct validation studies to develop cutoffs or thresholds that allow stutter artifacts to be reliably distinguished from faithfully amplified genomic DNA. The result has been the publication of numerous validation studies that address the prevalence of $n-4$ stutter (i.e., [17]) with relatively consistent observation of stutter artifacts that correspond to 10% or less of the signal strength associated with faithfully amplified templates from the 13 tetranucleotide repeat Combinded DNA Index System (CODIS) loci. Fewer studies on the prevalence of $n+4$ stutter have been conducted but a cutoff of 6% is generally considered to be sufficient to remove more than 95% of $n+4$ stutter peaks.

Technical Artifacts – Pull-up

Amplification products generated during this process are typically labeled with one of three or four different fluorescent dyes to facilitate the examination of several loci simultaneously. Signal associated with one dye color can sometimes give rise to the mistaken perception of signal in a different color in a phenomenon commonly known as *pull-up* or *bleed-through*. Remedies such as shorter injection times or reamplification with smaller amounts of template can usually be invoked to prevent signal strengths from exceeding the 4000 rfu saturation-level threshold. However, situations where these alternatives are not viable (e.g., in mixed samples where the genotype of an unknown secondary contributor is of interest) are commonly encountered in the course of forensic analyses. It is also possible that pull-up artifacts can be associated with primary peaks below saturation thresholds. As a result, significant portions of electropherograms can correspond to regions where legitimate amplification products cannot be reliably distinguished from pull-up artifacts.

Technical Artifacts – Spikes and Blobs

"Spikes" are narrow peaks usually attributed to fluctuation in voltage or the presence of minute air bubbles in the capillary. Spikes are usually seen in the same position in all or most of the dye colors used to generate DNA profile electropherograms. "Blobs" are false peaks thought to arise when some colored dye becomes detached from the DNA and gets picked up by the detector. Blobs are usually wider than real peaks and are typically only seen in one color. Spikes and blobs are not reproducible, which means that if

the sample is run through the genetic analyzer again these artifacts should not reappear in the same place. Hence, the generally accepted way to confirm that a potential spike or blob is an artifact is to rerun the sample. However, analysts often simply rely on their "professional experience" to decide which results are spurious and which are real. This practice can be problematic because no generally accepted objective criteria have yet been established to discriminate between artifacts and real peaks (other than retesting).

References

[1] Findlay, I., Taylor, A., Quirke, P., Frazier, R. & Urquhart, A. (1997). DNA fingerprinting from single cells, *Nature* **389**, 555–556.

[2] Oorschot, R.A.V. & Jones, M.K. (1997). DNA fingerprints from fingerprints, *Nature* **387**, 767.

[3] Thompson, W.C., Ford, S., Doom, T., Raymer, M. & Krane, D.E. (2003). Evaluating forensic DNA evidence: essential elements of a competent defense review. Part 1, *The Champion* **27**(3), 16–25.

[4] Thompson, W.C., Ford, S., Doom, T., Raymer, M. & Krane, D.E. (2003). Evaluating forensic DNA evidence: essential elements of a competent defense review. Part 2, *The Champion* **27**(4), 24–28.

[5] Leclair, B., Sgueglia, J.B., Wojtowicz, P.C., Juston, A.C., Frégeau, C.J. & Fourney, R.M. (2003). STR DNA typing: increased sensitivity and efficient sample consumption using reduced PCR reaction volumes, *Journal of Forensic Sciences* **48**(5), 1001–1013.

[6] Frégeau, C.J., Bowen, K.L., Leclair, B., Trudel, I., Bishop, L. & Fourney, R.M. (2003). AmpFLSTR® profiler plus™ short tandem repeat DNA analysis of casework samples, mixture samples, and non-human DNA samples amplified under reduced PCR volume conditions (25 μL), *Journal of Forensic Sciences* **48**(5), 1014–1034.

[7] DNA Advisory Board (DAB) (2000). Quality assurance standards for forensic DNA testing laboratories, *Forensic Science Communications* **2**(3), 1–14.

[8] Federal Bureau of Investigation (FBI) Laboratory (2005). *National DNA Index System (NDIS) Data Acceptance Standards*, http://forensics.marshall.edu/NEST/Nest%20PDFs/Documents/AppendB-NDIS-0505.pdf.

[9] Scientific Working Group on DNA Analysis Methods (SWGDAM) (2000). Short tandem repeat (STR) interpretation guidelines, *Forensic Science Communications* **2**(3), 1–14.

[10] Moretti, T.R., Baumstark, A.L., Defenbaugh, D.A., Keys, K.M., Brown, A.L. & Budowle, B. (2001). Validation of STR typing by capillary electrophoresis, *Journal of Forensic Sciences* **46**(3), 661–676.

[11] Rubinson, K.A. & Rubinson, J.F. (2000). Sample size and major, minor, trace, and ultratrace components, *Contemporary Instrumental Analysis*, Prentice Hall, Upper Saddle River, pp. 150–158.

[12] Applied Biosystems, Inc. (ABI) (2000). *Chemistry Reference for the ABI Prism® 310 Genetic Analyzer*, Applied Biosystems, Foster City.

[13] Anderson, N. (1989). Determination of the lower limit of detection [Letter], *Clinical Chemistry* **35**, 2152–2153.

[14] Thomsen, V., Schatzlein, D. & Mercuro, D. (2003). Limits of detection in spectroscopy, *Spectroscopy* **18**(12), 112–114.

[15] Arinbruster, D.A., Tillman, M.D. & Hubbs, L.M. (1994). Limit of detection (LOD)/limit of quantitation (LOQ): comparison of the empirical and the statistical methods exemplified with GC-MS assays of abused drugs, *Clinical Chemistry* **40**, 1233–1238.

[16] Gilder, J.R., Doom, T.E. & Krane, D.E. (2007). Run-specific limits of detection and quantitation for STR-based DNA testing, *Journal of Forensic Sciences* **52**(1), 97–101.

[17] Walsh, P.S., Fildes, N.J. & Reynolds, R. (1996). Sequence analysis and characterization of stutter at the tetranucleotide repeat locus vWA, *Nucleic Acids Research* **43**, 854–870.

[18] Butler, J.M. (2001). *Forensic DNA Typing*, Academic Press, San Diego.

Further Reading

Frégeau, C.J., Aubin, R.A., Elliott, J.C., Gill, S.S. & Fourney, R.M. (1995). Characterization of human lymphoid cell lines GM9947 and GM9948 as intra- and interlaboratory reference standards for DNA typing, *Genomics* **28**, 184–197.

DAN E. KRANE

Alterations: Erasures and Obliterations of Documents

Introduction

Forensic document examiners (FDEs) are routinely tasked with the examination of a document to determine if there has been an alteration to an entry or erasures, or to decipher an entry that has been obliterated by overwriting. The documents may be as varied

as a check that has been altered for a different payee and amount, a business contract in which some terms have been changed, or medical records in which the amount of a dosage was changed. The FDE uses techniques that are generally nondestructive, although as a last resort (and with the approval of all parties involved), destructive techniques may be used. Evidence of alteration, obliteration, or erasure is not necessarily pejorative. Changes to documents are often made in the normal course of business and any such evidence found has to be interpreted in the context of the case.

If the FDE finds that a document has been subject to alteration, erasure, or obliteration, his next task is usually to restore and decipher what the original entry was. This is accomplished by visual and instrumental means, though in several cases, no success is gained. ASTM International Standard Guide for the Examination of Altered Documents E2331-04 describes procedures that should be followed by FDEs in examinations involving altered documents. The guide includes some very useful definitions [1].

Equipment

The FDE uses a variety of instruments and lighting sources in the detection of alterations, obliterations, and erasures. Natural light together with low magnification (X5–X40) is used to examine the obverse and reverse sides of the document. The document is examined with backlighting and with oblique (or side) lighting.

The human eye can normally see light wavelengths in the range of 400–700 nm. This is called the *visible light range*. However, alternative wavelengths that are produced using filters and light sources are quite useful in detecting, deciphering, and restoring questioned entries. These wavelengths are in the ultraviolet (UV) range (200–400 nm) and the infrared (IR) range (700–1000 nm). UV and IR wavelengths are used in combination with a variety of filters to examine how papers and inks fluoresce or luminesce.

Specialized instruments have been developed for FDEs and other people who need to look at documents for alterations or security features. Foster & Freeman Ltd. (UK) is a company with offices in the United Kingdom and United States whose video spectral comparator (VSC) range of instruments is used in institutions such as banks to full-service Crime Laboratories. Projectina AG is a Swiss company with a similar line of instruments generally known as the *DocuCenter* or *Docubox*. These instruments allow the examiner to examine documents quickly with a wide range of wavelengths. It allows for easy documentation of evidentiary information resulting from the examinations. Some enterprising FDEs also build their own instruments and get good results [2].

Fuji introduced a UV/IR camera that allows an examiner to take direct digital photographs of effects seen in the UV and IR wavelengths [3]. Dichroic filters are often used to compare inks. The results complement those obtained with the other instruments [4].

Imaging software such as Adobe Photoshop is used to enhance faded entries and also differentiate between inks [5]. The LAB-Color Mode in Photoshop has shown some promise as an analytical tool for ink comparison [6].

Electrostatic detection devices (EDDs) are used to detect indented impressions in paper. There are several suppliers of such instruments such as the electrostatic detection apparatus (ESDA) manufactured by Foster & Freeman Ltd. and the indentation materializer (IMED) made by Kinderprint. These instruments work on the principle of electrostatics. When entries are handwritten on the top page of a writing pad, impressions of the handwriting are made on the pages below. Indentations have been reported to have been developed as far as eight pages down. Sequencing on impressions and ink strokes may disclose evidence that entries were not written in the time frame as they appear.

Research is being conducted using EDDs to detect impressions made by rollers within printers. This may be used to show that a page from a multipage contract was produced by a different printer than that used for the rest of the document [7].

Alterations

Alterations can be classified into two types: additive and subtractive (some authors sometimes present a third type: substitution, which in essence is a subtraction followed by an addition). Additive alterations are made by adding content to a document. The content may be as little as a single pen stroke to change a 1 to a 7, for example, or entire paragraphs or pages of writing or printing that changes the context or meaning of a document. In the case of checks in

which a check writing machine imprinted the amount, the alteration is conducted by simulating the impression of the check writer. This alteration can only "raise" the amount of a check, since the original entry involves perforation of the check stock.

Alterations can be quite subtle and a document such as a medical record can be altered by simply changing a decimal point to indicate a different dosage. This alteration would take place on one page amongst possibly hundreds. The FDE has to employ techniques that include overall visual examination, microscopic examination, alternative light sources with various filters, and careful measurement using specially designed rulers or grids.

Visual examination on the obverse side of a document may disclose entries that are out of alignment. In handwritten documents, this may be reflected in entries that are cramped into a small space and not in keeping with the remaining handwritten entries on the document. In some instances, an alteration can be shown when line crossings are out of the expected sequence. In handwritten documents, the reverse side of the document is examined with oblique lighting to determine if there is any difference in the degree of embossing made by the pen pressure from the obverse side. Differences may indicate that the document was prepared on different writing surfaces and therefore, possibly, at different times.

In typewritten documents, alterations may be detected if different fonts were used. If the alteration was made by addition of text after it was originally typewritten (even using the same typewriter and typing element), the use of special grids can show that the contested text was added later. The increased use of laser and ink-jet printers has made the detection of such additions more difficult. Whereas, in a typewriting case the tolerances would have been one-tenth or one-twelfth of an inch, for a laser printer it is about 1/7200 of an inch. A type of ruler called an *E-ruler*, which is used in the graphics industry, is the measurement tool for the font size and interlineal spacing. The line spacing and paragraph spacing together with the use of capitalized and bold fonts can make a difference in the measurements and conclusion as to whether an alteration by insertion has been made [8].

Occasionally a crude insertion is seen when different printers are used. Under magnification, it can quickly be determined when this is done.

Toner is a resinous material that sits on the paper and has a slight embossed feel. Ink-jet (for the most part) uses liquid ink that soaks into the paper. There are some ink-jet printers that use solid inks. These were introduced in 1991 by Tektronix.

On occasion, a page is substituted into a multipage document such as a will, contract, or medical record, for example. It may be an additional page, or more likely, a page that contains different information than the substituted page. An EDD is essential to these types of cases. Impressions from the pages above the questioned page may not be found on the questioned page as could be expected. Impressions from the questioned page may be found in areas where they should not be. If an impression of a signature and date on a page is found on a page with a later date, this is indicative that the questioned page had been backdated and possibly substituted. If the first page of a document is signed and impressions are found on pages three and four, but not page two, this will raise questions as to the sequence of events in which the document was prepared and signed.

Documents produced by two different printers may exhibit differences in the printing media (see Figure 1). Some toners have magnetic properties and some are nonmagnetic. These properties are determined with the aid of a magnetic viewer [9]. Different ink-jet inks may exhibit dissimilarities under IR wavelengths. Dissimilarities in drop size and



Figure 1 Two entries on the same contract. The majority of the contract was printed with toner (laser printer) (a). One line was added with an ink-jet printer (b) [Reproduced with permission from Rile & Hicks.]

distribution of the ink-jet droplets may also indicate different manufacturers [10].

Documents produced with different photocopiers can sometimes be distinguished by different "trash marks" left on the copies. These marks arise from blemishes on the copying platen, from rollers within the machine, from the drum and charging devices or cleaners [11]. Copies made from the platen and from the automatic document feeder on the same machine may also display different trash marks since the copying mechanism is different for both processes [12]. Different color copiers and certain high-end color laser printers also produce characteristics that allow copies or prints to be identified to a particular machine [13–15]. Thus page substitution can be proved if a different machine is used.

In multipage documents, marks left by items such as staples and paper clips can be of significant evidentiary value to proving page substitution. The number of staples and staple-holes through each page should be the same. Holes that are unaccounted for or not present when they should be, place doubt on the integrity of the page and the document as a whole. Similarly, paper clips or other binding devices and their impressions are useful in exploring the history of how a document was produced.

Additions can be produced by a single ink stroke. A difference in the width of the pen lines could indicate that a different writing instrument was used for the addition. IR reflectance and IR luminescence are very powerful tools in detecting different inks (see Figure 2).

It is important to note that the IR techniques cannot conclude that inks are the same. The technique can only differentiate different inks. Some different inks from different pens will have similar formulations and will give similar results under IR. Further information can be found in **Ink Analysis**.

Raman spectroscopy, thin-layer chromatography, gas chromatography, and liquid chromatography are other techniques that can be used for further ink differentiation. However, the majority of these techniques will involve some destructive testing and are best performed by specialized ink chemists (*see* **Ink Analysis**). The latter specialists can apply ink dating techniques, which may prove that a questioned entry was written at a later date or is contemporaneous with the nonquestioned entries [16, 17]. Security documents normally have specialized antialteration features some of which are incorporated during the manufacture of the paper stock and others during printing [18].

In questions of alteration, the FDE has to be very careful when presented with a photocopy for examination. It is fairly easy to use readily available imaging software to conduct "cut and paste" alterations. The original document is scanned and the image is then manipulated by electronically cutting or adding various portions of text or images [19, 20]. The changed document is then printed and offered as a genuine document. Signatures are probably the most popular choice for an alteration by addition to make a document appear to be genuine. If the original signature is found, it can be proved that it was used as the model for the "cut and pasted' document.



Figure 2 Infrared luminescence showing a check "raised" from 8000 to 80 000. (a) Visible entry on check and (b) entry viewed with infrared luminescence [Reproduced with permission from Foster & Freeman.]

Erasures

There are two types of erasures: mechanical and chemical. A mechanical erasure is one in which the entry is rubbed off using an abrasive substance. Erasure of penciled entries is perhaps the most common form of mechanical erasure. Ordinary black pencil has a marking core containing principally a mixture of graphite and clay. In writing, flakes of graphite form the black stroke. They rub off the point and adhere to the paper surface wedged between the paper fibers. The graphite cannot and does not penetrate the paper fiber as does ink [21]. The hardness of a pencil is a contributory factor in the ease of erasing its line. A soft black pencil leaves an intense black line with little pressure while a hard pencil leaves a lighter trace.

Examination of pencil erasures is time consuming and can involve use of oblique lighting and EDDs. IR reflectance is useful since graphite is opaque to IR and so the pencil traces appear darker. This makes the faint entry easier to read. It is often necessary to photograph the erased area with lighting from each side at a time. These photographs can then be electronically "stitched" to make a composite picture, which shows the entire erased area lighted from all sides. It is a useful technique both for decipherment and court display [22].

Mechanical erasure involves an abrasive rubbing over the paper surface. This results in the paper fibers being disturbed. This is easily seen with oblique light. The use of lycopode spores has also been shown to detect erasures [23]. However, this together with other powder and chemical means should be used as a last resort [21]. Excessive rubbing may wear away the fibers such that the paper in the area of the erasure becomes almost translucent. Backlighting quickly reveals this effect. UV light is also used to detect these erasures as there will be a difference in the reflection between the erased portion and the smooth paper. It should be noted that different erasers also have different effects on the paper. A soft art gum eraser will leave much less trace of an erasure than a typical rubber eraser.

Chemical erasures are normally used to remove ink lines or commercial printing. Typical solvents include bleach, acetone, brake fluid, and other cocktails of chemicals. Quite often, the use of the solvent is detectable by sniffing the document. Some documents have printing that is affected by the use of solvents and this effect will generally be seen very easily. UV light is probably the best means of detecting the stain left by a chemical used for an erasure. In some check-washing cases, there is no patent evidence that the check has been washed [24]. Comparison with other checks from the same check book may show some faintness of color in the washed check. Some security documents have printing that reacts to bleaching agents. This will make the erasure attempt patent. Examination with IR wavelengths may show some traces of the eradicated entries.

In many erasures, it is virtually impossible to decipher the original entries. It is sometimes possible to recover some of the entries by examining the embossing on the back of the document caused by the tip of the pen. It is important to note that when photocopies are examined, evidence of an erasure on the original may not have been reproduced on the copy.

Obliterations

An obliterated entry is obvious to the reader of a document. It may be done in desperation or at leisure at the time of writing. The former is more likely to result in a decipherment of the original entry since a different pen may be used.

Examinations with infrared reflectance (IRR) and infrared luminescence (IRL) are often successful in filtering the obliterating ink to reveal the original (see Figure 3). If the same or similar inks are used, then these techniques may not work. In this case, the FDE can examine the obliterated entry with high magnification and using an enlarged copy, go over the lines that comprise the obliteration. Whatever remains may constitute part or all of the

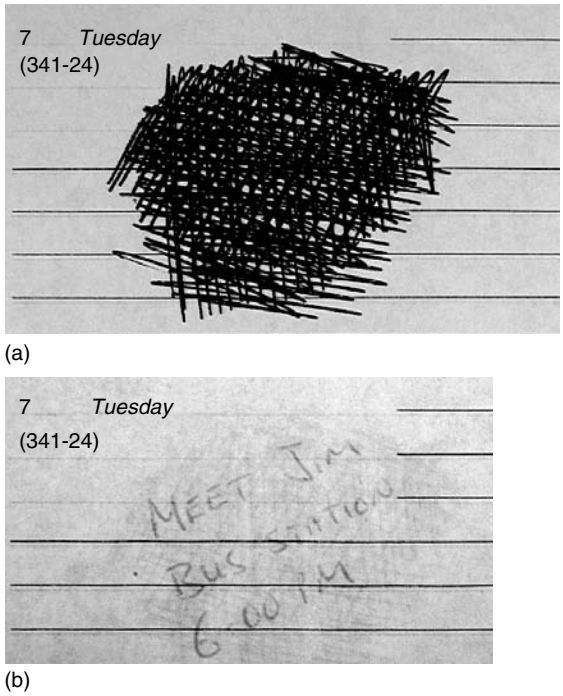


Figure 3 Decipherment of an obliterated entry. (a) Obliterated entry in a diary and (b) original entry deciphered using IRR [Reproduced with permission from Foster & Freeman.]

original entry. This requires some interpretation on the part of the examiner and the result may not be definitive.

A common form of obliteration is conducted with the use of the ubiquitous correction fluid. Sometimes, the obliteration is a cursory pass of the brush with the fluid. In other instances, the entry gets a severe pasting, which results in a thick coat of correction fluid. Examination with backlighting in the first instance is usually sufficient to reveal the obliterated entry. IR radiation and filters can be used to induce luminescence in the ink entry. This luminescence can be observed through the correction fluid. In circumstances where the correction fluid is exceptionally thick, substances such as petroleum ether or a xylene substitute can be used to remove some of the correction fluid [25]. A combination of techniques is generally required to reveal all or a portion of the obliterated entry.

In an examination of a photocopy, the FDE has to look for any evidence of obliterations. If a person carefully uses correction fluid, it may be impossible to tell on a photocopy. Some telltale signs are missing parts of lines or characters in the area of the obliteration. However, the FDE should not confuse normal dropouts caused by the photocopying process as evidence of obliteration.

References

[1] ASTM International Standard Guide for Examination of Altered Documents E 2331-04. (2004).

[2] Drexler, S. & Smith, G. (2002). Ink differentiation for the fiscally challenged, *Journal of the American Society of Questioned Document Examiners* **5**, 20–27.

[3] Richards, G. (1999). Beyond visible light, *Evidence Technology Magazine* **5**, 32–35.

[4] Richards, G. (2003). Dichroic filters: their use in questioned document examination, *Journal of the American Society of Questioned Document Examiners* **6**, 91–96.

[5] Herbertson, G. (2002). *Document Examination on the Computer*, Wideline Publishing, Berkeley, CA.

[6] Hammond, D. (2007). Validation of LAB-Color mode as a non-destructive method to differentiate blue ball pint pen ink. *Paper Presented at the 65th Annual General Meeting of the American Society of Questioned Document Examiners*, Boulder, CO.

[7] LaPorte, G. (2004). The use of an electrostatic detection device to identify individual and class characteristics on documents produced by printers and copiers – a preliminary study, *Journal of Forensic Sciences* **49**, 610–620.

[8] Kelly, J. & Lindblom, B. (eds) (2006). *Scientific Examination of Questioned Documents*, 2nd Edition, Taylor and Francis, Boca Raton, FL, pp. 159–175.

[9] Welch, J. (1985). Magnetic aspects of printing, photocopies and bank cards, *Journal of the Forensic Society* **25**, 343–347.

[10] La Porte, G. (2004). Modern approaches to the forensic analysis of ink jet printing – physical and chemical examinations, *Journal of the American Society of Questioned Document Examiners* **7**, 22–36.

[11] Kelly, J. (1983). *Classification and Identification of Modern Office Copiers*, The American Board of Forensic Document Examiners, Inc., Houston, TX.

[12] Frost, B. & Bell, J. (2007). The role of the slit glass in a photocopy system and its effect on photocopied documents. *Paper presented at the 65th Annual General Meeting of the American Society of Questioned Document Examiners*, Boulder, CO.

[13] Tweedy, J. (2001). Class characteristics of counterfeit protection system codes of color laser copiers, *Journal of the American Society of Questioned Document Examiners* **4**, 53–66.

[14] Li, C.K., Chen, W.C., Cheng, Y.S. & Leung, S.C. (2004). The differentiation of color laser printers, *Journal of the American Society of Questioned Document Examiners* **7**, 105–109.

[15] Li, C.K. & Leung, S.C. (1998). The identification of colour photocopies: a case study, *Journal of the American Society of Questioned Document Examiners* **1**, 8–11.

[16] Aginsky, V. (2006). Using TLC and GC-MS to determine whether inks came from the same manufacturing batch, *Journal of the American Society of Questioned Document Examiners* **9**, 19–27.

[17] Roux, C., Novotny, M., Evans, I. & Lennard, C. (1999). A study to investigate the evidential value of blue and black ballpoint inks in Australia, *Forensic Science International* **101**, 167–176.

[18] Ware, C. (2003). A new check security feature: Thermochromic ink, *Journal of the American Society of Questioned Document Examiners* **6**, 34–37.

[19] Hicks, F. (1995). Computer imaging for questioned document examiners: the benefits, *Journal of Forensic Sciences* **40**, 1045–1051.

[20] Hicks, F. (1995). Computer imaging for questioned document examiners: the potential for abuse, *Journal of Forensic Sciences* **40**, 1052–1054.

[21] Hilton, O. (1991). *Detecting and Deciphering Erased Pencil Writing*, Charles C. Thomas, Springfield, IL.

[22] Mohammed, L. & Williams, D. (2006). Using Adobe Photomerge® to prepare demonstration charts. *Poster presentation at the American Academy of Forensic Sciences Conference*, Seattle, WA.

[23] Ellen, D. (2006). *Scientific Examination of Documents: Methods and Techniques*, 3rd Edition, Taylor and Francis, Boca Raton, FL. p. 185.

[24] Kelly, J. & Lindblom, B. (eds) (2006). *Scientific Examination of Questioned Documents*, 2nd Edition, Taylor and Francis, Boca Raton, FL. p. 320.
[25] Kelly, J. & Lindblom, B. (eds) (2006). ibid. p. 332.

LINTON A. MOHAMMED

Alterations in Documents: Detection of ***see*** Alterations: Erasures and Obliterations of Documents

Alternative Specimens: Hair ***see*** Hair: Toxicology

Alternative Specimens: Oral Fluid ***see*** Oral Fluid Toxicology

Alternative Specimens: Sweat ***see*** Sweat: Toxicology

American Law Institute Standard of Insanity ***see*** Insanity: Defense

Amnesia ***see*** Dissociative Disorders

Amok: Running ***see*** Homicide: Multiple (Behavior)

Amphetamine

Introduction

The amphetamines represent a family of drugs that stimulate the central nervous system (CNS) and other parts of the body to produce euphoria and increased vigor and alertness. They are related chemically to the naturally occurring substance, ephedrine, which is found in *Ephedra* species. Related stimulants are also found in khat (*Catha edulis* grown in the horn of Africa and the Middle East, containing cathinone and other alkaloids), and mescaline. Mescaline is the active alkaloid in the peyote cactus (*Lophophora williamsii*) from Northern Mexico and Southern USA and other cactus species growing in subtropical and temperate areas of South America.

Pharmacologically, the amphetamines are all related to the hormone adrenaline (epinephrine) and the neurotransmitter noradrenaline (norepinephrine). Figure 1 illustrates the similarities of the naturally occurring stimulants ephedrine, cathinone, and mescaline as well as the hormones adrenaline and noradrenaline with amphetamine.

Some amphetamines are available legally, i.e., amphetamine for use in narcolepsy and attention deficit hyperactivity syndrome (ADHD); however, the majority of these drugs that are used in the community are from illegal sources. The amphetamines are produced in clandestine laboratories either from simple chemical precursors or from legally available drugs such as ephedrine and pseudoephedrine. The synthesis of designer drugs containing methylendioxy moiety is usually done using safrole (extracted from sassafras oils) as precursor.

Types of Amphetamines

Amphetamine, or sometimes known as *dex-amphetamine,* is the prototype drug in this class of drugs. Modifications on the benzene ring and nitrogen

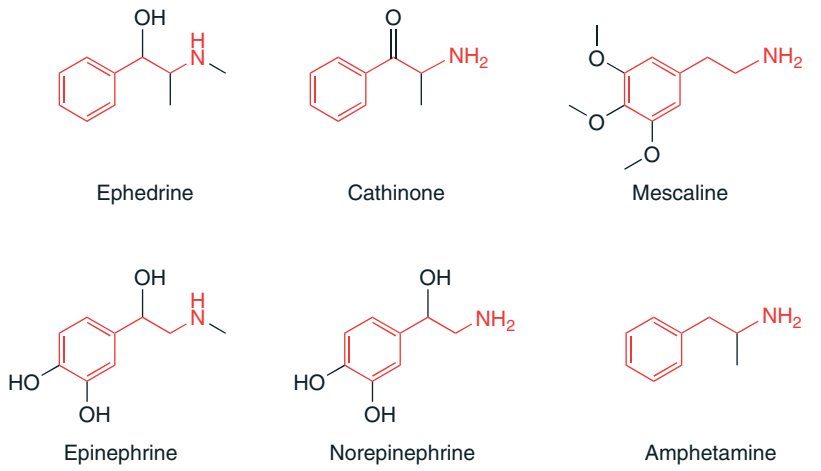


Figure 1 Chemical structures of selected stimulants and hormones

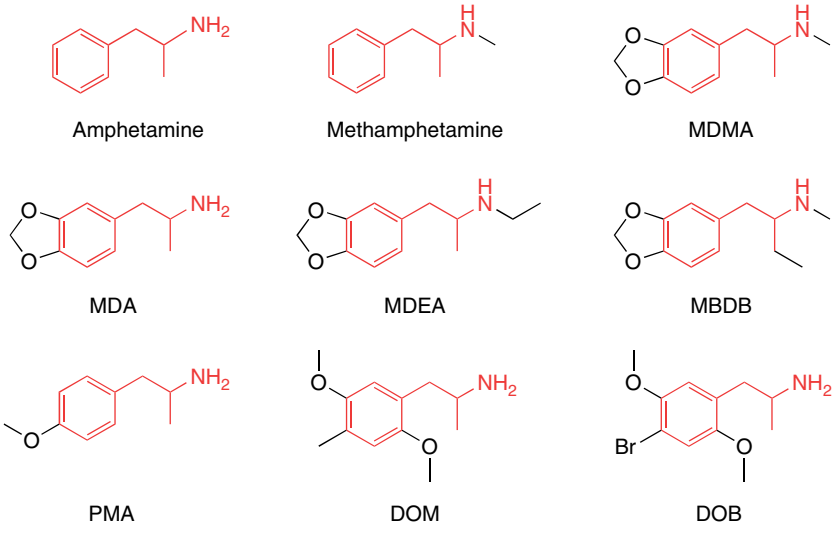


Figure 2 Chemical structures of selected stimulants. DOM, 4-methyl-2,5-dimethoxyamphetamine; DOB, 4-bromo-2,5-dimethoxy amphetamine; MDA, 3,4-methylenedioxyamphetamine; MDMA, 3,4-methylenedioxymethamphetamine; MDE, 3,4-methylenedioxyethylamphetamine; MBDB, N-methyl-benzodioxazoylbutanamine; PMA, 4-methoxyamphetamine

end have led to numerous "designer" stimulant drugs. The most common is methamphetamine (methylamphetamine or "speed"), which is the N-methylated form of amphetamine. Ring substitution with methylendioxy moiety leads to analogs including 3,4-methylenedioxy-methamphetamine (MDMA or ecstasy), 3,4-methylenedioxyamphetamine (MDA), 3,4-methylenedioxyamphetamine (MDE, or "eve"), or *N*-methyl-benzodioxazoylbutanamine (MBDB). Ring substitution with lipophilic groups (like methoxy groups) leads to analogs including 4-methoxyamphetamine (PMA), 4-methyl-2,5-dimethoxyamphetamine (DOM), or 4-bromo-2,5-dimethoxyamphetamine (DOB). Figure 2 illustrates the similarity in the chemical structures of few selected stimulants related to amphetamine.

Drugs with the dextro configuration (D) are more active than those with the levo (L) configuration.

For example, L-methamphetamine is less active than D-methamphetamine and is used as a decongestant in some countries (as L-desoxyephedrine), while the D-isomer is the abused form of the drug.

Abuse of Amphetamines

Amphetamines are arguably one of the most abused drugs. Amphetamine is the predominate member of this class abused in Western Europe while methamphetamine is the predominate member in Eastern Europe, Asia, Australia, and North America.

MDMA and related designer amphetamines are common in the night clubs worldwide due to their ability to promote empathy and a sense of well being. Ecstasy tablets are primarily composed of MDMA but can also contain other stimulants such as methamphetamine.

The United Nations Office on Drug Control (UNODC) estimates that approximately 34 million people consume amphetamines annually with 8 million consuming ecstasy [1]. The prevalence of use varies substantially from region to region and between types of amphetamines. However, the proportion of the population using the drug that is over 15 years of age generally ranges from about 1–4% with much higher numbers in younger males.

Amphetamines tablets come in several different forms, sizes, colors, and shapes [2]. Apart from the different appearances of the tablets, amphetamines can be used in different forms, as hydrochloride salt or as free base. The form used illustrates the common street names and the common way of intake of the drug. For example, methamphetamine used as hydrochloride salt is usually injected or taken orally and is known as *speed or ice*. When methamphetamine is used as a free base, it is often smoked and is known as *base or crystal meth*.

These products can be manufactured locally or imported through a criminal network. Low purity forms (powders and tablets) of methamphetamine are usually administered by snorting or injecting, and can be mixed with other drugs such as ketamine. A gluggy, pasty, or oily form is often brown or yellow due to the presence of iodine and other organic impurities, and crystal meth or ice is high purity, crystalline methamphetamine that comes in the form of large translucent/white crystals that are usually smoked or injected. Free base methamphetamine is required in higher doses compared to other forms and this in turn has a greater risk of harm and prospects of developing serious addiction [3].

Pharmacology and Disposition

Mechanism of Action and Effects

Amphetamines interact with noradrenaline- (norepinephrine-) and dopamine-containing nerves to either replace the monoamine neurotransmitter in the nerve ending and act as a false transmitter, or to facilitate the release of neurotransmitters from nerve endings. Mood and behavior can be affected by some amphetamines (e.g., MDMA) due to the release of serotonin at nerve endings in the brain. The sudden "rush" or "high" experienced with intravenous injections, or smoking, is caused by the sudden release of noradrenaline and related monoamines from nerve endings [4].

Amphetamines stimulate the heart and cause irregular rhythm. Increases in heart rate also occur. Amphetamines produce CNS effects such as movement disorders (locomotor activity) and can produce a stereotyped behavior. This behavior will manifest itself as repeated movement or agitation, unusual facial expressions or grimacing, pacing, grooming, etc. Overt aggression and a feeling of increased strength is also a frequent serious side effect of amphetamine use, particularly with repeated use.

Amphetamines can reduce fatigue but this is short-lived as tolerance sets in with repeated use leading to hypersomnolence. This is a serious risk factor in long-distance truck drivers using amphetamines to ward off the effects of long driving hours.

Chronic use of amphetamines can produce psychotic conditions, which is often associated with violent and irrational behavior.

Potency and Duration of Action

As one would expect, the various amphetamines have different potencies to cause an effect, and consequently each amphetamine is associated with a different range of common doses. These are tabulated in Table 1. Chronic abuse invariably involves higher doses than initiation doses in naïve users.

Depending on the dose and the type of amphetamine, the effects can last for at least one-half day and up to more than two days when higher

Table 1 Common doses for selected stimulants

Drug	Dose range (mg) [a]	Duration of action [b]	Common blood concentrations (mg l^{-1})[c]
Amphetamine	5–60	Half to one day	<0.2
Cathinone	50 or more	Several hours	<0.2
Ephedrine	to 240	Half day	up to 1
Methamphetamine	5 or more	Up to two days	<0.5
MDMA	50–150	Half to one day	<0.5
MDA	50–250	Half to one day	<0.5
MDE	50–250	Half to one day	<0.5
PMA	50–100	Half to one day	<0.5

MDA, 3,4-methylenedioxyamphetamine; MDMA, 3,4-methylenedioxymethamphetamine; MDE, 3,4-methylenedioxyethylamphetamine; PMA, 4-methoxyamphetamine

[a] Usual dose range

[b] Common duration of action (dependent on dose, pH of urine and a range of physiological factors)

[c] Common blood concentrations (after ingestion of usual dose)

doses are used or when excretion is reduced due to high acidity in the urine (Table 1). Acid urine can be caused by dietary influences or by the use of drugs that cause higher acidity.

The stimulant group of drugs, including legal stimulants and designer amphetamines, are generally well absorbed orally. This route tends to be the major route of administration. Peak concentrations of amphetamines are generally within 1 to 2 h postdose when taken orally.

Amphetamines are also injected intravenously, taken by nasal insufflation (snorting), or by smoking. Intravenous dosing provides the most efficient and quickest way of administering the drug. Amphetamines are smoked using a glass tube that is heated to volatilize the drug. Smoking or intravenous administration will produce much higher initial blood concentrations than oral administration.

Excretion

There are often substantial amounts of unchanged amphetamines excreted into urine. Hence urine is a useful screening specimen to determine the past use of these drugs.

Amphetamines are metabolized into substances that can be used to identify the type of amphetamine used, and if necessary targeted by analyses to increase the likelihood of detection. For example, amphetamine is hydroxylated on the α-carbon (carbon adjacent to ring) to norephedrine, which is in turn converted to benzoic acid.

Methamphetamine is metabolized to amphetamine (about 10%), which in turn is broken down further. MDMA is metabolized to another designer amphetamine, MDA, and to the dihydroxy metabolite (O-desmethylation). Consequently, for both methamphetamine and MDMA, the corresponding *N*-desmethyl metabolite is also targeted in the analyses. PMA is converted to the 4-hydroxy form and metabolized at the α-hydroxy to the corresponding norephedrine. Cathinone is metabolized by the oxidation of the keto group on the side chain to norephedrine and norpseudoephedrine.

The antiparkinson drug selegiline is metabolized to the weakly active L-isomer of methamphetamine [5]. The formation of amphetamines from legal substances applies to related drugs, e.g., fenethylline, clobenzorex, mefenorex, fenproporex, and benzphetamine. In these situations it will be necessary to use chiral assays to discriminate the isomers (dextro- or levo-) and/or other metabolites to determine if the source of the amphetamine is legal or illegal.

Toxicity

Life-threatening adverse reactions of amphetamines are well described. Initially, these can include agitation, fever, aggression, and violence, and can rapidly

lead to very high body temperatures, with elevated heart rate and blood pressure. Potentially fatal consequences of these effects can be rhabdomyolysis (liquefaction of muscles finally leading to disruptions in heart rhythm caused by hyperkalemia), intravascular coagulation, convulsions, profuse sweating, and hemorrhages in the heart and blood vessels.

Death resulting from excessive water consumption (caused by hyperthermia) following ingestion of ecstasy tablets at a nightclub presents another mode of death due to the use of amphetamines, particularly ecstasy-like amphetamines [6, 7]. PMA is a more toxic member, and while rarely seen has been linked to clusters of deaths [8, 9]. This may be caused by the phenotype of the enzyme CYP P450 2D6, which catalyzes the major metabolic step of PMA metabolism (*see* **Pharmacogenomics**). There is little correlation between concentration of amphetamines in blood and a likely toxic reaction [10]. Tolerance, route of administration, chronicity of use, and individual sensitivities regulate any unwanted responses to these dangerous drugs.

Prolonged use of amphetamines, particularly MDMA, cause long-lasting or even permanent nerve damage to dopamine- and serotonin-containing nerves in the brain leading to stereotyped behaviors and paranoid psychosis. Long-term use of amphetamines is also associated with damage of heart muscle that can manifest as cardiac arrhythmias leading to sudden death, although this may not be obvious in an external examination. Sudden death usually occurs only when preexisting heart disease is present or owing to brain hemorrhage that is induced by high blood pressure [11].

Interpretation of Results

As with other drugs of abuse seen in forensic cases, the interpretation of toxicology results of amphetamines requires a detailed understanding of the context in which the drug detection has occurred. This requires an understanding of the form of the drug used and its route of ingestion (e.g., intravenous or oral etc), whether the drug has been used in a continuing manner (e.g., acute or chronic use), and the circumstances of the case.

It would not be surprising to hear of a person suffering a "meltdown" at a night club (very high core body temperature) under the influence of MDMA or another designer amphetamine has been detected; however, such an adverse reaction does not occur in everyone using this drug even at blood concentrations reported of people exhibiting such toxicity.

Similarly, low concentrations of methamphetamine in long-distance truck drivers can lead to hypersomnolence and cause fatigue-based crashes after the strong stimulant effects of the drug have largely worn off, particularly in persons using the drug regularly. However, low concentrations may not have this effect in situations where single doses may have been used and no tolerance to the drug has developed.

Accordingly, drug concentration data in blood, and indeed other specimens, needs to be carefully evaluated based on the available circumstances, the known disposition of the drug, and of course on the actual detected amphetamine. It is equally true with amphetamine users as with other drug users that other drugs are often used concomitantly. These drugs may include other stimulants, cannabis, benzodiazepines, or even opiates. Multiple drug use can increase the risk of adverse reactions and toxicity. It is therefore essential that any drug screening encompasses a range of likely drugs (*see* **Postmortem Toxicology: Interpretation**).

Methods of Analysis

The class of drugs is readily detected by standard toxicology methods. Immunoassays are commonly used for initial testing since they require little or no sample preparation and are relatively cheap. Immunoassays are normally targeted to detect those members related to amphetamine (no substitution on the nitrogen including MDA) and those related to methamphetamine (N-substituted analogs including MDMA, MDE). These tests can be used on blood and urine and even oral fluid (saliva) [12, 13] (*see* **Toxicology: Initial Testing**; **Confirmation Testing: Toxicology**; **Oral Fluid Toxicology**).

In forensic matters confirmatory testing is usually conducted by gas chromatography followed by detection using mass spectrometry (GC–MS). Most of the amphetamines show poor mass spectral definition when underivatized, and hence it is common to chemically derivatize extracts with acetyl or perfluoracyl derivatives [14, 15]. More recently, liquid chromatography coupled to tandem mass spectrometry (LC–MS/MS) is being increasingly used as this

technique does not require chemical derivatization. Amphetamines are isolated from biological matrices either by liquid–liquid or by solid-phase extraction procedures [16, 17].

As mentioned above, some legal substances like selegiline or clobenzorex are metabolized to the weakly active L-isomer of methamphetamine. For differentiation of the intake of these substances from an illegal abuse of methamphetamine (D-isomer), chiral analysis is required. Methods for determination of isomer ratios have been described using GC–MS and LC–MS/MS. These techniques have been adapted for biological specimens such as blood, urine, saliva, and even hair [18–20] (*see* **Postmortem Toxicology: Laboratory Analysis**).

References

[1] Programme, U.N.D.C. (1996). *Amphetamine-type Stimulants: A Global Review*, *UNCDP Technical Series Number 3*, United Nations International Drug Control Programme, Vienna.

[2] Degenhardt, L. & Topp, L. (2003). Crystal meth use among polydrug users in Sydney's dance party subculture: characteristics, use patterns and associated harms, *International Journal of Drug Policy* **14**, 17–24.

[3] Morgan, P. & Beck, J.H. (1997). Hidden contexts of methamphetamine use in the United States, in *Amphetamine Misuse: International Perspectives on Current Trends*, H. Klee, ed, Harwood Academic Publishers, pp. 135–162.

[4] Drummer, O.H. & Odell, M. (eds) (2001). *The Forensic Pharmacology of Drugs of Abuse*, Arnold, p. 462.

[5] Romberg, R.W., Needleman, S.B., Snyder, J.J. & Greedan, A. (1995). Methamphetamine and amphetamine derived from the metabolism of selegiline, *Journal of Forensic Sciences* **40**(6), 1100–1102.

[6] Schifano, F., Oyefeso, A., Corkery, J., Cobain, K., Jambert-Gray, R., Martinotti, G. & Ghodse, A.H. (2003). Death rates from ecstasy (MDMA, MDA) and polydrug use in England and Wales 1996–2002, *Human Psychopharmacology* **18**(7), 519–524.

[7] Milroy, C.M., Clark, J.C. & Forrest, A.R. (1996). Pathology of deaths associated with "ecstasy" and "eve" misuse, *Journal of Clinical Pathology* **49**(2), 149–153.

[8] Ling, L.H., Marchant, C., Buckley, N.A., Prior, M. & Irvine, R.J. (2001). Poisoning with the recreational drug paramethoxyamphetamine ("death"), *The Medical Journal of Australia* **174**(9), 453–455.

[9] Felgate, H.E., Felgate, P.D., James, R.A., Sims, D.N. & Vozzo, D.C. (1998). Recent paramethoxyamphetamine deaths, *Journal of Analytical Toxicology* **22**(2), 169–172.

[10] Jones, A.W. (2007). Age- and gender-related differences in blood amphetamine concentrations in apprehended drivers: lack of association with clinical evidence of impairment, *Addiction* **102**(7), 1085–1091.

[11] Karch, S.B., Stephens, B.G. & Ho, C.H. (1999). Methamphetamine-related deaths in San Francisco: demographic, pathologic, and toxicologic profiles, *Journal of Forensic Sciences* **44**(2), 359–368.

[12] Apollonio, L.G., Whittall, I.R., Pianca, D.J., Kyd, J.M. & Maher, W.A. (2007). Matrix effect and cross-reactivity of select amphetamine-type substances, designer analogues, and putrefactive amines using the Bio-Quant direct ELISA presumptive assays for amphetamine and methamphetamine, *Journal of Analytical Toxicology* **31**(4), 208–213.

[13] Laloup, M., Tilman, G., Maes, V., De Boeck, G., Wallemacq, P., Ramaekers, J. & Samyn, N. (2005). Validation of an ELISA-based screening assay for the detection of amphetamine, MDMA and MDA in blood and oral fluid, *Forensic Science International* **153**(1), 29–37.

[14] Kudo, K., Ishida, T., Hara, K., Kashimura, S., Tsuji, A. & Ikeda, N. (2007). Simultaneous determination of 13 amphetamine related drugs in human whole blood using an enhanced polymer column and gas chromatography-mass spectrometry, *Journal of Chromatography B, Analytical Technologies in the Biomedical and Life Sciences* **855**(1), 115–120.

[15] Peters, F.T., Schaefer, S., Staack, R.F., Kraemer, T. & Maurer, H.H. (2003). Screening for and validated quantification of amphetamines and of amphetamine- and piperazine-derived designer drugs in human blood plasma by gas chromatography/mass spectrometry, *Journal of Mass Spectrometry* **38**(6), 659–676.

[16] Skrinska, V.A. & Gock, S.B. (2005). Measurement of 3,4-MDMA and related amines in diagnostic and forensic laboratories, *Clinical Laboratory Science* **18**(2), 119–123.

[17] Maurer, H.H. (2005). Multi-analyte procedures for screening for and quantification of drugs in blood, plasma, or serum by liquid chromatography-single stage or tandem mass spectrometry (LC-MS or LC-MS/MS) relevant to clinical and forensic toxicology, *Clinical Biochemistry* **38**(4), 310–318.

[18] Paul, B.D., Jemionek, J., Lesser, D., Jacobs, A. & Searles, D.A. (2004). Enantiomeric separation and quantitation of (+/−)-amphetamine, (+/−)-methamphetamine, (+/−)-MDA, (+/−)-MDMA, and (+/−)-MDEA in urine specimens by GC-EI-MS after derivatization with (R)-(−)- or (S)-(+)-alpha-methoxy-alpha-(trifluoromethy)phenylacetyl chloride (MTPA), *Journal of Analytical Toxicology* **28**(6), 449–455.

[19] Peters, F.T., Samyn, N., Kraemer, T., Riedel, W.J. & Maurer, H.H. (2007). Negative-ion chemical ionization gas chromatography-mass spectrometry assay for enantioselective measurement of amphetamines in oral fluid: application to a controlled study with MDMA and driving under the influence cases, *Clinical Chemistry* **53**(4), 702–710.

[20] Phinney, K.W. & Sander, L.C. (2004). Liquid chromatographic method for the determination of enantiomeric composition of amphetamine and methamphetamine in hair samples, *Analytical and Bioanalytical Chemistry* **378**(1), 144–149.
[21] Logan, B.K. (2001). Amphetamines: an update on forensic issues, *Journal of Analytical Toxicology* **25**(5), 400–404.
[22] Maurer, H.H., Bickeboeller-Friedrich, J., Kraemer, T. & Peters, F.T. (2000). Toxicokinetics and analytical toxicology of amphetamine-derived designer drugs ('Ecstasy'), *Toxicology Letters* **112–113**, 133–142.
[23] Baselt, R.C. (ed) (2005). *Disposition of Toxic Drugs and Chemicals in Man*, 5th Edition, Year Book Medical Publishers.
[24] Baselt, R.C. (2004). *Disposition of Toxic Drugs and Chemicals in Man*, 7th Edition, Year Book Medical Publishers.

OLAF H. DRUMMER, JOCHEN BEYER AND DIMITRI GEROSTAMOULOS

Amplified Fragment Length Polymorphism

Introduction

The tandemly repeated sequences of human genome, especially the hypervariable minisatellites, are suitable markers for forensic identification and population studies due to their highly polymorphic and multiallelic nature. Minisatellite loci are also known as *VNTRs*, a name that appropriately describes their structural conformation. When first used in forensic science, variation at VNTR loci was typically detected by restriction fragment length polymorphism (RFLP). Following the advent of the PCR, small VNTR loci analysed following PCR amplification were referred to as *amplified fragment length polymorphisms* (*ALFPs* or *AmpFLPs*). An example of an AmpFLP that was widely used in forensic science is HumD1S80 (pMCT118, GenBank accession number D28507). D1S80 is a VNTR locus located on chromosome 1p35–36 with a repeat unit of 16 bp [1, 2]. The number of repeats at D1S80 has been reported to vary from 13 to >50 [2]. The D1S80 minisatellite is a well-studied marker that has been used worldwide to differentiate genetic affinities among and between various populations [3, 4]. In forensic casework, AmpFLPs such as D1S80 have been successfully used to identify individuals [5, 6] and human remains [7] and also to investigate paternity [8]. Data on allele frequency distributions are necessary to support forensic analyses, so much of the population genetic data available emerged from forensic validation studies.

References

[1] Nakamura, Y., Carlson, M., Krapcho, K. & White, R. (1988). Isolation and mapping of a polymorphic sequence (pMCT118) on chromosome 1p (D1S80), *Nucleic Acids Research* **16**, 9364.
[2] Kasai, K., Nakamura, Y. & White, R. (1990). Amplification of a variable number of tandem repeats (VNTR) locus (pMCT118) by the polymerase chain reaction (PCR) and its application to forensic science, *Journal of Forensic Sciences* **35**, 1196–1200.
[3] Duncan, G., Thomas, E., Gallo, J.C. & Baird, L.S. (1996). Human phylogenetic relationships according to the D1S80 locus, *Genetica* **98**, 277–287.
[4] Walsh, S.J. & Eckhoff, C.I. (2007). Australian aboriginal population genetics at the D1S80 VNTR locus, *Annals of Human Biology* **34**, 557–565.
[5] Budowle, B., Chakraborty, R., Giusti, A.M., Eisenberg, A.J. & Allen, R.C. (1991). Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE, *American Journal of Human Genetics* **48**, 137–144.
[6] Gross, M., Carmody, G. & Guerrieri, R.A. (1997). Validation studies for the genetic typing of the D1S80 locus for implementation into forensic casework, *Journal of Forensic Sciences* **42**, 1140–1146.
[7] Fujita, Y., Kubo, S.-I., Tokunaga, I., Kitamura, O., Gotohda, T. & Ishigami, A. (2004). Influence of postmortem changes on DNA typing (D1S80, TH01, HLA DQA1, and PM typing system): case studies for personal identification, *Legal Medicine* **6**, 143–150.
[8] Tamaki, K., Huang, X.L., Yamamoto, T., Uchihi, R., Nozawa, H. & Katsumata, Y. (1995). Applications of minisatellite variant repeat (MVR) mapping for maternal identification from remains of an infant and placenta, *Journal of Forensic Sciences* **40**, 695–700.

SIMON J. WALSH

Amytal Interviews *see* Deception: Truth Serum

Analysis: Alcohol ***see*** Alcohol: Analysis

Analysis: Amphetamine ***see*** Amphetamine

Analysis: Benzodiazepines ***see*** Benzodiazepines

Analysis: Bloodstain Spatter/Pattern ***see*** Bloodstain Pattern Interpretation

Analysis: Bomb Pulse ***see*** Bomb-Pulse Dating

Analysis: Cannabis ***see*** Cannabis

Analysis: Cell Site ***see*** GSM Analysis and PDAs

Analysis: Cocaine ***see*** Cocaine

Analysis: Computer Network

Background

Criminals can use the Internet for a variety of purposes including distributed denial-of-service (DOS) attack, other attack or compromise of computer systems, identity theft, communication with accomplices, information gathering, illicit transfer of stolen intellectual property and/or contraband, and various forms of theft (services, goods, and funds). These activities can often be identified using network analysis. While media analysis is primarily concerned with the identification, retrieval, and recovery of data stored on a variety of media, network analysis is concerned with data transmitted from one computer to another (*see* **Computers**). The dominant networking protocol of the Internet is Internet Protocol (IP) and associated protocols are Transmission Control Protocol (TCP), and User Datagram Protocol (UDP). An excellent summary of these protocols is defined in Request For Comments (RFCs) 1180 [1]. TCP/IP works by encapsulating blocks of data into "packets". These packets have two parts: the "payload" or data portion; and the packet header, containing information needed by the TCP/IP itself, such as, source and destination IP addresses and port numbers. Owing to the nature of the TCP/IP protocol, there are several intermediaries (commonly known as *hops*) between the traffic source and its destination. This article identifies challenges to network analysis, common roles of compromised systems, and tools and techniques to conduct network analysis.

Challenges in Network Analysis

There are challenging aspects of network analysis, in addition to the challenges of media analysis. The principal challenges are as follows: data pertaining to an investigation is not located on one system but across multiple systems and attackers attempt to obfuscate the true source of network traffic. During preliminary stages of an investigation, the exact nature of the event may be unknown. In this case,

the investigator must consider any traffic that appears out of the ordinary.

Multiple Data Sources

Examiners usually cannot evaluate traffic in real time, but are limited to reviewing recorded traces of past activity. The examiner often must retrieve data from multiple sources, and attempt to reconstruct a plausible chain of events. The collection effort may include data from systems belonging to many different companies or organizations. In some instances, it may be necessary to obtain the cooperation of foreign governments if the network trail crosses national boundaries.

Lack of Collection Control. Network analysis is complicated by the fact that the examiner usually has little or no control over the way the data was initially collected. For example, different Internet service providers (ISPs) will have different equipment and logging mechanisms, and so the type and format of traffic logs will likely differ greatly. As a result, the examiner must deal with data in many formats including firewall logs; router logs; intrusion detection system (IDS) logs; network captures from network sniffers; binary, hexadecimal, or ASCII data formats; host-based logs (e.g., tcp wrappers, mail logs, telnet/ftp logs, system logs); and intruder network tools (e.g., attacker network captures, IP addresses of distributed DOS master servers). Logs created by similar products produced by different vendors are often dissimilar.

Incomplete Logs. Logging mechanisms are often configured poorly, disabled, or simply not capable of generating the detailed data desired. Some devices are capable of different levels of logging, and although capture of the desired information is possible, it may not be configured to do so. For example, many routers only log packets that are blocked. In the event of traffic of interest being passed, no record would exist in that router log. Owing to the large storage volume and performance impacts, many administrators enable detailed logging only when investigating a specific problem. In other cases, the device is not capable of logging the level of detail desired. A router's function is to forward the IP traffic, and its logs contain little or no insight about the protocols or data higher up in the network stack.

Falsified Logs. Any user with root (or administrative privileges) on a host has the potential to alter system logs. If an attacker has gained root privileges, logs may have been sanitized; that is, log entries may have been modified or deleted altogether. This is not an uncommon practice used to mask access to a compromised host. For example, the UNIX syslog program records user logins, including connections from other hosts. After gaining root privileges, an attacker could attempt to modify syslog (and possibly other logs) to remove relevant connections to the host. Similarly, an attacker who compromised a web server may attempt to erase records of network traffic from the server logs.

Source Address Obfuscation

Because of the way the TCP/IP protocol is implemented, we often cannot be certain that the source IP address of a monitored session is the true origin of the traffic in question. There are several techniques available to mask the true identity of the source host or make it difficult to actually locate the attacker.

Use of Open Access Points. With the proliferation of public Internet access points, many of them are used for nefarious purposes. These facilities can be public access points on a wired network as well as wireless access points (WAPs). While some control measures can be taken at these access points, many are left with minimal controls. Additionally, many WAP devices have minimal logging facilities and, in many cases, simple security measures (such as enabling encryption on the device) are not taken. Users can successfully connect from a distance, depending upon the location of the WAP device, the wireless card in the computer, external antenna usage, and intervening structural elements. Even WAP for home use are abused in this way, especially in areas where density is higher, like business districts, town houses, apartment buildings, and condominiums. Attackers can frequently change their IP addresses by utilizing multiple access points, making tracking the subject a difficult proposition.

IP Spoofing. IP spoofing is a technique in which an attacker sends traffic to a target while making it appear to have originated from a third party. This is common in DOS attacks, where there is no

need or desire to receive any response from the target. It is also used for network reconnaissance to mask the true source of probes. Determining the true origin of spoofed traffic requires tracing the path back through every intermediary router between destination and source, and may require the cooperation of multiple ISPs, telecom providers, and possibly foreign governments.

Proxy Servers and Anonymizer Services. Proxies are legitimately used by organizations to restrict access to needed Internet services while protecting the identity and security of internal hosts. The source address of traffic originating behind the proxy is replaced with that of the proxy. While a proxy server can provide the actual source address of the traffic, many are configured to hide this information. Identification of the true source host will require an examination of proxy server logs. Some proxies contain vulnerabilities or are poorly configured, resulting in use or control by unauthorized parties. Some proxies are actually networks, such as The Onion Routing (TOR) proxy, and have multiple entry and exit points, making tracing connections nearly impossible [2]. An anonymizer is a generic term used to describe some service that shields information about a user or a system. Anonymizer services act as proxies for common Internet services but provide additional functions, such as blocking of Hypertext Transfer Protocol (HTTP) cookie requests, providing false login credentials (pseudonyms), e-mail forwarding, and data encryption. Attackers can use such services to hinder identification.

Compromised Host Network. Attackers also mask their true origin by using a network of compromised hosts. In this method, the attacker never connects directly to his victim, but rather he uses one or more previously compromised hosts as intermediaries. In such cases, the source address of the attacking host is correct but the attacker is directing the attack from another host, and the intermediary host is simply used as a platform for the attack. To locate the true source of the attacker, the investigator must analyze the logs of one or more intermediary network systems, as the compromised system(s) logs may be altered or deleted. Attackers often share information on compromised hosts, so an attacker found to be currently using any given host may not be the one that originally gained access. A single compromised host may be used by multiple attackers at any given time. The discovery of many such hosts could cause the scope of an investigation to expand very rapidly.

Summary of Challenges

Certainly, the challenges listed are not all inclusive, but rather are designed to prompt the examiner to consider the types of problems that are encountered. Although there are many challenges, the situation is not hopeless. While the data provided by a single set of logs might be inadequate, data provided by multiple correlated logs may reveal the required information. For example, an attack on a web server leaves log traces on the web server, corporate firewall, and one or more intermediate routers. When malicious activity is ongoing, configuration changes to deployed systems will allow collection of additional information. Deployment of additional equipment or software tools on an appropriate network can collect relevant data. In most operating systems (OS) there are several logs that may record evidence of malicious activity. In any given incident, none, some, or all of these logs may have been modified. Not all attempts to modify logs are successful, and most modifications are easily detected. Many systems send log data to separate computer systems, which would store significant evidence for an examiner. The skill level of attackers varies widely such that even if some logs have been tampered with, useful information might still be obtained from other log files. The manner and level to which a compromised host has been modified may, in fact, help the examiner gauge the expertise, tools, and methods used by the attacker. Other systems may collect data about the attack and/or follow-on traffic, which the attacker is unable to manipulate, or else may be unaware.

Compromised Systems in Network Investigations

There are a number of ways that compromised systems may be used by an attacker, and understanding these uses is essential in order to use appropriate tools and methods to perform network analysis. The following paragraphs, although not a comprehensive list, provide concepts of how systems are used.

Backdoor/Trojan

A Trojan is a malicious program disguised as a benign application. Trojans usually provide a means for an attacker to access and/or control the compromised system. They are often used to harvest personal information (keystrokes, passwords, e-mail addresses, credit card info, etc.). Some Trojans send collected personal data to network locations accessible to the attacker. In order to protect access to a system, attackers often attempt to replace system files to hide files and processes on a system, making it difficult to identify the attacker's activities. In order to function, the Trojan (or the applications it installs) will generate network traffic.

Proxy Server

Compromised systems are often used as a proxy server or anonymizer, as described in the prior section. This allows the attacker to use a compromised system, instead of a commercial proxy or anonymizer, which may have records that could be obtained by legal authority. When used in this way, attacks or connections to other systems will show network traffic back to the compromised system. Each level of proxy provides another layer of anonymity to an attacker. Collecting and analyzing network traffic at a proxy server can identify the actual network information of an attacker.

Internet Relay Chat (IRC) Server

Internet Relay Chat (IRC) servers are systems providing both group communications (many to many) and individual communications (one to one) in discussion groups known as *channels*. They also provide the means to conduct data transfers via private messages. IRC is fully described in RFC 1459 [3]. IRC servers are relatively easy to set up and provide a simple effective solution for communications, which can even employ encryption. They are used to host distributed zombie or BOT (short for robot) networks because of the relative anonymity available. IRC servers are typically installed on victim machines. Network traffic analysis can identify the operator (attacker) of an IRC channel from BOT traffic data as well as gauge the size of the BOT network by enumerating the victims within a channel. Attackers also install an automated BOT client on compromised systems. The BOT can automatically send data, receive commands, and execute preconfigured actions. Figure 1 depicts a typical BOT network. Compromised hosts (victims) often include

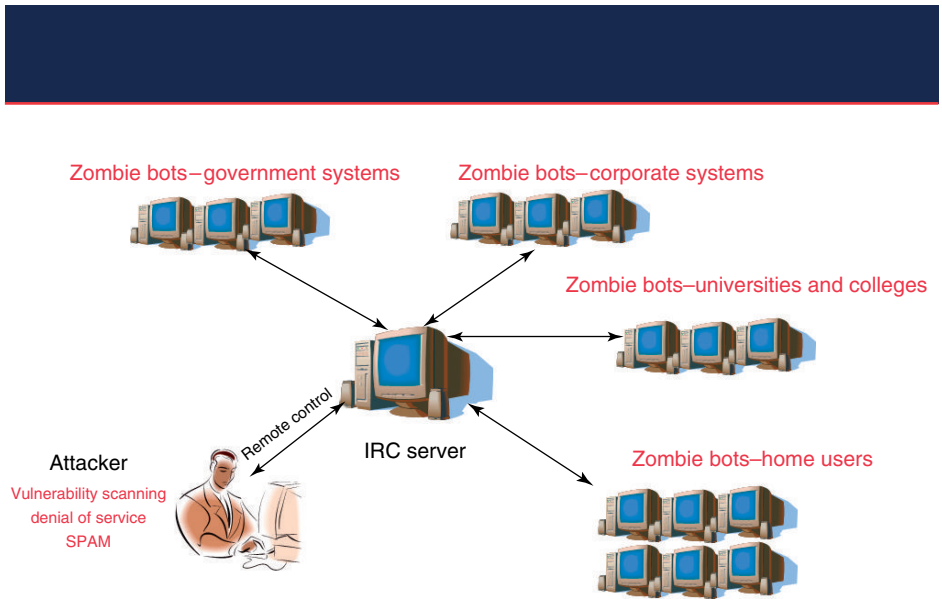


Figure 1 A typical IRC BOT Network

government, corporate, and university systems [4]. BOTs are used for activities that are easily automated such as vulnerability scanning, DOS, spam generation, etc. BOTs normally connect to an IRC server and report their status to an attacker "controlled" IRC channel. Automated processes controlled by an attacker can then connect to the IRC channel and direct the BOTs (individually or as a group) to perform specified actions, for example a distributed network DOS attack.

Web Server

Web servers use HTTP, which is one of the most prevalent application layer protocols on the Internet, used by millions of people daily while visiting various websites. HTTP can be used to provide static and dynamic content, or even to send and retrieve files like a File Transfer Protocol (FTP) server. HTTP is fully described in RFC 1945 [5]. Because HTTP is ubiquitous, it is normally allowed across different networks with little control or interference, and traffic volume may create problems when monitoring HTTP traffic. Attackers often host websites on compromised systems because of these attributes. However, since data would need to be retrieved from the sites, monitoring of traffic to a web server can disclose all systems connecting to it, and the types of connections can differentiate random visitors from an attacker.

Mail Server

Mail servers are used to distribute e-mail to users around the world. While there are several types of e-mail servers, most are compliant with Simple Mail Transfer Protocol (SMTP). SMTP is fully described in RFC 2821 [6]. Mail servers can be used to distribute mass mailing of advertising e-mails (often called *SPAM*) and malicious e-mails; however, most legitimate mail servers have some safeguards in place to preclude such abuse. An SMTP server that only sends e-mail requires little overhead and is often installed on victim systems. Analysis of the network traffic of such a server can identify control traffic to the SMTP server from the attacker.

FTP Server

FTP has been around since the early 1980s and is fully described in RFC 959 [7]. FTP was designed to enable systems to exchange files over TCP/IP networks. Today, attackers utilize compromised systems to operate FTP servers to host malicious code and large archives of illegally copyrighted software (commonly referred to as *warez*), music, and video content. Analysis of network traffic on an FTP server can identify control to the FTP server from the attacker.

Common Tools and Techniques

There are far too many tools and techniques that can be used to conduct network analysis to list them all in a brief article such as this. There are several detailed books and other resources that provide far more detailed information, which are listed in the last section of this article. Some of these provide case studies and sample data. A reduced set of tools and techniques is presented as an overview, and is by no means a complete enumeration of the types of techniques necessary for network analysis.

Tools

Examiners require tools that properly parse various file formats created by tools used in network analysis and investigation, store and retrieve data, and find correlations between the different data sources. Examiners also require knowledge and expertise in networks and networked systems. Examiners may need the ability to quickly modify existing tools (or create new tools) for analyzing unique file formats. Tools for performing network analysis and analysis of malicious binaries used in network attacks are subject to change over time. The following paragraph lists some functions/capabilities and tools that perform all or some of them that are used today.

Examiners can conduct static text analysis with the strings command (a utility built-in most Unix/Linux OS), the strings utility from Microsoft Sysinternals utilities [8], Hex Workshop [9], and WinHex [10]. For software disassembly and debugging, there are tools like IDA Pro [11] and OllyDbg [12]. For real-time OS monitoring, there are a number of UNIX utilities (free, top, sysstat, vmstat, lsof, etc.) and the Microsoft Sysinternals Utilities (such as regmon, filemon, procmon, and ProcessExplorer). For network monitoring, Snort [13] and Wireshark (formerly known as *Ethereal*) [14] are solid tools.

The use of virtual operating systems, using software such as VMWare [15] and Microsoft Virtual PC [16], is an excellent method for conducting multiple tests.

Network Traffic

Time is of the essence when obtaining network data and logs. Many ISPs maintain only logs necessary to support operational needs. For example, many ISPs utilize Dynamic Host Configuration Protocol (DHCP) to assign IP addresses to customer client systems, as there are often a limited number of IP addresses available. DHCP allows a provider to "reuse" an IP address for multiple systems. An explanation of DHCP can be found in RFC 2131 [17]. DHCP logs at a busy ISP may only be retained for days, or just hours, depending on the number of client connections and the ISPs IP address space. Router logs and other logs may simply be a rolling log that is limited by space, and thus may contain only the last several hours of connection data. Additionally, data stored on network servers by persons of interest can be moved, modified, or deleted should the attacker become aware of an investigation. It is essential to identify relevant networks and systems of interest, and initiate the process to have the ISP retain data until a formal legal order can be obtained from a court or other legal authority. In the United States, there are preservation requests that can be submitted to an ISP to retain such data for 90 days (which can be extended another 90 days) for investigative purposes [18].

In addition to logs, network traffic is often collected and stored in capture files. A common file format is the packet capture or "pcap" file format, used by many tools, such as TCPDump, based on libpcap [19]. Many other network capture formats exist and an examiner may need to be able to handle these other formats. Each captured packet includes a packet header followed by the raw captured bytes including the frame header, IP header, etc. The frame header type for the capture bytes is based on the "linktype" flag of the file header. This data in raw form is difficult to analyze as it is binary data and it is packet-based rather than session-based. Additionally, time is not stored the same way in all capture formats. In some formats, timestamps are assigned by the capture utility and is relative to the first packet. Time is often stored as a function of the operating system, and could be the number of seconds elapsed since midnight Coordinated Universal Time (UTC) of January 1, 1970, not counting leap seconds, or as the number of 100-ns intervals since January 1, 1601. However the raw network traffic does not have a time value other than dates/times that may be in the data portion of the packets. One can process the data with tools such as Snort to determine whether the capture reveals any known intrusion signatures. Sessions can be reassembled with other tools, such as Wireshark to extract session data such as an SMTP, FTP, or HTTP session. One must carefully document time/date offsets of systems from which data is collected, as well as time drift information (where available) in order to allow examiners to properly correlate data.

Examiners need to obtain information about the source of network events discovered by review of logs or network traffic data. An IP address can be resolved to a hostname using the "nslookup" or "host" command, and the hostname can also be resolved back to an IP address, using the same commands. Resolving a host in both directions provides a sanity check that may uncover domain name system (DNS) spoofing. An innovation which has recently become more widespread may complicate this procedure, known as *Dynamic DNS services*, which is described in RFC 2136 [20]. This allows systems without fixed IP addresses to utilize a fixed domain name. While not a new capability, it has become more widely available without cost and in an anonymous fashion. A "whois" query provides IP address block registration and domain name registration information. A "traceroute" provides information about the series of network connections between the source and destination system. Care must be taken, however, as traceroute and other active techniques are detectable by a knowledgeable target.

Peer-to-Peer (P2P) technologies are also utilized, not only to "chat" but also to share files. A simple P2P network directly connects one computer to another. In a more complex P2P network used for file sharing such as Kazaa, for example, users can search for files using a directory stored on a server. Most servers in this type of P2P network do not log the activity of users. The server identifies other computers where the file is available for download. The user's client can download files directly from another computer hosting the files desired. P2P networks are always changing as users connect and disconnect systems to the network. Analysis of logs maintained by the client software may provide information as to the location where files were downloaded. Additionally, network

traffic captures can readily identify connections made by the P2P client, although some clients can encrypt the data, the connection itself is visible.

Similar to P2P, a number of remote data storage techniques are available. These techniques use freely available services, such as GMail™ (Google's™ e-mail service) [21]. Using software clients, the user can specify a legitimate account at GMail, and utilize the e-mail account like a file system. This is not limited to GMail; however, it is a convenient example. One would have to conduct careful analysis of the HTTP traffic between the system and GMail and identify the nature of the traffic.

Binary Analysis

The root cause or catalyst for many network attacks today is malicious software or malware. Malware can exist in several forms – backdoor or key-logging Trojan; mass-mailing or self-propagating Worm, file-infecting, boot-sector, or macro virus; java applet or script; Active-X control; or other custom exploit. Peter Szor provided a comprehensive account of infection techniques, update procedures, and control capabilities for these and other classes of malware [22]. In many cases, network analysis directs an examiner to a suspicious executable or binary file on a system requiring evaluation to find leads to further an investigation. Binary analysis can be achieved using static and run-time analysis methodologies.

Static Analysis. Static analysis is the examination of a suspected binary without execution. It can extract human readable strings (i.e., e-mail addresses, IP addresses, registry keys, file names, usernames, hard-coded passwords, and application programming interface (API) calls) as an initial assessment to determine how a binary behaves when executed and clues to who created, authored, or controls the program. Figure 2 depicts an example of strings extracted from a sample binary file using a hex editor utility. Static analysis can also include disassembly and reverse engineering of malware. A disassembler is software used to recover object code from a binary executable. Object code is an intermediary format from source code, and is created by software compilers. Though

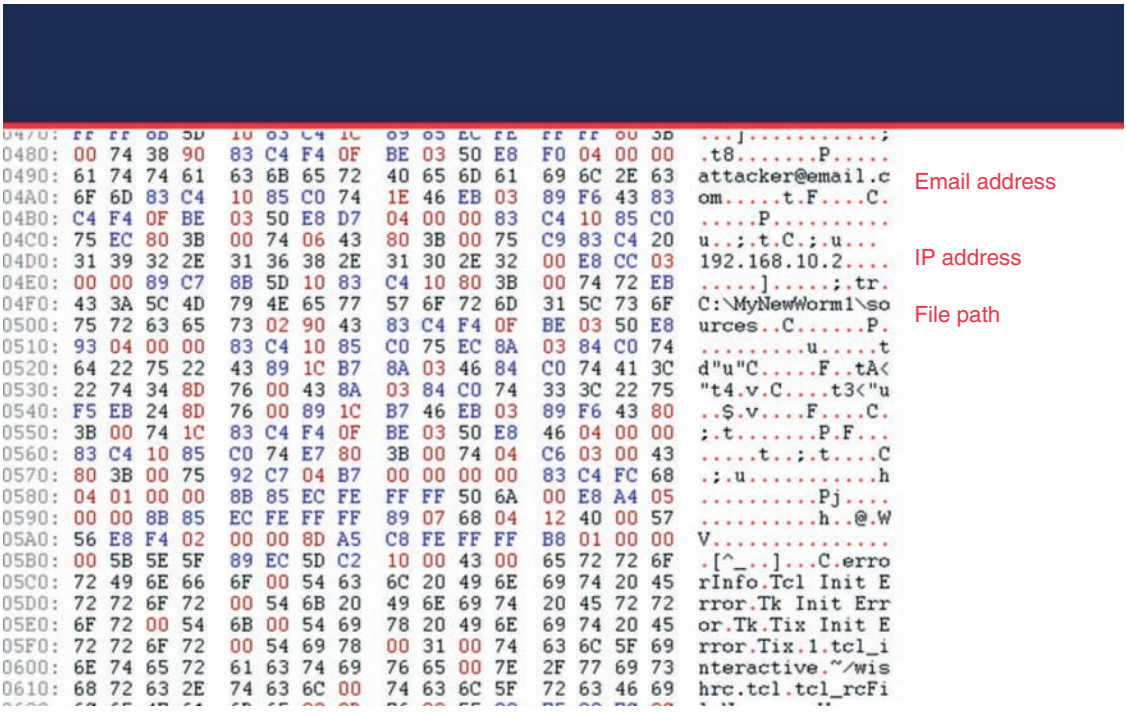


Figure 2 Example of strings extracted from malware

a very tedious process, disassembly is oftentimes a necessary process for forensic analysis of binary malicious code when source code is not available and the aforementioned strings analysis process proves not useful. Eilam details various reversing tools and disassembly techniques in his work [23].

Run-Time Analysis. Run-time analysis refers to the monitoring of the behavior of malware during execution. By behavior, we are referring to how the malware affects the OS and network during run-time. Malware will oftentimes delete and move existing files on a system, install or create new ones (i.e., Trojan key-logging files), or even make modifications to the system files like the Windows registry. Moreover, the tracking of newly opened network ports (i.e., Trojan backdoor) or data (i.e., passwords and credit card numbers) being surreptitiously exfiltrated from a system to an attacker across the network is critical when describing the behavior of malware. Microsoft Sysinternals hosts a suite of various tools for analyzing malware during run-time. In some cases, it may be necessary to analyze malware code and its memory contents while it is executing. Debuggers allow a malicious binary file to be monitored as each instruction is executed. An examiner may conduct "single-step" execution of malware one instruction at a time or set breakpoints at any time in order to view the values of code variables, registers, or memory locations. Debugging techniques enable the viewing of strings and data values that may have been otherwise encrypted, encoded, or obfuscated during static analysis.

Setting up an Analysis Environment. The presence of malicious binaries on a system is inherently unsafe. An examiner must give careful consideration to setting up a proper analysis environment for performing static and run-time analysis. The safest environment for such analysis is to establish an isolated or sandbox network. This enables malware to freely execute without fear of contaminating an actual network or sending traffic to the Internet. Virtual Machine software enables an emulated malware analysis environment to be created, saved, and refreshed in a quick and disposable fashion. An alternative solution involves creating and restoring malware analysis environments with incremental backup technology. This approach may be more viable if the malware is using antiemulation techniques, thus detecting if virtual machines, debuggers, or other techniques are being used.

Best Practices

Specified procedures, or protocols, for receipt, preparation, examination, analysis, and reporting of results should be in place. Adhering to a specified protocol/procedure will preclude mishandling, which could result in the scientific methodology being questioned in a legal proceeding. Additionally, data should be collected in a media-based methodology where possible, since media analysis may allow hidden, deleted, or altered data to be recovered during such analysis (*see* **Computers**). This recovered data may provide critical information to the examiner. In some cases, especially where critical servers are involved, it may not be possible to obtain complete disk images, and such cases should be documented carefully and thoroughly. Since network analysis normally requires examination of data from multiple sources, the examiner must use special consideration when performing correlation of data from these sources, as there could be time/date discrepancies, IP address changes due to network address translation or DHCP, etc. Additional important information required by the examiner would be network topology, system time deviation from a standard time for systems of interest, and actual configuration details of network devices such as firewalls, routers, and IDSs.

There are a number of excellent books and other publications that provide more detailed specific information on techniques for conducting e-mail, chat, file sharing, and other network based activities. A recent publication from the US National Institute of Standards and Technology (NIST) provides good overviews of these types of cases [24]. In addition to a solid explanation of typical investigative case scenarios (such as e-mail tracing, chat investigations, P2P networks, etc.), this document provides samples for a variety of legal requests under US Federal law. Other excellent resources (by no means a complete list) are Digital Evidence and Computer Crime [25], Handbook of Computer Crime Investigation [26], Windows Forensics and Incident Response [27], Computer Forensics: Incident Response Essentials [28], Incident Response: Computer Forensics (Second Edition) [29], Hacking Exposed: Computer Forensics [30], and WarDriving, Drive Detect and Defend: A Guide To Wireless Security [31].

References

[1] Socolofsky, T.J. & Kale, C.J. (1991). *A TCP/IP Tutorial*, RFC 1180, Internet Engineering Task Force, January 1991.

[2] Dingledine, R., Mathewson, N. & Syverson, P. (2004). Tor: the second-generation onion router. *Proceedings of 2004 Usenix Security Symposium*. San Diego, CA August 2004.

[3] Oikarinen, J. & Reed, D. (1993). *Internet Relay Chat Protocol*, RFC 1459, Internet Engineering Task Force, May 1993.

[4] Baylor, K. & Brown, C. (2007). *Killing Botnets: A View From The Trenches*, McAffee, Inc, (http://www.mcafee.com/us/local_content/white_papers/wp_botnet.pdf), October 2007.

[5] Berners-Lee, T., Fielding, R.T. & Nielsen, H.F. (1996). *Hypertext Transfer Protocol*, RFC 1945, Internet Engineering Task Force, May 1996.

[6] Klensin, J.C. (2001). *Simple Mail Transfer Protocol*, RFC 2821, Internet Engineering Task Force, April 2001.

[7] Prostel & Reynolds (1985). *File Transfer Protocol*, RFC 959, Internet Engineering Task Force, October 1985.

[8] Microsoft Windows SysInternals 2008. Website http://www.microsoft.com/technet/sysinternals.

[9] Hex Workshop http://www.bpsoft.com.

[10] WinHex http://www.winhex.com.

[11] IDA Pro http://www.datarescue.com.

[12] Olly Debug http://www.ollydbg.de.

[13] SNORT http://www.snort.org.

[14] Wireshark http://www.wireshark.org/.

[15] VMware http://www.vmware.com.

[16] Microsoft Virtual PC http://www.microsoft.com.

[17] Droms, R. (1997). *Dynamic Host Configuration Protocol*, RFC 2131, Internet Engineering Task Force, March 1997.

[18] This part of the electronic Communicaltion Privacy Act of 1986. (Public Law 99–508, Oct. 21, 1986, 100 Stat. 1848, 18 United States Code § 2510) Title 18 of the United States Code, Section 2703.

[19] LIBPCAP developed by the Network Research Group (1994). Information and Computing Sciences Division at Lawrence Berkley Laboratory, University of California, Barkeley, California.

[20] Vixie, P., Thomson, S., Rekhter, Y. & Bound, J. (1997). *Dynamic Updates in the Domain Name System*, RFC 2136, Internet Engineering Task Force, April 1997.

[21] GMail and Google are registered Trademarks of Google, Inc, Mountain View, CA.

[22] Szor, P. (2005). *The Art of Computer Virus Research and Defense*, Addison-Wesley.

[23] Eilam, E. (2005). *Reversing: Secrets of Reverse Engineering*, Wiley Publishing, Inc.

[24] NIST (2007). *Investigations Involving the Internet and Computer Networks*, US Government Printing Office, January 2007.

[25] Casey, E. (2004). *Digital Evidence and Computer Crime*, 2nd Edition, Academic Press, March 2004.

[26] Casey, E. (ed) (2001). *Handbook of Computer Crime Investigation*, Academic Press, October 2001.

[27] Carvey, H. (2004). *Windows Forensics and Incident Response*, Addison-Wesley, July 2004.

[28] Kruse II, W.G. & Heiser, J. (2001). *Computer Forensics: Incident Response Essentials*, Addison-Wesley Professional, September 2001.

[29] Prosise, C. (2003). *Incident Response: Computer Forensics*, 2nd Edition, McGraw-Hill, July 2003.

[30] Davis, C., Philipp, A. & Cowen, D. (2005). *Hacking Exposed: Computer Forensics*, McGraw-Hill/Osborne.

[31] Hurley, C., Thornton, F., Puchol, M. & Rogers, R. (Technical Editor) (2004). *WarDriving – Drive, Detect and Defend: A Guide to Wireless Security*, Syngress Pres.

DAVID W. BAKER AND THOMAS ERVIN

Analysis: Degraded Samples ***see*** DNA: Degraded Samples

Analysis: Drug Profiling ***see*** Drug Profiling

Analysis: Drugs ***see*** Drug Analysis

Analysis: Error ***see*** Error Rates in Forensic Methods

Analysis: Explosion Debris ***see*** Explosion Debris: Laboratory Analysis of

Analysis: Fire Debris ***see*** Fire Debris: Laboratory Analysis of

Analysis: Glass ***see*** Glass

Analysis: Hair, Drugs ***see*** Hair: Toxicology

Analysis: Ink ***see*** Ink Analysis

Analysis: Marijuana ***see*** Cannabis

Analysis: Naturally Occurring Poisons ***see*** Poisons: Detection of Naturally Occurring Poisons

Analysis: Neutron Activation

A Nondestructive Method

Neutron Activation Analysis (NAA), and variations such as Instrumental Nuclear Activation Analysis (INAA), Fast Neutron Activation Analysis (FNAA), and Epithermal Neutron Activation Analysis (ENAA) describe somewhat similar nondestructive methods for the analysis of concentrations of elements in small particles of trace evidence. NAA is a nuclear, as opposed to a chemical, spectrographic method of analysis. When the method first became popularized in the crime laboratory world of the United States in the 1960s. The high cost associated with conducting NAA examinations, as well as the discovery that the method performed poorly with some elements, caused it to fall somewhat from grace.

Also, as other less costly laboratory techniques suitable for the examination of minute specimens of trace evidence became available, NAA was effectively abandoned. When, early in the twenty-first century, the method began to be used in civil litigation through the services of some private and university-associated laboratories, NAA began to experience a modest revival.[a]

NAA is particularly an effective method for the analysis of bomb debris and explosives detection, gunshot residues and gunshot pellets, paint, and hair.

Basics of the Technique

NAA can determine the qualitative and quantitative elemental composition of a sample by bombarding the sample for a controlled period of time with an intense stream of nuclear particles, usually neutrons. Bombardment of a sample with neutrons will cause some of them to be absorbed by the nucleus of the bombarded atom. This bombardment ("neutron activation") causes an instability that produces radioactive species, or radionuclides, for about 70% of all elements.

The radioactive isotopes tend to return to a stable (nonradioactive) state over time. This process is called the *decay*, and the disintegration of the "unstable" atoms is accompanied by the emission of gamma rays – high-energy electromagnetic radiation. The gamma rays emitted by an activated sample are detected and measured by a gamma ray spectrometer, which indicates the energy and number of a given energy of each of the various rays being emitted. Thus, by comparing the data obtained with the data obtained from activating pure samples of the elements, the quantity of each element present in a specimen can be determined [1].

The energy of the gamma ray output is designated in million electron volts (Mev). The ability

of an atom's nucleus to capture bombarding neutrons also varies for each element. The probability that a stable nucleus will capture thermal neutrons is measured by its "cross section" and is expressed in "barns".

Some elements have low "barn" values, meaning that they capture thermal neutrons very reluctantly. An example of that is oxygen, whose "O" isotope has a cross section of only 0.0002 barn, which makes it virtually unusable for NAA. Cross sections of stable isotopes range from about 10^{-5} to 10^{5} barns. Higher barn values will be more sensitive in detecting elements. It is the combination of three values: half-life, decay energy (in Mev), and capture rate (in barns), which lies at the foundation of the NAA technique.

NAA, which was considered experimental for a long period, can today be considered a mature analytical technique that is being offered for forensic analyses in both criminal and civil cases by commercial as well as university-associated institutes.[b]

At the National Institute of Standards and Technology (NIST), INAA is used to detect and determine concentrations of up to 60 elements, derived from human as well as animal sources, along with specimens from botanical and geological sources.[c] In addition to forensic examinations of trace evidence samples that are of consequence to litigation, NAA/INAA are being applied to environmental studies to determine the source and nature of pollutants, the nature of biological tissues and fluids, the detection of impurities in pharmaceutical materials and food products, and for several analyses in nutritional epidemiological, industrial, as well as archeological studies.

End Notes

[a]The Federal Bureau of Investigation utilizes NAA for the examination of bomb and explosives debris, albeit sporadically.

[b]See, e.g., Elemental Analysis Laboratory at the Department of Chemistry, Texas A&M University; University of Missouri – Columbia Research Reactor Center; North Carolina State University, Service of the Nuclear Reactor Program; The Laboratory for Radiochemistry and Environmental Chemistry of the Paul Scherer Institute (PSI); General Activation Analysis, Inc. in Encinitas, CA., and many other facilities worldwide. Information is also available from the International Atomic Energy Agency at http://www-nds.iaea.org/pgaa/ on adopted database files.

[c]http://www.ncnr.nist.gov/instruments/nactanal.html (accessed Oct 2008), describing the experience at NIST with NAA and INAA.

Reference

[1] Guin, V.P. (1966). Neutron activation analysis and its forensic applications, *Proceedings of the 11th International Conference on Forensic Activation Analysis*, Texas A & M University.

Further Reading

Alfassi, Z.B. & Chien, C. (1995). *Prompt Gamma Neutron Activation Analysis*, CRC Press.

Campbell, J.A. & Bewick M.W.M. (1978). *Neutron Activation Analysis: A Review of the Method and Its Present and Future,* CRC Press.

Corliss, W.R. (1964). *Neutron Activation Analysis*.

Delftse, M. (1983). *Instrumental Neutron Activation Analysis – A Routine Method*, Delftste Universitaire Pers.

De Soete, D., Gijbels, R. & Hoste, J. (1972). *Neutron Activation Analysis*, Wiley-Interscience.

International Atomic Energy Commission (1973). *In vivo Neutron Activation Analysis*, IAEA.

Keisch, B. (1972). *The Atomic Fingerprint: Neutron Activation Analysis*; paperback, 2003 digitized 2007.

Nargolwalla S.S. & Przybylowicz, E.P. (1973). *Activation Analysis with Neutron Generators*, Wiley.

Parry, S.J. (2003). *Handbook of Neutron Activation Analysis*, Viridian Publishing.

United States Warren Commission (1964). *Investigation of the Assassination of President John F. Kennedy*, USGPO.

ANDRE MOENSSENS

Analysis: Opioids *see* Opioids

Analysis: Paint *see* Paint

Analysis: Paper ***see*** Paper Analysis

Analysis: Photographs ***see*** Visual Recognition Systems in Identification

Analysis: Postmortem ***see*** Postmortem Biochemical Examinations

Analysis: Soil ***see*** Soil: Forensic Analysis

Analysis: Toxicology ***see*** Toxicology: Initial Testing; Confirmation Testing: Toxicology

Analysis: Video ***see*** Visual Recognition Systems in Identification

Ancestry Determination from Skeletal Remains ***see*** Anthropology: Ancestry and Stature Determination

Animal Hair ***see*** Hair: Animal

Animation and Simulation Evidence: Computerized ***see*** Computer Animation and Simulation Evidence

Anthropology

Introduction

Forensic anthropology can be described as the analysis of human remains for the medicolegal purpose of establishing identity. It is a multi-disciplinary endeavor that applies the knowledge of biological anthropology and human osteology to cases where human remains are skeletonized, or where a detailed understanding of the growth and development, morphology, or reactive norms of the human body can assist other disciplines in positive identification. As such, it is perhaps best placed as a sub-discipline within the wider practice of forensic human identification, having applicability to both the living and the dead.

Anthropologists are routinely called upon to assist in the identification of individuals whose remains are severely decomposed, burned, disrupted, mutilated or otherwise rendered difficult to recognize. Their services have been of particular value in

three main areas: (i) investigations resulting from homicide, unexplained natural deaths, accidents, and mass fatalities; (ii) noncriminal events resulting in multiple deaths, e.g., arising from natural catastrophes; and (iii) war crimes investigations and genocide. However, the anthropologist may also be called upon to assist in the assessment of the living, providing confirmation of identifiers such as chronological age for the purposes of judicial accountability, immigration, or asylum status [1].

The context under which many anthropological cases are undertaken has changed during the last decade owing to a number of factors that have influenced not only on how anthropologists perform their job but also the type of jobs they are now requested to undertake. The twenty-first century has seen an increasing perception of risk involvement in mass fatality incidents including accidents, terrorist attacks, or natural disasters [2]. Terrorist activity resulting in mass fatalities, such as the September 11, 2001, bombing of the Twin Towers in New York, or the London bombings of July 7, 2005, has highlighted the requirement for national and international disaster victim identification (DVI) capability, and cemented the anthropologist's role as a significant component within the multi-disciplinary response facility. Thus, it may be argued that viewing the forensic anthropologist as someone who "deals with the analysis of human skeletal remains resulting from unexplained deaths" [3] is unrealistically restrictive given the multi-disciplinary nature of the demands of human identification in the twenty-first century.

Practitioners of forensic anthropology are specialists in human skeletal morphology, and are generally trained via human anatomy, physical or biological anthropology, or osteoarchaeology; professional accreditation is important in order to set professional entry standards, to monitor continuing professional development, and to oversee the maintenance of professional competence [4]. Practitioners may be registered with a suitable national organization, such as the American Association of Forensic Sciences (AAFS) in the US, or the Council for the Registration of Forensic Practitioners (CRFP) and the Forensic Science Society (FSS) in the UK.

Forensic Anthropology and Human Identification

The role of the forensic anthropologist focuses on applying detailed knowledge of the development, morphology, and variation of the human body as an aid to establishing, confirming, or indeed refuting personal identification. Four basic biological criteria form the basis of any primary identification: skeletal or developmental age, biological sex, living stature, and ancestry or racial affiliation [5]. These characteristics are supplemented by secondary traits, which include the following: body modifications [6]; evidence of surgical interventions [7], including scars and prosthetics; evidence of trauma; and chronic hard and soft tissue diseases that may leave characteristic macroscopic and microscopic lesions on surviving tissue [8]. Any of these additional factors can be of assistance when attempting to identify an individual and combinations of the above can be extremely useful when confirming identity. The anthropologist may also be called upon to advise on the manner of death, particularly in cases of skeletal trauma (blunt force, sharp force, or ballistic) or where physical processes such as burning, dismemberment, or explosive damage have disrupted the remains rendering quantification and identification problematical. They may also be called upon to corroborate estimations of the time elapsed since death (time-death interval) and assess taphonomic modification of human remains by taphonomic factors in the burial environment [9, 10].

The Stages of Forensic Human Identification

Is It Human? Differentiating Human from Nonhuman Bone

Frequently, bone is presented for identification from the sites of accidental discovery (i.e., stray surface finds, detritus from river, estuary or coastal contexts, or through excavation/construction work) or from the sites to which a presumption of violent or wrongful death may be attached through police intelligence. From a medicolegal standpoint, initially it is presumed that the remains are human and were deposited within a relevant timescale (i.e., 70 years before present in the UK or 50 years before present in the US). If they are determined to be human in origin,

and of recent deposition, then the forensic anthropologist applies the skills and techniques of skeletal analysis to establish or corroborate the presumptive identity of human remains. Thus, rapid determination of human or nonhuman origin is essential, both to determine whether a crime has been committed and to ensure that valuable police resources are not wasted investigating the disposal of nonhuman remains.

In practice, the anthropologist may not always be able to identify bone fragments to a particular species, but they should be able to determine whether they are human through macroscopic and/or microscopic evaluation depending on the elements of the skeleton presented for identification and the degree of fragmentation of the bones. Certain taxonomic groups may be difficult to differentiate because of similarities in overall structure, size, or bone quality; in general, the mammalian species with greatest likeness to those of humans can be found in larger mammals such as bears, deer, cows, pigs, and large dogs. Furthermore, skeletal remnants of smaller mammals and birds may be misidentified as fetal or juvenile human bones due to similarities in size and shape (Figure 1); perhaps more importantly, human fetal and juvenile bones may be misidentified as nonhuman due to the inexperience of the observer in working with such material.

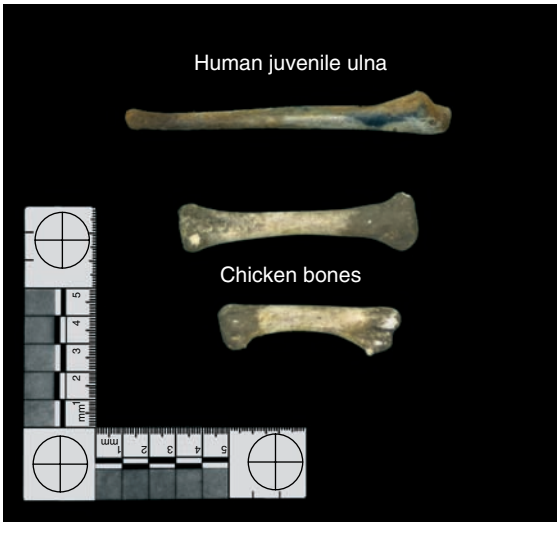


Figure 1 Comparison of human perinatal ulna with domestic chicken bones

Confusion primarily arises where only fragmentary bones exist, or the bones are charred, as morphological features are absent and cannot be used for identification. In cases where bone has been highly fragmented by trauma, postdepositional alteration, butchery/dismemberment, or burning, then identification through gross morphology alone may not always be possible. In such instances, the microscopic evaluation of bone histology [11], DNA analysis, [12], or immunological studies [13] may allow for taxonomic differentiation. Histological identification is predicated on the fact that different species both lay down bone and remodel it in different ways, the expression of which is dictated by functional demands (such as locomotion or body size) and the timing of major ontogenetic events set against lifespan. For instance, human bone displays tubular cellular structures that appear circular in cross section. These structures, known as *Haversian systems*, result from the remodeling of parallel layers that are deposited in a lamellar or concentric fashion to the cross-sectional axis of the bone; the extent of remodeling is proportional to the maturity of the bone.

Many animals, herbivores in particular, present a cellular structure made up of rectangular laminar plates, termed a *plexiform pattern*, which is only gradually replaced and overprinted by Haversian bone. Plexiform structures are a compromise between the requirements of structural rigidity (particularly, optimization for torsional forces) and the necessity for rapid bone formation and early maturation, resulting in large-diameter, thin-walled cross sections. The banding of osteonic units may also be used to differentiate human from nonhuman. While the presence of plexiform bone is a ready indication of nonhuman origin, it should be borne in mind that some animals present a combination of histological types (such as domestic and wild canids), which may make microscopic evaluation difficult. In such instances, the use of DNA analyses, immunoassay, or protein seriology may be more appropriate as a means of distinguishing between human and nonhuman bone. DNA (particularly, mitochondrial) can persist in bone for centuries, and recent studies have shown that highly repetitive DNA markers can be identified and used not only to determine human origin but also to identify individual species [12]. Similarly, specific coding regions (such as the mtDNA *cyt*-b gene) may be used to distinguish between taxa [14, 15].

Other discriminatory methods concentrate on the detection of species-specific proteins by utilizing the binding between an antibody and its antigen, a process known as *immunoassay*. Albumin is commonly used for detecting the origin of bone, as it is highly species specific, readily discriminates human from nonhuman bone, and can withstand high temperatures. Protein radioimmunoassays (pRIA) have also been used for discrimination, with some studies claiming 100% success rates in distinguishing human from nonhuman samples [13]. High levels of fragmentation, burning, and diagenic alteration by the burial environment adversely affect the discriminating abilities of these analytical methods [16].

Minimum Number of Individuals (MNI) Present

Accurate quantification of the number of individuals within an assemblage is the first stage of the individuation process. In cases of multiple burial of commingled remains, or fragmentation through explosive or thermal disruption, an accurate assessment of the minimum number of individuals (MNI) represented in an assemblage is paramount. However, quantification is compounded by the covarying factors of fragmentation and completeness. As a general rule, the more fragmented or disrupted the remains, the more difficult it will be to ensure full and accurate identification of the individual. Likewise, an individual represented by only a few bones of the skeleton will be more difficult to identify than one represented by a complete skeleton. As a consequence, there exists an inverse relationship between the degree of bone fragmentation and the number of individuals usually identified within an assemblage [17]; the competent anthropologist aims to reduce this identification bias through careful and detailed analysis of each and every fragment.

Quantification of MNI is based on a number of determinants, the most fundamental of which is nonduplication of skeletal parts. This requires accurate identification, seriation, and siding of all the bones or bone fragments present, with the proviso that there should be no duplication of skeletal elements if only a single individual is represented. For instance, the recovery of a frontal bone, right temporal bone, three to seven left ribs, a right femur, and two left humeral heads would indicate that at least two individuals are present due to the duplication of proximal humeri. It is, of course, entirely possible that 10 individuals are present as each recovered skeletal element could conceivably be derived from a different individual; determination of this would hinge on a variety of skeletal indicators coupled with an understanding of the context of skeletal recovery. Other skeletal markers that may assist in quantification include differing relative sizes of bones, differing stages of skeletal development, sex-discriminant markers, pathological conditions, or disjunction in color or state of preservation of the remains (the latter needs to be treated with some caution as bone color and preservation may differ within the same burial context). In general, such methods are prone to inaccuracy which has led to renewed interest in the methodology of quantification and sorting of commingled assemblages. New approaches that aim to be quantitative and more objective have been developed based on the recording of discrete anatomical zones [18], through the use of pair-wise maximum likelihood models, DNA analysis, or GIS-based (Geographical Information System) statistical models [19–21].

Sex Determination

Because most methods of age and stature determination are sex-dependant, determining the biological sex of an individual is the first component of individualized analysis. Like most primates, humans display a discrete pattern of morphological differentiation between males and females; this is termed *sexual dimorphism*. Some of these differences are associated with primary sexual characteristics of the reproductive system, which includes pelvic morphology, while others present a host of interrelated morphological, physiological, and behavioral features that become manifest with the onset of the hormonal surge at puberty. These are referred to as *secondary sexual features*.

In humans, secondary features manifest themselves in a variety of ways. Males, for example, tend to exhibit greater stature and body weight than females. They generally have more robust cranial and facial features, increased muscularity, greater strength, and speed. Females store more subcutaneous fat, while males have proportionately more muscle volume. Males have greater potential to develop muscle bulk and power output, while females have the potential to develop higher levels of stamina; this leads to definable differences in the form and

robusticity of their respective skeletons. Levels of sexual dimorphism can be affected by significant social phenomena including differential investment in parental effort, differential patterns of resource acquisition and utilization, and sexual division of labor [22–29]. For example, practices that are physically demanding may reduce sexual dimorphism if practiced equally between the sexes whereas they would tend to increase sexual dimorphism if practiced asymmetrically by one sex, with dimorphism arising as a consequence of divergence in muscular or skeletal robusticity. The level of sexual dimorphism exhibited by a population is thus a consequence of ecological, social, behavioral, dietary, and phylogenetic factors all finding varying degrees of expression. The result is that overall dimorphism represents a unique adaptation of the population to the environment, and is thus time, context, and population dependent [23, 26, 29–35].

As sexually dimorphic features are established during the adolescent growth spurt at puberty, juvenile remains must be assigned a sex with extreme caution; the remains of infants, children, and adolescents cannot be readily or accurately sexed on morphological grounds alone as primary and secondary sexual characters are yet to be established [36, 37]. While some success has been reported in the recognition of sex-discriminant characters in juvenile skeletons based on morphological assessment of the mandible and ilium [38–41] and the geometry of the developing cranium [42], these have yet to be shown to be sufficiently accurate for forensic purposes.

Sex assessment of adult skeletal remains is primarily based on a combination of size and shape differences between males and females (see Table 1). In general, these are most pronounced in the form of the bony pelvis, the cranial vault, and the mandible [5, 43], but are represented in most areas of the human skeleton to some degree [44]. Cranial characteristics (Figure 2a), such as larger brow ridges, increased sites of muscle attachment, pneumatized sinuses, rugose muscle entheses, and larger cheek bones and nasal apertures, are male characteristics linked to an extended period of facial growth, and the interaction between pituitary growth hormones and sex androgens. Shape-related, sexually dimorphic features of the pelvis (Figure 2b), such as the larger, more rectangular female pubic bone, the width of the greater sciatic notch, and the position and size of the acetabulum, result from differences between female and male hormone-induced growth patterns during adolescence. In general, the methods preferentially used in sex determination are morphologically (qualitatively) based and focus on the differentiation of cranial characters (the expression of the nuchal crest, mastoid process, glabella and supraorbital margin, and mental trigone) and pelvic morphology (greater sciatic notch, pubis morphology, and the sacroiliac articulation), although other characters such as variation in the ossification of the costal cartilages [45–47] have also been successful in sex discrimination (Figure 3). It is reported that accurate assignations of sex can exceed 95% from pelvis and skull, 90% from the pelvis and over 80% from the skull alone in the adult skeleton [48, 49]. However, these values should be considered optimal and context/population dependant.

However, quantitative methods have also been developed based on linear measurements of cranial and postcranial size variables. These have greater forensic preference as they are more readily scrutinized through statistical investigation, and as such, are not so heavily reliant on subjective evaluation (see [49]). Some of these utilize univariate discriminant functions to segregate male from female specimens around a predetermined sectioning point, while others employ multivariate discriminant functions including Mahalanobis distance and canonical variates analysis [49, 50, 52–65]. The linear measurements undertaken are primarily reflections of overall size and robusticity, and include measures of facial and cranial vault size (Table 2), humeral and femoral head size (Table 3), long bone lengths, and cross-sectional diameters of the appendicular elements. Computer-based multivariate programs have also been developed to facilitate the statistical discrimination of sex and ancestry measures in unknown-origin samples [66].

Age Determination

Determination of the chronological age of an individual is one of the more challenging areas of forensic anthropological enquiry, particularly once adulthood has been reached. In western society, it has become increasingly problematic to estimate the age of living individuals from visual examination let alone extrapolate this to the deceased. This difficulty can manifest itself in a number of ways with environmental stress,

Table 1 Morphological sex differences in the skull and bony pelvis[(a)]

Morphological trait	Male	Female
Skull		
Overall appearance	Large and rugged	Small and gracile
Supraorbital ridges	Prominent and rounded	Not prominent
Supraorbital margin	Rounded blunt margins	Thinner sharp margins
Orbital shape	Rectangular	Circular
Frontal profile	Sloping anteroposteriorly	Vertical
Frontal eminences	Bosses absent or not pronounced	Pronounced
Mastoid processes	Longer and more pneumatized	Short and small
Suprameatal crest	Extends as a bony crest past the auditory meatus	Root of zygoma does not continue past the auditory meatus
Parietal eminences	Bosses absent or not pronounced	Pronounced
Nuchal region	Muscle markings rugose and pronounced, nuchal crest with external hook may be present	Less rugose and pronounced. Protuberance not marked.
Palate	Larger, broader, tends to U shape	Small, tends to parabolic arc
Teeth	Large	Small
Mandibular symphysis	Pronounced trigonum mentale; projection of mental protuberance in midline with large mandibular tubercles	Small trigonum mentale; reduced projection of mental protuberance with small or absent mandibular tubercles
Mandibular profile	Squared and U-shaped	Narrow and more pointed
Mandibular ramus flexure	Present	Absent
Gonial flaring	Flaring present	Flaring absent
Gonial angle	More acute	More obtuse
Discriminant functions	DF > sectioning point (see Table 2)	DF < sectioning point (see Table 2)
Pelvis		
Pubic symphysis	Higher	Lower
Subpubic angle	Narrow V-shaped, acute angle	Wide U-shaped, obtuse angle
Subpubic concavity	Slight to no concavity	Concavity present
Ventral arc	Absent	Present as elevated ridge extending inferolaterally across ventral pubis
Medial ischiopubic ramus	Broad, flat and blunt. Slightly everted	Ridged, sharp edged, everted
Greater sciatic notch	Narrow and deep	Shallow and wide
Obturator foramen	Large and ovoid	Small and triangular
Acetabulum	Large, more laterally oriented	Small, more anteriorly oriented
Auricular surface	Depressed, wide	Raised, narrow
Preauricular sulcus	Not present or illusionary	Often present, well developed
Postauricular space	Narrow	Wide
Sacrum shape	Narrower, more curved, alae narrower than promontory	Wider, less curved, alae wider than promontory
Pelvic inlet shape	Heart shaped, narrow mediolaterally	Circular, elliptical, wide mediolaterally

[(a)] Source: [43, 44, 50, 51]

poor diet, and lack of exercise leading to premature aging, or the use of cosmetic skin and hair products or plastic surgery to reduce the apparent signs of aging in adults. Visual aging of adolescents can also be rendered problematical by the use of adult styles of makeup, hair and clothing, and the adoption of more "mature" patterns of social behavior. Thus, there can be a considerable discrepancy between the perceived

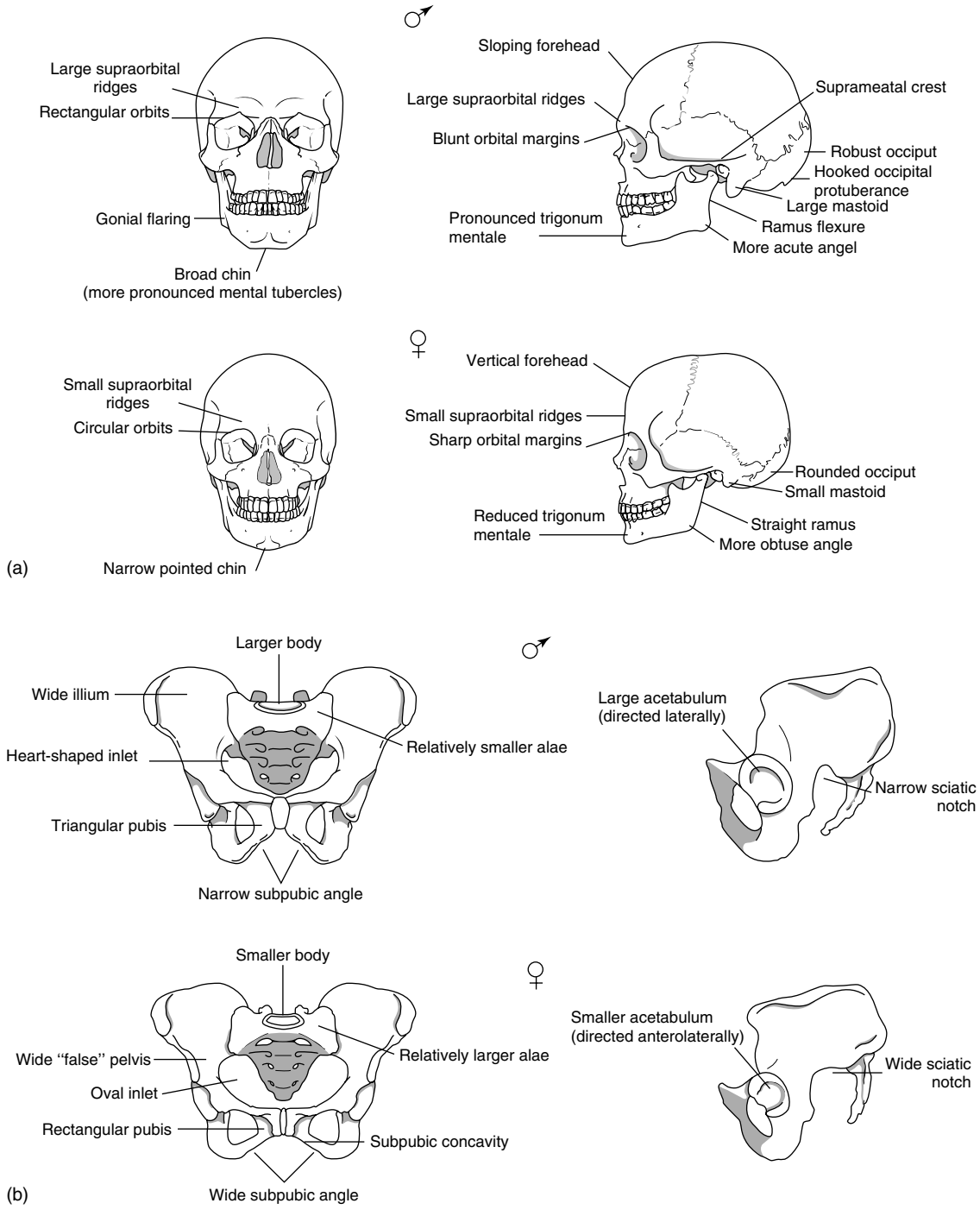


Figure 2 Principal patterns of human sexual dimorphism in the skeleton. (a) Sex differences in the skull. (b) Sex differences in the pelvic girdle [Reproduced with permission from Caroline Needham, University of Dundee.]

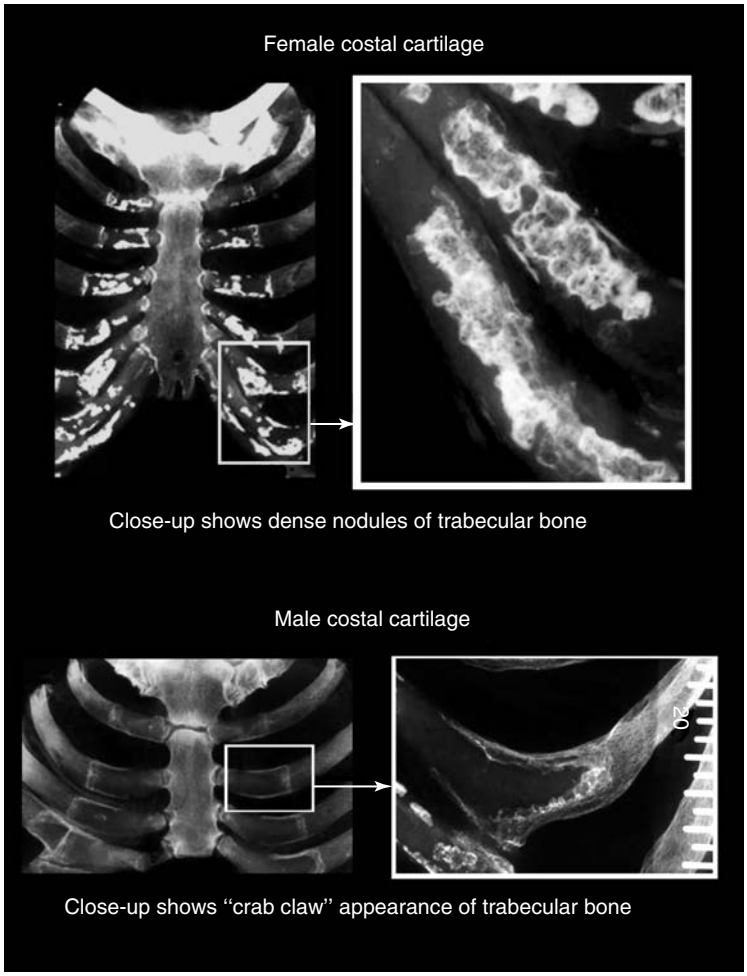


Figure 3 Sex differences in the ossification of costal cartilage

societal age of an individual and that stated on their official documentation.

Age studies have been used in a variety of situations including (i) identification of the individual as part of casework in both the living and the dead; (ii) study of the adequacy of growth of children in a population that is seen as an index of overall community health – poor growth is an indicator of unfavorable developmental conditions, diet or environmental stress; (iii) assessment of age-related social status milestones in life – coming of age, marriage, etc.; and (iv) construction of demographic profiles in an attempt to understand allochronic and/or diachronic patterning in mortality or funerary practices at the population level.

In biological terms, the aging process can be divided into three distinct phases: growth and development, equilibrium, and senescence. Natural variation in the timing and tempo of the aging process can lead to a divergence between biological (particularly skeletal) age and true or chronological age. As such, skeletal age should not be viewed as equivalent to chronological age at death. Skeletal age reflects the natural growth and development of the skeleton in response to diet, environment, and activity, rather than an absolute calendrical age; poor diet can be reflected in delayed maturation and growth yielding a "younger" skeletal age than real calendar years, and conversely, habitual activity, stress, or trauma can produce an aged, or 'older', skeleton than in reality.

Table 2 Discriminant functions for metrical sex determination of the skull for whites and blacks. Values below sectioning point female, above sectioning point male[a]

Whites						
Measurement variable	DF1	DF2	DF3	DF4	DF5	DF6
Maximum cranial breadth	3.107	3.400	1.800	–	1.236	9.875
Biauricular breadth	−4.643	−3.833	−1.783	–	−1.000	–
Basion-bregma height	5.786	5.433	2.767	–	–	–
Basion to nasion	–	−0.167	−0.100	10.714	–	7.062
Bizygomatic breadth	14.821	12.200	6.300	16.381	3.291	19.062
Basion to prosthion	1.000	−0.100	–	−1.000	–	−1.000
Nasion to prosthion	2.714	2.200	–	4.333	–	4.375
Maximum palate breadth	−5.179	–	–	−6.571	–	–
Mastoid length	6.071	5.367	2.833	14.810	1.528	–
Sectioning point	2672.39	2592.32	1296.20	3348.27	536.93	5066.69
Percentage correct	86.6%	86.4%	86.4%	84.5%	85.5%	84.9%
Blacks						
Measurement variable	DF7	DF8	DF9	DF10	DF11	DF12
Maximum cranial breadth	9.222	3.895	3.533	–	2.111	2.867
Biauricular breadth	7.000	3.632	1.667	–	1.000	–
Basion-bregma height	1.000	1.000	0.867	–	–	–
Basion to nasion	–	−2.053	0.100	1.000	–	−0.100
Bizygomatic breadth	31.111	12.947	8.700	19.389	4.936	12.367
Basion to prosthion	5.889	1.368	–	2.778	–	−0.233
Nasion to prosthion	20.222	8.158	–	11.778	–	6.900
Maximum palate breadth	−30.556	–	–	−14.333	–	–
Mastoid length	47.111	19.947	14.367	23.667	8.037	–
Sectioning point	8171.53	4079.12	2515.91	3461.46	1387.72	2568.97
Percentage correct	87.6%	86.6%	86.5%	87.5%	85.3%	85.0%

[a] Source: [3, 67]

This realization has profound bearing on many areas of inquiry with the effect that the criteria utilized to age an individual must be chosen and applied with care, caution, and understanding.

The developmental phase of aging (intrauterine to around 18 years of age) is largely under genetic and environmental influences, leading to well-understood and moderately predictable patterns and rates. During this phase, age determination is primarily based upon (i) well-understood and predictable rates for the mineralization, stages of crown and root formation, eruption and dehiscence of the dentition (Tables 4–6) [68–78]; (ii) timing of the appearance of primary and secondary ossification centers of the cranial and postcranial skeleton [36, 79]; (iii) the unification or fusion of the bones of the postcranial skeleton, which provides a reliable marker within relatively tight statistical margins (Table 7) [80–86]; (iv) correlation of the growth of individual cranial bones (pars basilaris, sphenoid, pars petrosa, tympanic ring, zygomatic, and maxilla) with gestational age [36, 79, 87]; and (v) correlation of developing long bone length against population-specific growth standards [88–94]. The latter method is relatively accurate in the intrauterine period, with fetal bone length closely linked to gestational age, but less accurate as developmental age increases because of the encroaching effects of environment and diet.

Once growth has ceased and skeletal and dental adulthood is reached, age determination becomes

Table 3 Metric determination of sex from selected measurements of the postcranial skeleton[a]

Measurement (mm)	Female	Intermediate	Male
Humerus			
Vertical head diameter	<43.0	44.0–46.0	>47.0
Radius			
Minimum head diameter	<20.0	–	>23.0
Maximum head diameter	<21.0	–	>24.0
Scapula			
Maximum length	<140.0	–	>170.0
Glenoid height	<36.0	–	>36.0
Femur			
Head diameter (whites)	<42.5	43.5–46.5	>47.5
Head diameter (blacks)	<40.0	43.0–44.0	>47.0
Tibia			
Proximal breadth (Terry whites)	<68.6	–	>76.0
Proximal breadth (Terry blacks)	<68.3	–	>77.1

[a] Source: [50, 53, 54]

more difficult as many of the criteria used are reflections of skeletal deterioration and 'wear and tear', and as such they are less predictive, highly variable in expression, and socially mediated [95, 96]. Ages derived from adult skeletons should ideally be considered in broad age categories when using standard techniques, namely young adult (roughly 18–25 years), middle adult (25–45 years), or mature adult (45+ years). In particular, the generalized onset of skeletal maturity at its lower limit and the statistical range of full skeletal adulthood at its upper limit define the lower age bracket, while the upper bracket marks the cutoff point at which current techniques fail to discriminate age-related changes in a time-dependant or statistically significant manner as evidenced from known-age archaeological populations or casework.

In young adults, several components of the skeleton do not fully complete growth until late in the second or the third decade, and thus can be used with some degree of reliability in age determination. These components include the jugular growth plate of the skull, the vertebral segments of the sacrum, the ring epiphyses of the vertebral centra, epiphyses of the scapula and pelvis, the costal notches of the sternum, and the medial epiphysis of the clavicle (see Table 7). Age determination in mature individuals has focused on markers of skeletal senescence, the majority of which are functionally related degenerative changes. Mainly, these have utilized four avenues of investigation [97]: morphological changes in the pubic symphysis [98–103], degeneration of the auricular surface of the sacroiliac joint [104–108], degenerative change in the costochondral junctions of the ribs [57, 58, 109–114], and the degree of closure of the ecto and endocranial sutures [115], although the latter method is considered too imprecise for forensic use (see Table 8 for details of auricular, costal, and pubic methods). Advances in accurate age estimation for adult specimens have been made with the development of tooth crown and root translucency, and radiographic, microscopic or histochemical analysis of teeth or bone [111, 116–121]. While many of these techniques are still in their infancy, they have been shown to have some efficacy in current practice. In general, most adult aging methods have been shown to exhibit wide variation in the accuracy of the age ranges produced when compared to known-age samples. However, they may be useful when used in conjunction with the incidence of other age-related degenerative criteria such as osteoarthroses, osteophytosis, and osteopenia in order to make an informed judgment of overall age-related skeletal change, particularly in individuals between 40 and 60 years of age. It should be borne in mind that all of these techniques have been found to

Table 4 Age estimation in years from the formation and mineralization of the permanent mandibular dentition

	I1	I2	C	P3	P4	M1	M2	M3
Male								
Initial cusp formation	–	–	0.6	2.1	3.2	0.1	3.8	9.5
Coalescence of cusps	–	–	1.0	2.6	3.9	0.4	4.3	10.0
Cusp outline complete	–	–	1.7	3.3	4.5	0.8	4.9	10.6
Crown 1/2 complete	–	–	2.5	4.1	5.0	1.3	5.4	11.3
Crown 3/4 complete	–	–	3.4	4.9	5.8	1.9	6.1	11.8
Crown complete	–	–	4.4	5.6	6.6	2.5	6.8	12.4
Initial root formation	–	–	5.2	6.4	7.3	3.2	7.6	13.2
Initial cleft formation	–	–	–	–	–	4.1	8.7	14.1
Root length 1/4	–	5.8	6.9	7.8	8.6	4.9	9.8	14.8
Root length 1/2	5.6	6.6	8.8	9.3	10.1	5.5	10.6	15.6
Root length 2/3	6.2	7.2	–	–	–	–	–	–
Root length 3/4	6.7	7.7	9.9	10.2	11.2	6.1	11.4	16.4
Root length complete	7.3	8.3	11.0	11.2	12.2	7.0	12.3	17.5
Apex 1/2 closed	7.9	8.9	12.4	12.7	13.5	8.5	13.9	19.1
Apex closed	–	–	–	–	–	–	–	–
Female								
Initial cusp formation	–	–	0.6	2.0	3.3	0.2	3.6	9.9
Coalescence of cusps	–	–	1.0	2.5	3.9	0.5	4.0	10.4
Cusp outline complete	–	–	1.6	3.2	4.5	0.9	4.5	11.0
Crown 1/2 complete	–	–	3.5	4.0	5.1	1.3	5.1	11.5
Crown 3/4 complete	–	–	4.3	4.7	5.8	1.8	5.8	12.0
Crown complete	–	–	4.4	5.4	6.5	2.4	6.6	12.6
Initial root formation	–	–	5.0	6.1	7.2	3.1	7.3	13.2
Initial cleft formation	–	–	–	–	–	4.0	8.4	14.1
Root length 1/4	4.8	5.0	6.2	7.4	8.2	4.8	9.5	15.2
Root length 1/2	5.4	5.6	7.7	8.7	9.4	5.4	10.3	16.2
Root length 2/3	5.9	6.2	–	–	–	–	–	–
Root length 3/4	6.4	7.0	8.6	9.6	10.3	5.8	11.0	16.9
Root length complete	7.0	7.9	9.4	10.5	11.3	6.5	11.8	17.7
Apex 1/2 closed	7.5	8.3	10.6	11.6	12.8	7.9	13.5	19.5
Apex closed	–	–	–	–	–	–	–	–

Reproduced from Ref. 73. © John Wiley & Sons, Inc, 1991, modified from Ref. 75

Table 5 Times of emergence in months of the deciduous dentition

	dI1	dI2	dC	dM1	dM2
Maxilla	10	11	19	16	29
±1 SD	8–12	9–13	16–22	13–19	25–33
Mandible	8	13	20	16	27
±1 SD	6–10	10–16	17–23	14–18	23–31

Reproduced with permission from reference [72]. © Malmo, 1962

Table 6 Times of alveolar emergence in years of the permanent dentition

Male	I1	I2	C	P3	P4	M1	M2	M3
Maxilla	6.2	7.3	11.2	9.8	11.1	5.3	11.4	17.7
±1 SD	0.86	1.29	1.21	1.41	1.60	0.74	1.09	1.52
Mandible	5.9	6.9	9.8	9.6	10.3	5.3	10.8	18.1
±1 SD	0.74	0.78	1.09	1.29	1.72	0.35	1.02	2.15
Female	I1	I2	C	P3	P4	M1	M2	M3
Maxilla	6.1	7.0	9.3	9.0	9.5	5.3	10.3	17.2
±1 SD	0.35	0.90	1.25	1.09	1.37	0.47	0.90	2.46
Mandible	5.8	6.5	8.8	9.1	9.2	5.0	9.9	17.7
±1 SD	0.43	0.55	0.63	0.90	1.64	0.39	1.06	2.34

Reproduced from Ref. 66. © Springer, 1998

be sex and context dependant, and prone to secular influences.

Stature

In the majority of cases, the calculation of the living height of an individual from skeletal remains is based on the fact that overall stature is correlated with long bone length assuming equality of body proportions; metric assessment allows the anthropologist to predict an individual's stature by regressing the lengths of individual limb segments. As body proportions vary by both race and sex (African-Americans, for example, have longer limb bones relative to height than US whites); thus, it is necessary to establish sex and race in order to use the correct regression formulae for the estimation of stature. The elements that are used to reconstruct individual stature include all of the major long bones (femur, tibia, fibula, humerus, radius, and ulna). The bones of the lower limb generally produce more accurate estimation of stature than those of the upper limb. A range of regression equations have been developed and are generally element and population specific; these include those of Badkur and Nath [122], Feldesman and Fountain [123], Genovés [124], Jantz *et al.* [125, 126], and Trotter and Gleser [127–131], amongst others [132–139].

Arguably the most accurate determination of stature (termed the *Fully Method*) is based on aggregated axial and paraxial measurements of the skeletal elements [140], which contribute to living height, but it is difficult to implement in many cases. The measurements comprise the height of the cranium, the total length of the vertebral column from C2 to L5, the anterior height of the first sacral vertebra, bicondylar length of the femur, the length of the tibia, and the height of the calcaneus and talus. Correction factors are applied for the missing soft tissue thicknesses and the method is obviously limited by the completeness of the skeleton under analysis [141, 142]. Estimation of living stature from fragmentary remains is generally considered too inaccurate and statistically inappropriate for forensic applications.

In addition to individualization, the reconstruction of stature and overall body size is routinely used to investigate secular change in archaeological populations through time [29, 138, 143–148]. Changes in stature tend to covary with changes in diet and calorific intake and are therefore potential indicators of periods of environmental or dietary stress, or conversely, of amelioration and excess. Depending on the pattern of expression, these changes may also reflect differential resource acquisition and utilization by sex or social group.

Ancestry or Ethnic Identity

The determination of ancestral or ethnic origin is one of the most problematical and contested areas of forensic and anthropological practice. Anthropologists are divided as to whether "races" actually exist, and most are in agreement that dividing the global population into three or four broad racial groups is not entirely justifiable on biological grounds alone [149–156]. The fundamental basis of the skeletal manifestation of "race" is predicated on the assumption that certain components of the body record evidence of ancestor-descendant relationships (phylogeny) with phenotypic similarity equating to genetic relatedness.

All modern humans share a relatively recent common African ancestor as suggested by the fossil

Table 7 Age estimation for adolescents and young adults from epiphyseal fusion of selected elements[(a)]

Element	Fusion	Age	
		Male	Female
Occipital	Fusion of sphenooccipital synchondrosis	13–18 years	11–16 years
	Closure of jugular growth plate	22–34 years	
Ethmoid	Ethmoid and vomer fuse	20–30 years	
Clavicle	Fusion of medial epiphysis	16–21 years, complete in all individuals by 29+ years	
	Fusion of lateral epiphysis	19–20 years	
Sternum	Sternebra 2 fuses to 3 and 4	11–16 years	
	Sternumebra 1 fuses to mesosternum	15–20 years	
	Sternum essentially complete	21+ years	
	All epiphyseal plaques in costal notches fused	25+ years	
	Xiphoid process commences fusion to mesosternum	40+ years	
Scapula	Coracoid and subcoracoid commence fusion to body	13–16 years	
	Coracoid and subcoracoid fused to body	15–17 years	
	Fusion of glenoid epihysis	17–18 years	
	Fusion of acromial and coracoid epiphysis complete	By 20 years	
	All epiphyses fused and adult form achieved	By 23 years	
Pelvis	Fusion of acetabulum	14–17 years	11–15 years
	Illiac crest commences fusion	17–20 years	
	Anterior iliac spine fused	By 20 years	
	Ischial and iliac crest fully fused	20–23 years	
Sacrum	Fusion between lateral elements and union of lower sacral segments	12–14 years	
	Sacrum complete	25+ years	
Humerus	Distal epiphysis fuses	12–17 years	11–15 years
	Medial epicondyle fuses	14–16 years	13–15 years
	Proximal epiphysis fuses	16–20 years	13–17 years
Radius	Proximal epiphysis fuses	14–17 years	11.5–13 years
	Distal epiphysis fuses	16–20 years	14–17 years
Ulna	Proximal epiphysis fuses	13–16 years	12–14 years
	Distal epiphysis fuses	17–20 years	17 years
Hand	Hook of hamate appears and fuses to body	10–12 years	
	Distal phalangeal epiphyses fuse	16 years	13.5 years
	Base of metacarpal 1 fuses	16.5 years	14–14.5 years
	Proximal and middle phalangeal epiphyses fuse	16.5 years	14–14.5 years
	Heads of metacarpals 2–5 fuse	16.5 years	14.5–15 years
Femur	Head fuses	14–19 years	12–16 years
	Greater trochanter fuses	16–18 years	14–16 years
	Lesser trochanter fuses	16–17 years	
	Distal epiphysis fuses	16–20 years	14–18 years
Tibia	Distal epiphysis fuses	15–18 years	14–16 years
	Proximal epiphysis fuses	15–19 years	13–17 years
Fibula	Distal epiphysis fuses	15–18 years	12–15 years
	Proximal epiphysis fuses	15–20 years	12–17 years
Foot	Calcaneal epiphysis commences fusion	11–14 years	10–12 years
	Distal and middle phalangeal epiphyses fuse	14–16 years	11–13 years
	Heads of metatarsals 2–5 fuse	14–16 years	11–13 years
	Proximal phalangeal epiphyses fuse	16–18 years	13–15 years
	Base of metatarsal 1 fuses	16–18 years	13–15 years
	Calcaneal epiphysis completes fusion	18–20 years	15–16 years

[(a)] Source: Data [36, 37]; table after [51]

Table 8 Comparison of three methods of age at death assessment in the adult skeleton[a]

Method	Phase	Mean	SD	95% CI
Auricular surface (Buckberry and Chamberlain, 2002)		Males and females		
Aging from auricular surface change can be applied regardless of sex and ancestry [107]. Buckberry and Chamberlain (2002) adapted Lovejoy's [108] method of scoring from which they treated changes to the auricular surface as a series of independent variables. The surface is assessed for five traits: transverse organization (five stages), surface texture (five stages), microporosity (three stages), macroporosity (three stages), and morphology of the apex (three stages) using a descriptive scoring system. A composite score is obtained from the sum of each trait, from which an age estimate can be obtained.	I	17.3	1.5	16–19
	II	29.3	6.7	21–38
	III	37.7	13.1	16–65
	IV	51.4	14.5	29–81
	V	59.9	13.0	29–88
	VI	66.7	11.9	39–91
	VII	72.8	12.7	53–92
Rib phase (Iscan, Loth and Wright 1984 and 1985)		White male		
This method uses age-related changes associated with the costochondral junction of the fourth rib, which correspond broadly to age but vary by sex and ancestry. Phase changes are noted in the morphology of the rib end, surface contour, rim edge and contour. In general, the trend is from a nonporous billowed surface with rounded margins to one which becomes more porous, trending from V- to U-shape in contour, with thin walls and a degraded irregular margin.	1	17.3	0.5	17–18
	2	21.9	2.1	21–23
	3	25.9	3.5	24–28
	4	28.2	3.8	26–31
	5	38.8	7.0	34–42
	6	50.0	11.2	44–56
	7	59.2	9.5	54–64
	8	71.5	10.3	65–78
		White female		
Resulting age ranges show little overlap between age classes, leading Loth and Iscan (2000: 245) to describe the technique as "... the most consistently reliable method for the determination of age at death in the adult skeleton". Unfortunately, the efficacy of the method has not been adequately tested; what work has been done suggests the technique is most useful in the 40- to 49-year age ranges [114]. The technique was established using a small sample (108 white males and 83 white females) from Broward County, Florida; as a result, the predicted age ranges should be considered minimal estimates of unknown confidence intervals [5].	1	14.0	–	–
	2	17.4	1.52	16–19
	3	22.6	1.67	21–25
	4	27.7	4.62	24–31
	5	40.0	12.22	34–46
	6	50.7	14.93	43–58
	7	65.2	11.24	59–71
	8	76.4	8.83	70–82
Pubic symphysis (Brooks and Suchey, 1990)		Male		
Age-related changes affecting the pubis have been recognized for many years [97, 98, 100, 103]. All methods perform poorly when tested on known-age samples outwith the demographic population on which the respective aging methods were established. For cases the Suchey–Brooks cast-based scoring system is considered most useful, although the statistical margins of age ranges for each phase are extremely large; in particular, these limit the accuracy of the technique when applied to individuals older than 30 years. The Suchey–Brooks method is based on a six-phase scoring system, which records a shift from pubic surface with billowed ridge and furrows to one of increasing granularity and to smoothing of the surface; the gradual formation of a ventral rampart; plateau formation at the dorsal margin; progressive lipping; the formation of bony outgrowths; and increasing irregularity and erosion.	1	18.5	2.1	15–23
	2	23.4	3.6	19–34
	3	28.7	6.5	21–46
	4	35.2	9.4	23–57
	5	45.6	10.4	27–66
	6	61.2	12.2	34–86
		Female		
	1	19.4	2.6	15–24
	2	25.0	4.9	19–40
	3	30.7	8.1	21–53
	4	38.2	10.9	26–70
	5	48.1	14.6	25–83
	6	60.0	12.4	42–87

[a] Source: Data from [98, 104, 109, 110]

[157–165] and archaeological records of early modern human morphology and behavior [166–169], coupled with assessments of African versus non-African genetic variability and phylogenetic relatedness [170–176]. This is corroborated by analyses of regional mtDNA relationships [177, 178], which root the mtDNA phylogeny firmly within Africa. Founder and maximum parsimony analysis indicates that all non-African populations are derived from a small founder population that most likely established itself to the West of the Indian subcontinent before bifurcating ca. 70 000 years ago into the two major divisional haplogroups, which segregate East Eurasia from West Eurasia. All subsequent human populations are ultimately derived from the African founder population, and share a common network of ancestor-descendant relationships; thus, all modern populations and "racial" groups are epiphenomena in the process of evolving and changing through admixture and hybridization, with any perceived morphological distinctions being the consequence of recent evolution. Consequently, the notion that modern humans can be classified into clear-cut subdivisions of monolithic racial origin groups is rejected by many anthropologists on both biological and ethical grounds [151]. While this standpoint may be true, it is clear that the human body does record phenotypic or genotypic traits that allow anthropologists to partition regional population groups [179].

In general, the most obvious ethnic differences in the living are either displayed in the face or the integumentary system, with the consequence that the skull is the most heavily studied skeletal element in "race" determination. Much of the early work focused on patterns of variation and the study of population movement, particularly in relation to archaeological "peoples", rather than individual discrimination that may be of use in a forensic context. Today, forensic anthropologists generally partition human cranial morphology into four primary skull morphotypes; whether these regional morphotypes equate to genetic or socially constructed "races" is a matter for continuing debate. The groups most commonly used are Caucasoid (including Europeans, Asians from the Indian subcontinent, North and East Africans, Arabs and Mediterraneans), Negroid (including West and southern Africans, and those of African descent such as African-Americans), Mongoloid (including southeast Asian and Palearctic Asiatics, and Native Americans), and Australoid (Australian Aborigines, Pacific Islanders, Fijians and Papuans).

Forensic assessment of ancestral association is carried out using both metric and morphological methods applied to both the cranial and postcranial skeletons [44, 51, 65, 67, 123, 180–186]. The principal cranial morphological differences between the groups are outlined in Table 9 (and see Figure 4) and it is clear that this type of analysis requires consistency and considerable experience observing and understanding skeletal variation between individuals, populations, and age groups. Metrical assessment of ancestry presents its own series of problems in determining morphological affinity, utilizing the main univariate or multivariate discriminant function analyses. Two recent computer-based multivariate methods have been developed on large database samples with the specific purpose of analyzing ancestry and partitioning an unknown specimen into pre-prescribed regional morphological groups. The FORDISC program [66] and CRANID [187] use metric measurements derived from archaeological/ethnographic crania, particularly the W.W. Howells dataset [188–190] or recent cases, in order to determine biological distance in an unknown-origin specimen. In practice, the strength of association of biological affinity is heavily influenced by the type of statistical contrasts used to discriminate between groups, and these programs need to be used with caution and a clear understanding of the underlying statistical and methodological principals.

Markers of Personal Identity

This supplementary set of characters differs from the first four primary factors (sex, age, stature, and ancestry) in that they are traits that are likely to be relatively unique to an individual, or that may be determined so with varying degrees of statistical certainty. They may be fixed characteristics under a high level of genetic control, or they may be subject to change during ontogeny through the addition, removal, or alteration of bodily tissues, and thus seen as phenotypically plastic in nature. Such characteristics may arise as a result of normal variation in development, through chance accident, or intentional modification, with the proviso that they are sufficiently discriminatory between individuals to act as markers of identity. These markers may include a suite of labile characteristics, which can aid in

Table 9 Craniofacial skeletal traits of four principal ancestral groups[a]

Morphology	Caucasoid	Negroid	Mongoloid	Australoid
Frontal profile	Sloping to upright profile	Sloping profile	Upright rounded profile	Sloping profile
Skull shape	Long	Long and low	Long	Short and high
Supraorbital ridges	Moderate-to-strong brow ridges with depressed glabella	Mild-to-moderate brow ridges	Little or no brow ridge expression	Massive brow ridges with glabellar expansion
Interborbital breadth	Narrow	Wide	Moderate	Wide
Face breadth	Narrow to wide	Narrow	Very wide	Wide
Face height	Medium to high	Low	High	Low
Cheek form	Receding malar region	Receding malar with some lateral projection	Strong anterior projection of malar (flat faced)	Strong lateral projection
Orbital shape	Angular orbits	Rectangular orbits	Round orbits	Rectangular orbits
Facial profile	Orthognathic	Pronounced prognathism	Moderate prognathism	Moderate prognathism
Nasal aperture	Narrow	Wide	Moderate	Very wide
Lower nasal margin	Prominent nasal spine with sharp (silled) nasal margin	Short nasal spine with smooth or guttered nasal base	Short nasal spine with sharp nasal margin	Guttering of nasal base
Nasal root	Steepled	Quonset hut	Tented	Steepled
Palate width	Narrow palate	Wide palate	Moderate palate	Very wide palate
Other characteristics	Prominent nose and chin. Deep canine fossa. Long pointed mastoid	Postbregmatic depression. Oblique gonial flare. Oblique posterior mastoid tubercle	Weak or absent canine fossa. Shoveled upper incisors	Marked occipital torus with no external occipital protruberance. Rocker-bottom jaw. Thick cranial vault bones

[a] Source: [51, 179, 191]

forensic individualization such as the uniqueness of skeletal or soft tissue structures, the pathological effects of disease or trauma, surgical intervention for medical or cosmetic purposes, or body modification by nonmedical personnel for aesthetic reasons. Such markers, whether assessed in combination or isolation, should provide a strong evidential basis for the recognition of individual identity.

In practice, anthropologists look for traits that set, for example, one Caucasian male, 25–35 years of age, of 170–175 cm living stature, apart from other missing persons that meet this common description. However, it is an important point to consider that while the fundamental parameters of age, sex, stature, and ancestral origin allow a reduction in potential matches, additional corroborative factors are only of value when preexisting comparative information is available. For example, the above case may show evidence that the male had suffered a broken left tibia that required surgical pining. While potentially individualizing, this information is limited in that it can only aid identification if suitable antemortem records exist to corroborate the match. In current practice, issues of identification are primarily addressed by individually sensitive biological markers such as DNA or fingerprints in preference to anthropological characteristics. However, in situations where such biological markers are unavailable (i.e., degradation of DNA in sea water), or where there is no potential of an antemortem match, then anthropological markers have utility. In cases where an assessment of personal identity fails to pinpoint a specific person, then facial reconstruction may be attempted in order to recreate the appearance of the unknown individual [191], which may then be matched against antemortem photographs of likely missing persons, if such exist.

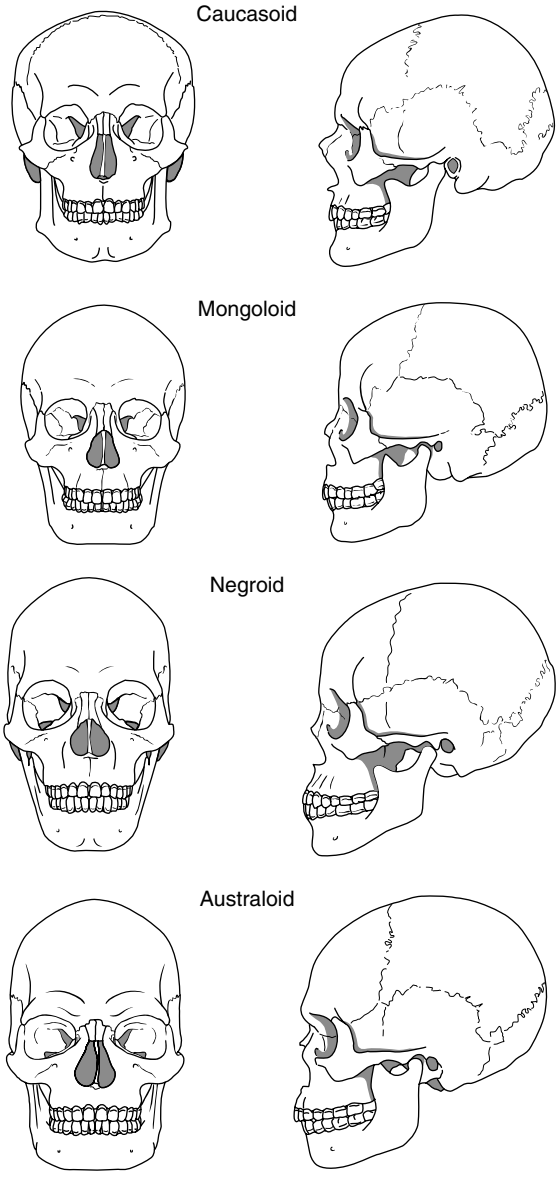


Figure 4 Frontal and lateral views of skulls showing ancestral group morphology [Reproduced with permission from Caroline Needham, University of Dundee.]

Individualizing Characteristics of Normal Human Variation. A number of soft and hard tissue structures have utility in cases of human identification, though in order to be of value such structures must not only be unique to the individual but must also remain stable over time.

Fingerprints or friction skin ridges are the most commonly used individualizing morphological structures. These are created during fetal development on the fingers, palms, toes, and feet. Friction skin ridge development is random, giving each finger or toe a unique set of surface shape characteristics; these are not considered to be genetically determined, and are therefore unique, probably even between identical twins [192].

The arborescent pattern of venous drainage and arterial supply is also thought to be individually unique. In particular, the pattern of epifascial veins superficially exposed at the dorsum of the hand and foot, the wrist, and the cubital fossa may have some utility in identification although this has yet to be statistically verified.

Palatal rugae are soft tissue ridges situated in the anterior part of the palatal mucosa on each side of the median palatal raphe and behind the incisive papilla. Rugae are useful in a forensic context in that they are somewhat protected from trauma by their intraoral location and are insulated from heat by the tongue and buccal fat pads. They have been suggested to be individually unique (even between twins) and minimally unaffected by factors of growth and development. The lengths of individual rugae have been demonstrated to change with age while shape remains largely stable. They have also been shown to resist the imposition of orthodontic treatment or the effects of antemortem tooth avulsion [193, 194].

Macrostructural and trabecular morphology of the sternum, the costochondral junctions, the lumbosacral spine, the pelvis, and skull are also considered to be highly individual. Areas of historical interest have included the size and shape of the frontal, paranasal and mastoid sinuses, endocranial vascular groove patterns, and ectocranial suture patterns [195–199]; in such cases, the comparison of morphological patterns by radiographic superimposition is generally used.

Pathological Conditions. The assessment of pathological variation or traumatic lesions may provide strong evidence as an aid to individuation. This can range from readily recognizable individualizing characteristics that may be known to or observed by informants, such as acne (which in extreme cases can cause significant scarring and skin discoloration), birthmarks, facial scar tissue, missing limbs, additional digits, or cleft palate, to characteristics that

may be cryptic during life such as degenerative joint disease or healed broken bones, the consequences of which may affect normal gait or mobility. The recognition of disease patterns may also assist in individuation, though in using such processes anthropologists are primarily looking at vectors that leave characteristic changes or lesions on the skeleton. The range of diseases that may be skeletally expressed includes trauma, congenital abnormalities (malformations produced by pathological changes during normal intrauterine development), circulatory disorders, joint diseases, infectious disease, diseases of the viscera, metabolic disease, endocrine disorders, hematological disorders, skeletal dysplasias, neoplastic disease, and various diseases and malformations of the dentition [200–202]. However, it is important to realize that not all disease processes leave skeletal lesions, and that any skeletal expression of a disease will be affected by the life stage (developmental age) of the individual concerned, the pathophysiology of the disease itself, and the duration for which the individual has been affected; as such, skeletal lesions of disease will only tend to be expressed in chronic rather than acute cases, and more often in adolescents or adults. They also have a continuum of appearance until they reach a stage of pathological stasis. Since similar skeletal lesions can be associated with different disease etiologies, any estimations based upon differential pathological diagnosis may be tenuous [203]; the difficulty involved in the diagnosis of pathological conditions from dry bone often requires the collaboration of anthropologists, physicians, and pathologists [204] to ensure accurate conclusions are reached.

Identification should never be based on the diagnosis of a pathological condition alone, although the presence of a condition can be useful as it may focus the list of possible identities where antemortem medical records are available. Furthermore, some pathological conditions are often more prevalent to a specific age group, sex or race, and may thus allow for the contextualization of remains of unknown-origin. For example, some disorders such as Legg–Calve–Perthes disease, osteochondritis dissecans, ankylosing spondylitis, or gout show a statistically higher and significant expression in males than females, and may assist in the determination of sex from skeletal remains.

Other distinctive pathological conditions of the skeleton may be due to infection from pathogenic bacteria and fungi, and many of the species involved have a distinct pattern of geographical and corporeal distribution, or more commonly affect a specific age range or sex. For example, hematogenous osteomyelitis is most commonly found in children and adolescents, with only 13% of reported cases involving adults [205] and presents four times more frequently in males [206]. Blastomycosis is caused by the inhalation of the fungus *Blastomyces dermatitidis*; it presents in a male-to-female ratio of at least 5 : 1 with a limited geographical dispersal. The disease is mainly limited to the cooler areas of the United States such as the Mississippi and Ohio valleys, the Mid-Atlantic states and North Carolina, as well as regions of Canada [207]. Hematological diseases such as sickle cell anemia (SCA) and thalassemia show striking differences in racial and geographical incidence, making their presence valuable for individualization of human remains. Since these disorders are present from birth and are associated with a high infant mortality rate [208], they are most frequently clinically observed in children. SCA predominantly affects Blacks of African descent, with a limited number of cases in Mediterranean populations such as Greece, Armenia, and southern Italy [208], whereas thalassemia has a complex geographical dispersal but is primarily found in Mediterranean regions, and is especially prevalent in those of Italian, Greek, and north African ancestry [207].

Trauma may also be a useful aid to individuation. Evidence of antemortem traumatic events may be left upon the skeleton following the decomposition process through the macroscopic recognition of antemortem fracture loci and/or skeletal scar tissue (the formation of a woven bone callus), or microscopically through the evaluation of hairline or cryptic (micro) fracture patterns. Traditionally, anthropologists have been of assistance in documenting and explaining traumatic patterning in relation to violent death through analyses of sharp force, blunt force, ballistic, or explosive trauma effects, the scope of which is outside the remit of this article (for discussion, see [209–212]). However, patterns of healed trauma can be useful in other ways. For example, the biological changes that occur in a postoperative context leave individualizing patterns within skeletal tissue, which continue for months and years, and can provide the anthropologist with a means of corroborating a presumptive identity, with the proviso that

sufficient antemortem clinical documentation exists in order to allow for comparison.

In such cases, the anthropologist is primarily looking at the effects of bone fracture. From a clinical and pathological perspective, a fracture is an interruption in the structural integrity of the bone, and may be variously present from a single crack, fissure, cortical break, dislocation, or bone avulsion, to a complete transverse break. Fractures occur through one (at least) of three mechanisms: (i) a single traumatic event with the application of excessive force; (ii) repeated stress; and (iii) abnormal weakening of the bone (this may arise through disease processes such as osteoporosis, osteogenesis imperfecta, rickets, or certain types of bone neoplasm). The cause of traumatic lesions can range from the accidental effects of crushes, trips or falls, to the consequence of deliberate action such as beatings, manual asphyxia, torture, suspension, electrocution, or the impact of ballistic projectiles or edged weapons. Traumatic lesions that occur antemortem will show evidence of callus formation and healing, and those which take place perimortem or postmortem will show no evidence of cellular recovery; consequently, it may be extremely difficult to distinguish between the two types of fracture in the perimortem and early postmortem periods.

Cultural Modifications. This class of evidence includes surgical interventions and body modifications. These may be undertaken by medically qualified personnel either for clinical or cosmetic purposes, or by the nonmedically trained for aesthetic reasons. Surgical intervention [7] includes surgical procedures such as amputation, orthopedic plating, breast augmentation, the insertion of cardiac pacemakers, or other foreign surgical devices such as staples or wiring (Figures 5 and 6) retained with the body. Some procedures, such as orthopedic joint replacements (Figure 7), require the implantation of prosthetics that may be individually numbered or carry manufacturer's identifying marks. Unfortunately, there are no universal stringent controls regarding the recording of these numbers; therefore, numbered prosthetics are only of use if the source of the prosthetic can be traced, or if specific matching evidence can be obtained, for example, in the form of radiographs. In addition to implantations, amputations can be easily recorded and can be of high value in an identification scenario in terms of reliability of evidence, as it is likely that informants may be aware of the loss of the body part. Scars resulting from surgery may aid in potential identification, particularly if they are in specific regions (for example

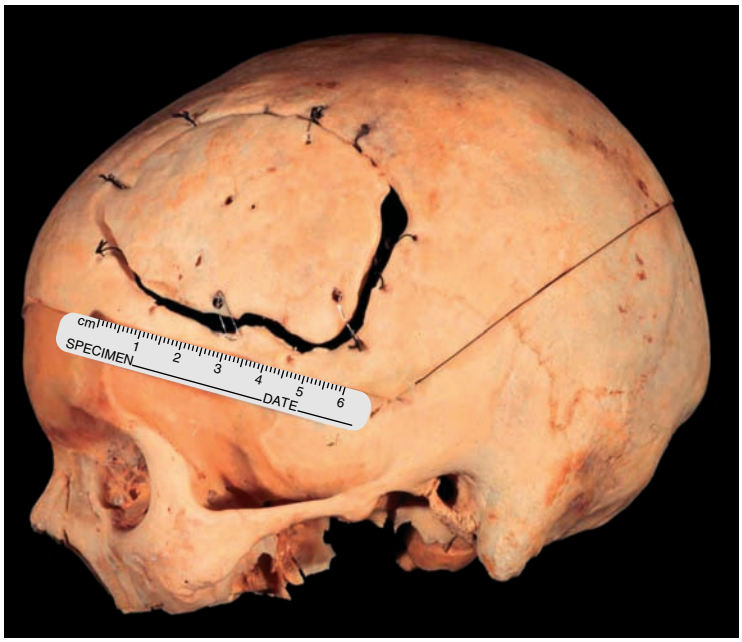


Figure 5 Surgical intervention to the cranium [Reproduced with permission from Michael Warren, University of Florida.]



Figure 6 Radiograph showing surgical artifacts of open-heart surgery



Figure 7 Radiograph showing surgical artifacts of stabilization of the proximal femur

resulting from appendectomies or caesarean sections) as they may guide an investigator toward a likely surgical specialist. The scale of surgical impact can also be important, with scarring and surgical interventions from major events such as car accidents less likely to be closely replicated between individuals, in addition to which most developed countries are likely to retain pre- and postoperative radiographs in such major trauma cases.

Body modification includes permanent and semipermanent tattooing, scarification, branding, cutting, subdermal implants (such as beads), or transdermal piercings such as rings, studs, or bars. There is a degree over overlap between surgical intervention and cosmetic body modification in that some of the more extreme types of alteration (tongue splitting, penile glans splitting, penilectomy, or osteogenic implants such as horns or crests) may involve anesthetics and utilize the services of a trained clinician. All of these classes of cultural/social modification or alteration may provide significant individualizing traits. While the increasing incidence of tattooing within the younger population can be viewed as limiting factors on their individualizing potential, those that contain personal information such as a name or badge of gang or clan membership, or that are extreme in expression or artistically unique, retain their individualizing value. Transdermal piercings including those to the earlobes, nipples, and genitals are now relatively common, but less common forms of alteration, including body sculpting or pocketing, are still relatively unique and likely to be memorable. Anthropologists must ensure that they have a current understanding of trends and developments in the techniques and social anthropology of body modifications (and the subcultural groups that adopt them), which may impact upon personalization of the human form.

Confirmation of Personal Identity

It is an important point to remember when considering the relative weighting attached to identification characteristics that while the fundamental four osteobiographical parameters of age, sex, stature, and ancestral origin allow a reduction in potential matches, additional corroborative factors are only of value when preexisting comparative information is available. As such, it may not always be possible to achieve a confirmed identity given insufficient

presumptive matches. To this end, we can define three categories of identity depending on the quality of the identification criteria utilized ([213] after [214]):

1. Positive or confirmed identity

This occurs when two sets of data are compared and enough specific markers found to conclude that the two sets of records were, in all probability, derived from the same individual. No irreconcilable differences should be noted. Unique markers may include DNA, fingerprints, dentition, and highly individualizing skeletal traits such as numbered surgical implants or radiographic superimposition of frontal sinus morphology.

2. Possible or presumptive identity

This occurs when several individual factors are considered, and although no single factor justifies the establishment of identity, the loadings or weightings attributed to the factors available are sufficient for a possible or presumptive ID. Such factors may include identifiable personal effects, visual recognition based on surviving soft tissues (facial form, body modifications, etc.), confirming osteobiographical data, skeletal pathologies, or other individualizing skeletal traits. In cases of BTB (believed to be), the principle of parsimony is invoked.

3. Exclusion

This arises when all the deceased in an identifiable category such as male or female have been identified, all surviving victims have been accounted for, and BTB have all been confirmed; those individuals who are unable to be matched to a list of presumptive identities are excluded from the identification. It is also possible to exclude with "absolute" certainty in certain cases, particularly where the remains of an unknown individual can be categorically excluded on the basis of an invariant or highly individualizing characteristic; for example, a set of unknown remains determined to be female and 18–25 years of age will clearly not be those of a missing 70-year-old male.

It is clear, therefore, that to limit the definition of forensic anthropology to the process of identification from the skeleton is outdated and inappropriate. The modern forensic anthropologist must be as comfortable working with the living as they are with the dead, and as familiar with the soft tissues as they are with the hard tissues of bones and teeth. It is a highly specialized area of expertise that requires thorough professional education and training.

Acknowledgment

The help and advice of Caroline Needham, Jennifer Randolph-Quinney, Christopher Rynn, and Caroline Wilkinson during the preparation of this manuscript was much appreciated and gratefully acknowledged.

References

[1] Scheuer, L. & Black, S. (2007). Osteology, in *Forensic Human Identification*, T. Thompson & S. Black, eds, CRC Press, Boca Raton, pp. 199–219.

[2] Hill, I. (2007). Physical appearance, in *Forensic Human Identification: An Introduction*, T. Thompson & S. Black, eds, CRC Press, London, pp. 87–97.

[3] Byers, S.N. (2008). *Introduction to Forensic Anthropology*, 3rd Edition, Allyn and Bacon, Boston.

[4] Black, S. (2003). Forensic anthropology–regulation in the U.K, *Science and Justice* **43**, 187–192.

[5] Klepinger, L.L. (2006). *Fundamentals of Forensic Anthropology*, John Wiley & Sons, Hobokoen.

[6] Black, S. & Thompson, T. (2007). Body modification, in *Forensic Human Identification*, T. Thompson & S. Black, eds, CRC Press, Boca Raton, pp. 379–399.

[7] Clarkson, J. & Schaefer, M. (2007). Surgical intervention, in *Forensic Human Identification*, T. Thompson & S. Black, eds, CRC Press, Boca Raton, pp. 127–146.

[8] Cunha, E. (2006). Pathology as a factor of personal identity in forensic anthropology, in *Forensic Anthropology and Medicine: Complementary Sciences from Recovery to Cause of Death*, A. Schmitt, E. Cunha & J. Pinheiro, eds, Humana Press, Totowa, pp. 333–358.

[9] Haglund, W.D. & Sor, M.H. (eds) (1997). *Forensic Taphonomy: The Postmortem Fate of Human Remains*, CRC Press, Boca Raton.

[10] Haglund, W.D. & Sorg, M.H. (eds) (2002). *Advances in Forensic Taphonomy: Method, Theory, and Archaeological Perspectives*, CRC Press, Boca Raton.

[11] Cuijpers, A.G.F.M. (2006). Histological identification of bone fragments in archaeology: Telling humans apart from horses and cattle, *International Journal of Osteoarchaeology* **16**, 465–480.

[12] Guglich, E.A., Wilson, P.J. & White, B.N. (1994). Forensic application of repetitive DNA markers to the species identification of animal tissues, *Journal of Forensic Sciences* **39**, 353–361.

[13] Lowenstein, J., Reuther, J., Hood, D., Scheuenstuhl, G., Gerlach, S. & Ubelaker, D. (2005). Identification of animal species by protein radioimmunoassay of bone fragments and bloodstained stone tools, *Forensic Science International* **159**, 182–188.

[14] Linacre, A. & Lee, J.C. (2005). Species determination: the role and use of the cytochrome b gene, in A. Carracedo, ed, *Methods in Molecular Biology: Forensic DNA Typing Protocols*, Humana Press, Totoawa.

[15] Matsuda, H., Seo, Y., Kakizaki, E., Kozawa, S., Muraoka, E. & Yukawa, N. (2005). Identification

of DNA of human origin based on amplification of human-specific mitochondrial cytochrome b region, *Forensic Science International* **152**, 109–114.

[16] Cattaneo, C., Di Martino, S., Scali, S., Craig, O.E., Grandi, M. & Sokol, R.J. (1999). Determining the human origin of burnt bone: a comparative study of histological, immunological and DNA techniques, *Journal of Forensic Sciences* **102**, 181–191.

[17] Lyman, R.L. (1994). *Vertebrate Taphonomy*, Cambridge University Press, Cambridge.

[18] Outram, A.K., Knüsel, C.J., Knight, S. & Harding, A.F. (2005). Understanding complex fragmented assemblages of human and animal remains: a fully integrated approach, *Journal of Archaeological Science* **32**, 1699–1710.

[19] Adams, B.J. & Byrd, J.E. (2006). Resolution of small-scale commingling: A case report from the Vietnam War, *Forensic Science International* **156**, 63–69.

[20] Adams, B.J. & Byrd, J.E. (eds) (2008). *Recovery, Analysis and Identification of Commingled Human Remains*, Humana Press, Totowa.

[21] Adams, B.J. & Konigsberg, L.W. (2004). Estimation of the most likely number of individuals from commingled human skeletal remains, *American Journal of Physical Anthropology* **125**, 138–151.

[22] Armelagos, G.J. & Van Gerven, D.P. (1980). Sexual dimorphism and human evolution: an overview, *Journal of Human Evolution* **9**, 437–446.

[23] Borgonini, S.M. & Repetto, E. (1986). Methodological considerations on the study of sexual dimorphism in past human populations, *Human Evolution* **1**, 51–66.

[24] Brace, C.L. (1972). Sexual dimorphism in human evolution, *Yearbook of Physical Anthropology* **16**, 31–49.

[25] Butendandt, O., Bidlingmaier, F. & Knorr, D. (1985). Genesis of sexual dimorphism in man: from chromosomes to pubertal hormones, in *Human Sexual Dimorphism*, J. Ghesquiere, R.D. Martin & F. Newcombe, eds, Taylor and Francis, London, pp. 161–168.

[26] Hall, R.L. (1982). Consequences of sexuality, in *Sexual Dimorphism in Homo Sapiens: A Question of Size*, R.L. Hall, ed, Praeger, New York, pp. 3–9.

[27] Jones, D. (1995). Sexual selection, physical attractiveness, and facial neotony, *Current Anthropology* **36**, 723–748.

[28] Plavcan, J.M. (2000). Inferring social behavior from sexual dimorphism in the fossil record, *Journal of Human Evolution* **39**, 327–344.

[29] Wolfe, L.D. & Gray, J.P. (1982). Subsistence practices and human sexual dimorphism of stature, *Journal of Human Evolution* **11**, 575–580.

[30] Alexander, R.D., Hoogland, J.L., Howard, R.D., Noonan, K.M. & Sherman, P.W. (1979). Sexual dimorphism and breeding systems in pinnipeds, ungulates, primates, and humans, in *Evolutionary Biology and Human Social Behaviour: An Anthropological Perspective*, N.A. Chagnon & W. Irons, eds, Duxbury Press, North Scituate, pp. 402–435.

[31] Brace, C.L. & Ryan, A.S. (1980). Sexual dimorphism and human tooth size differences, *Journal of Human Evolution* **9**, 417–435.

[32] Frayer, D.W. & Wolpoff, M.H. (1985). Sexual dimorphism, *Annual Review of Anthropology* **14**, 429–473.

[33] Leutenegger, W. (1982). Sexual dimorphism in nonhuman primates, in *Sexual Dimorphism in Homo Sapiens: A Question of Size*, R.L. Hall, ed, Praeger, New York, pp. 11–36.

[34] Quinney, P.S. & Collard, M.C. (1997). Sexual dimorphism in the mandible of *Homo neanderthalensis* and *Homo sapiens*: morphological patterns and behavioural implications, in A.G.M. Sinclair, E.A. Slater & J.A.J. Gowlett, eds, *Archaeological Sciences 1995*, *Oxbow Monographs in Archaeology*, Oxford, Vol. 65, pp. 420–425.

[35] Willner, L.A. & Martin, R.D. (1985). Some principles of mammalian sexual dimorphism, in *Human Sexual Dimorphism*, J. Ghesquiere, R.D. Martin & F. Newcombe, eds, Taylor and Francis, London, pp. 1–42.

[36] Scheuer, L. & Black, S. (2000). *Developmental Juvenile Osteology*, Academic Press, San Diego.

[37] Scheuer, L. & Black, S. (2004). *The Juvenile Skeleton*, Elsevier Academic Press, Amsterdam.

[38] Hunt, D.R. (1990). Sex determination in the subadult ilia: an indirect test of weaver's non-metric sexing method, *Journal of Forensic Sciences* **35**, 881–885.

[39] Schutkowski, H. (1987). Sex determination of fetal and neonatal skeletons by means of discriminant analysis, *International Journal of Anthropology* **2**, 347–352.

[40] Schutkowski, H. (1993). Sex determination of infant and juvenile skeletons. 1: Morphognostic features, *American Journal of Physical Anthropology* **90**, 199–205.

[41] Holcomb, S.M.C. & Konigsberg, L.W. (1995). Statistical study of sexual dimorphism in the human fetal sciatic notch, *American Journal of Physical Anthropology* **97**, 113–125.

[42] Bulygina, E., Mitteroecker, P. & Aiello, L. (2006). Ontogeny of facial dimorphism and patterns of individual development within one human population, *American Journal of Physical Anthropology* **131**, 432–443.

[43] Buikstra, J.E. & Ubelaker, D.H. (1994). *Standards for Data Collection from Human Skeletal Remains*, *Arkansas Archaeological Survey Research Series*, Vol. 44, Arkansas Archaeological Survey, Fayetteville.

[44] St. Hoyme, L.E. & Iscan, M.Y. (1989). Determination of sex and race: accuracy and assumptions, in *Reconstruction of Life from the Skeleton*, M.Y. Iscan &K.A.R. Kennedy, eds, Alan R. Liss, New York, pp. 53–93.

[45] McCormick, W.F. (1983). Ossification patterns of costal cartilages as an indicator of sex, *Archives of Pathology and Laboratory Medicine* **107**, 206–210.

[46] McCormick, W.F., Stewart, J.H. & Langford, L.A. (1985). Sex determination from chest plate roentgenograms, *American Journal of Physical Anthropology* **68**, 173–195.

[47] Torwalt, C.R.M.M. & Hoppa, R.D. (2005). A test of sex determination from measurements of chest radiographs, *Journal of Forensic Sciences* **50**, 1–6.

[48] Mays, S. & Cox, M. (2000). Sex determination in skeletal remains, in *Human Osteology in Archaeology*, M. Cox & S. Mays, eds, Greenwich Medical Media, London, pp. 117–130.

[49] Walker, P.L. (2008). Sexing skulls using discriminant function analysis of visaully assessed traits, *American Journal of Physical Anthropology* **136**, 39–50.

[50] Loth, S.R. & Iscan, M.Y. (2000). Sex determination, in *Encyclopedia of Forensic Sciences*, J. Siegel, G. Knupfer & P. Saukko, eds, Academic Press, London, pp. 252–260.

[51] Barker, C., Cox, M., Flavel, A., laver, J. & Loe, L. (2008). Mortuary procedures II–skeletal analysis I: basic procedures and demographic assessment, in *The Scientific Investigation of Mass Graves*, M. Cox, A. Flavel, I. Hanson, J. Laver & R. Wessling, eds, Cambridge University Press, Cambridge, pp. 295–382.

[52] Giles, E. & Elliot, O. (1963). Sex determination by discriminant function analysis of crania, *American Journal of Physical Anthropology* **21**, 53–68.

[53] Berrizbeeitia, E.L. (1989). Sex determination with the head of the radius, *Journal of Forensic Sciences* **34**, 1206–1213.

[54] Stewart, T.D. (1979). *Essentials of Forensic Anthropology*, Charles C. Thomas, Springfield.

[55] Guadi-Russo, E. (2007). Sex determination from the talus and calcaneus measurements, *Forensic Science International* **171**, 151–156.

[56] Holman, D.J. & Bennett, K.A. (1991). Determination of sex from arm bone measurements, *American Journal of Physical Anthropology* **84**, 421–426.

[57] Iscan, M.Y. (1985). Osteometric analysis of sexual dimorphism in the sternal end of the rib, *Journal of Forensic Sciences* **30**, 1090–1099.

[58] Iscan, M.Y. & Loth, S.R. (1986). Estimation of age and determination of sex from the sternal rib, in *Forensic Osteology: Advances in the Identification of Human Remains*, K.J. Reichs, ed, C.C. Thomas, Springfield, pp. 68–89.

[59] Josephson, S.C., Juell, K.E. & Rogers, A.R. (1996). Estimating sexual dimorphism by method-of-moments, *American Journal of Physical Anthropology* **100**, 191–206.

[60] Loth, S.R. & Henneberg, M. (1996). Mandibular ramus flexure: a new morphologic indicator of sexual dimorphism in the human skeleton, *American Journal of Physical Anthropology* **99**, 473–486.

[61] Lovell, N.C. (1989). Test of Phenice's technique for determining sex from the os pubis, *American Journal of Physical Anthropology* **79**, 117–120.

[62] Macho, G.A. (1990). Is sexual dimorphism in the femur a "population specific phenomenen"? *Zoologische Mophologie Anthropologie* **78**, 229–242.

[63] Relethford, J.H. & Hodges, D.C. (1985). A statistical test for differences in sexual dimorphism between populations, *American Journal of Physical Anthropology* **66**, 55–61.

[64] Uytterschaut, H.T. (1986). Sexual dimorphism in human skulls: a comparison of sexual dimorphism in different populations, in *Sexual Dimorphism in Living and Fossil Primates*, M. Pickford & B. Chiarelli, eds, Il Sedcesimo, Firenze, pp. 113–120.

[65] Walker, P.T. (2005). Greater sciatic notch morphology: sex, age, and population differences, *American Journal of Physical Anthropology* **127**, 385–391.

[66] Ousley, S.D. & Jantz, R.L. (2005). *FORDISC 3.0 ed*, University of Tennessee, Knoxville.

[67] Giles, E. & Elliot, O. (1962). Race identification from cranial measurements, *Journal of Forensic Sciences* **7**, 147–157.

[68] Liversidge, H.M., Herdeg, B. & Rosing, F.W. (1998). Dental age-estimation of non-adults. A review of methods and principles, in *Dental Anthropology, Fundamentals, Limits and Prospects*, K.W. Alt, F.W. Rosing & M. Teschler-Nicola, eds, Springer, Vienna, pp. 419–442.

[69] Liversidge, H.M., Lyons, F. & Hector, M.P. (2003). The accuracy of three methods of age estimation using radiographic measurements of developing teeth, *Forensic Science International* **131**, 22–29.

[70] Liversidge, H.M. & Molleson, T. (2004). Variation in crown and root formation and eruption of human deciduous teeth, *American Journal of Physical Anthropology* **123**, 172–180.

[71] Liversidge, H.M. & Molleson, T.I. (1999). Developing permanent tooth length as an estimate of age, *Journal of Forensic Sciences* **44**, 917–920.

[72] Liversidge, H.M. & Molleson, T.I. (2004). Variation in crown and root formation and eruption of deciduous teeth, **123**, 172–180.

[73] Haavikko, K. (1970). The formation and the alveolar and clinical eruption of the permanent teeth. An orthopantographic study, *Proceedings of the Finnish Dental Society* **66**, 101–170.

[74] Haavikko, K. (1970). The formation of the alveolar and clinical eruption of the permanent teeth, *Proceedings of the Finnish Dental Society* **61**, 101–170.

[75] Smith, B.H. (1991). Standards of human tooth formation and dental age assessment, in *Advances in Dental Anthropology*, M.A. Kelley & C.S. Larsen, eds, Wiley-Liss, New York, pp. 143–168.

[76] Moorrees, C.F., Fanning, E.A. & Hunt, E.E. (1963). Formation and resporption of three deciduous teeth in children, *American Journal of Physical Anthropology* **21**, 205–213.

[77] Moorrees, C.F., Fanning, E.A. & Hunt, E.E. (1963). Age variation of formation stages for ten permanent teeth, *Journal of Dental Research* **42**, 1490–1502.

[78] Lysell, L., Magnusson, B. & Thilander, B. (1962). Time and order of eruption of the primary teeth: a longitudinal study, *Odontologisk Revy* **13**, 217–234.

[79] Fazekas, I.G. & Kosa, F. (1978). *Forensic Fetal Osteology*, Akedemiai Kiado, Budapest.

[80] Krogman, W.M. & Iscan, M.Y. (1986). *The Human Skeleton in Forensic Medicine*, 2nd Edition, C.C. Thomas, Springfield.

[81] Schaefer, M. & Black, S.M. (2005). Comparison of ages of epiphyseal union in North American and Bosnian skeletal material, *Journal of Forensic Sciences* **50**, 777–784.

[82] Scheuer, J. & Black, S. (2000). Development and ageing of the juvenile skeleton, in *Human Osteology in Archaeology and Forensic Science*, M. Cox & S. Mays, eds, Cambridge University Press, Cambridge, pp. 9–21.

[83] Johnston, F.E. & Zimmer, L.O. (1989). Assessment of growth and age in the immature skeleton, in *Reconstruction of Life from the Skeleton*, M.Y. Iscan & K.A.R. Kennedy, eds, Alan R. Liss, New York, pp. 11–21.

[84] Ubelaker, D.H. (1989). The estimation of age at death from immature human bone, in *Age Markers in the Human Skeleton*, M.Y. Iscan, ed, Charles C. Thomas, Sprinfield, pp. 55–70.

[85] Albert, A.M. & Maples, W.R. (1995). Stages of epiphyseal union for thoracic and lumbar vertebral centra as a method of age determination for teenage and young adult skeletons, *Journal of Forensic Sciences* **40**, 623–633.

[86] Albert, A.M., Ricanek Jr, K. & Patterson, E. (2007). A review of the literature on the aging adult skull and face: Implications for forensic science research and applications, *Forensic Science International* **172**, 1–9.

[87] Kosa, F. (1989). Age estimation from the fetal skeleton, in *Age Markers in the Human Skeleton*, M.Y. Iscan, ed, Charles C Thomas, Springfield, pp. 21–54.

[88] Scheuer, J.L., Musgrave, J.H. & Evans, S.P. (1980). The estimation of late fetal and perinatal age from limb bone length by linear and logarithmic regression, *Annals of Human Biology* **7**, 257–265.

[89] Matsushita, K., Shinoda, K.-I., Akiyoshi, T. & Watanabe, H. (1995). Multivariate analysis of limb long bone growth during the human prenatal period, *The Tohoku Journal of Experimental Medicine* **176**, 109–120.

[90] Hoffman, J.M. (1979). Age estimations from diaphyseal lengths: two months to twelve years, *Journal of Forensic Sciences* **24**, 461–469.

[91] Olivier, G. & Pineau, H. (1958). Determination de l'age du foetus et de l'embryon, *Archives of Anatomy* **6**, 21–28.

[92] Olivier, G. & Pineau, H. (1960). Nouvelle determination de la taille foetale d' apres les longueurs diaphysaires des os longs, *Annales de Medeciene Legale* **40**, 141–144.

[93] Huxley, A.K. & Jimenez, S.B. (1996). Technical note: error in Olivier and Pineau's regression formulae for calculation of stature and lunar age from radial diaphyseal length in forensic fetal remains, **100**, 435–437.

[94] Chamberlain, A.T. (2006). *Demography in Archaeology*, Cambridge University Press, Cambridge.

[95] Cox, M. (2000). Ageing adults from the skeleton, in *Human Osteology in Archaeology and Forensic Science*, M. Cox & S. Mays, eds, Greenwich Medical Media, London, pp. 61–81.

[96] Molleson, T., Cox, M., Waldron, A.H. & Whittaker, D.K. (1993). *The Spitalfields Project Volume 2: The Anthropology–The Middling Sort*. CBA Research Report 86. Council for British Archaeology.

[97] Ascadi, G. & Nemeskeri, J. (1970). *History of Human Life Span and Mortality*, Akademiai Kiado, Budapest.

[98] Brooks, S. & Suchey, J.M. (1990). Skeletal age determination based on the Os pubis: comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods, *Human Evolution* **5**, 227–238.

[99] Gilbert, B.M. & McKern, T.W. (1973). A method for aging the the female os pubis, *American Journal of Physical Anthropology* **38**, 31–38.

[100] Todd, T.W. (1920). Age changes in the pubic bone: I, the white male pubis, *American Journal of Physical Anthropology* **3**, 285–334.

[101] Suchey, J.M. (1979). Problems in the aging of females using the Os pubis, *American Journal of Physical Anthropology* **51**, 467–471.

[102] Katz, D. & Suchey, J.M. (1986). Age determination of the male os pubis, *American Journal of Physical Anthropology* **69**, 435–436.

[103] McKern, T.W. & Stewart, T.D. (1957). *Skeletal Age Changes in Young American Males Analysed from the Standpoint of Age Identification*. Technical Report EP-45. Headquarters Quatermaster Research and Development Command, Natick.

[104] Buckberry, J. & Chamberlain, A. (2002). Age estimation from the auricular surface of the ilium: a revised method, *American Journal of Physical Anthropology* **119**, 231–239.

[105] Mittler, D.M. & Sheridan, S.G. (1992). Sex determination in subadults using auricular surface morphology: a forensic science perspective, *Journal of Forensic Sciences* **37**, 1068–1075.

[106] Mulhern, D.M. & Jones, E.B. (2005). Test of revised method of age estimation from the auricular surface of the ilium, *American Journal of Physical Anthropology* **126**, 61–65.

[107] Murray, K.A. & Murray, T. (1991). A test of auricular surface ageing techniques, *Journal of Forensic Sciences* **36**, 1162–1169.

[108] Lovejoy, C.O., Meindl, R.S., Mensforth, R.P. & Barton, T.J. (1985). Chronological metamorphosis of the auricular surface of the ilium: a new method for the determination of adult skleletal age at death, **68**, 15–28.

[109] Iscan, M.Y., Loth, S. & Wright, R. (1984). Age estimation from the rbs by pase analysis: white males, *Journal of Forensic Sciences* **29**, 1094–1104.

[110] Iscan, M.Y., Loth, S.R. & Wright, R.K. (1985). Age estimation from the rib by phase analysis: white females, *Journal of Forensic Sciences* **30**, 853–863.

[111] Stout, S.D., Dietze, W.H., Iscan, M.Y. & Loth, S.R. (1994). Estimation of age at death using cortical histomorphometry of the sternal end of the fourth rib, *Journal of Forensic Sciences* **39**, 778–784.

[112] Stout, S.D. & Paine, R.R. (1992). Brief communication: histological age estimation using rib and clavicle, *American Journal of Physical Anthropology* **87**, 111–115.

[113] Yoder, C., Ubelaker, D.H. & Powell, J.F. (2001). Examination of variation in sternal rib end morphology relevant to age assessment, *Journal of Forensic Sciences* **46**, 223–227.

[114] Loth, S.R. (1995). Age assessment of the Spitalfields cemetery population by rib phase analysis, *American Journal of Human Biology* **7**, 465–471.

[115] Meindl, R.S. & Lovejoy, C.O. (1985). Ectocranial suture closure: a revised method for the determination of skeletal age at death based on the lateral-anterior sutures, *American Journal of Physical Anthropology* **68**, 57–66.

[116] Lamendin, H., Baccino, E., Humbert, J.F., Tevernier, J.C., Nossintchouk, R.M. & Zerilli, A. (1992). A simple technique for age estimation in adult corpses: the two criteria dental method, *Journal of Forensic Sciences* **37**, 1373–1379.

[117] Baccino, E., Ubelaker, D.H., Hayek, L.-A.C. & Zerilli, A. (1999). Evaluation of seven methods of estimating age at death from mature skeletal remains, *Journal of Forensic Sciences* **44**, 931–936.

[118] Lucy, D., Pollard, A.M. & Roberts, C.A. (1995). A comparison of three dental techniques for estimating age at death in humans, *Journal of Archaeological Science* **22**, 417–428.

[119] Stout, S.D. (1989). The use of cortical histology to estimate age at death, in *Age Markers in the Human Skeleton*, M.Y. Iscan, ed, C.C. Thomas, Springfield, pp. 195–207.

[120] Stout, S.D. (1989). Histomorphometric analysis of human skeletal remains, in *Reconstruction of Life from the Skeleton*, M.Y. Iscan & K.A.R. Kennedy, eds, Alan R. Liss, New York, pp. 41–52.

[121] Stout, S.D. (1992). Methods of determining age at death using bone microstructure, in *Skeletal Biology of Past Peoples*, S.R. Saunders & M.A. Katzenberg, eds, John Wiley & Sons, New York, pp. 21–35.

[122] Badkur, P. & Nath, S. (1990). Use of regression analysis in reconstruction of maximum bone length and living stature from fragmentary measures of the ulna, *Forensic Science International* **45**, 15–25.

[123] Feldesman, M.R. & Fountain, R.L. (1996). 'Race' specificity and the femur/stature ratio, *American Journal of Physical Anthropology* **100**, 207–224.

[124] Genovés, S. (1967). Proportionality of the long bones and their relation to stature among Mesoamericans, *American Journal of Physical Anthropology* **26**, 67–78.

[125] Jantz, R.L. (1992). Modification of the Trotter and Gleser female stature estimation formulae, **32**, 452–466.

[126] Jantz, R.L., Hunt, D.R. & Meadows, L. (1995). The measure and mismeasure of the tibia: implications for stature estimation, *Journal of Forensic Sciences* **40**, 758–761.

[127] Trotter, M. (1970). Estimation of stature from intact long limb bones, in *Personal Identification in Mass Disasters*, T.D. Stewart, ed, Smithsonian Institute Press, Washington, DC, pp. 71–83.

[128] Trotter, M. & Gleser, G.C. (1952). Estimation of stature from long bones of American Whites and Negroes, *American Journal of Physical Anthropology* **10**, 463–514.

[129] Trotter, M. & Gleser, G.C. (1958). A re-evaluation of estimation of stature based on measurements of stature taken during life and of long bones after death, *American Journal of Physical Anthropology* **16**, 79–123.

[130] Trotter, M. & Gleser, G.C. (1977). Corrigenda to estimation of stature from long bones of American Whites and Negroes, *American Journal of Physical Anthropology* **47**, 355–356.

[131] Trotter, M. & Gleser, G.C. (1977). A corrigenda to 'Estimation of stature from long limb bones of American Whites and Negroes', *American Journal of Physical Anthropology* **47**, 355–356.

[132] Lundy, J.K. & Feldesman, M.R. (1987). Revised equations for estimating living stature from the long bones of the South African Negro, *South African Journal of Science* **83**, 54–55.

[133] Xiang-Quing, S. (1989). Estimation of stature from intact long bones of Chinese males, *Canadian Society of Forensic Sciences Journal* **22**, 123–126.

[134] Jason, D. & Taylor, K. (1995). Estimation of stature from the length of the cervical, thoracic, and lumbar segments of the spine in American whites and blacks, *Journal of Forensic Sciences* **40**, 59–62.

[135] Ousley, S. (1995). Should we estimate biological or forensic stature? *Journal of Forensic Sciences* **40**, 768–773.

[136] Petrovecki, V., Mayor, D., Slaus, M., Strinovic, D. & Skavic, J. (2007). Prediction of stature based in radiographic measurements of cadaver long bones: a study of the Croatian population, *Journal of Forensic Sciences* **52**, 547–552.

[137] Ross, A.H. & Konigsberg, L.W. (2002). New formulae for estimating stature in the Balkans, *Journal of Forensic Sciences* **47**, 165–167.

[138] Smith, S.L. (2007). Stature estimation of 3-10-year-old children, *Journal of Forensic Sciences* **52**, 538–546.

[139] Steyn, M. & Smith, J.R. (2007). Interpretation of antemortem stature estimates in South Africans, *Forensic Science International* **171**, 97–102.

[140] Fully, G. (1956). Une novell méthode de détermination de la taille, *Annales de Médecine Légale* **36**, 266–273.

[141] Raxter, M.H., Ruff, C.B. & Auerbach, B.M. (2006). Revision of the fully technique for estimating statures, *American Journal of Physical Anthropology* **130**, 374–384.

[142] Raxter, M.H., Ruff, C.B. & Auerbach, B.M. (2007). Technical note: revised fully stature estimation technique, *American Journal of Physical Anthropology* **133**, 817–818.

[143] Feldesman, M.R. & Kleckner, J.G. (1983). Femur/stature ratio and estimates of stature in Mid- and Late-Pleistocene fossil hominids, *American Journal of Physical Anthropology* **83**, 359–372.

[144] Feldesman, M.R., Kleckner, J.G. & Lundy, J.K. (1990). The femur/stature ratio and estimates of stature in mide- and late-Pleistocene fossil hominids, *American Journal of Physical Anthropology* **83**, 359–372.

[145] Feldesman, M.R. & Lundy, J.K. (1988). Stature estimates for some African Plio-Pleistocene fossil hominids, *Journal of Human Evolution* **17**, 583–596.

[146] Hens, S.M., Konigsberg, L.W. & Jungers, W.L. (2000). Estimating stature in fossil hominids: which regression model and reference sample to use? *Journal of Human Evolution* **38**, 767–784.

[147] Kondo, O., Dodo, Y., Akazawa, T. & Muhesen, S. (2000). Estimation of stature from the skeletal reconstruction of an immature Neandertal from Dederiyeh Cave, Syria, *Journal of Human Evolution* **38**, 457–473.

[148] Humphrey, L. (2000). Growth studies of past populations: an overview and an example, in *Human Osteoloy in Archaeology and Forensic Sciences*, M. Cox & S. Mays, eds, Cambridge University Press, Cambridge, pp. 23–38.

[149] Kennedy, K.A.R. (1995). But Professor, why teach race identification if races don't exist? *Journal of Forensic Sciences* **40**, 797–800.

[150] Brace, C.L. (1995). Region does not mean "race" – reality versus convention in forensic anthropology, *Journal of Forensic Sciences* **40**, 171–175.

[151] Sauer, N.J. (1992). Forensic anthropology and the concept of race: if races don't exist, why are forensic anthropologists so good at identifying them? *Social Science and Medicine* **34**, 107–111.

[152] Smay, D. & Armelagos, G. (2000). Galileo wept: a critical assessment of the use of race in forensic anthropology, *Transforming Anthropology* **9**, 19–29.

[153] Cartmill, M. (1999). The status of the race concept in physical anthropology, *American Anthropolgist* **100**, 651–660.

[154] Lieberman, L. & Reynolds, L.T. (1978). The debate over race revisited: an empirical investigation, *Phylon* **39**, 333–343.

[155] Washburn, S.L. (1962). The study of race, *American Anthropologist* **65**, 521–531.

[156] Littlefield, A., Lieberman, L. & Reynolds, L. (1982). Redefining race: the potential demise of a concept in physical anthropology, *Current Anthropology* **23**, 641–655.

[157] Brauer, G. (1981). New evidence on the transitional period between Neanderthal and modern man, *Journal of Human Evolution* **10**, 467–474.

[158] Brauer, G. (1989). The evolution of modern humans: a comparison of the African and non-African evidence, in *The Human Revolution: Behavioural and Biological Perspectives on the Origins of Modern Humans*, P. Mellars & C.B. Stringer, eds, Edinburgh University Press, Edinburgh, pp. 123–154.

[159] Brauer, G. (1992). Africa's place in the evolution of *Homo sapiens*, in *Continuity or Replacement: Controversies in Homo Sapiens Evolution*, G. Brauer & F.H. Smith, eds, A.A. Balkema, Rotterdam, pp. 83–98.

[160] Brauer, G. & Rimbach, K.W. (1990). Late archaic and modern Homo sapiens from Europe, Africa, and Southwest Asia: craniometric comparisons and phylogenetic implications, *Journal of Human Evolution* **19**, 789–807.

[161] Brauer, G., Yokoyama, Y., Falgueres, C. & Mbua, E. (1997). Modern human origins backdated, *Nature* **386**, 337–338.

[162] Stringer, C.B. (1985). Middle Pleistocene variability and the origin of late Pleistocene humans, in *Ancestors: The Hard Evidence*, E. Delson, ed, Alan R. Liss, New York, pp. 289–295.

[163] Stringer, C.B. (1989). The evolution of *Homo sapiens*: an examination of patterns in fossil hominid data, *L'Extinction* **7**, 121–128.

[164] Stringer, C.B. (1989). The origin of early modern humans: a comparison of the European and non-European evidence, in *The Human Revolution: Behavioural and Biological Perspectives on the Origins of Modern Humans*, P. Mellars & C.B. Stringer, eds, Edinburgh University Press, Edinburgh, pp. 232–244.

[165] Stringer, C.B. (1992). Replacement, continuity, and the origin of Homo sapiens, in *Continuity or Replacement: Controversies in Homo Sapiens Evolution*, G. Brauer & F.H. Smith, eds, Balkema, Rotterdam, pp. 9–24.

[166] Foley, R. & Lahr, M.M. (1997). Mode 3 technologies and the evolution of modern humans, *Cambridge Archaeological Journal* **7**, 3–36.

[167] Foley, R.A. & Lahr, M.M. (1992). Beyond "Out of Africa": reassessing the origins of *Homo sapiens*, *Journal of Human Evolution* **22**, 525–529.

[168] Brooks, A.S. (1996). Behaviour and human evolution, in Meikle W.E., Howell F.C., Jablonski N.G., eds. *Contemporary Issues in Human Evolution*, *California Academy of Sciences Memoir*, California Academy of Sciences, San Francisco, Vol. 21, pp. 135–166.

[169] Brooks, A.S., Helgren, D.M., Cramer, J.S., Franklin, A., Hornyak, W., Keating, J.M., Klein, R.G., Rink, W.J., Schwarcz, H., Smith, J.N.L., Stewart, K., Todd, N.E., Verniers, J. & Yellen, J.E. (1995). Dating and context of three middle stone age sites with bone points in the upper Semliki Valley, Zaire, *Science* **268**, 548–553.

[170] Cann, R.L. (1992). A mitochondrial perspective on replacement or continuity in human evolution, in *Continuity or Replacement: Controversies in Homo Sapiens Evolution*, G. Brauer & F.H. Smith, eds, A.A. Balkema, Rotterdam, pp. 65–74.

[171] Cann, R.L., Stoneking, M. & Wilson, A.C. (1987). Mitochondrial DNA and human evolution, *Nature* **325**, 31–36.

[172] Ayala, F.J. & Escalante, A.A. (1996). The evolution of human populations: a molecular perspective, *Molecular Phylogenetics and Evolution* **5**, 188–201.

[173] Armour, J.A.L., Anttinen, T., May, C.A., Vega, E.E., Sajantila, A., Kidd, J.R. *et al.* (1996). Minisatellite diversity supports a recent African origin for modern humans, *Nature Genetics* **13**, 154–160.

[174] Cvalli-Sforza, L.L., Menozzi, P. & Piazza, A. (1994). *The History and Geography of Human Genes*, Princeton University Press, Princeton.

[175] Hammer, M.F. (1995). A recent common ancestry for human Y chromosomes, *Nature* **378**, 376–378.

[176] Hammer, M.F. & Zegura, S.L. (1996). The role of the Y chromosome in human evolutionary studies, *Evolutionary Anthropology* **5**, 116–134.

[177] Richards, M. & Macaulay, V. (2000). Genetic data and the colonisation of Europe: genealogies and founders, in *Archaeogenetics: DNA and the Population Prehistory of Europe*, C. Renfrew, K. Boyle, eds, McDonald Institute for Archaeological Research, Cambridge, pp. 139–152.

[178] Behar, D.M., Villems, R., Soodyall, H., Blue-Smith, J., Pereira, L., Metspalu, E. *et al.* (2008). The dawn of human matrilineal diversity, *The American Journal of Human Genetics* **82**, 1130–1140.

[179] Iscan, M.Y., Loth, S.R. & Steyn, M. (2000). Determination of racial affinity, in *Encyclopedia of Forensic Sciences*, J. Siegel, G. Knupfer, P. Saukko, eds, Academic Press, London, pp. 227–235.

[180] Craig, E.A. (1995). Intercondylar shelf angle: a new method to determine race from the distal femur, *Journal of Forensic Sciences* **40**, 777–782.

[181] Gill, G.W. & Rhine, S. (eds) (1990). *Skeletal Attribution of Race: Methods for Forensic Anthropology*, Maxwell Museum of Anthropology, Albuquerque.

[182] Buck, T.J. & Vidarsdottir, U.S. (2004). A proposed method for the identification of race in sub-adult skeletons: a geometric morphometric analysis of mandibular morphology, *Journal of Forensic Sciences* **49**, 1–6.

[183] Iscan, M.Y. & Steyn, M. (1999). Craniometric assessment of population affinity in South Africans, *International Journal of Legal Medicine* **112**, 91–97.

[184] Gilbert, B.M. (1976). Anterior femoral curvature: its probable basis and utility as a criterion of racial assessment, *American Journal of Physical Anthropology* **45**, 601–604.

[185] Snow, C.C., Hartman, S.E., Giles, E. & Young, F.A. (1979). Sex and race determination of crania by calipers and computer: a test of the Giles and Elliot discriminant functions in 52 forensic science cases, *Journal of Forensic Sciences* **24**, 448–460.

[186] Edgar, H.J.H. (2005). Prediction of race using characteristics of dental morphology, *Journal of Forensic Sciences* **50**, 1–5.

[187] Wright, R. (2007). *CRANID 5.0 ed*.

[188] Howells, W.W. (1973). *Cranial Variation in Man*, Harvard University, Cambridge, Peabody Museum.

[189] Howells, W.W. (1989). *Skull Shapes and the Map*, Harvard University, Cambridge, Peabody Museum.

[190] Howells, W.W. (1995). *Who's Who in Skulls*, Harvard University Press, Cambridge.

[191] Wilkinson, C. (2004). *Forensic Facial Reconstruction*, Cambridge University Press, Cambridge.

[192] Galloway, V. & Charlton, D. (2007). Fingerprints, in *Forenisc Human Identification*, T. Thompson & S. Black, eds, CRC Press, Boca Raton, pp. 57–72.

[193] English, W.R., Robinson, S.F., Summit, J.B., Oesterle, L.J., Brannon, R.B. & Morlang, W.M. (1988). Individuality of human palatal rugae, *Journal of Forensic Sciences* **33**, 718–726.

[194] Kapali, S., Townsend, G., Richards, L. & Parish, T. (1997). Palatal rugae patterns in Australian Aborigines and Caucasians, *Australian Dental Journal* **42**, 129–133.

[195] Rhine, S. & Sperry, K. (1991). Radiographic identification by mastoid sinus and arterial pattern, *Journal of Forensic Sciences* **36**, 272–279.

[196] Culbert, W.C. & Law, F.M. (1927). Identification by comparison of roentgenograms of nasal accessory sinuses and mastoid processes, *Journal of the American Medical Association* **4**, 1634–1636.

[197] Kahana, T. & Hiss, J. (1994). Positive Identification by means of trabecular bone pattern comparison, *Journal of Forensic Sciences* **39**, 1325–1330.

[198] Kirk, N.J., Wood, R.E. & Goldstein, M. (2002). Skeletal identification using the frontal sinus region: a retrospective study of 39 cases, *Journal of Forensic Sciences* **47**, 318–323.

[199] Mann, R.W. (1998). Use of bone trabeculae to establish positive identification, *Journal of Forensic Sciences* **98**, 91–99.

[200] Ortner, D.J. (2003). *Identification of Pathological Conditions in Human Skeletal Remains*, Academic Press, London.

[201] Aufderheide, A.C. & Rodriguez-Martin, C. (1998). *The Cambridge Encylopedia of Human Paleopathology*, Cambridge University Press, Cambridge.

[202] Resnick, D. & Niwayama, G. (1988). *Diagnosis of Bone and Joint Disorders*, Saunders, Philadelphia.

[203] Cunha, E. (2006). Pathology as a factor of personal identity in forensic anthropology, in *Forensic Anthropology and Medicine: Complementary Sciences from Recovery to Cause of Death*, A. Schmitt, E. Cunha, J. Pinheiro, eds, Humana Press, New Jersey.

[204] Pickering, R.B. & Bachman, D.C. (1997). *The Use of Forensic Anthropology*, CRC Press, Boca Raton.

[205] Aufderheide, A.C. & Rodriguez-Martin, C. (1998). *The Cambridge Encyclopedia of Human Palaeopathology*, Cambridge University Press, Cambridge.

[206] Ortner, D.J. & Putschar, W.G.J. (1981). *Identification of Pathological Conditions in Human Skeletal Remains*, Smithsonian Institution Press, Washington, DC.

[207] Reichs, J.K. (1986). Forensic implications of skeletal pathology: ancestry, in J.K. Reichs, ed, *Forensic Osteology: Advances in the Identification of Human Remains*, Charles C. Thomas, Springfield.
[208] Ortner, D.J. (2003). *Identification of Pathological Conditions in Human Skeletal Remains*, 2nd Edition, Academic Press, San Diego.
[209] Kimmerle, E.H. & Baraybar, J.P. (eds) (2008). *Skeletal Trauma: Identification of Injuries Resulting from Human Rights Abuse and Armed Conflict*, CRC Press, Boca Raton.
[210] Black, S. (2005). Bone pathology and ante-mortem trauma in forensic cases, in *Encyclopedia of Forensic and Legal Medicine*, J. Payne-James, ed, Elsevier, London, pp. 105–113.
[211] Berryman, H.E. & Symes, S.A. (1998). Recognizing gunshot and blunt cranial trauma through fracture interpretation, in *Forensic Osteology: Advances in the Identification of Human Remains*, K.J. Reichs, ed, Charles C. Thomas, Springfield, pp. 333–352.
[212] Galloway, V. (ed) (1999). *Broken Bones: Anthropological Analysis of Blunt Force Trauma*, Charles C. Thomas, Springfield.
[213] Black, S. (2007). Introduction, in *Forensic Human Identification*, T. Thompson & S. Black, eds, CRC Press, Boca Raton.
[214] Jensen, R.A. (1999). *Mass Fatality and Casualty Incidents*, CRC Press, Boca Raton.

Related Articles

Archaeology
Bayesian Networks
Blunt Force Trauma
Botany
Death Penalty and Age
Death: Time of
Earprints: Interpretation of
Expert Opinion: United States
Expert Opinion: United Kingdom, Canada, and Australia
Entomology
Friction Ridge Examination (Fingerprints): Interpretation of
Firearm Examination: Ballistics
Facial Comparison
Gunshot Wounds
Identification and Individualization
Mitochondrial DNA: Profiling
Odontology
Palynology
Soil: Forensic Analysis
Wounds: Sharp Injury
Y-Chromosome Short Tandem Repeats

PATRICK RANDOLPH-QUINNEY, XANTHÉ MALLETT AND SUE BLACK

Anthropology: Age Determination of Remains

Subadults

When dealing with the remains of infants, children, or adolescents, the evaluation of the length of bones and degree of fusion of epiphyses, along with the eruption and mineralization of teeth, can be performed in several manners [1–13].

Fetal and neonatal osteology allows one by looking at the degree of development of the fetus to determine time of gestation. In particular, dental formation may be very specific. The first signs of development can be seen at 6 weeks *in utero*. At 16 weeks, deciduous incisors begin to mineralize; at 26 weeks mineralization of the cusps begins for the first two deciduous molars; and at 30 weeks anterior dental mineralization is marked, and incisors have three-fifths of the crown completed, and the crypts of the second deciduous molars already have five mineralized cusps. In the completed fetus, the cusps of the deciduous molars are mineralized and form a closed circle, and one can see the mineralization of the tip of the first cusp of the first permanent molar.

Similar growth milestones exist for the dimensions of small bones, such as those constituting the skull (Figure 1).

From birth to adolescence, age is then easy to determine, both skeletally and dentally. The simple observation of flat bones such as those in the cranium can give important indications. Skeletons of newborns are thin and fragile, and the bones of the face are minute. During growth, cranial bones gradually

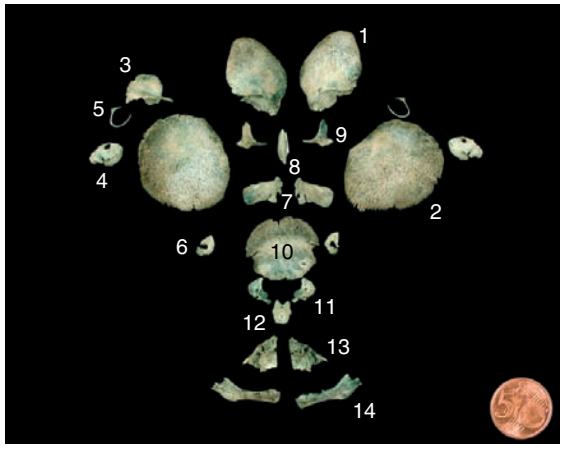


Figure 1 All the bones of a 5-week-old fetus: (1) frontal; (2) parietal; (3) squama of the temporal; (4) petrous portion of the temporal; (5) tympanic ring; (6) smaller wing of the sphenoid; (7) greater wing of the sphenoid; (8) vomer; (9) zygomatic; (10) occipital squama; (11) pars lateralis of the occipital; (12) pars basilaris of the occipital; (13) maxilla; and (14) hemimandible



Figure 2 Three crania in norma frontalis in different phases of growth. To the left a newborn (with no mandible); in the middle a 2-year-old child; and to the right a 10 year old

become thicker, particularly next to the sutures; the cranium quickly assumes adult proportions with a rapid growth of maxillae and mandible (Figure 2).

However, growth in the length of long bone diaphyses (Figure 3), together with the development of teeth, is a more precise indicator of growth.

Dental development and mineralization are perhaps the most precise indicators (Figure 4). Suffice it to say that teeth are less conditioned by environmental and pathological factors when compared to bone. Even when the teeth found are not *in situ* but are scattered, it is possible to achieve an accurate age by verifying the development. Macroscopic analysis



Figure 3 Metrical differences visible at different ages with these three femoral diaphyses: at the bottom a 6-month-old newborn; in the middle a 2-year-old; at the top an 8-year-old. Distal and proximal epiphyses have not been included. Long bones of the upper and lower limbs of infants are more or less of similar dimensions to adult metacarpals, but grow rapidly, so that at a certain length of the diaphysis one can obtain a corresponding age, with a margin error of 6–12 months

is generally insufficient and radiology should be performed in order to verify the exact extent of mineralization of roots.

In the skeleton of infants, long bone extremities (as well as other epiphyses – small "parts" of a bone, which will fuse during growth to the main portion of the same bone – Figure 5), where there is contact with growth cartilage, have a particularly beveled aspect.

Between 10 and 20 years of age, the fusion rates of epiphyses and the order in which they occur are especially useful. For example, proximal femur epiphyses fuse at 18 years of age, those of the iliac crest at 21, and those of the distal extremities of the humerus around 14 (Figure 6).

One should always keep in mind variations between different races. Studies have demonstrated that, for large epiphyses, there can be over two years' difference in age of fusion among different populations. Several years can go by between initial and complete fusion – this also could influence the error.

Adults

Determination of biological age in adult individuals is quite difficult and constitutes one of the most difficult tasks of forensic anthropology. Once dental and osseous development have occurred, very little remains on which one can accurately evaluate aging, if not pathology (arthrosis), degenerative changes,

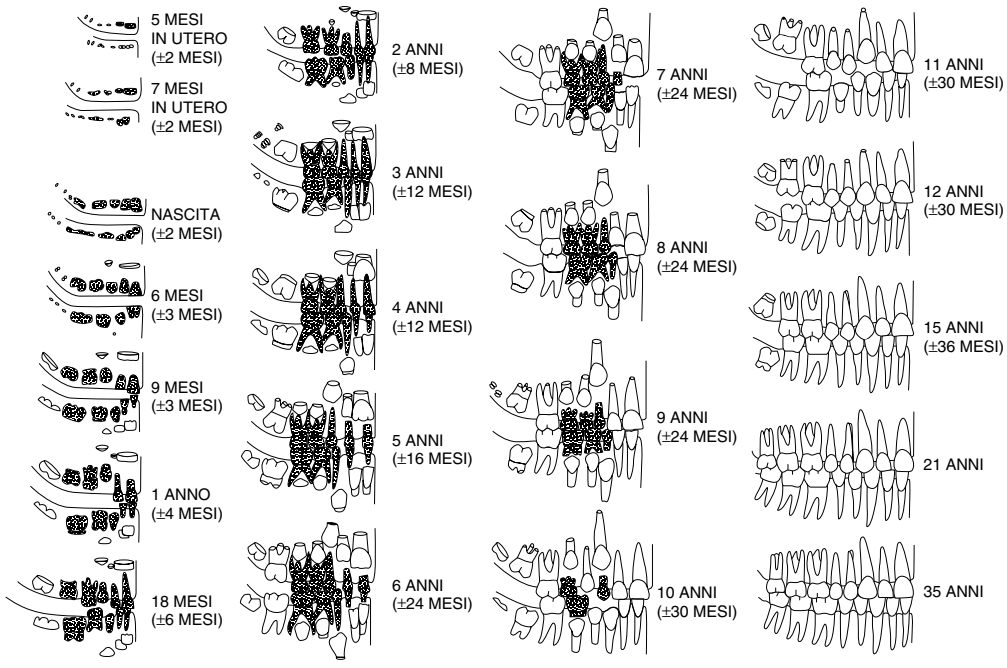


Figure 4 Diagram of dental development and corresponding age. The deciduous dentition is seen in gray and the permanent one in white

histological alterations, and chemical alterations of the components of bone and teeth.

General markers of old age do exist: osteoporosis, articular surface degeneration, osteophytes; however, these are severely influenced by pathology, nutrition, and interindividual variability.

Physical anthropology has established several methods for determining age in adults – macroscopic, microscopic, and biomolecular [14–77] (Figure 7), but most methods have error ranges, which encompass 10–15 years.

Macroscopic methods are the most important ones and consist principally in the degree of degeneration of selected articulations such as the pubic symphysis, osteochondral surface of the fourth rib, and auricular surface of the ilium. The rationale behind these anatomical sites is that they are articulations, which are equally stressed for all individuals, regardless of activity. Thus, they should reveal a degree of degeneration, which should be proportional only to the age and thus standardizable for aging procedures. Less reliable techniques are cranial bone suture analysis and radiological evaluation of bone rarefaction.

Pubic Symphysis

In young adults, the symphyseal surface of the pubic bones (i.e., the joint between the two pubic bones) has a billowed appearance, with deep and furrowed bone ridges running across the surface in parallel. As the person ages, the furrows gradually become filled and the surface becomes flatter and even. Around this surface, an oval outline is formed and an initial crest begins to form on the ventral margin. Then the surface starts to deteriorate. Numerous methods were devised for aging by pubic symphysis. The most acclaimed one, however, is the Suchey–Brookes method applicable to both male and female skeletons. This method divides the degenerative evolution into six stages, which correspond to mean age and age ranges. Phases are shown in Figure 8(a–f).

Fourth Rib

This method was developed by Iscan *et al.* and evaluates the profile of the osteochondral articulation of the rib, its depth, and the profile of the margin. With increase in age, the profile becomes more



Figure 5 Different phases of formation of the innominate bone. It consists of three portions (ileum, ischium, and pubis), which unite around the second year of life. Before adolescence, the innominate bone is made of three different bones: the ileum, whose principal nucleus of ossification appears at the second or third week of life *in utero*; the ischium, whose principal ossification nucleus appears at the fourth month *in utero*; and the pubis, which ossifies at around the fifth month *in utero*. The rami of ischium and pubis fuse at around 7–8 years of age. At 12 years of age, the cartilaginous strip, which separates the acetabulum from the three components, begins to ossify. Complete fusion occurs around 17 years of age. Other epiphyses of these bones, i.e., the iliac crest, the anterior superior iliac spine, and the ischiatic tuberosity, appear around puberty and fuse to the main body between 16 and 23 years of age. In the figure above one can see (1) ileum and ischium of a 5-month-old fetus; (2) ileum, ischium, and pubis of a 2-year-old child; (3) ileum, ischium, and pubis of a 11 year old where ischium and pubis are fused; (4) a 17 year old where the three main bones are fused but the lines of fusion are still partly visible within the acetabulum; (5 and 6) a 16–17 year old where the iliac crest has not yet fused to the rest of the ilium (note the billowing); (7) fusion of the iliac crest and ischiatic tuberosity has occurred; and (8) detail of the ischial tuberosity in fusion

irregular with margins, which go from a V form to a U form; depth increases and sternal insertion is more irregular (Figure 9a, b). These alterations are expressed in eight phases for males and females. One disadvantage of this method is that at times, with commingled remains, it is difficult to identify the fourth rib, or it may not be available.

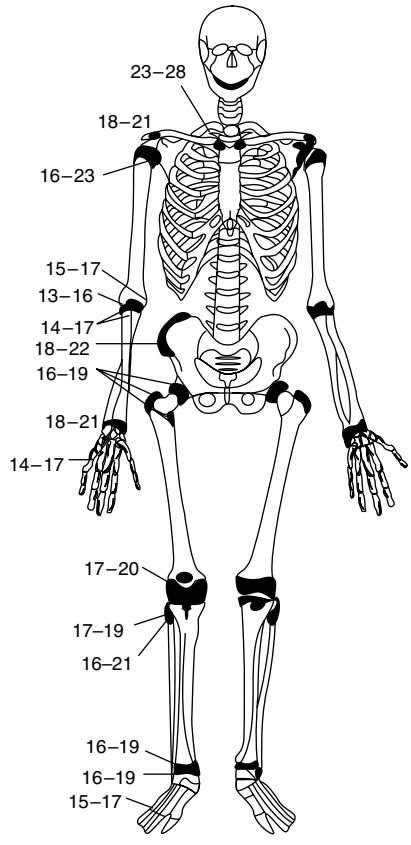


Figure 6 Diagram of a skeleton representing the age range of principal fusion events of the epiphyses with the respective diaphyses. One should always, however, keep in mind that these data are based on modern control populations, which belong to a specific race and social class, and therefore could be erroneous and at variation when applied on a different population

Auricular Surface

This method was devised by Lovejoy *et al.* in 1985 and is based, as for the pubic symphysis, on the degeneration, in eight phases, of the auricular surface of the ilium, according to the degeneration of the apex, the superior half of the surface, the inferior half, the retroauricular area, granulosity, density, and micro- and macroporosity of the surface.

Less accurate methods that involve the observation of suture closure both ecto and endocranial exist. These methods involve the evaluation of the degree of suture closure and give a score, which results in a specific age with a very large error range.

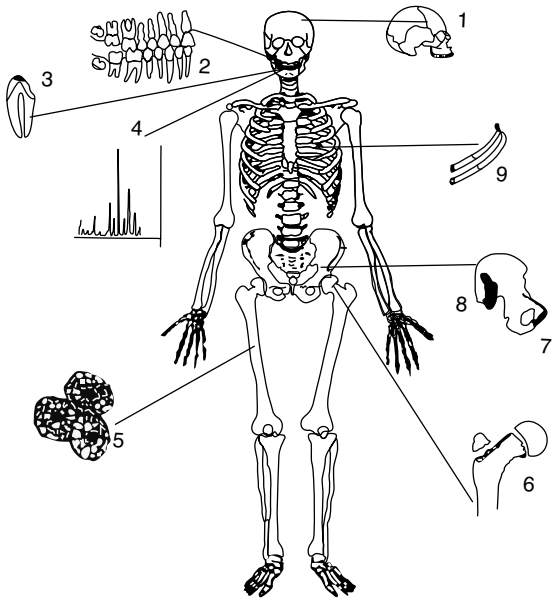


Figure 7 Schematic representation of various methods of aging adults: (1) cranial sutures; (2) third molar eruption; (3) dental transparency methods; (4) aminoacid racemization; (5) osteons; (6) fusion of last epiphyses to fuse; (7) pubic symphysis; and (8) auricular surface

Macroscopic alterations of teeth are also relevant. In particular, dentine translucency seems important – a variable that increases with age. As far as dental methods are concerned, the Lamendin is the most reliable among dental tests: it utilizes periodontal disease and translucency of root in monoradicular teeth throughout the following *formula*:

$$\text{Age} = (0.18P') + (0.42T') + 25.53 \qquad (1)$$

where $P' = P/L \times 100; T' = T/L \times 100; L$, teeth length; P, height of periodontosis; and T, height of root translucency.

The Lamendin test is considered by the literature the most accurate method of age estimation, especially between 40 and 60 years, but recent studies have demonstrated a variability in accordance with race and gender, requiring specific correction factors. It has, however, replaced the Gustafson–Johanson test, which is based on abrasion, secondary dentin deposition, periodontosis, root translucency, secondary cementum, root reabsorption, with a score from 0 to 3 for each element. However, new radiological methods, which study pulp chamber dimensions, are being introduced into the forensic scenario but still need testing.

Recently a "two-step procedure" has been developed for quickly achieving age: if the idea one has of the age of an individual by looking at the skeleton is that it is within the first three stages of the Suchey–Brookes pubic symphysis method, then this should be the method of choice. If the subject is classified in the last three phases, then one should switch to the Lamendin technique in order to assess age more accurately. Currently, age diagnosis at over 60 remains a problem.



Figure 8 (a–f) Appearances of the six different stages of the Suchey–Brookes method in which one can note the progressive alteration of the symphyseal surface. In order, fase I, much billlowing can be seen, average ages are expressed at 67% CI. phase I: in females 17–22, in males 16–21; phase II, reduction in billowing and a slight margin is beginning to form, females 20–30, males 20–27; phase III, billowing still visible, circumferential margin almost complete, females 23–39, males 22–35; phase IV complete margin, females 27–49, males 22–35; phase V, loss of billowing, degeneration starts, females 34–63, males 35–56; and phase VI alteration and degradation of the surface and margin, females 48–72, male 49–73



(a)



(b)

Figure 9 (a) The sternal end of a young fourth rib, where one can note the regular margin and round shape of the bottom. (b) The sternal end of an old fourth rib where one can note osteophytes, irregularity, and even the presence of calcified cartilage (yellow) at the margins

If macroscopic methods cannot be performed (for fragmentation, loss of the anatomical areas), one must then proceed to microscopic methods. The rationale behind the microscopic method is that since bone remodeling occurs constantly throughout life, an older bone will have more osteons and osteon fragments compared with a younger bone (Figures 10 and 11). Microscopic analysis of a bone section from long bone diaphyses can provide an age estimation throughout several methods, the most acclaimed being the Kerley–Ubelaker method [40], which counts the number of secondary osteons, of osteon fragments, and of primary osteons in a 100x microscopic field with a diameter of 1.62 mm; the method can be applied to tibial, femoral, and fibular diaphyses. The numbers are then inserted into an appropriate formula.

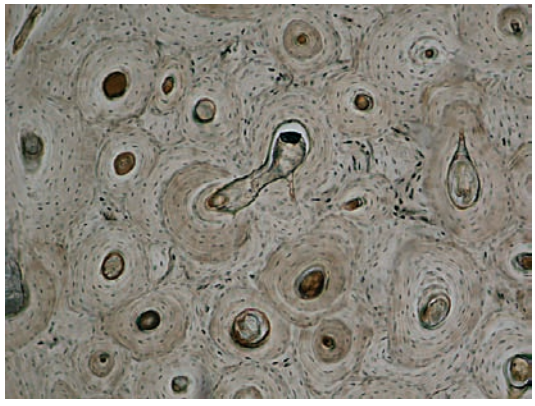


Figure 10 Bone section at a magnification of 100×: one can note, in this, thin section of femoral human bone and the presence of numerous osteons. The section belongs to an elderly individual

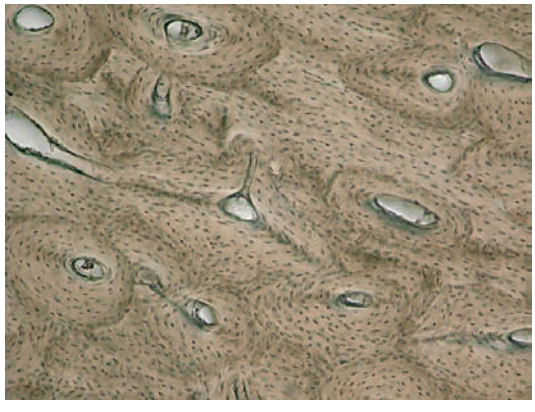


Figure 11 Similar section of a young individual. One can note the presence of a smaller concentration of osteons

Many other macroscopic and microscopic methods are available, among which are evaluation of the acetabulum and cementum annulation, respectively. These are, however, more difficult to apply and/or more costly, less user-friendly, or still need to be tested on larger populations.

Finally, there are also chemicophysical methods, which seem to yield a smaller error range. These are based on the racemization of aspartic acid. The aminoacids of which proteins are made of are present in nature in two stable forms: in the form of L-enantiomers and D-enantiomers. Living organisms synthesize proteins, which contain only L-aminoacids. In tissues such as teeth, D-aminoacids

will accumulate during life because of the racemization of L-aminoacids in the measure of about 0.1% every year. Levels of D-aminoacid seem to accumulate during life in a linear manner with age; thus, the evaluation of the L/D ratio is closely related to age. This technique requires the use of gas chromatography or high performance liquid chromatography (HPLC) and has already been tested on enamel, dentine, cartilage, and bone. The method, however, up to now, seems applicable only to relatively fresh teeth, and not on ancient or burnt material, is more expensive, and is more difficult to set up.

Finally, one should conclude the issue of aging adults with an invitation to caution, keeping in mind the large error ranges.

References

[1] Fazekas, I.G. & Kosa, K. (1978). *Forensic Fetal Osteology*, Akademiai Kiado Publishers, Budapest.

[2] De Salvia, A., Calzetta, C., Orrico, M. & De Leo, D. (2004). Third mandibular molar radiological development as an indicator of chronological age in European population, *Forensic Science International* **146**(Suppl S9-S12),.

[3] Harris, E.F. (2007). Mineralization of the mandibular third molar: a study of American blacks and whites, *American Journal of Physical Anthropology* **132**, 98–109.

[4] Blankenship, J.A., Mincer, K.K., Anderson, K.M., Woods, M.A. & Burton, E.L. (2007). Third molar development in the estimation of chronologic age in American blacks compared with whites, *Journal of Forensic Sciences* **52**, 428–433.

[5] Prieto, J.L., Barberia, E., Ortega, R. & Magana, C. (2005). Evaluation of chronological age based on third molar development in the Spanish population, *International Journal of Legal Medicine* **119**, 349–354.

[6] Maber, M., Liversidge, H.P. & Hector, M.P. (2006). Accuracy of age estimation of radiographic methods using developing teeth, *Forensic Science International* **159**(Suppl 1), S68–S73.

[7] Martin-de las Heras, S., Garcia-Fortea, P., Ortega, A., Zodocovich, S. & Valenzuela, A. (2008). Third molar development according to chronological age in populations from Spanish and Magrebian origin, *Forensic Science International* **174**(1), 47–53.

[8] Crowder, C. & Austin, D. (2005). Age ranges of epiphyseal fusion in the distal tibia and fibula of contemporary males and females, *Journal of Forensic Sciences* **50**, 1001–1007.

[9] Schaefer, M.C. & Black, S.M. (2005). Comparison of ages of epiphyseal union in North American and Bosnian skeletal material, *Journal of Forensic Sciences* **50**, 777–784.

[10] Scheuer, L. & Black, S. (2000). *Developmental Juvenile Osteology*, Academic Press, New York.

[11] Schmeling, A., Schulz, R., Reisinger, W., Mühler, M., Wernecke, K.D. & Geserick, G. (2004). Studies on the time frame for ossification of the medial clavicular epiphyseal cartilage in conventional radiography, *International Journal of Legal Medicine* **118**(1), 5–8.

[12] Schmeling, A., Geserick, G., Reisinger, W. & Olze, A. (2007). Age estimation, *Forensic Science International* **165**(2–3), 178–181.

[13] Slaus, M. & Tomicic, Z. (2005). Discriminant function sexing of fragmentary and complete tibiae from medieval Croatian sites, *Forensic Science International* **147**, 147–152.

[14] Ahlquist, J. & Damsten, O. (1969). A modification of Kerley's method for the microscopic determination of age in human bone, *Journal of Forensic Sciences* **14**, 2052.

[15] Baccino, E., Ubelaker, D.H., Hayek, L.A. & Zerilli, A. (1999). Evaluation of seven methods of estimating age at death from mature human skeletal remains, *Journal of Forensic Sciences* **44**(5), 931–936.

[16] Baccino, E. & Zerilli, A. (1997). The two-step strategy (TSS) or the right way to combine a dental (Lamendin) and an anthropological (Suchey-Brooks method) for age determination, *49th Annual Meeting of the American Academy of Forensic Sciences*, NewYork, February 17–22, 1997.

[17] Brooks, S. & Suchey, J.M. (1990). Skeletal age determination based on os pubis: a comparison of the Acsadi-Nemeskeri and Suchey Brookes methods, *Human Evolution* **5**, 227.

[18] Buckberry, J.L. & Chamberlain, A.T. (2002). Age estimation from the auricular surface of the ilium: a revised method, *American Journal of Physical Anthropology* **119**, 231–239.

[19] Martrille, L., Ubelaker, D.H., Cattaneo, C., Seguret, F., Tremblay, M. & Baccino, E. (2007). Comparison of four skeletal methods for the estimation of age at death on white and black adults, *Journal of Forensic Sciences* **52**(2), 302–307.

[20] Castellana, C. & Kosa, F. (2001). Estimation of fetal age from dimension of atlas and axis ossification centers, *Forensic Science International* **117**(1–2), 31–43.

[21] Cho, H., Stout, S.D., Madsen, R.W. & Streeter, M.A. (2002). Population-specific histological age-estimation method: a model for known African-American and European-American skeletal remains, *Journal of Forensic Sciences* **47**(1), 12–18.

[22] Djuric, M., Djonic, D., Nikolic, S., Popovic, D. & Marinkovic, J. (2007). Evaluation of the Suchey-Brooks method for aging skeletons in the Balkans, *Journal of Forensic Sciences* **52**(1), 21–23.

[23] Ericksen, M.F. (1991). Histologic estimation of age at death using the anterior cortex of the femur, *American Journal of Physical Anthropology* **84**, 171.

[24] Foti, B., Adalian, P., Signoli, M., Ardagna, Y., Dutour, O. & Leonetti, G. (2001). Limits of the

Lamendin method in age determination, *Forensic Science International* **122**(2–3), 101–106.
[25] Galera, V., Ubelaker, D.H. & Hayek, L.A. (1998). Comparison of macroscopic cranial methods of age estimation applied to skeletons from the Terry collection, *Journal of Forensic Sciences* **43**(5), 933.
[26] Ginter, J.K. (2005). A test of the effectiveness of the revised maxillary suture obliteration method in estimating adult age at death, *Journal of Forensic Sciences* **50**(6), 1303–1309.
[27] Greulich, W.W. & Pyle, S.I. (1959). *Radiographic Atlas of Skeletal Development of the Hand and Wrist*, Stanford University Press, Stanford.
[28] Huda, T.F.J. & Bowman, J.E. (1995). Age determination from dental microstructure in juveniles, *American Journal of Physical Anthropology* **97**, 135.
[29] Igarashi, Y., Uesu, K., Wakebe, T. & Kanazawa, E. (2005). A new method for the estimation of adult skeletal age at death from the morphology of the auricular surface of the ilium, *American Journal of Physical Anthropology* **128**, 324–329.
[30] Iscan, M.Y. (1989). *Age Markers in the Human Skeleton*, Charles C. Thomas Publisher, Springfield.
[31] Iscan, M.Y. & Loth, S.R. (1985). Age estimation from the rib by phase analysis: white females, *Journal of Forensic Sciences* **30**(3), 853–863.
[32] Iscan, M.Y. & Loth, S.R. (1986). Determination of age estimation from the sternal rib in white females: a test of the phase method, *Journal of Forensic Sciences* **31**(3), 990–999.
[33] Iscan, M.Y. & Loth, S.R. (1986). Determination of age estimation from the sternal rib in white males: a test of the phase method, *Journal of Forensic Sciences* **31**(1), 122–132.
[34] Iscan, M.Y., Loth, S.R. & Wright, R.K. (1984). Age estimation from the rib by phase analysis: white males, *Journal of Forensic Sciences* **29**(4), 1094–1104.
[35] Iscan, M.Y., Loth, S.R. & Wright, R.K. (1987). Racial variation in the sternal extremity of the rib and its effect on age determination, *Journal of Forensic Sciences* **32**(2), 452–466.
[36] Iscan, M.Y., Loth, S.R. & Wright, R.K. (1984). Metamorphosis at the sternal rib end: a new method to estimate age at death in white males, *American Journal of Physical Anthropology* **65**(2), 147–156.
[37] Kashyap, V.K. & Koteswara, N.R. (1990). A modified Gustafson method of age estimation from teeth, *Forensic Science International* **47**, 237–247.
[38] Klepinger, L.L., Katz, D., Micozzi, M.S. & Carroll, L. (1992). Evaluation of cast methods for estimating age from the os pubis, *Journal of Forensic Sciences* **37**, 763–770.
[39] Kerley, E.R. (1965). The microscopic determination of age in human bone, *American Journal of Physical Anthropology* **23**(2), 149–163.
[40] Kerley, E.R. & Ubelaker, D.H. (1978). Revisions in the microscopic method of estimating age at death in human cortical bone, *American Journal of Physical Anthropology* **49**(4), 545–546.
[41] Lamendin, H., Baccino, E., Humbert, J.F., Tavernier, J.C., Nossintchouk, R.M. & Zerilli, A. (1992). A simple technique for age estimation in adult corpses: the two criteria dental method, *Journal of Forensic Sciences* **37**(5), 1373–1379.
[42] Lovejoy, C.O., Meindl, R.S., Pryzbeck, T.R. & Mensforth, R.P. (1985). Chronological metamorphosis of the auricular surface of the ilium: a new method for the determination of adult skeletal age at death, *American Journal of Physical Anthropology* **68**(1), 15–28.
[43] Lucy, D. & Pollard, A.M. (1995). Further comments on the estimation of error associated with the Gustafson dental age estimation method, *Journal of Forensic Sciences* **40**(2), 222–227.
[44] Lucy, D., Aykroyd, R.G., Pollard, A.M. & Solheim, T. (1996). A bayesian approach to adult human age estimation from dental observations by Johanson's age changes, *Journal of Forensic Sciences* **41**(2), 189–194.
[45] Maples, W.R. & Rice, P.M. (1979). Some difficulties in the Gustafson dental age estimations, *Journal of Forensic Sciences* **24**(1), 168–172.
[46] Mann, R.W., Jantz, R.L., Bass, W.M. & Willey, P.S. (1991). Maxillary suture obliteration: a visual method for estimating skeletal age, *Journal of Forensic Sciences* **36**(3), 781–791.
[47] Masset, C. (1989). Age estimation on the basis of cranial suture, in *Age Markers in the Human Skeleton*, M.Y. Iscan, ed, Charles C. Thomas, Springfield.
[48] Meindl, R.S. & Lovejoy, C.O. (1985). Ectocranial suture closure: a revised method for the determination of skeletal age at death based on the lateral-anterior suture, *American Journal of Physical Anthropology* **68**(1), 57–66.
[49] Megyesi, M.S., Ubelaker, D.H. & Sauer, N.J. (2006). Test of the Lamendin aging method on two historic skeletal samples, *American Journal of Physical Anthropology* **131**(3), 363–367.
[50] Miles, A.E.W. (1963). The dentition in the assessment of individual age in skeletal material, in *Dental Anthropology*, D.R. Brothwell, ed, Pergamon Press, Oxford.
[51] Mincer, H.H., Harris, E.F. & Berryman, H.E. (1993). The A.B.F.O. study of third molar development and its use as an estimator of chronological age, *Journal of Forensic Sciences* **38**(2), 379–390.
[52] Murray, K.A. & Murray, T. (1991). A test of the auricular surface aging technique, *Journal of Forensic Sciences* **36**(4), 1162–1169.
[53] Mulhern, D.M. & Jones, E.B. (2005). Test of revised method of age estimation from the auricular surface of the ilium, *American Journal of Physical Anthropology* **126**(1), 61–65.
[54] Norris, S.P. (2002). Mandibular ramus height as an indicator of human infant age, *Journal of Forensic Sciences* **47**(1), 8–11.

[55] Oettle, A.C. & Steyn, M. (2000). Age estimation from sternal ends of ribs by phase analysis in South African blacks, *Journal of Forensic Sciences* **45**(5), 1071–1079.
[56] Ohtani, S., Yamada, Y., Yamamoto, T., Arany, S., Gonmori, K. & Yoshioka, N. (2004). Comparison of age estimated from degree of racemization of aspartic acid, glutamic acid and alanine in the femur, *Journal of Forensic Sciences* **49**(3), 441–445.
[57] Ohtani, S., Abe, I. & Yamamoto, T. (2005). An application of D- and L-aspartic acid mixture as standard specimens for the chronological age estimation, *Journal of Forensic Sciences* **50**(6), 1298–1302.
[58] Osborne, D.L., Simmons, T.L. & Nawrocki, S.P. (2004). Reconsidering the auricular surface as an indicator of age at death, *Journal of Forensic Sciences* **49**(5), 905–911.
[59] Pasquier, E., Pernot, L., Burdin, V., Mounayer, C., Le Rest, C., Colin, D., Mottier, D., Roux, C. & Baccino, E. (1999). Determination of age at death: assessment of an algorithm of age prediction using numerical three-dimensional CT data from pubic bones, *American Journal of Physical Anthropology* **108**, 261–268.
[60] Prince, D.A. & Ubelaker, D.H. (2002). Application of Lamendin's adult dental aging technique to a diverse skeletal sample, *Journal of Forensic Sciences* **47**(1), 107.
[61] Ritz-Timme, S., Cattaneo, C., Collins, M.J., Waite, E.R., Schutz, H.W., Kaatsch, H.J. & Borrman, H.J.M. (2000). Age estimation: the state of the art in relation to the specific demands of forensic practise, *International Journal of Legal Medicine* **113**(3), 129–136.
[62] Ritz-Timme, S., Schutz, H.W. & Collins, M.J. (1999). Evaluation of aspartic acid racemisation ratios in the human femur for age estimation, *Journal of Forensic Sciences* **44**, 874–876.
[63] Rouge-Maiilart, C., Telmon, N., Rissech, C., Malgosa, A. & Rouge, D. (2004). The determination of male adult age at death by central and posterior coxal analysis – a preliminary study, *Journal of Forensic Sciences* **49**(2), 208–214.
[64] Sahni, D., Jit, I., Neelam, B. & Sanjeev, G. (2005). Time of closure of cranial sutures in northwest Indian adults, *Forensic Science International* **148**, 199–205.
[65] Sarajlic, N. (2006). Age estimation based on sternal rib ends changes in Bosnian male population, *Medicinski Arhiv* **60**(6), 343–346.
[66] Sarajlic, N., Cihlarz, Z., Klonowski, E.E., Selak, I., Brkic, H. & Topic, B. (2006). Two criteria dental aging method applied to a Bosnian population: comparison of formulae for each tooth group versus one formula for all teeth, *Bosnian Journal of Basic Medical Sciences* **6**, 78–83.
[67] Schmitt, A., Murail, P., Cunha, E. & Rouge, D. (2002). Variability of the pattern of aging on the human skeleton: evidence from bone indicators and implications on age at death estimation, *Journal of Forensic Sciences* **47**(6), 1203–1209.
[68] Schmitt, A., Murail, P., Ha, E. & Rouge, D. (2002). Variability of the pattern of aging on the human skeleton: evidence from bone indicators and implications on age at death estimation, *Journal of Forensic Sciences* **47**(6), 1193–1199.
[69] Sitchon, M.L. & Hoppa, R.D. (2005). Assessing age-related morphology of the pubic symphysis from the digital images versus direct observation, *Journal of Forensic Sciences* **50**(4), 791–795.
[70] Solari, A.C. & Abramovitch, K. (2002). The accuracy and precision of third molar development as an indicator of chronological age in Hispanics, *Journal of Forensic Sciences* **47**(3), 531.
[71] Soomer, H., Ranta, H., Lincoln, M.J., Penttila, A. & Leibur, E. (2003). Reliability and validity of eight dental age estimation methods for adults, *Journal of Forensic Sciences* **48**(1), 149–152.
[72] Stout, S.D. (1998). The application of histological techniques for age at death determination, in *Forensic Osteology : Advances in the Identification of Human Remains*, K.J. Reichs, ed, Charles C. Thomas Publisher, Springfield.
[73] Stout, S.D. & Stanley, S.C. (1991). Percent osteonal bone versus osteon counts: the variable of choice for estimating age at death, *American Journal of Physical Anthropology* **86**, 515–519.
[74] Telmon, N., Gaston, A., Chemia, P., Blanc, A., Joffre, F. & Rougè, D. (2005). Application of the Suchey-Brooks method to three-dimensional imaging of the pubic symphysis, *Journal of Forensic Sciences* **50**(3), 507–512.
[75] Warren, M.W. (2000). Radiographic determination of developmental age in fetus and stillborns, *Journal of Forensic Sciences* **44**(4), 708–712.
[76] Yekkala, R., Meers, C., Van Schpdael, A., Hoogmartens, J., Lambrichts, I. & Willems, G. (2006). Racemization of aspartic acid from human dentin in estimation of chronological age, *Forensic Science International* **159**(Suppl 1), S89–S94.
[77] Yoder, C., Ubelaker, D.H. & Powell, J.F. (2001). Examination of variation in sternal rib end morphology relevant to age assessment, *Journal of Forensic Sciences* **46**(2), 223–227.

Related Articles

Anthropology: Aging the Living

Histology

Length Measurement

Odontology

CRISTINA CATTANEO

Anthropology: Aging the Living

The need to determine the age of living individuals constitutes a problem of increasing interest in our community, due to the progressively higher number of persons not in possession of any document of identity, who have immigrated illegally, and committed crimes, whose real age must be known in order to decide their imputability and whether they should be subject to trial as "of age" (e.g., 18 or more years of age). Many of these subjects actually do not know their real age; others, on the other hand, try lying about their real age in order not to go to an adult's prison. Judicial problems, which emerge, are aimed at determining whether the perpetrator of a more or less severe crime has reached "adult" age. This may be different in different countries; however, normally, the ages of 14, 16, 18, and 21 are the crucial thresholds for most American and European countries. The consequences are severe: if underage, the perpetrator will be brought to a juvenile prison, a much different environment with respect to a normal prison. Even in proceedings concerning adoption, it is sometimes important to assess age when no birth certificate is available.

Although such issues concern radiology, auxology, and pediatrics, they enter the realm of anthropology also since they deal with a classical anthropological domain such as aging. Aging has already been discussed in **Anthropology: Age Determination of Remains** with respect to human remains; however, the issue of aging the living deserves a specific section since usually more narrow error ranges are required – in other words, more accurate methods.

The anthropologist must learn to deal with concepts such as biological age, chronological age, skeletal age, and dental age. Chronological age is basically anagraphical age – the real age of a person. Biological age, on the other hand, is a statistical concept based on calculation of the degree of maturation of the body (e.g., sexual maturation), the skeletal system, or dentition. Thus, chronological age may not necessarily coincide with biological age, since they represent two completely different concepts. An expert witness, however, can only respond for biological age. An additional problem concerns race or ancestry. Even if progressive maturation of the skeletal and dental system follows a sequence, which is common to all individuals, different populations may reach different maturation stages at different times, and therefore at a different chronological age.

The evaluation of growth in individuals can be performed by the evaluation of anthropometric variables such as stature, weight, pubertal maturation, and bone and dental growth. Body development follows precise phases, which repeat themselves in the same manner for all individuals of the same sex, for the first 20 years of life. Somatic development, particularly dental and skeletal development, comprise well known and standardized stages. For this reason, it is preferable to choose to evaluate dental and skeletal development, by performing radiological exams on the person involved. And because of the specific legal requirements, aging has focused on those peculiar anatomical sites, which give more accurate results in the age ranges that are crucial to legal issues, such as the hand, wrist, and teeth. These methods not only allow one to give a rough estimate of age with a known error, but some, particularly the dental ones, may allow one to give the probability that the person has reached a certain age, for example, 18 years – which can be extremely useful for a judge who needs to know the entity of the risk he is taking. So, if a certain stage of development corresponds to a probability of having reached the 18th year of age, for example, of 90%, the judge will risk a "wrong" conviction only in 10% of cases.

In the last decade, forensic anthropology and forensic medicine, in general, have shown an increasing interest in this problem and in the reliability of methods for assessing biological age [1–13]. Among the methods most frequently used for skeletal assessment are those concerning the left hand wrist area, which can produce precise estimates up to the age of 16–17 years (Figures 1 and 2), at which time wrist maturation is complete in 90% of subjects. The development of teeth, and particularly of third molars, has been extensively studied for this purpose also.

A Study Group on Forensic Age Diagnostics has recently given guidelines and recommendations on this matter. It declares that in order to determine the age of living subjects, the assessment should include a physical examination, X ray of the hand and wrist, dental examination, and an orthopantomogram. computed tomography (CT) examination of the collar bones should be performed if the person is at least



Figure 1 X ray of the left hand of a 15-year-old individual. Notice how the epiphyses of phalanges, metacarpals, and ulna and radius are still unfused



Figure 2 X ray of the left hand of an 18-year-old individual. Notice how, with respect to the image in Figure 1, all epiphyses have fused

18 years old, since this is the only anatomical area, which will be able to give more precise information on the age range between 18 and 21.

A more difficult issue is aging living adult humans. This, once again, occurs in cases where individuals have no identity documents for the purpose of old-age pension or other civil acts. As has been said in the "aging human remains section", once maturation is complete, all one is left with are bone and dental degeneration. It is common experience that there are general markers of aging, such as wrinkles, white hair, posture etc., which, however, are not accurate due to the great interindividual variability. If the error range is large with human remains, where one can, however, attempt to look at pubic symphysis and other articular surfaces, perform osteon counts or look at dentine translucency, these methods are

not applicable in the live person. Some authors have suggested testing for amino acid racemization, although this implies performing a biopsy of the tooth. Other radiological methods, which observe the dimensions of the dental pulp chamber, are beginning to make their way; they, however, still need further testing.

Sometimes it is necessary to "age" pictures in order to provide reliable identikits of missing persons whose last picture may have been taken years before disappearing. It is, therefore, necessary to age that picture in order to perform the so-called face aging of photos. Aging process of photos has been carried out so far without a large-scale scientific approach. It is usually left to artists' subjective elaboration, although some researchers (few) are working toward the elaboration of software, which may use a slightly more scientific approach.

One of the main problems concerning the aging of living individuals represented in pictures is juvenile pornography. Once again, two-dimensional pornographic images can be the object of medicolegal/anthropological assessment. According to the country and legislation involved, the question frequently asked is the age of the child or adolescent (if, for example, under 10, 14, 16, or 18 years). This is a novel and very difficult aspect of age estimation, since facial and secondary sexual characteristics are extremely variable and do not necessarily represent chronological age. Since the late 1990s, in fact, authors have cautioned against the misuse of standard puberty stages, such as the Tanner ones, to estimate chronologic age. More scientific articles are stressing the differences in sexual maturation rates within different geographical areas. Pathologists, anthropologists, pediatricians, and gynecologists are really left with very little in order to fight juvenile pornography. In this sense, recent research in Europe is oriented toward the study of facial parameters, but is only at the beginning.

A great increase in diffusion of pedo-pornographic material has occurred in the last years, particularly due to the development of web technologies. Along with the technical progress, child pornography has increased proportionally. The misuse of the Internet as a (crime) weapon is a serious problem.

Technically, however, the difficulty lies in actually verifying whether the material is pedo-pornographic, i.e., if the subjects represented are – at present – below 18 (e.g., Italy, France, Canada, USA), 16 (e.g., Belgium, Switzerland, Netherlands, Great Britain) or 14 (e.g., Germany, Austria) years of age. Therefore, it often happens that specialists such as forensic pathologists, pediatricians, or anthropologist are called in by magistrates or the police in order to establish, on videotapes and photos, subject age. One of the most popular, but imprecise for these purposes, methods for age estimation is staging pubertal growth according to the Tanner criteria, utilized in clinical medicine to verify sexual maturation in adolescents; the Tanner method focuses on pubic hair and mammary glands in females, and on pubic hair, penis, and testicles in males. Use of Tanner staging for forensic purposes has been severely criticized. The author himself quotes that using the Tanner stages to estimate the probable chronological age is a "wholly illegitimate use of Tanner staging" and he wishes to "caution pediatricians and other physicians to refrain from providing "expert" testimony as to chronologic age based on Tanner staging ".

Great interindividual variability in sexual maturation is emphasized, bound to physiological and pathological factors, such as obesity, and to environmental influence. Another important limit is that Tanner staging focuses on clinical elements in order to establish which phase of sexual maturation the subjects have reached, not if the subject is of adult age. A further problem is that it is impossible to compare photographic staging with a complete medical examination, including inspection and palpation, since the method is applied to two-dimensional pictures. Furthermore, clues can be altered on pictures, e.g., by shaving the pubic or axillary hair. Dental eruption and development could be useful also; but frequently, detail of the images is not good enough to allow for analysis of dental outlines and features. Another method used in age diagnosis, not yet analyzed in depth, and apparently more subjective and disputable, is based on observation of facial traits. Morphological assessment of juvenile characters, however, may be extremely tricky. A more objective approach may be provided by facial metrics. However, research is at its initial stages.

References

[1] Cameriere, R., Ferrante, L., Belcastro, M.G., Bonfiglioli, B., Rastelli, E. & Cingolani, M. (2007). Age estimation by pulp/tooth ratio in canines by peri-apical X-rays, *Journal of Forensic Science* **52**, 166–170.

[2] Cameriere, R., De Angelis, D., Ferrante, L., Scarpino, F. & Cingolani, M. (2007). Age estimation in children by measurement of open apices in teeth: a European formula, *International Journal of Legal Medicine* **121**(6), 449–453.
[3] Cameriere, R., Brkic, H., Ermenc, B., Ferrante, L., Ovsenik, M. & Cingolani, M. (2008). The measurement of open apices of teeth to test chronological age of over 14-year olds in living subjects, *Forensic Science International* **121**(6), 449–453.
[4] Greulich, E. & Pyle, F. (1972). *Atlas of Skeletal Growth*.
[5] Parent, A., Teilman, G., Juul, A., Skakkebaek, N.E., Toppari, J. & Bourguignon, J. (2003). The timing of normal puberty and age limits of sexual precocity: variations around the world, secular trends and changes after migration, *Endocrine Reviews* **24**(5), 668–693.
[6] Roche, A.F., Cameron Chumlea, W. & Thissen, D. (1988). *Assessing the Skeletal Maturity of the Hand-Wrist: Fels Method*, Charles C. Thomas Publisher, Springfield.
[7] Rosenbloom, A.L. & Tanner, J.M. (1998). Misuse of Tanner puberty stages to estimate chronologic age, *Pediatrics* **102**(6), 1494.
[8] Schmeling, A., Reisinger, W., Geserick, G. & Olze, A. (2006). Age estimation of unaccompanied minors Part I. General considerations, *Forensic Science International*, **159**(Suppl), S61–S64.
[9] Schmeling, A., Olze, A., Reisinger, W. & Geserick, G. (2001). Age estimation of living people undergoing criminal proceedings, *Lancet* **358**, 89.
[10] Schmeling, A., Reisinger, W., Loreck, D., Vendura, K., Markus, W. & Geserick, G. (2000). Effects of ethnicity on skeletal maturation: consequences for forensic age estimations, *International Journal of Legal Medicine* **113**, 253.
[11] Tanner, J.M., Whitehouse, R.H. & Cameron, N. (1983). *Assessment of Skeletal Maturity and Prediction of Adult Height (TW2 Method)*, Academic Press, London.
[12] Tanner, J.M., Whitehouse, R.H., Marshall, W.A., Healy, M.J. & Goldstein, H. (1975). *Assessment of Skeletal Maturity and Prediction of Adult Height (TW Method)*, Academic Press, London.
[13] Sun, S.S., Scubert, C.M., Chumlea, W.C., Roche, A.F., Kulin, H.E., Lee, P.A. & Ryans, A.S. (2002). National estimates of the timing of sexual maturation and race differences among US children, *Pediatrics* **110**(5), 911–919.

Related Articles

Anthropology: Age Determination of Remains
Length Measurement
Odontology

CRISTINA CATTANEO

Anthropology: Ancestry and Stature Determination

Race or Ancestry

The word race seems to bring with it considerable discussion among biologists, sociologists, and even politicians and intellectuals. It can be criticized as being politically and biologically incorrect – and has been by many in the past decade. However, one cannot ignore that differences between populations do exist and that these may be fundamental for the identification of the victim of a crime. Thus, forensic anthropologists need to use "racial" traits in order to help investigating authorities.

So-called racial or ethnic characteristics can be considered as those differences that can be observed between populations and not among single individuals. Weight, for example, is extremely variable within a specific population; skin color and the type of hair may be less variable. From an evolutionary point of view, when a species lives in different environments, the best adapted genotypes are selected for survival. In the present sociocultural reality, thanks to the exchange of genetic heritage due to the increase in communication and acculturation, we have seen a progressive integration of different groups of populations – and it is therefore more difficult to distinguish specific racial groups. Ancestry is thus difficult to assess, since there is a large variation within races but also a great deal of overlapping between them. Nonetheless, the presence of concentrations of extreme expressions of some body traits can suggest affiliation with a particular major racial group.

The three main racial groups, easy to identify on human remains, are Caucasoids, Negroids, and Mongolids. Generally speaking, Caucasoids show high pigmentation variability, including dark skin and white skin types and red-haired subjects, thin and smooth hair, orthognathous profiles, and prominent chins, and height is often included within intermediate stages. Negroids show a variable coloring of skin from brown to black, dark hair and irises, variable height; the cranium is long, without prominent supraorbital reliefs, marked zygomas, alveolar prognathism, and receding chins. Mongolids have

yellow–brown skin color, dark and straight hair, variable height, high orbits, prominent zygomas, and shovel-shaped upper incisors.

Classical authors stress that an indication of racial affinity may be reached by metrical measurements of bone districts; the main measurements utilized in race diagnosis are the schelic index (lower limb length/height × 100), the nasal index (nasal width/nasal length × 100), morphological facial index (*nasio-gnathion* length/*zygion–zygion* length × 100), palate metrical analysis, and femural neck angle. These measurements show a high variability among different ethnic groups and change in conformity with geographical context. Prognathism (the forward jutting of the jaw) has an importance in determination of main racial groups. High degrees of prognathism are a common characteristic of Negroids, whereas low degrees of prognathism are observed in Mongolids. Histological analysis of skin in the case of putrefied and mummified remains provides some information on race, and has a great importance where skin color cannot be recognized, because of decomposition or conservative *postmortem* processes. Distribution of melanin pigments is also useful in remains where soft tissue residues are still present. Caucasoid and Negroid groups show a regular and continuous increase of the melanin stratum, whereas Mongolids have a typical melanin-flared disposition in the Malpighian stratum. Hair microscopic analysis also can be useful and used to distinguish between main races by observing its cross section: a round section is typical of Mongolids; oval sections of Caucasoids; and an elliptic section of Negroids (Figure 1).

Some information on race can also be reached throughout the analysis of bone and dental morphology. Caucasoid (White) crania tend to have receding cheek bones, which give a more pointed appearance

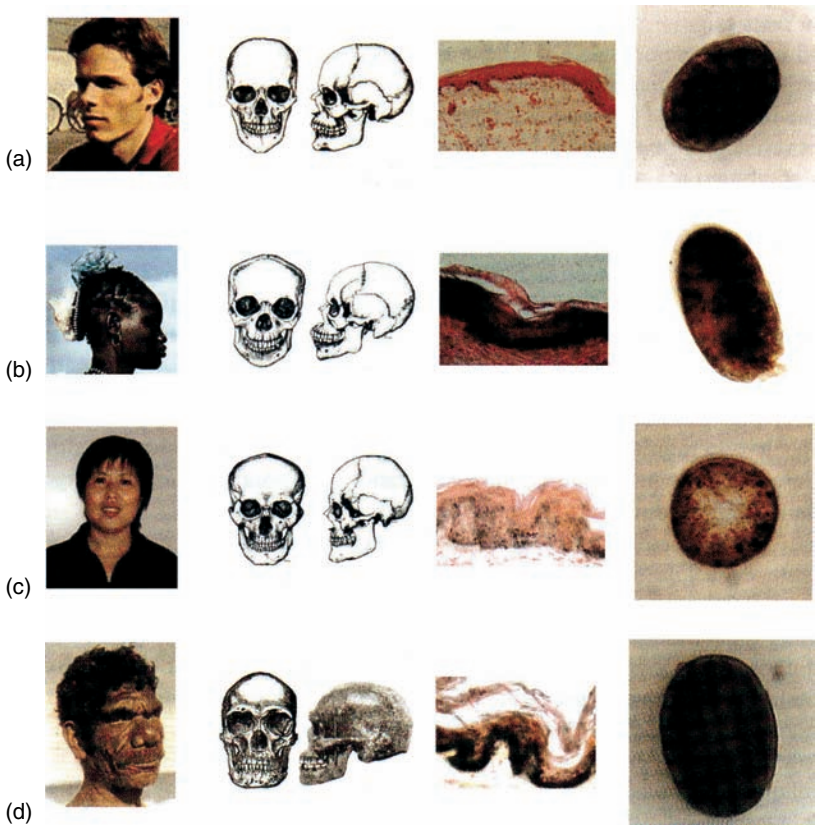


Figure 1 Each strip shows the appearance of face, cranium, skin histology, and hair microscopic section of the main ethnic groups in forensic anthropology: (a–d) Caucasoid, Negroid, Mongolid, and Australomelanesoid



Figure 2 (a) Cranium of a Caucasoid woman. (b) A Negroid woman. Notice the greater nasal aperture, absence of a nasal sill and more rectangular orbits in the Negroid



Figure 3 Lateral view of Caucasoid cranium (a) and Negroid cranium. (b) One can appreciate the marked prognathism in the Negroid

to the face. The nasal aperture is narrow and has a prominent and thick lower border (nasal sill) and a marked nasal spine. The palate is relatively narrow and the suture between maxilla and palatine bones seems to be curved. Negroids (Blacks) have more rectangular orbits. Nasal aperture is large and there is lack of a nasal sill; the nasal spine is absent or very small. The palate is wider and rectangular, and there is marked prognathism (Figures 2, 3a, b). This is the protrusion of the maxillary and mandibular region and can be evaluated via the "pencil test". This consists in trying to place the nasal sill, the mandibular symphysis, and the occlusal surface of the upper incisors on the same line. If one achieves this, then the cranium should be Caucasoid, if not, it is Negroid. Mongolids tend to have forward projecting malar bones. Nasal aperture has a slightly pointed lower margin. Orbits are circular. Facial flattening is also evident. A simple test for this is: position the cranium face up and balance a pencil across the nasal aperture; try to insert a finger between the maxillary bone and the pencil, if you can do it without moving the pencil it is probably Caucasoid, if not, it is Mongolid. Mongolids frequently have head-to-head occlusion of teeth.

Dental peculiarities may also indicate race. Diastemas (the space between the two central upper incisors) is more frequent among Negroids and

shovel-shaped incisors among Mongolids, although these are not exclusive traits.

Metric parameters of the cranium and other bones also may be helpful in determining race. The University of Tennessee has developed an interactive computer program called *Fordisc*, which performs a diagnosis of race from several cranial measures. The method is user-friendly but has some limits: the program compares the specific cranial measurements with a control sample taken from the US population. Its applicability in other parts of the world has only been partly tested. However, innumerous studies exist on racial variability of different parts of the skeleton, which are, year after year, building up information for the delicate problem of diagnosing ancestry from human remains [1–19].

Recently, researchers have focused not only on race but also on detecting signs of the country of provenance of that person. In fact, some radioisotopes typical to specific countries are stored in the bones of living individuals. The presence of such elements may be good indicators of whether that person spent a considerable time of his or her life in Eastern Europe as opposed to North Africa or other regions, for example. This represents a very interesting field whose developments may bring more information on the country of origin of an unidentified skeleton.

Stature

Height can be assessed by standardized mathematical formulae developed by several authors according to race and gender, and based mainly on long bone length, simply measured with an osteometric table (Figure 4). The bones most frequently used are femur, fibula, tibia, humerus, radius, and ulna. Formulae applied to each bone in different races are shown in Table 1.

Literature has also focused on the use of several other different bones, tested within different populations, as well as with equations including several bones [20–37].

Anomalies and Pathology

The more information one can obtain from the skeleton, the better for identification purposes. For example, dental work (*see* **Odontology**) and bone calluses can be fundamental in building a precise biological profile, which may lead to identification, just like tattoos and scars will be useful for well-preserved cadavers. Anthropologists are also very keen on recording nonmetric traits, which are anatomical anomalies, such as the presence of Wormian bones, metopic suture, or of double articular facets. Although these may be important in identifying the subject, if antemortem X rays show the presence of the same anomaly, they still do not have an extreme relevance in forensic anthropology, although they should always be recorded. The forensic anthropologist generally focuses on osteological signs, which give crucial details on the habits or health of a person. An important discipline in this sense is paleopathology [38]. This is the science that studies the signs of disease on skeletal remains and is very useful in telling us whether a person suffered from severe anemia, was lame because of an old bone fracture, or had tuberculosis. Below is a list of some disease categories detectable on bone, which may turn out useful for the biological profile.



Figure 4 An osteometric table to measure the length of bones – one of the classical anthropologist's main tools (along with sliding and spreading calipers)

Table 1 Main formulae for height estimation (from Trotter and Gleser) where F, T, H, R, Fi, and U are the maximum length expressed in centimeters for femur, tibia, humerus, radius, fibula, and ulna, respectively

	Femur	
	Males	Females
Caucasoids	$2.38 \times F + 61.41 \pm 3.27$	$2.47 \times F + 54.10 \pm 3.72$
Negroids	$2.11 \times F + 70.35 \pm 3.94$	$2.28 \times F + 59.76 \pm 3.41$
Mongolids	$2.15 \times F + 72.57 \pm 3.80$	
	Tibia	
Caucasoids	$2.52 \times T + 78.62 \pm 3.37$	$2.90 \times T + 61.53 \pm 3.66$
Negroids	$2.19 \times T + 86.02 \pm 3.78$	$2.45 \times T + 72.65 \pm 3.70$
Mongolids	$2.39 \times T + 81.45 \pm 3.27$	
	Humerus	
Caucasoids	$3.08 \times H + 70.45 \pm 4.05$	$3.36 \times H + 57.97 \pm 4.43$
Negroids	$3.26 \times H + 62.10 \pm 4.43$	$3.08 \times H + 64.67 \pm 4.25$
Mongolids	$2.68 \times H + 83.19 \pm 4.25$	
	Radius	
Caucasoids	$3.78 \times R + 79.01 \pm 4.32$	$4.74 \times R + 54.93 \pm 4.30$
Negroids	$3.42 \times R + 81.56 \pm 4.30$	$2.75 \times R + 94.51 \pm 5.05$
Mongolids	$3.54 \times R + 80.71 \pm 4.60$	
	Fibula	
Caucasoids	$2.68 \times Fi + 71.78 \pm 3.29$	$2.93 \times Fi + 59.61 \pm 3.57$
Negroids	$2.19 \times Fi + 85.65 \pm 4.08$	$2.49 \times Fi + 70.90 \pm 3.80$
Mongolids	$2.40 \times Fi + 80.56 \pm 3.24$	
	Ulna	
Caucasoids	$3.70 \times U + 74.05 \pm 4.32$	$4.27 \times U + 57.76 \pm 4.30$
Negroids	$3.26 \times U + 79.29 \pm 4.42$	$3.31 \times U + 75.38 \pm 4.83$
Mongolids	$3.48 \times U + 77.45 \pm 4.66$	



Figure 5 Amputation of first and second fingers of left hand



Figure 6 Signs of neurosurgery on the cranial vault of a young woman. One can notice the bone fragment originally removed probably to drain a hemorrhage and then reassembled to the rest of the cranium

Congenital disease: It is rare nowadays to see important congenital diseases on bone – examples are achondroplasia, cleft palates, and spina bifida. Trauma: aftermaths of amputations (Figures 5 and 6) or fractures (and therefore more or less recent calluses) along with prosthetic devices (Figure 7) are quite frequent, and it may be easy to determine



Figure 7 The presence of a prosthetic device in a left femur

the age of a particular fracture so that one can be more informed in determining when the incident happened. Degenerative disease/Arthrosis: Osteophytes (Figure 8) and/or any other proliferative reaction of articular surfaces, along with lytic reactions of bone, such as the presence of small depressions of the vertebra called *Schmorl's nodes*, may be indicative of discopathies or arthropathies to specific anatomical areas. Metabolic and neoplastic disease: bones and teeth can show signs of general stress (Figure 9), iron deficiency (Figure 10), scurvy, rickets, though rare. Moreover, metastatic bone lesions may sometimes be seen in individuals who had secondary cancer. Infectious disease: bone infections following bone trauma can be seen, such as osteomyelitis. More rarely, and in underdeveloped countries, syphilis, tuberculosis (Figure 11), or leprosy can leave specific signs on bones.



Figure 8 A lumbar vertebra with a marked osteophyte (a lip of bone growing upward from the vertebral body rim), indication of arthrosis



Figure 9 Enamel hypoplasia. These incisors and canines, which belong to a child (the teeth are still incomplete as the root is missing), show enamel hypoplasia, i.e., thin depression lines running across the width of the crown. This is a general stress marker, which shows that at some point a "systemic" incident (such as a severe disease) occurred that temporarily stopped dental growth

The disciplines of osteology, orthopedics and, in particular, paleopathology therefore supply a plethora of information on how to extrapolate disease, stress markers or simple anatomical anomalies, such as nonmetric traits, from human remains, which may be incredibly useful for building the biological profile and crucial for subsequent identification.



Figure 10 Cribra orbitalia. The fine porosity visible on the ceiling of the orbits is known as *cribra orbitalia*, a sign of severe anemia

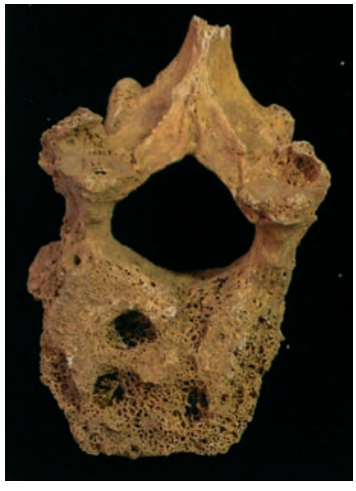


Figure 11 Thoracic vertebra bearing cavities due to tubercular abscesses, a clear sign of infection

References

[1] Ballard, M.E. & Trudell, M.B. (1999). Anterior femoral curvatures revisited: race assessment from the femur, *Journal of Forensic Sciences* **44**(4), 700–707.

[2] Bidmos, M. (2006). Adult stature reconstruction from the calcaneus of South Africans of European descent, *Journal of Clinical Forensic Medicine* **13**(5), 247–252.

[3] Birkby, W.H. (1966). An evaluation of race and sex identification from cranial measurements, *American Journal of Physical Anthropology* **24**(1), 21–27.

[4] Brace, C.L. (1995). *R*egion does not mean "race"- reality versus convention in forensic anthropology, *Journal of Forensic Sciences* **40**, 171.

[5] Buck, T.J. & Vidarsdottir, U.S. (2004). A proposed method for the identification of race in subadult skeletons: a geometric morphometric analysis of mandibular morphology, *Journal of Forensic Sciences* **49**(6), 1159–1164.

[6] Burris, B.J. & Harris, E.F. (1998). Identification of race and sex from palate dimension, *Journal of Forensic Sciences* **43**, 959.

[7] Byers, S.N., Churchill, S.E. & Curran, B. (1997). Identification of Euro-Americans, Afro-Americans, and Amerindians from palatal dimensions, *Journal of Forensic Sciences* **42**(1), 3–9.

[8] Chibba, K. & Bidmos, M.A. (2007). Using tibia fragments from South Africans of European descent to estimate maximum tibia length and stature, *Forensic Science International* **169**(2–3), 145–151.

[9] Craig, E.A. (1995). Intercondylar shelf angle: a new method to determine race from the distal femur, *Journal of Forensic Sciences* **40**(5), 777–782.

[10] De Mendonca, M.C. (2000). Estimation of height from the length of long bones in a Portuguese adult population, *American Journal of Physical Anthropology* **112**(1), 39–48.

[11] Duray, S.M., Morter, H.B. & Smith, F.J. (1999). Morphological variation in cervical spinous processes: potential applications in forensic identification of race from the skeleton, *Journal of Forensic Sciences* **44**(5), 937–944.

[12] Edgar, H.J. (2005). Prediction of race using characteristics of dental morphology, *Journal of Forensic Sciences* **50**(2), 269–273.

[13] Giles, E. & Elliot, O. (1962). Race identification from cranial measurements, *Journal of Forensic Sciences* **7**, 147.

[14] Holliday, T.W. & Falsetti, A.B. (1999). A new method for discriminating African-American from European-American skeletons using postcranial osteometrics reflective of body shape, *Journal of Forensic Sciences* **44**(5), 926–930.

[15] Iscan, M.Y. & Steyn, M. (1999). Craniometric determination of population affinity in South Africans, *International Journal of Legal Medicine* **112**, 91.

[16] Jantz, R. (2001). Cranial change in Americans: 1850–1975, *Journal of Forensic Sciences* **46**, 784.

[17] Patriquin, M.L., Steyn, M. & Loth, S.R. (2002). Metric assessment of race from the pelvis in South Africans, *Forensic Science International* **127**(1–2), 104–113.

[18] Ross, A.H., McKeown, A.H. & Konigsberg L.W. (1999). Allocation of crania to groups via the "new morphometry", *Journal of Forensic Sciences* **44**(3), 584–587.

[19] Sauer, N.J. (1992). Forensic anthropology and the concept of race: if races don't exist, why are forensic anthropologists so good at identifying them? *Social Science and Medicine* **34**(2), 107–111.

[20] Jantz, R.L., Hunt, D.R. & Meadows, L. (1995). The measure and mismeasure of the tibia: implications for stature estimation, *Journal of Forensic Sciences* **40**(5), 758–761.

[21] Jantz, R.L. (1992). Modification of the Trotter and Gleser female stature estimation formulae, *Journal of Forensic Sciences* **37**(5), 1230–1235.
[22] Jason, D.R. & Taylor, K. (1995). Estimation of stature from the length of the cervical, thoracic, and lumbar segments of the spine in American whites and blacks, *Journal of Forensic Sciences* **40**(1), 59–62.
[23] Klepinger, L.L. (2001). Stature, maturation variation and secular trends in forensic anthropology, *Journal of Forensic Sciences* **46**(4), 788–790.
[24] Mall, G., Hubig, M., Büttner, A., Kuznik, J., Penning, R. & Graw, M. (2001). Sex determination and estimation of stature from the long bones of the arm, *Forensic Science International* **117**(1–2), 23–30.
[25] Meadows, L. & Jantz, R.L. (1992). Estimation of stature from metacarpals lengths, *Journal of Forensic Sciences* **37**(1), 147–154.
[26] Munoz, J.I., Linares-Iglesias, M., Suarez-Penaranda, J.M., Mayo, M., Miguens, X., Rodriguez-Calvo, M.S. & Concheiro, L. (2001). Stature estimation from radiographically determined long bone length in a Spanish population sample, *Journal of Forensic Sciences* **46**(2), 363–366.
[27] Ozaslan, A., Koç, S., Ozaslan, I. & Tudcu, H. (2006). Estimation of stature from upper extremity, *Military Medicine* **171**(4), 288–291.
[28] Pelin, C., Duyar, I., Kayahan, E.M., Zagyapan, R., Agildere, A.M. & Erar, A. (2005). Body height estimation based on dimensions of sacral and coccygeal vertebrae, *Journal of Forensic Sciences* **50**(2), 294–297.
[29] Pelin, I. & Duyar, I. (2004). Estimating stature from tibial length: a comparison of methods, *Journal of Forensic Sciences* **48**(4), 708–712.
[30] Petrovecki, V., Mayer, D., Slaus, M., Strinovi, D. & Skavi, J. (2007). Prediction of stature based on radiographic measurements of cadaver long bones: a study of the Croatian population, *Journal of Forensic Sciences* **52**(3), 547–552.
[31] Radoinova, D., Tenekedjiev, K. & Yordanov, Y. (2002). Stature estimation from long bone lengths in Bulgarians, *Homo* **52**(3), 221–232.
[32] Rao, K.V., Gupta, G.D. & Sehgal, V.N. (1989). Determination of length of human upper limb long bones from their fragments, *Forensic Science International* **41**(3), 219–223.
[33] Steel, D.G. (1970). Estimation of stature from fragments of long limb bones, in T.D. Stewart, ed, *Personal Identification in Mass Disaster*, National Museum of Natural History, Washington, DC.
[34] Steel, D.G. & McKern, T.W. (1969). A method of assessment of maximum long bone length and living stature from fragmentary long bones, *American Journal of Physical Anthropology* **31**, 215.
[35] Trotter, M. (1991). Estimation of stature from intact limb bones, in T.D. Stewart, ed, *Personal Identification in Mass Disaster*, Smithsonian Institution, Taraxacum, Washington, DC.
[36] Trotter, M. & Gleser, G.C. (1977). Corrigenda to estimation of stature from long bones of American whites and Negroes, *American Journal of Physical Anthropology* **47**, 355.
[37] Trotter, M. & Gleser, G.C. (1952). Estimation of stature from long bones of American whites and Negroes, *American Journal of Physical Anthropology* **19**, 213.
[38] Roberts, C. & Manchester, K. (1995). *The Archaeology of Disease*, 2nd Edition, Alan Sutton Publishing Limited, Cornell University Press, Ithaca, New York.

Further Reading

Albanese, J. & Saunders, S.S. is it possible to escape racial typology in forensic identification? in A. Schmitt, E. Cunha & J. Pinheiro, eds, *Forensic Anthropology and Medicine*, Humana Press, pp. 281–316.
Gill, G.W. (1984). A forensic test case for a new method of geographical race determination, in T.A. Rathbun & J.E. Buikstra, eds, *Human Identification: Case Studies in Forensic Anthropology*, Charles C. Thomas Publisher, Springfield.
Gill, G.W. (1995). Challenge on the frontier: discerning American Indians from whites osteologically, *Journal of Forensic Sciences* **40**(5), 783–788.
Gill, G.W. (2001). Racial variation in the proximal and distal femur: heritability and forensic utility, *Journal of Forensic Sciences* **46**(4), 791–799.
Giles, E. (1993). Modifying stature estimation from the femur and tibia, *Journal of Forensic Sciences* **38**, 758.
Giles, E. & Vallandigham, P.H. (1991). Height estimation from foot and shoeprint length, *Journal of Forensic Sciences* **36**(4), 1134–1151.
Hauser, R., Smolinski, J. & Gos, T. (2005). The estimation of stature on the basis of measurements of the femur, *Forensic Science International* **147**(2–3), 185–190.
Holland, T.D. (1992). Estimation of adult stature from fragmentary tibias, *Journal of Forensic Sciences* **37**(5), 1223–1229.

Related Articles

Anthropology
Length Measurement
Species Determination of Osseous Remains
Sex Determination of Remains

CRISTINA CATTANEO

Antiandrogen Treatment *see* Sex Offenders: Treatment of

Anticonvulsants *see* Seizures: Behavioral

Antisocial Personality Disorder *see* Psychopathy

Applications of Forensic Toxicology *see* Toxicology: Forensic Applications of

Archaeology

Background

Archaeology is now becoming an increasingly familiar discipline in scenarios involving the search for and recovery of buried human remains [1], including the identification of mass graves and recovering victims [2]. The work has also been extended into nonhuman materials, for example, buried drugs, firearms, and stolen goods. Forensic archaeology is founded on the long development of "traditional" archaeology in the study, analysis and interpretation of buried soil formations, and in the use of electronic survey techniques. It requires a wide understanding of landscape mapping, natural geological phenomena, and the use of aerial photographic imagery. This thorough grounding in field archaeology is essential, as it brings with it an awareness of other disciplines that may need to be integrated as part of a recovery operation, for example, palynology (*see* **Palynology**), entomology (*see* **Entomology**), pedology (*see* **Soil: Forensic Analysis**), conservation, etc. Archaeology may also include additional specialisms in fields such as geophysical survey (e.g., ground-penetrating radar (GPR)) or physical anthropology (*see* **Anthropology**), but both these are usually considered individual fields of expertise in their own right.

Many archaeologists have completed some basic training in physical anthropology and will be familiar with which elements of the human skeleton are most useful for determining basic identification factors of age, sex, and stature, etc., and will therefore be able to take particular care in their recovery. This familiarity with hard tissue also usually extends to being able to distinguish between human and animal bone – an especially useful aspect of the discipline when it comes to assessing the significance of buried or concealed bone material discovered during construction work, gardening, house renovation, or surface remains handed in by dog walkers or other members of the public. Moreover, in forensic search scenarios where human remains may have become disarticulated, scavenged, or spread, the ability to determine this distinction *in situ* is both cost effective and time saving.

Forensic archaeology can be summarized as having input in criminal investigation in the following areas:

- Landscape analysis: the use of maps, aerial photographs, and shallow subsurface geophysical survey.
- Excavation: the recovery of buried individuals through excavation by understanding of stratigraphy and implementation of sampling, planning, and recording systems.
- Skeletal hard tissue: knowledge of skeletal components for determining individuality, and for distinguishing between human and animal materials.
- Associated environmental sciences: the value of soils, seeds, pollens, entomology, flora and fauna, etc., and their relevance in identifying habitats where incidents occurred.

The shift from "traditional" to forensic archaeology is not straightforward and requires a wide understanding of scene protocols, the evidential requirements of other specialists, and the nature of the legal constraints within which forensic work takes place. Recognition of the role of archaeology within criminal investigation has been a gradual process. There is often a popular misconception that "forensic" archaeology is predominantly concerned with historical or antiquarian matters such as the medical investigation of Egyptian mummies, the facial reconstruction of

ancient skulls, or analysis of the stomach contents of bog bodies. However, these simply involve the application of science (on occasions forensic science) to archaeological remains. "Forensic" archaeology, as with any other forensic discipline, is contemporary and pertains to matters of court; it normally involves working with police forces in the resolution of serious crime, usually murder. In fact many early police investigations searching for buried remains might best be described as archaeological scenarios, even although archaeologists were not involved. Archaeological input to crime investigation has been growing.

A number of fairly typical scenarios, illustrating the point at which archaeology may be utilized, and the archaeological methodologies involved, are outlined below. The majority involve locating clandestine burials and the recovery of victims, the elimination of areas where human remains are alleged to have been buried, or the occurrence of "stray" human material found, for example, during building operations. Within the context of mass graves, issues of genocide and human rights abuse are likely to be encountered, and the discipline has to address the types of evidence that may require recording. Mass grave work is relatively new in both the United States and the United Kingdom [1, 3].

Method

Contrary to some popular opinion, archaeology is not about finding objects. The discipline is largely concerned with locating specific buried remains and in recovering them in a manner that will include and maximize associated evidence, avoid contamination, and allow a process of reconstruction to take place. Archaeology is a destructive process and therefore unrepeatable; the key element is to ensure that an adequate record of the excavation is made. There are various strategies for ensuring that this is achieved, but there is no "by-the-book" method of approach. The method of each scenario has to be taken on its merits according to prevailing conditions and involves a minimum standard of process, record, and observation of strata. These factors are more complex in mass graves for which a number of standard operating procedures (SOPs) have been written [4].

There is often a general misunderstanding that ground subsurface is composed of an amorphous mixture of soils into which burials may be concealed, and from which they can be exhumed by coarse digging. This is far from the case: most shallow subsurface soils are composed of sequences of discrete layers (stratigraphy) representing different events in past time; these layers may be naturally formed, culturally formed, or a mixture of both. The key aspect of any archaeological intervention is an understanding of the nature and character of these soil layers. The layers may differ by virtue of color, physical property, geological character, or inclusion, and they may represent a range of different anthropogenic and natural processes (Figure 1). Archaeology is mostly about interpreting these layers, which can be created by a large range of activities, including construction, demolition, burning, cultivation, clearance, occupation, etc., or through natural formation processes. Successful interpretation of these layers enables an understanding of what occurred in a given place over time. In effect, interpreted correctly, these layers can be used to provide a statement of what went on in a particular place – it gives that location a context in which those events occurred. Moreover, because individual layers are superimposed, they also exhibit a local chronology – it is possible to show which is the earliest and which is the latest layer in that place – in effect providing a "snapshot" as well as a context for the layers in question. Any intervention into the ground, by clandestine burial or even by forensic investigation, will create a new set of layers within the existing context and add to the chronology of what was there before. These are the features with which the forensic archaeologist will need to be specifically concerned.

It makes no difference whether there are a few layers or a few hundred layers in any given location; and it makes no difference whether the layers are one thousand years old or a few days old – the principles of context and relative chronology remain exactly the same. In fact, even with recent criminal burials there are often layers that can be identified with specific events, such as the laying of terraces or patios, upcast from ditch clearance, minor building activity involving spreads of sand or other materials, even cultivation and manuring, or natural phenomena such as leaf fall. These events will be quite closely dateable, particular those aspects of construction work that require Local Authority permission or building consent. Dated layers that lie above and seal the grave are denoted as giving a *terminus ante quem* (time before which) for the date of the grave;

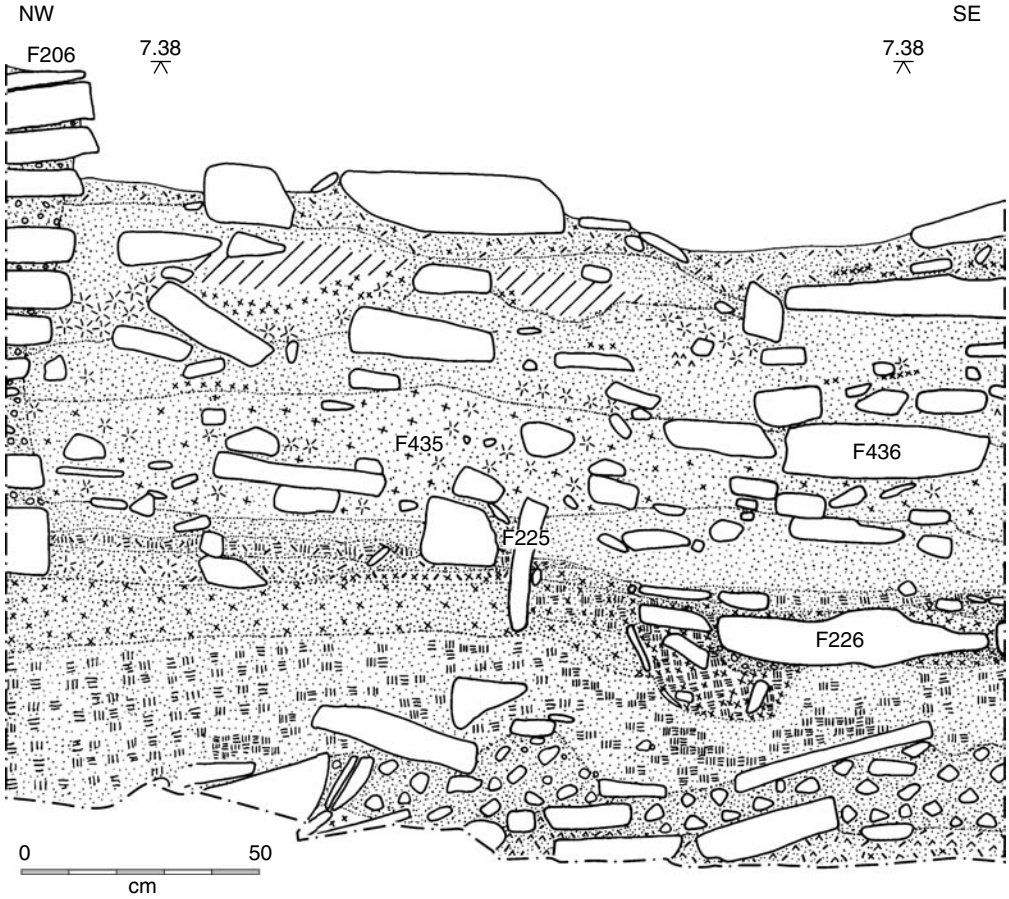


Figure 1 Illustration of different superimposed layers on an archaeological site

dated layers that are cut into by the grave can give a *terminus post quem* (time after which) for the date of the grave.

Stratigraphy

As the ground subsurface consists of a series of layers in varying degrees of complexity, any victim buried there is therefore contained within a fixed stratigraphic environment. The edges of any grave (i.e., the sides and base of the cutting made to create the grave, and the top after the grave has been filled in) effectively represent a sealed deposit in which the individual and any evidence pertaining to that individual is retained (Figure 2). It is best treated as a crime scene as it may contain evidence for the identity of the individual, the manner and cause of death and, depending on the layers above and below, the possible chronology of burial. Moreover (and less well recognized) is the fact that this deposit will also contain any evidence that may link the victim to an offender by virtue of traces of material (fibers, shoe marks, cigarette ends, etc.) or items added to the grave fill by the offender for concealment. It is of paramount importance that the integrity of that sealed deposit is recognized throughout the recovery operation; otherwise, not only will the validity of the evidence recovered (including material that may offer links with the offender) be suspect, but there will also be significant potential for contamination.

Many human remains are discovered during building work or land clearance, and there is an inevitable tendency for any bones to be removed

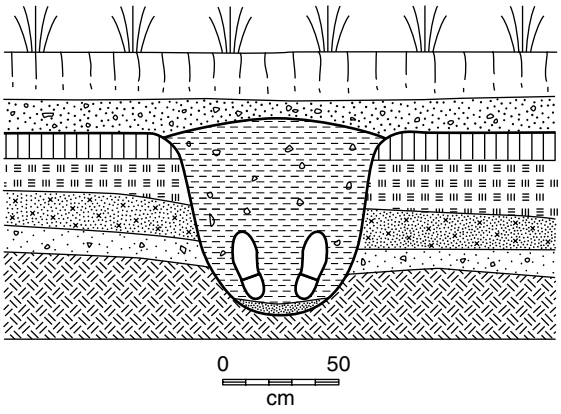


Figure 2 Reconstruction of the layers likely to surround a simple grave

for examination, usually to check if they are animal or human. By removing them from their layer, the integrity of the grave is effectively broken, their context is lost, and the chronology distorted. As a result, the material is largely unusable. However, the integrity of the grave can be maintained if the material is left exactly where it was found. The outline of the grave can be established by careful excavation and the contents removed appropriately. Case studies have shown that retaining the integrity of a burial can be important in a variety of practical situations: for example, if articulated remains are discovered during construction or drainage operations. By leaving them *in situ* an archaeological assessment of their stratigraphic context can be carried out and decisions can be made regarding whether the material is modern (and therefore of potential forensic interest) or ancient (and therefore of only archaeological or historic interest) before recovery commences.

As another example, there are sometimes incidents where human remains have been found partly buried; here it will be necessary to determine whether the remains were originally buried (i.e., by a third party) and may since have been exposed or scavenged, or have become buried (i.e., by a process of natural soil accumulation, hillwash, etc.). In the former, the matter may require criminal investigation, in the latter it may not. Any distinction between the two can be easily destroyed by clearing or exposing the remains by crime scene personnel, or by removal of the material, even in part, by the forensic pathologist. Retaining the integrity of the situation is critical to any understanding of what happened; hence the primary investigation and recovery is best carried out archaeologically.

Search

The initial search for missing persons assumes that the individuals concerned are alive in the first instance and may be missing because of medical or psychological conditions as much as to disorientation, abduction, or murder. The shift of search from a living to a deceased target is a gradual one and also necessitates a shift in the techniques deployed. In the United Kingdom, these techniques are listed and detailed in a comprehensive police advice manual [5]. This article is concerned only with searching for clandestine burials, which can involve the deployment of a range of specific techniques and an equally wide range of expert personnel (see also [6]). Most of these searches are highly planned and coordinated. Searches initially commence with noninvasive methods in order to target specific areas, to eliminate others, and to define search boundaries. The information (intelligence) that underpins this process may include a range of factors: offender and geographical profiling, witness accounts, last sightings, database statistics derived from similar scenarios (including disposal patterning), and the suspect's movements. General and specialized maps can be used to define access and feasibility of burial, aerial photographs may be commissioned, and appropriate areas can be defined. Aerial photography may be able to identify areas of disturbance from soil change, shadow, or crop mark, although on a scale much smaller than most archaeological features observed from the air, and in concealed rather than open locations.

All searches require some starting point or location of interest. Search is also about eliminating areas from the search, either because they are not considered feasible (e.g., open rocky ground) or because they have been searched without result. A search team should eliminate an area only if it can do so with a high degree of confidence. This level of confidence is achieved by using a sequence of search techniques, including complementary techniques. No one ever wants to find out that a burial was subsequently discovered in an area that had already been eliminated.

Once these "desktop" techniques have been concluded, target areas will be identified for more

detailed attention. These areas may consist of straightforward tracts of land, individual gardens, or specific, small points of interest identified from the air. Graves dug into the ground will invariably leave a trace, even years later. Ground disturbance is likely to have an effect on the surface vegetation, perhaps supporting stronger growth or different species or presenting different flowering times compared to the undisturbed surrounding ground. These are longer-term effects. In the shorter term, trampled vegetation and soil may be apparent, as will the local disposal of excess soil that is unable to fit back into the grave. There may also be a topographical effect in that eventual consolidation of the grave fill will cause a depression in the ground surface; this may later become exacerbated when the buried body collapses through the decay process. Both vegetational and topographical change may require skilled observation. Members of the public, so often photographed "helping" in many early enquiries, may unwittingly eliminate vulnerable evidence of this type.

The same effect of disturbance that can create vegetational and topographic change may also lead to the creation of a geophysical signature that differs from the surrounding, undisturbed ground. In effect, the search techniques used so far have been focused on identifying the disturbance caused by the grave, rather than the body lying within the grave. Archaeological geophysics (as opposed to geological or engineering geophysics) is particularly suited to work of this kind, as it shares the same requirements and sensitivity for identifying small, shallow subsurface targets [7, 8]. It does, however, require there to be a reasonable "background" of nondisturbed ground or of consistent geology in order for the grave to be picked up as being anomalous within that background. Even in wide open areas, geophysical techniques are pushed to their limits to detect grave-size targets, and this may become virtually impossible in many typical search environments, notably rear gardens that display various types of surface and are overgrown and cluttered with material both surface and buried that may interfere with the method.

Of the three main methods used in archaeology, resistivity and GPR are probably the most effective. Magnetometry, which is highly sensitive to local magnetic and ferrous effects, has been found to be less useful if only in view of the background "noise" present in many of the locations targeted. Resistivity is particularly useful in open landscapes such as lawns or areas of turned soil, but is likely to require data processing in order to enhance any anomalies detected. GPR, on the other hand, is better in confined spaces and has the ability to detect through dense surfaces. Thus, in contexts where patios, driveways, or swimming pools become targets of search, GPR is a helpful technique in being able to minimize heavy reinstatement costs. It also has the advantage of showing anomalies in "real time" and allows targeting of specific features as the survey progresses. However, there are now several software packages that allow all geophysical techniques to be processed and filtered as well as being converted into 3-D images, which allow the operator to maximize the data available.

It is important to stress that all geophysical survey methods have advantages and limitations. Electrical, magnetic, and electromagnetic techniques all respond to, or are affected by, different phenomena and the various systems are best used to complement each other rather than being used as individual options. For example, a grave in a particular soil environment might not be detected using resistivity, but may be picked up using magnetometry. Experimentation has emphasized the difficulty in detecting graves using geophysics, particularly soon after a burial has taken place. Equally, it can take a skilled and experienced operator to recognize the display effects of a burial evident in a radar scan depending on time since disposal. A key element here is the decay dynamics of the buried victim and its effect on detection methods. Features in traditional archaeology (such as pits, walls, ditches, etc.) tend to be inert, whereas a recently buried victim will undergo a process of decay, which ultimately (and under normal circumstances) will culminate in skeletonization. The speed and extent of this decay process and resulting effects (taphonomics) depend on numerous factors such as depth of burial, wrapping, wetness, temperature, soil environment, oxygen, bacteria, etc., and these in their turn may alter the efficacy of the various detection techniques available. In traditional archaeology, geophysical survey using resistance or magnetometry methods normally operates within 20 m × 20 m grids, allowing up to 10 grids to be covered in a working day taking measurements at 1 m intervals across the grid; in forensic work, readings are usually taken at 0.5 m intervals in order to ensure that any grave is

adequately sampled, and this will take commensurately longer.

The detection of likely targets either by geophysics, from the air, by fieldcraft, evidence of scavenging, or by observation of disturbed ground can then be tested by "ground truthing". This can entail trial trenching or, less invasively, the use of a cadaver dog. Cadaver dogs are trained to detect the scent of decaying human flesh but ideally require the ground to be "vented" first. This involves the handler piercing the ground in the suspect area often to a depth of around 0.6 m with a narrow metal probe in order to allow any scent to rise to the surface. Often the ground is vented with a series of holes, or in a line, which enables the handler to lead the dog into, and across, the suspect area downwind. If the dog responds in a particular place, further action can be taken. One option available is to become more invasive in order to be confident that no remains have been buried. For example, in a garden where a number of targets have been found by geophysics but have not been responded to by the dog, a mechanical excavator might be employed to strip away the topsoil. Used skillfully, machines have a valuable role to play in work of this kind. The use of a wide (toothless) ditching bucket can clear a garden down to undisturbed substrates in a matter of a few hours, allowing any disturbance to be investigated, usually by half-section. Given their destructive nature, machines are best used as a secondary rather than as a primary approach. Used primarily, they may destroy much evidence if any burial is present; used secondarily, they can achieve elimination of an area.

Probing can also be used, although this can have damaging effects on a buried victim. Used systematically, it may be possible to identify softer (disturbed) substrates from harder (undisturbed) substrates. This method can also be extended to the use of a corer, which will allow a column of earth to be removed in order to assess the nature of buried deposits. Both methods have destructive implications.

Any features discovered, by dog or other methods, can be excavated by hand, usually by half-section. This allows the excavator to identify the size, depth, and nature of the feature as rapidly as possible (there may be many features requiring investigation), as well as showing in section how the feature was infilled (Figure 3). If the feature turns out to be a grave, the nature of the infilling may be significant. During the half-section, spoil will be recovered either by contexts or (more likely) in arbitrary spits and retained in case the feature is found to contain human remains. In such an eventuality, the "search" process will have already interfaced with the recovery exercise: the likely extent and depth of the grave will be known, the nature of the infill identified, and the taphonomics established. Taphonomy is a relatively new area of study that can affect both search and recovery, and has been the subject of extensive focus [9, 10]. These factors will provide valuable information and guidance for the remainder of the recovery process. More significantly, both loss of evidence and the opportunity for contamination to have occurred will have been minimized.

Search exercises should be logged and produced in report form. This log should include times, dates, areas, methods, decisions made, and locations eliminated. In the event that no victim has been recovered, it provides an invaluable record for future "cold case" reviews in which new technologies and methods might be applied. Many older cases that now come under review with no such log, or even accurate record of where any "digging" took place, may need to be started from scratch.

Recovery

Excavation is normally geared in the first instance to identifying any grave outline, determining stratigraphic relationships between the grave fill and the surrounding ground, and setting up a recording regime. Fixed points need to be identified, and plans produced. In most, but not all, cases it may be preferable to half-section the grave in order to identify any particular pattern of infill in the exposed section. The section might, for example, show that a grave had been carefully backfilled with specific materials in order to minimize discovery. This might be used to demonstrate in court that the perpetrator was in a sound state of mind in disposing of the body, with resultant implications for motive and outcome. Most grave fills, however, are simple mixes of backfilled substrates and require excavating in numbered "spits" with spoil usually retained in sealed containers such as large plastic dustbins for later sieving. Graves often contain items deliberately concealed by the perpetrator, for example, soiled clothes or a weapon, as well as items conveniently discarded such as cigarette butts (useful for DNA) or newspaper (useful for dating).

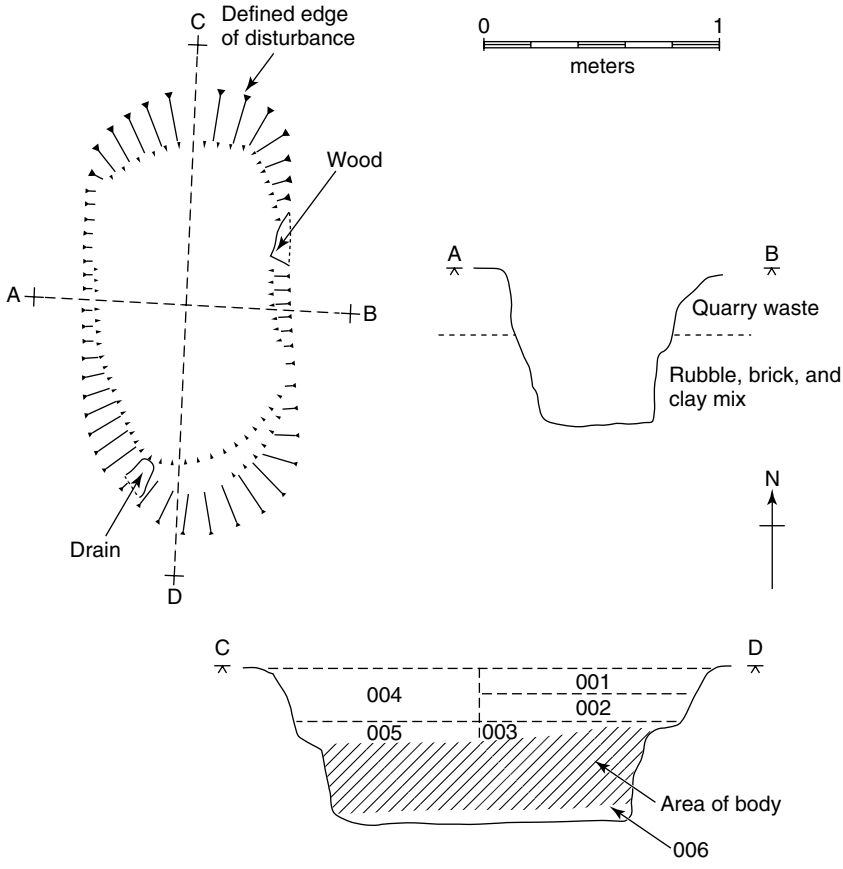


Figure 3 The half-sections and plan of a burial showing the layer numbering used

Forensic recovery is no different from any other type of archaeology in being destructive and therefore requiring the practitioner to excavate according to a research design in order to answer specific questions. Unlike traditional archaeology, however, these types of question tend to pertain to an individual event. The interrogation might consider how the grave was dug, what tools were used, whether it was excavated in a hurry or had been carefully prepared or even left open, whether it contained an unusual fill or any foreign material, or whether it contained material transferred from the offender into the fill. This is not conclusive evidence, but it can be used to develop further enquiries.

This type of evidence needs to be looked for and recovered as a matter of priority even if it entails excavating in unusual or unorthodox ways. Equally, the victim needs to be uncovered and removed in a manner that will retain any forensic evidence on the body itself, avoid contamination from nongrave contexts, and prevent any additional trauma to the body itself. This often has to be undertaken in confined, screened, or tented areas, sometimes in cellars or narrow gardens, and can present practical difficulties that have to be surmounted *ad hoc* in order to retain the integrity of the grave (above). Space permitting, access to the body may need to be facilitated by digging out a platform from the side of the grave in order that work on the body can take place at the same level. No matter how the job is done, at the end of the day the method undertaken needs to be able to recover essential data according to the highest possible standards in order for it to be presented in court and defended under cross-examination. Often this will need ingenuity and inventiveness, especially if a suspect is being held in custody with specific

time constraints. Irrespective of physical constraints, it is essential to produce basic archaeological records in both written and graphical terms.

Throughout this process, the archaeological work will need to be coordinated with the collection of other forms of evidence, for example, any necessary sampling for entomological or environmental purposes [11, 12]. The use of a stratigraphic methodology (real or in "spits") may allow such sampling to be coordinated in a manner that will allow all samples to be gathered within a common spatial framework, allowing for greater integrity of results.

Mass Graves

Opinions vary as to what constitutes a mass grave, but essentially it is one that contains numerous individuals, sometimes hundreds, and is usually the result of genocide or civil war, although they can result from the hurried consequence of natural disasters. Their excavation tends to fall into two types: the recovery of evidence in order to obtain convictions for war crimes, or the need to recover individuals in order to obtain personal identification and facilitate repatriation to families. Mass graves resulting from genocide or civil war are discussed here. They have occurred in several parts of the world, but most notably in Africa, the Balkans, and the Middle East. They can pose problems other than simply of scale, often through political interference, cultural differences, lack of appropriate facilities and equipment, and unfamiliarity with archaeological processes by those nominally responsible for excavation. Archaeologists and anthropologists may be part of a team but may not necessarily be in a position to exercise control over the way bodies are recovered and treated. They may need to devise minimum standards with which they are comfortable and which can be presented and defended in court.

Mass graves tend to be identified from the air, or located on the basis of witness accounts. Concealment is not necessarily a high priority in regions where the government itself has carried out atrocities, although some are very deliberately shielded from international gaze, for example, by redeposition of rubble. Witness accounts are probably the most important methods since the same witness(es) may also provide evidence of the people responsible. Trial trenching, usually by machine, can be used to refine the specific location. Geophysics remains a useful technique, but requires a large background area against which the anomaly caused by the grave is visible. Given that many areas around mass graves are known to be mined, this is not always feasible. However, once the grave has been found, geophysics may be able to provide information regarding density or nature of deposition within the grave itself.

Mass graves tend to be dug and infilled by machine. As such, they tend to be of characteristic width (typically around 3 m, which is the width of a large machine bucket) and have a ramp at one end. Tire tracks may still be evident on the surface and, if the grave was also an execution site, there may be cartridge cases and other material to be recorded at the grave edge. Many such graves may be over 2 m deep and will necessitate consideration of collapse and flooding during excavation. Other hazards may include live ordnance, or devices planted in the grave by perpetrators, and possibly local hostility. Only in some cases are the victims laid out individually or bagged. In some instances of genocide the bodies are carried to the grave in wagons and tipped in by machine causing a series of discrete depositions of contorted and intermingled remains. Some graves are secondary, i.e., the bodies have been removed from a primary grave, and therefore may contain evidence of the primary grave itself, or even of an original execution site. The process of moving human remains by machine inevitably causes some body parts to become detached; it may also entail differential states of decay between individual machine depositions within the grave. The net effect is that the archaeologist can be required to separate out tightly compressed individuals from each other, while at the same time identify soil change, record artifacts and belongings, identify evidence for abuse, and maintain a comprehensive system of recording.

Given the varied states of decay and disassociation of remains, recording a mass grave can require a high degree of anthropological knowledge. It will almost certainly require a system that can record in three dimensions and log data simultaneously. One software system, for example, allows for the taking of points on the main joints of each individual, thereby creating "stick-people" who can be viewed and rotated as appropriate on the computer screen. This method can sometimes assist the relocation of body parts to individuals. The nature of a mass grave is such that the individual victims are the key

"artifacts" and therefore bespoke recording proformas are likely to be required. Individuals are numbered uniquely, and this number should remain with each individual from excavation to mortuary, via autopsy and any other sampling (e.g., for DNA), through to eventual reburial. The same number will be used for clothing and any other associated material (e.g., jewellery) or objects that may support presumptive identification. Recording will also need to be made *in situ* of blindfolds, cable ties, and other evidence of abuse in case these become detached in the recovery process. Record will also need to be made of other materials within the grave, for example, artifacts such as projectiles or cartridge cases, botanical remains that may have been transferred with the victims, and soils or geological traces that may also provide links between graves and the point of execution. Tire marks and machine teeth marks within the grave itself may provide evidence of type of vehicle, or even the specific vehicle used.

Acknowledgment

Substantial parts of this article were published by the author in "Forensic Archaeology", *Encyclopedia of Archaeology*, 1396–1403, copyright Elsevier 2007.

References

[1] Hunter, J. & Cox, M. (2005). *Advances in Forensic Archaeology*, Routledge, London.
[2] Hunter, J. & Simpson, B. (2007). Preparing the ground: archaeology in a war zone, in *Forensic Archaeology and Human Rights Violations*, R. Ferllini, ed, Charles C. Thomas, Springfield, pp. 266–292.
[3] Connor, M. & Scott, D.D. (eds) (2001). *Archaeologists as forensic investigators: Defining the Role.* Dedicated Issue of Historical Archaeology, California University of Pennsylvania, Pennsylvania, Vol. 35, p. 1.
[4] Cox, M., Flavel, A., Hanson, I., Laver, J. & Wessling, R. (2007). *The Scientific Investigation of Mass Graves: Towards Protocols and Standard Operating Procedures*, Cambridge University Press, Cambridge.
[5] ACPO (Association of Chief Police Officers for England and Wales) (2006). *Practice Advice on Search Management and Procedures*, CENTREX, Wyboston.
[6] Killam, E.W. (1990). *The Detection of Human Remains*, Charles C. Thomas, Springfield.
[7] Pye, K. & Croft, D.J. (eds) (2004). *Forensic Geoscience*, *Special Publication 232*, Geological Society, London.
[8] Cheetham, P. (2005). Forensic geophysics, in *Advances in Forensic Archaeology*, J.R. Hunter & M. Cox, eds, Routledge, London, Chapter 3.
[9] Haglund, W.D. & Sorg, M.H. (eds) (1996). *Forensic Taphonomy*, CRC Press, Boca Raton.
[10] Haglund, W.D. & Sorg, M.H. (eds) (2002). *Advances in Forensic Taphonomy*, CRC Press, Boca Raton.
[11] Erzinclioglu, Z. (2000). *Maggots, Murder and Men*, Harley Books, Colchester.
[12] Brown, A.G., Smith, A. & Elmhirst, O. (2002). The combined use of pollen and soil analyses in a search and subsequent murder investigation, *Journal of Forensic Sciences* **47**(3), 614–618.

JOHN HUNTER

Arson *see* Firesetting

Arson Investigation: Misconceptions and Mythology

Introduction

This article explores the development and promulgation of the "mythology" of arson investigation. There is no reason to believe that anyone ever set out to promulgate something that was not true. It is likely that many myths came about as a result of unwarranted generalizations.

Some myths arose because of intuitively "obvious deductions". The notion that gasoline burns hotter than wood is an appealing one, as is the notion that a narrow V-pattern indicates a "rapid fire". The problem is that the term "rapid" is never defined, thus making it impossible, in many cases, to actually design an experiment to test a specific hypothesis about the significance of a particular fire artifact. Even when an artifact can be shown to be of no value by direct evidence, resistance to change and a culture of "circular citations" can allow the myth to persist.

Many of the myths were gathered by the Law Enforcement Assistance Administration (LEAA) and published in *Arson and Arson Investigation: Survey*

and Assessment [1], and although they were reported with appropriate cautionary language, on many occasions the cautions were not heeded. And when the "indicators" were listed by what should have been the ultimate authority, the cautions were lost. No less an authority than the National Bureau of Standards (NBS then – now National Institute of Standards and Technology NIST) published the *Fire Investigation Handbook* [2], which stated that crazed glass meant rapid heating, shiny alligator blisters meant that a fire burned "faster than normal", and narrow V's indicate "fast-developing, hot fires".

In the 1980s, one American text after another referred to the NBS publication or to another publication that cited the myths published in the LEAA report. These circular citations continued in books still in print. Interestingly, many of the myths never gained much credibility in the United Kingdom because the major "go to" textbook, Cooke and Ide's *Principles of Fire Investigation* [3], either did not repeat the myths or provided an accurate interpretation of the significance of indicators such as crazing and spalling.

In 1985, when the National Fire Protection Association (NFPA) Standards Council became increasingly concerned about the validity of fire investigations, it appointed a Technical Committee to address the issue. Seven years later, the Committee and NFPA produced the first edition of NFPA 921, *Guide for Fire and Explosion Investigations.* There was, perhaps predictably, resistance to this from some fire investigators. If what was printed in that document were actually true, it meant that hundreds or thousands of accidental fires had been wrongly determined to be incendiary fires. No investigator wants to admit to the possibility that they had caused an innocent person to be wrongly convicted, or a family to be wrongly denied their life savings.

In 1998, the Technical Committee on Fire Investigations, in response to public pressure, removed the word "misconception" from the titles of several paragraphs in the chapter on pattern development in the optimistic but mistaken belief that previous editions of the document, which were still not accepted in many quarters, had relieved the profession of these misconceptions.

The myths are slowly dying out (or being "Dauberted" out), but there are still practitioners who use them today. Examples of the continued promulgation and application of the mythology since 2000 are presented, as is the debunking of the myths as set forward in NFPA 921.

The Development and Promulgation of Myths

It is a failure to address a serious problem concerning the training and education of fire investigators that causes the myths to persist. The purpose of this article is to discuss those myths and, to the extent possible, attempt to understand why they came into being, and why some of them still persist.

No single reason exists to explain why a myth develops. No reason exists to believe that any investigator deliberately set out to promulgate something that was not true. It is likely that most myths originated as a result of unwarranted generalizations. For example, an investigator might observe that in a garage fire, a pattern of spalling surrounds the remains of a gasoline container and makes an association of gasoline with spalling. The next time he sees spalled concrete, he infers that gasoline must have been involved.

Paul Kirk, one of the leading forensic scientists of his time, wrote in the first edition of *Fire Investigation* [4] on examining melted metals.

> Whenever any residues of molten metal are present at the fire scene, they will reliably establish a minimum temperature for the point of their fusion in the fire. The investigator may use this fact to advantage in many instances, because of the differences in effective temperature between simple wood fires and those in which extraneous fuel, such as accelerant is present [4].

To this day, investigators sometimes infer the presence of accelerants when they observe a melted aluminum threshold.

The notion that crazed glass indicates that the glass was rapidly heated was appealing enough that Brannigan, Bright, and Jason, three respected fire researchers at the National Bureau of Standards (now NIST), published it in the Fire Investigation Handbook [2] Some authors have declared that crazed glass is sufficiently useful that the size of the crazing cracks can indicate proximity to the area of origin [5].

It is the unchallenged publication and continued promulgation of myths that ensures their longevity. If an "arson school" decides to use a text containing

the mythology in its training courses, hundreds of investigators can be exposed to this false "gospel". Those who take few refresher courses, fail to keep up with the literature, and those who attend few meetings may never be exposed to updated ideas and new research.

Much of the mythology about fire investigation was collected by the Aerospace Corporation, under a contract to the LEAA in a 1977 booklet entitled *Arson and Arson Investigation: Survey and Assessment.* To their credit, the authors of this survey pointed out, "Although burn indicators are widely used to establish the causes of fires, they have received little or no scientific testing." They recommended, "a program of carefully planned scientific experiments be conducted to establish the reliability of currently used burn indicators. Of particular importance is the discovery of any circumstances which cause them to give false indications (of, say, a fire accelerant)." In a remarkably prescient statement, they added, "A primary objective of this testing would be to avert the formidable repercussions of a court ruling on the inadmissibility of burn indicators on the grounds that their scientific validity had not been established." Despite this prediction, serious challenges to the myths did not become common until NFPA 921 was published. Part of the reason for the acceptance of the mythology may be that no less an authority than the National Bureau of Standards gave its blessing to many of the myths. In Section 1.1 of the *Fire Investigation Handbook*, which two National Fire Academy staffers are credited with contributing, most of the myths from the LEAA report were reprinted without a single caution of the type found in the original survey report. Having the imprimatur of such an august body as NBS, fire investigators and textbook authors believed (incorrectly as it turns out, but who knew?) that the myths had been scientifically tested.

The LEAA study provides a good basis for the study of the myths of fire investigation. Here is the list from the survey.

Alligatoring effect:

Checking of charred wood, giving it the appearance of alligator skin. Large rolling blisters indicate rapid intense heat, while small flat alligatoring indicates long, low heat. (This myth was repeated in the *NBS Handbook.*)

Crazing of glass:

Formation of irregular cracks in glass due to rapid intense heat – possible fire accelerant. (This myth was repeated in the *NBS Handbook.*)

Depth of char:

Depth of burning of wood – used to determine length of burn and thereby locate the point of origin of the fire.

Line of demarcation:

Boundary between charred and uncharred material. On floors or rugs, a puddle-shaped line of demarcation is believed to indicate a liquid fire accelerant. In the cross section of wood, a sharp distinct line of demarcation indicates a rapid, intense fire. (This myth was repeated in the *NBS Handbook.*)

Sagged furniture springs:

Because of the heat required for furniture springs to collapse from their own weight (1150 °F) and because of the insulating effect of the upholstery, sagged springs are believed to be possible only in either a fire originating inside the cushions (as from a cigarette rolling between the cushions) or an external fire intensified by a fire accelerant.

Spalling:

Breaking off of pieces of the surface of concrete, cement or brick due to intense heat. Brown stains around the spall indicate the use of a fire accelerant [1].

In addition to the misconceptions listed in the LEAA report, the following myths have also been widely promulgated:

Fire load:

Knowing the energy content (as opposed to the energy release rate) of the fuels in a structure was believed to allow an investigator to calculate the damage that a "normal" fire should produce in a given time frame.

Low burning and holes in the floor:

Because heat rises, it was widely believed that burning on the floor, particularly under furniture, indicated its origin on the floor.

V-pattern angle:

The angle of a V-pattern was supposed to indicate the speed of the fire. (This myth was printed in the *NBS Handbook.*)

Time and temperature:

By estimating the speed of a fire, or establishing the temperature achieved by a fire, it was believed that an investigator could determine whether it was accelerated.

The Overarching Myth

Before the discussion of the individual indicators used to determine whether a fire behaved "normally", we first consider the myth that makes all the rest of the myths plausible. This is the notion that accelerated fires burn at a higher temperature than unaccelerated fires. It is true that accelerated fires burn with a higher heat release rate (*see* **Fire: Dynamics and Pattern Production**) but there is no measurable difference in temperature.

A corollary myth about accelerated fires is really about accidental fires. Accidental fires generally start small and slow. The myth is that even after a slow smoldering fire makes the transition to flaming combustion, it remains small and slow. This is simply false, as anyone who has watched a small cigarette fire on a piece of upholstered furniture can attest. After making the transition to flaming combustion, the sofa or chair can release sufficient energy to bring a room to flashover in less than 5 min [6].

Many of the myths about fire investigation were addressed in the first two editions (1992 and 1995) of NFPA 921. In the chapter on fire patterns, there were several paragraphs entitled "Misconceptions about … (char, spalling, V-patterns, inverted cone patterns)". While the Technical Committee felt it important to shine a spotlight on these myths, many in the fire investigation community railed against the notion that any investigator had ever harbored any misconceptions about anything. They insisted, and the committee acquiesced to a change in the 1998 edition, that section titles be changed to "Interpretation of …", as if removing the word "misconception" would remove the misconception.

Alligatoring

The *Fire Investigation Handbook* contains some useful information, but it starts out with a myth-filled chapter on how to determine origin and cause [2]. Chapter 1 of the *Handbook* states:

> In determining whether the fire was a slowly developing one or a rapidly developing one, the following indicators may be used: a) Alligatoring of wood – slow fires produce relatively flat alligatoring. Fast fires produce hump-backed shiny alligatoring.

The 1982 [7] International Fire Service Training Association (IFSTA) smanual unequivocally states:

> If alligatoring is large, deep, and shiny, the fire spread extremely rapidly. Large alligatoring should be considered an indication of the nearby presence of a flammable or combustible liquid.

Nowhere it is stated what is the difference between a "fast" fire and a "normal" fire. The lack of a definition of these subjective words not only renders the "indicators" of a fire's progress meaningless but also makes it nearly impossible to design an experiment that tests the indicator's usefulness. The U.S. Army's Field Manual 19–20, *Law Enforcement Investigations* [7] provides a slightly different interpretation of alligatoring when it states:

> When wood burns, it chars a pattern of cracks which looks like the scales on an alligator's back. The scales will be the smallest and the cracks the deepest where the fire has been burning the longest or the hottest. Most wood in structures char at the rate of 1 inch in depth per 40 to 45 minutes of burning at 1400° to 1600° Fahrenheit – the temperature of most house fires. *(Thus combining three misconceptions in a single paragraph!)*

O'Connor's *Practical Fire and Arson Investigation* [9] stated:

> Deep alligatoring (large rolling blisters) on an exposed wooden surface ordinarily indicates an intense, rapidly moving body of flame. This condition may be associated with the use of an accelerant.

The second edition of the book (1997) is far more cautious, the authors having apparently become aware if the new scientific understanding of this subject [10]. The newer text states:

> It has been suggested that the presence of large shiny blisters (alligator char) and the surface appearance

> of char, such as dullness, shininess or colors have some relation to the presence of liquid accelerant as the cause, but no scientific evidence substantiates this. The investigator is advised to be very cautious in using wood char appearance as an indicator of incendiarism.

They have not completely given up, however. The 1997 text shows a photo of "a heavy rolling char . . . caused by the rapid intense movement (extension) of heat and flame."

Randall Noon, in his 1995 *Engineering Analysis of Fires and Explosions*, wrote:

> In the same way that a hunting guide interprets signs and markers to follow a trail of game, a fire investigator looks for signs and markers which may lead to a point of origin. For example, a fast, very hot burn will produce shiny type wood charring with large alligatoring. A cooler, slower fire will produce alligatoring with smaller spacing and a duller appearing char.

Noon then goes on to explain "scientifically" why this should be so [11].

> As heat impinges on the piece of wood, the water in the surface material will evaporate and escape from the wood. The rapid loss of the water at the surface is also accompanied by a rapid loss of volume, the volume which the water formerly occupied. The wood surface then is in tension as the loss of water causes the wood to shrink. This is the reason why wood checks or cracks when exposed to high heat or simply dries out over time. Of course, if the heat is very intense, more of the water "cooks" out, and the cracking or alligatoring is more severe.

The scientific-sounding explanation, although groundless, is the kind of exposition in many books that repeat the myths and has enhanced their credibility and thus their longevity.

The final word on this and most other myths may be found in NFPA 921. Here is what it says about alligatoring:

> **6.5.5 Interpretation of Char.** The appearance of the char and cracks has been given meaning by the fire investigation community beyond what has been substantiated by controlled experimentation. It has been widely stated that the presence of large shiny blisters (alligator char) is proof that a liquid accelerant was present during the fire. This is a misconception. These types of blisters can be found in many different types of fires.
>
> There is no justification that the appearance of large, curved blisters is an exclusive indicator of an accelerated fire. Figure 6.5.5, showing boards exposed to the same fire, illustrates the variability of char blister.
>
> **6.5.5.1** It is sometimes claimed that the surface appearance of the char, such as dullness, shininess, or colors, has some relation to the use of a hydrocarbon accelerant or the rate of fire growth. There is no scientific evidence of such a correlation, and the investigator is advised not to claim indications of accelerant or fire growth rate on the basis of the appearance of the char alone.

The referenced figure is a photograph taken by Monty McGill, which was first shown in *Kirk's Fire Investigation, Second Edition*. It is the definitive evidence that debunks the myth of the shiny alligator. (In the United Kingdom, the term *"crocodiling"* is more widely used.) Figure 1 shows a fire damaged wood wall that exhibits three different kinds of alligatoring, all the result of exposure to the same fire.

Crazed Glass

The *NBS Handbook* stated, "Window glass fragments in large pieces with heavy smoke deposits usually indicate slowly developing fires. Crazed or irregular pieces with light smoke deposits indicate a rapid buildup of heat" [2]. Both statements are false, but crazing is our focus for now.

The Army's Field Manual, *Law Enforcement Investigations*, states, "As a general rule, glass that contains many cracks indicates a rapid heat buildup. Glass that is heavily stained indicates a slow, smoky fire" [8].

IFSTA's *Fire Cause Determination* [12] stated:

> A window with small crazing (minute cracking), and perhaps with light smoke accumulation, is probably near the point of origin, its condition suggesting intense and rapid heat buildup. Large crazing and a heavy smoke accumulation suggest slow heat buildup and remoteness from the point of origin.

The IFSTA manual may have been the source used in *Practical Fire and Arson Investigation*, [13, 14], which repeats the notion that crazing implies a "rapid and intense" heat buildup and that if the crazing is "small", it is close to the area of origin. A larger crazing pattern, on the other hand, "implies that it may have been located in an area some distance away



Figure 1 Burned wooden wall exhibiting char blisters (alligatoring) of different sizes and different levels of gloss, all as a result of exposure to the same fire

from the point of origin." The misconception about crazing follows an extensive discussion of the types of glass that an investigator may encounter, complete with softening points, chemical compositions and applications. The reader thus is led to believe that the writers know all about glass.

In a series of laboratory experiments conducted by the author, it was conclusively demonstrated that crazing is never caused by rapid heating and can only be caused by rapid cooling [15]. The glass shown in Figure 2 was heated slowly to about 500 °C (932 °F), and then treated with water. It was possible to actually "write" in the glass using a wet cotton swab.

It is interesting to note that crazing of glass as an indicator of rapid heating is a myth that never gained much credibility in the United Kingdom. This is almost certainly because in the United Kingdom, the most widely read fire investigation text, *Principles of Fire Investigation,* correctly identified "the appearance of many small conchoidal fractures on one surface of the glass", as being the result of rapid cooling from extinguishment water [3]. The authors of that text did not use the term "*crazing*". The absence of the crazing myth in the United Kingdom lends credence to the proposition that it is publication in apparently respectable texts that is



Figure 2 Author's initials written in crazed glass by applying water on a cotton swab to hot glass. The area of crazing at the lower left was caused by the application of water from a spray bottle

responsible for the perpetuation of the mythology of arson investigation.

Depth and Location of Char

In 1979, Aetna Life & Casualty [16] published a brochure-style handbook, authored by John Barracato, who was hired by Aetna after a long career as a fire investigator for the City of New York. This pamphlet espoused as many myths as any publication ever printed on the subject. On the subject of depth of char, the booklet entitled *fire . . . is it arson?* states:

> The speed at which a fire burns is an important indicator of its cause. A fire not involving accelerant (such as gasoline or other flammable liquid) burns at the rate of 3/4 inch per hour into pine wood. The investigator should ask the fire department how long and intensely the fire burned, then carefully inspect any charred wood to see if there is a reasonable correspondence between the length of time the fire burned and the degree of damage it caused.

Exactly this type of analysis was put forward in the case of *Commonwealth v. Han Tak Lee* [17]. The investigator in that case made the following observations:

(a) the fire burned for a total of 28 min;
(b) fire burns 1 in. in 45 min (note: this is a more commonly cited charring rate than Barracato's 3/4 in. h^{-1}); and
(c) 2 by 10 s, were completely consumed.

Therefore, the fire must have been accelerated because it would take 4.5 h to burn through a 2 by 10. This of course, assumed that the fire would only burn in one dimension, as opposed to attacking the wood from both sides. He estimated the time required to burn a 2 by 4 at 1 h 43 min at 1780 °F [17]. The investigator was misled by two myths, one is that the depth of char could be used to determine reliably the time of burning, and two is a uniquely-held belief that only one side of the wood would be attacked by the fire. If we assume, for the sake of argument, that fire burns 1 in. in 45 min, then it should take only 34 min to burn through a 1 1/2 in. piece of wood, assuming it is attacked from both sides. This investigator had both his premise and his implementation of that premise wrong. NFPA 921 unequivocally states that depth of char measurements should not be relied on to determine the duration of the burning [18].

Lines of Demarcation

This is one of the more complex myths in fire investigation because, in some instances, lines of demarcation can be used to tell exactly what happened, whereas in other instances, lines of demarcation are just lines. The threshold question is whether the compartment where the lines occur experienced full room involvement. Let us be clear. There are times when a fire pattern is so obviously caused by an ignitable liquid that further analysis truly is "the icing on the cake". Figure 3 shows just such a situation. Once a fire progresses to the full room, however, it is no longer valid to make a determination using visual clues alone, and there are some who maintain this should never be done. Sharp, continuous, irregular lines of demarcation between burned and unburned areas are frequently cited as evidence of the use of ignitable liquids. It is true that ignitable liquids can produce such patterns on carpeting, and many arson seminars include staged fires that are extinguished early, so that investigators can learn to recognize "pour patterns". What is not evident from these incipient test fires is what happens after the room becomes fully involved.

Lines of demarcation can occur for no apparent reason. The intensity of radiation falls off as the square of the distance from the source to the target, so at some point, perhaps a sharply defined point, insufficient energy exists to maintain combustion. This property, as well as the random nature of some burning, can result in sharp lines of demarcation. The fire pattern shown in Figure 4 is an example of one that is subject to misinterpretation. No ignitable liquids were used to set the test fire, yet there are "sharp continuous, irregular lines of demarcation

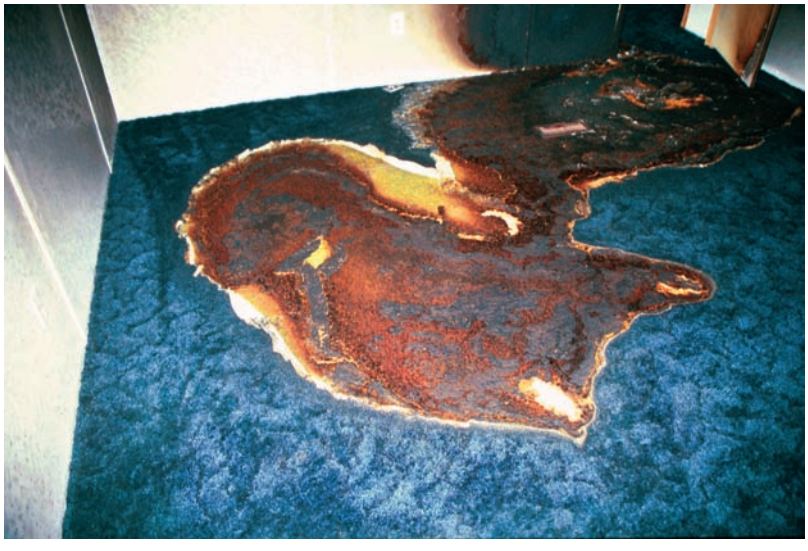


Figure 3 A rare case of an "obvious pour pattern". Only the carpet burned. There was no other fuel present. The carpet shows a "doughnut" pattern, and a sample of the carpet tested positive for the presence of a medium petroleum distillate



Figure 4 A hole in the floor and an irregular fire pattern produced by a test fire in which no ignitable liquids were used

between burned and unburned areas". There is even a hole burned through the floor, and there are still those who argue (incorrectly) that a fire will not burn downward without "help". A naïve person could be easily persuaded that this pattern was created by ignitable liquids burning on the floor. Illustrating the photograph with lines and labels makes it even more convincing, as shown in Figure 5.

Protection patterns can be produced by irregularly shaped pieces of gypsum drywall, which fall from the ceiling and provide protection to whatever floor they land upon. Clothing on the floor has also been known to produce alternating areas of exposure and protection. Figure 6 shows a protection pattern that might, if not carefully examined, be called a "*trail*". Such patterns are not uncommon in residences where the house is not tidy.

In a ground-breaking study of burn patterns caused by burning pools of gasoline and kerosene, Putorti demonstrated that even on wood and vinyl floors, the

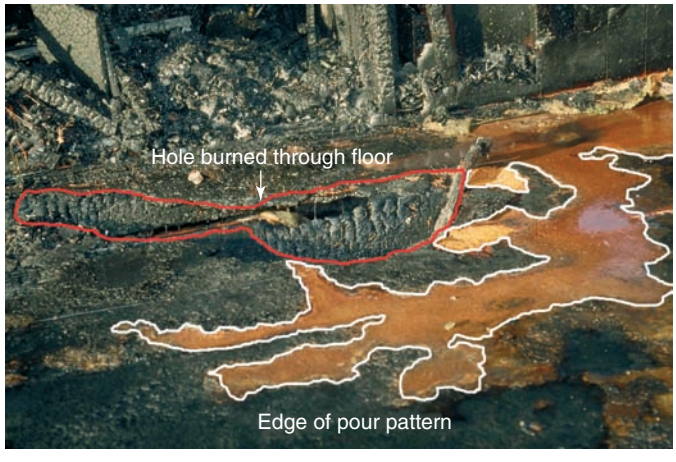


Figure 5 The same fire pattern seen in Figure 4, but annotated to show convincingly, albeit incorrectly, that the pattern was created by pouring ignitable liquid on the floor



Figure 6 "Trailer" pattern created by clearing a path through clothing on the floor of a mobile home burned in a test fire. Close examination of the edges of the pattern reveals the true nature of the sharp lines, but the pattern has the appearance of a trailer. The key to distinguishing real trailers lies in understanding that the unburned sides of the trail should have been burned in this fully involved compartment. Rather than trying to explain the burning, the investigator should explain the lack of burning. In this fire, the only valid explanation is that something protected the unburned areas

edges of the patterns produced are not necessarily sharp [19]. The only definitive pattern he found that could reliably be associated with the use of ignitable liquids was the "doughnut" pattern on carpeting, caused by protection at the center of the pattern by the presence of liquid fuel that had not yet evaporated.

Lines of demarcation in the cross section of charred wood have been cited since 1980 as an indicator of the speed of a fire. The *Fire Investigation Handbook* stated, "A distinct line between charred and uncharred portions indicates a rapidly developing fire. Lack of a distinct line usually indicates a slow cooking process, thus, a slowly developing fire" [20]. O'Connor [13] and O'Connor and Redsicker [14] both provide a diagram of a cross section of a piece of lumber showing a sharp line of demarcation indicating a rapid spread, and a gradual line of demarcation indicating a slow-burning fire. DeHaan [21] states, "One indicator that is more reliable [than the surface appearance of char] is the appearance of the charred wood in cross section. When a charred beam is cut crosswise, the gradation between the charred layer and the underlying undamaged wood is more gradual with a slowly developing fire." He then goes on to provide a perfectly reasoned analysis of why this should be so, but like O'Connor, provides neither data (though he also provides a drawing) nor a definition of what is meant by "sharp", "gradual", "fast", or "slow". DeHaan cautions that a fast-developing fire may or may not be accelerated. Nonetheless, this is the type of "data" that an investigator may use to incorrectly "eliminate" a smoking fire, since smoking fires are not "fast-developing". (Actually, once a smoldering fire started by a cigarette makes the transition to flaming combustion, the speed of fire growth is not distinguishable from a fire ignited by an open flame.)

Some of the more frequently debated sections of NFPA 921 deal with determinations made by observing lines of demarcation. While it is silent on the observation of cross sections, the document contains a whole section devoted to caution in the interpretation of burn patterns on the floor. NFPA 921 contains more cautions on this subject than on any other. The reason for the abundance of cautions on the subject of interpreting lines of demarcation is simple – the errors caused by this particular misinterpretation have been legion.

Sagged Furniture Springs

This myth has been subjected to different interpretations. The Aetna booklet f*ire ... is it arson?* [16] advised fire investigators to photograph furniture springs, "because their appearance can help the investigator to document the area of origin. Severely sagging springs can indicate that a flammable liquid was involved and created heat intense enough to cause the springs to sag" [22]. Carter, on the other hand, writing in *Arson Investigation* [23] stated that collapsing all or part of a coil spring showed the indication of a cigarette that caused the fire.

In the Han Tak Lee case, smoking in bed was ruled out because the bedsprings had lost their temper. Clearly, this is an area of much confusion. In 1989, Tobin and Monson, two Federal Bureau of Investigation (FBI) laboratory scientists, subjected furniture springs, both loaded and unloaded (with and without weights on them), to different fire conditions, and basically concluded that the condition of the springs is of little probative value in fire investigation [24]. DeHaan correctly states that varying degrees of spring damage can provide some insight into the progress of a fire, but cautions that the collapse of springs cannot be reliably used to determine whether a fire was incendiary [25]. NFPA 921 states that the value of analyzing the furniture springs is to compare the differences in the springs to other areas of the mattress, cushion, or frame. Comparative analysis of the springs can assist the investigator in developing hypotheses concerning the relative exposure to a particular heat source.

Spalling

There exists no more misunderstood and misused indicator than concrete spalling. It has been the pivotal "indicator" in many major fire cases, and has been the subject of numerous contentious articles in the *Fire and Arson Investigator*. To this day, arson cases are made on the basis of spalled concrete. Kennedy's *"Blue Book"* [26] had the following comments on spalling:

> Spalling caused by flammable liquids burning is usually found at low levels because the flammable liquid vapors are heavier than air and tend to go down.
>
> ... Regardless of the composition of the concrete or brick, the indicator is the spalled area or areas indicating the burning of accelerants ....
>
> ... The spalling temperatures are usually much higher than the temperatures found in the normal dwelling or commercial building fire. Therefore, we know that accelerants were used.

IFSTA's *Fire Cause Determination* [27] provided the following statement on spalling in 1982:

> Concrete floors and assemblies that have spalling should be examined closely. The spalling may be an indicator of the use of accelerants. If the accelerants had adequate time to soak in before ignition, the spalling will follow the flow pattern of the liquid. Spot spalling is not a clear indicator of the use of accelerants. Further, it is not unusual for spot spalling to result from severe fire exposure.

This semi-cautious language is typical of what has been written about spalling. Skeptics have always questioned the relationship between ignitable liquids and spalling. These include Harvey French in 1979, Fred Smith and Jack Mitchell in 1981, Bruce Ettling in 1984, Charles Midkiff in 1990, and Bernard Beland 1993. Some fire investigators simply ignored all these skeptics, and ploughed on with their case-making.

One of the largest insurance bad faith awards in Alabama history was the result of a fire investigator, who relied on a "trail of spalling" in addition to other "indicators" to conclude that the cause of the fire was arson. It did not help that the fire chief testified that he stood on the "trail" before the fire reached the basement. Nor was the court impressed with the shape of the "trail" when it learned that its shape resulted from the investigator shoveling a trail. When the slab was completely cleared, it was found that the entire slab had spalled, and no "trail" of any kind had ever existed [28].

Fire investigators have argued endlessly about the characteristics of an accelerant-induced spall versus a naturally occurring one. A brownish, or pinkish halo around the hole was thought to indicate the presence of burning hydrocarbons. Numerous slides and photos were exchanged, but in the end, the consensus was that the skeptics were right. It is actually quite simple to cause spalling without ignitable liquids. A stack of oak pallets ignited with newspaper on a concrete slab will almost invariably cause the explosive destruction of the surface through spalling. Even the most diehard defenders of spalling as an indicator of accelerant use would be unlikely to claim that the spalling shown in



Figure 7 Deep spalling on the ceiling of a parking garage

Figure 7 on the ceiling of a parking garage was the result of flammable liquids burning on the ceiling.

Cook and Ide [3] reported that the color change was probably the result of the dehydration of yellow colored hydrated iron oxides, which turned pink or reddish brown approximately at 300 °C.

Most of the writings on spalling today wishfully refer to misconceptions that were formerly held. NFPA 921 has, since its inception, warned about misinterpreting spalling. Overall, 921 states that the importance of spalling to the fire investigator lies in the documentation and analysis of a heat source.

Fuel Load

In one of the early attempts to bring a quantitative approach to the practice of fire investigation, French [29] described a method by which a fire investigator could determine whether the fire behaved in a "normal" manner. He described the process as follows:

> The heat energy production of fuels is extremely important to any competent fire investigator in determining fire load in the premises or equipment under investigation, again in respect to its potential in affecting temperature rise and spread and the time spectrum . . . .
>
> Fire load of any given space may be established by knowing the type of combustibles in storage, their calorific heat producing capacity in Btu's per lb., the total weight of the combustibles in storage, and the square-foot capacity of the space.
>
> The formula is as follows: multiply the calorific contents in Btu's per lb. by the total weight of the contents or materials in pounds. Then, divide the result by the area in square feet. The answer is fire load per square foot.
>
> National Bureau of Standards and American Standards as well as the National Fire Protection Association and British time/temperature curves are in general agreement as to what temperature rise may be expected in various occupancies, with known fire loads, particularly during the first two hours of combustion.
>
> For example, with sufficient oxygen to support continuing combustion, fires in buildings may be expected to attain 1000 to 1200 °F during the first 5 to 10 minutes, accelerating on the curve to approximately 1500 °F in the first half hour and with temperatures reaching the order of 1700 °F at one hour.

Carroll, writing in *Physical and Technical Aspects of Fire and Arson Investigation*, adopted a similar approach, but instead of fire load, he urged investigators to use the flame-spread index described earlier in his text [30]. The vaguely defined process was described in two paragraphs as follows:

> Using this flame spread index (available from Underwriters Laboratories, Inc.), a fire investigator can

determine the comparative rate of how fast a fire should or should not have spread under normal circumstances by comparing the burning rates of known fires and the standard ASTM (American Society of Testing and Materials) time/temperature fire exposure chart shown in Figure 10.

Figure 10 shows the temperature acquired as a function of time, which has been found to be the average temperature 8 feet off the floor.

Carroll also wrote, "By knowing the fire load of the building, i.e., the material available for the creation of heat, a reasonable approximation of the highest temperatures attained can be made and compared with temperatures to be expected had an accelerant been used."

Neither Carroll nor French understood that the "standard time/temperature curve" had nothing to do with the behavior of any fire. The standard time/temperature curve describes the way in which a furnace should be operated in order to compare the fire resistance of various building assemblies. Unfortunately, for many fire victims, including Han Tak Lee in Pennsylvania, this approach resulted in numerous determinations of a fire behaving "abnormally", and this in turn resulted in wrongful prosecutions and convictions.

This particular misconception about fire behavior did not receive as much acceptance as some of the other myths, possibly because it involved mathematics. There were enough practitioners using this "quantitative" approach; however, that the NFPA Technical Committee on Fire Investigations felt the need to address the issue in NFPA 921 in the chapter on basic fire science:

5.4.1 General. The term fuel load has been used to indicate the potential severity of a fire and has been expressed in terms of Btu (British thermal unit) or pounds of fuel per square foot of floor area. An example is provided in 5.4.1.1.

5.4.1.1 The Btus were expressed in wood equivalent based on 8000 Btu per pound. The fuel load was determined by weighing the fuel in a room and converting the weight of plastic to 2 pounds of wood using 16 000 Btu per pound as a value for plastic (1 pound of plastic equals 2 pounds of wood). The total Btus (or pounds of fuel) were divided by the area of the room floor. While this approach can be a measure of the total available if all the fuel burns, **it does not depict how fast the fire will develop once it starts.** *(Emphasis added).*

5.4.1.2 The rate of fire growth as determined by witness statements is highly subjective. Many times witnesses are reporting the fire growth from the time of discovery, which cannot be directly correlated to ignition time. The rate of fire growth is dependent on many factors besides fuel load, to include fuel configuration, compartment size, compartment properties, ventilation, ignition source, and first fuel ignited. The rate of fire growth as reported by witnesses is not reliable or supported independent evidence of an incendiary fire.

5.4.2 Heat Release Rate

5.4.2.1 Total fuel load in the room has no bearing on the rate of growth of a given fire in its preflashover phase. During this period of development, the rate of fire growth is determined by the heat release rate (HRR) from burning of individual fuel arrays. The HRR describes how the available energy is released. This quantity characterizes the power – energy released per unit time (Btu/sec or kilowatts) – and is a quantitative measure of the size of the fire. A generalized HRR curve can be characterized by an initial growth stage, a period of steady-state burning, and decay. The largest value of the HRR measured is defined as the peak heat release rate.

These values should only be considered as representative values for comparison purposes. Fuel items with the same function (e.g., sofas) can have significantly different HRRs. The actual heat released rate for a particular fuel item is best determined by test.

Despite this warning about fuel package variability, some allegedly scientific fire investigators insist on their right to estimate the HRR of an item of furniture, and then opine whether it will have sufficient radiant energy to ignite a nearby item. Usually the estimate is at the low end of the HRR scale, which may range from 500 to 2 500 kW for an armchair for example. The investigator then "deduces" that because the fire did spread, there must have been multiple points of origin.

Low Burning and Holes in the Floor

A common misconception is that because heat rises, fire burns up and out and will not burn downward unless it has "help".

In Carroll's [30] text, he discusses multiple low points and states, "The discovery of a low point should not be considered at the end of the search, since more than one low point may be discovered. This is particularly true in arson fires." He states further that, "Every effort should be made to determine whether multiple low spots are accidental or deliberate. If they have been set in an incendiary effort, these would be considered as evidence of arson."

Carroll states, "Significant differences in char depths at two different low points would indicate an accidental low point" [31]. If the low points or holes in the floor are all about equally charred, the use of multiple holes to indicate multiple points of origin could arguably be justified by referring to this treatise.

The IFSTA manual, *Fire Cause Determination*, espoused a similar misinterpretation when it stated, "Low levels of charring are good indicators of a flammable liquid having been used. For example, accidental fires are unlikely to burn the bottom edge of furniture or the bottom edge of the door" [32].

The Army's Field Manual [33] succinctly restates the popular myth

> Liquid accelerants leave evidence of low burn. That is, they show burning on the floor of the structure. A normal fire chars only the upper portion of a room. Floor damage in natural fires is usually limited to about 20% of the ceiling damage. Low burn, shown by complete charring of large areas of the floor or the baseboards, is not natural.
>
> Nor is fire burning downward natural. Fire burning downward is a prime indicator of the use of a flammable accelerant. Patterns burned in wood floors or holes in a floor may show that an accelerant was used.

Kirk [34] was one of several workers who disagreed with the notion that holes observed in a burned floor necessarily indicated the presence of an ignitable fluid. In 1969, he wrote:

> In many instances, the lowest burn is a floor surface or region directly under a floor. These points are sometimes difficult to evaluate and often lead to error in interpretation. For example, there is a hole burned in the floor in a region away from any walls or other objects that could carry a fire upward by providing fuel in the path of the flames. It is not uncommon for the investigator to assign the cause to the use of a flammable liquid. Such an interpretation is more often incorrect than otherwise. On a tight floor, it is always incorrect, unless holes or deep cracks are present. **Lacking such conditions, flammable liquids never carry fires downward.** (*Emphasis in the original*).

This important point was apparently lost by the investigator who testified that he was able to discern nine separate points of origin in the room shown in Figure 8. The ceiling and roof over this room were entirely consumed, but the investigator saw no connection between the holes burned in the floor. Numerous other myths were introduced in this 1989 criminal prosecution [17].

NFPA 921 has dealt with this myth in a straightforward admonition:

> **6.17.2.2**. Like other areas of low burning, holes in the floor can be produced by the presence of ignitable liquids, glowing embers, or the effects of flashover or full room involvement.

It is this warning in NFPA 921 that drew the ire of many fire investigators who believed they had been properly trained to recognize artifacts indicative of the use of liquid accelerants, even in fully involved compartments. There is no doubt that they had been trained, but that training had no validity.

The Angle of the "V"

The "V" angle myth has been published by many authors. It goes like this: The sides of a "normal" conical fire plume are angled 15° from vertical. The faster a fire burns, the slower it will spread laterally and the closer the angles will be to vertical. Conversely, the slower a fire burns the further the angle will tilt from vertical. Like most of the myths presented in this article, no scientific support exists for this myth; however, it is a deceptively appealing notion. The *NBS Fire Investigation Handbook* stated that the V-pattern should be examined to determine whether the fire was a slowly developing one or a rapidly developing one. Without defining rapid and slow, the intent was apparently to let the indicator do the defining [20]. The authors stated:

> Fire patterns – a wide or diffuse V pattern generally indicates a slowly developing fire. A narrow sharply defined V pattern generally indicates a fast developing, hot fire.

The U.S. Army's (1985) [8] *Field Manual* states:

> Fire burns up and out. It leaves a V-shaped char pattern on walls and vertical structures. A fire which is hot and fast at the point of origin will leave a sharp V pattern. A slow fire will produce a shallow V.

The new, allegedly updated, Army *Field Manual* [35] repeats the same misinformation.

Carroll [36] took a more quantitative approach, at least as far as the angles were concerned:



Figure 8 Fire pattern on a floor allegedly showing nine separate points of origin. The investigator also opined that 60 gallons of liquid accelerant were applied to the floor, though the laboratory tests all came back negative

> A normal fire, consuming wood, plastic or electrical insulation, would burn with a "V" pattern of approximately 30° measured vertically. If an accelerant was used, or if highly combustible material was involved, the "V" would be narrower as the temperature of the fire increased, due to the additional heat content of the accelerant or flammable liquid. This would cause a faster rise of heat and flame, resulting in a "V" pattern of approximately 10° depending on the heat flux generated by the accelerant.

Noon [11] devotes an entire section of his book to burning velocities and V patterns. This discussion is accompanied by equations to help the investigator determine the ratio of the upward burning rate to the lateral burning rate by finding the tangent of the angle of the V.

O'Connor [13] and O'Connor and Redsicker [14] repeat the story:

> The breadth or width of the V (also called the *funnel pattern*) is affected by (and hence, indicative of) the buildup, progression, speed and intensity of the fire. An intense rapidly moving fire produces a narrow the pattern whereas a slow, less intense fire produces a wide V pattern. The angles of the boundaries average between 10 and 15°.

In the third edition of *Kirk's Fire Investigation*, DeHaan urged caution in the interpretation of the pattern angles when he stated, "Although it is sometimes claimed that the more vertical the sides of the V, the faster the initial fire (and therefore the more suspicious), one can appreciate that the nature of any wall covering and conditions of ventilation have important effects on the shape of the pattern, and must be taken into account" [37]. By the fourth edition, he came right out and said, "The angle and width of the V are not dependent on the rapidity of ignition of the fuel" [38].

Since it was first published in 1992, NFPA 921 has contained a section on the interpretation of V patterns, and describes the equation of angles with speed as a "misconception". The current (2004) edition makes the following statement:

> **6.17.2.2.** The angle of the borders of the V pattern does not indicate the speed of fire growth; that is, a wide V. does not indicate a slowly growing fire, or a narrow V does not indicate a rapidly growing fire.

It is therefore surprising that when the U.S. Army revised their *Field Manual* in 2005 and stated that the revision contained information from NFPA, they would continue to include the V angle myth despite

the fact that NFPA 921 had disparaged this myth through five editions over 13 years.

Time and Temperature

A fire that burns "hotter than normal" or "faster than normal" is thought to indicate an accelerated fire. Actually, fire temperature and the perceived speed of the fire are not valid indicators of a fire's cause.

A major misconception underlying many false determinations of arson is that the temperature achieved by a particular fire can help an investigator to evaluate whether a fire was "normal" or "abnormal", with an abnormal fire being attributed to incendiary activity. Higher than "normal" temperatures indicate that a set fire is such an appealing notion that even Paul Kirk appeared to agree with the concept, as previously discussed. To this day, investigators sometimes infer the presence of accelerants when they observe a melted aluminum threshold.

Kennedy [39], discussing the melting of copper, wrote:

> Copper fuses at 1980 °F, which is very high. Therefore, if we find fused copper or beaded copper wires, we are immediately alerted because we have an unusually high temperature that must be explained. The normal burning of a structure would not cause temperatures in the 2000 °F range, which is necessary to fuse or melt copper.
>
> What could cause copper to fuse or melt? The burning of an accelerant such as flammable liquids, natural or LP gases or electrical shorts or arcing are some of the causes for "high heat" – meaning excessive temperatures.

Barracato [40] summed up the time and temperature equation as follows:

> Fires which burn through entire floor sections, destroy large support beams in a relatively short period of time, or melt articles located in the area of origin such as metal, copper, aluminum or glass, are unusual. It takes tremendous heat to cause such damage and unless there is a rational explanation for the heat buildup – for example, if the room was used to store a highly flammable material – it's very probable that the fire was intentional and an accelerant was used.

Carroll, as discussed previously under "fire load", presented the standard time/temperature curve from ASTM E119, *Standard Test Methods for Fire Tests of Building Construction and Materials*, and stated that it can be used as a basis in comparing the burning rates in structures. Carroll also stated, "if accelerant or other chemicals are present, temperatures can reach higher than on the standard fire curve" [41].

On the basis of the research that began in the late 1970s and continues until today, it is now well understood that there is no valid definition of "normal" fire spread and also that the ASTM time/ temperature curve has little relationship to the behavior of a "normal" fire.

In the second edition of *Kirl's Fire Investigation* [38], DeHaan somewhat moderated Paul Kirk's enthusiasm for interpreting melted metals, but still left readers with the suggestion that an abnormal fuel load, such as provided by an accelerant, will increase temperatures [42].

Although the third, fourth, and fifth editions of *Kirk's* include a discussion of the fact that gasoline burns at essentially the same temperature as wood, it still states that temperature can be used to determine the presence of "enhanced draft conditions or unusual fuel loads", when the data support only the former. Nonetheless, the modern text of *Kirk's* at least recognizes that increased ventilation causes increased temperatures.

NFPA 921, beginning with the first edition and continuing until today contains an admonition about placing too much stock in the perceived temperature of a fire. In the 2004 edition of the document, both temperature and speed of the fire are addressed, and investigators are warned to be cautious when interpreting temperature and speed. The following statement appears:

> **6.8.2.2** Wood and gasoline burn at essentially the same flame temperature. The turbulent diffusion flame temperatures of all hydrocarbon fuels (plastics and ignitable liquids) and cellulosic fuels are approximately the same, although the fuels release heat at different rates.

The speed at which a fire progresses is frequently used to imply that the fire is incendiary. While it is true that an accelerated fire burns faster than an unaccelerated fire, at least in its initial stages, serious caution is required when confronted with information about how rapidly a fire spread. Most observations about the "speed" of a fire are provided by eyewitnesses, but there have been reported instances of an investigator looking at a destroyed structure and by knowing the time from alarm to extinguishment,

opining that the amount of destruction could not have occurred in that timeframe unless the fire had "help". These conclusions are usually based on the misconception that the wood has a fixed burning rate, such as the often quoted "1 in. in 45 min".

A study by the editors of *Fire Findings*, published in 1995, revealed that witnesses might make very different observations about a fire, even if their sightings are only a few minutes apart [43]. This study, which involved the actual burning of a two-story house, resulted in the findings that fires may only appear to start rapidly and tended to dispel the widely held (but incorrect) belief that if a fire appears to start quickly, accelerants must have been involved. If an eyewitness only notes the existence of a fire at the point where it breaks out a window, the progress thereafter will be rapid indeed, regardless of the cause.

In the fifth edition of *Kirk's Fire Investigation*, DeHaan published a time/temperature curve for a fire in a "typical furnished room" with no accelerants that shows flashover occurring just 210 s (3.5 min) after ignition. This curve looks nothing at all like the "standard time/temperature curve" from the ASTM test method [21].

Thus, while it is true that accelerants tend to make fires burn more rapidly, a rapidly burning fire does not necessarily indicate the presence of accelerants.

Conclusion

Fire investigation involves the comparison of the investigator's "expectations" with his perception of the behavior of the fire. If those expectations are not properly "calibrated", the result will be numerous errors. In the seventeenth century, when the scientific community was first getting organized, it was understandable that misconceptions about fire, such as the phlogiston and caloric theories, should exist. What is surprising is that after three centuries of scientific examination of fire, myths have been added rather than dispelled.

Many of the myths have been taught to the individuals who are now in control of the fire investigation industry. These myths were promulgated by the National Fire Academy and by other entities involved in the training of fire investigators, and given the stamp of approval of one of the most prestigious and credible US government agencies. To date, the misinformation has not been officially repudiated or "recalled". While most responsible training organizations no longer teach the myths, there are still plenty of fire investigators who have not been back to school since it was learned scientifically that the "indicators of arson" that they learned were false. Further, the libraries of most fire investigators contain numerous texts that are filled with this misinformation. Even worse, there are still a few popular speakers and agencies training new investigators with the old mythology.

The sheer number of misconceptions and their widespread publication in learned and not-so-learned treatises indicate that fire investigation, as a profession, still has very far to go.

References

[1] Aerospace Corporation (1977). *Arson and Arson Investigation: Survey and Assessment*, National Institute of Law Enforcement and Criminal Justice, LEAA, USDOJ, p. 87.

[2] Brannigan, F., Bright, R. & Jason, N. (1980). *Fire Investigation Handbook*, *NBS Handbook 134*, p. 5.

[3] Cooke, R. & Ide, R. (1985). *Principles of Fire Investigation*, The Institution of Fire Engineers, Leicester, p. 134.

[4] Kirk, P. (1969). *Fire Investigation*, John Wiley & Sons, p. 145.

[5] O'Connor, J. & Redsicker, D. (1997). *Practical Fire and Arson Investigation*, CRC Press, p. 107.

[6] Lentini, J. (1992). The lime street fire: another perspective, *The Fire and Arson Investigator* **43**(1), p. 52.

[7] IFSTA (1982). *Fire Cause Determination*, 1st Edition, Fire Protection Publications, OSU, p. 48.

[8] U. S. Army (1985). *Law Enforcement Investigations*, Field Manual 19–20, p. 220.

[9] O'Connor, J. (1986). *Practical Fire and Arson Investigation*, CRC Press, Boca Raton, p. 88.

[10] O'Connor, J. & Redsicker, D. (1997). *Practical Fire and Arson Investigation*, CRC Press, Boca Raton, p. 99.

[11] Noon, R. (1995). *Engineering Analysis of Fires and Explosions*, p. 131.

[12] IFSTA (1982). *Fire Cause Determination*, 1st Edition, Fire Protection Publications, OSU, p. 46.

[13] O'Connor, J. (1986). *Practical Fire and Arson Investigation*, CRC Press, Boca Raton, p. 76.

[14] O'Connor, J. & Redsicker, D. (1997). *Practical Fire and Arson Investigation*, CRC Press, Boca Raton, p. 77.

[15] Lentini, J. (1992). Behavior of glass at elevated temperatures, *Journal of Forensic Sciences* **37**(5), 1358.

[16] Barracato, J. (1979). *Fire . . . is it arson?* Aetna Life & Casualty, p. 15.

[17] Commonwealth of Pennsylvania v. Han Tak Lee, Court of Common Pleas of Monroe County, 43rd Judicial District, No. 577 Criminal, 1989. Report of Daniel Aston, 1990.

[18] NFPA 921 (2004). *Guide for Fire and Explosion Investigations*, National Fire Protection Association, Quincy, p. 34.
[19] Putorti, A. (2000). *Flammable and Combustible Liquid Spill/Burn Patterns*, NIJ Report 604–00, U.S. Department of Justice, Office of Justice Programs, National Institute of Justice.
[20] Brannigan, R., Bright, R. & Jason, N. (1980). *Fire Investigation Handbook*, *NBS Handbook 134*, p. 6.
[21] DeHaan, J. (2002). *Kirk's Fire Investigation*, 5th Edition, Prentice Hall, p. 42.
[22] Barracato, J. (1979). *Fire . . . is it arson?* Aetna Life & Casualty, p. 23.
[23] Carter, R. (1978). *Arson Investigation*, Glencoe Press, Encino, p. 97.
[24] Tobin, W. & Monson, K. (1989). Collapsed spring observations in arson investigations: a critical metallurgical evaluation, *Fire Technology* **25**(4), 317.
[25] DeHaan, J. (2002). *Kirk's Fire Investigation*, 5th Edition, Prentice Hall, p. 212.
[26] Kennedy, J. (1977). *Fire-Arson Explosion Investigation*, Investigations Institute, Chicago, p. 392.
[27] IFSTA (1982). *Fire Cause Determination*, 1st Edition, Fire Protection Publications, OSU, p. 48.
[28] USAA v. Wade, 5344 So. 2d 906 Ala (1989).
[29] French, H. (1979). *The Anatomy of Arson*, Arco Publishing, New York, p. 36.
[30] Carroll, J. (1979). *Physical and Technical Aspects of Fire and Arson Investigation*, Charles C. Thomas, p. 105.
[31] Carroll, J. (1979). *Physical and Technical Aspects of Fire and Arson Investigation*, Charles C. Thomas, p. 105.
[32] IFSTA (1982). *Fire Cause Determination*, 1st Edition, Fire Protection Publications, OSU, p. 81.
[33] U. S. Army (1985). *Law Enforcement Investigations*, Field Manual 19–20, p. 225.
[34] Kirk, P. (1969). *Fire Investigation*, John Wiley & Sons, p. 74.
[35] U. S. Army (2005). *Law Enforcement Investigations*, FM 3-19-13, p. 7–3.
[36] Carroll, J. (1979). *Physical and Technical Aspects of Fire and Arson Investigation*, Charles C. Thomas, p. 103.
[37] DeHaan, (1991). *Kirk's Fire Investigation*, 3rd Edition, Prentice Hall, p. 91.
[38] DeHaan, (1997). *Kirk's Fire Investigation*, 4th Edition, Prentice Hall, p. 148.
[39] Kennedy, J. (1977). *Fire-Arson Explosion Investigation*, Investigations Institute, Chicago, p. 396.
[40] Barracato, J. (1979). *Fire . . . is it arson?* Aetna Life & Casualty, p. 15.
[41] Carroll, J. (1979). *Physical and Technical Aspects of Fire and Arson Investigation*, Charles C. Thomas, p. 54.
[42] DeHaan, J. (1983). *Kirk's Fire Investigation*, 2nd Edition, Prentice Hall, p. 173.
[43] Sanderson, J. (1995). Fire timing test results: fIRES may only appear to start rapidly, *Fire Findings* **3**(3), 1.

JOHN J. LENTINI

Arson Investigation: Mythology

see Arson Investigation: Misconceptions and Mythology

Arson Investigation: Pyromania

see Firesetting

Asphyxia

Introduction

Asphyxia is a classical term in forensic medicine that has come to have varied definitions and meanings. Etymologically, asphyxia means "pulseless", but in common forensic usage it has come to denote deaths from the acute cerebral deprivation of oxygen. On this basis, asphyxia can be unambiguously defined as death from rapid cerebral anoxia or hypoxia. The concept of asphyxia, and more particularly whether asphyxial deaths have specific postmortem signs, has been extensively discussed in the literature and classical textbooks of forensic pathology. There are long-standing myths about asphyxia that continue to be difficult to dispel.

The first myth about asphyxia relates to a misunderstanding of what constitutes the nature of the asphyxia. Specifically, asphyxia is not a cause of death but rather a mechanism of death. On this basis, the opinion that "asphyxia" has occurred provides no specific medicolegal information, unless the cause of the asphyxia can be ascertained (e.g., strangulation). Thus, without specific knowledge about how the "asphyxia" has occurred, there is no medicolegal value to providing a diagnosis of asphyxia. This closely relates to the most important myth about asphyxial deaths: that there are diagnostic postmortem signs of asphyxia. It is now known that there are no specific gross or microscopic changes that will allow the determination of asphyxia as a mechanism

Table 1 Classification of rapid anoxial deaths

Mechanism	Classical terminology	Examples
Reduced oxygen tension in the respirable atmosphere	Environmental hypoxia	Rebreathing within a plastic bag sealed overhead
Obstruction of the mouth and nose	Upper airway obstruction	Choking; gagging; suffocation/smothering
Interference with the mechanics of breathing or chest compression	Positional asphyxia	Heavy load compressing chest wall and limiting respiratory excursion; acute ethanol intoxication with fall into posture causing hyperflexion of the neck; hogtying in prone position
Neck compression	Pressure on the neck	Strangulation; hanging; neck holds
Chemical inhibition of oxygen utilization for metabolism	Chemical anoxia	Cyanide and carbon monoxide poisoning

of death. The best that can be accomplished using present methods of forensic medicine is to determine the cause of asphyxiation from direct evidence at the autopsy and then infer that the mechanism of death was asphyxial. Otherwise, it is not possible to determine that a death is asphyxial without resorting to circumstantial information. The classical signs of asphyxia including cyanosis, fluidity of blood, and visceral congestion and petechiae have been thoroughly debunked and are now considered to be unreliable indicators of a rapid anoxial death.

Another important and persistent myth about rapid anoxial deaths relates to the pathogenesis and medicolegal significance of ocular petechiae. Although it was once believed that petechiae were caused by asphyxia, this is not supported by available evidence. It is now generally accepted that petechial hemorrhages of the conjunctivae and face are due to rupture of small blood vessels due to increased jugular venous pressure rather than due to a tissue effect of hypoxia. Thus, in the context of rapid anoxial deaths, ocular petechiae are most frequently observed in cases where the cause of hypoxia coincidentally involves impairment or obstruction to the venous return from the head (e.g., compression of the neck with jugular venous compression or chest compression with increased intrathoracic and central venous pressure). Ocular petechiae are best interpreted as an indication of a pathophysiological process that has resulted in impaired venous return from the head. Thus, petechiae are not sure evidence of "asphyxia". This is demonstrated in routine medicolegal autopsy practice, where one of the most frequent causes of ocular petechiae is sudden cardiac death from ischemic heart disease.

Overall, rapid anoxial deaths are best classified by considering the level at which oxygen supply is lost and therefore unavailable for metabolism (Table 1). The deficiency in oxygen may be due to reduced oxygen tension in the respirable atmosphere; obstruction of the mouth and nose; interference with the mechanics of breathing; pressure on the neck; and chemical inhibition of oxygen utilization for metabolism. Traditionally, the term *mechanical asphyxia* has included any mechanism that interferes with oxygen delivery to the brain by application of an external force on the body, or the action of gravity. In the case of pressure on the neck, the most frequent means by which cerebral anoxia occurs is through reduction in cerebral arterial perfusion due to compression of the carotid arteries.

Drowning is best viewed as a rapid anoxial death due to airway obstruction by water. However, other pathophysiologic effects also occur. Therefore, drowning is considered separately.

Upper Airway Obstruction

Partial or complete obstruction of the nose, mouth, or upper aerodigestive tract created by a physical barrier to airflow can result in rapid anoxial death. The physical barrier maybe fixed in position such as in the application of a gag over the mouth and

nose, or occlusion of the upper aerodigestive tract by physical material (e.g., foreign body placed in the mouth or bolus of food). The physical barrier may be transient such as in smothering and suffocation.

Postmortem determination that death has occurred from upper airway obstruction can be difficult unless the physical barrier is still *in situ*, or if injuries are sufficiently characteristic to infer that a physical barrier was present. Homicidal upper airway obstruction sometimes occurs in infants. This poses problems to detection since the paucity of autopsy findings can result in the misdiagnosis of sudden infant death syndrome.

Gagging

Gagging can be defined as obstruction of the mouth and/or nose by a fixed physical barrier that is held in place by tying (e.g., knotted rag) or adhesion (e.g., wide tape such as "duct tape"). At autopsy, the main findings of gagging relate to the nature of the gag, injuries produced by the gag, and the secondary effects of the gagging process. The nature of the gag is easy to ascertain if the gag is left *in situ*. However, if the gag has been removed (after death but before the autopsy), then the prior presence of a gag can only be inferred from the injuries, secondary effects of the gag, and circumstantial information.

Gags are generally fixed into position by tying to the head or adhesion to the skin, if the gag is a tape. The completeness of occlusion is determined by the range of the gag, i.e., if it covers both the mouth and nose. Sometimes the obstructing effects of the gag are complicated by the presence of a separate foreign body placed in the mouth (e.g., cloth or sock in the oral cavity with duct tape over the mouth), which may also be referred to as a *gag*. The absorptive properties of the gag (e.g., cloth) and the presence of an intraoral component may further cause the gag to swell, and obstruction of the airway may occur progressively rather than acutely. In addition, the tongue may be displaced into the posterior oropharynx, further compromising the airway. If an adhesive tape-type gag has been removed, residue can sometimes be collected from the skin to indicate that a gag was present.

Injuries produced by the gag relate to direct damage to the corners of the mouth and oral mucosa by compression of tissue by the gag. Sometimes there may be facial injuries radiating from the corners of the mouth.

The secondary effects of gagging include aspiration of gastric contents. Since the fixed barrier over the mouth inhibits expelling vomitus, the airway will become occluded and death will occur by airway obstruction. This is often the immediate cause of death in cases of gagging when the mouth is obstructed but the nose is open. However, great care must be used when determining if gastric contents in the air passages have resulted from true aspiration rather than transfer during movement of the body to the mortuary. The other main secondary effect of gagging is more contextual and situationally defined. Sometimes elderly people with ischemic and/or hypertensive heart disease die suddenly while gagged, but without complete occlusion of both the mouth and nose. These fatalities are usually due to sudden cardiac death from underlying heart disease that is exacerbated by gagging, rather than the obstructing effects of the gag itself.

Smothering and Suffocation

Smothering and suffocation may be used interchangeably and are best defined as rapid anoxial death by facial compression with external obstruction of the mouth and/or mouth by a nonfixed or transient barrier. Smothering and suffocation are often used to denote transient obstruction of the nose and mouth by hands and fingers, or the application of a soft object to compress the face and thereby obstruct the external air passages. However, suffocation is also sometimes used to denote the rapid anoxial death that results from the application of an enclosing barrier over the head, such as a sealed plastic bag. The main challenge in determining whether death has occurred from smothering or suffocation is that injuries to the mouth, nose, lips, and face that are specific and reliable indicators of facial compression are not consistently present. At the extremes of age or when there is a reduced level consciousness, smothering or suffocation can occur with no external marks. In contrast, in responsive and resistant people, the injuries to the face can be extensive and are often associated with other injuries. When findings are present, the injuries are typically abrasions or contusions centered on the nose and mouth, as well as lacerations on the inner surface of the labial mucosa. Another pitfall is the overlap in perimortem injuries caused by emergency

medical intervention such as intubation, which may cause focal injuries to the lips and face. This can be particularly problematic in infants and children.

The gross and microscopic findings in cases of smothering and suffocation are controversial. It has been claimed that hyperinflation of the lungs (presumably due to expiration against a closed airway), pulmonary hemorrhage, and interstitial pulmonary emphysema are highly correlated with smothering and suffocation. The evidence to support these claims is anecdotal and the specificity of these findings has not been established. The finding that has been most extensively studied is pulmonary hemorrhage. It is clear that extensive acute pulmonary hemorrhage can be observed in individual cases of smothering in infancy. However, it has also been shown that pulmonary hemorrhage is found in infants who die suddenly from natural causes. There is some evidence that sublethal smothering maybe correlated with hemosiderin-laden macrophages in the lung, leading to the inference that a prior episode of smothering causing pulmonary hemorrhage occurred.

Choking

Choking is obstruction of the upper aerodigestive tract by food. Choking is almost invariably associated with two factors: a predisposing condition or circumstance that results in impaired swallowing and a bolus of incompletely masticated food that has a sufficient diameter to occlude the lumen of the airway. The predisposing condition or circumstance is usually acute ethanol intoxication, neurodegenerative or neuromuscular disease, poor dental hygiene, or the rapid and poorly controlled ingestion of food. The foodstuffs that are frequently observed in cases of choking are usually firm, such as meat. The main pitfalls in the autopsy diagnosis of choking usually relate to the history. If emergency medical intervention has occurred, the bolus of food may have been removed and no trace may be found at autopsy. Alternatively, endotracheal intubation can displace the bolus further down into the trachea. In addition, it is important to differentiate among postmortem transfer of gastric contents from the stomach into the trachea, aspiration of gastric contents, and choking. In choking, the obstruction is a discrete piece of material that is firmly wedged into a narrow portion of the upper aerodigestive tract, typically the laryngeal aperture.

Positional Asphyxia

Positional asphyxia is rapid anoxial death that occurs as a result of external compression of the body or postural/gravitational effects on the body that impair breathing at the level of the neck or chest wall. The key to the determination of positional asphyxia is the history including information regarding the position of the body at the scene, since positional asphyxia cannot be ascertained purely from autopsy findings. Positional asphyxia can be arbitrarily divided into crush (traumatic), postural, and restraint asphyxia. These categories are classically defined and relate to the common circumstances of how the "positional" elements of the "positional asphyxia" arose, rather than representing fundamentally different entities. Other forms of positional asphyxia include oddities such as burking (e.g., homicidal chest compression by sitting on the chest and impairing breathing).

Crush Asphyxia

Crush or traumatic asphyxia occurs when the chest is compressed by large external forces such as a considerable weight, or when the chest is wedged into a tight space, thus preventing expansion during inspiration. In many cases of crush asphyxia, there is a well-demarcated line of florid congestion and petechiae that is cephalad (toward the head) to the main horizontal line of chest compression. The nature of the chest compression may be reflected in injuries to the chest wall such as bruising, abrasion, and rib fractures. However, in many cases, the compression is caused by a flat and broad surface so no chest wall injuries may occur. Crush asphyxia occurs frequently in industrial mishaps and in some road traffic fatalities.

Postural Asphyxia

Postural asphyxia occurs when the fixed gravitational posture of the body precludes breathing or respiratory air transmission. This is distinct from crush and restraint asphyxia due to the lack of an external force or agent actually inhibiting breathing. The most frequent underlying cause for postural asphyxia is acute ethanol intoxication. In this circumstance, the person collapses and becomes unconscious in a position that impairs breathing. The positions that are highly correlated with postural asphyxia are

neck hyperflexion, prone and head down positioning, wedging between furniture or other fixed structures, and complex positions that may indicate many different physical impediments to breathing. There are no distinct pathological findings related to postural asphyxia unless there is an element of direct injury to the neck or chest. The distribution of postmortem hypostasis often provides information on body position, but scene information and toxicological findings are key to the diagnosis.

Another specific form of postural asphyxia is hogtying in the prone position. In this circumstance, an individual's ankles and wrists are tied together behind the back and the person is placed with the body weight concentrated on the chest. This position impairs movement of the chest wall during inspiration and the ligature restraints inhibit a change of position. This form of postural asphyxia may be aggravated by gagging or preexisting cardiopulmonary disease such as ischemic heart disease. This form of postural asphyxia overlaps with restraint asphyxia.

Restraint Asphyxia

Restraint asphyxia is considered by many forensic pathologists as a variant of positional asphyxia that occurs during physical restraint by another party. This is a controversial area that has not reached a consensus, since there are often factors that complicate a precise understanding of the cause and mechanism of death in these cases. Despite the controversy, it is generally accepted that individuals who are restrained, typically in the *prone* position during a bout of agitation and/or bizarre behavior, can die suddenly and unexpectedly. It is not uniformly accepted that these cases are rapid anoxial deaths due solely to the restriction of breathing or other forms of mechanical interference with neck or chest. The majority of these cases occur in the context of excited delirium from acute decompensation of a chronic psychiatric illness or acute cocaine intoxication. Therefore, one of the competing hypotheses to explain these deaths is sudden cardiac arrhythmia related to the excited delirium. On a practical level, this often becomes a challenge in the individual case if there are injuries on the neck that could relate to neck compression.

In the typical case, an individual is restrained by law enforcement officers in the prone position. The individual usually manifests excited delirium with agitation, bizarre, inappropriate and often confrontational behavior, apparent great strength, and hyperthermia. The individual usually becomes unresponsive in the prone position after a violent struggle that may involve the deployment of pepper spray, a taser, or physical techniques aimed at subduing resistance. Clearly, it is important to differentiate death *during* restraint from death *by* restraint, but this is frequently impossible.

Pressure on the Neck

Rapid anoxial death often results from pressure applied to the neck. The different forms of neck compression are determined by the physical means of application of the pressure. The main challenge for the forensic pathologist is realizing that the pathology of neck compression is not entirely specific. Thus, the injuries that result from pressure on the neck reflect both the nature of the compressive force and the physical effects of the compression. For example, the types of internal damage from hanging and strangulation can be the same since both compress the neck, but the distribution of the findings and the precise nature of the skin injuries can be used to differentiate between hanging and strangulation. Furthermore, the postmortem diagnosis of neck compression has three well-defined pitfalls. First, the postmortem findings of neck compression form a spectrum from no or minimal injury to extensive or maximal injury, the former often providing a barrier to detecting subtle or concealed homicides by neck compression (i.e., the under-diagnosis pitfall resulting in "missing" cases). Second, internal injuries of neck compression can be mimicked by postmortem artifacts including hypostatic hemorrhages (i.e., the over-diagnosis pitfall resulting in "over-detecting" cases). Third, interpreting the medicolegal significance of postmortem findings of neck compression is entirely contextual, unlike many deaths by physical injury. For example, the autopsy findings in a case of hanging cannot always be differentiated from postmortem suspension of the body. Similarly, the injuries in cases of nonfatal and fatal strangulation are often the same. In these circumstances, determining the relevance of the autopsy findings to the cause of death relies on understanding the circumstances of the case and exclusion of other causes of death.

Table 2 Pathology of anterior neck compression

Site	Classical injury pattern	Pitfalls and emerging issues
Skin	Abrasions; contusions; ligature marks	Hypostatic hemorrhages mimic contusions; neck folds mimic ligature marks
Fat and muscle	Subcutaneous, intramuscular and myofascial hemorrhage	Hypostatic hemorrhages; dissection artifacts (requires dissection in avascular field)
Hyoid	Fractures (most involve greater cornua)	Infrequent in infants, children, and young women
Laryngeal cartilage	Fractures (most at the base of superior cornua)	More frequent than hyoid fractures
Laryngeal mucosa	Mucosal hemorrhage and intralaryngeal mucosal hemorrhage	Specificity not completely studied; frequently observed in nonviolent deaths
Pharynx	Intramural hemorrhage	Specificity not completely studied; identical to Prinsloo–Gordon hemorrhage
Carotid arteries	Intimal tears	Frequency not well studied

Pathology of Neck Compression

There are five cardinal postmortem signs or hallmarks of neck compression: ocular petechiae; injuries on the skin of the anterior neck; injuries to strap muscles and soft tissues of the neck; injuries of hyoid-larynx complex; and miscellaneous or variable findings (Table 2). However, these hallmarks cannot be considered as diagnostic criteria for neck compression. It seems that straightforward diagnostic criteria for the postmortem diagnosis of neck compression cannot be proposed because of the inherent variability of the processes involved.

Ocular petechiae are related to the compression of the jugular venous system with rupture of small mucosal blood vessels as described above (Figure 1). Injuries on the skin of the neck are often quite helpful in determining the nature of the neck compression (e.g., manual or ligature strangulation, or hanging). Internal injuries including strap muscle and soft tissue hemorrhages are often helpful in diagnosis, particularly if the cutaneous findings are minimal (Figure 2). Fractures of the hyoid bone or larynx occur in less than 50% of female victims of strangulation and are probably more frequent in male victims. The presence of such fractures is often viewed as good evidence of neck compression. The laryngeal cartilage is more often fractured than the hyoid bone. The most frequent fracture is situated at the base of the superior cornu of the thyroid cartilage. In addition, there are often mucosal hemorrhages in the larynx and pharynx. Hemorrhages and microfractures in the laryngeal cartilage have been reported but are not well studied. Intimal tears of the carotid arteries can be seen in some cases of neck compression.



Figure 1 *Ocular petechiae*. These hemorrhages are related to the compression of the jugular venous system with rupture of small mucosal blood vessels

Hanging

Hanging is rapid anoxial death from ligature compression of the neck with (or "accompanied by") suspension of the body. The pressure on the neck is caused by gravitational tension on the ligature. The



Figure 2 *Strap muscle hemorrhages*. These hemorrhages are related to the blunt injury (bruising) of the soft tissues of the neck by applied force

pathology of hanging is highly variable and it relates to several factors including the nature of the ligature, degree of suspension, and relative position of the ligature. The surface injury on the neck in hanging is largely related to the physical properties of the ligature. If the ligature is soft and broad (e.g., suicidal bedsheet hanging in jail) there may be no mark on the neck. If the ligature is narrow and rough (e.g., a braided rope), the ligature mark may be a deep abraded furrow with a corresponding woven pattern within the ligature mark (Figure 3). Mostly, the ligature mark is a well-demarcated curvilinear abrasion consisting of a continuous component and an intermittent or incomplete component that is angled toward the point of suspension. Typically, this is evident by a rising or sloping nature of the ligature mark on the side(s) of the neck, depending on the precise position of the ligature and the point of suspension. Intense pressure on the neck may result in protrusion and postmortem drying of the tongue and perimortem epistaxis. It is classically stated that the main difference between hanging and ligature strangulation

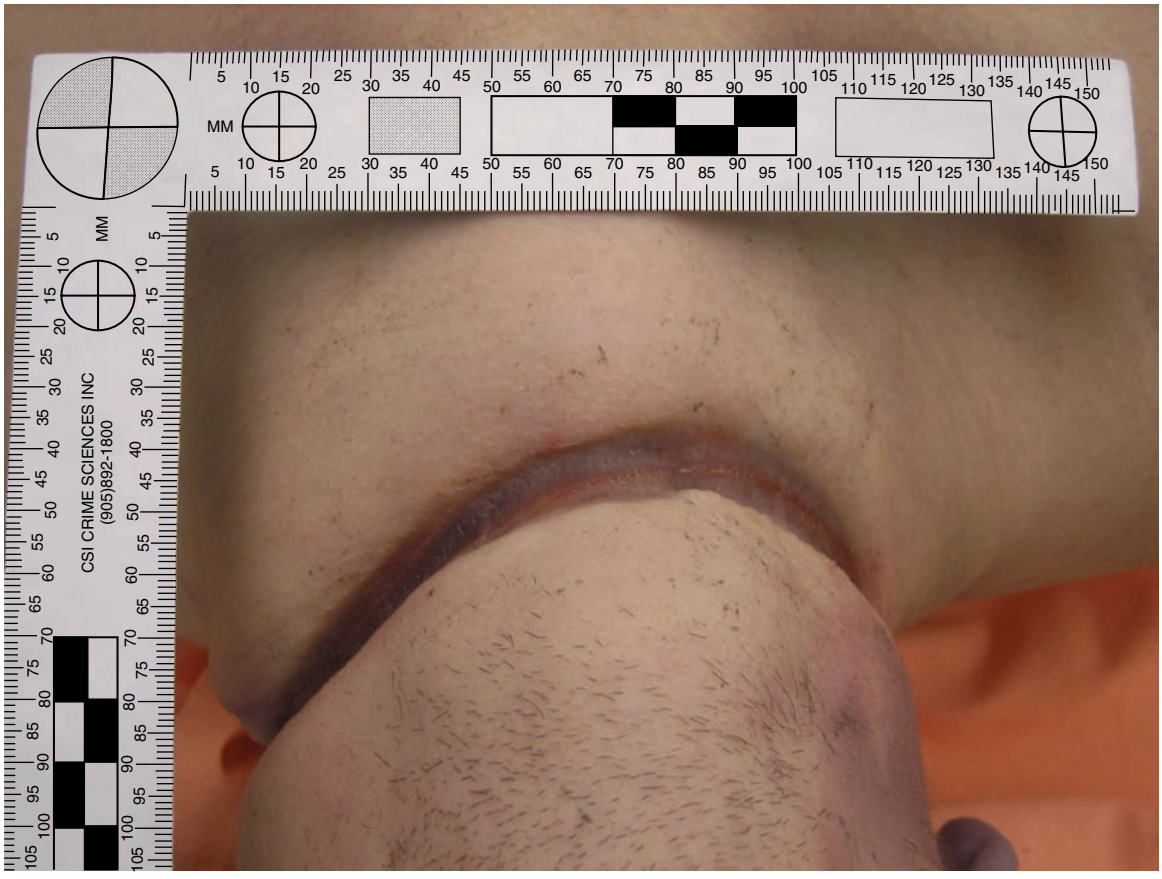


Figure 3 *Ligature mark in hanging*. The ligature mark is an indented abrasion or furrow caused by pressure applied to the skin by the weight of the body during suspension

is the horizontal nature of the ligature mark in the latter. However, this is an overstatement. Hanging may occur with relatively horizontal ligature mark, depending on the precise physical arrangements of the body and ligature. However, in general, ligature marks with hanging are typically not uniformly horizontal in orientation.

Internal injuries associated with hanging may range widely but are typically minimal. If subcutaneous or intramuscular hemorrhage is present, it is usually associated with the ligature mark, or is positioned over a prominence of the larynx or hyoid bone. Fractures of the hyoid and larynx may be detected and are often directly beneath a deeply indented ligature furrow or a knot.

Specific forensic variants of hanging include autoerotic hanging, atypical hanging with incomplete suspension, "long-drop" (or judicial) hanging, and so-called delayed hanging. In autoerotic hanging, hypoxia is induced by regulated neck compression (partial suspension by hanging, with release of the pressure on the neck by standing or kneeling) during masturbation. If the neck compression leads to unconsciousness, then the suspension may be accidentally fatal. In atypical or incomplete suspension hanging, there is often a physical process that results in neck compression by gravitational means that either involves a fixed surface or a ligature. For example, a toddler may be accidentally suspended in the drawstrings of a window blind. If there is complete suspension there may be florid ocular and facial petechiae. In "long-drop" (or judicial) hanging, there is often fracture of the upper cervical vertebra. Finally, if an individual is successfully resuscitated

from cardiopulmonary arrest after hanging, death may occur later as a delayed death following hanging. This is often observed in suicidal hangings in prison with immediate medical intervention. At autopsy, there may be no findings other than hypoxic–ischemic encephalopathy.

A rare occurrence is homicidal hanging. Homicidal hanging is often associated with incapacitating injury or intoxication, or other evidence of violence. Homicidal hanging must be differentiated from postmortem suspension of the body after strangulation, which is also rare.

Strangulation

Strangulation is pressure on the neck by the application of force using hands, a ligature, or other firm or unyielding material(s). Classically, strangulation has been subclassified by the mode of neck compression. Manual strangulation is pressure on the neck using hands, and ligature strangulation is neck compression achieved by application of a ligature around the neck. Other forms of strangulation include compression of the neck with an object (e.g., broom handle), appendage (e.g., foot, crux of elbow [sometimes called *mugging*]), or compression of the neck onto an immobile structure. Thus, the term *strangulation* has sometimes been used rather broadly. Furthermore, strangulation is complex and may involve more than one mode of compression. This is most frequently encountered with combined manual and ligature strangulation, when the hands are used to compress the neck and the collar of the victim's shirt is used as a ligature.

In general, the mode of strangulation is determined by the nature of injuries on the skin of the neck. In manual strangulation, the injuries are typically discoid bruises, abrasions, and sometimes small curvilinear (crescent-shaped) abrasions caused by fingernails. In ligature strangulation, there is typically a linear abrasion on the neck representing a ligature mark. Internally, the injuries are highly variable but often include strap muscle injuries. The least frequent classical injury is fracture of the hyoid bone.

One specific variant of strangulation is the neck hold or neck restraint. The context is usually restraint, sometimes by law enforcement officers or medical personnel. In addition, this can be observed in death during participation in martial arts including judo. In these cases, the neck hold is usually applied to restrain rather than kill, but death may occur. The autopsy findings are highly variable ranging from few findings to bilateral fractures of the superior cornua of the thyroid cartilage.

One popular concept associated with strangulation is sudden death associated with vasovagal inhibition or reflex cardiac arrest. It was once thought that compression of baroreceptors (carotid body) or irritation of the vagus nerve could cause sudden cardiac arrest. Although this mechanism of death from neck compression has some anecdotal support, it is probably better viewed as part of the folklore of forensic pathology rather than a robust concept based on scientific evidence.

Diagnostic Pitfalls

The main diagnostic pitfall in cases of neck compression relates to errors of overinterpretation and failure to use proper techniques when dissecting the anterior neck. It has been known for decades that the best approach to the internal examination of the neck is a layered bloodless dissection following venous decompression achieved by removal of the brain and evisceration of the thoracoabdominal organs. This ensures that no "hemorrhages" are created by extravasation of blood from congested blood vessels during soft tissue dissection. However, despite adequate technique, additional pitfalls can be encountered. If the body is in the prone position at the scene and postmortem hypostasis is well established in the neck, hypostatic hemorrhages may form and mimic bruising of strap muscles and soft tissue. Such "pseudobruises" are an important pitfall since the false diagnosis of strangulation can be made if the true nature of the "hemorrhages" are not identified. Another less frequent pitfall is misdiagnosis of the natural (synchondrotic) joints of the hyoid bone as fractures and mistaking triticeous cartilages for fractures of the superior cornua of the thyroid cartilage.

There are several open questions in the pathology of neck compression. These include the differentiation of hypostatic hemorrhages (pseudobruises) from "real" bruises, the combinatorial specificity of the "signs" of neck compression, the search for methods to detect "rapid anoxial deaths" in general, the incidence of hyoid–laryngeal fractures in neck compression, and the specificity of lesser well-known laryngeal findings in cases of neck compression.

Environmental Hypoxia

Oxygen desaturation of the respirable atmosphere may cause rapid anoxial death. This typically occurs by the utilization of oxygen in a closed space without regeneration of the oxygen supply, or displacement of oxygen by another gas. Confinement in a closed space with limited air supply may result in a reduction of oxygen tension due to respiration and replacement by carbon dioxide (environmental hypercapnia). In addition, the respirable atmosphere in a confined space may be saturated by another metabolically inert gas, usually a noble gas (e.g., argon, helium). Such gases have been called *asphyxiant gases*. There are no gross or microscopic postmortem findings that allow for the diagnosis of environmental hypoxia. The diagnosis is made by assessment of the scene and history and exclusionary findings at autopsy.

Chemical Anoxia

Anoxia may occur at the level of intermediary metabolism using oxygen. Thus, a blockage in either oxygen transport or oxygen utilization may cause anoxia at the tissue level. This occurs through chemical agents that poison specific metabolic steps that utilize oxygen. The most common anoxic poison is carbon monoxide. The high-affinity binding of carbon monoxide to hemoglobin (carboxyhemoglobin) results in the failure of oxygen transportation on hemoglobin and the deprivation of oxygen to tissues. Cyanide acts in a similar manner but binds to a cytochrome protein, thereby inhibiting energy production by the electron transport chain in the mitochondrion.

Carboxyhemoglobin imparts a bright cherry red color to blood and tissues at autopsy. Similarly, cyanide can cause pink–red discoloration of postmortem hypostasis. If cyanide is ingested (e.g., hydrocyanic acid), the gastric mucosa may have a bright red color and display acute hemorrhagic gastropathy. Deaths by carbon monoxide poisoning often occur from accidental or suicidal exposure to carbon monoxide as an exhaust gas from fuel combustion (e.g., automobile exhaust), or smoke produced in fires. Cyanide poisoning can occur by ingestion of cyanide usually in the context of homicide or suicide. Occasionally, cyanide poisoning may occur by the inhalation of cyanide in fire gases, since synthetic fibers and plastics may liberate cyanide upon combustion.

Conclusion

Asphyxial or rapid anoxial deaths are due to oxygen deprivation of the brain. The most common mechanisms of rapid anoxial deaths are mechanical, those that interfere with cerebral arterial perfusion, breathing, or airflow through the nose, mouth, or upper aerodigestive tract. Rapid anoxial deaths present technical challenges to the pathologist due to the nonspecificity of autopsy findings and the need to consider circumstantial information. Although some myths have been dispelled (e.g., the postmortem signs of asphyxia), there remain open questions that need to be addressed to provide a firmer evidence-based foundation and a more scientific approach to the postmortem diagnosis of rapid anoxial death.

Further Reading

Adelson, L. (1974). *The Pathology of Homicide*, Charles C. Thomas Publisher, Springfield.

Becroft, D.M. & Lockett, B.K. (1997). Intra-alveolar pulmonary siderophages in sudden infant death: a marker for previous imposed suffocation. *Pathology* **29**(1), 60–63.

Bell, M.D., Rao, V.J., Wetli, C.V. & Rodriguez, R.N. (1992). Positional asphyxiation in adults. A series of 30 cases from the Dade and Broward County Florida Medical Examiner Offices from 1982 to 1990. *American Journal of Forensic Medicine and Pathology* **13**(2), 101–107.

Betz, P. & Eisenmenger, W. (1996). Frequency of throat-skeleton fractures in hanging. *American Journal of Forensic Medicine and Pathology* **17**(3), 191–193.

Bullock, M.J. & Diniz, D. (2000). Suffocation using plastic bags: a retrospective study of suicides in Ontario, Canada. *Journal of Forensic Sciences* **45**(3), 608–613.

Channa Perera, S.D. & Pollanen, M.S. (2007). Sudden death due to sickle cell crisis during law enforcement restraint. *Journal of Forensic and Legal Medicine* **14**(5), 297–300.

DiMaio, V.J. (2000). Homicidal asphyxia. *American Journal of Forensic Medicine and Pathology* **21**(1), 1–4.

Ely, S.F. & Hirsch, C.S. (2000). Asphyxial deaths and petechiae: a review. *Journal of Forensic Sciences* **45**(6), 1274–1277.

Hanzlick, R. (2001). Pulmonary hemorrhage in deceased infants: baseline data for further study of infant mortality. *American Journal of Forensic Medicine and Pathology* **22**(2), 188–192.

Koiwai, E.K. (1987). Deaths allegedly caused by the use of "choke holds" (shime-waza). *Journal of Forensic Sciences* **32**(2), 419–432.

Milroy, C.M. (1999). Munchausen syndrome by proxy and intra-alveolar haemosiderin. *International Journal of Legal Medicine* **112**(5), 309–312.
Nikolic, S., Micic, J., Atanasijevic, T., Djokic, V. & Djonic, D. (2003). Analysis of neck injuries in hanging. *American Journal of Forensic Medicine and Pathology* **24**(2), 179–182.
O'Halloran, R.L. & Frank, J.G. (2000). Asphyxial death during prone restraint revisited: a report of 21 cases. *American Journal of Forensic Medicine and Pathology* **21**(1), 39–52.
Pollanen, M.S. (2001). Subtle fatal manual neck compression. *Medicine Science and the Law* **41**(2), 135–140.
Pollanen, M.S. (2000). A triad of laryngeal hemorrhages in strangulation: a report of eight cases. *Journal of Forensic Sciences* **45**(3), 614–618.
Pollanen, M.S., Channa Perera, S.D. & Clutterbuck, D. Hemorrhagic lividity of the neck: controlled induction of postmortem hypostatic hemorrhages. *American Journal of Forensic Medicine and Pathology*, in press.
Pollanen, M.S., Chiasson, D.A., Cairns, J.T. & Young, J.G. (1998). Unexpected death related to restraint for excited delirium: a retrospective study of deaths in police custody and in the community. *CMAJ* **158**(12), 1603–1607.
Pollanen, M.S. (2005). Deciding the cause of death after autopsy–revisited. *Journal of Clinical Forensic Medicine* **12**(3), 113–121.
Prinsloo, I. & Gordon, I. (1951). Post-mortem dissection artifacts of the neck: their differentiation from ante-mortem bruises. *South African Medical Journal* **25**(21), 358–361.
Rao, V.J. & Wetli, C.V. (1988). The forensic significance of conjunctival petechiae. *American Journal of Forensic Medicine and Pathology* **9**(1), 32–34.
Reay, D.T., Fligner, C.L., Stilwell, A.D. & Arnold, J. (1992). Positional asphyxia during law enforcement transport. *American Journal of Forensic Medicine and Pathology* **13**(2), 90–97.
Sauvageau, A. & Racette, S. (2006). Autoerotic deaths in the literature from 1954 to 2004: a review. *Journal of Forensic Sciences* **51**(1), 140–146.
Takahashi, M., Matsukawa, K., Nakamoto, T., Tsuchimochi, H., Sakaguchi, A., Kawaguchi, K. & Onari, K. (2007). Control of heart rate variability by cardiac parasympathetic nerve activity during voluntary static exercise in humans with tetraplegia. *Journal of Applied Physiology* **103**(5), 1669–1677.
Uzün, I., Büyük, Y. & Gürpinar, K. Suicidal hanging: fatalities in Istanbul retrospective analysis of 761 autopsy cases. *Journal of Forensic and Legal Medicine* **14**(7), 406–409.
Wick, R., Gilbert, J.D. & Byard, R.W. (2006). Café coronary syndrome-fatal choking on food: an autopsy approach. *Journal of Clinical Forensic Medicine* **13**(3), 135–138.
Yukawa, N., Carter, N., Rutty, G. & Green, M.A. (1999). Intra-alveolar haemorrhage in sudden infant death syndrome: a cause for concern? *Journal of Clinical Pathology* **52**(8), 581–587.

MICHAEL S. POLLANEN

Asphyxiophilia *see* Autoerotic Deaths

Assault *see* Child Sexual Abuse Accommodation

Assault: Baby *see* Shaken Baby Syndrome

Assault: Sexual, Drugs *see* Drug-Facilitated Sexual Assault

Assault: Sexually Motivated

Introduction

Sexually motivated assault occurs in all countries and all jurisdictions. The legal, cultural, and societal definitions of sexually motivated assault vary widely. Table 1 shows the current legal definitions of certain types of sexual assault in England and Wales. However, wherever a claim of such an assault is made, it is essential to justice that appropriate medical assessment of both complainant and suspect is undertaken. Such an assessment may identify injuries and the appropriate interpretation of such injuries (or lack of them) is a key to assisting the judicial process. Above all, it must never be forgotten that the initial examination and assessment is only the first part

Table 1 Some definitions of certain types of sexual assault in England and Wales (Sexual Offences Act 2003)

1 **Rape**

(1) A person (A) commits an offence if –
  (a) he intentionally penetrates the vagina, anus or mouth of another person (B) with his penis,
  (b) (B) does not consent to the penetration, and
  (c) (A) does not reasonably believe that (B) consents.

(2) Whether a belief is reasonable is to be determined having regard to all the circumstances, including any steps (A) has taken to ascertain whether (B) consents.

(3) Sections 75 and 76 apply to an offence under this section. [These Sections apply to evidential and presumptive conclusions about consent]

(4) A person guilty of an offence under this section is liable, on conviction on indictment, to imprisonment for life.

Assault

2 **Assault by penetration**

(1) A person (A) commits an offence if –
  (a) he intentionally penetrates the vagina or anus of another person (B) with a part of his body or anything else,
  (b) the penetration is sexual,
  (c) (B) does not consent to the penetration, and
  (d) (A) does not reasonably believe that (B) consents.

(2) Whether a belief is reasonable is to be determined having regard to all the circumstances, including any steps (A) has taken to ascertain whether (B) consents.

(3) Sections 75 and 76 apply to an offence under this section.

(4) A person guilty of an offence under this section is liable, on conviction on indictment, to imprisonment for life.

3 **Sexual assault**

(1) A person (A) commits an offence if –
  (a) he intentionally touches another person (B),
  (b) the touching is sexual,
  (c) (B) does not consent to the touching, and
  (d) (A) does not reasonably believe that (B) consents.

(2) Whether a belief is reasonable is to be determined having regard to all the circumstances, including any steps (A) has taken to ascertain whether (B) consents.

(3) Sections 75 and 76 apply to an offence under this section.

(4) A person guilty of an offence under this section is liable –
  (a) on summary conviction, to imprisonment for a term not exceeding 6 months or a fine not exceeding the statutory maximum or both;
  (b) on conviction on indictment, to imprisonment for a term not exceeding 10 years.

of a process where the examinee requires specialist medical advice and perhaps treatment, after the examination for the alleged assault, and the capability must exist for appropriate medical, psychological, and social support.

In the United Kingdom most adult assessments are undertaken by forensic physicians, and for pediatric patients, joint examinations with a forensic physician and pediatrician are recommended. In other jurisdictions, examinations may be undertaken by other trained healthcare personnel as part of their workload (e.g., sexual assault nurse examiners and forensic pathologists) [1–3]. Such an assessment of both complainant and suspect must be independent, compassionate, impartial, and nonjudgmental. The medical examiners' role is to identify, collect, and

interpret medical information. It is up to the court to apply that evidence to the entirety of the case. Complaints of sexual assault may relate to recent or older events, sometimes from years ago. Allegations of historic abuse may be disclosed many years later. In some cases, therefore, examination of genitalia may have no relevance. The need for an examination in sexual cases must be clear, in order that inappropriate and unnecessary examinations are not undertaken.

Principles of Examination

The principles expressed here are those that are generally applicable worldwide. Some of those principles may be limited by financial considerations, but if these impact on dignity and impartiality such financial considerations should be regarded as professionally and legally unacceptable. It is appropriate that complainants and suspects should be examined by doctors (in the United Kingdom – generally forensic physicians) or healthcare professionals with specific forensic expertise and experience. The appropriate level of these are by no means agreed or adopted throughout the world [2]. The examinee should be able to choose the gender of the examiner. A minority of complainants are male [4]. Examiners of suspects and complainants need to abide by the principles of consent (including for examination, sample taking and production of reports for courts) and confidentiality (with regard to personal medical information that may be unrelated to the allegation). The examiner must be trained in and understand and be familiar with medical care, the retrieval of evidence, and the preservation and interpretation of evidence.

All examiners need to be familiar with the range and frequency of "normal" sexual practice – which may well be dictated by factors including country, culture, sexual orientation, ethnicity, and religion. In one British study, and in another study, 30% of males and 26% of females ($n = 11\,161$) had experienced penile vaginal intercourse before the age of 16 years [5] Other studies have examined incidence of practices in consensual settings and found for fellatio (penis placed in mouth, sexual stimulation by sucking plus ejaculation) an incidence of 55% (Evans) cunnilingus (female genitalia licked, sucked, or rubbed by lips or tongue) 66–72% [6] and anilingus (anus is licked, sucked, or rubbed by lips or tongue) perhaps 15% ($n = 526$) [7]. In nonconsensual settings the following incidences have been considered – fellatio 34% (male–male $n = 104$), and 78% (female–male $n = 1403$) [8], cunnilingus −22% ($n = 1403$) [8], and no data are available for anilingus.

Social factors may also result in altering sexual behavior and in cultures where alcohol and drugs are used widely, risky sexual behavior may be increased in both males and females [9]. Concerns regarding drug-facilitated sexual assault are clear [10]. The evidence shows that alcohol is the drug that is most commonly associated with sexual assault [11–13]. A clear understanding is required of the nature and the purpose of the examination, appropriate documentation of findings, the reasons for sampling, and the current appropriate methods of sampling.

Assessment and Sampling

The assessment of a complainant or suspect of sexual assault is divided into four main parts – the history taking, the clinical examination, sampling, and postassault management. The examiner should ensure that all these aspects are addressed. Subsequently, the examiner is required to produce a report of findings for the court and may then be asked to attend court to be questioned on, their procedures and the interpretation of their findings.

Table 2 summarizes the key considerations for an assessment in allegations of sexual assault.

Table 2 Key considerations for medical assessment – complainant and suspect

Need for immediate medical care (may override forensic needs)
Timing of the assessment (acute or historic allegation)
Location of the assessment (e.g., hospital, victim examination suite)
Medical examination
Introduction (nature and purpose explained)
Consent
History of the allegations (complainant and others)
Medical, gynecological, and sexual history (if relevant)
Drug and alcohol history (if relevant)
General clinical examination (developmental, height, weight, nongenital injuries, and condition of nails)
Anogenital examination
Sampling
Documentation (e.g., photographs and colposcopic images)
Chain of custody
Report writing and interpretation

Table 3 Specific relevant information from history

How was each injury sustained
Object or weapons used (e.g., sex aid, erect penis, fingernails, and bite)
When was the injury sustained
Has injury been treated
Any preexisting illnesses (e.g., skin disease and bleeding diathesis)
Any regular physical activity (e.g., contact sports)
Any regular medication (e.g., anticoagulants, steroids)
Handedness of complainant and suspect
Use of drugs and alcohol
Clothing worn

Documentation of injuries and forensic samples can assist in confirming, or otherwise, the nature of sexual acts alleged, the possible identification of an unknown assailant, and potential links with other offenses. Such information may assist in determining the consensuality or nonconsensuality of any sexual contact.

In taking a history, it is important to establish as much information as possible that allows proper interpretation of marks, injuries, or scars that may be identified on a complainant or suspect. Table 3 identifies key features, which may require determination, dependent on the allegations made.

Characteristics of any injury (anogenital or nongenital) need to be documented. Relevant characteristics that may be important are listed in Table 4. A consistent injury classification system should be used that is both reproducible and understandable to medical and lay individuals [14]. It is also important that examiners understand the limitations of interpretation – for example of bruise coloration [15, 16].

Table 4 Injury characteristics

Location
Pain, tenderness, reduced mobility
Type (e.g., bruise, incised wound, abrasion, and laceration)
Size
Shape
Color
Orientation
Age (if known)
Causation (if known)
Time recorded
Temporary nature of injury (is reexamination recommended after 24 h?)

Once sites of possible contact or injury have been recorded and documented clearly, the forensic swabs from the sites should be taken [17–20]. Standardized modular sampling kits, which contain all appropriate items for sampling should be used if available. These assure standardization and sterility. Forensic samples may include swabs from skin, vagina, anus, rectum, mouth, and elsewhere. Other biological samples of relevance may include hair, nails, blood, urine, and saliva. Disposable proctoscopes and specula should also be retained for forensic information following use.

All samples should be labeled with the date and time taken, the source of the swab (e.g., penile shaft), the examiner and the examinee name. Tamper-evident seals should be used and appropriate storage identified for the type of sample. A clear chain of custody of sample must be established.

It is important to consider taking samples from all parts of the body where contact may have occurred, as recovery of biological samples may assist the investigative process [21]. Types of contact may include (in addition to general injuries such as punches or grip marks) licking, kissing, sucking, biting, ejaculating, and penetrating. These sites of contact may be visible, but some may not be visible and additional techniques have been applied to identify extraneous substances (e.g., semen) on skin, with limited success [22]. Appropriate sampling of such sites can improve positive sample recovery assist in identification by DNA recovery [23, 24].

Findings associated with sexual contact

It is a common, but wrong assumption that sexual assault of any nature results in injury to the victim, whether adult or child [25]. Sexual assault (including rape) may often occur without any visible physical injury – genital or extragenital. Conversely, consensual sexual activity (sexual activity between consenting adults) may result in injury to the body and genitalia. Thus the presence or absence of injuries, in association with allegations of sexual assault do not necessarily indicate by themselves whether the particular activity was consensual or nonconsensual. It is important therefore to understand that medical findings alone cannot imply either consensuality or nonconsensuality but must be taken as part of the overall body of evidence if a case goes to court. The

presence or absence of findings can assist in supporting (or otherwise) accounts of what occurred.

A number of studies in adults have been published, which explore the type of injury seen and the frequency with which the injuries have been noted following clinical examination of complainants of sexual assault. Some of these are discussed here. The method of examination (e.g., the use of a colposcope) may increase the number of "positive" findings [26]. The incidence of nongenital injury is generally higher than the incidence of genital injury [25]. A study of 819 women presenting after sexual assault showed 52% had general body and 20% anogenital trauma. A total of 41% had no injury [27]. One study [28] documented injuries in 249 women, who alleged that they had been assaulted. Of these, 80 had sustained injuries and 169 had not. These injuries when present are shown in Table 5:

The authors of this study concluded that in sexual assault "Most women do not have visible genitoanal injuries. The risk of sustaining genitoanal injury during a sexual assault is higher among women without prior sexual intercourse experience and among women exposed to anal penetration. The severity of the assault is a poor predictor of genitoanal injury."

There are few data of the incidence and type of genital and nongenital injuries of females and males following consensual sexual activity. Consensual sexual activity has been shown to result in the full gamut of blunt trauma injuries and consequently bruises, abrasions, and lacerations to genital and nongenital sites may be present [29]. The severity and site of the injury varies according to many factors including the type of sexual activity, relative sizes of participants, position of participants, previous sexual activity, and intoxication of the participants. The accounts between two participants may differ substantially in the description of the degree of force used and the medical findings may add more weight to one account over the other. The same factors that may influence the degree and severity of injury and the anatomical site of the injury are the same as for nonconsensual activity and thus may include some or all the following – each of which may have differing influences at different times – age of the complainant, type of sexual activity, relative positions of the participants, previous sexual activity, and degree intoxication of either or both of the participants. Consensual insertion or attempts at insertion of a finger or fingers, penis, or any other object into the vagina may result in bruises, abrasions, and lacerations of the labia majora, labia minora, hymen, and posterior fourchette. Consensual digital vaginal penetration may result in accidental fingernail damage or injury to parts of the female genital tract, which may not be noticed by either party. Death has been reported following such injury during sexual contact [30]. The effect of tampon use on the appearance of the hymen has been studied. In 300 females, sexually active subjects (81%) were significantly more likely than tampon users and pad users to have "complete clefts" in the lower hymen between the 2 o'clock and 10 o'clock positions; tampon users were not significantly different from pad users (11% *versus* 5%) [31]; and vaginal injuries may result in bleeding. If penile vaginal penetration follows the digital vaginal penetration then it seems likely that any injury or bleeding may be exacerbated. Cases have been reported where consensual vaginal sexual activity resulted in injury causing bleeding which necessitated internal iliac artery ligation [30, 32, 33].

The incidence of injuries after consensual and nonconsensual sexual intercourse has recently been studied. Anogenital injuries occurring in adolescent females (13–17 years old) after consensual and nonconsensual sexual intercourse [34] were documented in a retrospective, matched case–control study to assess anogenital injuries in female adolescents

Table 5 Injuries after sexual assault[(a)]

Genitoanal injury	*n*	Yes (%)	No (%)
Penetration			
Anal	38	20 (52.6)	18 (47.4)
Vaginal	150	47 (31.3)	103 (68.7)
Attempted vaginal/anal	14	4 (28.6)	10 (71.4)
Digital/pawing of genitals	18	1 (5.6)	17 (94.4)
No recollection	29	8 (27.6)	21 (72.4)
Reported physical violence			
Grievous bodily harm/attempted strangulation	28	9 (32.1)	19 (67.9)
Actual bodily harm	46	11 (23.9)	35 (76.1)
Restrained	104	34 (32.7)	70 (67.3)
None	44	15 (34.1)	29 (65.9)
No recollection	16	5 (31.3)	11 (68.7)
No information	11	6 (54.5)	5 (45.5)

[(a)] Reproduced from Ref. 28 © Elsevier, 2005

presenting to a nurse examiner clinic during a four-year study period. Adolescents were selected for inclusion in the study if they reported consensual sexual intercourse (CSI) and agreed to a medicolegal examination. Control subjects were victims of alleged sexual assault (nonconsensual sexual intercourse (NCSI)) matched to cases by age and prior sexual intercourse experience. Genital trauma was documented using colposcopy with nuclear staining and digital photography. Overall, 49% (25/51) of CSI subjects reported no prior sexual intercourse experience. CSI and NCSI were comparable in terms of race, time to physical examination, alcohol use, and frequency of genital injuries (73% *versus* 85%, $p = 0.069$). The mean number of documented anogenital injuries in CSI subjects was $1.9 + 1.5$. These injuries commonly involved the hymen, fossa navicularis, and posterior fourchette. NCSI subjects had a greater number of anogenital injuries ($2.6 + 2.0$; $p < 0.02$), typically involving the fossa navicularis, labia minora, and hymen. The most common type of injury in both groups was laceration (39% *versus* 41%); however, NCSI subjects had a greater incidence of anogenital abrasion, bruising, and edema ($p = 0.035$). The authors concluded "Anogenital trauma was documented in 73% of adolescent females after consensual sexual intercourse *versus* 85% of victims of sexual assault. The localized pattern and severity of anogenital injuries were significantly different when compared with victims of sexual assault, the presence of anogenital trauma suggests that penetration has occurred and implies nothing about consent".

Another study [35] reviewed adolescent complainants of sexual assault, comparing virgin and nonvirgin groups. About 224 patients with a mean age of 14.8 years were studied of which 81 were "virgins" and 97 had been sexually active prior to the assault. The virgin group took longer to present for examination than the nonvirgin group (90 h compared to 44 h). Of all clients, 51% had a nongenital injury. These tended to be minor. In the nonvirgin group 32% had a genital injury. In the virgin group, 53% had a genital injury, however only 32% had the type of genital injury that would leave permanent evidence of penetration (i.e., if examined more than several weeks later). Alcohol use prior to assault was common. The authors concluded "Genital and or body injuries are not routinely found in adolescents after an allegation of rape or sexual assault even when there has not been previous sexual experience. The absence of injury does not exclude the possibility of intercourse, whether with or without consent."

If children alone are considered, sexual abuse is rarely diagnosed on the basis of only physical examination or laboratory findings [36]. Even when the perpetrator admits to penetration of the child's genitalia, physical findings may be absent on examination [37–41]. Abuse may leave no physical evidence, and mucosal (skin or anal/rectal, vaginal surfaces) injuries often heal rapidly and completely [42–45]. Examination of the healing of hymenal injuries in prepubertal and pubertal girls shows that hymenal injuries in both groups healed rapidly and, with the exception of more extensive lacerations, left no evidence of previous injury [46]. Heger and colleagues have concluded from a review of over 2000 children that "research indicates that medical, social, and legal professionals have relied too heavily on the medical examination in diagnosing child sexual abuse. History from the child remains the single most important diagnostic feature in coming to the conclusion that a child has been sexually abused. Only 4% of all children referred for medical evaluation of sexual abuse have abnormal examinations at the time of evaluation. Even with a history of severe abuse such as vaginal or anal penetration, the rate of abnormal medical findings is only 5.5%" [47].

Anal intercourse is part of the normal sexual repertoire of many heterosexual and homosexual couples. Consensual anal intercourse may (in the same way as vaginal sex) be pain free and discomfort free and would not normally leave any residual injury. Repeated anal intercourse results in easier penetration over a period of time. Lubrication such as saliva or KY Jelly can be used to ease penetration.

Nonconsensual anal intercourse if done without force, with or without lubrication, and without physical resistance on the part of the person being penetrated may leave no residual injury and may be pain free. The effects of drugs and alcohol may make penetration easier [48]. In an individual who is otherwise used to anal intercourse no pain or discomfort may be experienced.

The likelihood of pain, or injury in nonconsensual anal intercourse may be increased (i) in someone who has not experienced anal intercourse, (ii) in the absence of lubrication, (iii) if force is used, and (iv) if there is great disparity between the size of the anus (which varies little in the adult) and the penis (which may vary a lot).

The types of injury that can be caused by forced anal intercourse are stretch injuries that damage the lining of the anal canal or the skin surrounding the anus (the perianal region). The stretching can cause damage to the blood vessels supplying the local tissues – and such damage to blood vessels can cause blood to leak out into the tissues resulting in bruises – or can tear the surface of the anal canal and the perianal region causing fissures, tears, or lacerations. Because of the nerve supply, such fissures, tears, or lacerations (the terms are used interchangeably), even if minor, can be exquisitely painful. The incidence of fissures in nonconsensual anal intercourse is uncertain as there are only a few studies that have explored these issues, and there is some overlap of definition of terms such as fissure, tear, or laceration. One study [49] found fissures in 16% of cases after anal intercourse. Another study found tears in 19 of 55 patients reporting anal contact. In both these studies patients were generally examined within 72 h of the incident. The incidence of bruising noted is considerably less and is reported as occurring in between 2 and 4% of cases. In no cases were major sphincter disruptions observed. Regarding repeated acts of anal penetration, study results conflict. A study of 129 heterosexual women who gave histories of anal intercourse showed no report of gross fecal incontinence [50] while another [51] found that although the anoreceptive males had lower resting anal canal pressures, there were no complaints of fecal incontinence; however, another study [52] found a significant increase in fecal incontinence or urgency. The term anal fissure refers to breaks or a tear in the skin around the anus. Fissures may be acute (i.e., developing and healing within a couple of weeks) caused by some form of trauma, or chronic (i.e., persistent and not healing) perhaps being initiated by trauma and prolonged by certain diseases or illnesses. Chronic constipation, requiring straining to defecate, associated with the passage of hard stool, is one cause of acute fissure development and defecation has been noted to cause a superficial abrasion of the anal verge [19]. Other causes of acute and chronic fissure include sexually transmitted disease, diarrhea, inflammatory bowel disease (e.g., Crohn's disease), and skin diseases. Another cause of the development of anal fissures is the passage of objects into the anal passage for sexual purposes, this may include items such as dildos or a penis. In the absence of repeated trauma, any fissures, tears, or lacerations would be expected to heal within two weeks or so and leave no residual marks.

Management Postassault

It is important that medical needs are appropriately attended to after an assault. The follow-up of a complainant may be essential, particularly when considering potential infectious complications (e.g., sexually transmitted infections (STI), hepatitis, and HIV). The usual ethical principles of consent and confidentiality apply, but these may be confounded, at least temporarily by other factors, such as drugs or alcohol or mental health considerations. The forensic physician must weigh up the priorities regarding medical care and forensic sampling and determine which, from the examinee's viewpoint is in their best interests. Treatment may involve screening for infection either in the community or in hospital, emergency contraception, and prophylaxis against STI, which may be a significant risk. Often, particularly in the case of HIV, specialist advice may need to be sought. It is essential that as much relevant information is supplied to the treating doctor as possible, so that an informed approach can be made [53]. Additionally, facilities should be in place for future medical, social, and psychological support.

Conclusions

Complaints of sexual assault are common. A forensic physician (or other appropriate healthcare professional) has a duty to undertake a full assessment in an appropriate nonjudgmental, impartial, and appropriate manner; false complaints do occur. Many investigating officers from police or other agencies may not be conversant with the needs and limits of medical forensic assessments.

The principles for examination of all complainants (and suspects) of sexual assault are clear. They require an understanding of medicine, anatomy, physiology, and science. The forensic physician must ensure that evidential findings are not over- or misinterpreted. The health of the individual being examined is paramount, forensic issues must be considered secondary to the well-being of the examinee. The examinee's best interests are served by a full and detailed history and examination with appropriate sample taking and clear and accurate documentation

of findings. The examiner should remember basic medical principles and ensure that the information they provide to the court is accurate and relevant and unbiased.

References

[1] Celbis, O., Gokdogan, M.R., Kaya, M. & Gunes, G. (2006). Review of forensic assessments of female referrals to the branch of legal medicine, Malatya region, Turkey – 1996–2000, *Journal of Clinical Forensic Medicine* **13**, 21–25.

[2] De Munnynck, K., De Houwer, L., Bronselar, K., Hanssens, M. & Van de Voorde, W. (2006). Medicolegal approach to sexual assault victims: the Belgian situation, *Journal of Clinical Forensic Medicine* **13**, 211–214.

[3] Hassan, Q., Bashir, M.Z., Mujahid, M., Munawar, A.Z., Aslam, M. & Marri MZ. (2007). Medicolegal assessment of sexual assault victims in Lahore, *Journal of the Pakistan Medical Association* **57**, 539–542.

[4] Thompson, C. (2006). Review of 212 individuals attending a city center genitourinary medicine clinic following acute sexual assault, *Journal of Clinical Forensic Medicine* **13**, 1861–1888.

[5] Wellings, K., Nanchahai, K., Macdowall, W., McManus S, Erens, B., Mercer, C.H., Johnson, A.M., Copas, A.J., Korovessis, C., Fenton, K.A. & Field, J. (2001). Sexual behaviour in Britain: early heterosexual experience, *Lancet* **358**, 1843–1850.

[6] Johnson, A.M., Wadsworth, J., Wellings, K. & Field, J. (1994). Heterosexual practices, *Sexual Attitudes and Lifestyles*, Blackwell Scientific Publications.

[7] Bolling, D.R. (1977). Prevalence, goals and complications of heterosexual anal intercourse in a gynecologic population, *The Journal of Reproductive Medicine* **19**, 120–124.

[8] Keating, S.M. & Higgs, D.F. (1992). Oral sex – further information from sexual assault cases, *Journal of Forensic Science Society* **32**, 327–331.

[9] Cashell-Smith, M.L., Connor, J.L. & Kypri, K. (2007). Harmful effects of alcohol on sexual behaviour in a New Zealand university community, *Drug and Alcohol Review* **26**, 645–651.

[10] Dorandeu, A.H., Pages, C.A., Sordino, M.-C., Pépin, G., Baccino, E. & Kintz, P. (2006). A case in South Eastern France: a review of drug facilitated sexual assault in European and English-speaking countries, *Journal of Clinical Forensic Medicine* **12**, 253–261.

[11] Scott-Ham, M. & Burton, F. (2005). Toxicological findings in cases of alleged drugfacilitated sexual assault in the United Kingdom over a 3-year period, *Journal of Clinical Forensic Medicine* **12**, 175–186.

[12] Scott-Ham, M. & Burton, F. (2006). A study of blood and urine alcohol concentration in cases of alleged drug-facilitated sexual assault in the United Kingdom over a 3-year period, *Journal of Clinical Forensic Medicine* **13**, 107–111.

[13] Hurley, M., Parker, H. & Wells, D.L. (206). The epidemiology of drug-facilitated sexual assault, *Journal of Clinical Forensic Medicine* **12**, 181–185.

[14] Payne-James, J.J., Crane, J. & Hinchliffe, J. (2005). Injury assessment, documentation and interpretation, in *Clinical Forensic Medicine. A Physician's Guide*, M.M. Stark, ed, Humana Press.

[15] Langlois, N.E.I. & Gresham, G.A. (1991). The aging of bruises: a review and study of the colour changes with time, *Forensic Science International* **50**, 227–238.

[16] Munang, L.A., Leonard, P.A. & Mok, J.Y.Q. (2002). Lack of agreement on colour description between clinicians examining childhood bruising, *Journal of Clinical Forensic Medicine* **9**, 171–174.

[17] Allard, J.E. (1997). The collection of data from findings in cases of sexual assault and the significance of spermatozoa on vaginal, anal and oral swabs, *Science & Justice* **37**, 99–108.

[18] Girardin, B., Faugno, D.K. & Howitt, J. (2003). Adult sexual assault: practical management, in *Forensic Medicine: Clinical and Pathological Aspects*, J.J. Payne-James, A. Busuttil & W. Smock, eds, Greenwich Medical Media.

[19] Rogers, D. & Newton, M. (2005). Sexual assault examination, in *Clinical Forensic Medicine. A Physician's Guide*, 2nd Edition, M.M. Stark, ed, Humana Press.

[20] Nadesan, K. (2005). Evidential sample collection, in *Encyclopedia of Forensic & Legal Medicine*, J.J. Payne-James, R.J. Byard, T. Corey & C. Henderson, eds, Elsevier.

[21] Rutty, G. (2001). An investigation into the transference and survivability of human DNA following simulated manual strangulation with consideration of the problem of 3rd part contamination, *International Journal of Legal Medicine* **116**, 170–173.

[22] Wawryk, J. & Odell, M. (2005). Fluorescent identification of biological and other stains on skin by the use of alternative light sources, *Journal of Clinical Forensic Medicine* **12**, 296–301.

[23] Sweet, D., Lorente, M., Lorente, J.A., Valenzuela, A. & Villanueva, E. (1997). An improved method to recover saliva from human skin: the double swab technique, *Journal of Forensic Science Society* **42**, 320–322.

[24] Sweet, D., Lorente, J.A., Valenzuela, A., Lorente, M. & Villanueva, E. (1997). PCR-based DNA typing of saliva stains recovered from human skin. *Journal of Forensic Science Society* **42**, 447–451.

[25] Payne-James, J.J. & Bassindale, C. Sexual Offences: injuries and findings after sexual contact, In: *Encyclopedia of Forensic & Legal Medicine*, Payne-James, J.J., Corey, T., Henderson, C., & Byard, R.W., eds, Elsevier, 2005.

[26] Adams, J., Girardin & D., Faugno, D. (2001). Adolescent sexual assault: documentation of acute injuries using photocolposcopy, *Journal of Pediatric and Adolescent Gynecology* **14**, 175–180.

[27] Sugar, N.F., Fine, D.N. & Eckert, L.O. (2004). Physical injury after sexual assault: findings of a large case series,

American Journal of Obstetrics and Gynecology **190**, 71–76.

[28] Hilden, M., Schei, B. & Sidenius, K. (2005). Genitoanal injury in adult female victims of sexual assault, *Forensic Science International* **154**, 200–205.

[29] Payne-James, J.J. & Bassindale, C. (2005). Injuries and findings after sexual contact, in *Encyclopedia of Forensic and Legal Medicine*, J.J. Payne-James, R. Byard, T. Corey & C. Henderson, eds, Elsevier.

[30] Sadler, D.W. & Pounder, D.J. (1998). Fatal air embolism during consensual intercourse in a non-pregnant female, *Journal of Clinical Forensic Medicine* **5**, 77–79.

[31] Emans, S.J., Woods, E.R., Allred, E.N. & Grace, E. (1994). Hymenal findings in adolescent women: impact of tampon use and consensual sexual activity, *Journal of Pediatrics* **125**, 153–160.

[32] Sivalingham, N. & Rajesvaran, D. (1996). Coital injury requiring internal iliac artery ligation, *Singapore Medical Journal* **37**, 547–548.

[33] McGolgin, W.S., Williams, L.M., Sorrells, T.L. & Morrison, J.C. (1990). Hemoperitoneum as a result of coital injury without associated vaginal injury, *American Journal of Obstetrics and Gynaecology* **163**, 1503–1505.

[34] Jones, J. & Rossman, R. (2003). Anogenital injuries in adolescents after consensual sexual intercourse, *Academic Emergency Medicine* **10**, 1378–1383.

[35] White, C. & McLean, I. (2006). Adolescent complainants of sexual assault; injury patterns in virgin and non-virgin groups, *Journal of Clinical Forensic Medicine* **13**, 172–180.

[36] Kellogg, N., The Committee on Child Abuse and Neglect (2005). The evaluation of sexual abuse in children, *Pediatrics* **116**, 506–512.

[37] Muram, D., Child sexual abuse–genital tract findings in prepubertal girls. I. (1989). The unaided medical examination, *American Journal of Obstetrics and Gynecology* **160**(2), 328–333.

[38] Kerns, D.L. & Ritter, M.L. (1992). Medical findings in child sexual abuse cases with perpetrator confessions [abstract], *American Journal of Diseases of Children* **146**, 494.

[39] Heger, A., e Ticson, L., Guerra, L., Lister, J., Zaragoza, T., McConnell, G., Morahan, M. (2002). Children referred for possible sexual abuse: medical findings in 2384 children, *Child Abuse and Neglect* **26**, 645–659.

[40] Adams (1994).

[41] Finkel, M.A. (1989). Anogenital trauma in sexually abused children, *Pediatrics* **84**, 317–322.

[42] McCann, J., Voris, J. & Simon, M. (1992). Genital injuries resulting from sexual abuse: a longitudinal study, *Pediatric* **89**, 307–317.

[43] McCann, J. & Voris, J. (1993). Perianal injuries resulting from sexual abuse: a longitudinal study, *Pediatrics* **91**, 390–397.

[44] Heppenstall-Heger, A., McConnell, G., Ticson, L., Guerra, L., Lister, J. & Zaragoza, T. (2003). Healing patterns in anogenital injuries: a longitudinal study of injuries associated with sexual abuse, accidental injuries, or genital surgery in the preadolescent child, *Pediatrics* **112**, 829–837.

[45] Royal College of Paediatrics and Child Health (2008). *The Physical Signs of Child Sexual Abuse: An Evidence-Based Review and Guidance for Best Practice*, Royal College of Paediatrics and Child Health/Faculty of Forensic & Legal Medicine, Royal College of Physicians, London.

[46] McCann, J., Miyamoto, S., Boyle, C. & Rogers, K. (2007). Healing of hymenal injuries in pre-pubertal and adolescent girls: a descriptive study, *Pediatrics* **119**, 1096–1106.

[47] Heger, A., Ticson, L., Velasquez, O. & Bernier, R. (2002). Children referred for possible sexual abuse: medical findings in 2384 children, *Child Abuse and Neglect* **26**, 645–659.

[48] Norfolk, G. (2005). Accidental anal intercourse: does it really happen? *Journal of Clinical Forensic Medicine* **12**, 1–4.

[49] Manser, T. (1991). Findings in medical examinations of victims and offenders in cases of serious sexual offenses – a survey, *Police Surgeon* **38**, 4–27.

[50] Evans, B.A., Bond, R.A. & Macrae, K.D. (1998). Sexual behaviour in women attending a genitourinary medicine clinic, *Genitourinary Medicine* **64**, 43–48.

[51] Chun, A.B., Rose, S., Mitrani, C., Silvestre, A.J. & Wald, A. (1997). Anal sphincter structure and function in homosexual males engaging in anoreceptive intercourse, *The American Journal of Gastroenterology* **92**, 465–468.

[52] Miles, A.J.G., Allen-Mersh, T.G. & Wastell, C. (1993). Effect of anoreceptive intercourse on anorectal function, *Journal of the Royal Society of Medicine* **86**, 144–147.

[53] Jawad, R. & Welch, J. (2005). Management postassault, in *Encyclopedia of Forensic & Legal Medicine*, J.J. Payne-James, R. Byard, T. Corey & C. Henderson, eds, Elsevier.

Further Reading

Exline, D.L., Smith, F.P. & Drexler, S.G. (1998). Frequency of pubic hair transfer during sexual intercourse, *Journal of Forensic Sciences* **43**, 505–508.

Lincoln, C.A., McBride, P.M., Turbett, G.R., Garbin, C.D. & MacDonald, E.J. (2006). The use of an alternative light sources to detect semen in clinical forensic medical practice, *Journal of Clinical Forensic Medicine* **13**, 215–218.

McCann, J., Miyamoto, S., Boyle, C. & Rogers, K. (2007). Healing of hymenal injuries in prepubertal and adolescent girls: a descriptive study, *Pediatrics* **119**, 1094–1106.

Slaughter, L., Brown, C.R.V, Crowley, S. & Peck, R. Patterns of genital injury in female sexual assault victims, *American Journal of Obstetrics and Gynecology* **176**, 609–616.

Stephenson, T. & Bialas, Y. (1996). Estimation of the age of bruising, *Archives of Disease in Childhood* **74**, 53–55.

JASON PAYNE-JAMES

Assessment: Capacity ***see*** Capacity Assessment

Assessment of Ability to Live Independently ***see*** Capacity for Independent Living

Assessment of Independent Living Capacity ***see*** Capacity for Independent Living

Auditory Hallucinations ***see*** Hallucinations

AURA ***see*** Seizures: Behavioral

Authentication of Documentary Evidence ***see*** Documents: Authentication of

Autoerotic Asphyxiation ***see*** Autoerotic Deaths

Autoerotic Deaths

Introduction

Deaths associated with sexual activity have been classified by homicide detectives as "death by sexual misadventure" [1], and by the research literature as "sexual fatalities" [2]. The term *sexual fatality* refers to *a broad array of deaths* that involve a sexual element (see Table 1).

This article focuses on sexual fatalities known as *autoerotic deaths*, which often require a postmortem behavioral analysis or psychological autopsy to assist the medical examiner. Autoerotic deaths are unanticipated deaths that occur while the victim is engaged in solitary sexual activity [3], most commonly masturbation [4]. Autoerotic deaths are most frequently caused by a failure of the "escape" mechanism the victim designed to stop his cerebral hypoxia immediately before unconsciousness.

The term *asphyxiophilia* has been used to describe sexual arousal to the state of oxygen deprivation [5]. It is not specifically listed in the Diagnostic and Statistical Manual of Mental Disorders, 4th ed.–Text Revision (DSM-IV-TR) under paraphilias; however, it could be classified as a "Paraphilia Not Otherwise Specified" [6]. Since this type of death is often associated with ligature strangulation and other forms of self-harm, there is the potential for the medical examiner to misidentify autoerotic deaths as suicides and homicides.

The medical examiner's goal is to classify deaths into one of the four categories or "modes": natural, accidental, suicide, or homicide (*NASH*) [7]. When a death cannot be immediately classified, it is often officially referred to as *undetermined*. Deaths that are associated with some element of sexual activity can occasionally present a confusing picture, resulting in a finding of an undetermined or "equivocal" death. An "equivocal death" is one in which the

Table 1 Examples of sexual fatalities [2]

• Myocardial infarction or stroke during coitus or masturbation
• Autoerotic asphyxiation
• Accidental asphyxiation by a partner
• Masochistic sexual practices involving electrocution or bondage

manner of death is questionable, or the circumstances surrounding the death are otherwise unclear [8]. Sexual fatalities in which the decedent was also using substances may be particularly difficult to classify [9]. Research has strongly suggested that the majority of autoerotic deaths are due to AEA [3].

Autoerotic Asphyxiation

AEA is the practice of self-induced strangulation, ostensibly for the purpose of enhancing the pleasure of masturbating [5, 10]. It has been reported that the hypoxia induced by the reduced cerebral blood flow may produce an altered or enhanced sexual experience [11]. Of critical importance is that the practitioner's intent is sexual pleasure, not suicide. Cases combining suicide with AEA have been reported, but are believed to be quite rare [2]. It has been estimated that AEA is responsible for approximately 500–1000 deaths annually in the United States [12, 13]. However, such estimates may be conservative, as cases of fatal AEA are subject to underreporting and misidentification. The majority of victims are Caucasian males [14].

Most deaths resulting from AEA are due to an accidental, unintended failure of a release mechanism in the apparatus designed to produce cerebral hypoxia [3, 15]. The physiological mechanisms observed in fatal AEA can be categorized into four basic processes (see Table 2).

Bilateral neck compression can quickly produce unconsciousness via the carotid sinus reflex. The carotid sinus reflex is triggered by continuous pressure on the carotid artery in the area of the carotid bifurcation. Stimulation of the carotid sinus reflex induces vascular hypotension and a reduced heart rate, which significantly reduces cerebral blood flow. Airway obstruction may result from gags, inhaled vomitus, plastic bags drawn tightly over the victim's head, or other forms of suffocation. AEA victims may also bind themselves tightly enough around the torso that chest expansion is greatly reduced, resulting in fatal hypoxia. Finally, inhaled chemicals such as amyl nitrite may displace or exclude the availability of oxygen in the lungs.

Table 2 Physiological mechanisms of fatal autoerotic asphyxiation [3, 15]

1. Neck compression
2. Airway obstruction
3. Chest compression
4. Oxygen deprivation

Table 3 Pathological findings in AEA [3]

- Petechial hemorrhages of conjunctiva, sclera, and face
- Pulmonary congestion
- Congestion of viscera
- Occasional thyroid cartilage fracture

On autopsy, several pathological findings are commonly found (see Table 3) [3]. These include petechial hemorrhages of the conjunctiva, sclera, and face. The viscera is often diffusely congested, and pulmonary congestion is common. There may occasionally be a fracture of the thyroid cartilage, depending upon the nature and placement of the neck ligature. However, fractured thyroid cartilage is more often associated with manual strangulation. The presence of seminal discharge has been described as unhelpful in distinguishing fatal AEA from other modes of death [4].

Criteria have been proposed to assist in the determination of the AEA death (see Table 4) [16]. There must be a physiologic mechanism present which the victim had used for sexual arousal. Evidence of solitary sexual activity and sexual fantasy aids should be present, such as pornography or sexually oriented props. There will often be evidence of prior autoerotic practice, either from findings at the scene, or from collateral interviews. Finally, there should be an absence of suicidal intent, which may require a comprehensive psychological autopsy (*see* **Psychological Autopsy**) [7, 8].

Other death scene findings have been described as pathognomonic for fatal AEA (see Table 5) [13]. The victim will often be either partially supported by the ground, or will otherwise have some structure nearby to stand on and alleviate neck compression. Neck ligatures, the most common method

Table 4 Proposed criteria for autoerotic death [16]

1. Physiologic mechanism for obtaining sexual arousal
2. Solitary sexual activity
3. Sexual fantasy aids (pornography, sex toys, and mirrors)
4. Prior autoerotic practice
5. Absence of suicidal intent

Table 5 Common findings in fatal AEA [13]

- Body partially supported by ground
- Ligature with self-rescue mechanism (slip knot, etc.)
- Bondage items and/or sexual masochistic behavior (genitals, nipples, etc.)
- Male wearing female attire
- Protective padding between ligature and body

for AEA, will have been designed by the victim to have some type of self-rescue mechanism. This may be a slip knot or an interconnection between ligatures.

Male victims of fatal AEA may be found wearing various pieces of female attire (cross-dressing), or bondage type attire such as leather, chains, and rubberized garments. Victims who practiced masochistic type sexual behaviors may show evidence of prior self-mutilation. Examination of the ligature often reveals the presence of protective padding between the ligature and the neck. This is done by the AEA practitioner to prevent tell tale neck abrasions and bruising.

AEA in Women

The extant research literature suggests that women victims of fatal AEA are far less frequent than men [17]. Thus, the body of research providing information about female autoerotic deaths is relatively small. It is believed that fatal AEA in women may have a less obvious presentation than in men [18]. This may be due to the fact that the majority of women in early studies did not employ sexual props, special clothing, or extra ligatures. AEA in women may be distinguished simply by evidence of solitary sexual activity, neck compression with a self-rescue mechanism, padding around the neck, and the absence of indicators of suicide or depression.

Women are more likely than men to employ only a simple ligature around the neck, which is controlled by body position movement [13]. Other methods described in cases of fatal AEA in women include suffocation by plastic bag over the head and inhalation of ether [19]. A case involving foreign body insertion has also been described [16]. Most of the research supports the finding that women may be less inclined than men to use pornography or fetishistic items during AEA. For this reason, the female autoerotic death scene may pose a greater challenge in distinguishing accidental death due to sexual misadventure from suicide and homicide.

Atypical Autoerotic Asphyxiation

While ligature strangulation about the neck is the most frequently reported method in fatal AEA, atypical cases may be encountered. Such cases may involve asphyxia due to inhalants, plastic bags, or suffocation due to a variety of forms of external compression of the thorax. Atypical cases of AEA have been defined on the basis of (i) the nature of the autoerotic activity, (ii) the cause of death, and (iii) the relationship between the two [20].

Atypical AEA leading to death is believed to account for about 10.3% of cases in the literature [14]. Atypical AEA fatalities can be classified into five broad categories: electrocution (3.7%), binding/overdressing/body wrapping (1.5%), foreign body insertion (1.2%), atypical asphyxia method (2.9%), and miscellaneous (1.0%) [14]. Table 6 lists some of the most commonly reported mechanisms of atypical AEA deaths.

Some cases may be atypical even when evidence of ligature strangulation is present. For example, the hypoxia induced may result in a fall and fatal head injury. Other causes may include aspiration, cardiovascular disease, and syncope due to inhalant use. Acute cerebral hypoxia may be produced by the inhalation of amyl and butyl nitrite, known colloquially as *poppers*. The use of other inhaled substances, such as N_2O, ether, and chloroform has also been described [4, 22].

The use of electrocution has been found to be the cause of death in some cases. Typically, the victim will have constructed a wiring device that draws on household current. The mechanism of death

Table 6 Mechanisms of death in atypical autoerotic fatalities [14, 21]

- Cardiovascular
- Electrocution
- Inhalants
- Suffocation from “binding”
- Hyperthermia
- Sepsis due to foreign body
- Hemorrhage

in such cases is often attributed to fatal cardiac arrhythmia, although respiratory arrest is also a possibility [23]. Cases have been reported involving electrocution of the genitals, nipples, and anal cavity [4, 15].

Binding

The sexual fetish known as *binding* or *body wrapping*, involves tightly wrapping the body in various materials such as leather, rubber, or plastic wrap. It is a relatively rare form of atypical AEA death, accounting for approximately 1.5% of published cases [24]. Besides suffocation due to thoracic compression, binding may lead to fatal hyperthermia when the body can no longer effectively thermoregulate with sweating because of the tightly-wrapped binding materials.

Aquaeroticum

Another rarely reported autoerotic phenomenon, aquaeroticum, involves asphyxia from submersion under water. Typically, the victim will engage in masturbation while submerged, either with or without bindings or other attire. For example, one reported case of a 25-year-old man described his homemade plastic body suit and complex bondage [25]. The victim was found submersed, tied underwater to a boat, and had been using a homemade diving apparatus for air supply. It was determined that his death resulted from a faulty air-supply device.

Breath Control Play

Breath control play is a term used by it's practitioners to describe the sexual practice of asphyxiation by strangulation or suffocation of one person by another during sexual intercourse [10]. Breath control play may be done in conjunction with other fetish activities. Cases of unintended homicide due to mutually consenting breath control play have been described [15]. Such cases have the potential to be misidentified as either a simple AEA fatality or as a sexual murder.

Asphyxial Games in Juveniles

In cases of autoasphyxiation in children and adolescents, the phenomenon of asphyxial games should be considered. Asphyxial games, while not new to the juvenile population, may be increasing in overall lethality. It has been hypothesized that this increased lethality may be due to the increasing use of ligature, playing the "game" alone, and a societal trend of "extreme" sports and activities among youth [26]. Some slang terms used by juveniles to describe asphyxial games are listed in Table 7. The goal of juvenile asphyxial games is to produce a euphoric hypoxia. Thus, asphyxial games in juveniles do not ordinarily have a sexual theme or purpose.

Table 7 Slang terms for asphyxial games [27]

- Space monkey
- Space cowboy
- The choking game
- Black hole
- Black out
- Flatlining
- Funky chicken
- Gasp
- Knock out
- Rising sun
- Tingling
- Suffocation roulette

The limited research available on juvenile asphyxial games suggests that the typical age range is 9–15 years, and the male to female ratio is approximately 2 : 1 [26]. Two general typologies of asphyxial game playing in juveniles have been proposed. One group is described as athletic, action-oriented, and being average to above average students. This group may turn to asphyxial games due to their thrill-seeking nature. The second group is described as a "subset of younger adolescents with anxiety and/or depression who, having a desire for drugs and alcohol... but limited access to such agents, turn to this kind of behavior as a substitute" [26].

Postmortem Behavioral Analysis and Psychological Autopsy

Many cases of autoerotic fatality require a comprehensive investigation consisting of (i) a thorough death scene investigation, (ii) a complete autopsy, (iii) toxicology, (iv) collateral interviews, (v) behavioral analysis, and (vi) a psychological

Table 8 Postmortem behavioral analysis

- Examination of neck
- Examination of ligatures, bindings, and restraints
- Careful analysis of release mechanism
- Decedent's attire
- Presence of sexual aids
- Evidence of prior autoerotic behavior
- Visual inspection of death scene
- Decedent's possessions
- Decedent's journals, emails, and writings
- Forensic analysis of decedent's personal computer
- Review of medical examiner's findings
- Toxicologic analyses
- Review of any relevant medical or psychiatric records
- Collateral interviews of decedent's family and social contacts

autopsy to determine the presence of suicidal intent [2–4, 7] (*see* **Psychological Autopsy**). The postmortem behavioral analysis requires gathering the decedent historical information from collateral sources, and analyzing this data in conjunction with the physical evidence from the death scene [2].

It is important to note that areas of the decedent's dwelling other than the death scene may have to be searched for the purpose of analyzing the decedent's possessions. Evidence of interest in and collection of autoerotic materials and pornography may be found in areas that are not readily visualized. Table 8 gives a list of important elements of the postmortem behavioral analysis.

Forensic computer analysis is a very helpful and often necessary part of the postmortem behavioral analysis. The victim's emails and Internet activity may reflect an interest in autoerotic activity. Websites explaining and promoting dangerous autoerotic activity currently abound on the Internet (e.g., http://www.zeroair.com) [27]. Over the past several years, there have been increasing reports of autoerotic fatalities that were directly associated with Internet use [28, 29].

During the behavioral reconstructive phase, it may be necessary for investigators to recreate complex bindings, knots, ligatures, or body positions in an effort to see if the decedent was physically capable of creating the mechanisms discovered at the death scene. Significant challenges to the investigation may arise. For example, cases of family members altering the death scene out of shame or impulse have been described [4]. In addition, cases in which suicide notes were actually part of a sexual fantasy may result in a mistaken conclusion of suicide [13].

Another challenge may involve decedents who are found with both hands bound. Such cases may easily be mistaken as "suspicious" for a homicide [30]. However, a thorough postmortem behavioral analysis which includes a reconstruction of the victim's position and bindings may resolve the case. A thorough postmortem behavioral analysis should focus on the presence or absence of certain key features of the death scene (see Table 9).

Investigators should attempt to find evidence of prior AEA experience in the deceased [3, 13] In addition to collateral interviews of the victim's social

Table 9 Key features of the autoerotic death scene [3, 4, 13]

Location	• Secluded, isolated • Bedroom, basement, and garage • Locked room
Victim position	• Partially supported by ground • Occasionally suspended • Suspension point within victim's reach • Genital area exposed • Victim's hand on or near genitals
Injurious agent	• Ligature • Electric current • Gag • Inhaled chemicals • Restrictive device or container
Self-Rescue mechanism	• Slip knot • Ability to stand erect to relieve pressure • Connection between ligatures • Knife
Bondage	• Ropes, chains • Handcuffs • Bondage of genitalia • Miscellaneous bindings
Masochism	• Injury inflicted on genitalia • Self-induced burns • Electrical wires • Insertion of sharp objects • Miscellaneous self-mutilation
Attire	• Nude or partially nude • Female clothing • Leather or rubber materials
Protective padding	• Between ligature and body surface
Sexual paraphernalia	• Pornographic literature, videos • Fetish items, female undergarments • Sexual devices (vibrators, dildos)

contacts, investigators should search for suspension point indentations or marks, suggesting prior use of ligature. Permanently affixed protective padding suggests prior experience with AEA. It has been reasoned that increasing complexity of the asphyxial mechanism, bindings, and sexual props suggests that the AEA practices have become more elaborate over time [4].

References

[1] Walton, R. (2006). *Cold Case Homicides: Practical Investigative Techniques*, CRC Press, Boca Raton.

[2] Hazelwood, R., Dietz, P. & Burgess, A. (1982). Sexual fatalities: behavioral reconstruction in equivocal cases, *Journal of Forensic Sciences* **27**(4), 763–773.

[3] Shields, L., Hunsaker, D. & Hunsaker, J. (2005). Autoerotic asphyxia: part I, *The American Journal of Forensic Medicine and Pathology* **26**(1), 45–52.

[4] Hazelwood, R., Dietz, P. & Burgess, A. (1981). The investigation of autoerotic fatalities, *Journal of Police Science and Administration* **9**(4), 2.1–2.10.

[5] Jenkins, A. (2000). When self-pleasuring becomes self-destruction: autoerotic asphyxiation paraphilia, *International Electronic Journal of Health Education* **3**(3), 208–216.

[6] American Psychiatric Association (2000). *Diagnostic And Statistical Manual of Mental Disorders*, 4th Edition, Text Revision, APA, p. 576.

[7] Scott, C., Swartz, E. & Warburton, K. (2006). The psychological autopsy: solving the mysteries of death, *The Psychiatric Clinics of North America* **29**(3), 805–822.

[8] Snider, J., Hane, S. & Berman, A. (2006). Standardizing the psychological autopsy: addressing the Daubert standard, *Suicide and Life Threatening Behavior* **36**(5), 511–518.

[9] Baden, M. & Hennessee, J. (1990). *Unnatural Death: Confessions of a Medical Examiner*, Ballantine Books, New York, pp. 178–186.

[10] http://www.autoerotic-asphyxiation.com/ (accessed Dec 2008).

[11] Resnik, H. (1972). Eroticized repetitive hangings: a form of self-destructive behavior, *American Journal of Psychotherapy* **26**, 4–21.

[12] Garza-Leal, J. & Landron, F. (1991). Autoerotic asphyxial death initially misinterpreted as suicide and a review of the literature, *Journal of Forensic Sciences* **36**(6), 1753–1759.

[13] Hazelwood, R., Dietz, P. & Burgess, A. (1983). *Autoerotic Fatalities*, Lexington Books, Lexington.

[14] Sauvageau, A. & Racette, S. (2006). Autoerotic deaths in the literature from 1954 to 2004: a review, *Journal of Forensic Sciences* **51**(1), 140–146.

[15] Shields, L., Hunsaker, D., Hunsaker, J., Wetli, C., Hutchins, K. & Holmes, R. (2005). Atypical autoerotic death: part II, *The American Journal of Forensic Medicine and Pathology* **26**(1), 53–62.

[16] Hazelwood, R., Burgess, A. & Groth, A. (1981). Death during dangerous autoerotic practice, *Social Science and Medicine* **15E**, 129–133.

[17] Sauvageau, A. & Racette, S. (2006). Female autoerotic deaths – still often overlooked: a case report, *Medical Science and the Law* **46**(4), 357–359.

[18] Byard, R., Hucker, S. & Hazelwood, R. (1993). Fatal and near-fatal autoerotic aspyxial episodes in women: characteristics based on a review of nine cases, *The American Journal of Forensic Medicine and Pathology* **14**(1), 70–73.

[19] Behrendt, N., Buhl, N. & Seidl, S. (2002). The lethal paraphiliac syndrome: accidental autoerotic deaths in four women and a review of the literature, *International Journal of Legal Medicine* **116**(3), 148–152.

[20] Dietz, P. & Hazelwood, R. (1982). Atypical autoerotic fatalities, *Medicine and Law* **1**, 307–319.

[21] Byard, R., Eitzen, D. & James, R. (2000). Unusual fatal mechanisms in nonasphyxial death, *The American Journal of Forensic Medicine and Pathology* **21**(1), 65–68.

[22] Byard, R., Kostakis, C., Pigou, P. & Gilbert, J. (2000). Volatile substance use in sexual asphyxia, *Journal of Clinical Forensic Medicine* **7**, 26–28.

[23] Knight, B. (1996). *Electrical Fatalities. Forensic Pathology*, 2nd Edition, Arnold, London, pp. 319–331.

[24] Schellenberg, M., Racette, S. & Sauvageau, A. (2007). Complex autoerotic death with full body wrapping in a plastic body bag, *Journal of Forensic Sciences* **52**(4), 954–956.

[25] Sauvageau, A. & Racette, S. (2006). Aqua-eroticum: an unusual autoerotic fatality in a lake involving a homemade diving apparatus, *Journal of Forensic Sciences* **51**(1), 137–139.

[26] Andrew, T. & Fallon, K. (2007). Asphyxial games in children and adolescents, *The American Journal of Forensic Medicine and Pathology* **28**(4), 303–307.

[27] Site accessed on: 2/12/2008, http://www.zeroair.com/.

[28] Hitchcock, A. & Start, R. (2005). Fatal traumatic asphyxia in a middle-aged man in association with entrapment associated hypoxyphilia, *Journal of Clinical Forensic Medicine* **12**(6), 320–325.

[29] Vennemann, B. & Pollak, S. (2006). Death by hanging while watching violent pornographic videos on the Internet – suicide or accidental autoerotic death? *International Journal of Legal Medicine* **120**(2), 110–114.

[30] Gorniak, J.M., Sudimack, J.R., Stanforth, J.R. & Lewis, B.J. (2007). Hanging deaths with bound hands: what is the manner? *American Journal of Forensic Medicine Pathology* **28**(3), 232–234.

JAMES L. KNOLL, IV AND MICHAEL M. BADEN

Automated Fingerprint Identification System

Introduction

Definition

Automated fingerprint identification system (AFIS) refers to a computer system developed in order to assist in establishing the identity of an individual through fingerprints. AFIS technology is used for forensic and nonforensic (other government uses and commercial) applications, but this article concentrates only on the forensic applications.

Early Fingerprint Classification

At the end of the nineteenth century, William Herschel and Henry Faulds set out the principles of the forensic use of fingerprints and fingermarks: the use of fingerprints and fingerprint collections for the identification of offenders and the use of fingermarks to establish a link between a crime scene or an object and an individual. In the literature, confusion exists between the term print and mark. We will use a uniform terminology: the finger dermatoglyphics and their standard rolled or flat inked or scanned impressions are named *fingerprints* while the recovered or lifted traces are named *fingermarks*. *Marks* is a preferred usage in many countries to designate impressions that, in other countries (primarily the United States and Canada) would be characterized as *latent* impressions. In criminal records, the standard impressions are collected using forms named 10-print cards (see Figure 1).

The first system of fingerprint classification was introduced by Juan Vucetich in Argentina in 1891. The development and practical application of dactyloscopy for forensic use became known in law enforcement after the publication of the first manual and a system of fingerprint classification by Francis Galton. This led to the acceptance of fingerprints in Great Britain. Then, in 1900, Edward Richard Henry modified the classification system of Galton, which became the most widely used system under the name of Galton–Henry. The volume of paper-based 10-print card collections increased progressively during the twentieth century and workability was maintained by increasing the sophistication of the classification system, leading to a trade-off between selectivity and reliability.

AFIS Technology

Development

With the advent of computers in the mid-1960s, work on the automation of fingerprint identification started. Manual searching of the 10-finger hard-copy card collections reached its limits, in terms of workload and efficiency. The USA and Japan concentrated on the automation of the high-volume 10-print work load, while France and the UK focused more on the automation of fingermark identification. After a decade of effort, digitization of the 10-print card and automatic extraction of minutiae became effective enough for the USA and UK to produce automatic fingerprint reader systems. This advancement eventually permitted the digitization of the 10-print cards, thus permitting the storage of the standard impressions and the demographic data of individuals (e.g., name, citizenship, and date of birth) in a computerized database [1, **Databases**].

Image Capture. Currently, flat-bed document scanners are used for the digitization of the 10-print cards, but the capture of standard impressions is also possible by applying and rolling the finger friction ridge skin directly on fingerprint live-scan equipment (*see* **Biometric Devices**). The fingermarks are captured from their initial support with digital cameras or are scanned from analog pictures. Image-processing techniques are then applied to segment each fingerprint of a 10-print card as an individual image and to enhance the image parameters. Finally, the images are stored as 256 gray-level (8-bit-GS) images at a resolution of 500 or 1000 pixels per inch (ppi) and compressed at a ratio of about 1 : 15 using wavelet-based algorithms such as wavelet scalar quantization (WSQ) or Joint Photographic Expert Group 2000 (JPEG2000 or JP2). The current international American National Standards Institute (ANSI)-Nattional Institute for Standards and Technology (NIST) fingerprint/fingermark image exchange file format stores JP2 8-bit-GS, 1000 ppi images.

Feature Extraction. In parallel to digitization, research has concentrated on classification. The first approaches translated the manual pattern

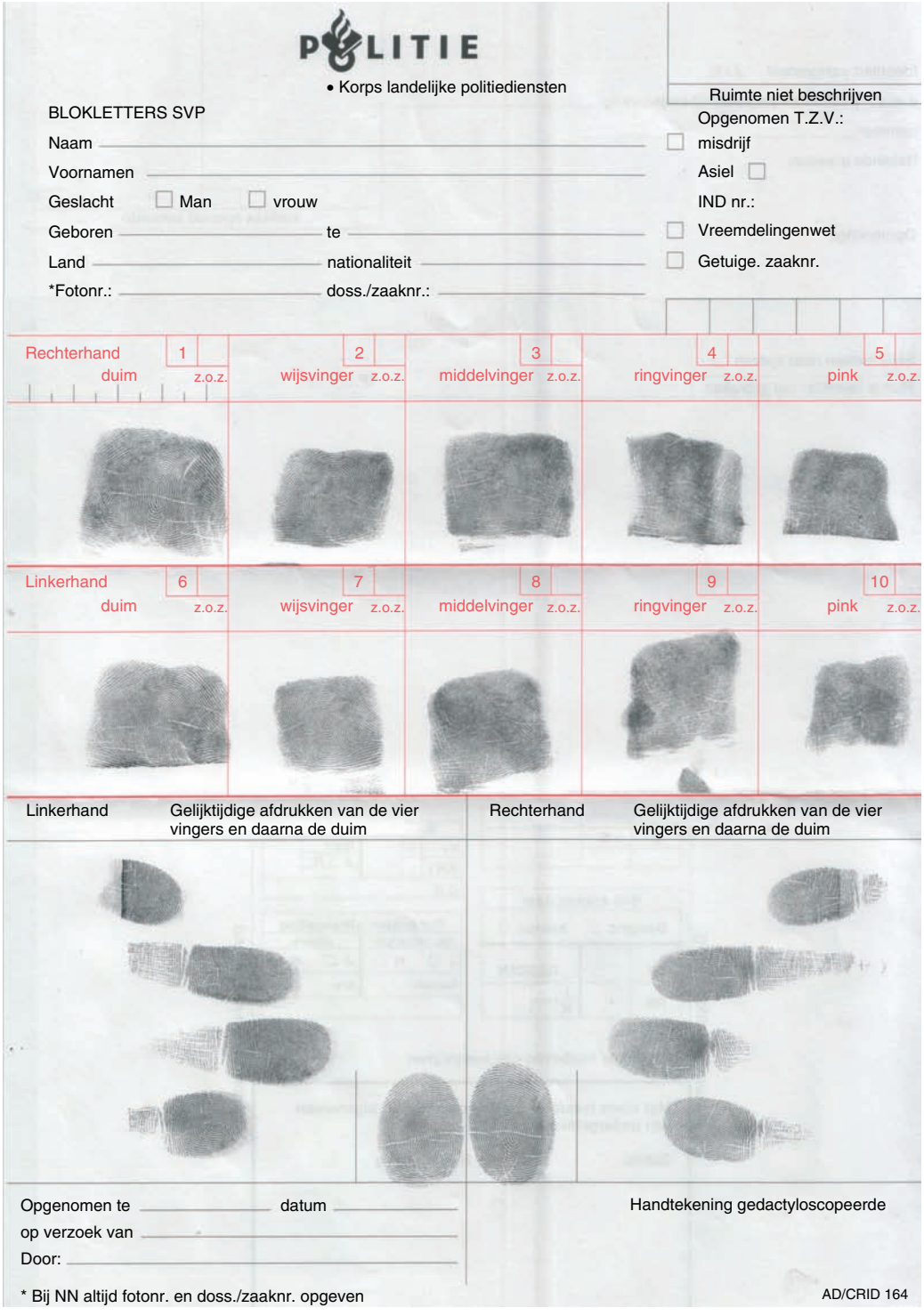


Figure 1 Example of 10-print card

classification into computer-friendly codes based on the shape of the papillary ridge flow; this shape is named *general pattern* and classified as first-level detail. This operation resulted in a potential elimination of the manual search and filing errors, but did not offset the original pattern type assignment errors [2]. Later, the fingerprint pattern classification was automated using the ridge direction matrix extracted from the fingerprint images. However, this suffered from the same type of assignment errors [3]. A solution was found with the more precise automatic minutiae detection and the comparison of minutiae configurations. Minutiae are points of termination (ridge ending) or bifurcation of the papillary ridges; they are also named *Galton points* and classified as second-level detail (see Figure 2a and b).

Current AFIS systems use the position (x and y coordinates) and the tail angle (θ) of the minutiae as the core components of comparison, followed by the use of extended characteristics of the minutiae such as their basic type (ridge ending or bifurcation), the ridge count between pairs of minutiae or the topology of the minutiae in combination with other features. The crucial elements in the success of this approach are the extreme typicality and robustness offered by the topology of the minutiae configurations and the development of robust image-processing techniques to enhance the papillary ridge structure and clarity.

The main stages of the feature extraction components consist in an orientation field estimation, by using information about local average directions of the ridges (gradient or ridge-valley algorithm), a ridge detection, by using either a thresholding algorithm given a gray-scale image, the ridge-valley algorithm, or a gray-level histogram decomposition, and a thinning of the ridges, by means of algorithms based on mathematical morphology [4]. These techniques allow for an accurate minutiae designation on high-quality fingerprint and fingermark images, as the ridges have well-defined frequency and orientation in local areas. On low-quality fingerprint and fingermark images, automatic minutiae designation remains a challenge and still does not equal the ability of human friction ridge examiners to designate the visible minutiae.

Feature Comparison. Numerous methods have been developed to automatically compare fingermark and fingerprint images; however, they can be classified into two approaches: the correlation-based and the structural feature-based comparison. The correlation-based comparison relies on global pattern matching; it consists in using translation and rotation to find the best superposition to compute the maximum correlation between two images. The structural feature-based comparison is based on the matching of extracted features; it consists in searching alignment between minutiae extracted from a fingermark

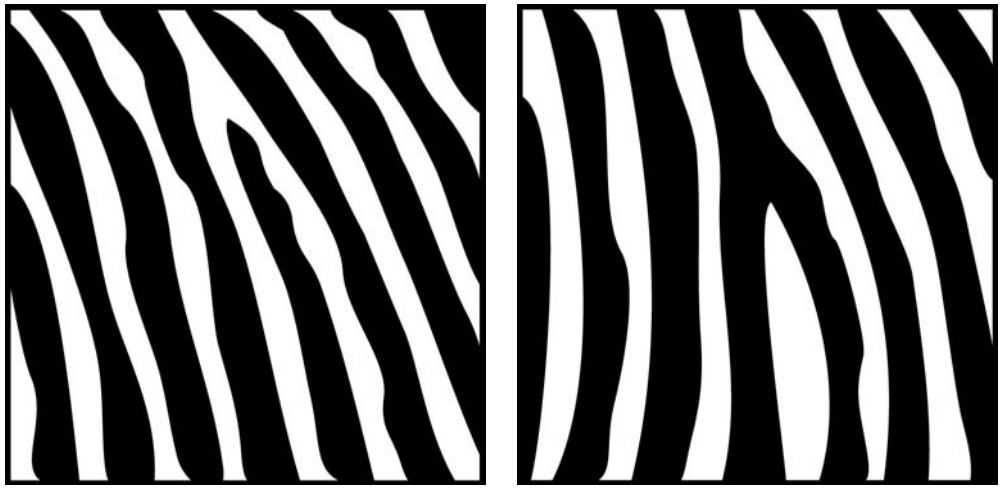


Figure 2 (a) Minutia or Galton point, basic type ridge ending. (b) Minutia or Galton point, basic type bifurcation

and a fingerprint image and in finding the maximum number of minutiae pairing. The structural feature-based approach is more robust but requires more computation than the correlation-based comparison. In general, the result of the comparison is expressed as a scalar number representing a statistical distance or a proximity index between a fingermark and a fingerprint [5].

Forensic AFIS systems exploit these two approaches to optimize their performance, but the problem remains complex. This difficulty is mainly due to the uncontrolled conditions of production of fingermarks during criminal activity and the large variability of different impressions of the same finger, named within-finger variability. The main factors responsible for this within-finger variability are the ridge skin condition, the transfer of information with loss from a complete three-dimensional pattern to a partial two-dimensional pattern, and the nonlinear distortions of the papillary ridges. These distortions result from the skin plasticity and from the movement of the finger during the production of the fingermark. These factors do not modify the topological relationships of the papillary ridges and of the minutiae configurations, but they modify the shape of the papillary ridge flow and the absolute distance between any pair of minutiae. The key to fingerprint features comparison is to exploit the topological invariance of the minutiae configurations despite the factors responsible for the within-finger variability. The current sophistication of the algorithms allows for state-of-the-art AFIS technologies to reach a false rejection rate (FRR) lower than 2% for a false acceptance rate (FAR) of 0.0001% in a task of identity verification comparing inked rolled fingerprints and partial fingermarks containing 15 paired minutiae.

Forensic Uses of AFIS Technology

AFIS technology was initially developed to assist the friction ridge examiners with computers in the process of verification of the identity of individuals through their fingerprints. This process consists in searching the 10 fingerprints of an individual in the database of standard impressions to verify if (s)he is already present in the database and, if present, to verify the rightness of his/her demographic data. AFIS technology has achieved enough maturity to offer an identity verification process, that is, virtually error-free from the technological point of view, although clerical mistakes in the database or in the running of the process can never be excluded.

In the 1990s, the semiautomatic processing of fingermarks was made possible by the improvement of both AFIS and computer technologies. Allying the manual minutiae extraction of the friction ridge examiners and the automatic comparison of the AFIS allowed the partial automation of forensic investigation and forensic intelligence using fingermarks. For forensic investigation, lists of potential sources for fingermarks could be produced automatically from the digitized 10-print card databases. For forensic intelligence, links between digitized fingermarks could be generated automatically.

In the 2000s, the improvement of the computer mass-storage, in terms of size and affordability, initiated the constitution of large-scale palmprint databases. It has allowed for an extension of the forensic investigation and forensic intelligence processes based on the use of palm marks. In most countries, the development of large-scale palmprint databases is an ongoing process.

Future Challenges for AFIS Technology in Forensic Science

Further Automation. Minutiae designation on low-quality fingerprint and fingermark images remains a computer-assisted process, combining the outstanding but subjective human pattern recognition ability and the more objective but also more limited computer ability. Both the human inconsistencies and the limits of the computer affect the performance of the feature extraction and, as a consequence, the performance of the forensic investigation and intelligence processes based on the use of finger- and palm-marks. Thus, a feature extraction process at once reliable and completely automated is desirable. Steps in this direction can be achieved not only by refining the minutiae extraction process but also by enriching the feature vector with other available, measurable, discriminatory, permanent, and robust features, like, e.g., the automatic count of the number of ridges between the minutiae.

Scalability and Interoperability. The scalability of the paper-based 10-print collections was limited by the necessary trade-off between selectivity and reliability imposed by manual classification systems, and the coexistence of several systems around the world

limited their interoperability at international level. The implementation of AFIS solved the problem of scalability for the national fingerprint databases but the interoperability problem remains as the first generations of AFIS incorporate feature vectors encoded using proprietary formats. Currently, the problem of interoperability between different types of AFIS is being solved partially by the widespread use of the ANSI-NIST biometric exchange file format, but the predominant use of proprietary formats for the feature vectors remains. This improvement opens a new challenge in terms of scalability, with the possibility to extend the interoperability of the AFIS at a global level.

Forensic Evaluation Based on AFIS Technology. Forensic evaluation of fingermarks and fingerprints consists mainly in the inference of identity of the source of a fingermark and a fingerprint. Currently, this task remains exclusively performed by friction ridge examiners. They apply the analysis, comparison, evaluation, verification (ACE-V) procedure and, in some countries, a numerical standard to substantiate three types of qualitative opinion: identification, exclusion, or inconclusive (*see* **Identification and Individualization**). As their evaluation is deterministic, friction ridge examiners also make an implicit use of their own subjective probabilities of the rarity of the features used for identification. They refine these subjective probabilities through training and experience.

In forensic research, the inference of identity of the source of a fingermark and a fingerprint is also envisaged, combining AFIS technology, digitized fingerprint, and fingermark databases as well as a scientific methodology; namely the likelihood ratio approach based on the Bayes theorem. This new approach aims to offer a uniform framework and a transparent methodology to the friction ridge examiners, and to assist them in producing a logical, testable, and quantitative evaluation of the fingerprint evidence [6, **Friction Ridge Examination (Fingerprints): Interpretation of**].

References

[1] Berry, J. & Stoney, D.A. (2001). History and development of fingerprinting, in *Advances in Fingerprint Technology*, H.C. Lee & R.E. Gaensslen, eds, CRC Press, Boca Raton, 1–40.

[2] Allen, R., Sankar, P. & Prabhakar, S. (2005). Fingerprint identification technology, in *Biometric Systems: Technology, Design and Performance Evaluation*, J. Wayman, A., Jain, D. Maltoni & D. Maio, eds, Springer-Verlag, London, pp. 21–61.

[3] Moore, R.T. (1991). Automatic fingerprint identification systems, in *Advances in Fingerprint Technology*, H.C. Lee & R.E. Gaensslen, eds, Elsevier Science Publishing, New York, pp. 163–191.

[4] Yager, N. & Amin, A. (2004). Fingerprint verification based on minutiae features: a review, *Pattern Analysis and Application* **17**, 94–113.

[5] Uchida, K. (2005). Fingerprint identification, *NEC Journal of Advanced Technology* **2**(1), 20–27.

[6] Becue, A., Champod, C. and Margot, P. (2007). Fingermarks, bitemarks and other impressions (barefoot, ears, lips) – A review: 2004 to 2007, *15th Interpol Forensic Science Symposium*, Lyon, France, pp. 745–800.

DIDIER MEUWLY

Automatism as a Defense to Crime

Automatism, from the Greek *automatismos* or self-action, refers to unconscious, involuntary behavior. The terms *automatism* and *unconsciousness* are sometimes used interchangeably. Either word has been used to mean that individuals have some degree of unawareness of their thoughts or behaviors. Automatism may be used as a defense against criminal responsibility, in essence to reduce culpability, based on the idea that a person cannot be held liable for actions that do not stem from full awareness of thoughts and behaviors.

In the case *People v Huey Newton* [1], the court held that "unconsciousness, when not self-induced say, as by voluntary intoxication, is a complete defense to a criminal act even though the defendant's acts seem very goal oriented". Lord Denning in *Bratty v Attorney General of Northern Ireland* [2] recognized two classes of automatism. The first class is a complete lack of consciousness and the second "an absence of control of the mind over actions". The latter applies to persons conscious of, but not in control of, their own actions.

The causes of automatism used in criminal defense include somnambulism [3], hypnotism [4], cerebral concussion, delerium from fever or drugs, diabetic shock [5], and epileptic blackouts [6, 7].

In American courts there has been no consistent interpretation of the automatism defense. Some courts have viewed automatism as a variation of the insanity defense, whereas other courts have accepted it as a complete affirmative defense. A great deal rests on the court's interpretation because an insanity defense usually results in hospitalization, whereas a successful strict affirmative defense results in acquittal [8].

In several cases in the United States, automatism has been considered a variation of the insanity defense (*see* **Insanity: Defense**). The Model Penal Code developed by the American Law Institute in 1962 and updated in 1981 defined the standard for the insanity defense as follows: "A person is not responsible for criminal conduct if at the time of such conduct, as a result of mental disease or mental defect, he lacks substantial capacity either to appreciate the criminality of his conduct or to conform his conduct to the requirements of law" [9].

In *People v Higgins* [10], Mr Higgins entered a plea of not guilty by reason of insanity contending that at the time of the offense he was "in the throes of an epileptic rage or furor, and consequently was laboring under such a defect of reason as not to know the nature of the quality of the act or that it was wrong". Mr Cooley, in a similar insanity defense application of automatism, presented evidence that at the time of committing the offense he was suffering from epilepsy and consequently was unaware of his actions [11]. The instructions to the jury in this case were "The law presumes every man sane until . . . the contrary is shown by the evidence; and, before the defendant can be excused on the ground of insanity, the Jury must believe from the evidence that the defendant was at the time of the killing without sufficient reason to know what he was doing, or had not sufficient reason to know right from wrong, or that, as a result of mental unsoundness, he had not then sufficient will power to govern his actions, by reason of some insane impulse which he could not resist or control". In these cases, the defense relies on the notion that a person's state of mind is incapable of knowing right from wrong, or if they know it, they have no volitional control over their behavior.

Other courts have recognized automatism as a complete affirmative defense, distinct from the insanity defense. The absence of a clear mental disease or defect, a requirement of the insanity defense in all jurisdictions, demands a different approach. The complete defense conceptualization is based on *actus reus*, the guilty act component of the two part test of criminality (the other being *mens rea*, guilty mind) must be voluntary. The Model Penal Code states that "A person is not guilty of an offense unless his liability is based on conduct which includes a voluntary act or the omission to perform an act of which he is physically capable." The code offers the following examples of involuntary acts: reflex or convulsion, bodily movement during unconsciousness or sleep, conduct during hypnosis or resulting from hypnotic suggestion, bodily movement that otherwise is not a product of the effort or determination of the actor, either conscious or habitual [9].

In *State v Caddell* [12], the Supreme Court of North Carolina stated that "the defenses of insanity and unconsciousness are not the same in nature, for unconsciousness at the time of the alleged criminal act need not be the result of a disease or defect of the mind; as a consequence, the two defenses are not the same in effect, for a defendant found not guilty by reason of unconsciousness, as distinct from insanity, is not subject to commitment to a hospital for the mentally ill". In *People v Newton* [13], the court stated that "unconsciousness is a complete defense that negates the defendant's capacity to commit any crime at all".

Likewise, a complete defense based on the absence of *mens rea* and on the inability to form some intent has been applied in court. This approach again differs from the insanity defense because it implies that an automatism is not the same as a mental disease or defect. In *People v Grant* [14], after the jury had rejected an insanity defense, the appellate court granted an epileptic defendant a remand to allow the jury to consider an automatism defense. Illustrative of the difficulty in clarifying *mens rea* in such defenses, the jury was allowed to consider prior actions as reflective of the degree of awareness of criminal behavior to impute the *mens rea* of the actual offense.

Regardless of the type of automatism defense, the insanity defense, which asserts a lack of knowing or volitional control of wrongful behavior, or a complete affirmative defense due to the absence of voluntary *actus reus* or *mens rea*, the court is challenged to determine the degree to which a person has access to and awareness of their thoughts and behaviors. Ultimately a judgment must be made as to whether the defendant had *enough* ability, along the continuum of awareness, to access into consciousness their criminal thoughts and behaviors. According to Candeub [15] "this judgment is crucial for determining culpability because only an actor who can be expected to consciously respond to reason can be morally and legally culpable". To simply offer that somnambulism, for example, is completely involuntary does not necessarily match the science. Somnambulism, and the behaviors associated with somnambulism, seems to be neither fully voluntary nor fully involuntary [16]. Thus, even defendants who following some otherwise criminal behavior due to somnambulism are able to later recall and discuss their intents and acts may be less culpable, or even lack complete culpability, if in the judgment of the court those intents and acts occurred outside the degree of awareness necessary for responsibility, *actus reus* or *mens rea*.

In the British system, a distinction is made between sane and insane automatism [8]. British courts use the *McNaughten* [17] standard for the insanity defense. The *McNaughten* standard states that, for a successful insanity defense, it must be proven that "at the time of committing the offense, the accused was laboring under such a defect of reason, from disease of the mind, as not to know the nature and quality of the act he was doing or if he did know it, that he did not know he was doing what was wrong". Within this context automatism is conceptualized as a "defect of reason", which may or may not be as a result of a "disease of the mind". If the "defect of reason" is a result of a "disease of the mind", it is defined as *insane automatism*, which meets the *McNaughten* standard. Consequently, an insanity defense is appropriate. If the "defect of reason" is not as a result of a "disease of the mind", it is considered a sane automatism, which may then serve as a complete defense and potential cause for acquittal.

The British system attempts to resolve the issue of degree of awareness of thoughts or actions through an understanding of the cause of the purported unawareness. If the trier of fact concludes that the degree of unawareness meets some minimal threshold, then the question becomes, "What caused the unawareness, a disease of the mind or not?" In *R v Sullivan* [18] Mr Sullivan contended that, while he had caused grievous bodily harm to the victim, he had done so while recovering from a minor epileptic seizure and that, therefore a defense of automatism was appropriate. The trial judge ruled that the defense was one of insanity not automatism. On appeal the House of Lords agreed, stating "if the effect of a disease was to impair those faculties so severely as to have the consequence that the accused did not know what he was doing, or, if he did, that he did not know that it was wrong, he was 'insane' in the legal sense. Accordingly, it did not matter whether the cause of the impairment was organic, as in epilepsy, or functional, or whether the impairment itself was permanent". The court further defined sane automatism as "temporary impairment of the mental faculties, not being self-induced by consuming drink or drugs, resulting from some external physical factor such as a blow to the head causing concussion or the administration of an anesthetic for therapeutic purposes".

Unfortunately, the Sullivan court's effort to interpret a disease of the mind as due either to internal factors or external factors only creates a subsequent problem. The argument that insane automatism is caused by internal factors and sane automatism by external factors is problematic. Consider, for example, the case of *R v Quick* [19] in which a diabetic in a hypoglycemic state injured an individual. The trial judge ruled that the appropriate defense was one of insanity. The Court of Appeals disagreed stating that it was a defense of "non-insane automatism, involuntary conduct which is not brought about by a disease of the mind but through other factors". The court considered it an "affront to common sense to regard a person as mad whose symptoms can be rectified by a lump of sugar".

Matters are further complicated because a judge may also consider the likelihood of recurrence of the criminal behavior as a factor in deciding whether the appropriate defense an insanity defense or a complete affirmative, that is, whether the appropriate disposition is hospitalization or acquittal. In *Bratty v Attorney General of Northern Ireland* [2], the court held that "any mental disorder which has

manifested itself in violence and is likely to recur is a disease of the mind. At any rate it is the sort of disease for which the person should be retained in hospital rather than be given an unqualified acquittal".

In summary, automatism, as a defense against criminal responsibility is problematic. Courts in both the United States and Britain have not had a consistent interpretation of the automatism defense.

References

[1] People v. Huey Newton, 8 Cal Rptr. 394 (Cal. 1970).
[2] *Bratty v Attorney General of Northern Ireland* AC 386 (1963).
[3] Fain v. The Commonwealth, 78 Ky. 183 (Ky. 1879).
[4] People v. Worthington, 105 Cal. 166 (Cal. 1894).
[5] Corder v. Commonwealth Ky. 278 S.W.2d 77.
[6] Smith v. Commonwealth Ky. 268 S.W. 2d 937.
[7] People v. Magnus, 92 Misc. 80, 155 N.Y.S. 10113.
[8] Schopp, R.F. (1991). *Automatism, Insanity And The Psychology Of Criminal Responsibility, A Philosophical Inquiry*, Cambridge University Press.
[9] Model Penal Code, American Law institute 1962, updated (1981).
[10] People v. Higgins, 5 N.Y.2d 607 (NY. 1959).
[11] Cooley v. Commonwealth of Kentucky, 459 S.W 2d 89 (Ky. 1970).
[12] State v. Caddell, 287 N.C. 266: 215 S.E. 2d 348; (1975).
[13] People v. Newton, 8 Cal. App 3d 359: 87 Cal. Rptr. 394 (Cal. 1970).
[14] People v. Grant, 6 Ill. App. 3d 125, 360 N.E.2d 809 (1977).
[15] Candeub, A. (2002). Consciousness and culpability, *Alabama Law Review, Fall* **54**, 113.
[16] McCall Smith, A., Shapiro, C. (1997). Sleep law: a challenge in law and medicine, *Forensic Aspects of Sleep*, C. Shapiro & A. Mc.Call.Smith, eds, Wiley.
[17] McNaughten 10 C&F 200 (1843).
[18] R v. Sullivan, 2 Alll E.R. 673, 675–676 (1983).
[19] *R v Quick* QB 910 (1973).

CHERYL A. HILL AND RYAN FINKENBINE

Automatisms ***see*** Trauma Causation: Analysis of Automotive; Seizures: Behavioral

Autopsy

Historical Overview

Anatomical Dissection

The knowledge of the early history of human anatomy is scanty and the possibility to dissect human bodies varied greatly at times. Very early Indian medical literature indicates a knowledge of anatomy that may have been acquired by dissecting bodies of young children as the Hindu scriptures enjoined that bodies of persons more than two years old had to be burned [1].

The first known school of anatomy was in Alexandria around 320 B.C., where human dissections were carried out by famous anatomists, such as Herophilus of Chalcedon (335–280 B.C.) and Erasistratus of Iulis (c. 310–250 B.C.) to determine the normal structure of organs and the changes made by disease [2].

According to several reports, occasional "autopsies" were performed already in the twelfth and thirteenth centuries. The English chronicle by William of Malmesbury (c. 1080/1095-c.1143) mentions one ordered by the Norwegian King Sigurd Jorsalfar (1090–1130) in 1111, while stopping in Byzantium (present Istanbul) on his return from Jerusalem, to find out whether the cause of death of many of his followers had been liver damage by too strong wine [3].

In 1391, King John I of Aragon gave the University of Lérida in Spain the permission to dissect, once every 3 years, the body of a criminal. In Vienna the first anatomical dissection took place in 1404, in Prague somewhat later in 1460 and at the University of Paris in 1478 [4].

The first known postmortem examination on the American continent was performed in Hispaniola in 1533 and recorded by Fernandez de Oviedo in his General and Natural History of the Indies [5].

The publication of the great textbook of anatomy *De Humani Corporis Fabrica* in 1543 by Andreas Vesalius (1514–1564), the "father of anatomy", marked the beginning of an overthrow of traditional Galenic anatomy and theories, although it took more than 300 years until the new concepts of pathogenesis of diseases and cellular pathology started to take over from the old school of thought.

Medicolegal Autopsy

The application of medicine in the administration of justice began long before medicine started to develop into a modern science. The medicolegal investigation of deaths was introduced very early from the requirements of the judicial system as it was of interest to the society to know whether a person had died as a result of violence or due to natural causes. The oldest known instructions by the authorities about external examination of corpses have been found in China and dated to the Ching dynasty (~2000 B.C.). During Song dynasty (960–1279 A.D.), a decree enacted in 995 A.D., provided that a government official, though not a physician, had to investigate a violent or otherwise suspicious death within four hours. It is not known whether autopsies were ever performed in ancient China [6].

The earliest medicolegal autopsies took place in Italy, probably in the middle of the thirteenth century, at the University of Bologna. One was recorded by Guglielmo de Placentinus Saliceto, or by William of Saliceto (1210–1276 or 1277), a surgeon and a teacher, on the medical faculty there and is mentioned in his book *Surgery* [7]. Whether this really was an autopsy or just an external examination of the corpse, is questionable. Some authors have attributed the first medicolegal autopsy to Bartolomeo da Varignana, who, in addition to his professional and academic activities, served the municipality of Bologna in a medicolegal capacity. The public prosecutor had ordered an autopsy in 1302 in a case suspected of having died due to poisoning [8, 9].

According to O'Neill, medical historians have long agreed that the fundamental motive for the initiation of academic dissections in Western Europe was forensic. The decretals of Pope Innocent III contain several cases in which such examinations had provided the evidential basis for a papal verdict [8].

In Europe the principles of medicolegal investigation of deaths were developed on the basis of the criminal codes of the sixteenth century: the Bamberg Code in 1507, compiled by Johann Freiherr von Schwarzenberg and which was soon adopted by a number of other German states. These along with the best known criminal statutes in Europe, Constitutio Criminalis Carolina, the Caroline Code of 1532, obliged the courts to use medical expertise in certain medicolegal issues such as abortion, bodily harm, homicide, or medical malpractice.

The "father of the Swiss legal medicine" Felix Platter I (1536–1614) from Basle has been said to have performed more than 300 autopsies. In France the first medicolegal autopsy was performed by Ambroise Paré (1510–1590) in 1562. The autopsy of King Stephen Báthory of Poland in 1586 was probably the first one performed in the Eastern Europe [10].

Our knowledge about the old autopsy procedures is poor and, apart from a few exceptions, detailed written autopsy reports are relatively recent.

In addition to codification of the principles of medicolegal investigations, some countries went further providing detailed instructions as to the performance of medicolegal autopsy. By far the best example of this is the Austrian decree of 1855. The objective of medicolegal investigation of the deceased and the duty of medicolegal expert has probably never been defined with such brevity and clarity than in Section 1 of the Austrian decree. It is noteworthy that it is still valid legislation in Austria.

"The medicolegal examination of the deceased is of great importance and therefore it is an imperative duty of the expert, that the examination must be carried out with utmost accuracy, as very often, the honor, freedom, property, and life of the person accused of a criminal act, as well as the rule of law, depend upon it". (Authors' translation of Section 1 of the Austrian decree on medicolegal examination of corpses from January 28, 1855, Figure 1.)

The decree gives detailed instructions in 134 paragraphs for the performance of medicolegal autopsy. The Prussian edict of 1875 is similar but not as detailed as the Austrian one and both of them can, so far, be considered as the culminating point of legislation dealing with the performance of medicolegal autopsy.

Clinical Autopsy

Early attempts to solve medical problems through autopsies can be found already in the Middle Ages e.g., in the chronicle of Fra Salimbene. He referred in his description to one dissection performed by an unnamed Cremonese physician, of a victim of the malady that swept through the cities of Northern Italy in 1286. Similarly, victims of bubonic plague had been dissected in Perugia during the outbreak of an epidemic in 1348 [8].


Nachdruck. 233

Reichs-Gesetz-Blatt

für das

Kaiserthum Oesterreich.

Jahrgang 1855.

VIII. Stück.

Ausgegeben und versendet am 13. Februar 1855.

26.

Verordnung der Ministerien des Innern und der Justiz vom 28. Jänner 1855,

womit die Vorschrift für die Vornahme der gerichtlichen Todtenbeschau erlassen wird.

Die Ministerien des Innern und der Justiz finden für die Vornahme der gerichtlichen Todtenbeschau die nachstehende Vorschrift zu erlassen:

Erstes Hauptstück.

Von der gerichtlichen Todtenbeschau überhaupt.

§. 1.

Figure 1 Front-page of the Austrian decree on medicolegal investigation of corpses January 28, 1855

However, hospital or clinical autopsy became meaningful first after the introduction of modern concepts of pathogenesis of diseases by Carl von Rokitansky (1804–1878) and cellular pathology by Rudolf Virchow (1821–1902).

Gradually in the beginning of the nineteenth century increased attention was paid to the actual autopsy technique. In 1846 Rudolf Virchow, then prosecutor in Berlin, insisted on regularity, method, and definitive technique. The classical techniques, which are still in use today are, more or less modifications of those introduced by Rokitansky (organs are dissected *in situ*), Virchow (organs are removed one by one), Ghon (thoracic and cervical organs, abdominal organs, and urogenital system are removed as organ blocks) and Letulle (cervical, thoracic, abdominal, and pelvic organs are removed as one organ block), among others [11].

In 1872 Francis Delafield's *A Handbook of Post-mortem Examination and Morbid Anatomy* was published in New York and German and English editions of Rudolf Virchow's book on autopsy technique were published in 1876.

The Present Autopsy

There is considerable variation in the types and standards of medicolegal systems throughout the world from nonexistent to well-functioning organizations with sophisticated autopsy practices. Generally the lack of forensic and medicolegal services is a characteristic of developing countries, usually including a poor judiciary and educational system, but the awareness of the importance of high standard medicolegal autopsy for the protection of individual rights varies a lot even in industrialized countries due to historical and/or political reasons.

Most countries have not provided laws to regulate the actual performance of medicolegal autopsies. There exist both international recommendations and national guidelines by the competent authorities or professional bodies such as the Royal College of Pathologists in the United Kingdom or minimum standards by individual forensic institutes.

Various international bodies have been interested in achieving harmonized and internationally recognized rules concerning the performance of medicolegal autopsies. In Europe the "Sevilla Working Party on the Harmonization and Standardization of Forensic Medicine (SWP)" was established in 1985 by Professor Luis Frontela Carrerras of Seville, Spain, carrying out groundwork and collaborating with the Council of Europe to create minimum standards for autopsy protocols. The Minnesota Lawyers International Human Rights Committee had taken the initiative to draft international standards for the investigation of suspicious deaths, including autopsy. In May 1989, UN Economic and Social Council adopted in its resolution 1989/1965 "the Principles on the Effective Prevention and Investigation of Extralegal, Arbitrary, and Summary Executions", which had been created by cooperation with intergovernmental and nongovernmental organizations. In 1991, the General Assembly of the United Nations endorsed the Model Autopsy Protocol of the United Nations. SWP's work was continued by the European Council of Legal Medicine (ECLM), whose document "Harmonisation of the Performance of the Medicolegal Autopsy" was adopted by the General Assembly in London in 1995 and which largely served as basis for the Pan European, Council of Europe – Recommendation No. R (99) 3 On the Harmonisation of Medico-Legal Autopsy Rules and Its Explanatory Memorandum, which was adopted by the Committee of Ministers in February 1999 [12].

In spite of the invention of new medical imaging techniques such as computed tomography (CT), multislice computed tomography (MSCT), or magnetic

resonance imaging (MRI) clinical autopsy has been shown to have maintained its clinical, educational, and epidemiological value and remained an essential factor in the quality assurance of medical care [13]. In spite of this, there has been a progressive decline in autopsy rates in recent decades in most industrialized countries on all continents [14–16].

In 1971, the Joint Commission on Accreditation of Hospitals (JCAH) in the United States ended its requirement for minimum rates of autopsy for certification. Until then, a mandatory 20% autopsy rate had been required for accreditation of postgraduate training. The idea had been that each institution should set its own rate but the dropped requirement seems to have had a negative impact on autopsy rates and, according to an estimate, the national autopsy rate of nonforensic deaths may have fallen roughly to 5% [17].

There are several reasons for this decline and the emphasis of contributing factors varies according to the country. Clinicians may be too reliant on the new diagnostic techniques and poorly motivated to obtain consent from the relatives to perform an autopsy. Pathologists are less enthusiastic, and owing to low autopsy rates, less experienced to perform autopsies and this task is often delegated to inexperienced trainees. Further factors that probably have contributed to the decline of autopsies are fear of malpractice litigation, if errors in clinical judgment are detected, and cost-cutting pressures, when autopsies do not produce income [18, 19].

Objectives of Autopsy

Autopsy is a detailed systematic external and internal examination of a corpse carried out by a pathologist or one or more medicolegal experts to ascertain the underlying and possible contributing causes of death and, depending on the jurisdiction, also the manner of death. Before the pathologist can begin the examination, he must be sure that he has been authorized to perform the autopsy on that particular body. As infection risks are common in mortuary and autopsy may involve also other risks, such as electrical, chemical, and radiological hazards, these must be assessed and necessary health and safety precautions taken. The autopsy and all ancillary investigations must be carried out in compliance with the relevant legislation and possible national guidelines in a manner consistent with medical ethics and respecting the dignity of the deceased.

An autopsy is performed to achieve one or more of the following objectives:

- to identify or record characteristics that may assist in identifying the deceased;
- to determine the cause of death or, in the newborn, whether live birth occurred;
- to determine the mode of dying and time of death, where necessary and possible;
- to demonstrate all external and internal abnormalities, malformations, and diseases;
- to detect, describe, and record any external and internal injuries;
- to obtain samples for any ancillary investigations;
- to obtain photographs or retain samples for evidential or teaching use;
- to provide a full written report and expert interpretation of the findings; and
- to restore the body to the best possible cosmetic condition before the release.

In addition to the anatomical dissection, there are basically two main types of autopsy and, although in principle, a single standard should be applicable to all autopsy examinations, this is not the case in practice, owing to different traditions and structures in education of medical specialists and wide variation in autopsy practices, from partial autopsy to full examination of all body cavities including ancillary investigations such as histology of all major organs, depending on the country and jurisdiction.

The clinical or hospital autopsy must, in most jurisdictions, be consented by the deceased person, before death, or the next of kin after death. It is usually performed to investigate details of a known disease process, e.g., the exact type or spread of a malignant disease or the effectiveness of therapy. It can also be performed for medical audit or research purposes.

The medicolegal or forensic autopsy necessitates, owing to legal issues involved, the highest possible standards of practice. It is ordered by the competent legal authority (a coroner, a medical examiner, a procurator fiscal, a magistrate, a judge, or the police) to investigate sudden unexpected, suspicious, unnatural, or criminal deaths. Also unidentified bodies or deaths occurring in special circumstances such as deaths in police custody or during imprisonment are often subjected to a medicolegal autopsy. In most

jurisdictions, permission of the next of kin is not required [20–22].

Autopsy Techniques

The basic technique of autopsy has remained the same since 1850s but various imaging techniques such as CT, MSCT, MRI, percutaneous needle biopsies, or endoscopic techniques have been applied, either for preautopsy screening purposes, or as so called "*minimally invasive autopsy*" if full autopsy is not possible.

Both clinical and medicolegal autopsy may involve different strategies and techniques depending on the questions they are expected to answer and pediatric autopsy technique is generally somewhat different from the one in adults. The Royal College of Pathologists (the United Kingdom) has published a set of Guidelines on Autopsy Practice and scenarios for specific types of death, to facilitate the improvement of the standards of autopsy performance and reporting [23].

The scope of medicolegal autopsy is often broader than that of clinical autopsy, with special emphasis on all information concerning the circumstances of death, often necessitating a visit to and investigation of the scene of death. Both types of autopsies should consist of full external and internal examination of the body including the dissection and investigation of all three body cavities.

The External Examination

In medicolegal autopsy, the examination of the clothing is often an essential part of the external examination, whereas in clinical autopsy it is generally not.

The description of the body includes age, sex, build, height, ethnic group and weight, nutritional state, color of the skin as well as other characteristics of the deceased including color of irises and sclerae, presence or absence of petechiae; color, length, density and distribution of hair, or any abnormalities such as scars or tattoos. All injuries must be described in detail, measured, and photographed. Postmortem changes relating to rigor mortis, hypostasis, and decomposition should be described in detail. The examination should be carried out carefully and systematically and include head, neck, trunk, upper and lower extremities, and the back and description of all body orifices.

The Internal Examination

The state of body cavities includes the description of the presence of gas (pneumothorax), fluids (effusions or exudates), or foreign bodies and the measurement of their volume, appearance of the internal surfaces, and anatomical boundaries as well as location and external appearance of organs.

The classical autopsy techniques vary mainly in the order in which the organs are removed. All organs have to be dissected, the outer appearance as well as the cut surfaces described and weights recorded. The hollow organs have to be opened and their content described and measured. All relevant vessels, arteries and veins as well as ducts have to be dissected. All abnormalities must be described by location and size.

Ancillary Investigations

Many pathological processes are not visible to the naked eye, therefore, postmortem histology is necessary and of great value, not only in confirming but also for excluding any pathological changes in a cause of death investigation. As a rule, complete histological examination of all major organs is considered an essential part of every postmortem and this principle has been adopted by many guidelines.

The need for other ancillary investigations depends on whether the cause of death has been established with the necessary degree of certainty, and if not, additional samples may have to be taken for toxicological or other investigations such as bacteriology and virology. For toxicology this may include peripheral blood, vitreous humor, cerebrospinal fluid, bile, hair samples, or other relevant tissues. Depending on jurisdiction, whenever the results of ancillary investigations are subject to a criminal investigation and may be used as evidence in a court of law, it may be necessary to prove that there has been an unbroken chain of evidence from sampling through tests to trial. Also possible restrictions owing to national legislation must be considered, when retaining tissues at autopsy.

Sometimes special procedures and modifications of normal dissection techniques are necessary. Imaging techniques or conventional X-ray should be performed before autopsy, if there is suspicion of e.g., air embolism, nonaccidental injury in children, or a

death involving firearms or explosives. Where neck trauma is suspected, the brain and the organs of the chest cavity have to be removed prior to the dissection of the neck, to drain the blood from the area to avoid artifactual bleeding. Postoperative autopsies may present various problems with medicolegal implications, such as complications of anesthesia, surgical intervention, or postoperative care. However, detailed description of these special dissection procedures and techniques is beyond the scope of this presentation.

Autopsy Report

The report is an essential part of the autopsy, but its structure may vary according to the type of autopsy and sometimes may be limited to a list of postmortem diagnoses and conclusions as to the causes of death in the case of a hospital autopsy.

Medicolegal autopsy reports are usually more detailed and structured consisting of a preamble with personal particulars of the deceased, with the date, time, and place of the autopsy as well as the personal data and qualifications of the pathologist and other relevant information to the circumstances concerning the autopsy, summary of the background information about the deceased, findings of the external, internal and supplementary investigations as well as commentary and conclusions with the causes of death and also manner of death depending on the jurisdiction (usually within noncoronial/medical examiner type of medicolegal systems). The report should be unambiguous, comprehensive, and intelligible also to the nonmedical reader. In addition to the factual, positive and negative gross, microscopical and analytical findings, the pathologist should conclude with a discussion of the significance of the findings. Where the findings are of uncertain nature and there are several competing causes, the pathologist should try to give an opinion as to their probability.

References

[1] Ó Neill, Y. V., Sushruta Samhita (The Compendium of Sushruta). *The Lancet* (1912). **180**(4650), 1025–1026.

[2] King, L.S. & Meehan, M.C. (1973). A history of the autopsy. A review, *American Journal of Pathology* **73**(2), 514–544.

[3] Grøn, F. (1934). Lot Sigurd Jorsalfar en av sine menn obdusere i Bysans? *Norsk Magasin for Lægevidenskaben* **95**, 1405–1418.

[4] Rabl, R. (1952). Die Wertung der Sektionen im Wandel der Zeiten, *Virchows Archiv* **321**, 142–162.

[5] Jimenez, F.A. (1978). The first autopsy in the New World, *Bulletin of the New York Academy of Medicine* **54**(6), 618–619.

[6] Tz'u, S. (1981). *The Washing Away of Wrongs*, The University of Michigan Center for Chinese Studies, Ann Arbor.

[7] Cunha, F. (1941). William of Saliceto – The School of Bologna, *American Journal of Surgery*, **52**, 144–149.

[8] O'Neill, Y.V. (1976). Innocent III and the evolution of anatomy, *Medical History* **20**(4), 429–433.

[9] Siraisi, N.G. (1977). Taddeo Alderotti and Bartolomeo da varignana on the nature of medical learning, *Isis* **68**(1), 27–39.

[10] Maksymowicz, K. (2008). Forensic medicine in Poland, in *Forensic Medicine in Europe*, B. Madea & P. Saukko, eds, Schmidt-Römhild, Lübeck.

[11] Ludvig, J. (2002). Principles of autopsy techniques, immediate and restricted autopsies and other special procedures, in *Handbook of Autopsy Practice*, J. Ludvig, ed, Humana Press, Totowa, p. 592.

[12] Nancy, G. Siraisi (2000). Council of Europe, Recommendation no. R (99) 3 of the Committee of Ministers to member states on the harmonization of medico-legal autopsy rules. *Forensic Sci Int*, **111**(1–3), 5–58.

[13] Burton, J.L. & Underwood, J. (2007). Clinical, educational, and epidemiological value of autopsy, *Lancet* **369**(9571), 1471–1480.

[14] Svendsen, E. & Hill, R.B. (1987). Autopsy legislation and practice in various countries, *Archives of Pathology and Laboratory Medicine* **111**(9), 846–850.

[15] Sugiyama, T., Fujimori, T. & Maeda, S. (1991). Autopsy rates in medical schools and hospitals in Japan, *IARC Scientific Publications* (112), 245–252.

[16] Jeganathan, V.S., Walker, S.R. & Lawrence, C. (2006). Resuscitating the autopsy in Australian hospitals, *ANZ Journal of Surgery* **76**(4), 205–207.

[17] Shojania, K.G. & Burton, E.C. (2004). The persistent value of the autopsy, *American Family Physician* **69**(11), 2540–2542.

[18] Yesner, R. (1986). Quality assessment of the autopsy, *American Journal of Clinical Pathology* **86**(2), 250.

[19] Hasson, J. & Schneiderman, H. (1995). Autopsy training programs. To right a wrong, *Archives of Pathology Laboratory Medicine* **119**(3), 289–291.

[20] Saukko, P. & Pollak, S. (2000). Postmortem examination: procedures and standards, in *Encyclopedia of Forensic Sciences*, J.A. Siegel, P.J. Saukko & G.C. Knupfer, eds, Academic Press, San Diego, Vol. 3, pp. 1272–1275.

[21] Saukko, P. & Knight, B. (2004). *Knight's Forensic Pathology*, Edward Arnold, London.

[22] Saukko, P. & Pollak, S. (2005). AUTOPSY, procedures and standards, *Encyclopedia of Forensic and Legal Medicine*, Elsevier, Oxford, p. 166.

[23] The Royal College of Pathologists (2002). *Guidelines on Autopsy Practice*, The Royal College of Pathologists, London, p. 34.

Further Reading

Lu, G.D. & Needham, J. (1988). A history of forensic medicine in China, *Medical History* **32**(4), 357–400.

Related Articles

Cardiac and Natural Causes of Sudden Death
Crime Scene Documentation
Gunshot Wounds
Histology
Natural Causes of Sudden Death: Noncardiac
Report Writing for Courts
Shaken Baby Syndrome
Species Determination of Osseous Remains
Toxicology: Analysis
Wounds: Sharp Injury

PEKKA J. SAUKKO AND STEFAN POLLAK

Autopsy: Psychological *see* Psychological Autopsy

Autoradiograph

Introduction

In the founding 'DNA fingerprinting' technique developed by Sir Alec Jeffreys [1, 2], restriction fragment length polymorphism (RFLP) techniques were used to isolate highly polymorphic minisatellite loci. The DNA fragments produced by restriction endocnuclease digestion were electrophoresed on agarose gels and then transferred to a nitrocellulose or nylon membrane via Southern blotting [3]. Following immobilization to the membrane (by UV fixation for example), an oligonucleotide probe was attached and then labeled with a radioisotope, commonly ^{32}P. Autoradiography was the process used for visualization. The radioactive object (in this case, the labeled probe hybridized to the DNA fragments in the membrane that were themselves produced through RFLP analysis) is placed in contact with high-speed X-ray film. The energy released from the decay products of the radioisotope is absorbed by the silver halide grains in the film emulsion to form a latent image. A chemical development process amplifies the latent image and renders it visible on the film [4]. This technique was time consuming and involved the use of hazardous radioactive reagents and has been superseded by contemporary PCR-based fluorescent detection methods.

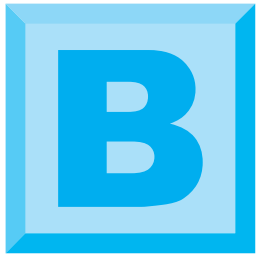
References

[1] Jeffreys, A., Wilson, V. & Thein, S.L. (1985). Hypervariable 'minisatellite' regions in human DNA, *Nature* **314**, 67–73.
[2] Jeffreys, A.J., Wilson, V. & Thein, S.L. (1985). Individual-specific 'fingerprints' of human DNA, *Nature* **316**, 76–79.
[3] Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis, *Journal of Molecular Biology* **98**, 503–517.
[4] Budowle, B., Smith, J., Moretti, T. & DiZinno, J. (2000). *DNA Typing Protocols: Molecular Basis and Forensic Analysis*, Eaton Publishing, Mattick.

Related Articles

Variable Number Tandem Repeats

SIMON J. WALSH



Back Calculation: Alcohol ***see*** Alcohol: Analysis

Ballistics ***see*** Firearms: Overview; Writing Instruments and Printing Devices

Battered Child Syndrome

Introduction

The battered child syndrome is a clinical concept that was first enunciated by Kempe and coworkers in 1962 [1]. The term was used to describe the classic physical findings in children who have experienced severe abuse, often in the hands of adult caretakers. In the years following Kempe's work, the existence of the battered child syndrome was validated by several researchers [2]. A form of the syndrome, which involves life threatening head trauma to very young children, is now known as the shaken baby syndrome (*see* **Shaken Baby Syndrome**).

There have been recent attempts to expand the scope of the battered child syndrome beyond Kempe's classic description of physical abuse to include other forms of severe child maltreatment, such as emotional and verbal abuse. However, such attempts remain controversial [3]. In no other area has the reputation of the syndrome grown than in the legal arena where it has been applied in two broad ways. First, in response to the work of Kempe and his colleagues, there has been greater awareness of child abuse that has led to the promulgation of more stringent child protection laws. Now in many jurisdictions in the world, it is a serious crime to physically abuse children. As a result, prosecutors often use the battered child syndrome to convict perpetrators of severe child physical abuse. An increased awareness of the consequences of the syndrome has also resulted in mandatory child abuse reporting laws that are now in place in all 50 states of the United States (US) of America, including the District of Columbia (DC).

The second legal ramification is that the syndrome has been offered as a justification for parricide by some children accused of murdering their parents [3].

This review covers the origin of the battered child syndrome, its common features, detection, management, and legal implications.

Historical Perspective

The problem of child abuse has plagued mankind since ancient times. For many centuries, maltreatment of children was widely practiced and readily

condoned. Child maltreatment was often rationalized as an effective means of discipline. It was also envisioned as an indispensable tool required for adequate education of children. In addition, in the medieval ages, physical discipline of children was utilized to appease the gods or cast out evil spirits. It was not until the late sixteenth century that some scholars started to write in protest of such maltreatment of children [4]. By the nineteenth century, there was a growing movement in the western world aimed at drawing governmental attention to the plight of the abused child. This led to the adoption of a "Children's charter" by President Hoover in 1930. The aim of this charter was to make every child "wiser, safer, better, and happier", as well as to protect children from "abuse, neglect, exploitation, or moral hazard" [4].

The existence of severe, unexplained, physical injuries in children was long recognized before Kempe and his colleagues published their paper. For example, in 1946, Caffey [5] described a series of cases in which the affected children suffered subdural hematoma (blood in one of the coverings of the brain) and fractures of the long bones. He concluded that although these injuries were traumatic in nature, the mechanism by which they occurred was "obscure". Despite the suspicion that children who presented with unexplained physical injuries might be victims of abuse, physicians were reluctant to get involved and almost never confronted parents when children presented with suspicious injuries. Public interest in the matter was aroused in 1955 by the work of Wooley and Evans [6] who pointed out that some x-ray findings in children with physical injuries were intentional in nature. Later on, Altman and Smith [7] suggested that traumatic injury in some children might have been caused by parents or older siblings. The work of Kempe and colleagues led to greater governmental efforts to protect children from serious physical abuse. This resulted in the enactment of mandatory child abuse reporting laws that have now been adopted by all the states in the United States of America.

Definition

The battered child syndrome is a severe and often fatal form of child physical abuse. In the US, child abuse is defined by the Child Abuse Prevention and Treatment Act (CAPTA) of 1974 as "Any recent act or failure to act on the part of a parent or caretaker which results in death, serious physical or emotional harm, sexual abuse or exploitation, or an act or failure to act which presents an imminent risk of serious harm" to a child under the age of 18 [8]. Kempe and colleagues defined the battered child syndrome as "a clinical condition in young children who have received serious physical abuse, generally from a parent or foster parents" [1]. Kempe and coworkers seemed to suggest that such injuries were the product only of an intentional act. However, Terr and colleagues (as reflected in the US act mentioned earlier) explained that child battering might also result from "an act of carelessness on the part of the parent or guardian" [9].

Prevalence

There are no reliable data that specifically address the prevalence of battered child syndrome as a condition separate from child abuse. However, if the scope is broadened to include all forms of child maltreatment, in the US, according to the National Child Abuse and Neglect Data System (NCANDS) [10], 16.5 children per 1000 were found to be victims of abuse in 2005. Child neglect was the most common form of abuse. Physical abuse, which accounted for 16.6% of all reported cases of abuse, was the second most common form of child abuse. According to the Public Health Agency (PHA) of Canada [11], the rate of substantiated Child Maltreatment was 21.71 per 1000 in 2003, of which 10% involved physical harm. In England, 24 children in every 10 000 were abused in 2001, of which 7300 were victims of physical abuse [12]. According to the United Nations (UN) Secretary General's report on violence against children published in October 2006, about 80–98% of children suffer some form of physical punishment at home, a third of which is severe enough to result in permanent physical impairments. It can thus be inferred from the data presented above that child physical abuse is a serious worldwide problem [13].

Common Features

Kempe and his colleagues reported that the presentation of the battered child syndrome varies depending

on the severity of the physical injury to the child. In mild cases, the condition can be overlooked because the injuries are often subtle [1]. Severe cases are associated with life threatening injuries. The most common manifestations of the condition include skull fracture and subdural hematoma as well as multiple fractures usually involving the long bones. Subdural hematoma, which may occur with or without skull fracture, is the most common cause of death in battered children [1, 14]. Other common injuries found in battered children include, skin lesions such as abrasions, bruises, lacerations, welts, scalds, and major burns [15–17]. Occasionally, battering may involve suffocation, drowning, gunshot wound, stabbing, or poisoning with harmful chemicals and fumes [13, 16]. Smith and Hanson [18] reported in 1974 that in a sample of 134 battered children, 47% presented with intracranial and intraocular hemorrhages (bleeding into the brain and the substance of the eye), 42% had recent or healed fractures while 38% had other injuries such as burns and scalds. In children who present with fractures, the most commonly involved bones are the skull, ribs, the humerus (upper arm bone), femur (bone in the thigh), and the tibia (leg bone) [15, 16].

Another common feature of battered child syndrome is that a large majority of affected children show evidence of neglect, such as failure to thrive or growth retardation (delayed or interrupted child growth). They often appear unkempt, malnourished, and deprived. In one sample, 16% of battered children had experienced neglect, while about 17% were previously hospitalized for failure to thrive [16]. One study reported that 30% of battered children have been exposed to some form of physical or emotional neglect [2].

Finally, in a few cases, the battered children may present with nonspecific medical complications such as fever, anemia, and seizures [14, 15, 17].

Risk Factors for Victimization

Although battered child syndrome can occur at any age, it has been more commonly identified and described in younger children. Several authors report that the most common age when battered child syndrome is suspected is usually sometime between the second and third birthday [1, 14, 17]. Under most circumstances, the affected child might have suffered for several months or years before the abuse is suspected or confirmed. Although some authors report that males are more frequently battered than females, some other studies suggests that there is no sex differentiation in its prevalence [4, 13, 18]. The race distribution of battered child syndrome also varies depending on the study. Some studies report that battered child syndrome is more common in nonwhite (mainly black and Hispanic) children, while other studies suggest a higher prevalence in Caucasian children [4]. There are also studies that report no racial difference in the occurrence of battered child syndrome. Low birth weight babies and children with developmental delay are generally more likely to be battered. In addition, battered children have been reported to have an increased incidence of birth defects including spinal bifida, hydrocephalus, encephalocele, and congenital dislocation of the hips [16].

Battered children are often perceived as difficult to cater for, stubborn, or demanding. They are also less likely to wake up during the night. They are generally “excitable” and “lively” with a tendency to become tired during the day [16]. Affected children may have siblings who also have experienced some form of maltreatment [16, 17].

Profile of the Perpetrator

Available evidence suggests that biological parents are more likely to engage in serious physical maltreatment of children than other caretakers [1, 17]. Other common perpetrators include adoptive parents, stepparents, and foster parents. Siblings, other relatives, and baby sitters are also frequent abusers of children. It is unusual for a stranger to physically assault children because the degree of injury sustained by the children often indicate a close relationship with the perpetrator [19]. In the US, according to NCANDS, one or both parents were involved in 78.9% of child abuse fatalities in 2004 while nonparent caretakers were involved in 10.7% of cases [10].

Both males and females can be perpetrators of battered child syndrome. Some of the older studies suggested that women are more likely to be involved in fatal incidents involving very young children. These findings have been attributed to the fact that women are more intricately involved in the day-to-day care of children. This then increases the possibility that frustrations related to perceived

difficulties or inadequacies in their care of these children often resulting in violent behavior directed at the children [20]. More recent data suggest that fathers and other male caretakers are more likely to be involved in fatal cases of battered child syndrome.

Perpetrators of battered child syndrome are more likely to be young adults, usually in their early to mid 20s [10]. Majority of perpetrators are either high school graduates or high school dropouts. However, battered child syndrome has been reported even in families with highly educated parents. There is no clear racial majority amongst perpetrators of battered child syndrome. There is some indication that perpetrators of battered child syndrome have a higher incidence of psychiatric disorders than the general public. The most commonly implicated psychiatric disorders include mood disorders such as depression and bipolar disorder, psychotic illnesses such as schizophrenia. Often times, the perpetrators are experiencing significant stress in their own personal lives that they then act out on the affected child. They may perceive the child as "illegitimate" or they may attack the child in response to a perception of rejection by the other parent [17].

Detection and Diagnosis

The detection of battered child syndrome requires a high index of suspicion. It might go unsuspected for a long time in cases involving subtle injuries. It should be suspected when a child presents with serious physical injuries that are inconsistent with the explanation provided by the parent or caretaker. It should also be suspected when there is a delay between the time the injury occurred and when medical help is sought for the child [21].

Most cases of battered child syndrome are usually seen for the first time in the emergency room or the office of the family physician or pediatrician. The physician who has first encounter with the child should take a detailed history including a description, by the involved parent or caretaker of the cause of the injury. The history should be "elicited in a nonaccusatory manner" [22]. The parent should generally be interviewed separately from the child whenever possible. Where possible, the physician should obtain collateral information from other members of the family. The involved parent often blames the child or some other persons that live in the same household for the injury. The perpetrator could also explain that the injury is the result of some preexisting medical condition. In some cases, the perpetrator may report that the injury was accidental and explain that they never meant to cause harm to the child. Some may justify their actions by saying that parents have the right to discipline a child in any manner so as to get them to conform to rules [21].

When an accidental cause is given for the injury, the involved parent or caretaker may report that the child accidentally fell from a bed, couch, swing, or countertop. In one instance, I was told by a man suspected of battering his 4-month-old baby to death that the injuries occurred when he accidentally dropped the child after bumping into a doorpost. These accidental falls are often used to explain injuries such as fractures, dislocation, head trauma, and skin lesions. However, given that these falls occur from short distances, the degree of damage done to the child is usually inconsistent with the proposed mechanism of the accident. This is because a lot of these falls occur from distances of about 4 feet or less, while it has been estimated that a fall has to occur from at least 20 feet for it to cause fatal injuries to a child. An alternative accidental explanation for the cause of serious head injury to battered children is that the injury occurred while the perpetrator was gently tossing the child in a playful manner. However it has been shown by some studies that a greater amount of force is required to cause serious injury to a child than is generated during such play. In cases of burn injury, the perpetrator may explain that the injuries resulted from an accidental splash of some hot fluid such as water or that the child accidentally fell in the bathtub. Such claims are often contradicted by the fact that most of the burns are symmetrical and do not show the usual splash marks that would be expected if a child accidentally falls in the bathtub. Also, the severity of the burns is often inconsistent with the mechanism of the burn as proposed by the perpetrator [21].

The involved parent or caretaker could also blame other adults in the house for causing the injury to the child. In such cases, it is often very important to determine who was the last person seen with the child as this would give a clue as to the real perpetrator of the crime. Occasionally, other children (such as the child's siblings) may be blamed for causing the injury to the affected child. In such cases, there is

often a need to establish whether the accused child has the physical capability of causing such injuries to the affected child [21].

The first step in confirming suspected cases of battered child syndrome is to obtain a thorough physical examination, with clear documentation of findings by a trained pediatrician. Physical examination may reveal evidence of previous physical abuse such as skin lesions in different stages of healing. The location of the skin lesions may suggest ongoing physical abuse. Accidental injuries to children often involve the exposed parts of the body such as the shin, knee, and elbow. When injuries affect "protected" parts of the body such as the buttocks, thighs and torso, physical abuse should be suspected. Often times, the pattern of the skin lesions may suggest physical abuse. For example, bruises may show the mark of belts, cords, or other objects used to inflict the injury. Physical examination should also include a specialized examination of the eyes called fundoscopy. This examination helps to establish the presence of retinal hemorrhage (ruptured blood vessels inside the eye) that strongly suggests that the child has been battered [1, 22].

Finally, an x-ray examination of the whole body is usually required to establish the diagnosis of battered child syndrome. This may reveal fractures, usually of the long bones, in different stages of healing. This is the hallmark of the condition as it is indicative of repeated nonaccidental traumatic injury. The most common fractures that have been associated with battered child syndrome include fractures affecting the metaphysis of long bones, posterior rib, scapular, spinous process, and sternum [1, 15, 22].

There are some medical conditions that need to be excluded before the diagnosis of battered child syndrome can be confirmed. These conditions include inherited blood conditions, inherited or acquired skin lesions, infectious skin lesions, congenital metabolic abnormalities, congenital insensitivity to pain, and accidental trauma [22].

Management

The initial goal in the management of battered child syndrome is to terminate the suspected abuse and to prevent further harm to the child. To this end, the first step is to attend to the medical needs of the child. If the child has life threatening injuries such as subdural hematoma and fractures, immediate hospitalization for acute treatment is a must. Failure to urgently address these injuries could result in fatalities. When the child's injuries are minor and nonlife threatening, there is a need to determine whether it is safe for the child to return home with the parents. This decision often requires some collaboration with child protective service agencies. As such, after addressing the child's medical needs, the next step is to report suspected cases to the appropriate governmental agency responsible for investigating and prosecuting child abuse. In the US, all 50 states and the DC have mandatory reporting laws that require professionals (doctors, nurses, social workers, teachers, lawyers, etc.) and other childcare providers to report every suspected case of child abuse to law enforcement agencies. It should be noted that what constitutes child abuse varies from one jurisdiction to another. For this reason, reporting a suspected case of battered child syndrome does not imply guilt until adequate investigation by law enforcement and conviction of the alleged perpetrator in a court of law. Generally, any professional that makes a report in good faith is protected by law from claims of breach of confidentiality or other civil damages though rules may vary with jurisdiction [1, 22].

In most cases, it is unsafe for the children to go back home since the suspected perpetrator is a parent. When such is the case, the child is usually admitted to the hospital for further medical management and to gather more information about the nature and extent of the suspected abuse. Occasionally, there may be a need for child protective services to obtain emergency temporary custody to ensure that the child does not return home with the suspected caretaker.

Once admitted to the hospital, the treatment of battered children requires an interdisciplinary approach. There is often a need to involve different medical specialties, depending on the variation and severity of the injury sustained by the child. Neurosurgical involvement is indicated in cases involving serious head injury, while an orthopedic surgeon would be needed if the child has multiple fractures. Pediatricians or family physicians should handle most of the more general medical care of the child. Social workers and child psychologists should be involved in providing parental counselling, communication with various agencies, and individual therapy for the child. Even after the child has been treated and ready to be

discharged, caution should be exercised, as a premature return of the child to his or her parents' parents, could result in retaliatory abuse that may ultimately be fatal. Sometimes there may be a need to place these children with relatives or in foster care pending the outcome of criminal investigations [1, 22].

Prognosis

According to Kempe and Helfer [23] the prognosis of battered child syndrome depends on the severity of the child's injuries and the "type and quality of intervention" provided once the diagnosis is made. The outcome is generally more positive if the child's medical needs are immediately addressed and if the child is protected from further maltreatment. Permanent brain damage may occur in children with severe head injury. This in turn could lead to mental retardation or poor intellectual/educational performance as the child becomes older. Inadequately attended fractures may result in permanent deformities. The emotional damage resulting from repeated abuse could cause permanent psychic scarring, with the result being a child who has difficulties having empathetic feelings for others. The child's prognosis is often improved when the parents are willing to seek therapy for themselves and the child.

Legal Implications

There are three important legal implications of the battered child syndrome in the US.

1. The battered child syndrome has become a useful tool for law enforcement agencies in the prosecution of suspected cases of serious child physical abuse. This is because the disorder has classical features that are well understood. Moreover, it can be easily documented with physical examination and radiographic evaluations, thus providing strong evidence in court for successful prosecution of suspected perpetrators. Prosecution of suspected cases often require expert testimony. The medical expert testimony is usually needed to show that the affected child did indeed suffer from battered child syndrome. Such expert testimony can be contentious especially when the expert reaches an opinion based on prior nonaccidental physical injuries. This issue was addressed in the US Supreme Court case of *Estelle v McGuire* (1990) [24]. In that case, the Court found that expert testimony based on prior non- accidental injuries did not violate the due process rights of the defendant.
2. Physicians who fail to report suspected cases of battered child syndrome may be successfully sued for medical malpractice. The first case involving such claims was *Landeros v. Flood* [25], a 1976 Supreme Court of California case. In that case, a suit was brought against an emergency room physician who discharged a child to the care of an abusive parent despite evidence that the child was a victim of abuse. The California Supreme Court concluded that contrary to the ruling of a lower court, the child had a claim, since another physician discovered the abuse the next time the child was taken to the emergency room for similar complaints. In a more recent case, *Cooper Clinic v. Barnes* [26], the Arkansas Supreme Court reversed a jury's decision to award $2.5 million to a father for a doctor's failure to report possible abuse by the parents. The jury's decision was overturned because the case was brought against the clinic and not the doctor. The Arkansas mandatory reporting law identified specific individuals who are required to report cases of abuse, but not institutions.
3. The battered child syndrome has been proffered as a defense in some cases of parricide. Most of the cases involved have been contested only at the state level. The US Supreme Court has yet to address the issue of the battered child syndrome as a credible self-defense theory for the murder of a parent by a child. In some of the state cases mentioned above, the argument often involves the admissibility of expert testimony on the battered child syndrome. In the first of such cases, *Jahnke v. State* [3, 27], the trial court's decision to bar expert testimony on past abuse as an explanation for self-defense was upheld by the Wyoming Supreme Court. In another case, *State v. Nemeth* [28] the Ohio Supreme Court found that although there was some controversy involving the use of the battered child syndrome as self-defense justification for parricide, expert testimony should be allowed. More recently, the Minnesota Supreme Court ruled, in *State v. MacLennan* [29, 30], that the admission of

expert testimony should be limited to general description of the syndrome and not on whether the child actually suffers from the condition (*see* ***Daubert v. Merrell Dow Pharmaceuticals***).

Summary

The battered child syndrome is an extreme and often fatal form of child physical abuse. The hallmarks of the condition are well established and most physicians should easily recognize it. A high index of suspicion is often required in cases involving minor injuries. Prompt report to the relevant investigative authority and prevention of further maltreatment of the child is very important once the diagnosis is made. Some of the legal ramifications of the battered child syndrome include prosecution of suspected perpetrators, civil litigation due to failure to report, and its use as justification for a self-defense theory in cases of parricide.

References

[1] Kempe, H.C., Silverman, F.N., Steele, B.F., Droegemueller, W. & Silver, H.K. (1962). The battered-child syndrome, *Journal of the American Medical Association* **181**(1), 105–112.

[2] Lauer, B., Ten Broeck, E. & Grossman, M. (1974). Battered child syndrome: review of 130 patients with controls, *Pediatrics* **54**(1), 67–70.

[3] Hart, J.L. & Helms, J.L. (2002). Factors of parricide: allowance of the use of battered child syndrome as a defense, *Aggression and Violent Behavior* **8**, 671–683.

[4] Helfer, R.E. & Kempe, C.H. (eds) (1968). *The Battered Child*, The University of Chicago Press, Chicago and London, pp. 3–17.

[5] Caffey, J. (1946). Multiple fractures in the long bones of infants suffering from chronic subdural hematoma, *American Journal of Roentgenology* **56**(2), 163–173.

[6] Woolley, P.V. & Evans, W.A. (1955). Significance of skeletal lesions in infants resembling those of traumatic origin, *Journal of the American Medical Association* **158**, 539–543.

[7] Altman, D.H. & Smith, R.L. (1960). Unrecognized trauma in infants and children, *Journal of Bone and Joint Surgery* **42**, 407–413.

[8] Child Welfare Information Gateway (2005). *Child Abuse and Neglect Fatalities: Statistics and Interventions*, Available from, http:// www.childwelfare.gov (accessed June 2006).

[9] Terr, L.C. & Watson, A.S. (1968). The battered child rebrutalized: ten cases of medical-legal confusion, *American Journal of Psychiatry* **124**(10), 126–133.

[10] National Data Analysis System (2005). *Child Abuse and Neglect*, Available on the Internet at, http://www.ndas.cwla.org (accessed Jan 2006).

[11] Public Health Agency of Canada (2003). *Canada Incidence Study of Reported Child Abuse and Neglect (CIS)* Available at, http://www.phac-aspc.gc.ca/media/nr-rp/2005/pdf/cis100405_e.pdf (accessed in 2006).

[12] The Department of Health (2001). *Children and Young People on Child Protection Registers*, England, Available at, http://www.performance.doh.gov.uk/HPSSS/TBL_C2.HTM (accessed in 2006).

[13] United Nations (UN) *The United Nations Secretary General's Study on Violence Against Children*. Available at, http://www.violencestudy.org (accessed in 2006).

[14] Brown, R.H. (1975). The battered child syndrome, *Journal of Forensic Science* **21**(1), 65–70.

[15] Behrooz, A., Torg, J.K. & Sussman, S. (1974). Manifestations of the battered child syndrome, *The Journal of Bone and Joint Surgery: American Volume* **56**, 1159–1166.

[16] Smith, S.M. & Hanson, R. (1974). 134 battered children: a medical and psychological study, *British Medical Journal* **3**, 666–670.

[17] Sussman, S.J. (1968). The battered child syndrome, *California Medicine* **108**(6), 437–439.

[18] Helfer, R.E. & Kempe, C.H. (eds) (1968). *The Battered Child*, The University of Chicago Press, Chicago and London, pp. 19–40.

[19] Helfer, R.E. & Kempe, C.H. (eds) (1968). *The Battered Child*, The University of Chicago Press, Chicago and London, pp. 33–34.

[20] Helfer, R.E. & Kempe, C.H. (eds) (1968). *The Battered Child*, The University of Chicago Press, Chicago and London, pp. 31–32.

[21] Kreston, S.S. (2002). Prosecuting battered child syndrome, *National District Attorneys Association* [Online], **15**(7), Available at http://www.ndaa-apri.org/publications/newsletters/update_volume_15_number_7_2002.html (accessed in 2006).

[22] Pressel, D.M. (2000). Evaluation of physical abuse in children, *American Family Physician* **61**(10), 3057–3064.

[23] Kempe, H.C. & Helfer, R.E. (eds) (1972). *Helping the Battered Child and his Family*, J. B. Lippincott Company, Philadelphia and Toronto, pp. 69–115.

[24] Estelle v. McGuire, 502 U.S. 62 (1991).

[25] Landeros v. Flood, 17 Cal.3d, 399 (1976).

[26] Cooper Clinic v. Barnes, Arkansas 05–1166 (2006).

[27] Jahnke v. State, 682 P.2d 991 (1982).

[28] State v. Nemeth, 82 Ohio St (1998).

[29] State v. MacLennan, 702 N.W.2d 219 Minnesota.3d, 202 (2005).

[30] Amato, J.M. (2006). The battered-child syndrome, *Journal of American Academy of Psychiatry and the Law* **34**(3), 414–415.

SYLVESTER SMARTY

Battered Spouse Syndrome

The legal concept of battered woman syndrome (BWS) evolved from Walker's [1] initial description and is used in the defense of individuals who kill an abusive partner. Walker defined *BWS* as having three major components: a specific "cycle of violence", learned helplessness, and typical psychological sequelae. The "cycle of violence" begins with the "tension" phase, in which tension builds in the abusive partner. Second is the acute battering incident that serves to rid the aggressor of his built-up tension. Following the acute aggression phase is the "honeymoon" or "contrite" stage during which the batterer shows remorse for the aggression. This ends when the tension phase renews. Walker [1, 2] also theorized BWS as encompassing learned helplessness behavior, suggesting that the victim becomes more passive as she sees fewer solutions to her abusive situation. Finally, Walker noted several common psychological characteristics of battered women, including low self-esteem, guilt, denial of feelings of terror and anger, and traditionalist attitudes regarding gender roles, among others.

Since Walker's publications, systematic studies of the effects of battering on women have been done (e.g., [3]); however, Walker's theory continues to largely inform legal cases concerning battered women. Currently, the more inclusive phrase "battered s*pouse* syndrome" is sometimes used in an effort to recognize the occurrence of woman on man, woman on woman, and man on man intimate partner abuse (see e.g., [4]) (*see also* **Aggression: Gender Differences in**).

Legal Use of Battered Woman Syndrome

Initially, BWS evidence was used to support temporary insanity defenses and was successful in obtaining some acquittals. One of the first and most well known of these cases is the "burning bed" (*Michigan v Hughes*, 1977). In this case, Francine Hughes was physically, verbally, and sexually abused by her husband. After attempts to leave and divorce him failed, she set fire to the bed in which he was sleeping, killing him. She was found not guilty by reason of insanity based on BWS evidence [5].

Because most battered women fail to meet the cognitive impairment criterion needed for defenses of temporary insanity, BWS is now rarely used as part of an insanity defense. Instead, BWS has been used as an adjunct to self-defense claims [6]. Often, the killing does not occur in the course of an acute battering episode. In these cases, BWS has provided a "bridge" between the context of the killing and claim of self-defense. Walker's [1, 2] theory suggests that women perceive imminent danger, even when she is not actively incurring physical abuse. Battered women are said to be able to "read" their abuser and discern danger even when physical violence is not occurring. This aspect of BWS legitimizes a battered woman's killing as a form of self-defense when she kills outside of an acute battering episode [1, 2, 7]. In the context of self-defense, testimony regarding BWS is used to increase the credibility of an individual's claims of abuse and to support the reasonableness of the woman's actions [8]. These aims can be met via witness and defendant testimony regarding the abuse. They are also met using experts who testify regarding BWS.

Admissibility and Effect of BWS Testimony

As of 1996, all states had admitted testimony of BWS in some form [9]. When expert testimony regarding BWS has been rejected it has typically been on grounds that (i) it is irrelevant to claims of self-defense, (ii) it is unnecessary in aiding jurors to understand the situation of battered women, (iii) it is not well-supported by the scientific community, or (iv) its probative value is outweighed by its potentially prejudicial effects [8].

In the 1990s, public dismay was expressed regarding the "abuse excuse" [10, 11]. This largely stemmed from the still-present perception that women who kill their abusers are often "let go" without repercussion for their actions. In reality, most battered women who kill plead guilty to or are convicted at trial of serious charges, often murder or manslaughter, and serve significant time in prison [8]. In an examination of 100 battered women who killed their spouses Ewing found that nine plead guilty to charges of murder, manslaughter, or criminally negligent homicide. In 28

cases, the women were not sentenced to jail because of acquittal, findings of not guilty by reason of insanity or dropped charges. The remaining 63 women went to trial and were found guilty of charges ranging from reckless homicide to first-degree murder.

Concerns Regarding the Legal Use of Battered Woman Syndrome

The first problem surrounding BWS evidence is its questionable fit within self-defense doctrine. Self-defense implies, among other elements, that the individual believes he or she is in imminent danger of sustaining bodily harm. The fear of imminent danger must be reasonable, in that a "rational" individual would also perceive imminent harm [12]. Many women kill their spouses outside an acute battering incident but use "cycle of violence" and learned helplessness theories to fit their actions into the self-defense paradigm. The legitimacy of using BWS to bolster claims of self-defense has been questioned and fears of stretching self-defense doctrine immeasurably have been raised (e.g., [7]). Ewing [8] suggests an alternative to continuously broaden the scope of self-defense to include incidents of battered spouses who kill. He proposes the "psychological self-defense", in which the use of deadly force would be legitimate in cases where an individual acts to protect his or her psychological well-being. Ewing's theory purports that a reasonable person could discern imminent risk from her history of interactions with the abuser. If recognized by the courts, the psychological self-defense would preserve the spirit of self-defense laws but allow for the context of the battered woman's actions to be presented in the court.

Second, concerns have been voiced that BWS is more advocacy than science. Some experts testify regarding BWS in cases that clearly are outside the realm of self-defense, thus, misapplying the syndrome and stretching their credibility as experts [9]. In other words, while intimate partner abuse should be recognized by the courts, it should not come at the expense of "good science."

Third, similar to other syndromes (e.g., rape trauma syndrome and homosexual panic) the validity of BWS has been challenged on the basis that its presentation lacks specificity; moreover, research has not supported the notion that battering causes a unified set of symptoms in the majority of women. Legally, the theories of learned helplessness and the "cycle of violence" proposed by Walker are the most recognized and utilized aspects of BWS [13]. Many researchers deny that there is sufficient evidence that battered women experience the three stages she proposes (e.g., [7]). Several others (e.g., [7]) have questioned Walker's methodology in studying the effects of battering. Moreover, Walker's own data showed that 62% of battered women's relationships did not include the entire "cycle of violence" [9].

Similarly, researchers have disputed the occurrence of learned helplessness, emphasizing that many battered women use physical, verbal, and legal means to escape abuse [7, 9]. Further, killing one's abusive partner is in direct opposition with learned helplessness. Research has also shown that battered individuals present with varied symptoms, some of which directly oppose or are not specified under Walker's characterization of BWS. Because of the varied presentation of battered women, some have even argued that BWS does not meet the definition of a "syndrome" [6].

Finally, in most states, decisions regarding the admissibility of scientific evidence and expert testimony are based on the criteria set forth in the United States Supreme Court decision in *Daubert et al. versus Merrell Dow Pharmaceuticals, Inc.* (1993). *Daubert et al.* sets forth that, in order to be admissible in court, scientific evidence must be: (i) amenable to and have been scientifically tested, (ii) subject to peer review, (iii) have a known error rate, (iv) have limits regarding its definition, and (v) have widespread acceptance within the relevant scientific community. While BWS has been the subject of research and peer review, it lacks definitional specificity and is not widely accepted (in a unified form) by the scientific community. Thus, the appropriateness of presenting BWS evidence in court is questionable.

Conclusion

BWS describes the characteristics of some abused women and is often used to support claims of self-defense when individuals kill an abusive partner. Currently, courts rely heavily on Walker's conceptualization of BWS; however, many challenges have been made to this definition. It has been argued that BWS lacks definitional specificity and that it is inadequate to describe most battered women. BWS has

also been challenged on the grounds that it should not be used to support claims of self-defense and that it is not *Daubert*-worthy.

References

[1] Walker, L.E. (1979). *The Battered Woman*, Harper & Row, New York.

[2] Walker, L.E. (1984). *The Battered Woman Syndrome*, Springer, New York.

[3] Dutton, D.G. (1995). *The Domestic Assault of Women: Psychological and Criminal Justice Perspectives*, UBC Press, Vancouver.

[4] Hatcher, G.R. (2003). The gendered nature of the battered woman syndrome: why gender neutrality does not mean equality, *Annual Survey of American Law* **59**, 21–50.

[5] McNulty, F. (1980). *The Burning Bed: The True Story of Francine Hughes – A beaten wife Who Rebelled*, Harcourt Brace Jovanovich, New York.

[6] Follingstad, D.R. (2003). Battered woman syndrome in the courts, in *Handbook of Psychology: Forensic Psychology*, A.M. Goldstein, ed, John Wiley & Sons, Hoboken, Vol. 11, pp. 485–507.

[7] Faigman, D.L. (1986). The battered woman syndrome and self-defense: a legal and empirical dissent, *Virginia Law Review* **72**, 619–647.

[8] Ewing, C.P. (1987). *Battered Women Who Kill: Psychological Self-defense as Legal Justification*, Lexington Books, Lexington.

[9] State Justice Institute (1996b). Validity of "Battered woman syndrome" in criminal cases involving battered women, *The Validity and Use of Evidence Concerning Battering and its Effects in Criminal Trials: Report Responding to Section 40507 of the Violence Against Women Act*, U.S. Departments of Justice and Health and Human Services (NCJ Publication No. 160972), U.S. Department of Justice, Rockville.

[10] Dershowitz, A.M. (1994). *The Abuse Excuse*, Little Brown, Boston.

[11] State Justice Institute (1996a). Trend analysis: Expert testimony on battering and its effects in criminal cases, *The Validity and Use of Evidence Concerning Battering and its Effects in Criminal Trials: Report Responding to Section 40507 of the Violence Against Women Act*, U.S. Departments of Justice and Health and Human Services (NCJ Publication No. 160972), U.S. Department of Justice, Rockville.

[12] LaFave, W.R. & Scott, A.W. (1986). *Criminal Law*. West Publishing, St. Paul.

[13] Biggers, J.B. (2003). A dynamic assessment of the Battered Woman Syndrome and its legal relevance, *Journal of Forensic Psychology Practice* **3**, 1–22.

TRACY A. THOMAS, CARL CLEGG AND WILLIAM J. FREMOUW

Battered Woman Syndrome *see* Battered Spouse Syndrome

Battered Woman's Reality

Battered Women's Syndrome: Woman's Reality

Some 20 years ago in *Zecevic v DPP*, a judge of the highest Australian court stated that the question to be asked in determining if an individual acted in self-defense was "whether the accused believed upon reasonable grounds that it was necessary in self-defense to do what he did" [1]. In looking at the judicial outcomes for battered women who killed a violent partner in Australia over the past two decades, it is evident that they often confront a legal system that anticipates the factual paradigm of a male–male confrontation, and a court that is accustomed to interpreting "reasonable grounds" in terms of a "barroom brawl" scenario and not within a context of domestic violence antecedents.

Why Expert Evidence is Necessary

The term "*domestic violence*" can underestimate seriousness and criminality of violence within the home (*see also* **Battered Spouse Syndrome**). "Domestic" can imply private or relatively unimportant in contrast to the "real" world outside the home. This is further illustrated by the words used by some judges as umbrella labels to describe in their judgments the assaults against a woman that her defense had presented in arguing for self-defense or provocation. Phrases such as "problems of a marriage", "domestic dispute", "difficult relationship", "domestic argument", "matrimonial discord", and "stormy" relationship tend to trivialize the violent antecedents and neutralize the role of perpetrator by allocating responsibility to both parties [2].

Consequently, in most of the Australian cases in which battered women have been acquitted to date, the homicide took place immediately (as defined in seconds) after their partner assaulted them [3]. There

are exceptions such as the acquittal of a defendant who shot her violent partner in the back. Her defense team argued that the shooting was a preemptive strike by using the state of Queensland's code definition of an assault: one can be defending oneself against "situations in which violence is merely threatened as long as there is actually or apparently a present ability to implement the threat . . ." [4]. Critically (and atypically), the judge had instructed the jury that the assault that the accused was defending herself against was the general nature of the relationship and all of the threats made toward her over an extended time period.

Accordingly, the view is held by some that the current Code provisions in Australian jurisdictions should be able to incorporate the context of long-term domestic violence without law reform but that "many solicitors, judges, and/or juries do not possess the requisite knowledge of the dynamics of domestic violence to be in any position to judge the availability of the defense generally and its applicability in the particular case" [5]. Thus, expert evidence is necessary to properly convey the social reality of the battered woman [6].

Expert evidence can serve another purpose. Direct questioning, fact-finding, and impartial/impersonal communication styles come at the expense of a more trusting and sensitive environment in which the women can feel safe when telling their stories. Beside aboriginality or immigrant status, both gender and the reality of living in a violent relationship can make such a style of lawyering and adjudication riddled with problems for the battered woman. She often lacks the self-confidence to present what society defines as credibility. The expert can not only help judges and jurors to learn more about the violent antecedents in the specific case but also by substantiating her evidence, the battered woman's testimony may be seen as more credible.

The main way that such expert testimony has been introduced in courts has been through raising or introducing battered woman syndrome (BWS). However, this psychologizing of women's experiences has not been without its detractors in Australia.

Objections to BWS Evidence

Some are concerned that BWS is constructed as a psychological affliction, which "affects the way you perceive things" [7]. The woman's individual state of mind and the medically explicable "irrationality" of her actions are the focus of medical and psychological testimony. Instead of explaining her action as the result of rational self-preservation decision-making, BWS presents the killing as a decision made from a state of "learned helplessness" or dependency [8].

In Australia this has been even more the case with an emphasis on the abnormality or the subjective aspect of BWS as compared to Canada where BWS evidence has been adduced to explain the objective component of self-defense – why a particular accused's response was ordinary or reasonable given the circumstances of violence the woman had survived [9].

Another concern is the potential risk that a formal or informal test of the "reasonable battered woman" will evolve that categorizes the effects of domestic violence into the traits of a singular personality type. This makes it even more difficult for courts to interpret and understand the actions of battered women whose personalities deviate somewhat from this stereotype; those who were not passive in their relationship and/or those who fought back, and/or had sought outside help. For that reason, current descriptions and applications of the syndrome have been criticized for isolating gender as a category of analysis and determinant of behavior while ignoring other systems of power and identity.

Specifically, there is concern that BWS denies that for many, the experience of being a woman, and in particular a battered woman, is incomplete without incorporating their experiences of race, religion, or class. For instance, BWS has not usually covered the particularities of being a battered Torres Strait Island woman or Aboriginal woman. The possible divergence of an indigenous woman's experiences from the stereotype of the dependent, irrational battered woman is evident, particularly with regard to their roles as "heads of households, responsible for the financial and the emotional survival of their families as primary kin-keepers" [10]. As a result, many do not conform to a psychological profile of dependency and passivity.

Perhaps then the critical question is not whether expert evidence is necessary, but the nature of the expertise and evidence. As discussed next, it is possible to use people, either from the "coalface" or academics, who do not syndromize but situationalize [11].[a] The main obstacle in using battered woman's reality BWR is the incompatibility of a traditionally

male dominated court and the acceptance of refuge workers and other nonmedicos or nonscientists as expert witnesses (*see* **Alcohol: Behavioral and Medical Effects**).

One Australian jurisdiction, Victoria has made significant reforms to its Crimes Act (1958) in 2005, clearing a pathway for evidence of family violence and its effects.[b] In other jurisdictions the admission of evidence about domestic violence is left to the magistrate or judge. Since expert testimony is generally only admitted if the subject contains material beyond the ken of the jury, the use of the label "syndrome" may be critical for admissibility.

Battered Woman's Reality (BWR) Evidence

Having nonpsychologizing experts explain the experience of living in domestic violence can assist the jury and the judge in understanding what is reasonable behavior in a domestic violence situation. Diversity of responses to ongoing violence is emphasized with the aim neither to further reinforce stereotypes of women as helpless or passive nor to pathologize their action.

The expert's report may commence with a section on domestic violence in general. The situational expert explains that the domestic violence ending in death more often contains alcohol abuse, death threats, threats with weapons, more severe battery, and sexual violence toward the woman [12–15]. She/he describes the effects of living under the constant threat of violence, categorizing the violence and its impact as "battered woman's reality." If relevant to the particular facts, the court learns that these homicides are usually preceded by an unusual incident: something done by the male that is outside his customary repertoire of violence. He may either have threatened the children's lives, began to sexually assault one or more of the children, or the mother might have just learned about the latter. This has been referred to as "*the turning point*". Something occurs that simply extends beyond the boundaries of what the woman is accustomed to [16].

Then, from a scrutiny of all pertinent records of interview, the expert report can focus on the case at hand, highlighting the pertinent background variables such as a history of childhood abuse, battering increasing in severity over time, intermittent periods of contrition and relative "peace" in the relationship, sexual abuse, death threats, and a precipitating event.

The expert may also be able to discern and label violent behavior by the batterer that has not been named as such by the defendant. For example, in one case, in the records of interview with the police and with the solicitor, the woman was asked specific questions about antecedent abuse.

"Were you hit?"

"Was your husband physically violent toward you?"

She responded "no" to these and other direct questions. The expert though identified a pattern of ongoing physical violence inflicted by the deceased. In the transcript of an interview with her lawyer, the defendant had mentioned that when she slept later than her partner, he squeezed her breasts and nipples so hard that it brought tears to her face. In addition, it was not uncommon for him to jump on her as she lay on the bed. These were not actions that the defendant labeled as hitting or bashing. In part, that is owing to the language and the different images that specific words such as "violent" conjure for different people, which is why screening questions need to be designed very carefully. And, in part, it can be attributed to one of the effects of ongoing violence. It may become so normative that its victims trivialize or normalize it to the point of invisibility. Thus, the situational expert may penetrate the family system's denial, naming not only some physical acts as violence but also emotional, economic, and sexual patterns of control.

How the Situational Expert Addresses the Issue of Immediacy Answering the Question, "How Can It be Self-Defense if He was Asleep or Passed Out?"

The BWR expert discusses how the woman's action was preemptive, explaining that if she did try to defend herself seconds after the violent partner had attacked, she would probably end up seriously injured or dead. Such an "immediate" response is not reasonable behaviour for someone who is, in all likelihood, neither as strong nor aggressive, not as skilled in assault, ans particularly a person who has been beaten down (externally and internally) usually for an extended period of time. Further, immediacy implies that there is one discrete precipitating incident. The expert explains that such a construction is . . .

> . . . (W)holly inappropriate because the danger that women who are habitually and seriously abused face

> is not so much embodied in a single attack as in the day to day experience of living under continuous threat [17].

Additionally, through describing the dynamics of domestic violence, the expert can help the court to understand that for the battered woman, the threat may have felt immediate. The violence is often erratic and therefore unpredictable. This unpredictability can be conducive to feelings of terror – simply not knowing when the batterer will attack.

How the Situational Expert Responds to the Issue of "No Other Recourse" Answering the Question, "Why didn't She Just Leave?"

The situational expert explains that for a battered woman, leaving the relationship may not be easy. It seems that the worse the violence inflicted, the harder it is for many to comprehend the woman's inertia or her return. Yet paradoxically, the worse the violence, the greater her inability to leave. The BWR evidence may equate the violent home with a political hostage context; most people can understand that the more serious the psychological and physical brutality, the greater the capacity of the captors to exert control. The expert shows that these controlling behaviors are often evolutionary and insidious. Disempowerment is perpetuated through humiliation and degradation and the woman's increased shame and isolation. The latter is often at least in part the consequence of the violent partner's exertion of social control – forcing the woman to cut her ties with family and friends.

Further, leaving the relationship does not necessarily mean the violence stops. In fact, leaving can place her life in more danger as the male feels his power is threatened. Accordingly, one fourth of the intimate partner homicides in Australia from 1989 to 2002 involved estranged couples; in 84% males were the perpetrators [18].

How the BWR Expert Explains the Woman's Motives and Emotions to the Court as a Normal Response to the Situation, Answering the Question, "Why is not She Remorseful?"

One less visible barrier to the battered woman's use of self-defense may be her lack of what are considered as "reasonable" emotions such as remorse that accompany or follow the homicide. It has been argued that self-defense may not in fact be appropriate for all battered women who kill since in some instances they may have an affect that does not correlate with the "reasonable subjective state of mind" to accompany an act of self-defense [19]. However, an argument that seeks "reasonable" motivation and affect fails to understand the unique dynamics of domestic violence and the intricacies of the relationship between the abuser and his "hostage" (*see* **Stockholm Syndrome**).

The BWR expert explains that if a political hostage killed his captor after enduring years of emotional and mental torture and terror, most people would accept that it would be "reasonable" for the hostage to be experiencing a wide gamut of emotions at the time of killing. Self-preservation and the desire for freedom and safety could easily be accompanied by a bonding with the perpetrator, anger, vindictiveness, and revenge.

In conclusion, this article offers a fact-based alternative to psychodynamic-based expert testimony that assists the trier of fact not through theories related to behavior syndromes but through background information that places conduct in a factually appropriate context.

End Notes

[a.] For example in *R v Gadd* (Unreported Qld Sup Ct, 31 March 1995) a social worker testified. Also the author, identifying as a sociolegal academic has provided expert reports in Family Court, State Supreme Courts, Immigration Tribunal and in *Winnett v Stephenson* (ACT Magistrates Court, unreported, 19 May 1993).

[b.] Section 9 AH eliminates the immediacy and proportionality requirements and also states that in circumstances where family violence is alleged a person may believe, and may have reasonable grounds for believing, that his or her conduct is necessary ...

References

[1] *Zecevic v DPP* (1987). 162 Commonwealth Law Reports 645.

[2] Easteal, P. (2003). Violence against women in the home: kaleidoscopes on a collision course?, *QUT Law and Justice Journal* **3**(2), 250–273. See at http://www.law.qut.edu.au/about/ljj/editions/v3n2/index.jsp.

[3] For a review of the recent research throughout Australia see Roth, L. (2007). *Provocation and Self Defence in*

Intimate Partner and Homophobic Homicides, Briefing paper No 3/07, NSW Parliamentary Research Service.
[4] *R v Sternquist* (unreported, Cairns Circuit Ct, 18 June 1996, Derrington J) S 245, discussed by Colvin, E., Linden, S. and Bunney, L. (1998). *Criminal Law in Queensland and Western Australia*, 2nd Edition, Butterworths, p. 252.
[5] Kift, S. (2001). Defending the indefensible: the indefatigable Queensland criminal code provisions on self-defence, *Criminal Law Journal* **25**, 28.
[6] Schuller, R. (2004). *The Impact of Expert Testimony in Trials of Battered Women Who Kill Psychiatry, Psychology and Law* found that respondents in Australia and Canada were more likely to support a woman's claim of self-defense if they had been provided with either BWS or situational expert testimony. For the two measures on which there was a differential impact on the basis of the nature of the testimony, it was the situational, or what Schuller calls the social/agency evidence that produced a more favorable outcome than BWS.
[7] Martin., C.J., Angel and Mildren, J.J. (1996). As stated by the defense lawyer in the Northern Territory case of *Secretary v DPP* (unreported, NT CCA, 2 April 1996), per, p. 13.
[8] Stubbs, J. & Tolmie, J. (1994). Battered woman syndrome in Australia, *Women, Male Violence and the Law* 210.
[9] Stubbs, J. & Tolmie, J. (1995). Falling short of the challenge? A comparative assessment of the Australian use of expert evidence on the battered woman syndrome, *Melbourne University Law Review* **23**(3), 709.
[10] Pettman, J. (1992). *Living in the Margins: Racism, Sexism and Feminism in Australia*, Allen & Unwin, Sydney, p. 65.
[11] Easteal, P., Hughes, K. & Easter, J. (1993). The reasonable battered woman and duress: educating the judiciary, *Alternative Law Journal* **18**(2), 139.
[12] Browne, A. (1987). *When Battered Women Kill*, The Free Press, New York.
[13] Ewing, C. (1987). *Battered Women Who Kill*, D.C. Heath and Co.
[14] Walker, L. (1989). *Terrifying Love*.
[15] Easteal, P. (1993). *Killing the Beloved*.
[16] Blackman, J. (1989). *Intimate Violence*, Columbia University Press, New York.
[17] Stubbs, J. & Tolmie, J. (1994). Battered woman syndrome in Australia, *Women, Male Violence and the Law* 196.
[18] Mouzos, J. and Rushford, C. (2003). Family Homicide in Australia, *Trends and Issues in Crime and Criminal Justice*, at http://www.aic.gov.au/publications/tandi2/tandi255.html.
[19] Hubble, G. (1997). Feminism and the battered woman: the limits of self-defence in the context of domestic violence, *Current Issues in Criminal Justice* **9**(2), 113.

PATRICIA EASTEAL

Bayesian Networks

Introduction

Forensic literature has pointed out the utility of methods that deal with formal analysis of inference and decision making [1]. Notably, it has been underlined that complex frameworks of circumstances, situations involving many variables, require a logical assistance [2]. Methods of formal reasoning have been proposed to assist the forensic scientist to understand all of the dependencies that may exist among different aspects of the evidence [3]. Graphical models, in a probabilistic environment (i.e., conducted through the use of a formalism known as *Bayesian networks* (*BNs*)) provide a valuable aid for representing relationships among characteristics in situations of uncertainty. They assist the user not only in describing a complex problem and communicating information about its structure but also in calculating the effect of knowing the truth of one proposition or piece of evidence on the plausibility of others. Moreover, the graphical nature of such models facilitates the formal discussion and clarification of probabilistic arguments [4].

Bayesian Networks

Definition

A BN is a compact model representation for reasoning under uncertainty that formally combine elements of graph and probability theory. BNs allowing their user to define a pictorial representation of assumed probabilistic relationships among a set of variables,[a] deemed to be relevant for a particular inferential problem.

In a BN, random variables are represented by nodes,[b] whereas directed edges (arcs) are used to express assumed relationships among the nodes. Both nodes and edges are combined in order to form a directed acyclic graph (DAG), or a digraph for short. Assuming discrete variables, the strength of the relationship between the variables is quantified by conditional probability distributions associated with each node: for a variable B with parents $A_1, \ldots, A_n$, there is a conditional node probability table $\Pr(B|A_1, \ldots, A_n)$, whereas a table containing unconditional probabilities $\Pr(A)$ is assigned to a

variable A that has no parents. If the conditional relationships implied by the structure of a BN hold for a set of variables $A_1, \ldots, A_n$, then the joint probability distribution $\Pr(A_1, \ldots, A_n)$ is given by the product of all specified conditional probabilities:

$$\Pr(A_1, \ldots, A_n) = \prod_i \Pr(A_i \mid \text{par}(A_i)) \qquad (1)$$

where par(A_i) represents the set of parental variables of A_i. Equation 1 is called the *chain rule* for BNs [5, 6] and formally defines what a BN means: a representation of the joint probability distribution for all the variables. An example is presented at the end of next section.

Structure

For the building of BNs, it is important to remind that they do not represent the flow of information, but serve as a direct representation of a part of the real world [6]. This means that through the use of a BN, a human expert can graphically and numerically articulate his subjective (not arbitrarily) view of a real world system. Therefore, the task of modeling as well as a model that is obtained as a result of this process will principally be influenced by both, the properties and the experts' individual view, perception, and ultimately, extent of understanding, of a domain of interest. As Lindley [7] wrote:

> A model is merely your reflection of reality and, like probability, it describes neither you nor the world, but only a relationship between you and that world. It is unsound to refer to the true model.

Even if the finding of an appropriate representation of a case under examination seems to be, to some extent, an art form, note that it can be guided by scientific and logical considerations.

The basic building blocks of Bayesian nets are the three possible elementary connections among three nodes A, B, and C (see Figure 1).

Such a pictorial scheme allows the scientist to directly translate qualitative information into dependence relations among variables under specified circumstances, specifying the relevance of the variables in the context of the case.

It is necessary to examine the properties implied by the structures. This would mean that it is required to observe concepts such as *d*-separation properties,

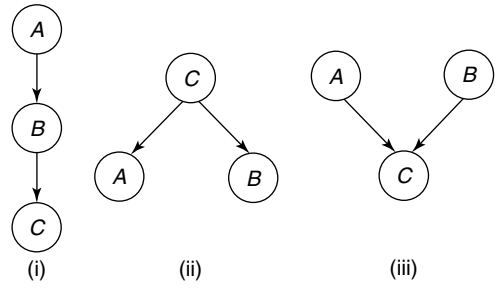


Figure 1 Basic connections in Bayesian networks: (i) serial, (ii) diverging, and (iii) converging connection

that stands in close relation to the notions of dependence and (conditional) independence. The point is that by drawing a certain graphical structure whatsoever, we will automatically be encoding such concepts in our representation. This is why we should be aware of their significance, and, furthermore, we are required to provide justifications for them, if we would like to argue for the appropriateness of the structure we propose.

d-separation, where *d* denotes directional, is a graphical criterion [8] that designates the blocking (or stopping) of the flow of information (or of dependencies) between variables that are connected through a sequence of arcs independent from the direction of the arrows (a chain). Consider this concept in case of the three basic connections presented in Figure 1.

- In serial and diverging connections, a path is said to be d-separated if the middle variable is instantiated.[c] In a serial connection, let A be the proposition that the suspect is the offender, B the proposition that the blood stain found on the crime scene comes from the suspect, and C that the suspect's blood sample and the blood stain from the crime scene share the same DNA profile. Then A is relevant for B and B for C but, given B, the cause of the presence of blood could be different from A. Therefore, evidence may be transmitted unless the state of B is known. Diverging connection is appropriate when we judge that knowledge of the truthstate of a certain event A provides relevant information for another event B and that knowledge of the truthstate of a third event C, which is relevant for both A and B, screens off A from B. For example, let C be the suspect who has been in contact with the victim, A the bloodstain on the suspect's clothes that

comes from the victim, and B the bloodstain on the victim that comes from the suspect. Evidence may be thus transmitted unless the state of C is known.

- In converging connections, a path is called *d-separated* as long as the intermediate variable, or one of its descendants, has not received evidence. For opening the chain, it is not necessary that the intermediate variable is instantiated. For example, let A be the proposition that the suspect is the offender and B the proposition that the bloodstain found on the scene of the crime comes from the offender: knowledge that one of these events occurred would not provide information about the occurrence of the other, but if it is true that C (the blood stain found on the crime scene comes from the suspect), then A and B become related. Converging connections in Bayesian nets are particularly important, because they represent a very common pattern of reasoning: conditional dependence or explaining away [9]. C has been observed; then if A is observed, it explains the observation of C. There is no longer the need to invoke other 'causes'.

Stated otherwise, if two variables in a network are d-separated, then changes in the value (known in the context as *truthstate*) of one variable will have no impact on the truthstate of the other variable. If two variables are not d-separated, they are called *d-connected* [6].

Consider the chain rule in case of the three basic sequential connections that are possible in BNs (see again Figure 1). For a path from A to C via B, as shown in Figure 1(i), $\Pr(A, B, C) = \Pr(A) \cdot \Pr(B|A) \cdot \Pr(C|A, B)$ can be reduced to $\Pr(A, B, C) = \Pr(A) \cdot \Pr(B|A) \cdot \Pr(C|B)$. For a diverging connection, the joint probability can be written as $\Pr(C, A, B) = \Pr(C) \cdot \Pr(A|C) \cdot \Pr(B|C)$, whereas in a converging connection it would be $\Pr(A, B, C) = \Pr(A) \cdot \Pr(B) \cdot \Pr(C|A, B)$.

Reasoning

It can be further explored how to use a BN to reason about the domain. In particular, when we observe a new information (a state of a variable in the net is instantiated), it is of interest to observe its effect on other nodes (i.e., query nodes). This process is called *inference* and it is relied to the computation of posterior probability distributions for these query nodes (*see* **Evidence Interpretation: a Logical Approach**).

Given that BNs offer a representation of probability distributions over variables, this imply that any set of variable can be instantiated and their effect of other variables can be measured as schematically presented in Figure 2. For the sake of illustration, the terms *cause* and *effect* are introduced to describe a simple two-nodes BN: cause → effect.

- The cause 'produces' the effect, i.e., knowing that the cause happened, it can be foreseen that the effect will occur or might probably occur, too. This is a predictive line of reasoning. For example, given that the suspect is the source of the stain (say, a binary variable S with states true instantiated), even before any laboratory tests have been performed, the probability of a match between the stain's DNA profile and the suspect's profile (say, variable E) increase (note that if S is known, then the probability of E is just given by the specified probability $\Pr(E \mid S)$ in the conditional probability table relates to node E). Note that this estimate represents the numerator of the likelihood ratio. By the instantiation of state false of variable S, scientists obtain the denominator of the likelihood ratio (*see* **Case Assessment and Interpretation**).
- The effect does not "produce" the cause, but knowing that the effect occurred, one may infer that the cause probably occurred. This is a line of reasoning against the causal direction that can also be termed *diagnostic*. For example, given that the test shows a match between the stain's DNA profile and the suspect's profile, then the belief that the suspect is the source of the stain is updated. If E is known, probability of S is updated *via* Bayes' theorem.

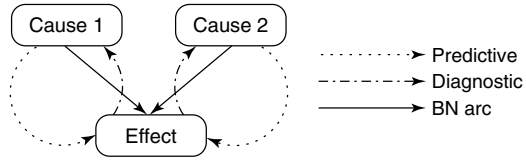


Figure 2 Predictive and diagnostic reasoning in Bayesian networks [Reproduced with permission from Elsevier.]

Bayesian Networks in Judicial and Forensic Context

There are different ways in which BNs are used as a modeling technique. Legal scholars focus on BNs as a means for structuring cases as a whole, whereas forensic scientists concentrate primarily on the evaluation of selected issues that pertain to scientific evidence. Many studies with an emphasis on legal applications thus rely on graphical models as a method for the retrospective analysis of complex and historically important causes 'célèbres', such as the Collins case [10], the Sacco and Vanzetti case [11], the Omar Raddad case [12], or the O.J. Simpson case [13]. Another contribution proposes BNs to clarify fallacious arguments [14].

Forensic applications of BNs [1] range from offender profiling [2], single [15, 16] and complex [3, 17, 18] configurations of different kinds of trace evidence as well as inference problems involving the results of DNA analysis. The latter is an important category that covers studies focusing on small quantities of DNA [19], cross-transfer evidence [20], relatedness testing with or without missing data on relevant individuals [21, 22], or mixed stains [23, 24]. Interest has also been focused on case preassessment where the role of the forensic scientist consists of assessing the value of the expected results, prior to laboratory examination [1, 25]. BNs evaluate the probability of possible outcomes in various cases together with their respective weight.

These conceptual and practical studies represent a relatively rich variety of inferential topics. The reported works jointly support the idea that graphical probability models can substantially improve the evaluation of likelihood ratios used for the assessment of scientific evidence. In particular, they allow their user to engage in probabilistic analysis of much higher complexity than what would be possible through traditional approaches that mostly rely on rather rigid, purely arithmetic developments.

In many applications, the size of problems to be analyzed using BNs may be important; scientists are seeking ways to reduce the complexity of model building. Object-oriented Bayesian networks (OOBNs) [26] facilitate hierarchical construction utilizing small modular networks (network fragments) as building blocks. Examples of such models are presented in [24, 27, 28].

What about the Numerical Specification?

If we cannot get any numbers, the method should be rejected? Note that numbers are not of primary importance; as mentioned by de Finetti, "a rough qualitative appreciation of probabilities is sufficient and worthy for practical purposes" [29]. In fact, reasoning with graphical models is about clarifying one's thinking.

Operationally, an abstraction of quantitative BNs, the so-called qualitative probabilistic networks, (QPNs), and sensitivity analysis can be used for inference purposes. In the former, qualitative relationships are used instead of numerical relations, and allow the user to rapidly gain an idea as to the qualitative effect a variable may have on another. On some occasions, this information may be entirely sufficient; on other occasions it may represent a valuable preliminary information before more sophisticated methods will be employed for eliciting probabilities[d] [31]. In the latter, information about the effect that one variable (or the joint effect that two or more variables) may have on a second is studied. Such indications may be helpful for deciding on the relative importance of efforts needed for the elicitation of some probabilities.

Forensic examples of such analysis are presented and discussed in [1, 32, 33].

What about the Decision-Making Process?

As mentioned earlier, BNs provide a coherent environment in which beliefs about target variables can be reevaluated in the light of newly acquired evidence. This constitutes a fundamental prerequisite for decision making under uncertainty. In fact, inference and decision are connected because the results of the former are the point of departure of the latter [34]. BNs can be extended to incorporate basic ingredients necessary to perform Bayesian decision analysis, that is, decision and utility nodes, representing actions available to the scientists and values for possible consequences of these actions, respectively [6]. Forensic examples of this approach are presented in [35–37].

Conclusions

BNs are a widely applicable formalism for a concise representation of uncertain relationships among variables of interest in a domain (in this case, forensic

science). They provide assistance and support in the rigorous evaluation of evidence according to the laws of probability, which guarantee rationality in reasoning under uncertainty.

BNs take an important part in forensic science practice and research, notably as

- a cognitive tool for the structuring and logical guiding of mental processes;
- a coherent inferential concept for drawing inferences to hypotheses based on evidence;
- a flexible and dynamic approach allowing knowledge to be transparently revised upon newly acquired information.

End Notes

[a.] Literature often refers to causal relationship.
[b.] Nodes represents random variables, where the random variable may be either discrete, with a finite set of mutually exclusive states which themselves can be categorical, discrete or continuous.
[c.] A variable is called *instantiated* if its state is changed from unknown to known.
[d.] Often the conditional distributions include parameters that are unknown and must be estimated from data. This technique is called *parameter learning*. A forensic example is given by [30].

References

[1] Taroni, F., Aitken, C.G.G., Garbolino, G. & Biedermann, A. (2006). *Bayesian Networks and Probabilistic Inference in Forensic Science*, John Wiley & Sons, Chichester.

[2] Aitken, C.G.G. & Gammerman, A. (1989). Probabilistic reasoning in evidential assessment, *Journal of the Forensic Science Society* **29**, 303–316.

[3] Dawid, A.P. & Evett, I.W. (1997). Using a graphical method to assist the evaluation of complicated patterns of evidence, *Journal of Forensic Sciences* **42**, 226–231.

[4] Garbolino, P. (2001). Explaining relevance, *Cardozo Law Review* **22**, 1503–1521.

[5] Cowell, R.G., Dawid, A.P., Lauritzen, S.L. & Spiegelhalter, D.J. (1999). *Probabilistic Networks and Expert Systems*, Springer, New York.

[6] Jensen, F.V. (2001). *Bayesian Networks and Decision Graphs*, Springer, New York.

[7] Lindley, D.V. (2000). The philosophy of statistics, *The Statistician* **49**, 293–337.

[8] Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers, San Mateo.

[9] Wellman, M.P. & Henrion, M. (1991). Qualitative intercausal relations, or explaining "explaining away", *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*, Morgan Kaufmann Publishers, pp. 535–546.

[10] Edwards, W. (1991). Influence diagrams, bayesian imperialism, and the Collins case: an appeal to reason, *Cardozo Law Review* **13**, 1025–1074.

[11] Kadane, J.B. & Schum, D.A. (1996). *A Probabilistic Analysis of the Sacco and Vanzetti Evidence*, John Wiley & Sons, New York.

[12] Levitt, T.S. & Blackmond Laskey, K. (2001). Computational inference for evidential reasoning in support of judicial proof, *Cardozo Law Review* **22**, 1691–1731.

[13] Thagart, P. (2003). Why wasn't O.J. convicted? emotional coherence and legal inference, *Cognition and Emotion* **17**, 361–383.

[14] Fenton, N. & Neil, M. (2000). The 'jury observation fallacy' and the use of Bayesian networks to present probabilistic legal arguments, *Mathematics Today Bulletin of the IMA* **36**, 180–187.

[15] Garbolino, P. & Taroni, F. (2002). Evaluation of scientific evidence using Bayesian networks, *Forensic Science International* **125**, 149–155.

[16] Biedermann, A. & Taroni, F. (2006). A probabilistic approach to the joint evaluation of firearm evidence and gunshot residues, *Forensic Science International* **163**, 18–33.

[17] Taroni, F., Biedermann, A., Garbolino, P. & Aitken, C.G.G. (2004). A general approach to Bayesian networks for the interpretation of evidence, *Forensic Science International* **139**, 5–16.

[18] Taroni, F. & Biedermann, A. (2005). Inadequacies of posterior probabilities for the assessment of scientific evidence, *Law, Probability and Risk* **4**, 89–114.

[19] Evett, I.W., Gill, P.D., Jackson, G., Whitaker, J. & Champod, C. (2002). Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks, *Journal of Forensic Sciences* **47**, 520–530.

[20] Aitken, C.G.G., Taroni, F. & Garbolino, P. (2003). A graphical model for the evaluation of cross-transfer evidence in DNA profiles, *Theoretical Population Biology* **63**, 179–190.

[21] Dawid, A.P., van Boxel, D.W., Mortera, J. & Pascali, V.L. (1999). Inference about disputed paternity from an incomplete pedigree using a probabilistic expert system, *Bulletin of the International Statistical Institute* **58**(Book 1), 241–242.

[22] Dawid, A.P., Mortera, J., Pascali, V.L. & van Boxel, D. (2002). Probabilistic expert systems for forensic inference from genetic markers, *Scandinavian Journal of Statistics* **29**, 577–595.

[23] Mortera, J., Dawid, A.P. & Lauritzen, S.L. (2003). Probabilistic expert systems for DNA mixture profiling, *Theoretical Population Biology* **63**, 191–205.
[24] Cowell, R.G., Lauritzen, S.L. & Mortera, J. (2004). *Identification and Separation of DNA Mixtures Using Peak Area Information*, Technical Report 25, Cass Business School, London.
[25] Biedermann, A., Taroni, F., Bozza, S. & Aitken, C.G.G. (2008). Analysis of sampling issues using bayesian networks, *Law, Probability and Risk* **7**, 35–60.
[26] Koller, D. & Pfeffer, A. (1997). Object-oriented bayesian networks, *Proceedings of the Thirteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-97)*, Providence, Rhode Island, pp. 302–313.
[27] Dawid, A.P., Mortera, J. & Vicard, P. (2007). Object-oriented bayesian networks for complex forensic DNA profiling problems, *Forensic Science International* **169**, 195–205.
[28] Cavallini, D. Corradi, F. (2006). Forensic identification of relatives of individuals included in a database of DNA profiles, *Biometrika* **93**, 525–536.
[29] de Finetti, B. (1973, 1993). Bayesianism: its unifying role for both the foundations and the applications of statistics, in *Bruno de Finetti, Probabilit'a e Induzione*, P. Monari & D. Cocchi, eds, Clueb, Bologna, pp. 467–490.
[30] Dawid, A.P. (2003). An object-oriented Bayesian network for estimating mutation rates, *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics* Key West, Florida.
[31] Neapolitan, R.E. (2004). *Learning Bayesian Networks. Prentice Hall Series in Artificial Intelligence*, Pearson Prentice Hall, Upper Saddle River.
[32] Biedermann, A., Taroni, F., Del'emont, O., Semadeni, C. & Davison, A.C. (2005). The evaluation of evidence in the forensic investigation of fire incidents (Part II): practical examples of the use of Bayesian networks, *Forensic Science International* **147**, 56–69.
[33] Biedermann, A. & Taroni, F. (2006). Bayesian networks and probabilistic reasoning about scientific evidence when there is a lack of data, *Forensic Science International* **157**, 163–167.
[34] de Finetti, B. (1970). Logical foundations and measurement of subjective probability, *Acta Psychologica* **34**, 129–145.
[35] Taroni, F., Bozza, S. & Aitken, C.G.G. (2005). Decision analysis in forensic science, *Journal of Forensic Sciences* **50**, 894–905.
[36] Taroni, F., Bozza, S. & Biedermann, A. (2006). Two items of evidence, no putative source: an inference problem in forensic intelligence, *Journal of Forensic Sciences* **51**, 1350–1361.
[37] Biedermann, A., Bozza, S. & Taroni, F. (2008). Decision theoretic properties of forensic identification: underlying logic and argumentative implications, *Forensic Science International* **177**, 120–132.

FRANCO TARONI

Behavior and Alcohol Intoxication *see* Alcohol: Behavioral and Medical Effects

Behavior and the Law *see* Behavioral Science Evidence

Behavioral Genomics *see* Genomics and Behavioral Evidence

Behavioral Science Evidence

Consideration for the relationship between individual differences, conduct, and culpability has existed since before recorded history. Throughout most of early history, human beings lived in extended family groups or tribes with no formal law, as we know it today. These tribal groups developed folkrights or customs for dealing with unmistakable differences between the individual members of the group. For example, if a man "fell out of his senses or wits and killed someone", his kinsmen were expected to compensate the victim [1]. People recognized what we today consider mental illness and physical handicaps. Often the mentally ill were regarded as having special powers. At other times, as appropriate, the mentally and physically disabled were given special dispensations and others in the group assumed their responsibilities and obligations.

In the eleventh century, major attempts were made to organize tribal and folk groups into nations. This is always a difficult undertaking. Tribes, by whatever name, are controlled by people generally known to

all of the members. Like a family, social dynamics can be worked out face-to-face; there is no need for formal law; and each member can be taken as a unique individual and treated accordingly. While tribes vary considerably, and some can be totalitarian and even belligerent, most people feel comfortable in small social networks, where they can plead their own case and find unique acceptance. For this reason, much of the world is still at best made up of alliances between local groups. Often these folkgroups or clans unite in religious or nationalistic factions, and these factions both impose some form of discipline and authority, while often also engaging in feuds or warfare against each other.

Great Britain attempted its first serious transition from tribalhood to nationhood following the Norman Conquest in 1066. It was a slow process, and it was not until the Magna Carta or "Great Charter" in 1215 that anyone really started to think about government under a rule of law. Henry de Bracton was born about the time the charter took effect, and went on to become perhaps the greatest single influence on modern law. Bracton was a priest – there were no lawyers or law schools at that time – and he soon became a judge and legal scholar.

Having studied folk customs, Bracton understood that law had to recognize individual differences and their impact on culpability. Bracton introduced the legal term *mens rea*. Law, Bracton reasoned, involved two components. The first is physical and considers questions such as whether a person actually did the act that, for example, defines a crime. The second is mental; *mens* being the mind or intellect, and *rea* being the answerability or accountability [2].

Bracton shaped a sophisticated body of legal thought that is still influential today [3, 4]. Special measures were taken for those who could not participate in their defense, parents were warned to take care of and treat leniently their "evildoer" offspring, and at least one court acquitted a defendant who had committed his crime under the influence of an improper prescription ordered by his physician.

Nevertheless, this was a balanced justice. Children and those suffering from a mental disease (what we would call today a *psychosis* that reaches the legal test of insanity) or mental defect (for example, mental retardation) were accountable for their *Torts* [5]. Torts are civil breaches of interpersonal conduct. The mentally ill can and occasionally do commit torts, and the law in most jurisdictions still holds them or their guardians accountable for the consequences.

Despite all these sophisticated legal provisions, most people were fully accountable for their conduct. Conduct is the fixed legal standards that specify what people should and should not do. The legal requirements that govern conduct did and still do provide tremendous latitude in individual behavior, but virtually all nations have laws that specify what is and what is not acceptable conduct. For this limited number of actions that are illegal, there is very little flexibility except for the special responsibility exception formalized by Bracton.

Folkway, on the other hand, did contain provisions for individual difference; a way to provide dispensations based upon the individual and the circumstances. Modern nations, working to create a society under law, could not ignore this need for and tradition of special dispensations. However, in the British courts and those of most nations to this day, findings of fact are restricted *only* to whether the accused did or did not commit the act that defines a legal offense. In fact, in many nations including England, the burden of proof, whether nation must prove guilt or the accused prove innocence, rests with the accused, though the burden of proving sanity may shift back to the prosecution, once the accused makes a prima facie case that he was insane.

The granting of special dispensation was reserved as an unyielding prerogative of royalty. Thus, in England and elsewhere, a defendant is first tried and found guilty of a criminal act. The guilty party must then appeal for royal dispensation based upon their individuality and circumstances. While this may seem complicated, it was consistent with folkways that left such matters to elders or other leaders of the tribe. Royalty was determined to retain the power of dispensation because it is one of the greatest powers in governance. Used wisely, it can demonstrate fairness, further respect, and earn undying gratitude.

Some royalty grant audiences during which dispensations are pleaded and awarded; usually with witnesses to the largess. This, however, can become a time consuming process, and in England the duty was delegated, except in high-profile political situations, to a court of chancery. A special court with a strong religious influence, chancery's legacy can be seen today in the American court of equity. These no longer involve a separate court, but regular courts

can convene in equity and deal with matters that reach beyond the letter of the law.

Insanity

Insanity is not a medical or psychiatric term. Insanity is a *legal* concept and must be determined as a question of fact. The cornerstone legal definition was formulated in England in 1843, and draws directly from Bracton. M'Naghten's Rule stemmed from Queen Victoria's outrage with a "not guilty" verdict in the case of Daniel M'Naghten, debate in both houses of Parliament, and a joint meeting of the high justices and the House of Lords. The Rule holds that "to establish a defense on the grounds of insanity, it must be clearly proved that, at the time of committing the act, the party accused was laboring under such a defect of reason, from disease of the mind, as *not to know the nature and quality of the act he was doing, or if he did know it, that he did not know he was doing wrong"* [6].

Since M'Naghten, attempts have been made to alter the law to make it easier or more difficult to enter a successful insanity defense. On the heels of M'Naghten, the Massachusetts Court added what came to be known as the "*irresistible impulse*" *test*, whereby an accused should be found not guilty if the jury determined he was irresistibly driven to committing the crime by the mental disease or defect of which he suffered [7]. In 1954, the US federal courts adopted the Durham Standard [8] also known as the "*product*" *test*. The Court held that neither M'Naghten nor the irresistible impulse test provided satisfactory criteria for determining sanity. The *Durham* court stated that "a defendant is not criminally responsible if his unlawful act was the "product" of mental disease or defect".

Durham presented a number of problems. The new test made it much easier to be found not guilty by reason of insanity. On the other hand, the test also provided for indefinite commitment in a mental hospital, and defendant Durham ended up spending more time behind bars than if he had been found guilty of the crime.

Finally, in 1981, John W. Hinckley, Jr. was found legally insane at his trial for attempting to assassinate President Ronald Reagan, an act by which Hinckley sought to impress actress Jodie Foster. The public saw Hinckley as an indolent self-indulgent drifter from a wealthy family. In response Congress enacted the Insanity Defense Reform Act of 1984 [9] that returned to Bracton's roots; producing a new definition requiring that at the time of the commission of the acts constituting the offense, the defendant, *as a result of severe mental disease or defect, was unable to appreciate the nature and quality or the wrongfulness of his acts.* Mental disease or defect would not otherwise constitute a defense, and the burden of proving the defense of insanity by clear and convincing evidence rested with the defendant.

Not all jurisdictions follow the US federal standard, and a few, including the state of New Hampshire, still have an irresistible impulse test on the books. Regardless of the test, however, jury verdicts reflect a societal consensus that the folkviews upon which Bracton based his legal formulations continue to this day. Nevertheless, insanity is not a simple issue (*see* **Insanity: Defense**; **Temporary Insanity**; **Treatment, Mandated: Mental Health**; **Treatment, Right to Refuse: Mental Health**; **Sex Offenders: Treatment of**).

Diminished Responsibility

When the United States declared their independence and sought to create a nation of their own, they faced a unique problem. The colonies rejected the monarchy, and were adamant about the separation of church and state. In fact, they even distrusted their own government and devised a set of checks and balances that restrained central power. This effectively precluded the delegation of dispensations to a monarchy or to a religious power. Faced with this problem, the new nation turned back to history and, as in early folkgroups, delegated the responsibility for special dispensations to the people who should best know and understand the parties in litigation: a jury of their peers.

No longer would justice require a two-step process with a trial for findings of the fact and a chancery to consider mental state and motive. The trial court would now address both. This required a completely new definition of laws; specifying not only the physical act(s), which define the violation of conduct, but specific adverbs to indicate required mental states.

The obvious examples are crimes that contain in their definition mental qualifications, such as "deliberate", "intentional", "wanton", and "premeditated".

Societies include these adverbs in their legislation not just to express their belief that the acts are unusually heinous, but to indicate that to meet the *legal* definition of such a crime, and therefore be subject to punishment for committing the crime, the perpetrator must meet both the *physical* requirement of committing the specified act or omission and must have possessed the requisite *mental* state during the events leading to and during the act or omission.

These expanded definitions, which invite behavioral evidence, are often referred to as *opening the door to "diminished capacity" defenses*. However, just as with insanity, if the defendant fails to satisfy the mental state requirements of the law, the law is simply not applicable. Although it is true that the added language involves states of mind, the states of mind and motives considered in an insanity defense, as opposed to other legal situations, are qualitatively different.

An insane defendant is accorded a unique status and could, for example, be institutionalized. The only way to hold a sane defendant accountable is to charge her with the correct crime in the first place. This may require, for example, reducing a charge from murder to manslaughter. This was effectively what a court of chancery did in retrospect. In the United States, however, matching the charge to both the physical and mental requirements of the law falls entirely within the trial court action. Moreover, while American criminal law, in particular, contains state of mind adverbs, mental state also plays a role in many civil causes of action. For example, the tort of "intentional infliction of emotional harm" requires specific intent – a direct correspondence between state of mind and the wording of the law. An insane person, on the other hand, is assumed to suffer from a disease or defect that deprives them of the capacity to conform to the law.

Virtually all litigation in any nation still requires consideration of both the physical and mental aspects of the alleged conduct. In most jurisdictions other than the United States, however, the mental aspects are still left to sovereigns or religious officials [10]. Therefore, if you look at the editors and contributors to the major international academic journals on law and behavior [11], virtually all are based in the United States or Canada. Other academic experts in this particular field also tend to have been educated or originally lived in North America.

The Judiciary

Most disputes involve *versions* of reality, much as with memory and belief, which must typically be accepted as narrative rather than historical truth. Only rarely do courts have access to the physical evidence necessary to reconcile beliefs to a single, certain, truth. Science may be the business of truth, but law is the business of justice. It is the acceptance of a quality of justice – respect for the court and its procedures – rather than any objective verifiable standard of truth, which determines whether the citizens of a nation resolve their disputes through law or through the shedding of blood in the streets.

The kings of Bracton's time recognized this or learned in blood the lessons of unjust rulership. Courts today, and particularly those that must go to the heart of justice that always rests in human individuality and motive, also understand that justice can never be entirely separated from public sentiment. The term *forensic* derives from the *forum* or arena of public debate – places like the courts in which matters of broad public concern are openly debated, exchanged, weighed, and assessed.

In earlier days, people came to view the court when it was the only public activity in town. Today, they follow trials on television. Certainly public debate can inspire bad ideas as well as good, but one of the strengths of open societies is their marketplace of ideas. Courts air a wide range of behaviors from exemplary to outrageous. Societies define their standards through every individual's constant measurement of their conduct against all other examples of human behavior including that of criminals, and generally arrive at a reasonable balance. While jurors define justice in a particular case through their verdicts, societies define justice through their debate of verdicts, by acting upon their conclusions as future jurors, and as they did in response to the Hinckley verdict.

The Judiciary, of which the trial courts are one part, is the branch of government most responsible for assuring stability and continuity of conduct, expectations, and social structure. The levels of appeal built into the judicial system, and the accumulation of common law – the past precedents that shape future holdings – provide a buffer against societal fads and the passions of the moment. Nevertheless, as Bracton recognized, law must gradually evolve with the evolution of societal convictions. If law does not heed

evolving public sentiment, triers of fact override the letter of the law through what is known in the legal vernacular as *jury nullification* [12–20].

As we saw in the post-Hinckley example, law related to any given area waxes and wanes in response to current beliefs. The public, and appropriately so, is generally willing to consider new information and scientific theories. Shifts in the law often occur in response to new or "novel" theories; particularly theories of human behavior. Justice operates in a social context, and in the 1960s the American public was heavily impacted by psychoanalysis with its view of the unconscious and therefore on limitations in individual responsibility. Over the past several decades, psychoanalytic therapy has been supplanted by cognitive behavioral therapies that maximize rather than minimize individual responsibility [21].

The impact of the evolution of "novel" scientific theories can be seen through the use of the legal research tool known as *case congregation analysis* [22]. In each case congregation, disputes progress along a path shaped not only by science but also by social fears and broad national political forces. Often, but not always, science prevails. Clearly, there can be no doubt that resolution of underlying scientific issues is a prerequisite to settlement of a complete legal case congregation, notwithstanding the reality that early cases are typically tried well before the scientific issues have been fully delineated, much less resolved. However, achievement of a national consensus on a seemingly scientific issue often turns out to require a social consensus on moral or cost/benefit issues, which transcend science [23].

Science

Although law is the province of justice, science should be the province of truth. The law has steadily moved toward requiring higher standards of proof and procedure in scientific expert testimony (*see* ***Daubert v. Merrell Dow Pharmaceuticals***; ***General Electric v. Joiner***; ***Kumho Tire v. Carmichael***; ***Weisgram v. Marley***).

In *Daubert* [24], the most recent major US Supreme Court case to address the standards for expert scientific testimony, the issue of fact in dispute was whether the drug Bendectin was a teratogen. Accepted scientific procedures have been established to determine whether a substance causes birth defects. However, science is not simply a set of accepted facts. Science is a *process* through which empirical truth is sought using scientific methods, adversarial debate, and peer review. In *Daubert*, the US Supreme Court focused upon scientific debate and just how the reasoning and methodology underlying evidence and testimony can be established as scientifically valid, as well as how best to determine whether that reasoning or methodology can be properly applied to the facts in dispute.

In some instances, the facts at issue in a given trial can be addressed through behavioral science expert testimony on the basis of scientific standards applicable to the best of the hard sciences (*see* **Deception: Truth Serum**). Typically, however, behavioral science experts testify about factual issues that are less comparable to those addressed by the physical sciences. In the vernacular of the field, the task is frequently to help "separate the mad from bad or the lazy from the crazy". It is not true that physical science operates in a judicial context any less influenced by issues of policy and societal pressures. Both behavioral and physical sciences cases are subject to these pressures, and their case congregations – the course of cases over time, which address a given issue of fact – reflect the evolution of public consensus. The distinction is that the physical sciences lend themselves more easily to methodologies that eventually yield scientific conclusions.

Nevertheless, the behavioral sciences appropriately address issues of fact intrinsic to the law. The law, and particularly the structure of law in the United States, is about individual differences. Questions of motive and state of mind are, by definition, unique to a given person at a specific time. Science can determine that a given drug is or is not a teratogen, and that scientific conclusion applies to that drug in any future litigation. Behavioral science expert testimony typically addresses issues of fact that relate to the conduct of a unique individual, at a given time and place, under a particular set of circumstances.

The behavioral sciences do follow scientific methodologies, but the application of a given methodology varies with the question being asked. The challenge is to match the correct methodology to the question being asked by the trier of fact. Behavioral science experts typically bring to a jury experience and expertise in assessing individuals. Triers of fact weigh that experience and look to

conclusions grounded in experience and training. Therefore, a proposal to introduce behavioral science expert testimony generally raises a question not of admissibility, but of credibility [25] (*see* **Expert Opinion in Court: a Comparison of Approaches**).

It is often suggested that the fact that two or more behavioral scientists can testify in the same case with opposite opinions is proof that behavioral science is not really a science at all. It is just as frequent that physical scientists testify with opposing positions, particularly in cases that are early in a case congregation sequence. Being a scientist is not about knowing the answers, but about following a systematic process to find the right answer to the question at hand.

Scientists, like the population at large, view the world through their own perspectives and biases. The rules of science are designed so that other scientists can detect and refute premature or flawed conclusions. As long as scientists follow the rules of science, the search for truth is built into the structure of the discipline. Science, in this respect, is no less adversarial than law.

In most physical sciences, replication is both a goal and proof that a finding has been scientifically established. This does not apply to behavioral case assessments. Under law, every person is unique. The fact that the vast majority of other people may fit a particular pattern is irrelevant if there is any chance that the person being evaluated may be an exception. The reality that behavioral experts do not always agree often reflects nothing more than one expert arguing the rule, and another the exception.

Expert Testimony

In a country such as the United States, where motive and mindset are integral to the facts that must be addressed by the jury at trial, behavioral science testimony is by far the most frequent form of expert testimony. Because it is so routine, and often the expert witnesses are government employees or others already working within the justice system, exact figures are unknown. Estimates suggest that from 30 to 60% of all expert testimony involves issues of motive or mental state.

The relationship between an expert and her attorney/client also follows a set of well-established rules. Experts must be objective to assist the trier of fact. Experts cannot accept contingent fees based on the outcome of a case. Their job is not to formulate trial strategy, which is the job of counsel who is bound by rules of privilege that may limit the type and amount of information that can be shared even with the expert. The topic is beyond the scope of this article, and an expert should conduct the research necessary to master the rules and art of expert witness practice.

As a general rule, expert witnesses testify only at trial. Appeals courts address questions of law and usually all questions of fact are resolved at trial and not revisited on appeal (*see* **Expert Opinion: Appeal v. Trial**). Moreover, experts usually testify about, but do not tell the trier of fact, the answer to the ultimate issue upon which the case turns. For example, if the case turns on whether a defendant is or is not legally insane, an expert may assist the jury to understand the mental state and functioning of the defendant; but will not be asked to state a conclusion as to whether the particular defendants actually is insane. This should be left to the trier of fact (*see* **Ultimate Issue Evidence by Experts**).

The legal system works when the people upon whom the system depends do their job.

Prerogatives of the Trier of Fact

Just who is or is not telling the truth in a legal situation in which the facts at issue are material to the outcome of the case is a question of fact usually left to the trier of fact – the jury or in some instances a judge. The judiciary is very reluctant to take such decisions away from the trier of fact, and therefore does not permit expert testimony to substitute for a party taking the stand for direct assessment by a jury. Similarly, although science has clearly demonstrated that eyewitnesses may provide the most unreliable forms of evidence, we leave their evaluation to the jury and even make it difficult (depending upon jurisdiction) to provide expert testimony on the reliability of eyewitness testimony [26] (*see* **Eyewitness Testimony**; **Eyewitness Lineups: Identification from**).

For the same reason, the courts generally do not allow admission of the results of lie detector tests or other new high-technology deception technologies [27]. Not only are they of questionable scientific accuracy, they are limited by the same human memory constraints as hypnosis and truth serums (*see*

Deception: Detection of; **Deception: Detection of and Brain Imaging**).

"Truth serum" and hypnosis evidence also impinge upon the prerogatives of the finder of fact, and for this reason alone have been held inadmissible as legal evidence (*see* **Deception: Truth Serum**; **Hypnosis and Memory**).

Issues and Controversies

The behavioral sciences, and related disciplines including psychopharmacology, are advancing at a progressive rate and today bear little resemblance to where they were fifty or even thirty years ago. The emphasis today is upon evidence-based practice, and the scientific literature in the field now holds its own against the standards of medicine and the hard sciences (*see* **Psychopathology: Terms and Trends**; for recent developments in psychopharmacology, *see* **Psychopharmacology**; **Psychopharmacology: Child and Adolescent**).

During recent years, crime novels, movies, and television have featured advanced forensic sciences as well as behavioral science methodologies such as profile and syndrome identification. While the imitation is flattering, most of the content is pure fiction. The glamor and unfortunately too much of the substance and its application to routine law enforcement is vastly overrated. Moreover, much of the fiction is based upon outdated information and practices that have been discredited or are no longer followed (*see*, in particular, **Profiles: Psychological and Behavioral**; **Syndromes: Psychological**).

Courts devote considerable attention to issues of child custody (*see, for example,* **Visitation Rights**; **Parenting: Assessment of Capacity**; **Battered Child Syndrome**; **Children: as Witnesses**) and capacity (*see*, for example, **Capacity Assessment**; **Capacity to Stand Trial**).

Elder affairs have also moved into the courthouse, with issues ranging from protection to guardianship (*see*, for example, **Guardianships of Adults**; **Capacity for Independent Living**; **Elder Abuse: Policy**; **Elder Abuse: Risk**).

Risk assessment (*see*, in particular, **Homicide: Multiple (Behavior)**; **Risk Assessment: Patient and Detainee**; **Duty to Warn**), and threat assessment (*see*, for example, **Threat Assessment: School**; **Threat Assessment: Workplace**; **Violence Risk Assessment for Mental Health Professionals**) are specialties that have made significant advancements in recent years (*see* **Dangerousness: Risk of**).

Malingering (*see*, for example, **Malingering: Forensic Evaluations**), persuasion and undue influence (*see*, for example, **Interrogation**; **Confessions: Evidentiary Reliability of**; **Capacity to Waive Miranda Rights**), Jury dynamics (*see* **Jury Dynamics**) and the structure of justice itself (*see*, for example, **Therapeutic Jurisprudence**; **Mental Health Courts**) are topics of increasing relevance.

Even the conceptualization of mental health diagnosis and classification has ongoing relevance to the functioning of the judicial system. The *Diagnostic and Statistical Manual of Mental Disorders* [28], that in turn shapes the *International Classification of Diseases* [29], is at a point that is comparable to where birth defects were 30 years ago. Developed as research tools and containing a disclaimer that the classifications "may not be wholly relevant to legal judgments . . . that take into account such issues as individual responsibility, disability determinations, and competency", these references are widely used as a catalog of "scientific" entities that can be used to either strengthen a defense or assert a cause of action.

The classifications, however, are of limited legal utility, and counsel as well as experts should learn how to prevent their abuse. Nevertheless, the structures and definitions in these volumes serve the same purpose that a compendium of birth defects accomplished prior to the development of large-scale gene mapping technologies – it defines research and testing populations upon which basic science research can be performed, thus hopefully helping to transform behavioral sciences from clinical to more empirical specialties.

In the meantime, efforts to shift and realign definitions of behavioral conditions have become controversial and contentious. The ramifications are not simply academic. As examples, consider addiction and posttraumatic stress disorder. Addiction (*see,* for example, **Alcohol: Use, Abuse, Tolerance, and Dependency**; **Addictions**; **Substance Abuse**) has been defined (here simplified for purposes of illustration) as taking a drug and undergoing withdrawal upon its discontinuation, or alternatively as taking a drug, going successfully through withdrawal, and then returning to use of the drug. The distinction is significant. The first definition can include almost anyone, while the second is far more restrictive and

selective. Their alternative application in a legal context can change the outcome of the litigation.

Similarly, a wide range of definitions of posttraumatic stress disorder (PTSD) has been proposed and is being debated (*see* **Posttraumatic Stress Disorder**). At one end of the continuum, the condition is a rare anomaly. Most people are strengthened by life and mastery of adversity, while a small number, perhaps for genetic reasons, are susceptible to PTSD. At the opposite end of the continuum, virtually everyone meets the criteria for a PTSD diagnosis at some point in their life; and everyone with the diagnosis might then have a cause of action against someone else for creating the condition.

"Novel" behavioral theories also have had a tendency to create both "victims" and causes of action [30] (*see*, for example, **Syndromes: Psychological**; **Deception: Truth Serum**). Typically, these proposed conditions are advanced during a period of social change, anxiety, and as a seeming means of righting some historic wrong. They usually run a course in which an individualized argument has occasionally been successful in the past. A novel theory is then advanced to lend scientific respectability to the concept. The theory is introduced in litigation, unsuccessful until a landmark case brings the "science" to public attention. Following the publicity and the controversy that it engenders, successful cases rise sharply until overuse become apparent and a concerted effort is made by the opposition to discredit the theory. Once discredited, as demonstrated repeatedly through case congregation analysis, the rate of successful applications of the theory drops precipitously until successful cases reach the pretheory level. Often public outrage calls for the banning of such theories as "junk science". However, it is the abuse of the scientific imprimatur that must be discredited, and once a behavioral theory is discredited, it ceases to be used. Lack of success is more effective than legal prohibition at ending impropriety.

Conclusions

Human beings have understood and accommodated the relationship between individual differences, conduct, and culpability since before recorded history. Henry de Bracton, building on this foundation in the thirteenth century, created a modern judiciary that could unite disparate clans into a nation under law. Bracton also reasoned that law involved two components: the physical fact of whether a person committed a crime, and a corresponding requirement of mental accountability. The prerogative of granting dispensations on the basis of mental accountability was reserved for kings and religious officials until the United States rejected royal and religious rule, and enacted laws that transferred behavioral dispensations to trial juries. Behavioral evidence now accounts for the single largest category of expert testimony in the United States, and the academic literature on the topic is followed worldwide.

Insanity is not a medical or psychiatric term; it is a *legal* concept that must be determined as a question of fact. The behavioral sciences today bear little resemblance to where they were 50 or even 30 years ago. Questions of motive and state of mind are, by definition, unique to a given person at a specific time. The task of an expert witness is to assist the trier of fact. The reality that the vast majority of people may fit a particular pattern is irrelevant, if there is any chance that the person being evaluated may be an exception. The fact that behavioral experts may disagree often, reflects nothing more than one expert arguing the rule, and another the exception.

"Novel" behavioral theories and classifications have had a tendency to create both "victims" and causes of action. Typically, these proposed conditions are advanced during a period of social change, anxiety, and as a seeming means of righting some historic wrong. Case congregations provide a vivid record of shifts and corrections in the judicial process. Most problems are self-correcting, given an open forum and a public, both willing to accept new ideas in its effort to do the right thing, and also able to recognize mistakes.

References

[1] For the history and evolution of behavioral law, see Edwards, C.N. (ed) (2001). *Responsibilities and Dispensations: Behavior, Science, and American Justice*, Four Oaks Press, Dover.

[2] For a modern discussion of how law actually is a behavioral science, see Blumenthal, J.A, (2002). Law and social science in the twenty-first century, *Southern California Interdisciplinary Law Journal* **241**, 1–52.

[3] For an excellent one-volume history of international law, see the classic Wormser, R.A. (1962). *The Story of the Law*, Simon and Schuster, New York.

[4] For the best single source on the history of the law of criminal insanity in England, see Walker, N. (1968). *Crime and Insanity in England: The Historical Perspective*, Edinburgh University Press, Edinburgh.
[5] The exception being breach of contract since it is assumed that young children and the mentally disabled lack the capacity to enter into contracts.
[6] *M'Naghten's Case*, 10 Clark & Finelly 200, 8 Eng. Rep. 718 (1843).
[7] *Commonwealth v. Rogers*, 48 Mass. 500, 502 (1844).
[8] *Durham v. United State*, 214 F.2d 862 (1954).
[9] U.S.C. Title 18 § 17, 3006A, 4241.
[10] For a discussion of the role of behavioral concepts in American law see Piar, D.F. (2008). A welfare state of civil rights: the triumph of the therapeutic in American Constitutional Law, *William and Mary Bill of Rights Journal* **16**, 649–684.
[11] See, for example, *Behavioral Science & Law* (published in England) and *Law and Human Behavior* (published in Germany).
[12] Creagan, M.K. (1993). Jury nullification: assessing recent legislative developments, *Case Western Reserve Law Review* **43**, 1101–1150.
[13] Weinstein, J.B. (1993). Considering jury 'nullification': when may and should a jury reject the law to do justice, *American Criminal Law Review* **30**, 239–254.
[14] Schoop, R.F. (1996). Verdicts of conscience: nullification and necessity as jury responses to crimes of conscience, *Southern California Law Review* **69**, 2039–2116.
[15] Leipold, A.D. (1996). Rethinking jury nullification, *Virginia Law Review* **82**, 253–324.
[16] Niedermeier, K.E., Horowitz, I.A. & Kerr, N.L. (1999). Informing jurors of their nullification power: a route to a just verdict or judicial chaos? *Law and Human Behavior* **23**(3), 331–351.
[17] For a history of jury nullification, see Farnham, D. (1997). Jury nullification: history proves it's not a new idea, *Criminal Justice* **11**(4), 4–14.
[18] For a view of jury nullification in the context of mental health exotica, see Dorfman, D.N. & Iijima, C.K. (1995). Fictions, fault, and forgiveness: jury nullification in a new context, *University of Michigan Journal of Law Reform* **28**, 861–929.
[19] For an excellent discussion of jury nullification in the context of a war protest case see *United States v. Dougherty*, 473 F.2d 1113 (1971).
[20] For an overview of the jury system generally see Abramson, J. (1994). *We the Jury*, Basic Books, New York.
[21] Cognitive Therapy proper was developed in the 1960s by psychiatrist Aaron T. Beck as a treatment for depression. Later elaborated and expanded by his daughter, psychologist Judith S. Beck, the approached has been found applicable to the broad range of psychological disorders that are not overwhelmingly biological in origin. Beck, J.S. (1995). *Cognitive Therapy: Basics and Beyond*, Guilford Press, New York. Salkovskis, P.M. (ed) (1996) *Frontiers of Cognitive Therapy*, Guilford Press, New York; Beck later expanded the application of his theory to the treatment of anger, hostility and violence in his *Prisoners of Hate* (New York: HarperCollins, 1999). Today, the benefits of cognitive therapy are so widely recognized that in England, physicians are required to receive training in the technique as part of their clinical preparation. Behavior therapy has very little relationship to American behaviorism, but instead grew out of the tradition of British behaviorism which was far more cognitive, and came out of the discovery that patient symptoms often persisted because the patient had made inaccurate interpretations of the meaning of their behaviors or adopted counterproductive behaviors while trying to cure their initial complaint. In both Cognitive and Behavioral Therapy, the patient is taught, through structured exercises, to correctly interpret events and sensations, and adopt effective behaviors.
[22] See, for example, Galanter, M. (1990). Case congregations and their careers, *Law & Society Review* **24**(2), 371–395. Case congregations, distinct from the individual cases which compose the congregation much as individual bricks make up a brick wall, provide a graphic depiction of trends and allow legal strategists to place a given trial in broader perspective.
[23] For a review of case congregations in both the physical and behavioral sciences, see Edwards, C.N. (ed) (2005). Shifting sands of justice, *Responsibilities and Dispensation: Behavior, Science, and American Justice*, Four Oaks Press, Dover.
[24] *Daubert v. Merrell Dow Pharm., Inc.* 509 U.S. 579 (1993).
[25] For a discussion of this topic, see also Cole, S.A. (2007). Where the rubber meets the road: thinking about expert evidence as expert testimony, *Villanova Law Review* **52**, 803–839.
[26] For an argument that defendants should be allowed to use behavioral evidence more extensively, see George, J.A. (2008). Offender profiling and expert testimony: scientifically valid or glorified results? *Vanderbilt Law Review* **61**, 221–260.
[27] Thompson, S.K. (2007). A brave new world of interrogation jurisprudence? *American Journal of Law and Medicine* **33**, 341.
[28] American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition, Text Revision, APA Press, Washington, DC.
[29] WHO (1990). *International Classification of Diseases and Related Health Problems*, 10th Revision, World Health Organization, Geneva.
[30] For an analysis of science in the context of the law, see Edwards, C.N. (1998). In search of legal scholarship: strategies for the integration of science into the practice of law, *Southern California Interdisciplinary Law Journal* **8**(1), 1–38.

CARL N. EDWARDS

Behavioral Toxicology

Introduction

The ability of drugs to affect behavioral changes is arguably one of the most common issues confronting forensic practitioners investigating violent crimes. This has as much to do with the frequency of their use by sections of the community as to their often profound adverse effects on humans.

The risks of violent death including homicide and suicide are associated with the use of alcohol or illicit drug including those people living with drug-using persons but who themselves do not abuse alcohol or drugs [1, 2].

Drugs of most concern include alcohol, all of the amphetamine class of drugs, cocaine, the benzodiazepines and the hallucinogens such as ketamine and PCP, although many other drugs can have profound behavioral effects directly or can modify the behavioral effects of other drugs.

National household surveys on drug use conducted in the United States indicate that about 10 and 6% of males and females respectively have used illicit drugs in the previous month. This is mainly cannabis followed by cocaine [3]. The prevalence of drugs capable of causing behavioral changes is significant [4]. In victims of homicide, the most common is alcohol followed by the benzodiazepines, opiates, and stimulant drugs.

This article explores the notion of behavioral toxicity and its relevance to forensic medicine and provides some of the most common examples to illustrate the significance of this phenomenon.

What is Behavioral Toxicity?

This is a term used to define changes in behavior that impact adversely on rational thought and actions that may predispose to violence or other behaviors and have been caused by the effects of chemical substances. Most typically, these substances are illicit drugs, but this can also be caused by the inhalation of volatile substances, e.g., petroleum products, solvents, etc., the use of natural products that contain psychoactive compounds, or the misuse of prescribed medications (Table 1). The net effect of these behaviors is the association with criminal acts.

Table 1 Common drugs causing behavioral toxicity

Drug or drug class	Relative risk of behavioral toxicity
Alcohol	Moderate
Amphetamines	High
Anabolic steroids	Low to moderate
Benzodiazepines	Low to moderate
Cannabis	Low to moderate
Cocaine	High
LSD	Low to moderate
Mescaline	Low to moderate
PCP and ketamine	Moderate to high
Psilocin/Psilocybin	Low to moderate
Volatile substances (hydrocarbons, etc)	Moderate

This phenomenon is illustrated by excessive alcohol consumption, in which a person becomes disinhibited leading to an outburst of aggression following an argument resulting in harm to another person. While the actions cannot be blamed on the drug itself, the drug has modified the person's behavior to increase the propensity to violence and often also the severity of the violence. This is a well-known association with excessive alcohol use; however, with some drugs (e.g., benzodiazepines) changes in behavior are not so predictable.

Mechanisms of Action

Chemical substances may act on receptors in brain or bind to membranes affecting cellular functions in the brain modifying normal nerve function. The transmission of signals by nerves involves the use of chemicals known as *neurotransmitters*. These neurotransmitters include substances such as norepinephrine (noradrenaline), serotonin (5-hydroxy tryptamine or 5-HT), dopamine, *N*-methyl-D-aspartic acid (NMDA), and gamma-aminobutyric acid (GABA).

Interference in the actions of neurotransmitters can occur by blocking the effects of these substances on receptors located on adjacent tissue (postsynaptic sites) or the nerve ending itself (presynaptic sites) or by modifying the release, reuptake, or metabolism of these substances. For example, cocaine inhibits the reuptake of norepinephrine and dopamine and thereby effectively prolongs the effects of released neurotransmitter. Reuptake is essentially a process to recycle neurotransmitters in which active processes

exist within nerve terminals to reabsorb released neurotransmitter. Amphetamines can increase the amount of neurotransmitter released by the nerve ending thereby increasing the actions of the nerve impulse.

Benzodiazepines block the effects of released GABA. The binding of the benzodiazepine on this receptor has the net effect of increasing the movement of chloride through ion channels in membranes. This process reduces the stimulation of areas in the brain involved in regulating a range of behaviors and physiological functions including those involving emotions and memory.

Some drugs increase the actions of the neurotransmitter by inhibiting the enzyme responsible for metabolizing excess neurotransmitter, monoamine oxidase (e.g., the antidepressants moclobemide, phenylzine, etc).

Alcohol

Alcohol, or more accurately termed as *ethanol*, is by the far the most common drug affecting behavior. Alcohol is a complex drug in terms of its mechanism of action. It is a central nervous system (CNS) depressant and acts to depress various functions in the brain and the nervous systems generally. It resembles the volatile anesthetics such as chloroform, ether, and the modern halogenated anesthetics. Alcohol not only affects membrane function nonspecifically but also affects a number of excitatory and inhibitory amino acid transmitters such as GABA, glutamate, dopamine, and serotonin (5-HT).

Blood ethanol concentrations (BAC) in excess of 0.02% adversely affect coordination skills and cognitive function. These symptoms become quite obvious at BAC 0.1%, causing slurred speech, unsteadiness on the feet, poor judgment, and decision making. Disinhibition is most apparent over 0.1% BAC and is associated with aggression and with loud and outrageous behavior, risk taking, and criminal activity (*see* **Alcohol**) [5].

Amphetamines and Cocaine

This group of strong stimulants include (dex) amphetamine, methamphetamine, and the designer amphetamines such as 3,4-methylenedioxy-metham phetamine (ecstasy, (MDMA)),3,4-methylenedioxy-amphetamine (eve, (MDA)), **N-methyl-benzodio xazoyl-butanamine** (MBDB) and *p*-methoxy-amphe tamine (PMA). They all act as strong stimulants of the CNS and depending on the drug elevate heart rate, they increase anxiety, improve mood, and reduce the effects of fatigue. Ecstasy and some of the other designer amphetamines increase empathy and are often used in nightclubs to increase socialization. Owing to the stimulatory effects of these drugs on the CNS, persons become agitated and talkative. Other behavioral manifestations of strong stimulants include rapid or confused speech and aggressive behavior.

Persons driving motor vehicles using these strong stimulants display less vigilance and are less able to properly respond to multiple inputs (i.e., display impaired divided attention tasks). These drivers are also likely to drift out of the lane of travel, show erratic driving, and are associated with high speed collisions [6].

Prolonged use of amphetamines and cocaine can cause long-lasting or even permanent damage to dopamine- and serotonin-containing nerves in the brain leading to stereotyped behaviors and psychoses [7].

Consequently, it should be of no surprise that these drugs are associated with violence, particularly assault occasioning harm including homicides (*see* **Cocaine**; **Amphetamine**) [8].

Anabolic Steroids

Anabolic steroids are the drugs related to the male sex hormone testosterone and have androgenic activity, or an ability to increase muscle mass and tone. Numerous drugs are available in this group, mostly illicit, as solutions for injection, although some are taken orally. These steroids include stanozolol, nandrolone, methenolone, metandienone, oxymetholone, tenbolone, etc.

These drugs can be taken as small doses of two or more steroids, larger doses in cycles lasting one to three months, or as ever increasing doses as demanded. The net effect is a larger buildup of muscle mass compared to a drugfree situation for a given amount of body building.

A well-recognized side-effect associated with the use of anabolic steroids is the appearance of

mood disorders, irritability, and aggression. Controlled studies in athletes show a significant number of steroid users reporting major mood disorders, including mania, hypomania, and major depression. Psychoses, delusions, aggressive, and violent behavior are also strongly associated with active steroid users. Reports of fits of anger, assault, attempted murder are also linked to steroid use [9].

Personality profiles of men using anabolic steroids also show enhanced physical and verbal aggression and disinhibition [10].

Hallucinogens

Hallucinogens produce an altered perception of reality and include lysergic acid diethylamide (LSD), often sold as "trips" or "tabs". LSD users are often confused and exhibit illogical actions and may result in bizarre behaviors and paranoid delusions.

Phencyclidine (PCP) and its analogs thienyl cyclohexyl piperidine (TCP), phenyl cyclohexylpyrrolidine (PHP), phenyl cyclopentyl piperidine (PCPP), and cyclohexamine (PCE) produce an altered conscious state leading to hallucinations, dysphoria, symptoms of dissociation, and distortion of visual and other sensory signals. PCP shows some actions similar to LSD. Ketamine shares similar properties with PCP and is used widely by veterinary surgeons.

PCP is sold as a powder, flakes, or a liquid. In any of these forms, it can be impregnated in tobacco or cannabis cigarettes, or leaf material. Street doses of PCP range from 1 to 10 mg. Ketamine is sold as solutions for injection, powder, or tablet form. Ketamine is also commonly cut with other illicit drugs such as amphetamine, cocaine, and heroin.

Users of PCP can experience acute psychotic episodes leading to reckless and dangerous actions and it is not surprising that this drug is also associated with violence and dissociated behavior. In some persons, excessive PCP can lead to lethargy/stupor, acute brain syndrome, or even coma; however, surprisingly many can be alert and oriented [11].

Other common hallucinogens include the plant alkaloid mescaline and active ingredient in "magic mushrooms" psilocin/psilocybin. These substances produce physiological and psychological changes including anxiety, compulsive movements, and antisocial behavior (*see* **Poisons: Detection of Naturally Occurring Poisons**).

Benzodiazepines

Benzodiazepines are a large class of minor tranquilizers that are legally available, but are abused either by themselves or in combination with other drugs, often alcohol, amphetamines, and heroin. They include alprazolam, clonazepam, diazepam, flunitrazepam, and temazepam. The drugs sedate and reduce anxiety and in sufficient doses produce unsteady gait, slurred speech, and disorientation. They often affect retention of events while under the influence of the drugs, e.g., cause anterograde amnesia.

Benzodiazepines have adverse effects on driving skills through impairment of psychomotor and cognitive skills such as increased reaction times, poor lane control and poor lane tracking, and impaired divided attention tasks.

In some subjects, disinhibition may be observed with the excessive use of the drug either alone or in combination with other CNS active drugs. Alcohol is a common associated drug that intensifies these behavioral changes. Disinhibition of normal control mechanisms results in aggression and bizarre behavioral changes. These reactions, although not common, can produce disturbing changes when benzodiazepines are abused. In extreme cases, subjects may not have any conscious control over their actions (*see* **Benzodiazepines**).

Psychoses

Aggression can occur even without the direct effects of drugs, particularly in psychiatrically disturbed persons. Several psychotic disorders, including schizophrenia, may be associated with symptoms of acute agitation and aggression. Cannabis intoxication can also cause acute psychotic episodes.

Drug treatment can play an important role in the management of agitated persons. The most commonly used are the conventional antipsychotics and benzodiazepines, often by injection to achieve rapid results.

Tolerance and Withdrawal

Users of essentially all of the drugs mentioned previously will exhibit tolerance, or neuroadaptation if the drugs are used regularly. This applies particularly to the amphetamines and cocaine, although users of cannabis and other CNS depressants can also show

significant signs of drug dependence and tolerance if the drugs are constantly being misused. Chronic use of alcohol will also lead to dependence and tolerance.

Tolerance occurs as the human body becomes accustomed to the pharmacological and physiological effects of the drug causing drug users to compensate by consuming larger amounts of drug. They will consume doses several fold higher during their active phase than when they first started using the drug. Tolerance can occur after days to weeks of use.

Drug-dependent persons will by definition not only exhibit drug-seeking behavior but will also show signs of withdrawal following drug abstinence. Withdrawal symptoms can often be more profound than the direct effects of drug. For strong stimulants, abstinence symptoms are often severe and will be associated with behavioral changes. This includes fatigue (hypersomnolence or rebound fatigue following excessive stimulation of the CNS), depression, and suicidal behavior. Paranoid psychoses and other mood disorders can occur and can be particularly troublesome.

References

[1] Rivara, F.P., Mueller, B.A., Somes, G., Mendoza, C.T., Rushforth, N.B. & Kellermann, A.L. (1997). Alcohol and illicit drug abuse and the risk of violent death in the home, *The Journal of the American Medical Association* **278**, 569–575.

[2] Sinha, R. & Easton, C. (1999). Substance abuse and criminality, *The Journal of the American Academy of Psychiatry and the Law* **27**, 513–526.

[3] National Survey on Drug Use & Health (2006). http://www.oas.samhsa.gov/nsduh.htm (accessed October 7 2007).

[4] Drummer, O.H. & Odell, M. (2001). *The Forensic Pharmacology of Drugs of Abuse*, Arnold, London.

[5] Murdoch, D., Pihl, R.O. & Ross, D. (1990). Alcohol and crimes of violence: present issues, *The International Journal of the Addictions* **25**, 1065–1081.

[6] Logan, B.K. (2001). Amphetamines: an update on forensic issues, *Journal of Analytical Toxicology* **25**, 400–404.

[7] Steele, T.D., McCann, U.D. & Ricaurte, G.A. (1994). 3,4-Methylenedioxymethamphetamine (MDMA, "Ecstasy"): pharmacology and toxicology in animals and humans, *Addiction* **89**, 539–551.

[8] Logan, B.K., Fligner, C.L. & Haddix, T. (1998). Cause and manner of death in fatalities involving methamphetamine, *Journal of the Forensic Science* **43**, 28–34.

[9] Pope, H.G. & Katz Jr, D.L. (1994). Psychiatric and medical effects of anabolic–androgenic steroid use. A controlled study of 160 athletes, *Archives of General Psychiatry* **51**, 375–382.

[10] Galligani, N., Renck, A. & Hansen, S. (1996). Personality profile of men using anabolic–androgenic steroids, *Hormones and Behavior* **30**, 170–175.

[11] McCarron, M.M., Schulze, B.W., Thompson, G.A., Conder, M.C. & Goetz, W.A. (1981). Acute phencyclidine intoxication: incidence of clinical findings in 1,000 cases, *Annals of Emergency Medicine* **10**, 237–242.

OLAF H. DRUMMER

Benzodiazepines

Introduction

Benzodiazepines are a large class of drugs widely prescribed for the treatment of a variety of conditions ranging from sleep disorders, anxiety, muscle tension to some forms of epilepsy. The best known members include Normison, Valium, and Xanax. They are among the most prescribed drugs in the western world.

The benzodiazepines are widely abused, either by themselves or in combination with alcohol, cannabis, and the opiate-like drugs (opioids). Indeed the benzodiazepines are second to alcohol in the prevalence of misuse and subsequently are often detected in a range of forensic cases. The benzodiazepines are classified as central nervous system (CNS) depressants as are alcohol, barbiturates, and the opioids.

The degrees to which benzodiazepines are associated with illicit or nonprescribed use depend on the availability of licit and illicit supplies in particular jurisdictions. Most commonly the drugs are obtained legally through medical practitioners. Persons using more than the prescribed dose or using the drug in a nonprescribed manner (misuse) are often the subject of forensic investigations.

This article provides an overview of the most common members of this class of drugs, their mode of action, pharmacological effects, pharmacokinetics, and applications relevant to forensic science.

Types of Benzodiazepines and Their Medical Applications

Over 50 benzodiazepines are known worldwide although most countries may have only 10–20 of these registered for legal use. The prototype benzodiazepine is chlordiazepoxide of which analogs such as diazepam, nordiazepam, temazepam, and oxazepam are commonly available worldwide. Other analogs include bromazepam, flurazepam, and lorazepam. Clonazepam, flunitrazepam, and nitrazepam are 7-nitro substituted analogs that possess slightly modified properties. Further three-ring variants include alprazolam, estazolam, midazolam, and triazolam. These three-ring analogs are also frequently very potent benzodiazepines requiring much lower doses than the older analogs and are also metabolized by the body through different pathways (see Table 1).

Benzodiazepines are best described pharmacologically as minor tranquilizers, that is they sedate and relax and are used to treat insomnia and various forms of anxiety, including generalized anxiety disorder. Their applications in medicine are often determined by their duration of action.

Members of this class that are short acting tend to be used as hypnotics (induce sleep). The short action is necessary to prevent persons feeling drowsy after a sleep. The half-life, or time taken for the blood concentration to halve for hypnotics tends to be under 12 h, and is often much shorter. The shorter acting benzodiazepines (e.g., midazolam) are also used as induction agents for general anesthesia in some forms of day surgery such as endoscopies.

The longer acting analogs such as diazepam, alprazolam, etc. are used to treat various forms of anxiety and are often prescribed in conjunction with other conditions associated with mood changes, such as depression and psychoses. Posttraumatic stress, obsessive compulsive disorders, and panic attacks associated with phobias can be treated with long acting benzodiazepines. Many of these disorders can lead to significant physiological and behavioral changes often seen in forensic cases. The longer acting benzodiazepines, particularly diazepam, are also used to treat anxiety associated with alcohol withdrawal and muscle spasm, and some are used in the management of some epileptic conditions and to treat seizures following head trauma in intensive care settings [1].

Pharmacology and Pharmacokinetics

Mechanism of Action

Benzodiazepines, like many drugs interact with specific sites, known as *receptors* that result in a cascade of actions ultimately manifesting as the desired action. Benzodiazepines bind to receptors in the brain that increases the affinity of the neurotransmitter gamma-amino-butyric acid (GABA). This interaction increases the chloride (an electrolyte) transport

Table 1 Medical uses and properties of common benzodiazepines[(a)]

Benzodiazepines	Applications	Relative potency[(b)]	Relative half-life
Alprazolam	Sedative	Moderate	Moderate
Chlordiazepoxide	Sedative	Low	Long
Clonazepam	Anticonvulsant, sedative	Moderate	Moderate-long
Diazepam	Sedative, anticonvulsant, relaxant	Moderate	Long
Estazolam	Sedative	High	Moderate
Flunitrazepam	Hypnotic	High	Moderate
Flurazepam	Sedative	Moderate	Long
Lorazepam	Sedative, hypnotic	Moderate to high	Moderate
Midazolam	Sedative, hypnotic	Moderate to high	Short
Nitrazepam	Hypnotic	Moderate	Short to moderate
Oxazepam	Hypnotic	Low	Short
Prazepam	Sedative	Low	Long
Temazepam	Hypnotic	Low	Moderate

[(a)] Most common uses, although there will be some variation from country to country
[(b)] Relates to both the commonly prescribed dose and the associated blood/serum concentrations

Table 2 Route of metabolism and target analytes

Prototype benzodiazepine	Examples	Fate and target substance[a]
Alprazolam	Etizolam, midazolam, triazolam	1-Hydroxy-metabolite
Clobazam	Prototype only	Desalkyl-clobazam
Diazepam	Chlordiazepoxide, clorazepate, pinazepam	Nordiazepam, temazepam, oxazepam
	Temazepam	Temazepam, oxazepam
	Oxazepam	Oxazepam
Flunitrazepam	clonazepam, nimitazepam, nitrazepam	7-Amino metabolite
Flurazepam	Prototype only	Desalkyl-flurazepam

[a] Selected benzodiazepines or metabolites are listed in the order they are metabolized

through ion channels and reducing the arousal of the cortical and limbic systems in the CNS. This is a simplified explanation of their actions, but helps explain the overall effects of these drugs, and how other drugs can modify this response.

Some benzodiazepines and unrelated drugs with similar properties (e.g., zolpidem, zopiclone, zaleplon) bind at subtypes of these receptors causing slightly different effects. Some anesthetics (e.g., propofol) and some anticonvulsants bind at adjacent sites to the benzodiazepine receptor and exert their actions by mediating conformational changes at this receptor. Other drugs cause changes to interconnected receptor systems that ultimately affect the activity of this receptor. These include ethanol, some antidepressants, and the narcotic analgesics (opioids). These drugs therefore modify or even enhance the activity of benzodiazepines when consumed together.

Pharmacokinetics and Metabolism

As the preceding section on the medical applications of benzodiazepines demonstrated that the members of this large class of drugs have widely differing pharmacokinetic properties. As expected, the longer half-life members are detectable in biological specimens for longer than the shorter acting drugs.

Diazepam, arguably the most commonly used benzodiazepine in the world has a half-life of over one day. This means that it is likely to be present in detectable amounts in tissues such as blood for at least a few days since last use. This is accentuated since its major (and active) metabolite; nordiazepam has an even longer half-life (about four days). Diazepam and nordiazepam can therefore be detected in biological specimens for about two weeks depending on the dose administered and the sensitivity of the analytical method.

In contrast, midazolam has a half-life of a few to several hours and will only be measurable for about one to two days since last use. Table 1 summarizes the pharmacokinetic half-life of several examples.

The metabolism of benzodiazepines is essential to understand when considering their actions and their duration of action and detectability in various specimens. Nordiazepam is the major metabolite of diazepam and also of a number of other benzodiazepines. Table 2 illustrates some of the more common examples of benzodiazepines and their corresponding major metabolites.

The related drugs zolpidem, zopiclone, and zaleplon have different pathways of metabolism because they do not have the benzodiazepine-type structure. Zolpidem is converted to carboxy-metabolites that can be targeted in urine. The other members tend not to have metabolites that can be readily targeted for analysis.

Behavioral Effects and Drug Associated Crimes

Benzodiazepines in normal prescribed doses can in addition to their sedative and anxiolytic effects reduce cognitive performance and produce short term memory loss (anterograde amnesia).

Short term memory loss is common and is often seen in patients given a benzodiazepine for a minor procedure. This loss of memory may be partial in that persons do not recollect all what has happened while under the influence of a drug, and/or the recollection is not entirely accurate. Cognition (ability to think and reason) is adversely affected by benzodiazepines and is particularly significant for persons on higher doses or those using longer acting drugs [2].

Not all benzodiazepines produce the same degree of adverse effects. For example, the nonbenzodiazepine hypnotics zolpidem and zaleplon produce

less behavioral changes than the benzodiazepines. Clobazam is also less sedating than the conventional benzodiazepines such as those based on diazepam or alprazolam.

When some benzodiazepines are misused or mixed with other drugs affecting the CNS disinhibitory effects can manifest such as aggressive behavior, mania, and depersonalization. These symptoms are sometimes termed paradoxical since they are seemingly not expected for sedative drugs [3]. These symptoms are more likely in persons vulnerable to psychiatric disorders such as borderline personality disorder. The benzodiazepines most linked to these effects are alprazolam, clonazepam, flunitrazepam, and midazolam.

Benzodiazepines are used in drug-facilitated crimes not only for their sedative actions but also to induce amnesia and hence victims will have poor recall of subsequent events, e.g., following a sexual assault (*see* **Drug-Facilitated Sexual Assault** for further details).

The benzodiazepines are also one of the more common drugs associated with impaired driving. This is due to their sedative effects reducing speed and accuracy of hand-eye movements and other psychomotor tasks required for safe driving. The sedative actions also reduce vigilance and responsiveness to driving situations. Alcohol invariably increases impairment even at blood concentrations of 50 mg per 100 ml of blood (*see* **Drug-Impaired Driving** for further details).

Dependence and Tolerance

As for almost all drugs regular use of benzodiazepines leads to an adaptation to many of the pharmacological effects. Unwanted day-time sedation, memory deficits, residual effects on other cognitive and psychomotor functions are attenuated with long term use. In some situations, particularly when the drugs are abused increasing doses are used to offset the developing tolerance. Some users take several tablets at once far above normal prescribed doses.

Subjects who use benzodiazepines for weeks to months will develop a dependence that leads to drug seeking behavior and signs of withdrawal during abstinence. These symptoms include anxiety, insomnia, restlessness, and a range of behavioral changes. Benzodiazepines enhance the effects of heroin and cocaine, and reduce the impact of withdrawal symptoms of these drugs [4].

Toxicity

The more severe symptoms of benzodiazepines are not as life threatening than other drugs, e.g., alcohol, cocaine, amphetamines, and opiates; however, misuse of benzodiazepines can be life threatening.

An overdose of benzodiazepines typically causes unsteady gait, somnolence, slurred speech, and substantial cognitive deficits. Coma is common.

Most of the benzodiazepines are capable of causing death although this occurs rarely in the absence of other contributing factors, such as respiratory or heart disease. Most commonly, benzodiazepines reported as causing death are alprazolam, diazepam, flunitrazepam, flurazepam, nitrazepam, temazepam, and triazolam.

In many cases presentations are mixed drug ingestions with other CNS depressant drugs such as alcohol and the opiate class of drugs such as heroin, oxycodone, methadone, and morphine.

Interpretation of Results

As with almost all drugs there is little correlation between the blood concentration of benzodiazepines and the response. However, it is still useful to measure the blood concentration of benzodiazepine in specific cases since the concentration provides a guide as to the amount used and possible time since last use.

Prolonged or inappropriate storage can cause some loss of benzodiazepines (and their metabolites). This means that specimens should be kept as cold as physically feasible during transportation and storage. It is strongly recommended to store specimens frozen ($-20\,^{\circ}C$ or lower) even if blood is collected in tubes containing preservative (fluoride/oxalate).

In postmortem settings, benzodiazepines concentrations in blood and tissues remain relatively static since there is little postmortem drug redistribution within the body; however, prolonged exposure to the elements can result in loss of some benzodiazepines, e.g., temazepam and diazepam. The 7-nitrobenzodiazepines are unstable postmortem and convert to their respective 7-amino metabolites. These metabolites can be successfully targeted in such cases.

Table 3 Common analytical techniques for the measurement of benzodiazepines

Screening (initial testing)	Confirmatory testing
Immunoassay (e.g., ELISA[a], CEDIA[b], EMIT[c], TDx[d]) [6]	GC[e] (e.g., ECD[f] detector) [7] GC-MS(MS) [8]
HPLC[g] (UV, DAD[h], etc.) [9]	LC-MS(MS) [10]
GC (various detectors) [11]	
TLC[i] (urine only) [12]	

Table shows some examples of analytical methods.
[a] ELISA: enzyme linked immunosorbent assay
[b] CEDIA: cloned enzyme donor immunoassay
[c] EMIT: enzyme linked immunoassay technique
[d] TDx: fluorescent polarization assay
[e] GC: gas chromatography
[f] ECD: electron capture detector
[g] HPLC: high-performance liquid chromatography
[h] DAD: photodiode array or multiwavelength detection
[i] TLC: thin layer chromatography

Methods of Analysis

Initial Testing

Most clinical and forensic laboratories test for the presence of the class of benzodiazepines in the initial phase of their investigations. These immunoassays are effective for the more common members particularly the lower potency members that have relatively high concentrations in fluids, e.g. diazepam, temazepam, oxazepam, alprazolam. However, they will detect most of the class providing the concentration is sufficient [5] (see Table 3).

This technique using commercial kits enables all members of the class to be detected; however, owing to their differing immunoreactivities the sensitivity to different benzodiazepines and their metabolites vary, and for some may be quite poor. This applies particularly to the more potent members, e.g., lorazepam, triazolam, clonazepam etc.

Initial tests are commonly conducted in urine since the concentration of the benzodiazepine or its metabolite is often much higher than blood. However, immunoassays designed for blood/plasma, or even oral fluid, are commercially available. An example of this is the enzyme linked immunosorbent assay (ELISA).

When a class test is positive further (confirmation) tests are required to detect the benzodiazepine (or its metabolite) that is causing this positive response (*see* **Toxicology: Initial Testing** for further details).

Confirmation Testing

The definitive confirmation method in forensic toxicology is mass spectrometry. This can either be gas chromatography mass spectrometry (GC-MS) or liquid chromatography-mass spectrometry (LC-MS). LC-MS or tandem LC-MS methods are now dominating the measurement for this class of drugs due to its very high sensitivity and specificity and requires very little or no chemical modification to permit chromatographic analysis (see Table 3).

Depending on the specimen and the drug the target benzodiazepine (or metabolite) varies. For example, nordiazepam is the main target metabolite for diazepam and a range of other benzodiazepines metabolized to nordiazepam (e.g., chlordiazepoxide, clorazepate, pinazepam etc.).

In blood and most other specimens the parent drug is the target analyte; however, in urine, a metabolite is often the target. For diazepam, drugs metabolized to nordiazepam, and for temazepam, the most common metabolite is oxazepam and the oxazepam glucuronide. The 7-amino and 7-acetamido metabolites are the main metabolites for the 7-nitro benzodiazepines. The 1-hydroxy metabolite is the target for the tri-ring analogs alprazolam, etizolam, midazolam, and triazolam.

See **Confirmation Testing: Toxicology** for further details.

References

[1] Borchardt, M. (1999). Review of the clinical pharmacology and use of the benzodiazepines, *Journal of Perianesthesia Nursing* **14**, 65–72.

[2] Beracochea, D. (2006). Anterograde and retrograde effects of benzodiazepines on memory, *Scientific World Journal* **6**, 1460–1465.

[3] Mancuso, C.E., Tanzi, M.G. & Gabay, M. (2004). Paradoxical reactions to benzodiazepines: literature review and treatment options, *Pharmacotherapy* **24**, 1177–1185.

[4] Schweizer, E. & Rickels, K. (1998). Benzodiazepine dependence and withdrawal: a review of the syndrome and its clinical management, *Acta Psychiatrica et Neurologica Scandinavica. Supplementum* **393**, 95–101.

[5] (1997). *Recommended Methods for the Detection and Assay of Barbiturates and Benzodiazepines in Biological Specimens*, United Nations, New York.

[6] Augsburger, M., Rivier, L. & Mangin, P. (1998). Comparison of different immunoassays and GC-MS screening of benzodiazepines in urine, *Journal of Pharmaceutical and Biomedical Analysis* **18**, 681–687.

[7] Guan, F., Seno, H., Ishii, A., Watanabe, K., Kumazawa, T., Hattori, H. & Suzuki, O. (1999). Solid-phase microextraction and GC-ECD of benzophenones for detection of benzodiazepines in urine, *Journal of Analytical Toxicology* **23**, 54–61.

[8] Maurer, H. & Pfleger, K. (1987). Identification and differentiation of benzodiazepines and their metabolites in urine by computerized gas chromatography-mass spectrometry, *Journal of Chromatography* **422**, 85–101.

[9] Musshoff, F. & Daldrup, T. (1992). A rapid solid-phase extraction and HPLC/DAD procedure for the simultaneous determination and quantification of different benzodiazepines in serum, blood, and post-mortem blood, *International Journal of Legal Medicine* **105**, 105–109.

[10] Smink, B.E., Brandsma, J.E., Dijkhuizen, A., Lusthof, K.J., de Gier, J.J., Egberts, A.C. & Uges, D.R. (2004). Quantitative analysis of 33 benzodiazepines, metabolites and benzodiazepine-like substances in whole blood by liquid chromatography-(tandem) mass spectrometry, *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences* **811**, 13–20.

[11] Drummer, O.H. (1998). Methods for the measurement of benzodiazepines in biological samples, *Journal of Chromatography. B, Biomedical Sciences and Applications* **713**, 201–225.

[12] Jain, R. (1993). Simplified method for simultaneous determination of diazepam and its metabolites in urine by thin-layer chromatography and direct densitometry, *Journal of Chromatography* **615**, 365–368.

[13] Drummer, O.H. & Odell, M. (2001). *The Forensic Pharmacology of Drugs of Abuse*, Arnold, London.

[14] Drummer, O.H. (2002). Benzodiazepines: effects on human performance and behavior, *Forensic Science Review* **14**, 2–14.

[15] Moffat, A.C., Osselton, M.D. & Widdop, B. (eds) (2004). *Clarke's Isolation and Identification of Drugs*, (3rd edition) The Pharmaceutical Press, London.

[16] Baselt, R.C. (2004). *Disposition of Toxic Drugs and Chemicals in Man*, Year Book Medical Publishers.

OLAF H. DRUMMER

Best Evidence Rule

The best evidence rule is a judicial construct that has its origin in the common law of England. Its meaning is not widely understood by many forensic experts, who are likely to suggest that the "best evidence rule" provides that only the most probative – or the "best" – physical and original documentary evidence is admissible in court. This view contains at least two misunderstandings of the evidentiary principle. First, the rule does not apply to most forms of physical evidence. Second, the rule's name is a misnomer; it does not specify that only the "best" evidence of a fact must be produced.

First, the best evidence rule is a narrow principle. It applies only to writings, X rays, films, videotapes, photographs, motion pictures, and recordings, but is applicable only when the *content* of the evidence is sought to be established [1]. If it is not the content of a document, but only the fact of its execution or transmission that is sought to be established, the best evidence rule does not apply. It also does not apply to other forms of physical evidence [2]. Second, the rule would be more appropriately called *the original evidence* rule in that it expresses merely a preference for the original. In any case, where the content of a writing is sought to be established, the original document must be provided to the court. However, the rule can also be satisfied if a witness can satisfactorily account for the absence of the original document; if the reason for not producing the original is believed by the judge, then a foundation has been laid for the use of secondary evidence. In that event, a copy of the original, or even oral testimony of its content, can be admitted to prove the

content of the writing and the best evidence rule is satisfied [3].

There are two different circumstances in which the best evidence rule applies typically. The first is where one seeks to prove a legal event that can only be demonstrated by a writing. This applies to divorce decrees, deeds to property, a last will and testament, a contract between two or more parties, or other documents containing a legally operative document – a document that creates, alters, or destroys legal relationships. The content of these documents requires that the original be produced, or its absence satisfactorily explained.

A second courtroom scenario wherein the best evidence rule applies is where a witness's sole knowledge of an event comes from what is contained in a writing. The witness is aware of certain facts because he read about their occurrence in a document – a situation where the witness may be said to lack "personal knowledge," as lawyers are likely to argue. In that case, too, the original writing is required, if it is available. Only in the event of unavailability will a reasonable substitute be admitted.

The rule is also relaxed when it comes to proving the content of voluminous writings, recordings, or other evidentiary material, which cannot be conveniently examined in court. In that case, the material may be testified to in the form of a chart, summary, or calculation, provided access to the originals has been provided to the opposing party prior to trial [4].

Some more technical aspects of the best evidence rule are the responsibility of lawyers and usually do not affect expert witness testimony.

References

[1] (U.S.) Federal Rule of Evidence (FRE) 1001.

[2] For example, Federal Rule of Evidence (FRE) 1002 provides that "To prove the content of a writing, recording, or photograph, the original writing, recording, or photograph is required, except as otherwise provided for in these rules . . . " Further, FRE 1003 provides that duplicates are admissible to the same extent as originals unless a genuine question is raised as to the authenticity of the original, for the admissibility of duplicates.

[3] FRE 1004 describes the circumstances that are recognized as accounting satisfactorily for the absence of the original and permitting the use of secondary evidence. FRE 1005 permits public records to be put in evidence by production of a certified copy executed by the custodian of the records.

[4] See, Federal Rule of Evidence (FRE) 1006, which states: "The content of voluminous writings, recordings of photographs which cannot conveniently be examined in court may be presented in the form of a chart, summary, or calculation. The originals, or duplicates, shall be made available for examination or copying, or both, by other parties at reasonable time and place. The court may order that they be produced in court".

Related Articles

Chain of Possession of Tangible Evidence
Demonstrative Evidence
Documents: Authentication of
Foundation Testimony
Hearsay Evidence

ANDRE MOENSSENS

Bias: Observer *see* Interpretation: Observer Effects

Bias on Examination and Interpretation *see* Interpretation: Observer Effects

Biological Agents

Introduction

The term *biological agent* (*bioagent*) refers to *pests and pathogens*, and/or their products that can be used for criminal, terrorist, or warfare purposes. Generally, these are organisms or toxins that are not only microscopic in nature but also include visible

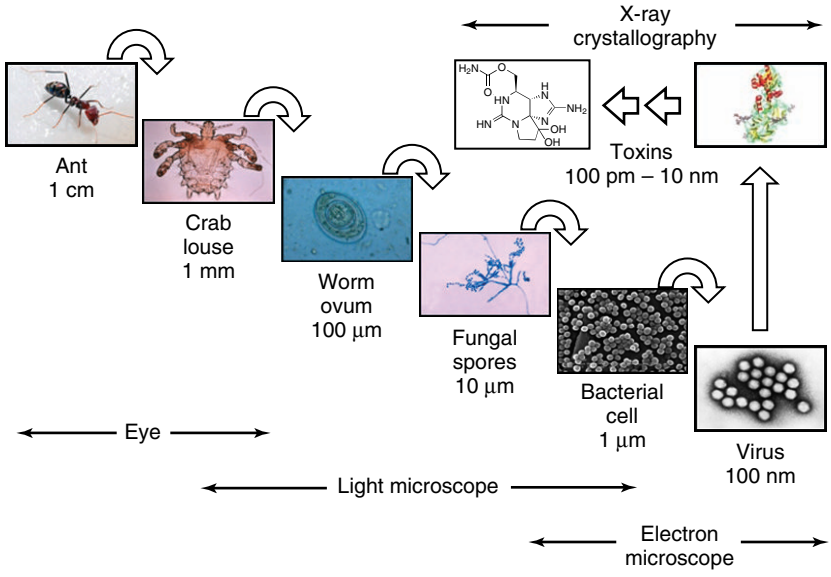


Figure 1 The relative size of biological agents and type of microscopy required to visualize them

organisms such as insects, worms, fungi, and plants. Figure 1 demonstrates the relative size of the different types of biological agents. Most biological agents have the potential to cause disease.

The manner in which biological agents are used or released is classified into four categories. Biowarfare refers to the *military use of biological agents*. Bioterrorism is the use, or threatened use, of biological agents by illicit organizations or individuals to intimidate or coerce a government, civilian population or persons, to further political or social objectives. The term *biocrime* refers to the *use, or threatened use*, of biological agents to further individual objectives. Examples of biocrimes include acts such as the deliberate infection or intoxication of individuals with intent to harm, incapacitate, or intimidate. The last category is bioaccident, which is the unintentional release of a biological agent. Depending on the incident, bioaccidents may be deemed as criminal acts if an accidental release is linked to illegal activities such as smuggling or illegal importation of goods. A good example of the latter is the citrus canker outbreak that occurred in Australia in 2004, which was alleged to have resulted from the illegal importation of plant material [1]. All categories have the potential to cause harm, create fear, generate social unrest, and/or restrict civil liberty. This article concentrates on the use of biological agents in acts of bioterrorism and intentional biocrimes.

Potential targets of bioterrorism and biocrime include humans, animals, and plants, however under a broader definition, ecological targets such as natural forests and marine parks, and industrial targets such as sewage plants and water distribution systems can also be included. Naturally, acts imposed upon human targets impose most fear, whereas those targeted toward agricultural plants and animals are likely to have the greatest economic impact. The release of biological agents may be overt, such as the anthrax letter attacks in the United States in 2001 [2, 3], or covert, such as the deliberate contamination of salad bars and salad dressing with *Salmonella* serotype Typhimurium by the Rajneeshee cult in the United States in 1984 [4].

Recognizing Bioterrorism and Biocrime

Acts of bioterrorism and biocrime can be difficult to detect and investigate. Covert acts have the potential to go unnoticed or be attributed to natural outbreaks. In general, the longer it takes to detect the act, the more complicated the ensuing investigation. A good example of this is the acts of bioterrorism conducted by the Aum Shinrikyo cult in Japan in 1993. The

Aum Shinrikyo cult is primarily remembered for the sarin attack on the Tokyo subway system in 1995; however, this cult had been experimenting with a number of biological agents including Ebola virus, botulism toxin and *Bacillus anthracis* for many years prior [4–6]. On a number of occasions in 1993 the cult aerosolized liquid suspensions of anthrax spores and cells over Tokyo [3, 7]. These releases remained unnoticed until after the sarin attack in 1995 and only emerged following the testimony of cult members and subsequent retrospective investigation. The investigation confirmed that the agent used was the *B. anthracis* Sterne 34F2 strain, a nonpathogenic strain used for the vaccination of animals against anthrax [6]. In this instance, the release had little potential to cause human illness; nevertheless, it demonstrates the possibility that a bioterrorist act could remain undetected for a number of years. Conversely, it is important that natural outbreaks are not attributed to acts of bioterrorism. The global severe acute respiratory syndrome (SARS) outbreak in 2003 [8] and the outbreak of West Nile virus in the United States in 1999 [9] are good examples of outbreaks that could have been falsely attributed to acts of bioterrorism.

The most common acts of bioterrorism and biocrime are hoaxes. The number of hoax incidents that followed the real anthrax letters of 2001 threatened to overwhelm emergency services across the globe; in the United States alone more than 15 000 anthrax hoaxes were made in the following 12 months [10]. Repeat offenders are common with one example of an individual in the United States who sent over 700 threatening letters containing white powder [10]. While hoaxsters cause no physical harm, they do cause considerable stress to their victims, contribute to a national climate of paranoia, cause disruption to services, and unnecessarily burden emergency services. The cost to governments and businesses is staggering. The main threat posed by hoaxes is that they provide a cloak under which real threats may be hidden [11].

Biological Agents as Weapons

There are many organisms and toxins that can be used as biological agents. These include pathogenic fungi, bacteria, rickettsiae, and viruses, as well as toxins from bacteria, fungi, plants, and venoms from animals. The biological agents listed in Table 1 are the core list recommended by the Australia Group for export control. The list assists countries within the Australia Group alliance to ensure that exports comply with the Biological and Toxins Weapons Convention of 1972 (http://www.opbw.org) and, although representative of many of the dangerous biological agents, it is far from exhaustive. Biological agents are designed to cause incapacitation (e.g., salmonellosis, cholera, and brucellosis) or inflict casualties (e.g., anthrax, plague, smallpox, and botulism). Most biological agents can be acquired from the environment and can be cultivated without highly specialized equipment by individuals with limited technical skills. This is particularly true for many fungi and bacteria. Viruses and rickettsiae are more difficult to cultivate and generally require more specialized equipment and technical skills. Venoms can be milked directly from animals, whereas toxins from plants, bacteria, and fungi generally require methods for cultivation and/or purification.

The effectiveness of any biological attack is dependant on the type of agent, the manner in which it is prepared, the method of delivery, and the location of delivery, as well as numerous climatic factors such as wind velocity, wind direction, humidity, temperature, degree of cloud cover, and rainfall [12]. The susceptibility of the host, and the presentation and lethality of disease is dependant on the route of exposure. For example, studies in mice show that ricin has an inhalational LD_{50} of $3-5\,\mu g\,kg^{-1}$ with death at 60 h, a subcutaneous LD_{50} of $24\,\mu g\,kg^{-1}$ with death at 100 h, and an oral LD_{50} of $20\,mg\,kg^{-1}$ with death at 85 h [13] (*Note*: the LD_{50} is the dose of toxin that causes death in 50% of victims.) Ricin can produce severe-to-fatal injury by contact with the eyes, but dermal exposure has little apparent toxicity [13]. Similarly, inhalational anthrax is dependant upon the deposition of spores in the lower airways. The estimated LD_{50} for *B. anthracis* is 2500–55 000 inhaled spores [2, 3]. Since only particles in the $0.5-5\,\mu m$ size range efficiently deposit in the lower airways when inhaled, crude preparations of anthrax powders and aerosolized suspensions are largely inefficient in causing inhalational anthrax (Figure 2). The so-called *weapons grade anthrax* contains purified spore-bearing particles $<5\,\mu m$ in size, treated with chemicals to minimize static and maximize dispersion [2, 3]. The preparation of weapons grade anthrax requires advanced biotechnical capabilities.

Table 1 Australia Group core list of biological agents for export control (July 2006) (http://www.australiagroup.net/en/biological_agents.html)

Viruses

1. Chikungunya virus
2. Congo–Crimean hemorrhagic fever virus
3. Dengue fever virus
4. Eastern equine encephalitis virus
5. Ebola virus
6. Hantaan virus
7. Junin virus
8. Lassa fever virus
9. Lymphocytic choriomeningitis virus
10. Machupo virus
11. Marburg virus
12. Monkey pox virus
13. Rift valley fever virus
14. Tick-borne encephalitis virus (Russian spring–summer encephalitis virus)
15. Variola virus (Smallpox)
16. Venezuelan equine encephalitis virus
17. Western equine encephalitis virus
18. White pox
19. Yellow fever virus
20. Japanese encephalitis virus
21. Kyasanur Forest virus
22. Louping ill virus
23. Murray valley encephalitis virus
24. Omsk hemorrhagic fever virus
25. Oropouche virus
26. Powassan virus
27. Rocio virus
28. St Louis encephalitis virus
29. Hendra virus (Equine morbillivirus)
30. South American hemorrhagic fever virus (Sabia, Flexal, and Guanarito)
31. Pulmonary and Renal syndrome-hemorrhagic fever viruses (Seoul, Dobrava, Puumala, and Sin Nombre)
32. Nipah virus

Genetic elements and genetically modified organisms

1. Genetic elements that contain nucleic acid sequences associated with the pathogenicity of any of the microorganisms in the list
2. Genetic elements that contain nucleic acid sequences coding for any of the toxins in the list, or their subunits
3. Genetically modified organisms that contain nucleic acid sequences associated with the pathogenicity of any of the microorganisms in the list
4. Genetically modified organisms that contain nucleic acid sequences for any of the toxins in the list or for their subunits

Rickettsiae

1. *Coxiella burnetti*
2. *Bartonella quintana* (*Rochalimea qunitana*, *Rickettsia quintana*)
3. *Rickettsia prowasecki*
4. *Rickettsia rickettsii*

Bacteria

1. *Bacillus anthracis*
2. *Brucella abortus*
3. *Brucella melitensis*
4. *Brucella suis*
5. *Chlamydia psittaci*
6. *Clostridium botulinum*
7. *Francisella tularensis*
8. *Burkholderia mallei* (*Pseudomonas mallei*)
9. *Burkholderia pseudomallei* (*Pseudomonas pseudomallei*)
10. *Salmonella* Typhi
11. *Shigella dysentaeriae*
12. *Vibrio cholerae*
13. *Yersinia pestis*
14. *Clostridium perfrigens*, epsilon toxin-producing types
15. Enterohemorrhagic *Escherichia coli*, serotype 0157 and other verotoxin producing serotypes

Fungi

1. Coccidioides immitis
2. Coccidioides posadasii

Toxins (and subunits thereof)

1. Botulism toxins
2. *Clostridium perfrigens* toxins
3. Conotoxin
4. Ricin
5. Saxitoxin
6. Shiga toxin
7. *Staphylococcus aureus* toxins
8. Tetrodotoxin
9. Verotoxin and shigalike ribosome inactivating proteins
10. Mycrocystin (Cyanginosin)
11. Aflatoxin
12. Abrin
13. Cholera toxin
14. Diacetoxyscirpenol toxin
15. T-2 toxin
16. HT-2 toxin
17. Modecin toxin
18. Volkensis toxin
19. Viscumin toxin



Figure 2 Scanning electron micrograph of anthrax spores (CDC PHIL Image 2266–Courtesy of Laura Rose)

When considering the type of agent that might be used for a biological attack, the mode of delivery and the scale of the attack should be considered. The biological agents that cause salmonella and cholera are more likely to be used for small-scale food and waterborne attacks, whereas those that cause anthrax and tularemia are more likely to be used for large-scale airborne dispersal. Smallpox and pneumonic plague are highly contagious, so person-to-person dispersal should be considered. *Yersinia pestis*, the etiological agent of plague, also has the potential to be used for a large-scale airborne attack while Variola virus, the causative agent of smallpox, is unlikely to be used in this manner because of the complexities involved with cultivating large quantities of agent. Similarly, ricin can be delivered *via* aerosol but is inefficient in comparison to *B. anthracis*; tonnes of purified ricin are needed to have the same impact as kilograms of weapons grade *B. anthracis* spores [7]. Ricin would be more effective for small-scale air, food, and waterborne attacks or *via* direct injection. Botulism toxin is the most toxic substance known. The lethal dose for an adult is estimated to be 0.09–0.15 μg intravenously or intramuscularly, 0.7–0.9 μg inhalationally, and 70 μg orally [14]. Since its lethal dose is in the order of hundreds to thousands times less than ricin, botulism toxin would be better suited to large-scale dispersal. The significance of botulism toxin as a bioweapon is demonstrated by the fact that Iraq chose to weaponize more botulism toxin than any other of its known agents. After the 1991 Persian Gulf war, Iraq admitted to the United Nations inspection teams to having produced 19 000 l of concentrated botulism toxin, more than is required to kill the entire human population by inhalation [14].

The Detection of Biological Agents

Unlike acts of terrorism involving chemical or radiological agents, the investigation of acts of bioterrorism or biocrimes is hampered by the lack of real-time detectors for the identification of biological agents. This provides a significant gap in the ability of first responders to determine risk when tasked with the initial investigation of potential biological releases. The technologies used for biological identification have not kept pace with those used for chemical and radiological identification. In part, this is because of the variety, complexity, and large size of toxins and organisms, but it is also due to the high level of sensitivity that is required for biological detection and identification. For example, the lethal dose for sarin is estimated as 1700 mg for a 70-kg man [15], which is approximately 7.3×10^{21} molecules,

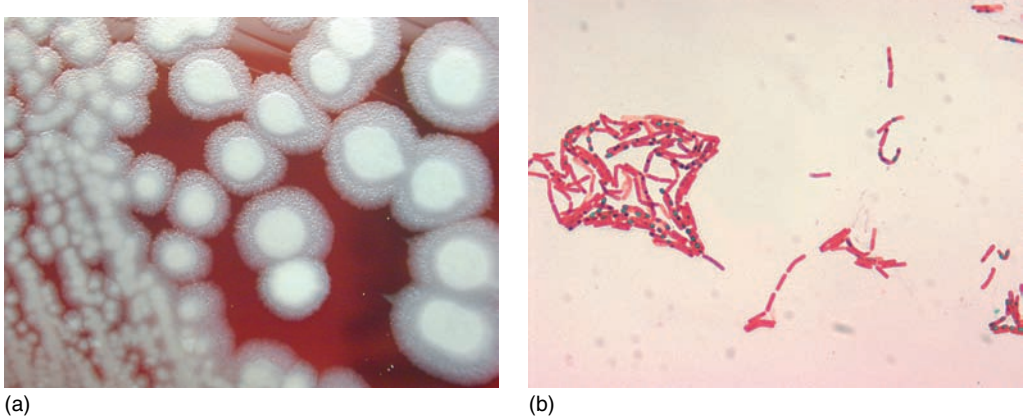
whereas the infectious dose for tularemia is as few as 10 cells [16].

Historically, the common procedures used for the identification of pathogens and toxins require the capture, concentration, and/or amplification of the biological agent. Culture, genetic tests, and immunological assays may all be used to achieve this. Culture is still the gold standard for pathogen identification, but generally requires a minimum of 24–48 h (Figure 3a). Biological staining and microscopy can provide results within minutes; however, these procedures are labor intensive, largely insensitive, and require considerable skill in the performance of techniques and the interpretation of results (Figure 3b). Immunofluorescence microscopy has demonstrated merit as a tool to assist the rapid diagnosis of melioidosis; however, its sensitivity in comparison to culture is poor [17]. Its application outside of a laboratory setting is unlikely.

Immunoassays

To date, immunoassays provide the fastest biological detection technology that is commercially available with most able to provide results within 5–15 min. Most immunoassays use an antibody-mediated capture of their target antigens in lateral flow, enzyme-linked immunosorbent assay (ELISA) or magnetic bead-type formats. Typically, detection and identification is by means of specific antibodies that are covalently bound with visible or fluorescent dyes or with enzymes that generate visible or chemiluminescent reaction products. Tests range in sensitivity and reliability from simple colorimetric handheld tickets through to the more complex portable fluorescence detectors and the laboratory-based ELISA plate readers [18–23] (Figure 4). The specificity of these technologies can be quite variable. Cross-reactivity with related species and/or nonpathogenic variants of the target species is not uncommon and is the main cause of false positives. Conversely, some immunoassays are specific for one or a few forms of the target and therefore have the potential for false negatives. The botulism toxin is an excellent example of the latter; with seven known immunologically different forms (and subvariants) there are no commercial immunological assays that provide detection for all forms of the toxin. Sensitivity can also be an issue, but most immunoassays are developed to be within, or close to, the infectious or lethal dose.

The major downfall of immunoassay technologies is that they are currently limited to one or a few targets in each test. A requirement to test for multiple targets can take hours if the tests cannot be performed in parallel. Notwithstanding that, immunoassay technologies have been developed to detect multiple targets [24, 25]. Some have been adapted for field detection as portable automated pods [26], but these are yet to receive widespread evaluation and acceptance. The hybridization of immunoassay and flow cytometry technologies provides yet another



(a) (b)

Figure 3 (a) Anthrax colonies on blood agar (CDC PHIL Image 1897–Courtesy of Larry Stauffer, Oregon State Public Health Laboratory). (b) Spore stain of anthrax (CDC PHIL Image 1896–Courtesy of Larry Stauffer, Oregon State Public Health Laboratory)

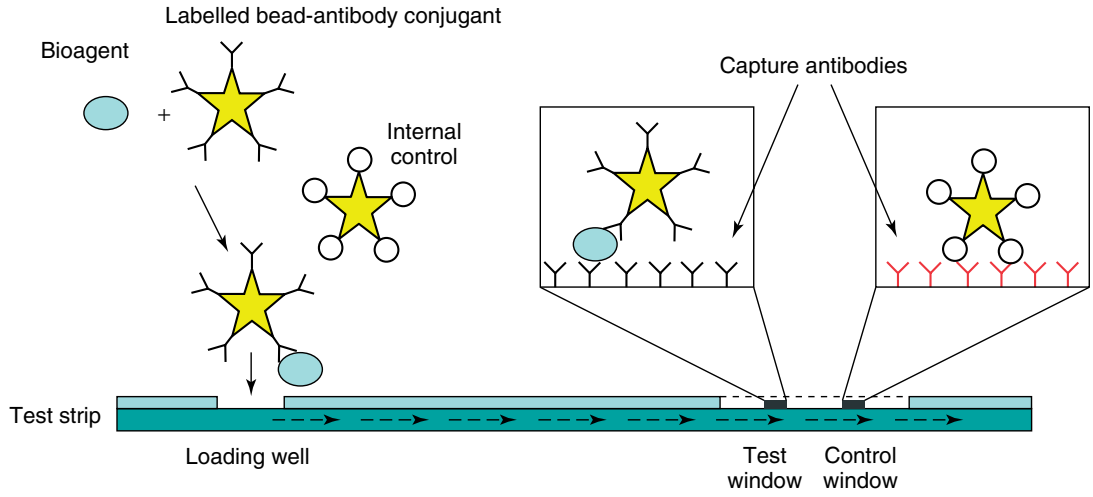


Figure 4 Lateral flow immunoassays

interesting technology that can provide quantitative assessment of multiple agents in a single sample [27]. Assays of this nature have great potential as field units, but currently still require complex and lengthy procedures in order to prepare samples for analysis.

Genetic Assays

Genetic techniques such as the polymerase chain reaction (PCR) and microarrays provide the exquisite sensitivity required for the detection of biological agents in the field and those samples submitted to the laboratory [28, 29].

Real-Time PCR. To date, real-time polymerase chain reaction (RT-PCR) technology has shown the greatest application for the genetic detection and identification of biological agents. RT-PCR combines amplification with simultaneous detection. It is based on the change in fluorescence in the PCR reaction tube, which is proportional to the amplified product. RT-PCR is as sensitive as traditional PCR, but less prone to false positives through contamination because the reaction tubes need not be opened for analysis. In addition, RT-PCR provides a mechanism through which an accurate quantitation of the target can be performed.

Most laboratory-based RT-PCR machines provide results within 90 min and use PCR tube or 96-well plate formats. The LightCycler (Roche Diagnostics GmbH, Mannheim, Germany), which uses capillaries, provides much faster thermal cycling but can be difficult to load and handle. The ruggedized advanced pathogen identification device (RAPID) (Idaho Technology, Salt Lake City, Utah) integrates the LightCycler technology but in a ruggedized portable format designed for field use. The SmartCycler (Cepheid, Sunnyvale, California), also designed for field use, uses a thin plastic capillary tube. The tube is much easier to load and manipulate than the capillaries used for the LightCycler and its design provides very rapid heat exchange allowing the completion of 40 PCR cycles within 15 min. An improved version of the SmartCycler, called the *GeneXpert system* (Cepheid, Sunnyvale, California), integrates the design of the SmartCycler but in a disposable test cartridge format that automates extraction, PCR setup, thermal cycling, and detection. Although portable, the RAPID, SmartCycler and GeneXpert systems, are still quite large and are best deployed in the field *via* a vehicle. The RAZOR thermal cycler (Idaho Technology, Salt Lake City, Utah), about the size and weight of a small car battery, is deployable by hand. In this system, amplification is performed inside wells within a thin flexible plastic pouch, which is slotted into the machine.

One of the major disadvantages of RT-PCR is the inability to multiplex numerous targets. Current technologies either run concurrent reactions for several different targets or multiplex only a few targets.

Multiplexing RT-PCR is technically complex because of the difficulties associated with predicting the interactions that take place among multiple primers, probes, target DNA, and amplified products. In addition, the technology is currently restrained by the number of fluorophores that can be analyzed simultaneously; presently, it is restricted to a maximum of five fluorophores. Another major disadvantage is the time, cost, and labor required to prepare samples for RT-PCR analysis. RT-PCR can be significantly inhibited by contaminants in the sample. Reliability and reproducibility is dependant upon the purification of the nucleic acid target away from potential inhibitors. The automation of extraction reduces the manpower required to process specimens but not necessarily the time. The hybridization of automated extraction and RT-PCR, such as used in the GeneXpert system, considerably simplifies the analysis process.

Microarrays. Typical microarrays consist of a bank of nucleic acid probes bound to the surface of a glass slide. Typical microarrays require a full working day or longer to complete. The advantage of microarrays is that the technology allows the simultaneous analyzes of hundreds to thousands of targets. Their application to diagnostic microbiology is potentially revolutionary [30]. Traditional microarray technology is lengthy, complex, and technically difficult. The integration of automation and software processing has largely removed human involvement into many of the repetitive tasks involved in the processing and interpretation of microarrays. Nevertheless, the traditional process is lengthy and ill-suited to field detection. However, the combination of nucleic acid amplification, lateral flow chromatography, and microarray technologies provides a novel platform from which microarrays can be developed for field deployment [31, 32]. This technology provides the versatility of multiplexing, with an unparalleled specificity afforded by combining amplification and hybridization technologies in series, while providing a level of sensitivity down to one or a few targets. Through miniaturization, automation, and improvements in nucleic acid preparation and amplification, it is likely this technology will be developed for real-time detection.

Real-Time Early Warning Biodetection Systems. A number of early warning systems have been developed for the real-time detection of bioaerosols. The most simple of the systems use light scattering for the detection of particles as they pass through a light beam and the natural fluorescence of biological molecules when energized with ultraviolet light (UV); the biological agent warning sensor (BAWS) [33], VeroTect sensor [34] and biological agent real-time sensor (BARTS) [35] are examples of these systems. These are true real-time sensors. As the particle passes through the UV light, beam light scatter is recorded enabling the determination of the particles shape and size. The presence of a biological particle is then confirmed from the emission spectrum; particles containing the amino acid tryptophan emit a characteristic fluorescence at 325 nm. Since all biological agents have proteins, all types of biological agents are detected. These systems are fully automated with continuous sampling and concurrent analysis, and are able to provide quantitative results. The downside is that the technology is unable to distinguish naturally occurring nonpathogenic organisms from pathogenic agents.

The more complex systems collect and concentrate airborne particles into a liquid sample that are then analyzed by genetic and/or immunoassays. Current technologies are not real time. The autonomous pathogen detection system (APDS) is an example of this type of technology, utilizing a combination of multiplexed immunoassay flow cytometry and then PCR on positives for the detection and identification of aerosols [26, 36]. The combination of immunoassay and PCR virtually eliminates the chance of false positives. The APDS is fully automated, continuously cycling between sampling and analysis. Although highly discriminatory, this technology is still comparatively slow, requiring a minimum of 20 min per analysis.

Emerging Real-Time Biodetection Technologies. A host of new technologies are being developed that have been reviewed elsewhere [23, 26, 33, 37–42]. Many of these are hybrids of existing technologies that improve specificity, sensitivity and speed, and/or automate and miniaturize processes. Most are not real-time processors. Two particularly interesting technologies that are true real-time processors are the microcantilever transducers and nanowire detectors. Both these technologies have theoretical and practical sensitivities down to a single target, can run in real time, can be designed to target any species, and can be multiplexed to detect multiple targets. Further,

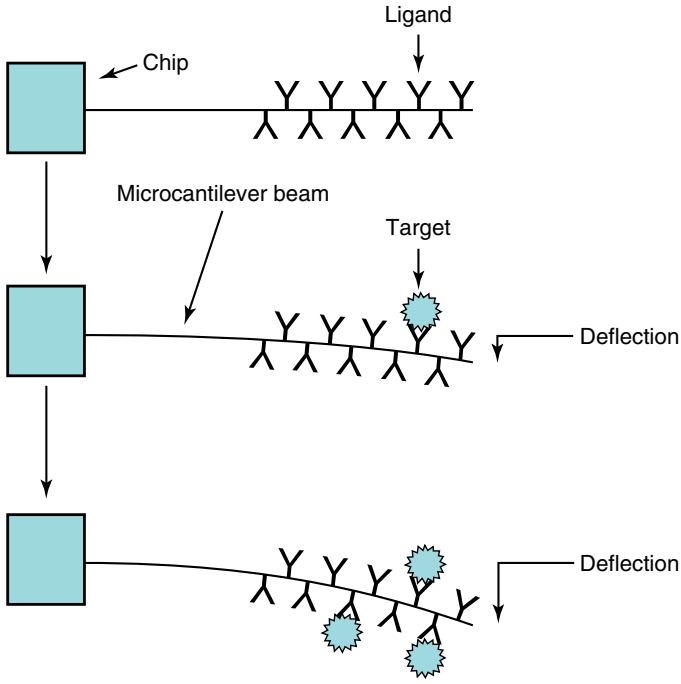


Figure 5 Microcantilever technology

both technologies are good for both chemical and biological targets, which can be detected simultaneously. A microcantilever is a minute cantilever beam that is coated with a ligand that is specific for a target molecule [26, 37, 38, 43, 44] (Figure 5). Any ligand can be used. Typical ligands include chelating agents, lectins, antibodies and antibody fragments, and aptamers and peptides. The binding of target particles to the beam causes deflection of the beam and changes the vibrational frequency within the beam. As more target binds the deflection, change in the vibrational frequency increases, thereby providing a mechanism for both the detection and quantitation of the target. Nanowire detectors work on a similar principle. A nanowire is a microscopic wire coated with a ligand. A change in resistance within the wire occurs as a consequence of a biorecognition event [26, 37, 38] (Figure 6). The level of resistance changes as more target binds, once again providing both detection and quantitation of the target. Since the microcantilevers and nanowires can be coated with affinity molecules for any target (biological or chemical), the opportunity for parallel sensing of multiple targets is enormous.

The Microbiological Investigation of Bioterrorism and Biocrime

The microbiological component of the forensic investigation has the potential to assist the investigation considerably, by providing information about the suspect(s), the source of the biological agent, the methods used to prepare the biological agent, the type of laboratory that was used, the potential location of the laboratory, and an estimate of the period over which the biological agent was prepared. Combined, this information can suggest opportunity and motive, thereby providing assistance to all the fundamental questions behind a successful criminal investigation; who, what, how, where, when, and why.

The first stage of the microbiological investigation is to locate the source of biological agent, determine the extent of contamination, and trace the movements of the agent. Standard microbiological techniques such as culture and viable counts are useful in this endeavor. In the investigation of the 2001 anthrax attack in the United States, these simple techniques were used to locate the letters and trace the time and vicinity in which the letters were posted

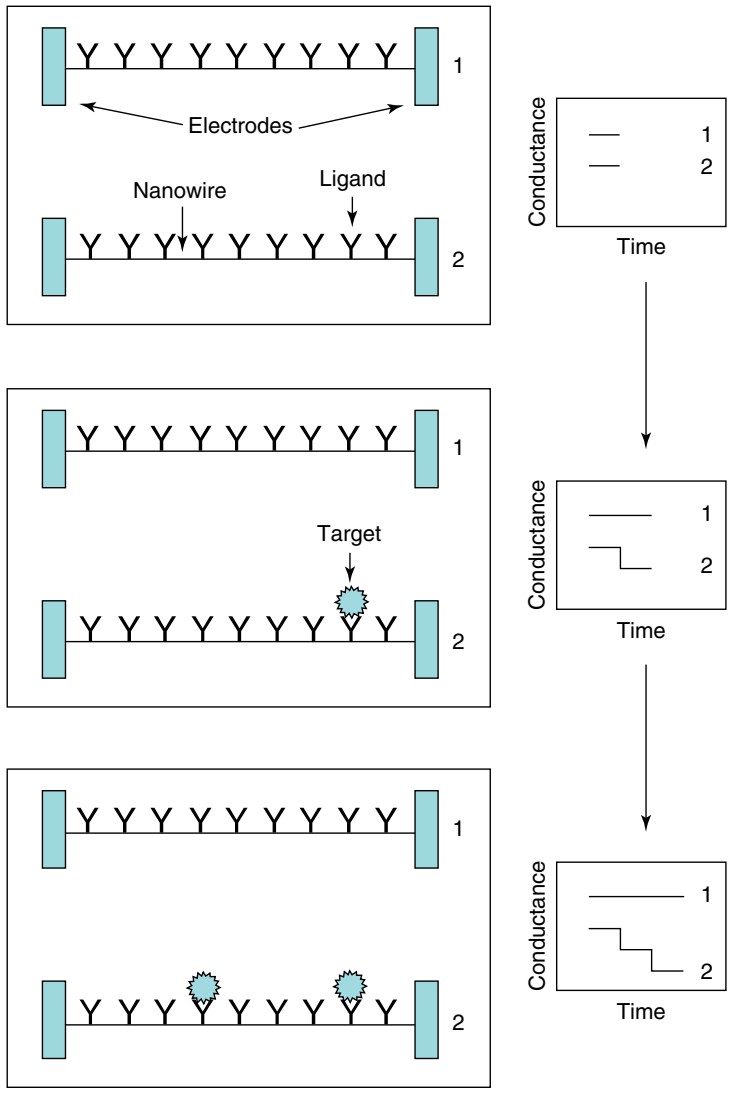


Figure 6 Nanowire technology

[45]. The next critical stage in the microbiological investigation is the identification and characterization of the biological agent. A myriad of cultural, physical, genetic, biochemical, enzymatic, proteomic, lipomic, and metabolic profiling procedures are available that can assist with the species and strain identification of the agent and should aim to provide the highest level of characterization as possible. Identifying the strain and establishing its identity at the site of the attack and other locations of interest allow police investigators to establish links between attacks and between the attack and the perpetrator [46–52]. The identity of the strain may also be used to trace the attack strain to isolate from culture collections, laboratories, organizations, natural outbreaks, or particular geographical regions.

Chemical and microscopic analysis of the original sample can provide a great deal of information about the manufacturing process. Traces of agar indicate that the agent was harvested from culture plates. Chemical analyses can indicate the type of culture medium that was used [53–56]. This can be

related back to a commercial source of the medium and potentially even back to a particular batch of medium. Stable isotopes of a number of elements like carbon, nitrogen, strontium, and lead provide an independent mechanism through which media can be traced [53–55, 57, 58]. The ratio of stable isotopes is directly linked to their geographical region of origin. This has shown great forensic utility for tracing the origins of illicit drugs such as heroin and cocaine, as well as to determine the authenticity of plant-derived food products [59, 60]. In this respect, stable isotope ratios have great application to the tracing of the biological agents such as ricin, abrin, and venoms that are extracted from plants and animals. Although the geographical origin of components within culture media may not necessarily relate to the geographical location at which the agent was cultured, it can be related back to the type of medium, the commercial source of the medium and potentially even the specific batch of medium, thereby indirectly providing evidence for the geographical location of the laboratory in which the biological agent was manufactured [53–55, 57, 58]. Isotopes of hydrogen and oxygen also show geographical influence (Figure 7). As components of water, these can be used as natural tracers to provide direct evidence for the location of the water source used by the laboratory during the manufacture of the biological agent [57, 61]. ^{14}C bomb-pulse dating can provide information about the date of manufacture of the biological agent within a precision of 1–2 years (postnuclear bomb, i.e., 1950 AD) [62, 63]. This can provide information about whether the biological agent has been manufactured recently or not, and if not, from where the biological agent may have been sourced.

Electron microscopy can be used to determine the purity and level of refinement of the biological agent, thereby providing important information about the manufacturing processes. The level of purity, degree of refinement, and presence of flow enhancers is evidence of intent by the perpetrator to prepare a weaponized agent. Chemical analysis for solvents, detergents, stabilizers, flocculants, and preservatives can provide additional information about the process used to manufacture the biological agent [49, 51, 65].

When examined as a whole, the results from these investigations provide a comprehensive profile of the perpetrator and the facilities used to manufacture the agent. It provides evidence for the time and vicinity of release of the agent, the source of the biological agent, the date of manufacture, the geographical location of the manufacturing facility and the manufacturing process. In turn, this provides information about the level of sophistication of the laboratory involved; whether it is likely to be a low-tech clandestine "back yard" laboratory or a well-funded high-tech laboratory. Speculations as to the type and sophistication of the laboratory, as well as the manufacturing equipment, the personal protective equipment (PPE), and the reagents and consumables, that are likely to be present can be made with confidence. Speculation can also be made about the level of experience and knowledge required to manufacture and disperse the agent, which in turn intimates prior career experience and educational background of the perpetrator and suggests access, opportunity, and motive. The medical history of the perpetrator should also be considered, given the potential for infection during manufacture and dispersal. The perpetrator may have been vaccinated or treated for infection, or may have developed symptoms associated with the improper handling of reagents used in the preparatory process.

Conclusion

In theory, acts of bioterrorism have the potential to be extremely decimating. Many of the biological agents such as *B. anthracis* and ricin can be acquired from the environment and prepared using crude equipment and reagents. Further, the cost involved in the small-scale manufacture of crude preparations is low and crude systems such as spray bottles or needles can be used to deliver the agent. In reality, the likelihood of a successful large-scale biological attack is low. History shows that successful attacks are generally targeted at one or a few individuals, and that attempts to deliver biological agents to large numbers of individuals have been unsuccessful [2–4, 66]. This is mostly due to the lack of knowledge, funds, and resources of terrorists and terrorist organizations, which has prevented the development of effective bioweapons.

Crude preparations may be relatively easy and cheap to prepare, but are largely ineffective unless specifically administered to a victim through close contact. Highly refined, weaponized biological agents are much more effective for large-scale attacks, but are very difficult and very costly to prepare, even

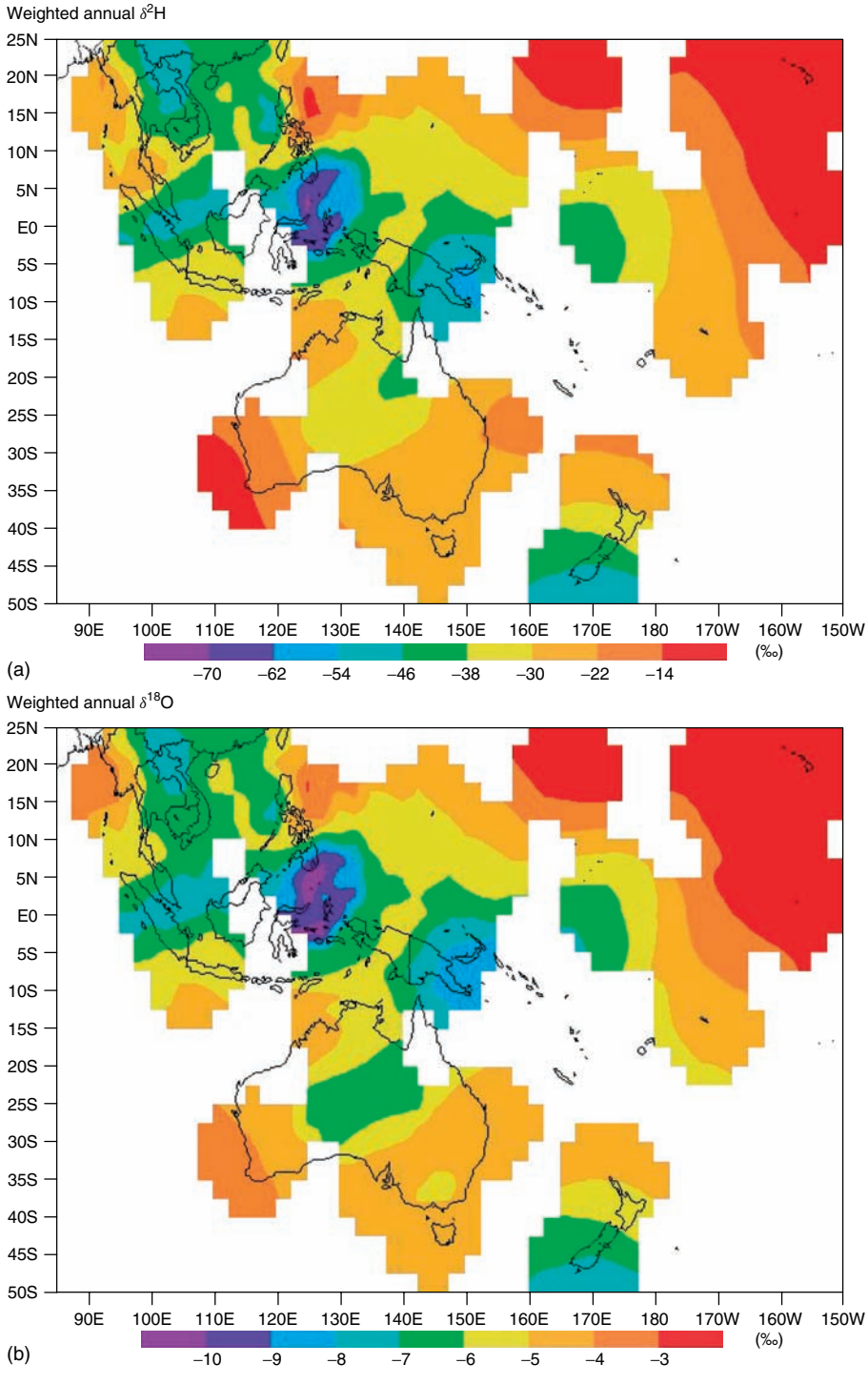


Figure 7 Weighted annual precipitation maps for (a) δ^2H isotope ratios and (b) δ^8O isotope ratios across the South Pacific region [64]

in small-scale volumes. The equipment, technology, knowledge, and financial capital required to prepare weaponized biological agents is not within the reach of most terrorists or terrorist groups.

The 2001 anthrax letter attacks in the United States are a good example of the ineffectiveness of a widespread attack using a crude mode of dissemination. The letters contained sufficient anthrax to kill thousands of people and they leaked as they passed through the postal system, subsequently resulting in the widespread contamination of the US postal system and the postal systems of most other countries across the globe. Yet the result was 22 cases of anthrax and five deaths [67].

When compared to the enormous loss of life caused by naturally occurring diseases, such as Human immunodeficiency virus acquired immune deficiency syndrome (HIV AIDS), malaria, and tuberculosis, the reality is that the threat from bioterrorism and biocrime is low. However, the consequence of even a small-scale attack on society is high.

References

[1] *The Administration by the Department of Agriculture, Fisheries and Forestry of the citrus canker outbreak*. Report to the Senate Rural and Regional Affairs and Transport Legislation Committee June (2006). ISBN 0 642 71657 9.

[2] Tasota, F.J., Henker, R.A. & Hoffman, L.A. (2002). Anthrax as a biological weapon: an old disease that poses a new threat, *Critical Care Nurse* **22**, 21–36.

[3] Inglesby, T.V., O'Toole, T., Henderson, D.A., Bartlett, J.G., Ascher, M.S., Eitzen, E., Friedlander, A.M., Gerberding, J., Hauer, J., Hughes, J., McDade, J., Osterholm, M.T., Parker, G., Perl, T.M., Russell, P.K. & Tonat, K. (2002). Anthrax as a biological weapon, 2002, *Journal of the American Medical Association* **287**, 2236–2252.

[4] Tucker, J.B. (1999). Historical trends related to bioterrorism: an empirical analysis, *Emerging Infectious Diseases* **5**, 498–504.

[5] Takahashi, H., Keim, P., Kaufmann, A.F., Keys, C., Smith, K.L., Taniguchi, K., Inouye, S. & Kurata, T. (2004). Bacillus anthracis incident, Kameido, Tokyo, 1993, *Emerging Infectious Diseases* **10**, 117–120.

[6] Keim, P., Smith, K.L., Keys, C., Takahashi, H., Kurata, T. & Kaufmann, A. (2001). Molecular investigation of the Aum Shinrikyo anthrax release in Kameido, Japan, *Journal of Clinical Microbiology* **39**, 4566–4567.

[7] Kortepeter, M.G. & Parker, G.W. (1999). Potential biological weapons threats, *Emerging Infectious Diseases* **5**, 523–527.

[8] CSR/WHO (World Health Organisation) (2003). *Severe Acute Respiratory Syndrome (Sars): Status of The Outbreak and Lessons Learned for The Immediate Future*, 20 May.

[9] Nash, D., Mostashari, F., Fine, A., Miller, J., O'Leary, D., Murray, K., Huang, A., Rosenberg, A., Greenberg, A., Sherman, M., Wong, S. & Layton, M. (2001). The outbreak of West Nile virus infection in the New York city in 1999, *New England Journal of Medicine* **344**, 1807–1814.

[10] Robbins, I.P. (2004). Anthrax hoaxes, *American University Law Review* **54**, 1–74.

[11] Sullivan, J.P. & Wirtz, J.J. (2008). Terrorism early warning and counterterrorism intelligence, *International Journal of Intelligence and CounterIntelligence* **21**, 13–25.

[12] Beeching, N.J., Dance, D.A.B., Miller, A.R.O. & Spencer, R.C. (2002). Biological warfare and bioterrorism, *British Medical Journal* **324**, 336–339.

[13] Augerson, W.S. (2000). Toxins, in *A Review of The Scientific Literature As It Pertains To Gulf War Illness, Chemical and Biological Warfare Agents*, Vol. 5, RAND, pp. 53–97 Accessed online 13-3-08 1254 hrs.

[14] Arnon, S.S., Schechter, R., Inglesby, T.V., Henderson, D.A., Bartlett, J.G., Ascher, M.S., Fine, A.D., Hauer, J., Layton, M., Lillibridge, S., Osterholm, M.T., O'Toole, T., Parker, G., Perl, T.M., Russell, P.K., Swerdlow, D.L. & Tonat, K. (2001). Botulism toxin as a biological weapon, *Journal of the American Medical Association* **285**, 1059–1070.

[15] Dwyer, A. (ed) (2003). *Jane's Chem-Bio Handbook*, 2nd Edition, Jane's Information Group, Serrey.

[16] Dennis, D.T., Inglesby, T.V., Henderson, D.A., Barlett, J.G., Ascher, M.S., Eitzen, E., Fine, A.D., Friedlander, A.M., Hauer, J., Layton, M., Lillibridge, S.R., McDade, J.E., Osterholm, M.T., O'Toole, T., Parker, G., Perl, T.M., Russell, P.K. & Tonat, K. (2002). Tularemia as a biological weapon, *Journal of the American Medical Association* **285**, 2763–2773.

[17] Wuthiekanun, V., Desakorn, V., Wongsuvan, G., Amornchai, P., Cheng, A.C., Maharjan, B., Limmathurotsakul, D., Chierakul, W., White, N.J., Day, N.P.J. & Peacock, S.J. (2005). Rapid immunofluorescence microscopy for diagnosis of melioidosis, *Clinical and Diagnostic Laboratory Immunology* **12**, 555–556.

[18] *MSA Biosensor Biological Agent Detector: Instruction Manual*, (2007). MSA North America, Pennsylvania.

[19] *Rapid Analyte Measurement Platform (RAMP): Operator's Manual*, (2006). Response Biomedical Corporation, Vancouver.

[20] Emanuel, P.A., Chue, C., Kerr, L. & Cullin, D. (2003). Validating the performance of biological detection equipment: the role of the federal government, *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science* **1**, 131–137.

[21] Chiao, D.J., Wey, J.J. & Tang, S.S. (2008). Monoclonal antibody-based enzyme immunoassay for detection of botulism neurotoxin type A, *Hybridoma* **27**, 43–47.

[22] Milanovich, F., Alabala, J., Colston, B., Langlois, R. & Venkateswaren, K. (2000). *Broad Based Biological Assay Using Liquid Based Detection Arrays*, UCRL-ID-141150, Lawrence Livermore National Library.

[23] Ligler, F.S. (2006). Biosensors for detection of bioterrorist threats, in *Optical Chemical Sensors*, F. Baldini, A.N. Chester, J. Homola & S. Martellucci, eds, Springer, Dordrecht, pp. 437–456.

[24] Khan, I.H., Kendall, L.V., Ziman, M., Wong, S., Medoza, S., Fahey, J., Griffey, S.M., Barthold, S.W. & Luciw, P.A. (2005). Simultaneous serodetection of 10 highly prevalent mouse infectious pathogens in a single reaction by multiplex analysis, *Clinical and Diagnostic Laboratory Immunology* **12**, 513–519.

[25] Swartzmann, E.E., Miraglia, S.J., Mellentin-Michelotti, J., Evangelista, L. & Yuan, P.M. (1999). A homogenous and multiplexed immunoassay for high-throughput screening using fluorometric microvolume assay technology, *Analytical Biochemistry* **271**, 143–151.

[26] Gooding, J.J. (2006). Biosensor technology for detecting biological warfare agents: recent progress and future trends, *Analytica Chimica Acta* **559**, 137–151.

[27] Park, M.K., Briles, D.E. & Nahm, M.H. (2000). A latex bead-based flow cytometric immunoassay capable of simultaneous typing of multiple pneumococcal serotypes (multibead assay), *Clinical and Diagnostic Laboratory Immunology* **7**, 486–489.

[28] Makino, S.-I., Cheun, H.I., Watarai, M., Uchida, H.I. & Takeshi, K. (2001). Detection of anthrax spores from the air by real-time PCR, *Letters in Applied Microbiology* **33**, 237–240.

[29] Fykse, E.M., Langseth, B., Olsen, J.S., Skogan, G. & Blatny, J.M. (2008). Detection of bioterror agents in air samples using real-time PCR, *Journal of Applied Microbiology* [Online]. DOI: 10.1111/j.1365–2672.2008.03750.x.

[30] Perreten, V., Vorlet-Fawer, L., Slickers, P., Ehricht, R., Kuhnert, P. & Frey, J. (2005). Microarray-based detection of 90 antibiotic resistance genes of gram-positive bacteria, *Journal of Clinical Microbiology* **43**, 2291–2302.

[31] Carter, D.J. & Bruce Cary, R. (2007). Lateral flow microarrays: a novel platform for rapid nucleic acid detection based on miniaturized lateral flow technology, *Nucleic Acids Research* **35**, e74. DOI: 10.1093/nar/gkm269.

[32] Chou, C.-F., Lenigk, R., Sadler, D., Changrani, R., O'Rourke, S., Liu, R. & Zenhausern, F. (2003). Rapid genotyping with integrated continuous-flow PCR and bioelectric detection. *Proceedings of the 7th International Conference on Miniaturized Chemical and Biochemical Analysis Systems*, Squaw Valley, October 2003, pp. 1203–1205.

[33] Primmerman, C.A. (2000). Detection of biological agents, *Lincoln Laboratory Journal* **12**, 3–32.

[34] Foot, V.E., Clark, J.M., Baxter, K.L. & Close, N. (2004). Characterising single airborne particles by fluorescence emission and spatial analysis of elastic scattered light. *Proceedings of SPIE – Optically Based Biological and Chemical Sensing for Defence*, October 2004, London, Vol. 5617, pp. 292–299.

[35] Retfalvi, L.A., Newman, E., Boryski, M. & Kacelenga, R. (2004). The challenges of effective biological agent detection in Homeland security applications. *Proceedings of the 8th International Symposium on Protection against Chemical and Biological Warfare Agents*, Gothenburg, June 2004.

[36] Hindson, B.J., McBride, M.T., Makarewicz, A.J., Henderer, B.D., Setlur, U.S., Smith, S.M., Gutierrez, D.M., Metz, T.R., Nasarabadi, S.L., Venkateswaran, K.S., Farrow, S.W., Colston Jr, B.W. & Dzenitis, J.M. (2005). Autonomous detection of aerosolized biological agents by multiplexed immunoassay with polymerase chain reaction, *Analytical Chemistry* **77**, 284–289.

[37] Chambers, J.P., Arulanandam, B.P., Matta, L.L. & Weis, A. (2008). Valdes. Biosensor recognition elements, *Current Issues in Molecular Biology* **10**, 1–12.

[38] Ferrari, M. (2005). Cancer nanotechnology: opportunities and challenges, *Nature Reviews Cancer* **5**, 161–171.

[39] Lim, D.V., Simpson, J.M., Kearns, E.A. & Kramer, M.F. (2005). Current and developing technologies for monitoring agents of bioterrorism and biowarfare, *Clinical Microbiology Reviews* **18**, 583–607.

[40] Ivnitski, D., O'Neil, D.J., Gattuso, A., Schlicht, R., Calidonna, M. & Fisher, R. (2003). Nucleic acid approaches for detection and identification of biological warfare and infectious disease agents, *BioTechniques* **35**, 862–869.

[41] Wang, J. (2007). Nanoparticle-based electrochemical bioassays of proteins, *Electroanalysis* **19**, 769–776.

[42] Bentzen, E., Wright, D.W. & Crowe, J.E. (2006). Nanoscale tools for rapid and sensitive diagnosis of viruses, *Future Virology* **1**, 769–781.

[43] Archibald, R., Datskos, P., Devault, G., Lamberti, V., Lavrik, N., Noid, D., Sepaniak, M. & Dutta, P. (2007). Independent component analysis of nanomechanical responses of cantilever arrays, *Analytica Chimica Acta* **584**, 101–105.

[44] Morshed, S. & Prorok, B.C. (2007). Tailoring beam mechanics towards enhancing detection of hazardous biological species, *Experimental Mechanics* **47**, 405–415.

[45] Beecher, D.J. (2006). Forensic application of microbiological culture analysis to identify mail intentionally contaminated with bacillus anthracisspores, *Applied and Environmental Microbiology* **72**, 5304–5310.

[46] Van Ert, M.N., Easterday, W.R., Simonson, T.S., U'Ren, J.M., Pearson, T., Kenefic, L.J., Busch, J.D., Huynh, L.Y., Dukerich, M., Trim, C.B., Beaudry, J., Welty-Bernard, A., Read, T., Fraser, C.M., Ravel, J. & Keim, P. (2007). Strain-specific single-nucleotide polymorphism assays for the bacillus anthracisAmes strain, *Journal of Clinical Microbiology* **45**, 47–53.

[47] Hoffmaster, A.R., Fitzgerald, C.C., Ribot, E., Mayer, L.W. & Popovic, T. (2002). Molecular subtyping of bacillus anthracis and the 2001 bioterrorism-associated anthrax outbreak, United States, *Emerging Infectious Diseases* **8**, 1111–1116.

[48] Keim, P., Van Ert, M.N., Pearson, T., Vogler, A.J., Huynh, L.Y. & Wagner, D.M. (2004). Anthrax molecular epidemiology and forensics: using the appropriate marker for different evolutionary scales, *Infection, Genetics and Evolution* **4**, 205–213.

[49] Budowle, B., Murch, R. & Chakraborty, R. (2005). Microbial forensics: the next forensic challenge, *International Journal of Legal Medicine* **119**(6), 317–330.

[50] Budowle, B., Johnson, M.D., Fraser, C.M., Leighton, T.J., Murch, R.S. & Chakraborty, R. (2005). Genetic analysis and attribution of microbial forensics evidence, *Critical Reviews in Microbiology* **31**, 233–254.

[51] Murch, R.S. (2003). Microbial forensics: building a national capacity to investigate bioterrorism, *Biosecurity and Bioterrorism: Biodefense Strategy, Practice and Science* **1**, 117–122.

[52] Hofstadler, S.A., Sampath, R., Blyn, L.B., Eshoo, M.W., Hall, T.A., Jiang, Y., Drader, J.J., Hannis, J.C., Sannes-Lowery, K.C., Cummins, L.L., Libby, B., Walcott, D.J., Schink, A., Massire, C., Ranken, R., Gutierrez, J., Manalili, S., Ivy, C., Melton, R., Levene, H., Barett-Wilt, G., Li, F., Zapp, V., White, N., Samant, V., McNeil, J.A., Knize, D., Robbins, D., Rudnick, K., Desai, A., Moradi, E. & Ecker, D.J. (2005). Tiger: the universal biosensor, *International Journal of Mass Spectrometry* **242**, 23–41.

[53] Horita, J. & Vaas, A.A. (2003). Stable-isotope fingerprints of biological agents as forensic tools, *Journal of Forensic Sciences* **48**, 122–126.

[54] Kreuzer-Martin, H.W., Chesson, L.A., Lott, M.J., Dorigan, J.V. & Ehleringer, J.R. (2004). Stable isotope ratios as a tool in microbial forensics–part 1: microbial isotopic composition as a function of growth medium, *Journal of Forensic Sciences* **49**, 954–960.

[55] Kreuzer-Martin, H.W., Chesson, L.A., Lott, M.J., Dorigan, J.V. & Ehleringer, J.R. (2004). Stable isotope ratios as a tool in microbial forensics - part 2: isotopic variation among different growth media as a tool for sourcing origins of bacterial cells or spores, *Journal of Forensic Sciences* **49**, 961–967.

[56] Cliff, J.B., Jarman, K.H., Valentine, N.B., Golledge, S.L., Gaspar, D.J., Wunschel, D.S. & Wahl, K.L. (2005). Differentiation of spores of Bacillus subtilis grown in different media by elemental characterization using time-of-flight secondary ion mass spectrometry, *Applied and Environmental Microbiology* **71**, 6524–6530.

[57] Kreuzer-Martin, H.W. & Jarman, K.H. (2007). Stable isotope ratios and forensic analysis of microorganisms, *Applied and Environmental Microbiology* **73**, 3896–3908.

[58] Kreuzer-Martin, H.W., Chesson, L.A., Lott, M.J. & Ehleringer, J.R. (2005). Stable isotope ratios as a tool in microbial forensics - part 3: effect of culturing on agar-containing growth media, *Journal of Forensic Sciences* **50**, 1372–1379.

[59] Rossman, A. (2001). Determination of stable isotope ratios in food analysis, *Food Reviews International* **17**, 347–381.

[60] Ehleringer, J.R., Casale, J.F., Lott, M.J. & Ford, V.L. (2000). Tracing the geographical origin of cocaine, *Nature* **408**, 311–312.

[61] Kreuzer-Martin, H.W., Lott, M.J., Dorigan, J. & Ehleringer, J.R. (2003). Microbe forensics: oxygen and hydrogen stable isotope ratios in Bacillus subtilis cells and spores, *Proceedings of the National Academy of Sciences* **100**, 815–819.

[62] Tuniz, C., Zoppi, U. & Hotchkis, M.A.C. (2004). Sherlock Holmes counts the atoms, *Nuclear Instruments and Methods in Physics Research* **213**, 469–475.

[63] Zoppi, U., Skopec, Z., Skopec, J., Jones, G., Fink, D., Hua, Q., Jacobsen, G., Tuniz, C. & Williams, A. (2004). Forensic applications of ^{14}C bomb-pulse dating, *Nuclear Instruments and Methods in Physics Research* **223–224**, 770–775.

[64] IAEA (2001). *GNIP Maps and Animations*, International Atomic Energy Agency, Vienna. Accessible at http://isohis.iaea.org.

[65] Velsko, S.P. (2005). *Physical and Chemical Analytical Analysis: A Key Component of Bioforensics*, UCRL-CONF-209735, Lawrence Livermore National Library.

[66] Frerichs, R.L., Salerno, R.M., Vogel, K.M., Barnett, N.B., Gaudioso, J., Hickok, L.T., Estes, D. & Jung, D.F. (2004). *Historical Precedence and Technical Requirements of Biological Weapons Use: A Threat Assessment*, SAND2004-1854. US Department of Energy Sandia Report.

[67] Jernigan, D.B., Raghunathan, P.L., Bell, B.P., Brechner, R., Bresnitz, E.A., Butler, J.C., Cetron, M., Cohen, M., Doyle, T., Fischer, M., Greene, C., Griffith, K.S., Guarner, J., Hadler, J.L., Hayslett, J.A., Meyer, R., Petersen, L.R., Phillips, M., Pinner, R., Popovic, T., Quinn, C.P., Reefhuis, J., Reissman, D., Rosenstein, N., Schuchat, A., Shieh, W.-J., Siegal, L., Swerdlow, D.L., Tenover, F.C., Traeger, M., Ward, J.W., Weisfuse, I., Wiersma, S., Yeskey, K., Zaki, S., Ashford, D.A., Perkins, B.A., Ostroff, S., Hughes, J., Fleming, D., Koplan, J.P. & Gerberding, J.L. (2002). Investigation of bioterrorism – related anthrax, United States, 2001: epidemiological findings, *Emerging Infectious Diseases* **8**, 1019–1028.

Related Articles

Chemical, Biological, Radiological, and Nuclear Investigations

PAUL E. ROFFEY AND BRONWYN C. MORRISH

Biological Stains

Introduction

Over the years, many techniques have been utilized to determine the type of biological material present at a scene or on clothing and the tissue source of that biological material. Profiling a stain using DNA methods has been established as the gold standard in associating a stain to an individual and has been accepted in courts throughout the world. However, it remains critically important to establish the biological source of the DNA profile. Generally, this means identifying semen or saliva in sexual assault cases and blood in assaults and homicides.

Blood

Blood is unique in its appearance in the form of a dried stain and has a distinct reddish-brown color, which quickly becomes recognizable by the experienced analyst. Tools such as an "alternate light source" can be utilized to assist the examiner in detecting trace amounts of blood, especially on dark clothing. Bloodstains absorb ultraviolet light and appear as dark stains on a dark background. However, most bloodstains can be located using a bright white light source, even on dark clothing.

After locating potential bloodstains, "presumptive color tests" are utilized, which turn their color in the presence of the molecule, "heme" [1]. Many chemicals suffice in this regard, the most popular being phenolphthalein, leukomalachite green, *o*-toluidine, and tetramethylbenzidine [2–5]. All these chemicals are colorless in a reduced ionic state. However, in an oxidized state, these colorless reagents turn color rapidly, turning pink, green, or blue green. This color conversion occurs in the presence of an oxidant, e.g., hydrogen peroxide, rapidly in the presence of a catalyst, such as heme (Figure 1). Heme consists of pyrole rings surrounding an atom of iron, and in the presence of protein molecules, comprises the molecule hemoglobin that, of course, is found in red blood cells.

The presumptive color test is an easy test to conduct. A dry, sterile, cotton-tipped swab is rubbed against the suspected stain. A single drop of the testing reagent is added to the swab, followed by a single drop of hydrogen peroxide. An immediate color change indicates that the stain may be blood. The reagents are added in this order for a reason. A color change only after the addition of the testing reagent indicates the presence of strong oxidants and the results can be inconclusive.

A swab moistened with distilled water can be used to collect a small amount of the stain for testing; however, this may rehydrate the stain increasing the chances of bacterial degradation.

A positive presumptive color test indicates that the stain may be blood. The tests are sensitive, reacting with bloodstains that cannot even be seen with the naked eye. However, they may react with other materials such as rust, copper, and vegetable peroxidases. Oxidation of the test reagents will occur over time, so the tests must be read within seconds.

Some laboratories conduct confirmatory tests on extracts from the stains to determine whether the stains are, in fact, blood. "Takayama" and "Teichman" are two such confirmatory tests [6–8]. Small cuttings from the suspected bloodstains are placed on a microscope slide, covered with a coverslip, and the reagent of choice is added to the edge of the coverslip. Capillary action draws the reagent under the coverslip allowing it to react with the stain. Unique rhombic-shaped crystals formed in the presence of the reagents confirm that the stain is blood. However, these reagents react with animal blood as well.

At this point in the analysis, some laboratories will conduct DNA testing (*see* **DNA**) on the bloodstains. The primers used in DNA profiling only react with human and upper primate DNA, and so obtaining a DNA profile indicates that the bloodstain is most likely of human origin. However, with the advent of monoclonal antibody technology, rather simple yet elegantly designed membranes exist that can quickly confirm the presence of human blood. The technology of immunochromatographic membranes that incorporate the use of monoclonal antibodies specific to human hemoglobin or glycophorin A has aided the field of forensic science [9, 10]. Approximately 100 µl of stain extract are added to a well in the cassette. The sample is absorbed onto a pad which contains mobile monoclonal antibodies. This antigen–antibody complex migrates down the membrane where it comes into contact with an immobilized monoclonal antibody. If the extract contains human blood, a visible line will appear indicating that the sample is human blood (Figure 2).

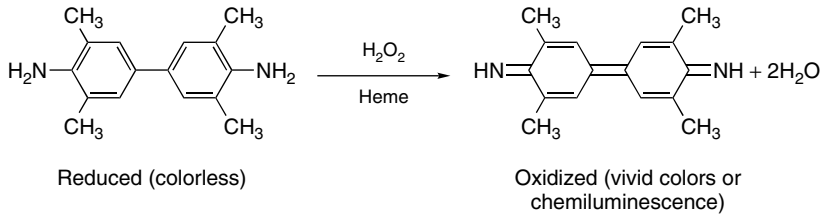


Figure 1 The catalytic conversion mechanism of presumptive color reagents for the presumptive detection of blood

A caveat exists here. Certain membranes are known to react with upper primate and ferret blood. Hence, this matter must be taken into consideration when conducting these tests and preparing reports.

Finally, after thorough analysis of the stain has been completed, a portion of the stain is cut out and is extracted (*see* **Extraction**) for DNA analysis. At the conclusion of the DNA testing, the forensic scientist may conclude that the bloodstain could have originated from a certain individual (based on accompanying statistics (*see* **Case Assessment and Interpretation**)), or exclude someone as a donor of that bloodstain.

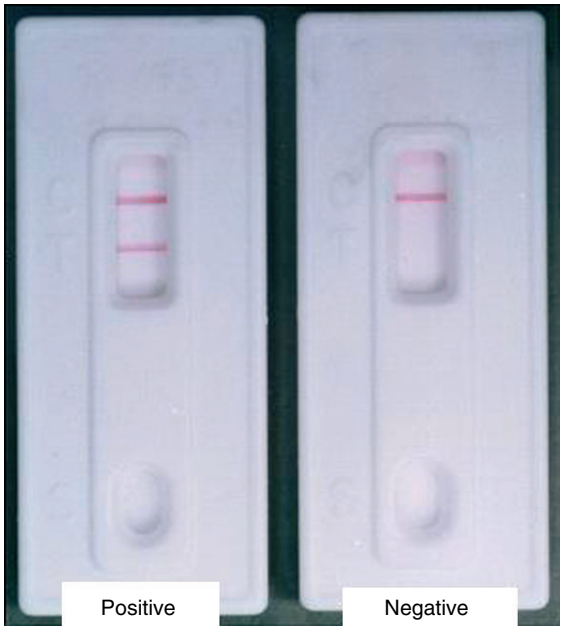


Figure 2 Immunochromatographic membrane testing of blood using monoclonal antibodies specific to human hemoglobin

Semen

Semen is the fluid that is discharged from the penis during ejaculation. It consists of a fluid portion called *seminal fluid*, a complex fluid that contains many proteins and chemicals required for the survival of the "spermatozoon". The spermatozoon is the reproductive cell of the male and consists of a head, neck, and tail region. A good analogy that helps to explain the constituents of semen is a swimming pool. The swimming pool with all of the bathers inside the pool would be analogous to semen. Remove the bathers, or in our analogy, the spermatozoa, and you are left with the pool water, or seminal fluid, with all of the chemicals.

The approach to the analysis of semen mimics that of bloodstain analysis in that articles of clothing and bedding are examined with an alternate light source. Contrary to bloodstains, semen stains "fluoresce" when exposed to ultraviolet light. The source of this fluorescence has been debated but is thought to be due to nonproteinaceous compounds found in semen. Whatever the source, dried semen stains frequently fluoresce intensely under ultraviolet light aiding in their detection.

After suspected semen stains are located, they are tested with a presumptive color test that detects the presence of "acid phosphatase" [11]. Acid phosphatase (*see* **Acid Phosphatase**) is an enzyme (*see* **Enzymes**) found in large amounts in semen and is responsible for the hydrolysis of phosphates resulting in energy for the sperm. Acid phosphatase can be found in much smaller quantities in other biological fluids including vaginal secretions and saliva, hence, the presumptive nature of this test.

A number of chemicals have been used over the years for the detection of acid phosphatase. Basically, all of them involve the hydrolysis of a substrate such as α-naphthyl phosphate by acid phosphatase. This reaction results in the formation of free naphthyl,

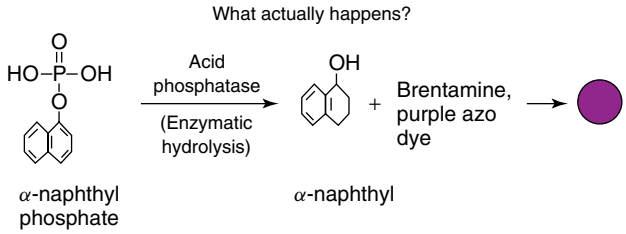


Figure 3 The enzymatic breakdown of α-naphthyl phosphate by acid phosphatase is the basis for the presumptive color test for seminal fluid

which reacts with an azo dye, such as *o*-dianisidine, resulting in a very dark pink color, generally within minutes (Figure 3). The test is very sensitive and reacts with as little as two units of acid phosphatase, known to exist in vaginal secretions. The stronger and faster the test result, the more likely the stain is from semen.

The time tested method for conclusive determination of semen is a microscopic examination and the observation of spermatozoa. The morphology of the spermatozoon is unique from other cells in the body. Using contrast stains such as hematoxylin and eosin or, the more popular nuclear fast red and picroindigocarmine, commonly referred to as *Christmas Tree stain*, the structure of the spermatozoon takes shape [12]. A flattened oval shaped head has its nucleus compacted into the lower half of the head which stains intensely with nuclear stains (Figure 4). An acrosomal cap on the top half of the head prevents staining and appears clear. A short midpiece is attached to the base of the head followed by a single, long flagellum, or tail, which moves the sperm cell. Some laboratories require that a tail, or a portion of a tail is found on the cell to confirm the presence of spermatozoa, but other laboratories allow the identification of a single head lacking a tail. It is known that the longer the time interval between assault and collection of samples, the less likely sperm with tails will be found. The number of spermatozoa decreases and often become misshapen (Figure 5). Typically, sperm can be found in 72-h postcoital swabs and occasionally up to 96 h, the longest surviving constituent of semen in the vagina (Figure 6).

There may be situations in which semen is found lacking spermatozoa (aspermic). This may be due to vasectomy, disease, drug use, and frequency of ejaculation. In such cases, semen can

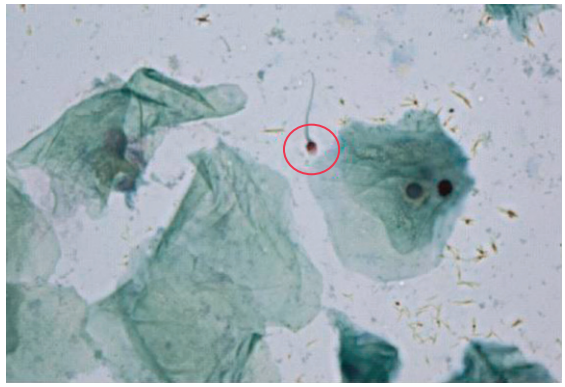


Figure 4 A single sperm cell with its tail among vaginal cells stained with the Christmas Tree reagents

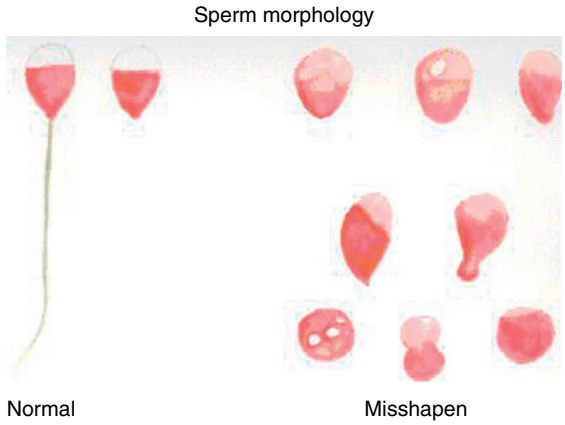


Figure 5 Morphological changes of sperm cells as they degrade

be confirmed using immunochromatographic membranes that utilize monoclonal antibodies specific for either prostate-specific antigen (PSA) [13–15] or

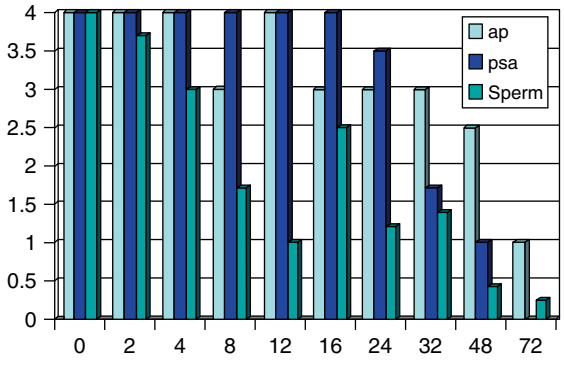


Figure 6 The relative levels of acid phosphatase, PSA, and sperm at various times postcoitus from vaginal samples. Hours postcoitus (x-axis) and relative levels (y-axis). (See Laux, D.L. Forensic detection of semen IV: the persistence of acid phosphatase, P30, and sperm postcoitus, *Midwestern Association of Forensic Scientists Meeting*, Columbus, Ohio, October 24, 2003)

"semenogelin". These proteins are highly specific to seminal fluid. PSA was initially believed to be found only in the prostate of males, but recent studies have shown that it can be found in female urine and breast milk [16, 17]. The amounts of PSA in these fluids is incredibly low compared to that of seminal fluid and proper extraction of the stain or swab can distinguish semen from these fluids [18]. Semenogelin seems to be only expressed by testicular tissue and so may become a more definitive source of seminal fluid [19].

It is known that DNA profiling can be accomplished on aspermic semen samples. There are other sources of nucleated cells in semen besides spermatozoa including epithelial cells lining the vas deferens, Sertoli cells, and white blood cells.

Saliva

Saliva stains are frequently found in sexual assault cases. Once again, preliminary examination is accomplished with an alternate light source. Saliva stains will fluoresce under ultraviolet light, albeit, not as bright as semen stains. The use of varying wavelengths and colored goggles increase the likelihood of finding saliva stains.

After locating possible saliva stains, the stains can either be mapped or cut out and directly analyzed. Mapping has the benefit of observing the location and distribution of saliva stains and allows for photographic documentation. Mapping can be simply accomplished by overlaying stains with moistened starch papers, pressing them against the stains, and developing them with an iodine solution [20]. Iodine reacts with starch to produce a dark purple color. Areas that do not turn purple and remain white indicate the presence of the enzyme "amylase" (Figure 7). Amylase is found in high amounts in saliva and starts the digestion of starches and carbohydrates. Amylase can be found in other biological fluids such as semen, feces, and in bacteria. However, the levels of amylase in saliva are 1000 times the levels found in other body fluids [21, 22]. In addition, amylase levels drop in dried stains as they sit at room temperature and thus, make it more probable that positive amylase results are due to the presence of saliva.

At this time, no agreed confirmation test for the presence of saliva exists. Immunochromatographic

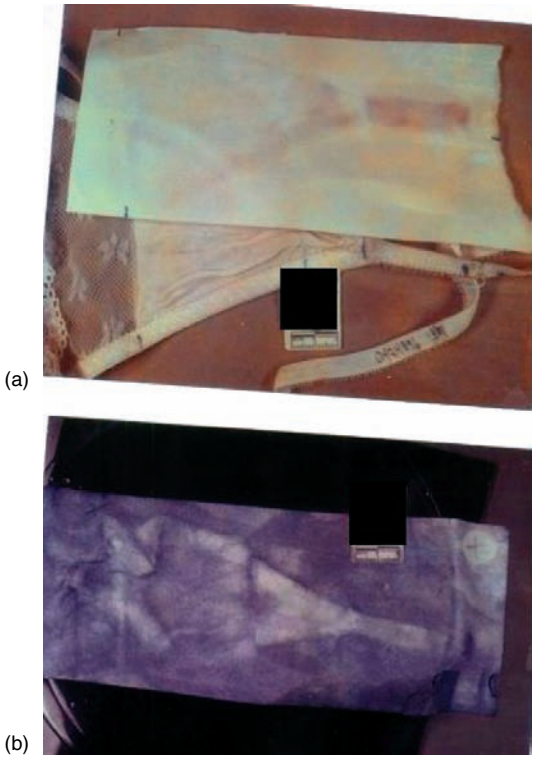


Figure 7 Mapping of underwear for acid phosphatase (a) and amylase (b). Purple-(dark) stained areas (a) demonstrate acid phosphatase distribution. (b) White (light) areas demonstrate the distribution of amylase

membranes are available that utilize monoclonal antibodies to amylase, however, these have been found to react with human feces, dog feces, and upper primate saliva [23].

DNA profiling results on saliva stains are mixed. It really depends on the number of nucleated cells that are found. Saliva is a liquid and stains are located by the reaction of amylase. A large amount of saliva may be present, but if the number of nucleated cells is small, a DNA profile may not be obtained.

Urine

Urine stains typically have a unique smell due to the presence of uric acid and urea. Occasionally, it may be necessary to determine the presence of urine and attempt DNA profiling on the stains. Assault cases have occurred in which the victims had been urinated upon. In addition, corrections officers have had liquids thrown upon them and determination of the source of the liquid becomes important.

No confirmatory test exists for the presence of urine. All of the tests rely on the determination of various components of urine such as uric acid, urea, and creatinine. "Creatinine" is a component of urine and exists at a relatively high concentration ($311-1790\,\mathrm{mg\,l^{-1}}$) [24]. The classic test for creatinine is called the *Jaffe test* [25]. A portion of the stained area along with a portion of the substrate lacking the stain (a "control" area) is cut from the garment or bedding. A drop of sodium hydroxide (NaOH) is added to the stain followed by a drop of concentrated picric acid. The formation of an intense orange color is a positive indication of creatinine (Figure 8). A clear demarcation between the control area and the orange color makes the determination easier. The test can be run on sterile cotton-tipped swabbings of the stained areas as well.

DNA results from urine stains have been obtained; however, the likelihood of obtaining results using genomic DNA is not high.

Feces

Fecal stains, such as urine stains, have an odorous component that can be used in the analysis. The unique smell is due to the chemicals skatole and indole, which are by products of bilirubin catabolism (part of the breakdown mechanism of hemoglobin).

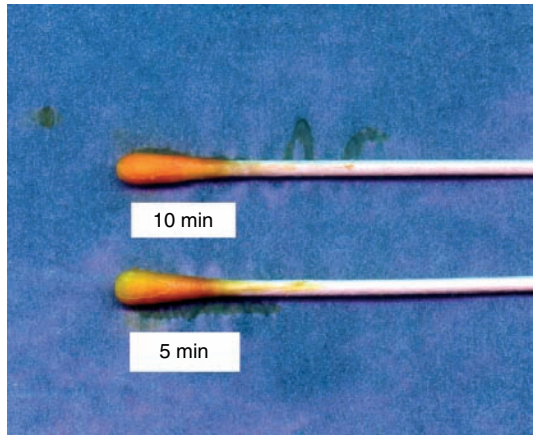


Figure 8 Creatinine color change at 5 and 10 min

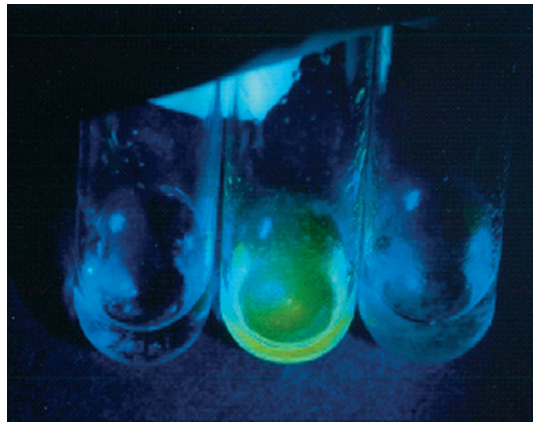


Figure 9 Urobilinogen test for the presumptive detection of fecal material. Tubes (from left) blank, positive fecal sample, and control

Fecal stains generally have a brown appearance and can be located visually. They can be confused with aged bloodstains and may contain traces of blood, so a positive reaction with a blood-testing reagent may occur.

Methods of analysis of fecal stains have centered on chemical and microscopic analyses. Microscopic analysis of an extract from the stain may reveal the presence of plant and animal materials, which would support the conclusion of feces. A simple chemical test can be performed on an extract of the stain dissolved in distilled water. Added to this extract are a few drops of ethanol followed by a

few drops of concentrated zinc chloride in ethanol. Under ultraviolet light, an intense apple green color results indicating the presence of "urobilinogen" [26], a breakdown product of bilirubin and a component of feces (Figure 9).

Note

The views expressed are those of the author and do not necessarily represent those of the Ohio Bureau of Criminal Identification and Investigation.

References

[1] Spalding, R.P. (2005). Presumptive testing and species determination of blood and bloodstains, in *Principles of Bloodstain Pattern Analysis: Theory and Practice*, S.H. James, P.E. Kish & T.P. Sutton, eds, CRC Press.

[2] Cox, M. (1991). A study of the sensitivity and specificity of four presumptive tests for blood, *Journal of Forensic Sciences* **36**(5), 1503–1511.

[3] Garner, D.D., Cano, K.M., Peimer, R.S. & Yeshion, T.E. (1991). An evaluation of tetramethylbenzidine as a presumptive test for blood, *Journal of Forensic Sciences* **36**(5), 1503–1511.

[4] Laux, D.L. (2005). Luminol, in *Principles of Bloodstain Pattern Analysis: Theory and Practice (practical Aspects of Criminal and Forensic Investigations)*, S.H. James, P.E. Kish & T. Paulette Sutton, eds, CRC Press, Boca Raton.

[5] Gross, A.M., Harris, K.A. & Kaldun, G.L. (1999). The effect of luminol on presumptive tests and DNA analysis using the polymerase chain reaction, *Journal of Forensic Sciences* **44**(4), 837–840.

[6] Takayama, M. (1905). Beitrag zur hamatoporphyrinprobe, *Vierteljahrschr Gerichtl Medicine Sanitaetswes* **29**((3F), Suppl.), 232–240.

[7] Teichman, L. (1853). Ueber die krystallisation des organischen bestandtheile des blutes, *Zeitschrift für Rationelle Medicin* **3**, 375–388.

[8] Hatch, A.L. (1993). A modified reagent for the confirmation of blood, *Journal of Forensic Sciences* **38**(6), 1502–1506.

[9] Hochmeister, M.N., Budowle, B., Sparkes, R., Rudin, O., Gehrig, C., Thali, M., Schmidt, L., Cordier, A. & Dirnhofer, R. (1999). Validation studies of an immunochromatographic 1-step test for the forensic identification of human blood, *Journal of Forensic Sciences* **44**(3), 597–602.

[10] Misencik, A. & Laux, D.L. (2007). Validation study of the seratec hemdirect hemoglobin assay for the forensic identification of human blood, *Midwestern Association of Forensic Scientists Newsletter*, **36**(2), 18–26.

[11] Laux, D.L. (2003). Forensic detection of semen I: the acid phosphatase test, *Midwestern Association of Forensic Scientists Newsletter* **32**, 6–10.

[12] Allery, J.P., Telmon, N., Mieusset, R., Blanc, A. & Rouge', D. (2001). Cytological detection of spermatozoa: comparison of three staining methods, *Journal of Forensic Sciences* **46**(2), 349–351.

[13] Hara, M., Inorre, T. & Fukuyama, T. (1971). Some physico-chemical characteristics of gamma-seminoprotein, an antigenic component specific for human seminal plasma, *Japanese Journal of Legal Medicine* **25**, 322–326.

[14] Li, T. & Beling, C.G. (1973). Isolation and characterization of two specific antigens of human seminal plasma, *Fertility and Sterility* **24**, 134–144.

[15] Sensabaugh, G.F. (1979). Isolation and characterization of a semen-specific protein from human seminal plasma: a potential new marker for semen identification, *Journal of Forensic Sciences* **23**, 106–115.

[16] Pollen, J.J. & Dreilinger, A. (1984). Immunohistochemical identification of prostatic acid phosphatase and prostate specific antigen in female periuretheral glands, *Urology* **23**(3), 303–304.

[17] Yu, H. & Diamandis, E.P. (1995). Protease prostate specific antigen in milk of lactating women, *Clinical Chemistry* **41**, 54–60.

[18] Laux, D.L. & Custis, S.E. (2004). Forensic detection of semen III. Detection of PSA using membrane based tests: sensitivity issues with regards to the presence of PSA in other body fluids, *Midwestern Association of Forensic Sciences Newsletter* **33**, 33–39.

[19] Sato, I., Kojima, K., Yamasaki, T., Yoshida, K., Yoshiike, M., Takano, S., Mukai, T. & Iwamoto, T. (2004). Rapid detection of semenogelin by one-step immunochromatographic assay for semen identification, *Journal of Immunological Methods* **287**, 137–145.

[20] Wurster, J.W. & Laux, D.L. (1990). A rapid amylase mapping procedure, *Midwestern Association of Forensic Sciences Newsletter* **19**, 48–49.

[21] Auvdel, M.J. (1986). Amylase levels in semen and saliva stains, *Journal of Forensic Sciences* **31**(2), 426–431.

[22] Rushton, C., Kipps, A., Quarmby, V. & Whitehead, P.H. (1979). The distribution and significance of amylase containing stains on clothing, *Journal of Forensic Science Society* **19**(53), 53–58.

[23] Misencik, A., Laux, D.L. & Smith, C. (2006). Validation study of rapid stain identification for human saliva, *Presentation at the Midwestern Association of Forensic Scientists Meeting*, Indianapolis, September 2006.

[24] Spierto, F.W., Hannon, W.H., Gunter, E.W. & Smith, S.J. (1997). Stability of urine creatinine, *Clinica Chimica Acta*, **264**, 227–232.

[25] Jaffe, M. (1886). Ueber den niederschlag, welchen pikrinsaure in normalen harn erzeugt and uber eine neue reaktion des kreatinins, *Zeitschrift fur Physiologische Chemie* **10**, 391–400.

[26] Llyod, J.B.F. & Weston, N.T. (1982). A spectrometric study of the fluorescence detection of fecal urobilinoids, *Journal of Forensic Sciences* **27**(2), 352–365.

DALE L. LAUX

Biological Swabs

Purpose

Swabs are used for removing biological samples from people and objects for subsequent analysis. Their most frequent applications are as follows:

- to take reference DNA samples from victims or suspects;
- to remove wet or damp biological fluids from the skin and orifices of victims of sexual assault or corpses; and
- to remove dry biological fluids from nonabsorbent surfaces such as skin, water-proof clothing, footwear, furniture, and flooring.

Reference DNA Samples

A buccal swab or scrape is frequently used for collecting DNA samples as it is a noninvasive method that does not require a medical examiner. Sampling involves repeated firm rubbing of the swab on the inside of the cheek to remove some of the epithelial cells. Some organizations opt to have the damp saliva transferred to FTA paper, which is commercially available, contains chemicals that enhance DNA stability. It allows for easy storage of DNA samples at room temperature and suits automated analysis as samples can be removed with a punch [1]. The absorption and release properties of the swab fabric are critical for the transfer process and organizations using FTA paper favor swabs with sponge-type material.

Removal of Biological Fluids

The forensic medical examiner uses dry swabs to remove wet or damp samples of semen, saliva, blood, or any other foreign tissues from orifices, genitalia, skin, and hair of complainants of sexual assault, suspects, or deceased persons.

Dry stains can be removed from skin surfaces at scenes or items examined within the laboratory by wetting the swab with sterile water and rubbing the stained area or the location where biological fluids are suspected to be present. The wet and dry swabbing technique [2] is more successful than wet alone. It is thought that this is due to the fact that the dry cells rehydrate and plump up after the wet swabbing and come away on the dry swab. This technique is less successful on absorbent surfaces such as clothing and soft furnishing.

Small pointed swabs are superior for swabbing small visible stains, especially blood, as the stain concentrates in a small area rather than excessive dilution across a larger swab head. Some laboratories recommend using this type of swab for removing foreign cells from under fingernails rather than cocktail sticks.

Dry or wet swabbing of clothing is used by some organizations to remove epithelial cells to determine the DNA of the wearer, rather than extracting pieces of the fabric. Other laboratories have replaced this technique with mini-taping (see below).

Preservation of Biological Material on Swabs

Damp swabs are an inhospitable environment for DNA and prone to bacterial and fungal growth. Hence, swabs need to be either air-dried or frozen after sampling. Concerns about the possibility of contamination while air-drying have increased with improved DNA sensitivity. Many laboratories freeze the swabs, even though this leads to long-term storage problems. Some organizations use cardboard or paper containers or modified swab containers with a removable plug, both of which allow drying after packaging. Air-drying instruments with internal racks to hold the swabs are popular in the United States.

Swab Types

The most commonly used swab consists of a shaft and handle (approximately 16 cm in length) with a piece of absorbent material at one end. Plastic shafts are preferable to wooden ones as the latter can absorb liquid during the DNA extraction process.

Although many different types of material are used in the swab heads, there is little published data on their relative efficiency for recovery and release of biological fluids. Windram *et al.* [3] report superior yields of DNA from foam popules relative to nylon, polyester, and cotton. Overall, recovery of biological material from swabs is remarkably poor. Unpublished work suggests that less than 25% of spermatozoa are

recovered from swabs. Gartside *et al.* [4] estimated an extraction efficiency of 1% or less for prostate-specific antigen from cotton swabs. Most research, to improve recovery of DNA from swabs, focuses on the optimization of extraction methods [5–7].

Although swabs produced for medical use are sterile, they are not necessarily DNA free. There are rare reports of contamination by manufacturing staff [8]. Treatment with ethylene oxide gas has been shown to be superior to UV or other radiation in removing DNA contaminants from swabs [9]. The poisonous property of the gas means that the sterilization process is limited to approved companies.

Use of Mini-Tapes to Remove Biological Samples

Swabbing of surfaces has been replaced in some laboratories by mini-taping. This method uses small pieces of adhesive tape to remove cells from clothing, weapon handles, and other items [10]. The increased recovery of foreign DNA from clothing is significant. Caution in the interpretation of results is required, due to the possibility of detecting DNA, following innocent secondary transfer.

References

[1] Tack, L.C., Thomas, M. & Reich, K. (2007). Automated forensic DNA purification optimized for FTA card punches and identifiler STR-based PCR analysis, *Clinics in Laboratory Medicine* **27**, 183–191.

[2] Sweet, D., Lorente, M., Lorente, J., Valenzuela, A. & Villanueva, E. (1997). An improved method to recover saliva from human skin: the double swab technique 2, *Journal of Forensic Sciences* **V42**(2), 320–332.

[3] Windram, K., Miller, W., Ward, D., Silenieks, E. & Henry, J. (2005). *Comparison of Swab Types for the Recovery of Trace DNA in Forensic Investigations, Biology Report*, R73, Forensic Science South Australia.

[4] Gartside, B., Brewer, K.J. & Strong, C.L. (2003). Evaluation of Prostate-Specific Antigen (PSA) extraction efficiency from forensic samples using the Seratec® PSA semiquant semiquantitative membrane test, *Forensic Science Communications* **5**(2), www.fbi.gov/hq/lab/fsc/backissu/april2003/gartside.htm.

[5] Allard, J., Baird, A., Davidson, G., Jones, S., Lewis, J., McKenna, L., Weston, C., Scrimger, D. & Teppett, G. (2007). A comparison of methods used in the UK and Ireland for the extraction and detection of semen on swabs and cloth samples, *Science & Justice* **47**(4), 160–168.

[6] Norris, J., Manning, K., Linke, S., Ferrance, J. & Landers, J. (2007). Expedited, chemically enhanced sperm cell recovery from cotton swabs for rape kit analysis, *Journal of Forensic Sciences* **52**(4), 800–805.

[7] Giles, R. (2008). *Improved Methods for the Elution and Extraction of Spermatozoa from Sexual Assault Swabs*, *Forensic Magazine*® RSS, www.forensicmag.com/articles.

[8] Sullivan, K., Johnson, P., Rowlands, D. & Allen, H. (2004). New developments and challenges in the use of the UK DNA database: addressing the issue of contaminated consumables, *Forensic Science International* **146**, S157–S176.

[9] Shaw, K., Sesardic, I., Bristol, N., Ames, C., Dagnal, K., Ellis, C., Whittaker, F. & Daniel, B. (2008). Comparison of the effects of sterilisation techniques on subsequent DNA profiling, *International Journal of Legal Medicine* **122**(1), 29–33.

[10] Hall, D. & Fairley, M. (2004). A single approach to the recovery of DNA and firearm discharge residue evidence, *Science & Justice* **44**(1), 15–19.

Related Articles

Biological Stains

LOUISE MCKENNA

Biological Warfare *see* Chemical, Biological, Radiological, and Nuclear Investigations

Biometric Devices

Introduction

First records of allowing human fingerprints as evidence in a court of law date back to late nineteenth century, when the Home Office of the United Kingdom accepted every individual's fingerprints to be unique [1]. Owing to the apparent potential of this in forensics, research in the field advanced, and methods

of comparing fingerprint marks became increasingly efficient. Nevertheless, the manual methods involved were still quite laborious and monotonous. So when the demand for the comparisons kept increasing, an effort was made to improve the efficiency of the methods by automating the systems. Advances in computer technology, sensor technology, and pattern recognition led to automated fingerprint identification systems (AFISs) in the past few decades [1, 2]. In the mean time, other disciplines such as image processing, computer vision, machine learning, and signal processing advanced as well, leading to other ways of comparing records of a variety of human traits considered to be useful in authenticating people. Examples of this are facial geometry, hand geometry, voice, vein pattern, and iris recognition. However, it was only in the 1990s that the common ground in the various disciplines was recognized enough to spawn an independent field called *biometrics* [3]. The International Biometrics Group (IBG) uses the following technical definition of biometrics [4]:

> The automated measurement of behavioral or physiological characteristics of a human being to determine or authenticate their identity.

This actually means that in biometrics, computers are used to check to what extent a particular human behavior or physical characteristic matches with similar records in a database. This enables machines to check if a person is who he says he is (verification), or if he can be matched with anyone from a database (identification). The matching reliability strongly depends on the technology and the methods involved. A comprehensive background on many aspects of biometrics can be found in literature [1–3].

Currently, the potential uses for biometrics are not limited to forensics alone, but include prevention of identity fraud, logical access control to information systems, such as computers or networks, and physical access control to facilities such as buildings or across borders. Terrorist attacks have had a distinct influence on investments in and the development of the biometrics market, which has been growing rapidly. The US Visit Program, currently in effect, requires travelers passing the US border control to have their two index fingers scanned and a digital photograph taken, to be compared to those stored in their "biometric passport" for authentication [5]. Such biometric passports, containing a chip with biometric data, are in the process of being installed in more countries around the world.

As the use of biometric systems becomes more widespread, the implications of it for the forensic practice should be investigated. It is of interest in the field of forensics to know about the reliability of the performance of various biometric systems. As such systems often involve computers, usage logs are common. What would the forensic expert say if biometric system log files are being used as an alibi, or to incriminate potential trespassers of the system? Or, how safe is an access control system that uses biometrics really? How difficult is it for unauthorized persons to hack their way into the system? Can criminal suspects hide their sensitive data from the forensic expert's eye by protecting it with biometrics?

This article discusses some of the technologies involved in biometric devices. In addition, we briefly deal with what is involved in evaluating a biometric system's reliability and then focus on a particular aspect of circumventing biometric devices using fake copies of biometric samples.

Human Characteristics Suitable for Biometric Systems

Table 1 lists biometric characteristics that have received attention in the research and development industry because of their potential for automatic identification and verification of persons [1, 6]. Not all of the biometric features from this table have yet yielded commercial applications, gained a foothold in the market, or indeed show promise of doing so in the near future.

There is a wide range of biometric products in the market. For this article, we have focused on the physical characteristics that are currently most commonly used as biometric identifiers in biometric devices, as well as some that show promise. Notably, facial geometry (2D and 3D) shows great promise, although this is not in the scope of this document. The workings of a biometric system and the technologies used in biometric devices that involve the selected biometrics are explained in the following section.

Biometric System Components

The major functional components of a simplified biometrics-based authentication system are (see Figure 1) [7] as follows:

Table 1 List of biometric characteristics having received attention because of their potential for automatic recognition of persons

Physical characteristic		
Biodynamic signature	Finger surface (3D)	Odor
Bioelectric field	Finger vein pattern	Palm print
Bite marks	Finger wrinkles	Pores
Bone sound transmission	Fingerprint	Reflection of acoustic waves in the head
Cardiac pulse	Hand geometry	Retinal pattern
Corneal surface topography	Hand pressure profile	Skin impedance
Dental geometry	Hand thermogram	Skin pattern
DNA	Hand vein pattern	Skin spectrum
Ear	Iris	Smile
Facial geometry (2D/3D)	Knuckle creases	Voice print
Facial thermogram	Lips	
Finger geometry	Nail	
Behavioral characteristic		
Dynamic grip recognition	Handwriting	Tapping
Eye movement tracking	Keystroke dynamics	
Gait	Mouse dynamics	

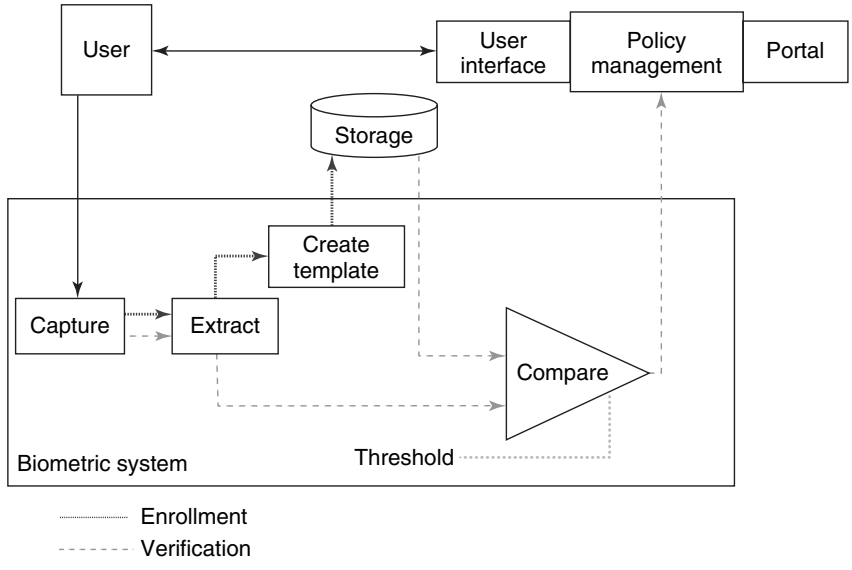


Figure 1 Functional components of a simplified biometric system [Reproduced with permission from Ref. 7. © CESG, 2002.]

1. Capture: acquisition of a raw biometric sample.
2. Extract: conversion of the raw biometric sample to an intermediate form.
3. Create template: conversion of the intermediate data into a user template for storage in a database.
4. Compare: comparison with the information in a stored reference template.

A biometric system has two principal modes: enrollment and verification/identification. Enrollment is the process in which new users are enrolled into the system database. The verification or identification mode involves the comparison of a user with one or more records in the database. The capture part (component 1) of both modes takes place using a biometric device. This device captures the relevant

human characteristic using some kind of sensor, for instance, a light-sensitive chip in a camera or an air-pressure-sensitive chip in a microphone. The data that is ultimately used in the comparison (using component 4) is always an abstraction of the actual biometric. It is an interpretation of sensor output, from which features are extracted (using component 2) that are believed to capture the essentials with which to discern between individuals. Using carefully designed algorithms, a captured fingerprint could, for instance, be abstracted to just the minutiae locations and types. For storing the data in a database in enrollment mode, the abstracted data is combined with general user information into a user template (component 3). This template is stored in the database. The comparison algorithm provides details on whether or not a match has occurred after computing a matching score and checking whether it passes a predefined matching threshold.

Depending on system requirements, the specifics of the system may deviate from this generic case. For instance, a database may hold just the essential abstracted data, or some intermediate form from which the abstraction can still be derived later. The latter can be useful when different methods and algorithms are available for extracting features. For instance, storing an image of a fingerprint instead of the minutiae information as extracted by a particular algorithm, allows different algorithms to be used on database records without having to recapture all the fingerprint data, thus increasing interoperability between different systems that use that same biometric sample as input. In addition, the physical location of the hardware performing the functions may vary, depending on the biometric system setup. A biometric device may be a standalone device, with integrated extraction, database, and comparison functions. The device could also be no more than a capture station, relaying all computations and storage to linked computers and software.

Biometric Device Technologies

Biometric devices come in a great variety. Next, we discuss the technology involved in a number of fingerprint scanner types, a simple iris scanner, a hand geometry scanner, and a hand vein pattern scanner.

Fingerprint Recognition

Fingerprint recognition using biometric devices involves taking fingerprint images from fingertips. Patterns of fingerprint ridges and valleys are detected and stored in a simplified data template that can be used for comparison with other fingerprint templates for verification or identification. The images are acquired with some kind of sensor. A variety of fingerprint sensor technologies are used in commercial fingerprint scanners [1, 6].

Optical Sensors.

- *Frustrated total internal reflection* (*FTIR*): This technology is based on the behavior of light at boundaries from one material to another. The finger is placed on the top side of a glass prism (see Figure 2) [1]. Light entering the prism from a light-emitting diode (LED) on one side of the prism is partially reflected at the contact surface and then captured via a lens with a light-sensitive chip (for instance, charge coupled device (CCD) or complementary metal oxide semiconductor (CMOS) image sensors) on the other side of the prism. Image contrast is caused by the fact that light is randomly scattered or absorbed at the points where skin ridges and the prism make contact, and totally reflected at the valleys, where no contact is made.
- *Surface-enhanced irregular reflection* (*SEIR*): This technology shares some characteristics with

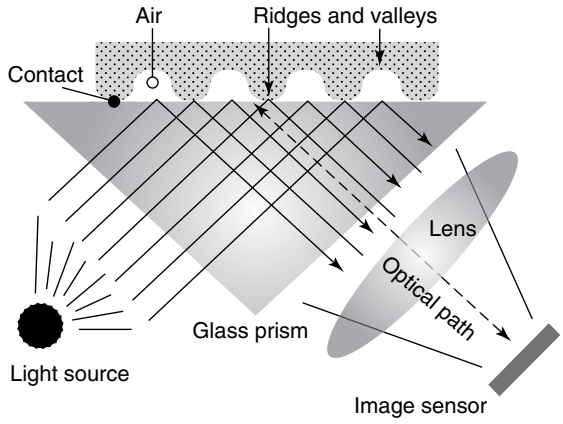


Figure 2 FTIR-based fingerprint sensor [Reproduced from Ref. 1. © Springer-Verlag, 2003.]

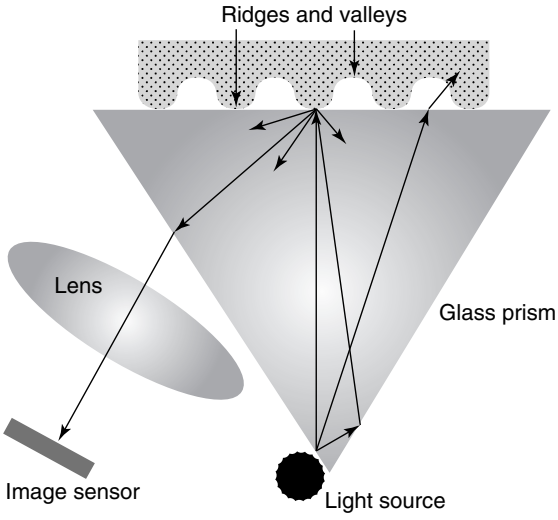


Figure 3 SEIR-based fingerprint sensor

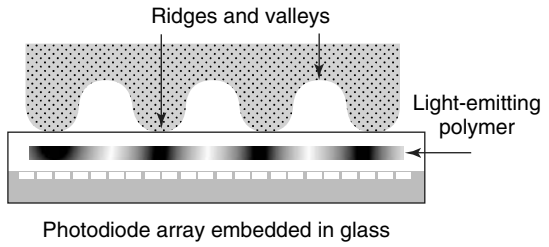


Figure 4 Electro-optical fingerprint sensor [Reproduced from Ref. 1. © Springer-Verlag, 2003.]

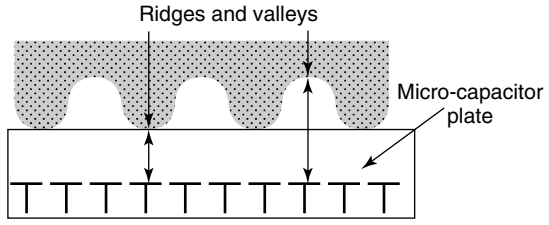


Figure 5 Capacitive fingerprint sensor [Reproduced from Ref. 1. © Springer-Verlag, 2003.]

FTIR technology. The contrast between ridges and valleys is also brought about by the different behavior of light after hitting the ridges and valleys. In this case, however, the light hits the contact surface perpendicularly (see Figure 3), scattering at the ridges but completely passing at the valleys; hence, no scattering occurs. The scattered light is collected by the image sensor and thus produces bright spots for ridges and dark spots at valleys. The sensor manufacturers claim that this technology gives higher contrast images than FTIR technology [8].

- *Electro-optical*: These sensors use a layer of light-emitting polymer, of which the light emission varies on the basis of the potential applied on one side (see Figure 4) [1]. When placing a finger on the polymer surface, ridges touch the polymer and valleys do not, causing the potential to vary across the surface. Thus, a luminous representation of the fingerprint is generated. A second layer, consisting of a photodiode array or a CMOS, converts the light pattern into a digital image.
- *Direct sensing/touchless*: In this case, a high-quality camera is used to focus on the fingertip and directly read the fingerprint. Usually some kind of mechanical support is present to facilitate presenting the finger at a set distance.

Solid-State Sensors.

- *Capacitive*: A capacitive sensor is a two-dimensional array of microcapacitor plates embedded in a chip. The finger skin acts as a second microcapacitor plate (see Figure 5) [1]. Small electrical charges are created between the array and the finger, of which the magnitude depends on the distance between the surfaces. As such, the resulting capacitance pattern represents the ridge and valley pattern of the fingerprint.
- *Electric field:* This type of sensor generates a small radiofrequency field, which is modulated by the highly conductive subsurface of the skin (live skin cell layer). A matrix of antennas receives the modulated analog small-amplitude signal, which is then further processed and digitized to obtain an image representing the contours of the live skin layer (see Figure 6) [9].
- *Piezoelectric*: These sensors make use of the piezoelectric effect. The sensor surface is made of a nonconducting dielectric material that generates small amounts of current when pressed. The amount of current depends on the pressure applied. When pressing a fingertip on the sensor,

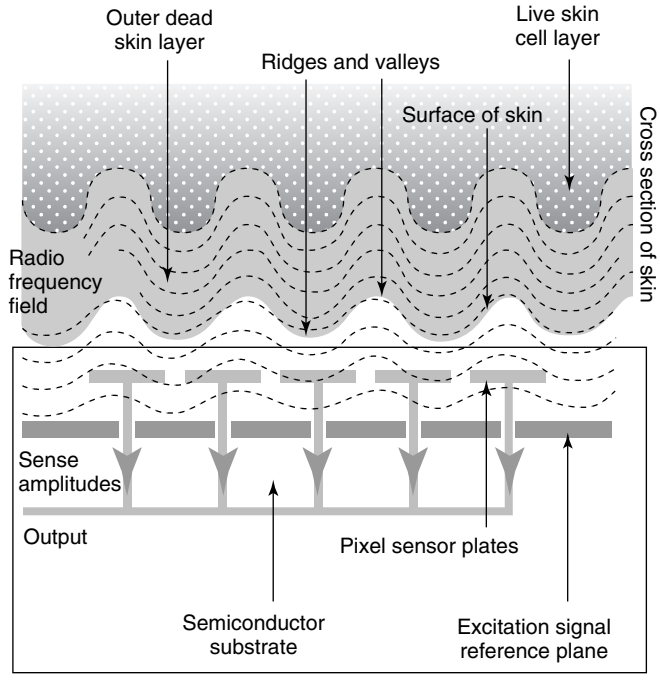


Figure 6 Electric-field fingerprint sensor [Reproduced with permission from Authentec Inc.]

the ridges will apply a higher pressure than the valleys as they are closer to the sensor surface. Typically, the sensor material uses some kind of threshold to determine whether or not a sensor element is "pressed", thus only enabling acquisition of binary images [1].

- *Thermal (sweep sensor)*: These sensors are made of pyroelectric material that generates current based on temperature differentials. Sweeping a finger across an electrically heated sensor allows measurement of heat flow to the skin, which is higher at direct contact with the sensor, thus allowing distinction between finger ridges and valleys [1, 6].

Other Sensor Types.

- *Ultrasonic*: This type of sensing is based on sending acoustic signals toward a fingertip and capturing the echo signal. As each change of impedance gives a partial echo, this technology can be used to image the subsurface of the skin [1] (see Figure 7).

Iris Recognition

Iris recognition technology uses a near-infrared light source and optical camera to capture an image of the iris [10]. Using pattern recognition algorithms, the image is converted to a template (also known as *IrisCode* [11]), which allows comparison with other enrolled templates (see Figure 8).

Hand Geometry Recognition

Hand geometry recognition is currently used mainly for physical access control applications. Typically, devices using this feature map the size and shape of a person's hand, compute a template from it, and verify this with a previously enrolled template.

A source of light illuminates two perpendicular mirrors and the reflected light is captured by the device. When a hand is inserted at the proper location (using guidance pegs), it partially blocks the light reflecting from the mirrors, thus creating a silhouette image on the sensor of both the top view and the side view of the hand (see Figure 9) [12].

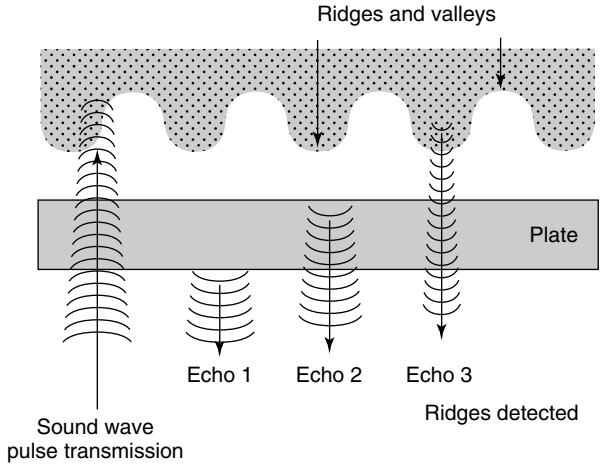


Figure 7 The principle of ultrasonic sensing of a fingerprint [Reproduced from Ref. 1. © Springer-Verlag, 2003.]

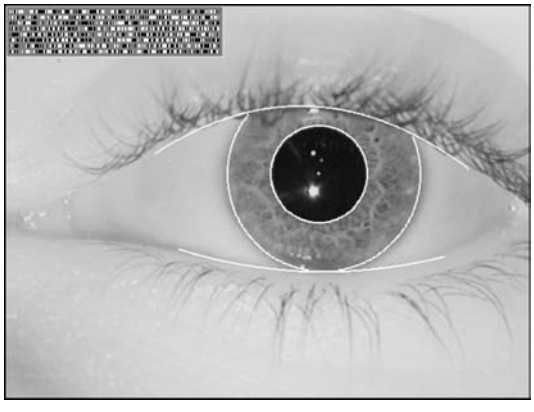


Figure 8 Iris pattern and IrisCode, J.G. Daugman, 2004 [Reproduced from Ref. 11. © IEEE, 2004.]

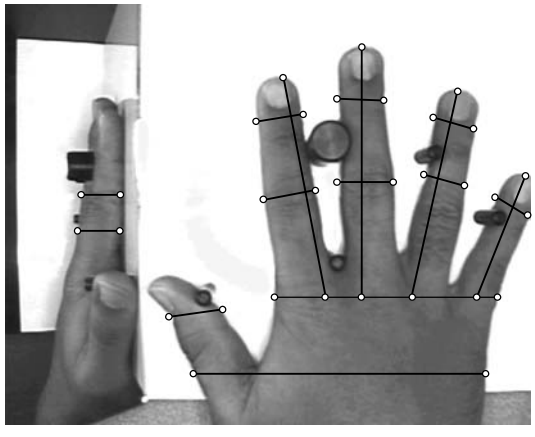


Figure 9 Hand side views used for hand geometry recognition [Reproduced from Ref. 12. © Michigan State University.]

Hand Vein Pattern Recognition

This technology is based on comparison of vascular pattern images of the back or palm of a hand. The devices use a near-infrared light source to shine on the hand and a near-infrared optical sensor system to capture the reflected light. Since near-infrared absorption and scattering properties of tissue and blood differ [13, 14], an image of the vein pattern is effectively captured. Image-processing algorithms are used to facilitate template creation and comparison (see Figure 10) [15].

Manufacturers claim that no nonbiometric patterns can be enrolled, as extensive checking is done of whether or not an actual, living hand is presented, for instance, by sensing the temperature pattern [16] or an active flow of hemoglobin through the person's veins [15]. In general, such checking of "aliveness" is referred to as *liveness detection*.

Reliability of Biometric Systems

There are a number of factors influencing the reliability of a biometric system. A proper evaluation of a biometric system should include the following tests [7]:

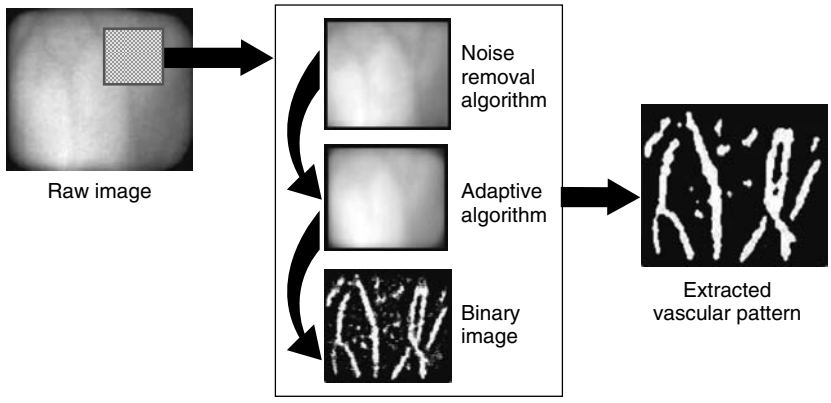


Figure 10 Hand vascular pattern extraction process flow

- *Environmental testing*: Biometric systems are affected by their physical environment. For instance, a device for recording facial images or iris patterns relies on ambient lighting conditions. Other influences include ambient sound levels, temperature, ambient electromagnetic noise, atmospheric humidity, dust, voltage supply variations, shock, and vibrations. The effects of such environmental influences should be investigated.
- *Statistical performance testing*: The performance of a biometric system is often described using particular error rates that can be evaluated by performing real-life automated biometric comparisons for which the ground truth is known. The larger the number of comparisons, the more accurate the calculated rates will be. Frequently reported rates include the following:
  - *False accept rate* (*FAR*): The frequency at which the system concludes that a claimed user identity is true, while in reality it is not.
  - *False reject rate* (*FRR*): The frequency at which the system concludes that a claimed user identity is not true, while in reality it is.
- *Vulnerability testing*: A biometric system has several aspects that introduce potential vulnerabilities. Data streams may, for instance, be intercepted and modified or a fingerprint sensor may be deceived by an artificial finger. Such vulnerabilities should be analyzed.

Environmental and statistical performance testing is beyond the scope of this article. Independent statistical performance evaluation reports are available on many specific biometric devices and comparison algorithms, for instance, through the IBG or the National Institute of Standards and Technology (NIST). The following paragraph deals with the locations in biometric systems that are potentially vulnerable to attack.

Biometric System Vulnerabilities

Biometric System Threat Locations

The Common Criteria Biometric Evaluation Methodology Working Group (CCBEMWG) [7] has put together an extensive list of potential threats to a general biometric system. An example of such a threat is that an authorized user's biometric, such as a fingerprint, could be copied, unknowingly, willingly, or unwillingly, and made into a fake fingerprint and presented to the biometric device sensor by an impostor. Or, a virus could be programmed to intercept or manipulate digital data being sent from any one component to another. Databases could be hacked and templates could be inserted or deleted. System administrators may make errors in setting up matching thresholds, or even be hostile and manipulate user records and privileges on purpose. Hardware components may be bypassed, deactivated, or otherwise tampered with. Many of the threats can be partly or completely eliminated by proper system design. Nevertheless, all threat locations should ideally be analyzed. This article deals mainly with the threat of fake biometric samples.

Biometric Device Spoofing

Most biometric devices can be fooled, also known as *spoofing*, using relatively simple means. This makes that these devices still require human supervision when used in high security applications to ensure that this weakness is not exploited. Furthermore, it diminishes the forensic reliability of user traces in unsupervised biometric access control systems, as they could originate from fake biometric samples. The following paragraphs show the methods that have been tested.

Fingerprint Spoofing

The following types of fake fingerprints were tested:

1. gelatin fingerprint from silicone mold;
2. super-soft plastic fingerprint from silicone mold;
3. wood glue fingerprint from digital image of fingerprint printed on sheet;
4. rubber fingerprint stamp from digital image of fingerprint.

The first two methods require cooperation of an authorized user, as an actual finger is needed to make an impression in silicone. The last two methods require a digital image of a fingerprint. For the spoofs made here, a rolled ink fingerprint was used as source, which was then digitized using a flatbed scanner. Alternatively, one could use conventional forensic methods to get a digital photograph of a residual fingerprint that an authorized user left on some object. This can potentially be used to steal a person's fingerprint and enter a biometric system without cooperation of the authorized user. The making of the molds and the artificial fingerprints is described in the following section.

Making the Silicone Mold. We used a two-component low-viscosity silicone paste for making a mold (see Figure 11). This material captures the details of a fingerprint well, and can be used multiple times for making an artificial finger without getting damaged. The procedure entails mixing the two components in the right proportions quickly but thoroughly, and then pressing the finger into the mixture the same way one would press a fingerprint sensor. After holding still for a few minutes until the paste has dried and hardened, careful removal of the finger leaves a good quality mold.

Making a Usable Digital Image of a Fingerprint. The third and fourth spoofing methods require the digital image of the fingerprint to be binary. Digital photographs of a fingerprint residue made visible by standard forensic methods and scans of a rolled fingerprint are usually in color or grayscale format; hence, some image processing must be done. For the

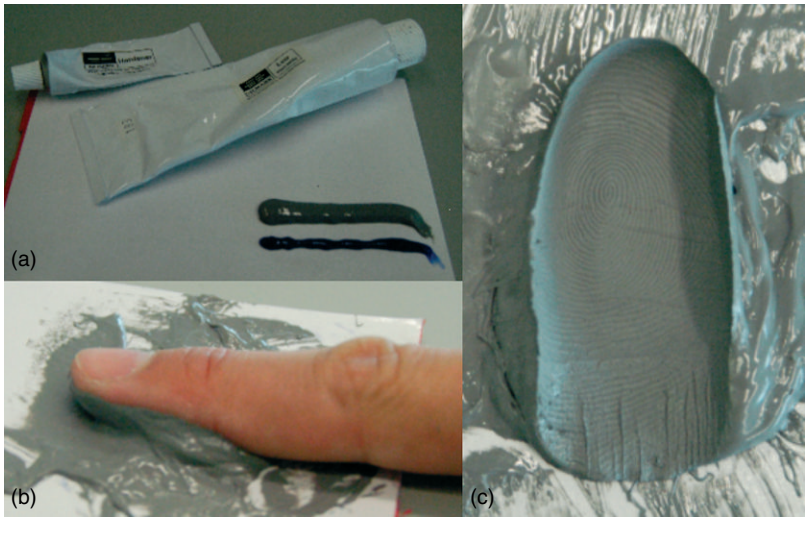


Figure 11 Making a silicone mold: (a) paste and hardener; (b) pressing finger in the mixture; and (c) the resulting mold

scanned image of the rolled fingerprint, we used an image-editing package to change the brightness and contrast with −60 and +100, respectively, effectively binarizing the image (see Figure 12). Depending on whether the image is to be used as input for a mold or a stamp, it may have to be inverted and/or mirrored.

Making the Gelatin Fingerprint. Making the so-called gummy fingers has been done extensively before [17–19]. The method involves dissolving "kitchen-quality" gelatin in near-boiling hot water and pouring it into a mold, such as the previously described silicone mold. When cooled down, the result is a flexible, transparent, yellowish cast of the finger. The main reason for using gelatin solution is that its electrical conductivity and moisture level resemble that of a human finger [17]. Sensor technologies based on such finger properties are therefore usually also responsive to this material.

A solution of approximately 55% water and 45% gelatin was used. Dissolving such relatively high amounts of gelatin in water without getting bubbles requires some practice. Gentle stirring and repeatedly cooling down and heating up the solution (while keeping below boiling point) may be necessary to allow bubbles to escape the mixture.

This spoof produced an acceptable result with all the tested scanners. It gave the best overall image quality of all the spoofs, based on visual inspection. Gelatin is cheap and available in almost every supermarket. However, durability is not very good. When kept in the open, it quickly dries and shrinks. When kept in a plastic bag it lasts longer but will become moldy. Keeping it in a refrigerator extends its lifetime. Also, it takes some practice to prepare the mixture properly. In our tests, user cooperation was needed for the mold. Note that there are other methods that can use the digital image of a fingerprint as source [19], thus only requiring a latent fingerprint.

Making the Super-Soft Plastic Fingerprint. Super-soft plastic (see Figure 13) is a nontoxic polymer that comes as a whitish opaque liquid. After heating, it turns solid when cooled down to room temperature. The result is a soft, flexible, transparent, and colorless plastic. It is typically used for creating fish lures, and may be found in some fishing shops.

A small amount of super-soft plastic is heated in a microwave just until the liquid turns completely colorless and transparent. One must keep a constant eye on the substance while heating, as it can easily overheat and consequently burn, turning yellowish and lumpy, diminishing its usability. The heated

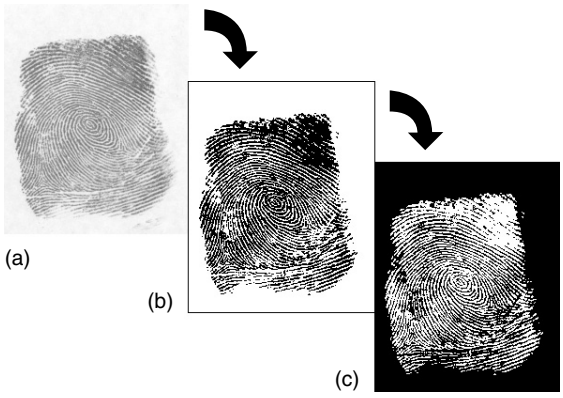


Figure 12 Processing a rolled ink fingerprint: (a) scanning; (b) binarizing; and (c) inverting

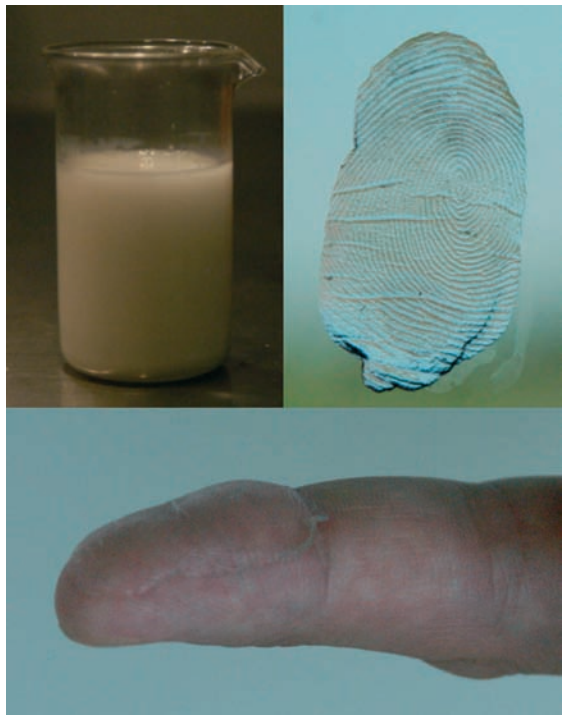


Figure 13 Super-soft plastic spoof

plastic is then poured into the silicon mold and left for some time to cool down and turn solid. For making a thin spoof, just pour in a small quantity and tilt the mold a bit to spread the liquid out over the mold surface.

This spoof was shown to be able to deceive the optical FTIR and SEIR scanners, as well as the ultrasound scanner. It can be used relatively inconspicuously when a thin specimen is stuck on a fingertip. However, because of the transparency of the material, the optical scanners sometimes show a faint impression of the underlying structure, which gets more obvious with a thinner spoof. Furthermore, it has good durability and it is very easy and cheap to produce. One does need a real finger to be able to make a proper mold.

Making the Wood Glue Fingerprint. This technique is based on the fact that when printing with a laser printer, the toner forms a relief that is deep enough to be useful for making a fingerprint mold. Water-proof wood glue is an appropriate material for the fake fingerprint, as it can be smeared out in a thin layer, dries up transparently, and returns to its dried-up shape well after bending or stretching.

This spoof can be made as follows [20] (see Figure 14):

1. Print the binarized, inverted image of the fingerprint on a plastic sheet using a laser printer.
2. Smear out a thin layer of ordinary wood glue across the image.
3. Let it dry until the glue becomes transparent.
4. Carefully peel off the layer of glue.
5. Cut the glue sheet to a size fitting your fingertip and stick it on. If it does not stick well enough by itself, use theatrical glue.

This spoof was able to deceive the FTIR and SEIR optical scanners, as well as the ultrasound scanner. The basic materials are cheap and readily available and it is quite easy to prepare this spoof. It is also the thinnest, most inconspicuous of the spoofs.

Making the Rubber Fingerprint Stamp. Professional stamp-making machines used by stamp factories need only a digital, binary image as input for making a stamp. A laser with a fixed intensity burns away the appropriate parts by passing the surface of a rubber sheet to leave a relief of the input image, up to a precision of 1000 dots per inch (DPI) (see Figure 15). The burn depth is determined by the movement speed of the laser. This speed is constant for a single stamp; thus, there can be no variation in the depth up to which the rubber is burnt away. A 25-W Trotec laser engraving machine was used to make our stamps. One can use natural as well as synthetic vulcanized rubber, the latter being more resistant to chemicals.

For a stamp to be used directly on a sensor, a binary image of a fingerprint will do. If a mold stamp

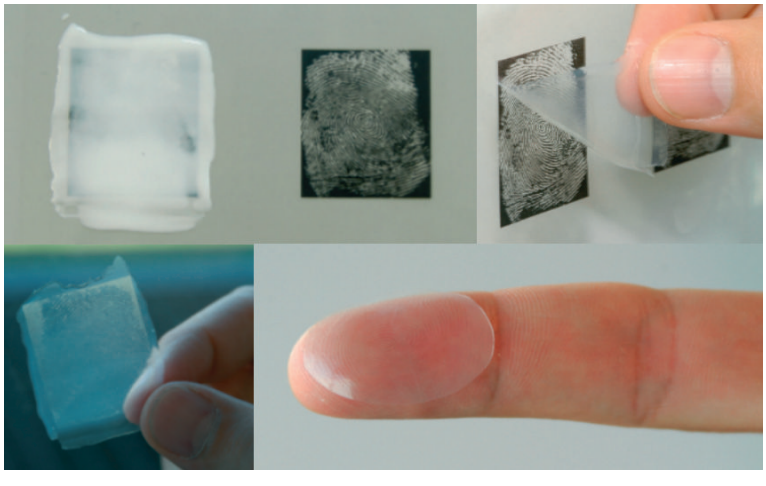


Figure 14 Making the wood glue fingerprint spoof [Reproduced with permission from Ref. 20. © Chaos Computer Club, 2004.]



Figure 15 A set of fingerprint stamp spoofs made of natural vulcanized rubber

is to be made, the image must also be inverted and mirrored. We found, however, that the method is not particularly fit for making molds, as only the surface of the protruding edges of the stamp is as smooth as the original material was, but the burnt-out parts are somewhat irregular.

The spoof was able to deceive all scanners except the ones that use electrical properties of the skin. One can, however, easily discern the spoof by visual inspection of the digital images. Durability is very good. They can be used over and over again. The fingerprint image that is needed can be made with some knowledge of basic image-processing software. Many stamp makers allow online ordering of stamps, where stamp images can be uploaded, making access to the technology quite easy. In addition, making stamps is quite cheap.

Spoof Capture Examples. Figure 16 shows examples of the images acquired when presenting each of the spoof types to an SEIR fingerprint scanner. Figure 17 shows images from eight different fingerprint scanners for the most successful artificial fingerprint: the gelatin spoof. The gelatin spoof passed verification with all software that we tested.

Iris Spoofing

For this test, a low-end desktop iris scanner was used with biometric software, which enables logging on to a computer using iris verification. A live iris was enrolled into the database and fake irises were presented to the camera. Our tests confirm previous successful tests [20] in which a basic iris scanner was fooled using a paper photocopy of an iris. We, however, used much lower resolution images than were thus far considered necessary. Note that high-end iris scanner such as those used at airports have several antispoofing measures incorporated, preventing the use of paper irises.

Paper Iris Method. All that is needed is a black-and-white low-resolution image of the authorized user's iris printed on paper (see Figure 18). As the scanner does check for the presence of a retina reflection, the pupil must be cut out to allow the user to look through the hole when presenting the paper iris. When taking a photograph of the iris, care must be taken not to "pollute" the iris region with reflections. One can photograph through a tube to get a good result, and when the photo is taken from the front, any flash reflection will be inside the pupil region.

In the tests, irises from both a 6 megapixel portrait photograph and a 6 megapixel close-up of the eye were used, which after rescaling to actual size resulted in a 240 and 500 pixels per inch (PPI)



Figure 16 Comparison of images from an optical SEIR scanner using a live finger and fake fingerprints

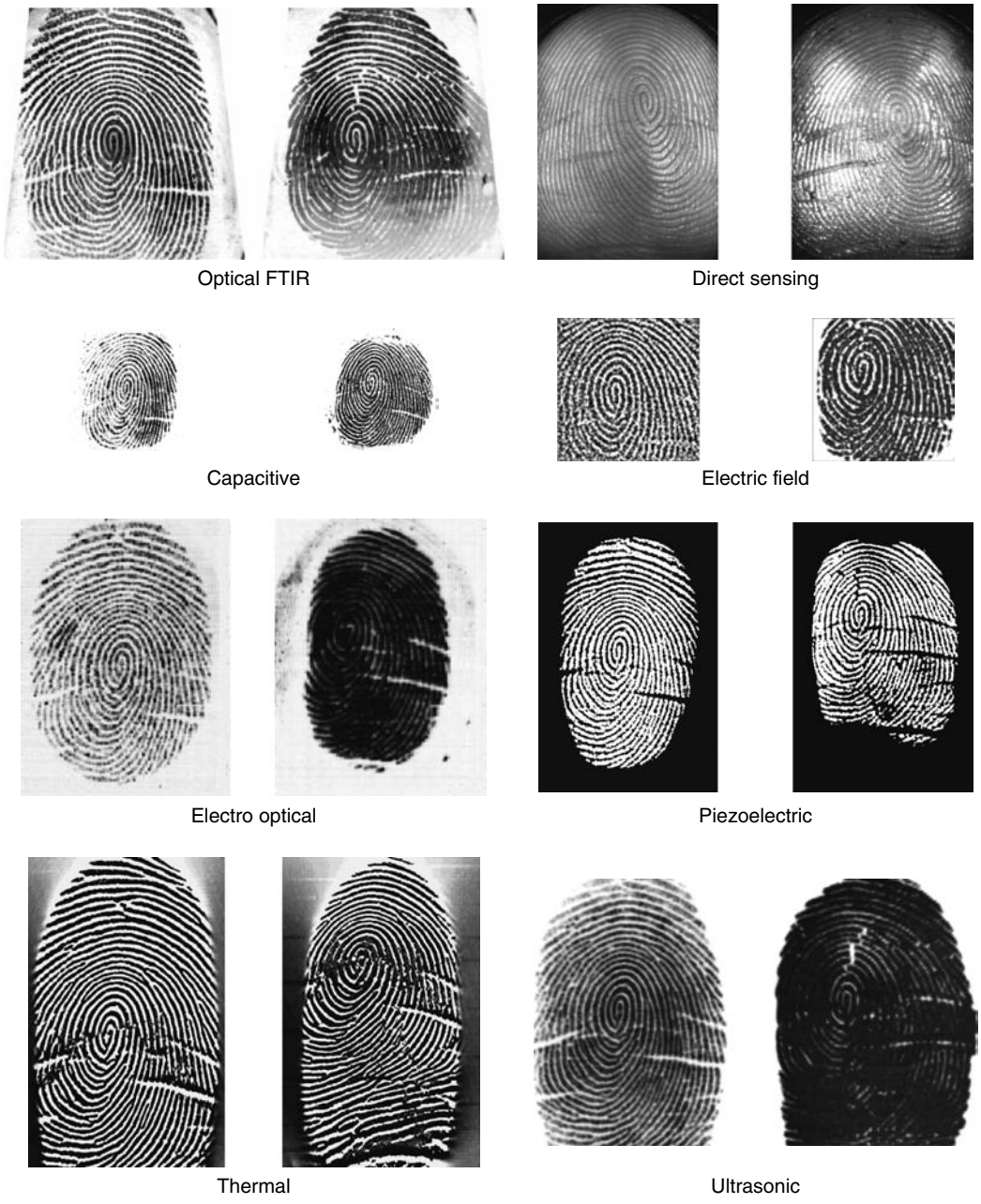


Figure 17 Images acquired using eight different fingerprint scanner technologies. The left of each pair shows a natural finger, the right shows a gelatin spoof

image, respectively. The images where printed using 600 DPI laser printer. After some practice, it was quite easy to pass the system using the fake irises, especially with the 500 PPI version. So far, we have only been able to spoof the device using copies of light eyes.

To test the usability of old photographs, a similar test was also done using an analog photograph of

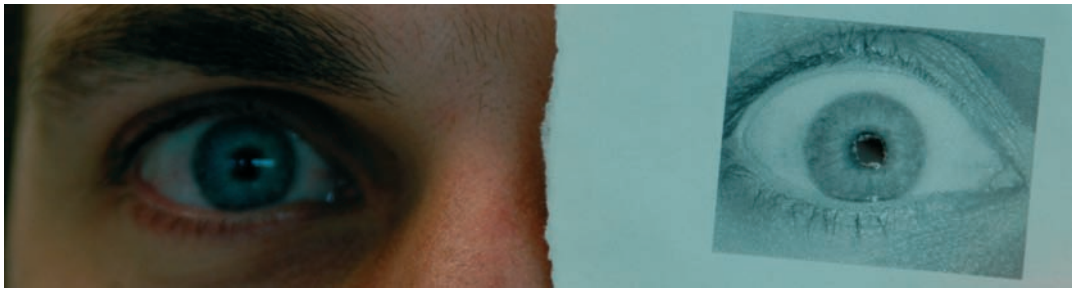


Figure 18 Original live iris and the paper spoof

the authorized user from his childhood (from over 20 years ago). The photo was scanned to digital format and slightly brightened to make the details stand out more. The paper spoof created with this image was also accepted for verification.

Hand Geometry Spoofing

Our tests have shown that a basic hand geometry recognition device can, in principle, be deceived using something other than an actual hand, when security level is set somewhat below the factory setting. Furthermore, an unauthorized user with approximately the same hand size as the authorized user scores reasonably well, confirming the well-known fact that hand geometry uniqueness can be an issue [21].

Paper Hand Method. The image extraction method used in hand geometry recognition suggests that it need not be presented with an actual 3D hand for getting a valid result, as only the side views are captured. Furthermore, judging from a hand geometry recognition patent [22] (see Figure 19), the influence of the side view is not expected to be so big. The figure shows that the side view of the hand accounts for a relatively small part of the information captured by hand geometry recognition devices.

Tests show that a top view silhouette of a hand cut out in paper can indeed be enough to pass the verification check of the device, albeit at a lowered security level. Thus, this is not a representation of practical use of the device.

The paper hand can be made as follows (see Figure 20):

1. Measure the sizes and relative positions of the guidance pegs.
2. Precisely draw the pegs in the proper configuration using drawing software and print it on a transparent sheet (alternatively, print it on paper, and then copy it on sheet).
3. Make a photocopy of the authorized user's hand with a photocopier. Use the sheet with the peg configuration to properly position the hand and fingers on the copier.



Figure 19 Typical image captured by a hand geometry recognition device

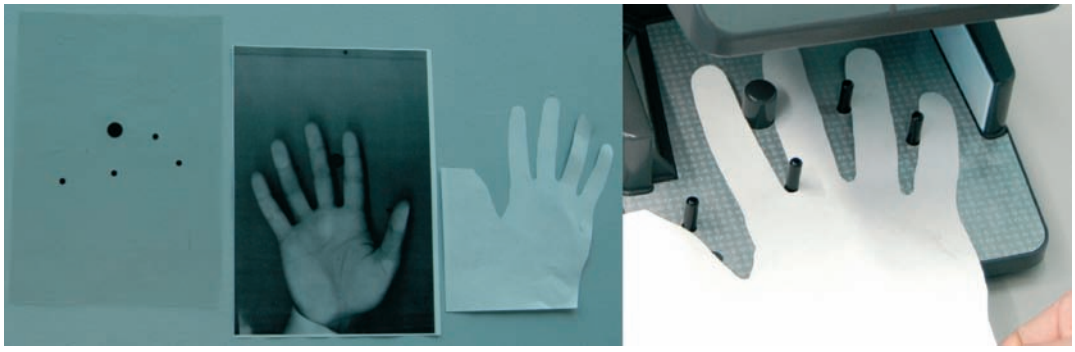


Figure 20 Hand geometry verification device spoofing at lowered security settings

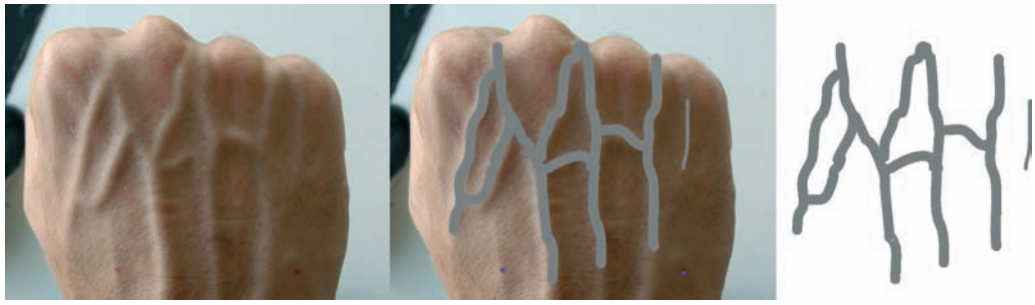


Figure 21 Obtaining a vascular pattern using a digital camera: (a) image of the back of a hand shot in daylight using a common digital photo camera; (b) visible veins manually traced using image-processing software; and (c) the drawn layer ready for printing

4. Cut out the paper hand.

It is expected that the quality of the spoof can still be improved by cutting a more accurate representation of the hand top view. In addition, adding the side view silhouette to the paper hand is expected to improve matching scores. Currently, however, an actual authorized hand gives better scores than the paper hand spoof.

Hand Vascular Pattern Spoofing

For hand vascular pattern spoofing, deceiving the biometric pattern capture and recognition part does not suffice. The liveness detection also has to be passed. With the device we tested, we were not successful in spoofing both at the same time. However, the liveness detection can be turned off by the system administrator. We found that with liveness detection turned totally off, a paper printout of the vascular pattern of an authorized user can be enough to pass the verification. Also, with liveness detection turned totally off, a paper printout of a person's vascular pattern can be enrolled. Then, when turning the liveness detection back on, the actual person can log into the system using his own hand. Hence, we can conclude that the biometric pattern recognition part can be spoofed. The method for making the partial spoof is elaborated below.

Vascular Pattern Printout Method (Partial Spoof). Basically, one must create a gray-and-white copy of the pattern of the veins on the back of the hand. Some people have clearly visible veins that lie close to the surface of the skin, in which case a normal digital photograph of the top view of the hand in enrollment attitude will do as a starting point (see Figure 21a). In that case, the following procedure can be followed:

1. Draw reference dots on the back of the hand at known distances from each other. These can be used later to scale the photograph to the right size.
2. Take a photograph of the top view of the back of the hand (Figure 21a). The hand should have the same attitude as it would have when enrolling.
3. Use image-processing software that can work with layers to manually trace the vein pattern (Figure 21b).
4. Print just the drawn layer in the actual size (scale the picture using reference dots). Color contrast of lines and background should not be too great. Gray value 128 on a white background works fine (Figure 21c).

If the veins are not sufficiently discernable in visible light, a camera with "nightshot" function can be used. Such cameras use a near-infrared lamp to illuminate the target and a sensor array that is sensitive to near-infrared light to capture the reflecting rays. The quality of the nightshot function varies with camera type. In this experiment, a Sony DCR-TRV9E digital video camera recorder was used with good results. The spoof can be made as follows (see Figure 22):

1. Cover the camera-mounted near-infrared lamp with some sheets of tissue. This blocks a portion of the light, which is necessary to prevent spots of overexposure that make postprocessing more difficult.
2. In a dark room, take a shot of the top view of the back of the hand using the nightshot function of the camera. The hand should have the same attitude as it would have when enrolling. One can include a measuring stick at the same height as the back of the hand to the shot to facilitate scaling later on.
3. Use either the previously discussed tracing method or a sequence of image-processing filters to obtain a two-color representation of the vascular pattern. An example of a filter sequence that was used in this test is as follows:
   (a) Apply a Gaussian blur on the image. Use a blur radius large enough to make the details fade. We used a radius of 25 pixels for a 720 × 576 pixel image.
   (b) Subtract the blurred image from the original. This reduces the more global differences in intensity that may be present due to uneven lighting conditions.
   (c) Binarize the image using a gray value threshold that leaves the veins well visible.
   (d) Apply a Gaussian blur on the image. Use a blur radius large enough to make pixels that make out the veins connect, but small enough to keep the vein lines standing out. We used a radius of 3 pixels for a 720 × 576 pixel image.
   (e) Binarize the image again using a gray value threshold that leaves the veins well visible.
   (f) Lower the contrast by changing black to gray (for example, gray value 128).
4. Scale the image, and thus it represents the actual size and print the resulting image.

The paper pattern can be stuck on someone's hand or anything else, such as a bottle. When presenting the spoof to the vascular pattern scanner, care must be taken to match the pattern position with the position



Figure 22 Processing a vascular pattern from a nightshot camera to a usable spoof: (a) nightshot camera image; (b) half-way processed; and (c) fully processed

of original live hand, as the devices can be very sensitive to shifting of the pattern.

Conclusion

It is evident that the current state of the art of biometric devices leaves much to be desired. A major deficit in the security that the devices offer is the absence of effective "liveness" detection. At this time, the devices tested require human supervision to be sure that no fake biometric is used to pass the system. This, however, negates some of the benefits these technologies potentially offer, such as high-throughput automated access control and remote authentication. On a positive note, biometric device technologies are constantly evolving; thus, current weaknesses may well be overcome in the near future.

The independent testing of biometric devices is still nontrivial as manufacturers tend to oversell the capabilities of their products. The latter can give a false sense of security, adversely affecting actual security if not recognized in time. It is an issue that we encounter in many forms of technology today; if it can be cracked, it will be cracked. Accepting this would need a different attitude of manufacturers, in which more of what is going on inside the device and the accompanying software is made public. It would allow potential users of biometric systems to better judge the fitness of such systems for their particular purposes.

From a forensic point of view, care should be taken when drawing conclusions from information extracted from access control systems that use biometric devices. The possibility that the system was compromised, falsely linking persons to events consequently, should be examined or at least noted in the forensic examination report.

References

[1] Maltoni, D., Maio, D., Jain, A.K. & Prabhaker, S. (2003). *Handbook of Fingerprint Recognition*, Springer-Verlag, New York.

[2] Jain, B., Bolle, R. & Pankanti, S. (1999). *Biometrics – Personal Identification in Networked Society*, Kluwer Academic Publisher.

[3] Bolle, R.M., Connell, J.H., Pankanti, S., Ratha, N.K. & Senior, A.W. (2004). *Guide to Biometrics*, Springer-Verlag, New York.

[4] International Biometrics Group website (2007). http://www.biometricgroup.com/US-VISIT.html (accessed 25 July 2007).

[5] Department of Homeland Security website (2007). http://www.dhs.gov (accessed 26 July 2007).

[6] Mainguet, J.F. (2007). *Page d'accueil du site de Jean-François Mainguet*, available at http://perso.orange.fr/fingerchip (accessed 29 July 2007).

[7] Common Criteria Biometric Evaluation Methodology Working Group (2002). *Common Criteria – Common Methodology for Information Technology Security Evaluation – Biometric Evaluation Methodology Supplement*, Version 1.0.

[8] SecuGen Biometric Solutions (2006). *SEIR™ Optic Technology*, Available at http://www.secugen.com/download/SGWP_SEIR.pdf.

[9] Authentec, Inc. website. (2006). Available at http://www.authentec.com.

[10] Daugman, J. & Downing, C. (2001). Epigenetic randomness, complexity, and singularity of human iris patterns, *Proceedings of the Royal Society of London. Series B. Biological Sciences*, Vol. 268, pp. 1737–1740.

[11] Daugman, J.G. (2004). How iris recognition works, *IEEE Transactions on Circuits and Systems for Video Technology* **14**, (1), 21–30.

[12] Biometric Research – PRIP MSU website (2006). http://biometrics.cse.msu.edu/.

[13] Institute for Biodiagnostics website (2006). *Tissue Absorption*, http://www.ibd.nrc-cnrc.gc.ca/english/spec_e_inVivo_absorption.htm.

[14] UCL Department of Medical Physics & Bioengineering website (2006). http://www.medphys.ucl.ac.uk/research/borl/research/NIR_topics/nirs.htm.

[15] Fujitsu Systems Business, Fujitsu Announces Global Launch of its Contactless Palm Vein Authentication Technology (2005). Available at http://www.fujitsu.com/th/en/news/recent/news_Palm_Vein.html.

[16] Tech-Sphere website (2006). http://www.tech-sphere.com.

[17] Blommé, J. (2003). *Evaluation of Biometric Security Systems Against Artificial Fingers*, Master's thesis LITH-ISY-EX-3514-2003, Department of Electrical Engineering, Linköping University, Linköping.

[18] Thalheim, L., Krissler, J. & Ziegler, P.M. (2002). *Körperkontrolle – Biometrische Zugangssicherungen auf die Probe gestellt. c't 11/2002*, p. 114. Available at http://www.heise.de/ct/english/02/11/114/.

[19] Matsumoto, T. (2004). *Gummy Finger and Paper Iris: An Update*, Available at http://www-kairo.csce.kyushu-u.ac.jp/WISR2004/presentation12.pdf, (accessed Oct 2004).

[20] Chaos Computer Club (2006). *How to Fake Fingerprints*, (accessed 2006). Available at http://www.ccc.de/biometrie/fingerabdruck_kopieren.xml?language = en.

[21] Gasson, M., Meintz M. & Warwick, K. (2005). *FIDIS Deliverable 3.2: A Study on PKI and Biometrics*, (accessed July 2005). Available at http://www.fidis.net.

[22] Recognition Systems, Inc. (1986). *Identification Apparatus*, European Patent EP0209317.

Related Articles

Automated Fingerprint Identification System

Facial Comparison

Friction Ridge Skin: Comparison and Identification

RIKKERT ZOUN

Biometrics: Face Comparison *see* Facial Comparison

Bitemark *see* Odontology

Black Rage Syndrome *see* Syndromes: Psychological

Blood: Detecting *see* Luminol

Blood: Luminol *see* Luminol

Blood: Tests for *see* Luminol

Blood Grouping

Introduction

Before Sir Alec Jeffrey's pioneering discovery of the power of DNA fingerprinting in 1984 [1] and its first use in a criminal case in 1986, human identification, relationship testing and population studies were done through assessment of protein variation seen in blood groups and various proteins found within body fluids.

Enzymes and other proteins exist in different isoforms and are observed as *phenotypes* (*see also* **Phenotype**) through the use of techniques such as electrophoresis. Different phenotypes may result from different genotypes, the genetic instructions that result in the different protein forms, or *alleles*, that are translated, but can also be produced through the influence of environmental factors, various subunit interactions or different posttranslational modifications.

Polymorphism (*see also* **Polymorphism: Genetic**) exists when more than one phenotype exists and is very common. It results from evolutionary forces relating to genetic selection, adaptation, and biodiversity. Polymorphisms are heritable factors and only change if a mutation occurs between parent and offspring. Unlike DNA short tandem repeat polymorphism, the current technique used for human identification, mutant events in gene products between parent and child are rare because a genetic mutation is likely to result in a dysfunctional or nonfunctional product associated with disease implications or with potentially lethal effects.

To be considered a polymorphism, the least common form should be seen in at least 1% of the population to distinguish this from products of random mutation events [2].

Protein polymorphisms are the result of changes in amino acid composition, which can affect both structure and function. Traditionally, these have been investigated using electrophoretic techniques. The charge exhibited by a protein depends on the relative content of basic and acidic amino acid residues and its environmental pH. At a low pH, basic amino acids gain protons (H^+) and become positively charged, and move toward the cathode in an electrophoresis tank. At a high pH, this is reversed: the acidic amino acids lose protons, acquire negative charges, and travel toward the anode.

The point at which net charge is zero is the *isoelectric point*, and is, therefore, the point at which electrophoretic mobility is zero. *Isoelectric focusing* (*IEF*) [3], like simple electrophoresis, is a technique for separating different molecules according to their electrical charge, but provides greater resolution by taking advantage of the fact that the electric charge of a protein varies according to the pH of its environment. For example, standard electrophoresis identifies only three variants of the polymorphic phosphoglucomutase (PGM) protein system, whereas IEF distinguishes 10 variants. A pH gradient can be produced in a medium with the use of *ampholytes,* molecules that have both basic and acidic groups. A variety of *polyampholytes* with different isoelectric points are introduced to a starch, agarose, or polyacrylamide gel, and an electric current applied, producing an immobilized medium, graded according to its pH. Proteins to be analyzed are then introduced into the process and are separated according to their isoelectric point. Use of this technique produces gel bands that are much sharper than other gels.

While *protein polymorphisms* are those nonmembrane bound proteins found in blood or body fluids, or intracellularly, *blood group polymorphisms* are restricted to blood cell surface *antigens*, although the ABO blood group is an exception to this, as the A, B, and H antigens of this system are found in various body secretions in some individuals (*secretors*). Blood groups are inherited characters of the red cell surface (antigens) that are detected by specific *alloantibodies* and the defining interactions are observed when an antigen–antibody reaction takes place in a group of red cells and they *agglutinate*.

Typing of samples by immunogenic testing or electrophoresis is not only very time consuming but also requires significant expertise, so that the processes are properly executed and the correct interpretation made. Problems can occur if rare variants are not recognized because of their unusual presentation, or if gel resolution is poor. All of these factors can result in typing errors. Although this form of typing can be speeded up with the use of microtiter plates, for example, this is not sufficient with the speed of throughput required in the expansion of forensic identification. In electrophoresis, although several samples can be introduced on to a single gel, in IEF only one system can be typed per gel because of the different pH gradients required, and in other systems specialized development techniques are often required to detect the particular proteins concerned. It is for these reasons that these systems have gradually been phased out in favor of DNA analysis.

DNA (*see also* **DNA**) is the genetic code that provides the instructions for proteins to be produced, according to the particular sequence of nucleotides along a DNA strand. In its natural state, DNA is a double-stranded molecule with two antiparallel polynucleotide strands, assembled into chromosomes. The particular nuclear DNA inherited by an individual is contributed to equally by both parents and results in individuals having two alleles for every protein sequence. Individuals who inherit the same allele (gene sequence at a particular locus), from both parents are said to be *homozygous* for that allele, whereas individuals who inherit different alleles are *heterozygous* at this locus.

Nucleotides are the basic building block of the DNA molecule and consist of a base, a deoxyribose sugar, and a phosphate group. Nucleotides differ only according to the base they carry. There are four different bases: the purines, A (adenine) and G (guanine), and the pyrimidines, C (cytosine) and T (thymine); the double strand of the DNA is stabilized with hydrogen bonds between complementary bases, A pairs with T, and G with C.

In a sequence of nucleotides, a group of three nucleotides (*a codon*) codes for a particular amino acid, which, on translation, determines the primary structure of a molecule. Proteins are the products of genes, or combination of genes, that are present within *exons* on DNA strands, being separated from each other with noncoding portions (*introns*). The sequence details are first converted into messenger ribonucleic acid (mRNA) and intron sequences removed to leave the exons in series. Proteins are synthesized through translation of the amino acid sequences which determine primary, secondary and tertiary structure of the protein. Protein polymorphisms can result from genetic changes and, in principle, classical markers can be predicted through examination of the genetic material.

Most protein polymorphisms are the results of single amino acid substitutions, often through missense mutations (*single nucleotide polymorphism* or *SNP*), in which one nucleotide is altered in a codon, leading to a different amino acid being produced. For such changes to be detectable using electrophoresis, this change has to alter the charge or isoelectric point of

the protein. Some mutations do not result in changes in amino acids as some codons code for the same nucleotide.

Polymorphisms can also be the results of insertions or deletions that produce *frame shifts*. These also disturb the codon sequences and hence lead to changes in amino acid sequences. Sometimes, recombination events may be unbalanced and can produce subtle protein changes that are not detected through examination of the genetic structure.

While DNA profiling has, understandably, replaced the use of classical markers in forensic identification, the ability to examine many of these polymorphisms with modern genetic methods means that these systems can be revisited and have already been utilized in "cold" case reviews. Molecular analysis further offers the opportunity to predict classical polymorphisms in small and degraded samples, and the technique is not restricted to a particular cellular source.

Blood Group Polymorphisms

While any variation detected in blood could be considered to be a blood group, the term is normally restricted to red blood cell surface antigens. These antigens are detected by *alloantibodies* (which are naturally occurring), or as a result of *alloimmunization* with red cells (through blood transfusion or from a fetus during pregnancy).

The entity of a blood group was first described by Landsteiner [4] in 1900 when he reported how the plasma of some individuals agglutinated the red cells of others. Not all antibodies lead to direct agglutination and it was only after 1945, when Coombs and others [5] developed the antiglobulin test, that many new antigens that describe a large number of different blood groups have been discovered. Only a few of these have traditionally been used by the forensic scientist.

Blood group antigen determinants are the result of two main types: proteins or carbohydrates. They have a number of fundamental differences: while most blood group antigens are synthesized in the red cell, some antigens result from adsorption on to the cell from plasma, and while most antigens are only detected on red cells, others are found throughout in secretions.

ABO Blood Group System

Pioneering research in the first quarter of the twentieth century led to a basic understanding of this blood group system in which it was shown that only three alleles at one locus could describe the four different blood groups, A, B, AB, and O (although many subtypes have since been defined).

The A and B red cell antigens are modifications of the H antigen. The H antigen is defined by a specific chain of carbohydrates produced by the O gene and individuals with this antigen form blood group O. Inheritance of the A gene or B gene, which codes for different glycosyl transferases (acetylgalactosaminyl-tranferase and galactosyl-transferase, respectively), enables the movement of particular terminal immunodominant sugars to the H antigen, leading to blood groups A and B, respectively [6]. In comparison with the other blood groups the H gene (fucosyltransferase1 FUT1) has a nucleotide deletion, which results in a truncated protein that has no active transferase. Some rare individuals lack the H gene (*Bombay blood* group) and therefore have no H antigen on their red cells, neither do they produce A or B antigens, even if they have the appropriate glycosyl transferase [7]. Table 1 defines the main characteristics of the ABO blood group system.

The H antigen is also present in body secretions of about 80% of Europeans (*secretor*) [8, 9], determined by the FUT2 gene [10]. Secretors of H substance also secrete A or B if they are of the appropriate group. Nonsecretors do not secrete H, A, or B. The ability to secrete H substance is also related to the phenotypes observed within another blood group system, Lewis [11].

ABO blood group typing of liquid blood may be undertaken by mixing red cells of an individual with appropriate antisera (*forward group*), or by mixing the plasma from an individual with

Table 1 Characteristics of the ABO blood group system

Blood group (phenotype)	Genotype	Red cell antigen	Serum antibody
A	AA or AO	A	Anti-B
B	BB or BO	B	Anti-A
AB	AB	A and B	None
O	OO	Absent	Anti-A,B

Table 2 ABO forward and reverse grouping
Forward grouping uses red cells from a person and antisera

Cells from blood group	Anti-A	Anti-B	Anti-AB
A	+	−	+
B	−	+	+
AB	+	+	+
O	−	−	−

Reverse grouping uses serum from a person and indicator red cells

Serum from blood group	A cells	B cells	AB cells	O cells
A	−	+	+	−
B	+	−	+	−
AB	−	−	−	−
O	+	+	+	−

appropriate red cells (*reverse group*). For example, cells from a blood group A individual are agglutinated by serum from a group B person, and a group O person, as both of them have anti-A in their serum. Serum from a blood group A individual has anti-B in his plasma and therefore, agglutinates the red cells of someone with blood group B, or AB. Table 2 illustrates the reactions.

In forensic analysis, blood stains may often be found dry on a substrate. In this case, an *absorption-elution* technique is used [12]. In this technique, portions of the bloodstained material are exposed to the different antisera. In the presence of the corresponding antigen and antibody, a complex forms within the stained material, which is then washed to remove any unbound antibody. Raising the temperature of the liquid in which the stained material is placed to about 50 °C–60 °C, dissociates antigen and antibody. If a complex had been formed in the initial part of the test, then antibody is now eluted into the surrounding fluid and addition of the appropriate red cells result in agglutination, visible in the fluid.

For example, if the dried blood comes from a group A individual, then mixing the stain with anti-A forms an antigen-antibody complex. After removal of excess antisera, the bound anti-A can be eluted and agglutinate A cells, but not B cells. In contrast, mixing the stain with anti-B does not result in an antigen–antibody complex, and any antisera present is washed away. Because no antibody has been bound, there is nothing to be eluted and no agglutination occurs, whatever the red cells added. Although the technique is very sensitive, antisera that work with fresh stains may not be sufficiently strong to detect antigens in older stains and adjustments may be made in the antiserum dilution used. Factors affecting the elution of antibodies from bloodstains are discussed in a publication by Lincoln and Dodd [13].

Determination of the ABO blood group can also be made through the typing of body fluids amongst individuals who are *secretors* of ABH substance. *Lewis* antigens are also tissue antigens, carbohydrates, which are adsorbed onto red cells. They require a FUT for the expression of the Le antigen, coded for by the FUT3 gene [14]. There are two alleles, Le^a and Le^b, expressed by individuals who have the Le gene. In 1948, it was noted by Grubb [15] that most people with the Le^a antigen, having Le (a+) red cells detected by anti-Le^a, did not secrete ABH substance (*nonsecretors*). A general rule has developed: Le (a + b−) cells come from nonsecretors; Le (a − b+) come from secretors, and Le (a − b−) cells come from either.

Grubb later proposed a theory that explained the different observations [16]. It was stated that the presence of Le^a was controlled by a single locus, Le (LE or FUT3 gene), while Le^b appeared to be an interaction between Lewis and the secretor gene, present on the same chromosome. People with the Lewis (Le) and secretor (Se) genes have both Le^a and Le^b in their secretions, but have Le (a − b+)

Table 3 Interaction between Lewis and secretor genes and resultant phenotypes

Genotype		Red cell	Body fluid		
Lewis	Secretor	Lewis phenotype	Secretor of H	Secretor of Le^a	Secretor of Le^b
LeLe or Lele	SeSe or Sese	Le (a − b+)	Yes	Yes	Yes
	Se^wSe^w or Se^wse	Le (a + b+)	Weak	Yes	Yes
	sese	Le (a + b−)	No	Yes	No
Lele	Any	Le (a − b−)	Yes or no	No	No

red cells. Individuals who have the Lewis (Le) gene, but are homozygous for se, the nonsecretor allele, have only Le^a in their saliva and Le (a + b−) red cells. Some secretor genes are weak (Se^w) allowing the expression of Le^a and Le^b, with Le (a + b+) red cells. Table 3 summarizes these variant phenotypes in a person whose blood group is O (expressing the H antigen only).

Although these substances can be found in a wide range of body fluids, forensic analysis is normally concerned with saliva or semen. Their detection requires an indirect technique because there are no cells to agglutinate. The detection of Le^a and Le^b is detected using an *absorption-elution* technique, as described earlier. This method can also be used to detect ABH in secretions but, because of the very high levels of these substances in secretors, greater care must be taken, and the methodology requires the use of a set of dilutions [17] and both *absorption-inhibition* and absorption-elution techniques.

The absorption-inhibition technique is less sensitive than the elution method, but is suitable for detection of ABH, particularly, because its concentration is relatively high and its implementation in a variety of fluids with different concentrations of soluble antigens is improved with use of several dilutions [18]. Antiserum is added to the fluid and antigen–antibody complexes are allowed to form. If antigen is present, then most of the antibody is bound. Subsequent addition of red cells containing the relevant antigen leads to agglutination if there is plenty of free antibody (and no antigen), or agglutination is absent, or weak, if antigen were originally present in the fluid. It must be remembered, however, that in dealing with forensic material that might be susceptible to bacterial contamination, spurious results may occur. *Escherichia coli*, a common contaminating bacterium, for example, produces B-like, and sometimes A-like substances that can lead to discrepant typing [19, 20].

For ABO grouping, antisera are needed against A, B, and H antigens. Anti-A and anti-B are available from the serum of people of blood groups B and A, respectively. Anti-H can be found as a lectin (seed extract) prepared from *Ulex europaeus*, which contains a phytohemagglutinin specific for the H antigen [21]. Table 4 illustrates the type of results that are obtained using these techniques in saliva.

Pre-DNA forensic cases often have serological typing undertaken and confirmation of these in material from cold cases can require a return to older techniques. This is not a problem for blood group polymorphisms because antisera and indicator cells remain readily available because of their use in blood

Table 4 Grouping of saliva using absorption-inhibition and absorption-elution methods

Group	Method	Anti-A then A cells	Anti-B then B cells	Anti-H then O cells
A secretor	Inhibition	−	+	−
	Elution	+	−	−
B secretor	Inhibition	+	−	−
	Elution	−	+	−
O secretor	Inhibition	+	+	−
	Elution	−	−	+
B nonsecretor	Inhibition	+	+	+
	Elution	−	−	−

transfusion laboratories. Nevertheless, it may be useful to use genetic methods that are more sensitive: detection of SNPs found at the ABO blood group locus has been found to provide reliable typing as low as 200 pg [22].

Cloning of the ABO gene in 1990 [23] has led to the development of various molecular techniques to predict ABO blood group phenotype. The coding region of the gene has seven exons over 18 kb with exons 6 and 7 containing most of the coding sequence. SNPs in exon 7 distinguish between groups A and B, whereas O is defined through SNPs in exon 6 and 7 [24]. While many different alleles have been described, reflecting mutations in the gene, most of these are not important in determining phenotype and SNP analysis can provide a cost-effective, sensitive, and rapid methodology for typing of many classical markers. cDNA, produced from isolated RNA, has shown that A and B cDNA differ by seven nucleotides that lead to four amino acid differences between the *A*- and *B*-transferases that produce the A and B antigens [25]. The O sequence is identical to in its first part, except that there is a single base deletion, which results in a frame shift and produces a premature "stop" signal, leading to a truncated protein produced with no transferase activity [26].

ABO frequencies differ across the world. Individuals originating from western and central Europe have high frequencies of both group O and A. In America, and in parts of African and Australia, group O is the most common. Indeed, in South America, some populations are mainly group O, suggesting that they were group O before the European invasion [9]. While the frequency of A is high in most of Europe, especially in the northwest, it is rare in virtually all of the other indigenous populations. Group B is most prevalent in central Asia with a clinal decline, east to west, across Europe [27]. Table 5 gives estimated frequencies of ABO phenotypes across three different population groups.

Table 5 ABO blood group frequency estimates across three populations (Caucasian, black, and South Asian) resident in the United Kingdom [28]

Blood Group	Caucasian	Black	South Asian
A	0.410	0.228	0.238
B	0.111	0.225	0.287
AB	0.027	0.043	0.087
O	0.452	0.505	0.388

MNS Blood Group System

The MNS blood group system was first fully described by Sanger and Race in the middle of the twentieth century [29]. M and N were first described by Landsteiner and Levine in 1927 [30], but were later found to be associated in some way with S [31]. About 86% of individuals who possessed the S antigen, also had the M antigen [32]; later, when the *s* allele was identified, family studies showed that the MN and Ss blood group systems were, indeed, closely linked (unlikely to be separated by recombination events) and so are usefully considered together as a complex blood group system [33]. Both MN and Ss are polymorphic in all populations and provide valuable additional discrimination in human identification of blood. In a Caucasian population, about 28% have MM, 22% have NN, and 50% have both: MN. Similarly, about 11% have SS, 45% ss, and 44% Ss.

Unlike in the ABO blood group system, antibodies to M and, especially, to N are not very common and, originally, typing was done in the cold using antibodies produced in rabbits [34]. The seeds of the *Vicea graminea* plant were also shown to have potent anti-N activity [35]. Anti-M and anti-N are cold-reacting and agglutinate at room temperature, like ABO. In contrast anti-*S* and anti-s are immune antibodies, active at 37 °, and so these antibodies can be identified within serum of exposed individuals. Anti-s is, however, very rare [36], and the presence of the s antigen is usually inferred from the absence of the S antigen in people of non-African origin.

S, with or without s, is associated with the U antigen, discovered by Wiener [37] in the mid-1950s, and individuals who express neither S nor *s* also tend to have absent or very low levels of U. Wiener proposed that a pair of alleles control the expression of U. Individuals without the U antigen are defined as"*u*", meaning an absence of U. Its association with Ss led to the definition of the S^u antigen for the silent gene at the *Ss* locus that also produces no antigen (S-s-U-, or S-s-u). Although U− red cells are almost always S-s-, about 50% of S-s- cells are U+, although the strength of the U antigen in these is variable, and it is generally considered as a variant allele of U. The presence of this variant U is almost exclusively

confined to people of African origin [38]. Individuals with S^u have a deficient or truncated glycophorin B GPB molecule [39]. The prevalence of u amongst blacks, albeit low, means that MNSs typing amongst this population should include the use of anti-*s*.

More recent methods have allowed the manufacture of *monoclonal* antibodies. Monoclonal antibodies are typically made by fusing myeloma cells (cancer cells that can be maintained *in vitro*) with the spleen cells (containing antibody-secreting cells but which die *in vitro*) from a mouse that has been immunized with the desired antigen, and the antibody reagent is produced by clones (*murine hybridomas*) [40].

The lipid membrane of red cells contains many proteins and glycoproteins [41]. Some are heavily glycosylated (are *glycophorins*) and two carry the MN and Ss determinants – GPA and GPB, respectively.

The M and N antigens differ according to the amino acid sequence of GPA, three nucleotide substitutions producing two amino acid changes. M individuals have serine at position 1 and glycine at position 5; in contrast N individuals have leucine and glutamic acid, respectively, in these positions [42]. The determination of the S and s antigens depends on a single amino acid substitution (Met-Thr) at position 29 [43]. Both of these SNP variants can be used in molecular assays to predict MNSs blood type from any nucleated material.

M and N are much less variable between populations across Europe, Africa, and East Asia, being around the 50% level [44]. High levels of M (over 90%) are found in Native Americans and low levels (down to 2%) are found in the Pacifics [27, 45]. S is less common in the Far East than in Europe and both *S* and *s* are almost always found in Caucasians, although about 1% of Africans lack both (S-s- or S^u) [27]. Table 6 shows the estimated frequency of the different linked MNSs blood groups across three population groups.

Table 6 MNSs blood group frequency estimates across three populations (Caucasian, black, and South Asian) resident in the United Kingdom [28]

Phenotype	Caucasian	Black	South Asian
MS	0.058	0.003	0.051
Ms	0.109	0.161	0.171
NS	0.005	0.009	0.009
Ns	0.137	0.249	0.102
MSs	0.154	0.057	0.176
NSs	0.047	0.035	0.052
MNS	0.025	0.025	0.029
MNs	0.219	0.356	0.216
MNSs	0.246	0.085	0.194
MS^u		0.003	
NS^u		0.009	
MNS^u		0.006	

Rhesus Blood Group System

The Rhesus blood group system is the most complex of the blood group systems consisting of 46 different antigens.

In 1940, Landsteiner and Wiener injected the blood of a Macacus rhesus monkey into guinea pigs and rabbits [46]. The antiserum obtained not only agglutinated rhesus monkey red cells but also approximately 85% of a panel of New Yorkers. A lack of association with any of the other known blood groups made Landsteiner and Wiener realize that they had detected a new blood group antigen. They named the antibody as anti-Rh and individuals whose blood was agglutinated by this reagent as Rh positive, in contrast with the nonreactors, which they termed Rh negative. Family studies showed that Rh positive was inherited as a dominant factor [47].

Prior to Landsteiner and Wiener's work, Levine and Stetson had published a case study about a woman who had delivered a stillborn baby [48]. Shortly after the birth, she received a blood transfusion from her husband and had a hemolytic reaction. It was proposed that she had become immunized by being exposed to a "foreign" fetal antigen inherited from the father, and the antibodies that were produced agglutinating the husband's red cells on their transfusion. This report was the catalyst for a series of papers in which Wiener and others showed that the antibody was the same anti-Rh factor that Landsteiner and Wiener had described and that it was owing to the immunization of an Rh-negative mother by red cells from an Rh-positive fetus, *in utero*. They also showed that antibody in the mother could cross the placenta and result in the condition of *erythroblastosis foetalis*, now known as *hemolytic disease of the newborn*, that can lead to severe anemia or death of the fetus (*hydrops fetalis*) [49]. Soon, this antibody was described in several cases of transfusion reaction in which ABO compatible blood had been given [50].

In fact, the anti-Rh that was produced in animals was soon shown to be different from that produced within humans, but by this time, the term Rh had become established in the literature [51].

Some antisera did not seem to react in the same way as others, and it was soon realized that there may be more than one Rh antigen. On both sides of the Atlantic, researchers were coming to very similar conclusions as to how these new antigens were related. Other antisera with different Rh specificities were reported: Wiener had three antisera that defined six alleles [52], and Race had four antisera that defined seven alleles [53]. Apart from anti-Rh, antibodies with different specificities were called *anti-Rh_1*, *anti-Rh_2* and *St*. This led to two different theories to describe and name the various Rh blood group antigens: that developed by *Fisher* and *Race* is perhaps the easier to understand [54], but can be fully related to the nomenclature proposed by *Wiener* [55].

Wiener's theory proposed multiple alleles at a single locus. He suggested that each gene encoded for an "agglutinogen" (antigen), which he called R^1, R^2, R^o, R^z, r, r', r'', and r^y, composed of several blood factors, Rh^o, rh', rh'', hr' and hr''.

Fisher's hypothesis was based on the observation that two of Race's four antibodies were antithetical [54]; everyone either reacted with one person or the other, but not both. He called these anti-C and anti-c, recognizing the C and c antigens. The other two antibodies did not share this pairing characteristic and were later shown to be anti-D and anti-E. The *Fisher–Race* hypothesis proposed three closely linked loci, C or c, D or d, and E or e, inherited as a *haplotype*. Apart from anti-d, all the other antibodies against each of these proposed antigens have been described. Although d is used in the nomenclature, it is now considered to be a silent allele. Anti-D was the same as the original anti-Rh, and thus it is just the presence or absence of the D antigen that defines whether someone is defined as Rh positive, or Rh negative.

While D, in the Fisher–Race nomenclature, has been shown to be the same as Wiener's blood factor Rh^o, rh' and rh'' are the same as C and c, and hr' and hr'' translate into E and e. Table 7 shows how these two nomenclatures are interchangeable and identify the eight haplotypes of Fisher–Race, or the eight Rh–Hr agglutinogens of Wiener. Agglutinogens named with an R indicate the presence of the D antigen; those named r do not have the D antigen.

In 1986, Tippett [56], on the basis of her observations of rare complexes with aberrant expression of the Rh antigens, proposed two genes, closely linked, one coding for D and the other coding for the CcEe antigens. This model has since been shown to be close to the truth.

Most Rh antibodies used for testing are immune, are IgG, and react with antigen optimally at 37 °C [9]. Preparation of antisera for typing requires very careful assessment to ensure that the antisera are of sufficient strength and specificity [57]. Some antibodies are "complete" and require only incubation with the antigen in saline to produce agglutination, whereas other "incomplete" antibodies require other treatment. Complete antibodies are normally IgM, and incomplete are IgG. IgM is a pentamer, and this makes it more able to bind to several antigen receptors on different red cells, facilitating agglutination. IgG, in

Table 7 Components of the Rh blood group system according to nomenclatures proposed by Fisher–Race and Wiener

Fisher–Race nomenclature	Wiener nomenclature			
Haplotype	Rh–Hr agglutinogen	Blood factors		
CDe	R^1	rh'	Rh^o	hr''
cDE	R^2	rh''	Rh^o	hr'
cDe	R^o	rh''	Rh^o	hr''
CDE	R^z	rh'	Rh^o	hr'
Cde	r'	rh'		hr''
cdE	r''	rh''		hr'
CdE	r^y	rh'		hr'
cde	r	rh''		hr''

contrast, is dimeric, and other methods can be used to detect the antigen–antibody complex.

Suspension of the red cells within a protein environment sometimes facilitates agglutination of these antibodies [57]. Use of *bovine albumin* at a 20%–30% concentration has been found to be the most effective, while avoiding problems of pseudoagglutination associated with other proteins. Treating red cells with a *proteolytic enzyme*, such as papain, trypsin, ficin, or bromelin, can reveal more antigen sites and reduce the negative charge on red cells, enabling cells to get closer to each other, making it easier for the small incomplete antibodies to bind [57]. *Blocking tests* can be used to detect incomplete antibodies, but cannot do so in a mixture of complete and incomplete antibodies unless the former are removed first. Incubation of Rh D positive red cells with excess incomplete anti-D in saline fills all of the antigen-binding sites, without agglutination – the cells are said to be "blocked" as no more antibodies can be bound. Subsequent attempts to produce agglutination with saline-acting anti-D fail. This latter technique is only really of use in the detection of antibody, rather than typing of the antigen [57]. The most commonly used method when using incomplete antibody is the *antihuman globulin* (*AHG*) or *Coombs test* [5]. Like the blocking tests, this can be used in a "direct" manner to detect the presence of antibody already on red cells, but its use in typing of antigen involves the indirect form of the test. This involves mixing of antiserum and test cells. If the antigen is present on the red cell, then the antibody binds. Rigorous washing of red cells removes any excess antibody and then the presence of antibody on the cells can be detected by addition of antihuman globulin, resulting in agglutination. A negative reaction clearly indicates absence of antigen.

Typing can sometimes produce inconclusive results. Stratton [58] described the D^u variant allele, which was positive with IgG anti-D antisera detected through the AHG test, but not with the use of IgM anti-D. D^u was not an antigen, but related to the number of D antigen sites available on a red cell. D^u is now called *weak D* [59].

Anti-D antisera are mostly IgG, but many also contain an IgM components [9] – they are *polyclonal.* The development of monoclonal antisera to replace polyclonal reagents has not been as simple as for the MNS system as the mouse hybridoma approach failed. Instead, human, or mixed human–mouse hybridomas are used. Typically, Epstein–Barr virus (EBV)-transformed lymphoblastoid cells [60], or CD40 activated lymphocytes [61], are fused with myeloma cells to produce a hybridoma.

Because of the large number of haplogroups, Rh blood typing is more useful than most other markers. Stains can be typed using the absorption-elution technique [17]. The method is not as simple as that used for ABO and MNS typing because the number of antigenic sites is not as high and more material must be used. It is also not possible to elute directly into a cell suspension because incubation at high temperature to disturb the antigen–antibody complex also destroys the ability for red cells to agglutinate. Generally, incomplete antibodies are used for detection, because they are of a higher titer and apparently more heat stable, along with enzyme-treated indicator cells.

Because large amounts of the stain must be used, and an essential component of the tests is the ability to wash out any unbound antibody, the composition of the stained fabric becomes important. Rh typing of denim is not normally successful, probably because of the "size" treatment of the material to facilitate the dye coloring. Detection of weak D forms are difficult, as is typing of stains that are old, and therefore, high avidity antisera and confirmatory controls must be employed [17].

Molecular analysis has shown that the Rh antigens are encoded by two closely linked genes on chromosome 1 (RHD and RHCE), producing the RhD and RhCcEe proteins, which are hydrophobic nonglycosylated proteins that appear to loop in and out of the red cell membrane 12 times [62]. The two polypeptides differ by about 35 amino acids but the gross gene structure is very similar. They each have 10 exons and have about 94% homology [63]. The greatest difference between the two genes is seen in intron 4. The two genes are orientated tail to tail, with their 3′ ends facing each other. On either side of the RHD gene are two homologous regions known as *Rhesus boxes* [64]. Deletion of RHD, the most common reason for the Rh-negative phenotype, occurs between two regions within each of the Rhesus boxes. This also explains why a *d* antigen, allelic to D, has never been found.

A variety of molecular techniques can now be used to predict Rh phenotype in a variety of nucleated material. Some methods exploit the presence or absence of RHD to detect Rh D phenotype, looking for particular sequences on RHD, but not RHCE [65];

another method uses sequence specific primers to identify different sized products from exon 7 [66], and a third method uses a single pair of primers across exon 4 of both genes producing 600 bp from RHD and 1200 bp from RHCE, with the Rh D negative phenotype being predicted by the absence of a band after electrophoresis of the polymerase chain reaction (PCR) products [67].

Errors may occur using these methods if variants are present and tests of more than one region of RHD are recommended [68]. Most Rh D negative Caucasian people in the United Kingdom are cde/cde and don't possess introns 4 and 10, but other Rh D negative haplotypes produce discordant results due to frameshifts and stop codons. Rh D negativity in about 65% of black Africans is due to an inactive RHD gene (RHD_{ψ}), either inherited as a homozygote, or as a heterozygote with the RHD deletion gene. RHD_{ψ} is due to a duplication of a region of intron 3 and exon 4, which causes a frameshift and potentially introduces a stop codon, but there is also a mutation in exon 6 that also produces a stop, ensuring that, despite the presence of the RHD gene, no RhD protein is produced. A multiplex approach has been described that detects exon 7 and intron 4 from RHD, RHD_{ψ}, and the C and c alleles of RHCE [69].

RHD_{ψ} is much lower in frequency in black African-Americans, but two other common variants seen in Rh D negative black Africans are seen more commonly amongst the black American population. This difference is likely to be due to owing to *founder effects* related to the particular African populations involved in the slave trade.

The Rh C/c polymorphism is associated with six SNPs in RHCE that produce four different amino acids [70]. Of particular importance is the T307C polymorphism in exon 2 of RHCE, determining C, but exon 2 is identical in RHD and RHCE, and so allele-specific primers for C are only useful in Rh D negative people. C/c genotyping can now be done in two PCR reactions to detect a C-specific region within intron 2 of RHCE and a second to detect the C nucleotide at 307 in exon 2, which defines c [71].

Both C and c are commonly seen in Caucasians (at 68% and 81%, respectively) [72], but C is lower and c higher amongst the African population, in contrast to East Asia, where the frequency of C approaches 100% and c is much lower [27, 45] C^w and C^x were initially thought to be alleles of C/c, but were subsequently found to be alleles of a high incidence antigen, *MAR* [73], owing to two different SNPs in exon 1 of RHCE producing two different amino acids in the RhCe protein [74]. The different proteins produced appear to result in conformational changes that lead to quantitative and qualitative abnormalities of $C.C^w$ is seen in about 2.6% of European populations, whereas C^x is much rarer, being seen in about only 0.12% of this population [9]. Both variants are, however, much more common in Finland [73] where their frequency is around 9% and 2%, respectively, and it is therefore not surprising that it was in this population that anti-MAR was first recognized.

Rh E/e is another pair of antithetical antigens and is associated with the C676G SNP in exon 5 of RHCE, producing a protein change in one of the extracellular loops of the polypeptide RhCcEe protein [70]. As in C, G676 on RHCE, determining e, has the same sequence on RHD, yet the latter does not express the e antigen; this suggests that the expression of the e antigen is a conformational difference. This also means that the e antigen cannot be defined in everyone by simply using the G676 marker, but detection can be combined with an RHCE specific nucleotide in exon 5 (A787) [75]. The E antigen can be defined by detecting the C676 SNP in exon 5 [76]. There is very little variation in E and e frequencies across populations [27], e being seen in 98% of the population with E present in around 30% [72].

These population frequency differences amongst the separate antigen pairs reflect the different haplotype frequencies between populations [27]. Rh-negative haplotypes are rare in South-East Asian populations and in indigenous Americans, whereas about 15% of white Europeans are Rh negative. CdE (r^y) is very rare in all populations, whereas cDe (R^o) is the commonest haplotype in Sub-Saharan Africa, being found in about 60%, in comparison with a frequency of less than 5% in other populations.

Inheritance of the Rh blood group is Mendelian, with each parent passing a haplotype to the child. The Rh blood groups therefore provide a powerful system in *parentage testing*, but they can be complex to interpret, and prediction of actual haplotypes that have been inherited is difficult from the phenotype [77]. The eight established haplotypes can be paired into 36 different genotypes, but using the five standard antisera only 18 phenotypes can be distinguished, and, because many phenotypes represent more than one genotype, only eight of these represent

a single genotype and the CcDEe phenotype could be explained by six different genotypes. For example, an individual with the phenotype cde, can only have the genotype cde/cde, and someone with the phenotype cdEe must be of genotype cde/cdE. This is clearly not the case for someone with a phenotype such as cDEe, which could be the result of cDE/cde or cDE/cDe, or cDe/cdE genotypes. In addition to doubts about the true genotype, prediction of the most likely genotype will depend on the population of origin, and it must be remembered that in the case of disputed paternity the tested man may not originate from the same population as the true father. The phenotype example cDEe is most likely to be cDE/cde, if both parents are of Caucasian origin. In an African origin population, cDE/cDe would be the most likely genotype. When using these markers for paternity tests, it is important to predict all possible genotypes of the child in order not to exclude a man whose likely genotype does not match with the likely genotype of the child.

Despite there being a large number of possible genotypes, many haplotype combinations are very rare. Table 8 gives the frequencies of observed phenotypes in three different population groups.

Protein Polymorphisms

Many enzymes (*see also* **Enzymes**) and other blood proteins exist in different polymorphic forms. *Isoenzymes* are enzymatically active proteins that catalyze the same reaction within a species, but have different physical or chemical properties. One such difference can lead to a variation in mobility when exposed to an electric field and this can be examined through *electrophoresis*.

A large number of polymorphic proteins and enzymes, found in many body tissues but normally analyzed within red cells or serum, have been utilized in forensic analysis and in the investigation of paternity [17]. Many of those analyzed for forensic purposes are red cell enzymes: these include adenosine deaminase (ADA), adenylate kinase (AK), carbonic anhydrase (CA), erythrocyte acid phosphatase (EAP), esterase D (EsD), glucose-6-phosphate dehydrogenase (GPD), glutamate-pyruvate transaminase (GPT) glyoxylase I (GLO), peptidase A (Pep A), 6-phosphogluconate dehydrogenase (PGD), and PGM. Others, found within red cells, or within serum, can also be useful in forensic analysis, but have more often been utilized as polymorphic markers (in conjunction with the markers already mentioned) in the investigation of paternity. This required the use of a large number of independent genetic systems to obtain sufficient power to answer questions of parentage prior to DNA testing. These systems include the following: α^1-anti-trypsin (PI), α^2-HS glycoprotein (α^2-HSG), group specific component (Gc), GM blood group factor (GM), haptoglobin (Hp), hemoglobin (Hb), KM blood group factor (KM), plasminogen (PLG), and transferrin (Tf) [28]. Many of these systems are discussed below.

Red cell enzymes lose their activity in stains more readily than red cell antigens and are, on the whole,

Table 8 Rh blood group frequency estimates across three populations (Caucasian, black, and South Asian) resident in the United Kingdom [28]

Phenotype Fisher–Race	Phenotype Wiener	Caucasian	Black	South Asian
CDe/cDE	R^1R^2	0.126	0.031	0.145
CDe/CDe	R^1R^1	0.163	0.016	0.448
cDE/cDE	R^2R^2	0.035	0.012	0.028
CDe/cde	R^1r	0.346	0.248	0.277
cDE/cde	R^2r	0.136	0.127	0.058
cde/cde	rr	0.171	0.046	0.025
C^wDe/cDE	$R^{1W}R^2$	0.001		
C^wDe/CDe	$R^{1W}R^1$	0.002		
C^wDe/cde	$R^{1W}r$	0.002		0.002
CDe/CDE	R^1R^Z	0.001		
cDe/cDe	R^o	0.007	0.508	0.009
Cde/cde	$r'r$	0.006	0.012	0.009
cdE/cde	$r''r$	0.004		

inactive after a few weeks. Analysis of these systems is through electrophoresis on some form of support medium. Those most commonly used in forensic analysis are:

Starch gels, traditionally made from potato starch, can be used on glass plates to form a thin opaque gel with a pore size approaching that of the protein molecules under investigation. It therefore acts as a molecular sieve. Detection of enzymes is through an assay that links function to a staining reaction and therefore may require that the enzymes are still functional after separation.

Acrylamide gels produce tough, flexible, clear gels in which the pore size can be adjusted by altering the percentage of acrylamide, polymerized through *cross-linking* with chemical reagents, or exposure to UV light, for example, creating density gradients. *Native gel electrophoresis* (*Native PAGE*) separates proteins according to their charge to mass ratio, which means that different proteins with the same molecular weight will migrate differently, as each protein has an isoelectric point and molecular weight according to is primary structure. *Sodium dodecyl sulphate polyacrylamide gel electrophoresis* (*SDS-PAGE*) denatures secondary protein structures and applies a negative charge proportional to its mass. This means that the charge to mass ratio is similar for all proteins, which migrate proportional to the size of the protein.

Adenosine Deaminase (ADA)

ADA deaminates adenosine to produce inosine. It is found in most tissues, particularly red cells, lymphocytes, and macrophages, but is not normally detected in semen or vaginal secretions. There are two main alleles produced at a single locus: ADA^1 and ADA^2 that lead to three phenotypes: ADA 1, ADA 2, and ADA 2–1 [78]. Rare variant alleles have also been described, and numbered 3, 4, and 5.

ADA in blood stains is analyzed through electrophoresis of lysates in thin layer starch gels at 4 °C [17]. The enzymes alleles are distributed in different zones, with the 1 variant moving further toward the anode. The enzyme variants are detected through an enzymic action: adenosine and subsequent detection chemicals (nucleoside phosphorylase, xanthine oxidase, Dimethylthiazol 2 yl diphenyltetrazolium bromide (MTT) tetrazolium and phenazine methosulphate (PMS)) are added to a 2% buffered agar solution, which is poured over the gel anodal to the origin. When set, the whole is covered and incubated at 37 °C for the enzymic reaction to take place.

Adenosine in the agar is converted to inosine by the ADA in lysate. Inosine is subsequently converted to hypoxanthine by nucleoside phosphorylase, added to the agar reagent. Hypoxanthine is then oxidized by xanthine oxidase and, in the presence of PMS, MTT tetrazolium, which has a yellow color, is reduced to MTT formazan, which is an insoluble blue compound. The bands therefore appear blue on a yellow background.

Typically, there are two or three bands, one band being common to all. ADA 2-1 has three bands. In ADA 1 the cathodal band is absent and in ADA 2 the cathodal band is more prominent than the central band and there may be some weak activity in the anodal band area. Variants with type 1 show bands that are more cathodal (types 3 and 4) or more anodal (type 5) than the normally observed bands. These variants are very rare.

Problems can arise due to oxidation of lysates although this can be reversed through treatment with β-mercaptoethanol [17]. Oxidation results in the faster anodal bands becoming stronger and the slower bands weaker, leading to mistyping in these situations. In material that is very old, or is, perhaps, infected with bacteria, a single slow moving band is seen. Phenotype frequencies of the ADA variants are shown in table 9.

ADA phenotypes can be predicted through analysis of a G/A SNP that results in an amino acid change from aspartic acid (Asp) to asparagine (Asn) [79].

Adenylate Kinase (AK)

AK catalyzes the transfer of a high energy phosphate group from one molecule of adenosine diphosphate (ADP) to another to produce one molecule of

Table 9 ADA frequency estimates across three populations (Caucasian, black, and South Asian) resident in the United Kingdom [28]

ADA	Caucasian	Black	South Asian
1	0.912	0.979	0.751
2–1	0.085	0.015	0.236
2	0.003		0.013
1-var		0.001	

adenosine triphosphate (ATP) and one molecule of adenosine monophosphate (AMP) with magnesium ions acting as a cofactor. It is found in most tissues, but is difficult to detect using electophoretic techniques in semen and vaginal secretions. There are two main alleles: AK^1 and AK^2 that result in three phenotypes: AK 1, AK 2–1 and AK 2, although the latter is very rare [80]. Variants have also been described.

Detection is through electrophoresis in a thin layer starch gel [17]. Typically, a cotton thread, soaked in a red cell lysate, is introduced into the gel and electrophoresed in the cold. After electrophoresis, enzyme detection reagents (ADP, glucose, magnesium chloride, hexokinase, G6PD, Nicontinamide adenine dinucleotide phosphate (NADP), MTT tetrazolium, and PMS) are introduced in an agar overlay with incubation in the dark at 37 °C.

AK in the gel converts ADP in the detection reagent to produce ATP and AMP. ATP acts on glucose to produce glucose-6-phosphate (G6P). Glucose-6-phosphate dehyrogenase (G6PD) then oxidizes G6P to produce 6-phosphogluconate and NADP is converted to Reduced nicontinamide adenine dinucleotide phosphate (NADPH). Reoxidation to NADP occurs along with reduction of MTT tetrazolium to the insoluble blue MTT formazan, with PMS acting as a cofactor. Like ADA, the enzyme bands of AK activity appear blue on a yellow background.

AK 2–1 appears as two strong bands and a weaker cathodal band. In AK 1 homozygotes, the most anodal band is absent; in AK 2 the most anodal band is the strongest, with a weak central band. A very weak cathodal band may be seen. Standard analysis is done at pH 5.0, but rare variants are better separated using pH 7.0. Problems can occur if the red cells are not properly lysed as this can produce weak shadow bands leading to interpretation difficulties.

The AK1 (red cell) gene is found on chromosome 9 and has seven exons [81]. The AK locus has been shown to be closely linked to the ABO locus [82] and this must be borne in mind if these markers are to be used together in analysis of parentage. Table 10 shows the frequencies of the various phenotypes across three populations.

Table 10 AK frequency estimates across three populations (Caucasian, black, and South Asian) resident in the United Kingdom [28]

AK	Caucasian	Black	South Asian
1	0.926	0.979	0.849
2–1	0.073	0.021	0.146
2	0.001		0.005

Erythrocyte Acid Phosphatase (EAP or ACP1)

The *EAP* enzyme is now more frequently known as soluble *acid phosphatase 1* (*ACP1*) and is distinct from the ACP2 gene (*see also* **Acid Phosphatase**). Although called 'red cell', there are reports of the enzyme being found in other tissues, which are also known to demonstrate a variety of other acid phosphatases. EAP acts as an acid phosphatase and as a protein tyrosine phosphatase. It hydrolyzes orthophosphoric monoesters to alcohol and orthophosphate, and protein tyrosine phosphate to protein tyrosine and orthophosphate. There are three common variants, EAP^A, EAP^B, and EAP^C, which give rise to six phenotypes: A, B, C, BA, CA, and CB [83]. Rare types R and D have also been described using starch gel electrophoresis [84]. SDS-PAGE shows that each allele encodes two electrophoretically different isozymes, 'f (fast)' and 's (slow)', which are produced in allele-specific ratios (2 : 1 for Af and As; 4 : 1 for Bf and Bs; 1 : 4 for Cf and Cs) [85].

Traditionally, EAP is analyzed in a thick (2 mm) gel, necessary to detect weaker bands [17]. Lysates of blood mixed with dithiothreitol (DTT) (a reducing agent to prevent oxidation complications of additional bands associated with storage) on filter paper or within stained fibers are electrophoresed with a cooling plate. Postelectrophoretic analysis involves placing filter paper, soaked with the reaction mixture, on the gel, followed by incubation within a closed box to prevent evaporation.

Detection involves EAP catalyzing the removal of a phosphate group from 4-methyl umbelliferyl phosphate to form methyl umbelliferone, which fluoresces under UV light. A complex pattern of bands is produced with movement from the origin toward the anode. The A band is strong and the most anodal and is associated with a second band, more cathodal, the latter being weaker in heterozygotes. B and C bands are seen as two bands in the same positions, flanking the weak cathodal A band. In B, however, the more anodal band is the stronger, and in C, it is the more cathodal band that is the strongest. The rare R variants are found beyond the anodal A band, whereas the D variant travels in a cathodal direction and is

Table 11 SNP typing of the ACP1 locus

Phenotype	SNP observed at the 15-3F locus	SNP observed at the 12-3S locus
A	C	T
B	C	C
C	T	C
BA	C	T/C (heterozygous at this locus)
CB	T/C (heterozygous at this locus)	C
CA	T/C (heterozygous at this locus)	T/C (heterozygous at this locus)

found on the other side of the origin. R is more often observed in people of African origin. C is rare in all populations and homozygous have not been observed in a large population series.

Distinction of C and CB can sometimes be problematic as differentiation depends on a weak B position band in the former, which is stronger in the latter. Similarly BA has strong bands in the B and A positions, with a weak band in the C position, in comparison with CA, which has strong bands in all three positions, but particularly in the C and A positions. Molecular analysis has shown that the coding portions of the B and C alleles are identical and it is the different ratios of the f and s isozymes that explain the identical electrophoretic mobility but different intensity of the bands. The A allele has a single amino acid substitution at residue 105.

The ACP1 gene is located on chromosome 2 and consists of seven exons. Alternative splicing of exons 3 and 4 account for the different f and s isozymes, which probably serve different functions within the cell [86]. The sequences of these isozymes are identical, apart from a 108 bp sequence within the reading frame.

The presence of SNPs at two locations (ACP1-15-3F and ACP1-12-3S) allow molecular phenotype prediction [71, 87]. The two SNPs are both C/T – (Asp)GA*C* to (Asp)GA*T*, and (Ser)AG*C* to (Ser)AG*T* predict the f and s isozymes. Table 11 shows how these SNPs can be interpreted:

Table 12 shows how these phenotypes vary in frequency across different populations.

Table 12 EAP (ACP1) frequency estimates across three populations (Caucasian, black, and South Asian) resident in the United Kingdom [28]

EAP (ACP1)	Caucasian	Black	South Asian
A	0.125	0.050	0.087
B	0.356	0.543	0.482
C			
BA	0.430	0.342	0.422
CB	0.057	0.006	0.008
CA	0.032	0.003	0.002
RB	0.001	0.045	
RA		0.012	

Esterase D (ESD)

A variety of different esterases exist. In particular, the ESD enzyme has been found to be identical to S-formylglutathione hydrolase (FGH) and hydrolyzes thiol esters of glutathione. There are two common variants: ESD^1 and ESD^2, producing three phenotypes: 1, 2 and 2-1 [88], although a large number of rare variants have been described [89].

Electrophoresis of lysates in DTT is achieved through placing stained cloth inserts in thin layer starch gels as relatively large amounts of material need to be applied [17]. Detection reactants are soaked in filter paper, which is placed on the gel, postelectrophoresis, followed by a short incubation at 37 °C. The gel is viewed under UV light to detect fluorescent bands.

Like EAP, ESD can catalyze the removal of a phosphate group from 4-methyl umbelliferyl phosphate to form the fluorescent methyl umbelliferone. Five bands are defined, numbered 1 to 5, from cathode to anode. Weak fluorescent bands are also seen, one cathodal of position 1 and the other coincident with position 5. These weak bands are related to other esterase enzymes. ESD 2–1 shows three bands, strong in position 2 and weak in position 1 and 3. ESD 1 has two bands, strong in position 1 and weak in position 2; ESD 2 also has two bands, strong in position 3 and weak in position 4.

While ESD 2 is more readily identified because of its very different mobility, distinguishing between

Table 13 ESD frequency estimates across three populations (Caucasian, black, and South Asian) resident in the United Kingdom [28]

ESD	Caucasian	Black	South Asian
1	0.773	0.848	0.539
2–1	0.177	0.139	0.360
2	0.011	0.009	0.079
5–1	0.032	0.003	0.016
5–2	0.006		0.006
5	0.001		
1-var	0.001		

ESD 1 and ESD 2–1 can be problematic as this depends on the relative strengths of the bands. This is particularly the case in stains, which often produce weak and diffuse results in which the faint bands are difficult to see. Additional bands also result after frozen storage, or in dried stains, but addition of more DTT can remove these and help reduce background fluorescence.

ESD is located on chromosome 13 and the variants are due to a polymorphism at a single locus. SNP typing can be used to differentiate the two variants: (Gly)C*G*A to (Glu)C*A*A, the former defining ESD 1, and the latter, ESD 2 [89]. Table 13 shows the frequency of the phenotypic variants in three populations.

Glyoxalase I (GLO-I)

GLO-I catalyzes the conversion of methyl glyoxal to S-lactolyl glutathione lactate using glutathione as a cofactor; glyoxylase II (GLO-II) subsequently converts this to lactic acid and glutathione. GLO-I is associated with cell division and has been found in a variety of tissues, including semen. Two main types exist: GLO^1 and GLO^2, leading to three phenotypes: 1, 2 and 2–1 [90].

Lysates in mercaptoethanol solution (which acts as an antioxidant) are electrophoresed in a cooled thin layer starch gel [17]. Although mercaptoethanol and DTT (used in other assays) are similar, they produce different GLO-I mobilities. Postelectrophoretic treatment involves placing filter paper soaked in detection reagents (methyl glyoxal, glutathione, iodine) on the gel and incubating at 37 °C, followed by removal of the filter paper and pouring of an indubiose agarose solution containing iodine over the gel.

GLO-I catalyzes the reaction that converts methyl glyoxal and glutathione to S-lactolyl glutathione. In the absence of free glutathione, the iodine reacts with starch giving a blue-black color. The remainder of the gel surface still has reduced glutathione on it, which has a reactive SH group that reacts with the iodine, leaving insufficient free iodine to react with the starch gel.

GLO-I type 2–1 has three bands, the central one being weaker. Type 1 shows just the more cathodal band, and type 2 the more anodal band. Problems can arise if the iodine is not properly dissolved to avoid blue coloration in the agar overlay. If stains are old, the cathodal band of the type 2–1 is weak and the anodal band stronger, which might, falsely, suggest the presence of a variant.

GLO-I is located on chromosome 6 and shows six exons [91]. It shows strong linkage with the Human leucocyte antigen (HLA) locus, so the results cannot be combined in the statistical assessment of paternity [92]. GLO-1 phenotypes can be predicted through use of an SNP at a single locus that leads to an amino acid substitution: (Ala)G*C*G to (Glu)G*A*G, the former defining type 1, and the latter, type 2 [93]. Table 14 shows the allele frequencies across three population groups.

Phosphoglucomutase (PGM)

PGM 1 catalyzes the reaction that transforms glucose-1-phosphate into glucose-6-phosphate. It is found in all human tissues. Three unlinked loci determine the production of PGM [94]; here we are concerned with only the first, which has four common variants: 1+, 1−, 2+, and 2−, producing 10 different common phenotypes. Rare variants 3+, 3−, 7+, 7− are seen in some oriental population groups.

Starch gel electrophoresis detected the types PGM_1-1 and PGM_1-2 [95], but isoelectric focusing in acrylamide gels enables these isozymes to be split

Table 14 GLO-I frequency estimates across three populations (Caucasian, black, and South Asian) resident in the United Kingdom [28]

GLO-I	Caucasian	Black	South Asian
1	0.214	0.112	0.060
2–1	0.486	0.423	0.350
2	0.300	0.465	0.590

Table 15 SNP typing of the PGM_1 locus. Using these SNPs it is not possible to distinguish between the 2 + 1− and 2 − 1+ phenotypes

Phenotype	SNP observed at the 723 locus	SNP observed at the 1320 locus
1+	C	T
1−	C	C
2+	T	T
2−	T	C
1 − 1+	C	T/C
2 + 1 − /2 − 1+	T/C	T/C
2 − 2+	T	T/C
2 + 1+	T/C	T
2 − 1−	T/C	C

into + and − forms, originally referred to as A and B forms [96]. Ampholines are used to produce a pH range between 5.0 and 8.0 within a UV-polymerized gel. Lysates in filter paper inserts are electrophoresed, and development of the postelectrophoretic gel is through an agar overlay with reaction chemicals (glucose-1 phosphate, glucose 6 phosphate dehydrogenase, PMS, MTT tetrazolium, and NADP) [97].

PGM catalyzes the reaction from glucose-1 phosphate to glucose-6 phosphate, which is then transformed to 6-phosphogluconate through the action of glucose-6-phosphate dehydrogenase [17]. At the same time NADP is reduced to NADPH. Using PMS as a transfer reagent, MTT tetrozolium is reduced to the insoluble blue MTT formazon and NADPH reoxidized to NADP. PGM activity bands are seen as blue bands on a yellow background.

The different allelic forms separate in the order 1−, 1+, 2−, 2+ from pH 7.0 through to pH 5.0 and phenotypes are readily identified from the band positions.

PGM_1 is found in chromosome 1 and the variants 1 and 2, and + and −, are due to reciprocal intragenic recombination between two SNP sites in exons 4 and 8, associated with amino acid substitutions: (Arg)*C*GT to (Cys)*T*GT and (Tyr)*T*AT to (His)*C*AT [98]. SNPs PGM_1 723 and PGM_1 1320 can be used to predict phenotype from genetic material.

Table 15 shows how SNP variants can be interpreted, and table 16 shows the phenotype frequencies across three population groups.

Table 16 PGM frequency estimates across three populations (Caucasian, black, and South Asian) resident in the United Kingdom [28]

PGM	Caucasian	Black	South Asian
1+	0.387	0.464	0.396
1 + 1−	0.186	0.166	0.152
1−	0.018	0.024	0.020
2+	0.024	0.033	0.063
2 + 2−	0.025	0.018	0.014
2−	0.003	0.003	
2 + 1+	0.227	0.199	0.241
2 + 1−	0.045	0.042	0.035
2 − 1+	0.071	0.045	0.065
2 − 1−	0.012	0.006	0.014
2var1−	0.001		

Group Specific Component (Gc)

Gc was first described by Hirschfeld [99] in 1959 using an immunoelectrophoretic technique to investigate the α_2 globulin of human serum. He recognized three different precipitation patterns that, family studies showed, were under genetic control. Immunoelectrophoretic patterns are characteristic precipitant bands that form in arcs in an agar gel [100]. Homozygote phenotypes Gc-1, Gc-2, and the heterozygote were described. Gc-1 was subsequently divided into "fast" and "slow" types [101], Gc-1F and Gc-1S, respectively, through use of the isoelectrofocusing technique, and a number of other variants were reported [102].

Gc was soon discovered to be identical to the vitamin D binding alpha globulin (VDAG) [103]. The high frequency of Gc-2 was associated with low levels of sunlight with the 2 gene being expressed in about 30% of the white population, in comparison with only about 7% of the black African population [104]. On the other hand, about 80% of the black African population has the Gc1F gene, in comparison with about 15–20% of other populations.

Thin layer acrylamide and *bis*-acrylamide gels are prepared using pH ampholines in the range 3.5–5.0 and 4.0–6.0 in a 5 : 1 ratio [105]. Addition of tetramethylethylenediamine (TEMED) and ammonium persulphate polymerizes the gel. Serum-soaked filter paper was placed close to the cathode and electrophoresed. Coomassie brilliant blue was used to stain the proteins and the rest of the gel destained.

The staining pattern produces between 1 and 4 bands. Gc-2 provides the maximum number of cathodal bands. Gc-1 provides a pair of more anodal bands, the Gc-1F being more anodal than the Gc-1S bands. Individuals with phenotypes involving both Gc-2 and Gc-1 produce three bands, and those with Gc-1F1S have four. A large number of variants have also been described. The technique has been used to successfully type blood stains in a collaborative study.

Gc is a single polypeptide; originally, two alleles were described Gc-1 and Gc-2, with Gc1F and Gc1S thought to be posttranslational [101]. Subsequent analysis has shown the gene to be present on chromosome 4, consisting of 13 exons, but the totality is not translated [106]. Exon 11 is responsible for the differences between the types with SNPs producing amino acid changes at positions 416 and 420 [102]. Table 17 summarizes these, and table 18 shows the phenotype estimated frequencies in three population groups.

Table 17 SNP typing of the Gc locus

Phenotype	SNP observed at the 416 locus	SNP observed at the 420 locus
1F	(Asp) GA*T*	(Thr) A*C*G
1S	(Glu) GA*G*	(Thr) A*C*G
2	(Asp) GA*T*	(Lys) A*A*g

Table 18 Gc frequency estimates across three populations (Caucasian, black, and South Asian) resident in the United Kingdom [28]

Gc	Caucasian	Black	South Asian
1S	0.303	0.018	0.354
1F-1S	0.169	0.196	0.213
1F	0.030	0.632	0.037
2-1S	0.317	0.033	0.183
2	0.087	0.003	0.091
2-1F	0.089	0.105	0.120
1S-var	0.002	0.003	0.002
1F-var		0.006	
2-var	0.003	0.003	

Haptoglobin (Hp)

Hp is a tetrachain glycoprotein, much like hemoglobin, and is only found in man. It is secreted in the liver as an α^2 globulin in serum and binds with hemoglobin after hemolysis. It is believed that this action reduces the loss of iron through urinary excretion during red cell breakdown, preventing kidney damage [107] There are two genes, Hp^1 and Hp^2, resulting in three phenotypes: Hp1, Hp2, and Hp2–1 [108]. In addition, there are a number of rare types, including Hp2–1M and Hp0 [109].

Two slightly different methods of detection are used, depending on whether serum or a blood stain is being used [17]. Serum is mixed with hemoglobin A (adult Caucasian red cells) to form a haptoglobin–hemoglobin complex, and soaked into a filter paper wick to be loaded onto a thick layer starch gel for electrophoresis. Alternatively, the bloodstain is eluted into buffer and used to load the gel. The gel is cut vertically so that the wick can be inserted into the full depth. After electrophoresis, the gel is sliced, horizontally, along its length. The open gel is stained with phenolphthalein, which is a colorless solution in alkaline, but in the presence of an oxidizing agent, such as the perioxidase of hemoglobin, a deep pink color develops. Reading should take place immediately because the color fades after a few minutes. Other stains, such as leucomalachite green of amido black can also be used.

The Hp1 band forms a single, thick band, close to the free hemoglobin band on its cathodic side. Free hemoglobin moves fastest from the origin. Hp2 and Hp2–1 show a cluster of bands that are further toward the cathode. The Hp1 component of Hp2–1 is in the same position as Hp1, but is much weaker, and this band is absent in Hp2. Hp2–1 has an additional band, closest to the cathode, which stains heavier than any of the Hp2 bands. Hp2–1M looks like the Hp2–1, but the cathodic cluster of bands is missing.

The haptoglobin gene is found on chromosome 16 and its variants seem to be associated with a duplication event, seen in Hp2, but not in Hp1, encompassing exons 3 and 4 [109]. Hp2–1M, is due to reduced Hp2 and an SNP is found in position 61 of the promoter region. Hp0 is an absence of haptoglobin (*ahaptoglobinemia*) and is believed to

Table 19 Interpretation of bands produced using primer pair amplification of the Hp gene

Primer pair	Purpose	Band present	Band absent
A	HpA and HpB – identifies Hp^1 by product size	Hp1, Hp2–1	Hp2
B	HpA and HpARev – identifies Hp^2 due to HpARev overlapping the duplication of Hp^1 in the Hp^2 gene	Hp2, Hp2–1	Hp1
C	HpA2 and HpB2 – identifies Hp^1 by product size	Hp1, Hp2–1	Hp2
D	HpC and HpD – identifies Hp^2 due to the section of Hp^2 amplified in the PCR product overlapping the duplication of Hp^1 in the Hp^2 gene	Hp2, Hp2–1	Hp1

Table 20 Hp frequency estimates across three populations (Caucasian, black, and South Asian) resident in the United Kingdom [28]

Hp	Caucasian	Black	South Asian
1	0.153	0.330	0.029
2–1	0.478	0.379	0.287
2	0.363	0.148	0.684
2–1M	0.001	0.076	–
0	0.005	0.067	

be due to SNPs at 61 and 101 within the promoter region [110].

Distinction of the main allelic types by molecular methods must target the duplication [111]. Seven primers defining forward and reverse sequences are used in four pairs for interrogation and confirmation of phenotype, and a set of SNPs are used to determine the common variant SNPs in the promoter region. Primers A and B target Hp^1, and primers C and D target Hp^2; confirmatory primers are also used. Table 19 shows how the results of the PCR using the primers are interpreted, and table 20 gives the phenotype frequency of the Hp markers across three populations.

Summary

This large collection of polymorphic antigens, enzymes, and other proteins have provided useful human identification markers over many years. The analysis of these markers generally, however, requires significant expertise, in comparison with DNA technology, and much of the equipment used and reagents of suitable quality are not so readily available.

As products of genes, classical markers are not subject to the same intergenerational mutations that can hamper investigations of parentage when DNA markers are used. In contrast, their use in criminal investigations is more limited, not only because of their poor availability in some tissues of interest, especially saliva and semen or, in the case of enzyme activity assays, they may be too old to be analyzed, but also because they are not a informative as DNA markers. While the use of DNA short tandem repeat analysis has overtaken classical markers in both forensic and parentage analyses, the greater understanding of the molecular basis of many of these classical markers means that it is now possible to predict most of the classical marker phenotypes in nucleated materials of any age and size, and this has already proven useful in analysis of old cases.

References

[1] Jeffreys, A.J., Wilson, V. & Thein, S.L. (1985). Individual-specific 'fingerprints' of human DNA, *Nature* **316**, 76–79.

[2] Strachan, T. & Read, A. (2003). *Human Molecular Genetics*, 3rd Edition, Garland Science.

[3] Garfin, D.E. (2003). Gel electrophoresis of proteins, in *Essential Cell Biology*, *Cell Structure, A Practical Approach*, J. Davey & M. Lord, eds, Oxford University Press, Oxford, Vol. 1, pp. 197–208.

[4] Landsteiner, K. (1900). Zur Kenntnis der antifermentativen, lytischen und agglutinietenden Wirkungen des Blutserums und der Lymphe, *Zentralbl Bakt* **27**, 357–366.

[5] Coombs, R.R.A., Mourant, A.E. & Race, R.R. (1945). A new test for detection of weak and 'incomplete' Rh

agglutinins, *British Journal of Experimental Pathology* **26**, 255–266.

[6] Lowe, J.B. (1995). Biochemistry and biosynthesis of ABH and Lewis antigens: characterisation of blood group-specific glycosyltransferases, in *Blood Cell Biochemistry*, J.-P. Cartron & P. Rouger, eds, Plenum Press, New York, Vol. 6, pp. 75–115.

[7] Kelly, R.J., Ernst, L.K., Larsen, R.D., Bryant, J.G., Robinson, J.S. & Lowe, J.B. (1994). Molecular basis for H blood group deficiency in Bombay (Oh) and para-Bombay individuals, *Proceedings of the National Academy of Sciences of the United States of America* **91**, 5843–5847.

[8] Oriol, R. (1995). ABO, Hh, Lewis and secretion: serology, genetics and tissue distribution, in *Blood Cell Biochemistry*, J.-P. Cartron & P. Rouger, eds, Plenum Press, New York, Vol. 6, pp. 36–73.

[9] Daniels, G. (2002). *Human Blood Groups*, 2nd Edition, Blackwell Science, Oxford.

[10] Costache, M., Cailleau, A., Fernandez-Mateos, P., Oriol, R. & Mollicone, R. (1997). Advances in molecular genetics of α -2- and α -3/4-fucosyltranferases, *Transfusion Clinique et Biologique* **4**, 387–382.

[11] Morgan, W.T.J. & Watkins, W.M. (1969). Genetics and biochemical aspects of human blood group A-, B-, H-, Le^a - and Le^b - specificity, *British Medical Bulletin* **25**(1), 30.

[12] Kind, S.S. (1960). Absorption-elution grouping of dried blood smears, *Nature* **185**, 397–398.

[13] Lincoln, P.J. & Dodd, B.E. (1973). An evaluation of factors affecting the elution of antibodies from blood stains, *Journal of Forensic Science Society* **13**, 37.

[14] McCurley, R.S., Recinos, A., Olsen, A.S., Gingrich, J.C., Szcwspaniak, D., Camerson, H.S., Krauss, R. & Weston III, B.W. (1995). Physical maps of human α (1,3) fucosyltransferase genes FUT-3-FUT-6 on chromosomes 19p13.3 and 11q21, *Genomics* **26**, 142–146.

[15] Grubb, R. (1948). Correlation between Lewis blood group and secretor status in man, *Nature* **162**, 933.

[16] Grubb, R. (1951). Observations on the human blood group system Lewis, *Acta Pathologica et Microbiologica Scandinavica* **28**, 61–81.

[17] Metropolitan Police Forensic Science Laboratory (1978). Body fluids 3.1–3.37, in *Biology Methods Manual*, Commissioner of Police of the Metropolis, London.

[18] Lee, H.C., Gaensslen, R.E., Pagliaro, E.M. & Novitch, B. (1988). Two-dimensional absorption-inhibition, *Journal of Forensic Sciences* **33**, 1127–1138.

[19] Jenkins, G.C., Brown, J., Lincoln, P.J. & Dodd, B.E. (1972). The problem of the acquired B in forensic serology, *Journal of Forensic Sciences* **12**, 597.

[20] Pereira, M. (1973). ABO grouping of decomposed tissue, *Journal of Forensic Sciences* **13**, 33.

[21] Sharon, N. & Lis, H. (1972). Lectins: cell-agglutinating and sugar-specific proteins, *Science* **177**, 949–959.

[22] Ferri, G., Bini, C., Ceccardi, S., Ingravallo, F., Lugaresi, F. & Pelotti, S. (2006). Minisequencing-based genotyping of Duffy and ABO blood groups for forensic purposes, *Journal of Forensic Sciences* **51**, 357–360.

[23] Yamamoto, F., Clausen, H., White, T., Marken, J. & Hakomori, S. (1990). Molecular genetic basis of the histo-blood group ABO system, *Nature* **345**, 229–233.

[24] Seltsam, A., Hallensleben, M., Kollmann, A., Burkhart, J. & Blasczyk, R. (2003). Systematic analysis of the ABO gene diversity within exons 6 and 7 by PCR screening reveals new ABO alleles, *Transfusion* **43**, 428–439.

[25] Yamamoto, F. & Hakomori, S. (1990). Sugar-nucleotide donor-specificity of histo-blood group A and B transferases is based on amino acid substitutions, *The Journal of Biological Chemistry* **265**, 19257–19262.

[26] Yamamoto, F. & McNeill, P.D. (1996). Amino acid residue at codon 268 determines both activity and nucleotide-sugar donor substrate specificity of human histo-group A and B transferases: *in vitro* mutagenesis study, *The Journal of Biological Chemistry* **271**, 10515–10520.

[27] Mourant, A.E., Kopec, A.C. & Domniewska, K. (1976). *The Distribution of Human Blood Groups and Other Polymorphisms*, Oxford University Press, London.

[28] London Hospital Medical College. *Phenotype and Gene frequencies of various red cell antigens, red cell antibodies and serum protein enzymes in over 1000 individuals self-declared as Caucasian, black African/Afro-Caribbean or South Asian (from the Indian sub-continent)*, Unpublished data.

[29] Sanger, R. & Race, R.R. (1951). The MNSs blood group system, *American Journal of Human Genetics* **3**, 332–343.

[30] Landsteiner, K. & Levine, P. (1927). A new agglutinable factor differentiating individual human bloods, *Proceedings of the Society for Experimental Biology and Medicine* **24**, 600–602.

[31] Walsh, R.J. & Montgomery, C.M. (1047). A new human iso-agglutinin subdividing the MN blood groups, *Nature* **160**, 504–506.

[32] Sanger, R., Race, R.R., Walsh, R.J. & Montgomery, C. (1948). An antibody which subdivides the human MN blood groups, *Heredity* **2**, 131–139.

[33] Greenwalt, T.J., Sasaki, T., Sanger, R., Sneath, J. & Race, R.R. (1943). An allele of the S(s) blood group genes, *Proceedings of the National Academy of Sciences of the United States of America* **40**, 1126–1129.

[34] Lansteiner, K. & Levine, P. (1928). On the inheritance of agglutinogens of human blood demonstrable by immune agglutinins, *The Journal of Experimental Medicine* **48**, 731–749.

[35] Duk, M. & Lisowska, E. (1981). *Vicia graminea* anti-N lectin: partial characterization of the purified lectin and its binding to erythrocytes, *European Journal of Biochemistry* **118**, 131–136.

[36] Levine, P., Kuhmichel, A.B., Wigod, M. & Koch, E. (1951). A new blood factor, s, allelic to S, *Proceedings of the Society for Experimental Biology and Medicine* **78**, 218–220.

[37] Wiener, A.S., Unger, L.K. & Gordon, E.B. (1953). Fatal haemolytic transfusion reaction caused by sensitization to a new blood factor U, *JAMA* **119**, 1444–1446.

[38] Wiener, A.S., Unger, L.J. & Cohen, L. (1954). Distribution and heredity of blood factor U, *Science* **119**, 734–735.

[39] Huang, C.-H., Johe, K., Moulds, J.J., Siebert, P.D., Fukuda, M. & Blumenfeld, O.O. (1987). δ glycophorin (glycophorin B) gene deletion in two individuals homozygous for the S-s-U- blood group phenotype, *Blood* **70**, 1830–1835.

[40] Kohler, G. & Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity, *Nature* **256**, 495–497.

[41] Blanchard, D. (1990). Human red cell glycophorins: biochemical and antigenic properties, *Transfusion Medicine Reviews* **4**, 170–186.

[42] Blumenfeld, O.O. & Adamany, A.M. (1978). Structural polymorphism within the amino-terminal region of MM, N and MN glycoproteins (glycophorins) of the human erythrocyte membrane, *Proceedings of the National Academy of Sciences of the United States of America* **75**, 2727–2731.

[43] Dahr, W., Beyreuther, K., Steinbach, H., Gielen, W., Krüger, J. (1980). Structure of the Ss blood group antigens. II. A methionine/threonine polymorphism within the N-terminal sequence of the Ss glycoprotein, *Hoppe-Seyler's Zeitschrift fur Physiologische Chemie* **361**, 895–906.

[44] Race, R.R. & Sanger, R. (1975). *Blood Groups in Man*, 6th Edition, Blackwell Scientific Publications, Oxford.

[45] Tills, D., Kopec, A.C., Tills, R.E. (1983). *The Distribution of the Human Blood Groups and Other Polymorphisms* Oxford University Press, Oxford, Suppl. 1.

[46] Landsteiner, K. & Wiener, A.S. (1940). An agglutinable factor in human blood recognized by immune sera for Rhesus blood, *Proceedings of the Society for Experimental Biology and Medicine* **43**, 233.

[47] Landsteiner, K. & Wiener, A.S. (1941). Studies on an agglutinogen (Rh) in human blood reacting with antirhesus sera and with human isoantibodies, *The Journal of Experimental Medicine* **74**, 309–320.

[48] Levine, P. & Stetson, R.E. (1939). An unusual case of intra-group agglutination, *JAMA* **113**, 126–127.

[49] Levine, P., Burnham, L., Katzin, E.M. & Vogel, P. (1941). The role of iso-immunization in the pathogenesis of erythroblastosis fetalis, *American Journal of Obstetrics and Gynecology* **42**, 925–937.

[50] Wiener, A.S. & Peters, H.R. (1940). Hemolytic reactions following transfusions of blood of the homologous group, with three cases in which the same agglutinogen was responsible, *Annales de Droit International Medical* **13**, 2306–2322.

[51] Levine, P., Celano, M.J., Wallace, J. & Sanger, R. (1963). A human 'D-like' antibody, *Nature* **198**, 596–597.

[52] Wiener, A.S. (1943). Genetic theory of the Rh blood types, *Proceedings of the Society for Experimental Biology and Medicine* **54**, 316–319.

[53] Race, R.R., Taylor, G.L., Cappell, D.F. & McFarlane, M.N. (1944). Recognition of a further common Rh genotype in man, *Nature* **153**, 52–53.

[54] Race, R.R. (1944). An 'incomplete' antibody in human serum, *Nature* **153**, 771–772.

[55] Wiener, A.S. & Wexler, I. (1963). *An Rh-Hr Syllabus*, 2nd Edition, Grune and Stratton, New York.

[56] Tippett, P. (1986). A speculative model for the Rh Blood groups, *Annals of Human Genetics* **50**, 241–247.

[57] Boorman, K.R., Dodd, B.E. & Lincoln, P.J. (1977). *Blood Group Serology: Theory, Techniques, Practical Applications*, 5th Edition, Churchill Livingstone, Edinburgh, London and New York.

[58] Stratton, F. (1946). A new Rh allelomorph, *Nature* **158**, 25–26.

[59] Agre, P.C., Davies, D.M., Issitt, P.D., Lamy, B.M., Schmidt, P.J., Tracy, M. & Vengelen-Tyler, V. (1992). A proposal to standardize terminology for weak D antigen, *Transfusion* **32**, 593–594.

[60] Kozbor, D., Roder, J.C., Chang, T.H., Steplewski, Z. & Koprowski, H. (1982). Human anti-tetanus toxoid monoclonal antibody secreted by EBV-transformed human B cells fused with murine myeloma, *Hybridoma* **1**, 323–328.

[61] Thompson, J.M., Lowe, J. & McDonald, D.F. (1994). Human monoclonal anti-D secreting heterohybridomas from peripheral B lymphocytes expanded in the CD40 system, *Journal of Immunological Methods* **175**, 137–140.

[62] Avent, N.D. (2000). Molecular biology of the Rh blood group polymorphisms, in *Human Blood Cells: Consequences of Genetic Polymorphisms and Variations*, M.-J. King, ed, Imperial College Press, London, pp. 105–147.

[63] Chérif -Zahar, B., Raynal, V. & Cartron, J.-P. (1997). RH gene structure: reassignment of two exon-exon junctions, *Blood* **89**, 4661–4662.

[64] Wagner, F.F. & Flegel, W.A. (2000). RHD gene deletion occurred in the *Rhesus Box*, *Blood* **95**, 3552–3558.

[65] Bennett, P.R., Le Van Kim, C., Colin, Y., Warwick, R.M., Chérif-Zahar, B., Fisk, N.M. & Cartron, J.P. (1993). Prenatal determination of fetal RhD type by DNA amplication, *The New England Journal of Medicine* **329**, 607–610.

[66] Wolter, L.C., Hyland, C.A. & Saul, A. (1993). Rhesus D genotyping using polymerase chain reaction, *Blood* **82**, 1682–1683.

[67] Simsek, S., Faas, B.H., Bleeker, P.M., Overbeeke, M.A., Cuijpers, H.T., van der Schoot, C.E. & von dem Borne, A.E. (1995). Rapid Rh D genotyping by polymerase chain reaction-based amplification of DNA, *Blood* **85**, 2975–2980.

[68] Flegel, W.A., Wagner, F.F., Müller, R.N. & Gassner, C. (1998). Rh phenotype prediction by DNA typing and its application to practice, *Transfusion Medicine* **8**, 281–302.

[69] Singleton, B.K., Green, C.A., Avent, N.D., Martin, P.F., Smart, E., Daka, A., Narter-Olaga, E.G., Hawthorne, L.M. & Daniels, G. (2000). The presence of an RHD pseudogene containing a 37 base pair duplication and a nonsense mutation in Africans with the Rh D negative blood group phenotype, *Blood* **95**, 12–18.

[70] Mouro, I., Colin, Y., Chérif-Zahar, B., Cartron, J.-P. & Le Van Kim, C. (1993). Molecular genetic basis of the human Rhesus blood group system, *Nature Genetics* **5**, 62–65.

[71] Inagaki, S., Yamamoto, Y., Doi, Y., Takata, T., Ishikawa, T., Imabayashi, K., Yoshitome, K., Miyaishi, S. & Ishizu, H. (2004). A new 39-plex analysis for SNPs including 15 blood group loci, *Forensic Science International* **144**, 45–57.

[72] Race, R.R., Mourant, A.E., Lawler, S.D. & Sanger, R. (1948). The Rh chromosome frequencies in England, *Blood* **3**, 689–695.

[73] Sistonen, P., Sareneva, H., Pirkola, A. & Eklund, J. (1994). MAR, a novel high-incidence Rh antigen revealing the existence of an allelic sub-system including C^w (Rh8) and C^x (Rh9) with exceptional distribution in the Finnish population, *Vox Sanguinis* **66**, 287–292.

[74] Mouro, I., Colin, Y., Sistonen, P., Le Pennec, P.T., Cartron, J.P. & Le Van Kim, C. (1995). Molecular basis of the RhC^w (Rh8) and RhC^x (Rh9) blood group specificities, *Blood* **6**, 1196–1201.

[75] Faas, B.H.W., Simsek, S., Bleeker, P.M., Overbeeke, M.A., Cuijpers, H.T., von dem Borne, A.E. & van der Schoot, C.E. (1995). Rh E/e genotyping by allele-specific primer amplification, *Blood* **85**, 829–832.

[76] Le Van Kim, C., Mouro, I., Brossard, Y.Y., Chavinié, J.P. & Colin, Y. (1994). PCR-based determination of Rhc and RhE status of foetuses at risk of Rhc and RhE haemolytic disease, *British Journal of Haematology* **88**, 193–195.

[77] Lawler, S.D., Bertinshaw, D., Sanger, R. & Race, R.R. (1950). Inheritance of the Rh blood groups: 150 families tested with anti-C-c- C^w -D-E and anti-e, *Annals of Eugenics* **15**, 258–270.

[78] Spencer, N., Hopkinson, D.A. & Harris, H. (1968). Adenosine deaminase polymorphism in man, *Annals of Human Genetics* **32**, 9.

[79] Hirschhorn, R., Yang, D.R. & Israni, A. (1994). An Asp8Asn substitution results in the adenosine deaminase (ADA) genetic polymorphisms (ADA 2 allozyme): occurrence on different chromosomal background and apparent intragenic crossover, *Annals of Human Genetics* **58**, 1–9.

[80] Fildes, R.A. & Harris, J. (1966). Genetically determined variation of adenylate kinase in man, *Nature* **209**, 261–262.

[81] Matsuura, S., Igarashi, M., Tanizawa, Y., Yamada, M., Kishi, F., Kajii, T., Fujii, H., Miwa, S., Sakurai, M. & Nakazawa, A. (1989). Human adenylate kinase deficiency associated with haemolytic anemia: a single base substitution affecting solubility and catalytic activity of the cytosolic adenylate kinase, *The Journal of Biological Chemistry* **264**, 10148–10155.

[82] Robson, E.B., Harris, H. & Smith, S.M. (1967). Data on the incidence, segregation and linkage relations of the adenylate kinase (AK) polymorphism, *Annals of Human Genetics* **31**, 237–242.

[83] Hopkinson, D.A., Spencer, N. & Harris, H. (1963). Red cell acid phosphatase variants: a new human polymorphism, *Nature* **199**, 969–971.

[84] Sorgo, G. & Brinkmann, B. (1974). Inheritance of the allele P^r of the red cell acid phosphatase in a Caucasian family, *Humangenetik* **24**, 155.

[85] Dissing, J., Johnson, A.H. & Sensabaugh, G.F. (1991). Human red cell acid phosphatase (ACP1): the amino acid sequence of the two isozymes Bf and Bs encoded by the ACP1*B allele, *The Journal of Biological Chemistry* **266**, 20619–20625.

[86] Bryson, G.L.M., Massa, H., Trask, B.J. & Van Etten, R.L. (1995). Gene structure, sequence and chromosomal localization of the human red cell-type low molecular weight acid phosphotyrosyl phosphatase gene, ACP1, *Genomics* **30**, 133–140.

[87] Dissing, J., Thymann, M. & Hopkinson, D. (2002). Simultaneous detection of ACP1 and Gc genotypes using PCR/SSCP, *Annals of Human Genetics* **67**, 81–85.

[88] Hopkinson, D.A., Mestriner, M.A., Cortner, J. & Harris, M. (1973). Esterase D: a new human polymorphism, *Annals of Human Genetics* **37**, 119–137.

[89] Tsuchida, S., Fukui, E. & Ikemoto, S. (1994). Molecular analysis of esterase D polymorphism, *Human Genetics* **93**, 255–258.

[90] Parr, C., Bagster, I.A. & Welch, S.G. (1977). Human red cell glyoxalase I polymorphism, *Biochemical Genetics* **15**, 109–114.

[91] Gale, C.P. & Grant, P.J. (2004). The characterisation and functional analysis of the human glyoxalase – a gene using methods of bioinformatics, *Gene* **340**, 251–260.

[92] Hansen, H.E. & Eriksen, B. (1979). HLA-GLA linkage analysis in 57 informative families, *Human Heredity* **29**, 355–360.

[93] Junaid, M., Kowal, D., Barua, M., Pullarkat, P., Brooks, S.S. & Pullarkat, R. (2004). Proteomic studies identified a single nucleotide polymorphism in glyoxylase I as autism susceptibility factor, *American Journal of Medical Genetics* **131a**, 11–17.

[94] Parrington, J.M., Cruickshank, G., Hopkinson, D.A., Robson, E.B. & Harris, H. (1968). Linkage relationships between the three phosphoglucomutase loci PGM(1), PGM(2) and PGM(3), *Annals of Human Genetics* **32**, 27–34.

[95] Hopkinson, D.A. & Harris, H. (1966). Rare phosphoglucomutase phenotypes, *Annals of Human Genetics* **30**, 167–181.
[96] Dykes, D.D., Kühnl, P. & Martin, W. (1985). PGM1 system: report on the International Workshop, October 10–11, 1983, Munich, West Germany, *American Journal of Human Genetics* **37**, 1225–1231.
[97] Tipler, T.D., Dunn, D.S. & Jenkins, T. (1982). Phosphoglumutase first locus polymorphism as revealed by isoelectric focusing in Southern Africa, *Human Heredity* **32**, 80–93.
[98] Yip, S.P., Lovegrove, J.U., Rana, N.A., Hopkinson, D.A. & Whitehouse, D.B. (1999). Mapping recombination hotspots in human phosphoglucomutase (PGM1), *Human Molecular Genetics* **8**, 1699–1706.
[99] Hirschfeld, J. (1962). The Gc-system: immune-electrophoretic studies of normal human sera with special reference to a new genetically determined serum system (GC), *Progress in Allergy* **6**, 155–86.
[100] Bearn, A.G., Bowman, B.H. & Kitchin, F.D. (1964). Genetic and biochemical consideration of the serum group-specific component, *Cold Spring Harbor Symposia on Quantitative Biology* **29**, 435–442.
[101] Svasti, J., Kurosky, A., Bennett, A. & Bowman, B.H. (1979). Molecular basis for the three major forms of human serum vitamin D binding protein (group-specific component), *Biochemistry* **18**, 1611–1617.
[102] Kofler, A., Braun, A., Jenkins, T., Serjeantson, S.W. & Cleve, H. (1995). Characterization of mutants of the vitamin-D-binding protein / group specific components: GC Aborigine (1A1) from Australian Aborigines and South African blacks, and 2A9 from South Germany, *Vox Sanguinis* **68**, 50–54.
[103] Daiger, S.P., Schanfield, M.S. & Cavalli-Sforza, L.L. (1975). Group-specific component (Gc) proteins bind vitamin D and 25-hydroxyvitamin D, *Proceedings of the National Academy of Sciences of the United States of America* **72**, 2076–280.
[104] Mourant, A.E., Tills, D. & Domaniewska-Sobczak, K. (1976). Sunshine and the geographical distribution of the alleles of the Gc system of plasma proteins, *Human Genetics* **33**, 307–314.
[105] Kueppers, F. & Harpel, B. (1979). Group-specific components (Gc) 'subtypes' of Gc1 by isoelectric focussing in US blacks and whites, *Human Heredity* **29**, 242–249.
[106] Witke, F.W., Gibbs, P.E.M., Zielinski, R., Yang, F., Bowman, B.H. & Dugaiczyk, A. (1993). Complete structure of the human Gc gene: differences and similarities between members of the albumin gene family, *Genomics* **15**, 751–754.
[107] Oliviero, S., Morrone, G. & Cortese, R. (1987). The human haptoglobin gene: transcriptional regulation during development and acute phase induction, *The EMBO Journal* **6**, 1905–1912.
[108] Smithies, O. (1955). Zone electrophoresis in starch gels: group variations in the serum proteins of normal human adults, *The Biochemical Journal* **61**, 629–641.
[109] Maeda, N. (1991). DNA polymorphisms in the controlling region of the human haptoglobin genes: a molecular explanation for the haptoglobin 2–1 modified phenotype, *American Journal of Human Genetics* **48**, 158–166.
[110] Teye, K., Quaye, I.K., Koda, Y., Soejima, M., Tsuneoka, M., Pang, H., Ekem, I., Amoah, A.G., Adjei, A. & Kimura, H. (2003). A-61C and C-181G Hp gene promoter polymorphisms are, respectively, associated with ahaptoglobinaemia and hypohaptoglbulinaemia in Ghana, *Clinical Genetics* **64**, 439–443.
[111] Koch, W., Latz, W., Eichinger, M., Roguin, A., Ley, A.P., Schomig, A. & Kastrati, A. (2002). Genotyping of the common haptoglobin Hp 1/2 polymorphism based on PCR, *Clinical Chemistry* **48**, 1377–1382.

DENISE S. COURT

Bloodstain Pattern Interpretation

Introduction

"All criminal investigation is concerned either with *people* or with *things*. Only people commit crimes, but invariably do so through the medium of things. It is these things that together constitute the broad field of physical evidence" [1]. Bloodstain patterns at crime scenes are often highly significant in allowing an accurate reconstruction of what prior events must have occurred to have produced them.

Dr Kirk, a true pioneer in scientific crime detection, was the inspiration for an early research in the geometric interpretation of bloodstain patterns. The report, *Flight Characteristics and Stain Patterns of Human Blood* [2], was the result of a Department of Justice funded research project on the topic. It also was the first publication on this subject in the western hemisphere. As a result of that report interest in, and acceptance of, bloodstain patterns, as evidence grew very rapidly.

During the last three decades, bloodstain evidence has become accepted as an important tool by the forensic community for the investigation of crimes wherein blood has been shed. Perhaps even more importantly, bloodstain evidence may be

used to exculpate an innocent suspect or defendant. Identification or elimination of an individual by DNA has become a routine event that is reported in the media almost on a daily basis. Eyewitness identification and other personal accounts of what they believed happened can be, and frequently are, inaccurate. The scientific reliability of physical evidence, in this case bloodstain pattern interpretation, is more accurate in reconstruction of prior events than evidence based upon the memory of an eye witness. However, bloodstain patterns, like other physical evidence, must first be detected, preserved, examined, interpreted, and properly presented in court to utilize its evidentiary value.

A Brief History of Bloodstain Pattern Interpretation

In 1971, this author was only able to locate 15 references on the subject of bloodstain pattern interpretation. However, considerably more research had been conducted in this discipline than was originally believed. In fact, well over 400 references on this subject have been found that were published before 1970; most are in forensic, academic, and medical libraries throughout the world.

In October 1971, this author wrote "Although it would seem that prior investigations of physical evidence would have resulted in a wealth of knowledge regarding the significance of bloodstains, such is not the case." This statement appears in the introduction of *Flight Characteristics and Stain Patterns of Human Blood*. Today, some 38 years later, it is clear that this statement was incorrect.

For those who would like to read a detailed book on the early literature in this discipline they could review *Segments of History, The Literature of Bloodstain Pattern Interpretation, Segment 00: Literature Through the 1800's* [3]. This reference includes reports of a murder case that occurred before 100 AD, another in London in 1514, the murder of his wife by a minister in Upstate New York in 1859, and several others prior to 1900.

The outstanding research of Dr Eduard Piotrowski [4] in 1895 was undoubtedly one of the most comprehensive works of the nineteenth century. He was involved in a beating death and conducted many experiments using rabbits, which he documented in great detail. His book included 22 color plates, which illustrated his experiments.

Unfortunately, it is not possible to include a synopsis of the classic works of Gross [5], Orsos [6], and Balthazard [7] in this article. Suffice it to say that many early forensic scientists recognized the potential value of bloodstain patterns but, for whatever reason, this subject was essentially never practiced in the United States until Dr Paul Leland Kirk became involved with the Dr Samuel Sheppard case in 1955. His affidavit [8] on bloodstains in that case appears to be the first in-depth use of evidence of this type in the reconstruction of a homicide.

Serious students seeking additional information on this subject will find some 450 references on bloodstain patterns in the first issue of *Quinnipiac Health Law* [9].

Characteristics of Liquid Blood

In order to understand why blood behaves as it does, one must have a basic understanding of two important properties of human blood; its surface tension and its viscosity. These properties of blood are sufficiently similar to those of water that it is unnecessary for the reader to formulate an entirely new concept to deal with blood as a liquid. Basically, it is within the common experience that everyone has had with water insofar as how it pours, splashes, and spatters. However, water is colorless and, therefore, it makes the casual observations of these phenomena more difficult than with blood. Bloodstains will always have a greater contrast over a light-colored background surface than water does.

Perhaps the most important physical property of blood that should be understood by the forensic investigator is its surface tension. To a large degree, this property explains why blood behaves as it does. Like all liquids, blood is held together by cohesive forces between its molecules called *Van der Waals force*. On the surface of a glass of water, or a lake, this intermolecular attraction produces a skin-like surface that is resistant to penetration or separation. Technically, this phenomenon is known as the *surface tension of the liquid*. It is this skin-like property of water that allows water bugs to walk across a pond. Likewise, if a razor blade or a small needle is carefully lowered onto the surface of water, neither will sink even though steel is 7.8 times denser than water. They are not "floating" in accordance with Archimedes' Principle, but rather they are supported by water's surface tension.

Surface tension across the surface of a liquid decreases its surface area. When small volumes of a liquid become detached from the main body of the liquid, such as from a medicine dropper, they assume a spherical geometry because such configuration minimizes surface area. The so-called teardrop shape, so often seen in the artist's rendition of falling raindrops, simply has not, does not, nor ever can exist in nature. Blood drops, like all other liquids, have a spheroid or ball shape, and not the "teardrop" shape. In his amazing book, *Flash*, Dr Harold Edgerton shows high-speed photography to show how drops are formed as their shape in free fall [10]. Figure 1 shows four such photographs.

The study of bloodstains is primarily concerned with what happens to blood once it has left the body. When blood spatters, drips, or is gushing from an artery, it must follow the laws of physics, specifically that of ballistics, which is the study of projectiles in motion. Therefore, interpretation of the significance of bloodstain evidence should be conducted by a physical scientist rather than a pathologist whose expertise is generally limited to the behavior of blood within the body. However, they should recognize the more common bloodstain patterns that can be present within the body of a deceased person.

Basic Pattern Principles

The study of bloodstain patterns must be directed toward the understanding of several basic pattern

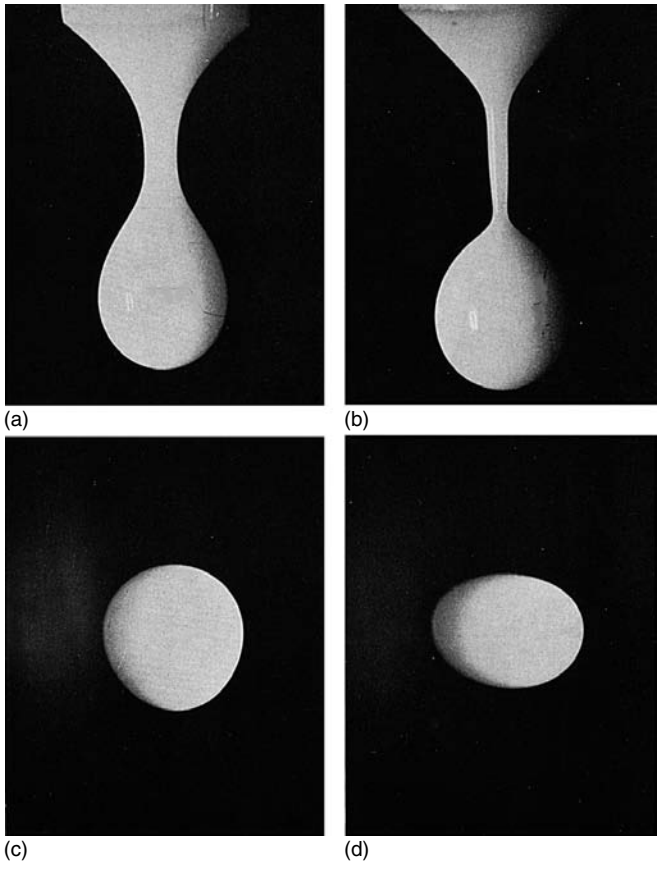


Figure 1 The formation of a drop of milk recorded by stroboscopic photography. (a) Weight of the forming drop begins narrowing the connecting source. (b) Surface tension still holds the drop before the drop falls free. (c) In free fall the drop immediately forms a spheroid or ball. (d) Because of its fluid nature and relatively large volume, the drop will oscillate in free fall

types and how they can be produced. These are discussed in detail later but, by way of an introduction to the significance of bloodstain patterns, the more significant considerations are listed below in a general order of decreasing significance:

1. The shape of the individual bloodstains, which may allow their origin to be determined in three dimensions. Under certain circumstances, this category can also include some transfer patterns.
2. The size of the individual bloodstains, which suggests the kind of energy that was available for their production if they were the result of an impact.
3. The distribution and concentration of a bloodstain pattern may suggest the distance between the origin of blood and the surface upon which it was deposited. This can be important in differentiating among impact spatter, expirated blood, and arterial patterns.

Blood Patterns within the Human Body

A frequent observation of a bloodstain pattern within the human body is the recognition of blanched or void patterns, which appear as the result of external pressure on compressed areas of the body, such as the back, arms, or legs. Such pressure prevents the uniform development of livor mortis in those areas and often can outline a common geometric shape that may be easily identifiable. When such patterns are observed, they should be immediately photographed since such images may fade quite rapidly depending upon how long it has been since the livor mortis developed.

Surface Texture Influence on the Degree of Bloodspatter

The degree of spatter that results when a drop of blood falls onto a surface is far more dependent upon the nature of the surface than it is on the distance the drop has fallen before it impacted the surface. Generally speaking, the harder and less porous the surface, the less spatter will occur. This is due to the fact that, even though the drop undergoes considerable geometric distortion upon impact because it spreads out and then contracts, the shape of the drop remains intact because of blood's high surface tension. Protrusions and rough texture of irregular and porous surfaces can rupture surface tension and result in increased spatter.

When free falling, the surface tension of a drop of blood is not unlike a rubber balloon filled with water. It can be "teased about" over a smooth surface to great degrees of geometric distortion, but the same balloon will likely rupture and discharge its contents if abraded in contact over a coarse surface such as sandpaper.

There are also distinctions that can be drawn between hard and soft porous surfaces. Raw wood and asbestos board are certainly porous, but they are also hard in contrast to newspapers or paper towels. Consequently, there will be fewer spatters on the raw wood and asbestos board. Likewise, irregularity of the target surface is an important factor that should not be overlooked. The two sides of a sheet of singularly corrugated cardboard may be of the same composition, but they usually differ markedly in texture.

Estimations of the distance of a drop of blood that has fallen from the extent of satellite spatter around the central bloodstain, or from "spines" that may also be produced, will usually not be accurate at all if surface texture of the target is not considered. Spines are the pointed streaks of blood that radiate away from the center of a bloodstain. As an example, the bloodstain shown in Figure 2, was produced by a drop of blood that fell over 80 ft before it struck a hard, smooth, glossy cardboard surface but no spatter whatever resulted. While it might seem that blood ought to spatter upon impact after falling so far, it does not. Surface tension of blood, and the fact that it reaches a terminal velocity when air resistance and gravitational attraction become balanced, prevents a drop from spattering *regardless of the height it falls before striking a smooth, hard surface*. In contrast, the bloodstain shown in Figure 3 was produced when a drop of blood fell only 18 in. before striking a desk blotter.

From Figures 2 and 3, it is obvious that correct interpretation of spattering in bloodstain patterns requires consideration of the surface texture of the target upon which it impacted.

Bloodstain Characteristics: Shape of Stain

The shape of bloodstains produced when individual drops of blood strike a surface, whether simultaneously or sequentially, will be either round or

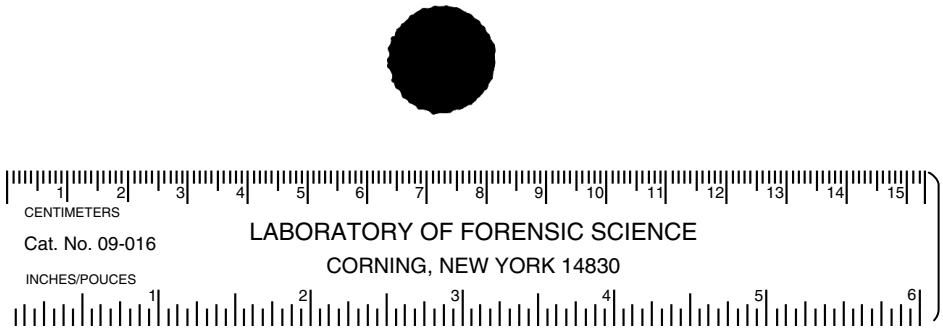


Figure 2 The bloodstain produced when a typical 0.05-ml drop of human blood fell 80 ft before striking a hard, smooth, glossy cardboard surface

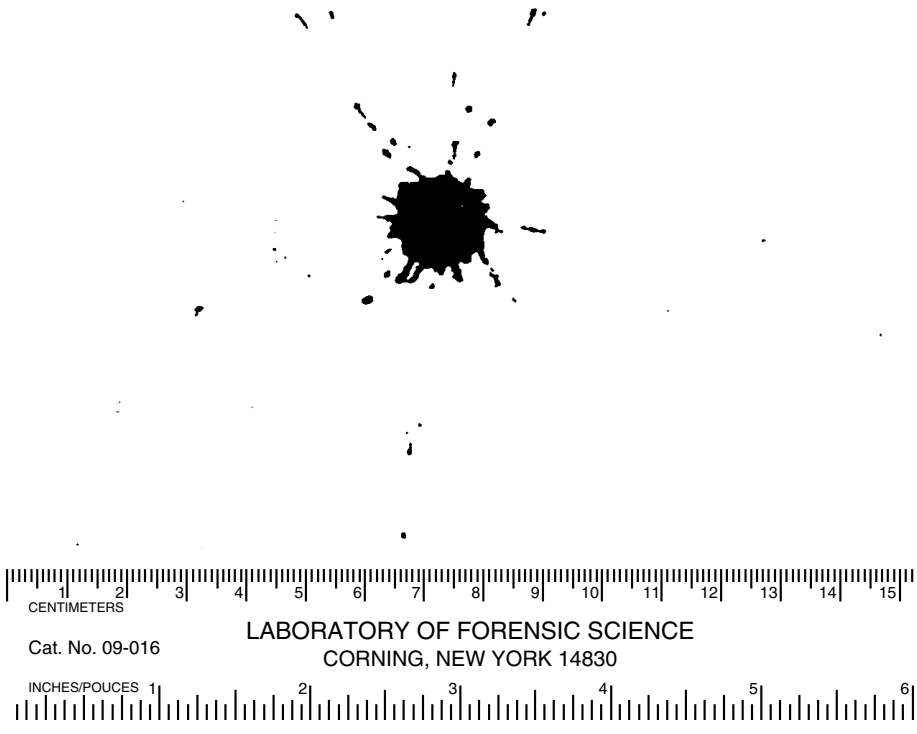


Figure 3 The bloodstain produced when a typical 0.05-ml drop of human blood fell 18 in. before it struck a soft desk blotter. Note that extensive spatter resulted when the blotter's fibers ruptured its surface tension

elliptical. The horizontal direction these blood drops were traveling prior to impact may often be determined from the shape of the resulting bloodstains. Establishing such directionality is possible even when there are no edge scallops, spines, or satellite spatters. Measurements are not necessary to determine the direction of flight of a drop of blood immediately prior to an angular impact onto a flat surface when one end of the ellipse displays a point, which looks like the artist's incorrect rendition of a tear drop. This point indicates its forward direction of travel prior to impact. The bloodstains shown in Figure 4 resulted from droplets that were traveling from left to right prior to their striking a flat horizontal surface.

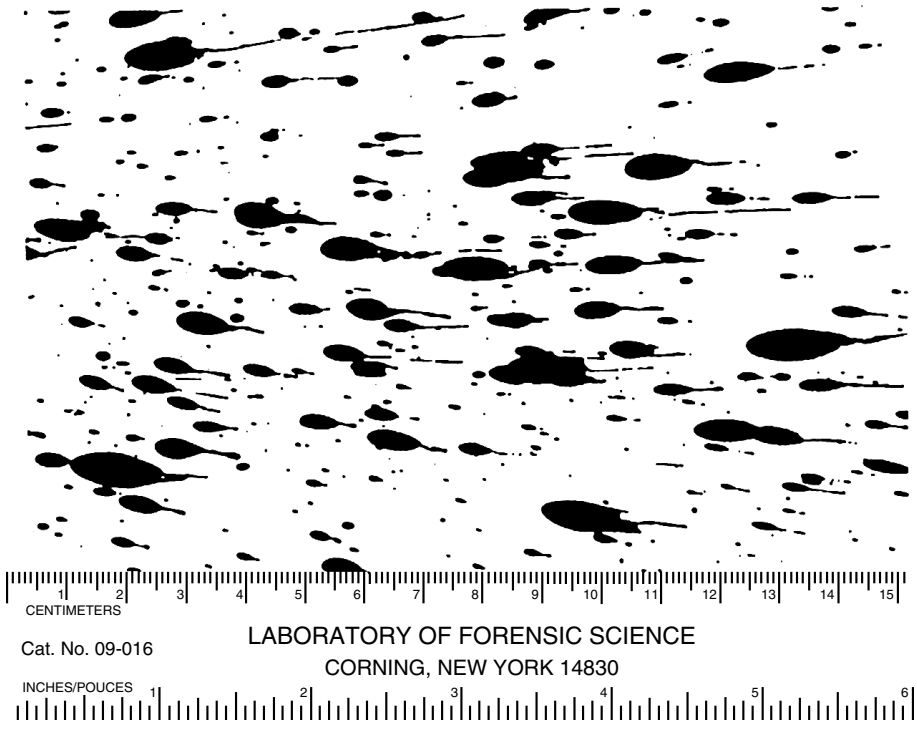


Figure 4 This pattern resulted from many droplets of blood that were traveling from left to right prior to their striking a flat level surface

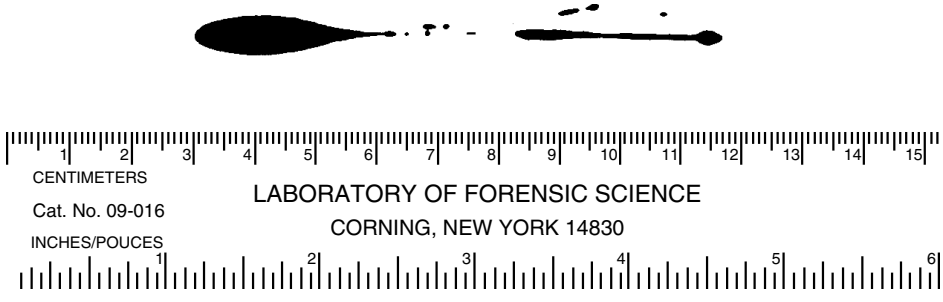


Figure 5 The parent bloodstain is shown on the left and the smaller wave cast-off is on the right. The blood drop that produced this pattern was traveling from the left to right before it struck a horizontal smooth cardboard target at an angle of approximately 16°

A smaller droplet may sometimes be thrown from a larger parent drop upon impact to produce a "wave cast-off". An example of this is shown in Figure 5 where the parent drop is on the left and the wave cast-off is on the right. The obvious direction of travel of the blood drop prior to its impact on the surface was from left to right.

The two-dimensional origin of a bloodstain pattern may be established by drawing straight lines through the long axis of the bloodstains. These lines represent the direction of each blood drop's trajectory prior to its striking the surface. The intersection of these lines represents the origin of bloodspatter in a two-dimensional configuration as shown in Figure 6. The

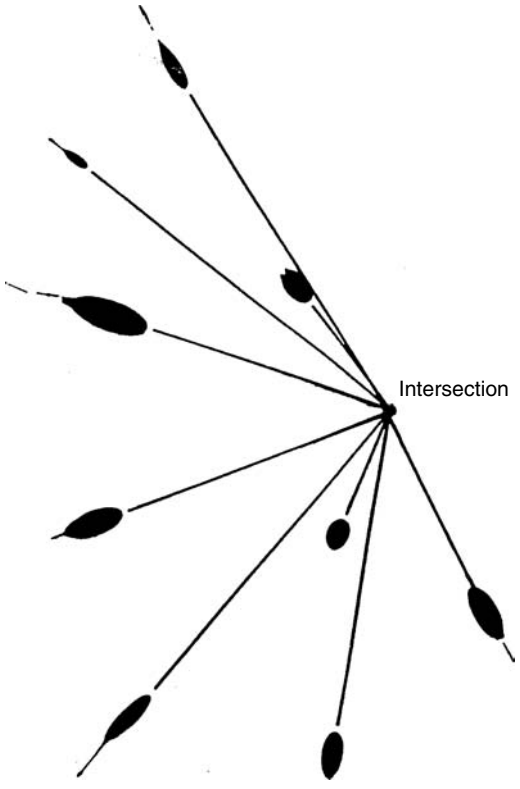


Figure 6 The origin of bloodspatter on a surface can be determined by establishing a convergence of lines drawn through the long dimension of several bloodstains. This shows a two-dimensional origin

actual origin of blood spatter would have been at some point above this intersection.

Tracing bloodstain directionalities on flat surfaces to find their general area of convergence does not provide sufficient information to determine the height of the origin of bloodspatter above that surface. To achieve this, the impact angle of several individual bloodstains must be calculated and projected back above their convergence on the surface, thereby constructing a three-dimensional model.

Estimation of the height above or away from a surface may usually be established by calculation of the impact angle of each bloodstain and projecting their trajectories back to an axis that is placed 90°, or normal, to the intersection of the previously established two-dimensional origin. This is discussed in greater detail in the next section.

Impact Angle Considerations

Blood that falls onto a flat surface that is not horizontal will produce an elliptical rather than a round bloodstain. Its degree of distortion from a circle is inversely proportional to the impact angle. That is, as the impact angle decreases from 90° to 10°, the bloodstain's shape becomes progressively elongated. Figure 7 shows nine bloodstains that resulted when single drops of blood struck a smooth surface at impact angles of 10, 20, 30, 40, 50, 60, 70, 80, and 90°, respectively.

There is a mathematical relationship between the shape of the bloodstain and the angle the blood drop that produced it struck the surface. Therefore, the impact angles may be determined quite accurately using the bloodstain's width-to-length ratio. This ratio has a trigonometric relationship to the angle of impact. The calculation is elementary and uses standard trigonometric tables. Simply divide the width of an elliptical bloodstain by its length. The result will always be one or less and is the sine of the impact angle. The formula to use for this calculation is:

$$\text{arc sine of } \frac{\text{width of bloodstain}}{\text{length of bloodstain}} = \text{impact angle}$$

Accuracy in the determination of impact angle depends on the care that is taken in the measurement of the length and width of each bloodstain. However, since the fluid nature of blood does not produce a perfect ellipse when it impacts a surface at an angle other than 90°, the pointed end of the bloodstain must be "rounded off" so it is symmetrical with its opposite end as may be seen in Figure 8. The more acute the angle of impact, the more pointed the forward end of the ellipse will be. It points in the direction of its travel prior to impact.

When evaluating bloodstains of <5 mm on their long dimension measurements should be made with some type of 20× pocket microscope which has a built-in reticule. Regardless of their size, only bloodstains that are well defined should be measured.

By determining the angles of impact for several bloodstains in conjunction with their distances from the origin (on a two-dimensional plane), it is possible to estimate that the distance of the origin of spatter was from the stained surface. This origin would be either the height above a floor or the distance from a wall or ceiling. A three-dimensional model may be

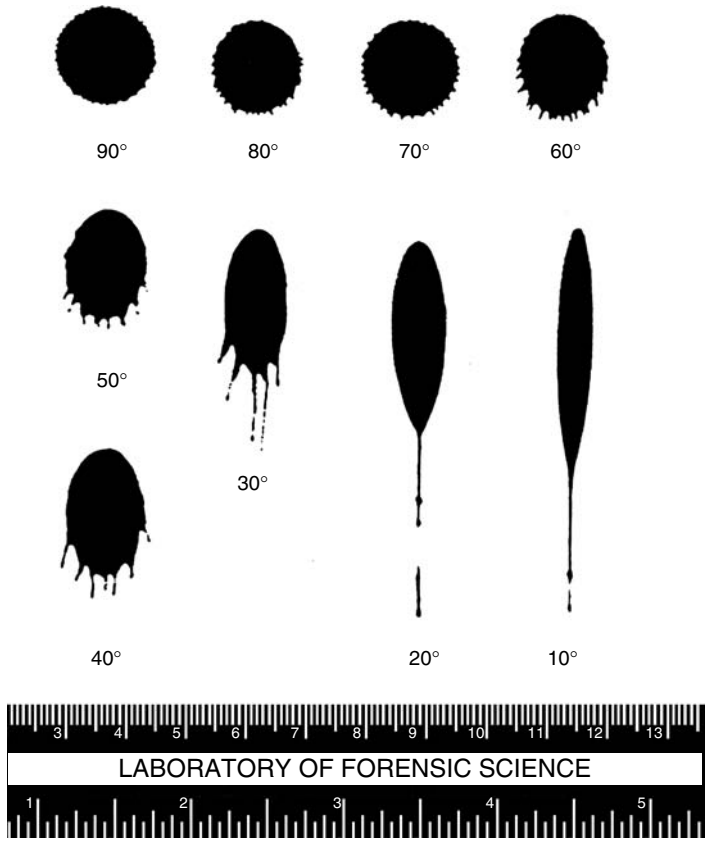


Figure 7 Patterns produced when typical drops of blood that fell 42 in. onto hard, smooth, cardboard at the indicated impact angles

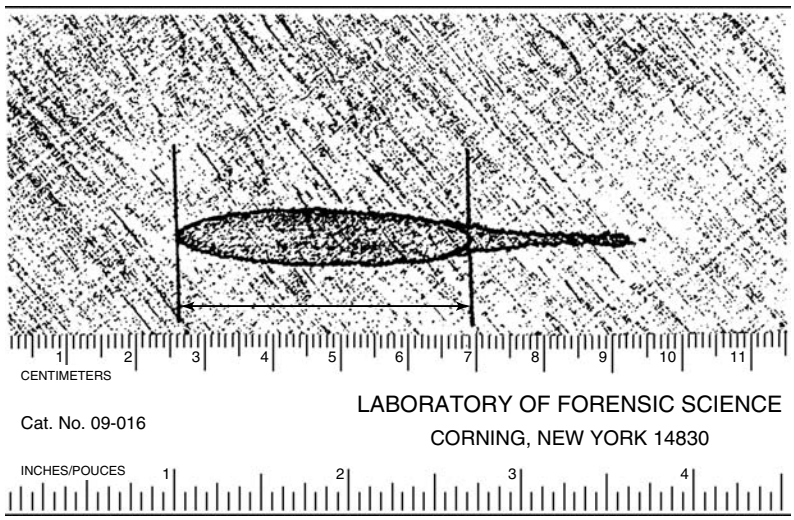


Figure 8 A curved line is traced over the bloodstain to identify the length of the ellipse at its pointed end

constructed at the scene from bloodstains using string, tape, and a protractor; or it may be done back at the laboratory from measurements taken at the scene. Both the location and geometry of several bloodstains are necessary to prepare an accurate model or graph to identify the origin.

It must be emphasized that when measurements and angles are used to establish the origin or origins in space, not only will the actual origin be somewhere below the point or points of convergence but it must also be remembered that the investigator is actually determining a space and not a point of origin. Since the practical application of establishing this space, which is more accurately identified as a volume, it is only establishing whether the victim was standing, sitting, or kneeling, on or just above the floor, in a chair or on a bed; pinpoint accuracy is not necessary nor can it ever be obtained. Wounds are never a pinpoint size.

Computer programs have been developed that can calculate the exact origin of spatter from bloodstains on the walls, floor, or ceiling of crime scenes. While this is a practical application of computer technology, such precise measurements will seldom be necessary. If the origin of bloodspatter can be established within the volume of a grapefruit, or even a basketball, and that origin is either at a standing height of 5–6 ft above the floor, at a level consistent with sitting on the piece of furniture in question, or lying on the floor, a satisfactory reconstruction will be achieved in all but the most unusual circumstances.

Bloodstain Characteristics: Size of Stain

Most information on this subject frequently neglected two essential considerations in their estimation of the distance a drop of blood had fallen before it struck a surface. These are the volume of the drop of blood and the nature of the target surface.

There is a very wide range of "wettability" of surfaces. Soft porous surfaces, such as a piece of cloth, will often have a large bloodstain that results from liquid blood soaking into the fabric. Conversely, blood that falls onto glass, or some other nonporous surface, and does not "wet" the surface, will result in smaller bloodstains as it contracts before it dries because of its internal cohesive forces. The result is that a smaller bloodstain will be produced than would have been the case had the same drop of blood fallen the same distance onto a tile floor, a wooden surface or a tight pile carpet. The cohesive property of blood may be easily and dramatically demonstrated by simply letting drops of blood fall onto a smooth, hard surface such as the top of a coffee can. Rub a very thin layer of oil or grease onto one half of the can top and leave the other half clean. Allow a drop of blood to fall 3 ft or so to one surface at a time. The drop that strikes the "greased surface" will immediately begin contracting until it is approximately one-half its original diameter. The drop that strikes the clean side will not contract at all because it "wets" the surface. This is an excellent way to demonstrate the terms "*wettability*" and "*surface tension*" to a lay jury: a "jury friendly" exercise, which helps them understand why many bloodstain patterns are produced as they are.

The resulting edge characteristics of a bloodstain are far more dependent upon the texture of the target surface upon which the drop lands than they are upon the distance it fell before impact. Therefore, conclusions as to the significance of "spines" or spatters as a function of distance fallen cannot be reliable unless texture of the target surface is considered. Bloodstains on glass will not spatter regardless of the distance they fell before impact, but bloodstains on wood or paper spatter to various degrees because they have wide variations in their surface texture.

Provided the volume of a drop of blood was known to be that of the 0.05-ml drop, which is typical, it is possible to estimate the distance it fell prior to impact. However, in almost 50 years of study this author has yet to find a practical application for such information.

Blood Spattered from Low-Velocity Impact (Dripped and Splashed Blood)

Blood dripping into blood produces a characteristic stain pattern. Liquid-to-liquid impact of this type results in many very small droplets being spattered upward and out away from the center of the blood pool that is being formed. This action is somewhat like a water fountain wherein small droplets are projected up and away from its center and they return to the surface from whence they came at obtuse angles. A typical drip pattern is shown in Figure 9.

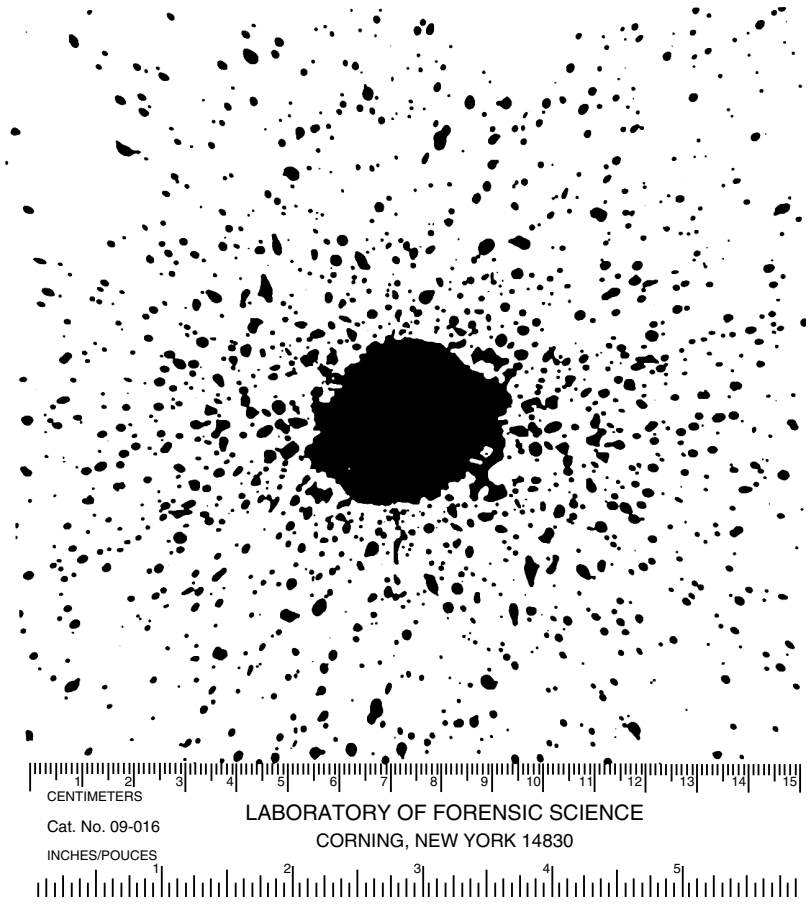


Figure 9 A typical drip pattern produced as blood fell into blood. Several drops of blood were released to form this low-velocity impact pattern

Blood acted upon by a low-energy, low-velocity force, such as when someone steps into a small pool of blood on a floor, will be splashed away from the point of impact. The drops that are produced will strike the surrounding area at very acute angles that are more streaks than ellipses. The resulting bloodstain pattern will usually have these long streaks radiating out away from the original source of blood as shown in Figure 10. Blood cannot have an upward trajectory from impact by the underside surface of a flat object such as a shoe. As a result the splashed blood can only skim along the surface before the individual droplets streak to a halt. Splashed bloodstain patterns are most often formed when larger volumes of blood are available for impact of this type.

Blood Spattered from Medium-Velocity Impact

During a beating, a stabbing, or other violent act it is common for blood to be shed. As blood accumulates on the surface of the victim's clothing or bare skin, if additional blows are made to the same area blood will usually be spattered. The force of the blows, with or without a knife blade, will usually have sufficient energy to overcome blood's surface tension releasing hundreds of very small droplets. The resulting spatter pattern is described as medium-velocity impact spatter. It is important to understand that the velocity of the weapon is usually higher than the velocity of blood, which leaves the main blood source, usually the victim's body. The preponderance

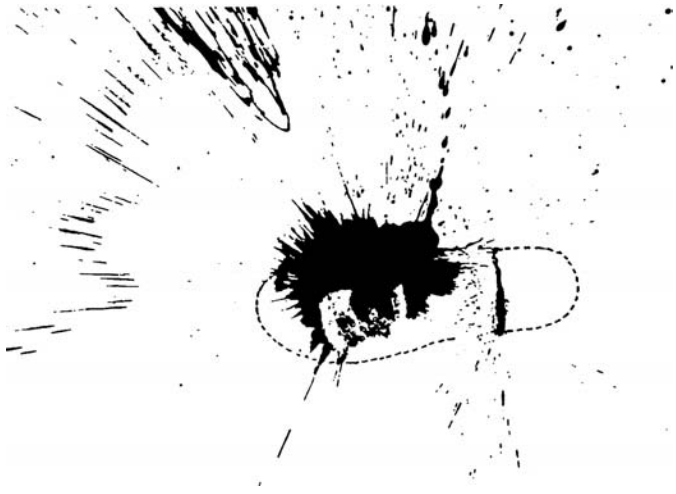


Figure 10 Low velocity, surface splashing caused by stepping into a pool of blood

of bloodstains resulting from medium-velocity impact spatter have diameters of 1 mm or larger as shown in Figure 11. It must be noted that while the diameter of most bloodstains produced by these mechanisms is well over this dimension, some are also much smaller.

Homicides due to beating are usually quite characteristic because of the large amount of blood that spatters. It is usually more than from a stabbing, cutting, or shooting. Exceptions, of course, do exist. If a broad axe were used as an instrument for delivering multiple blows or if a baseball bat repeatedly struck a victim's head in the same location the amount of bloodspatter would be considerable. In fact, multiple blows being struck to the head will create more bloodspatter than from any other common mechanism except gunshot wounds.

Those who administer beatings will often become spattered with blood. The absence of bloodspatter on the accused's clothing does not, however, prove nonparticipation as many defense lawyers would like to believe. To the contrary, it is possible to beat someone and not receive any backspatter whatsoever.

Many earlier researchers conducted beating experiments and reported that they found that blood was not always backspattered on garments worn by the perpetrator even though considerable blood was spattered elsewhere [4–6, 11–14]. Some of them pointed out that assailants could also have worn outer protective garments or had simply changed their clothes after they beat their victim. In many cases, it has also been reported that the perpetrator was nude when committing the crime and afterward simply showered and got dressed so they had no blood on their body or clothing.

Scientists should concern themselves with what they "see" rather than speculate as to why they do not see something they think they should. Nevertheless, when blood spots in the medium-velocity impact range are discovered on a garment, they are usually somewhat below the resolution of the average juror's eyesight. Some simple means of providing a "road map" identifying the location of these small spots of blood should be considered so a jury can understand not only their location but also their size and distribution. One convenient method is simply to place small white three ring binder reinforcers around each of the bloodstains and then to photograph the garment. Reinforcers should never be glued onto the garment but just placed around the bloodstain long enough to photograph the garment. An example of this technique is illustrated in Figures 12 and 13. In this case, a farm laborer beat his employer's wife to death with a hammer and a wrench. His work shirt, shown in Figure 12, was spattered with over 250 small bloodstains. Note the blood spots over his shirt pocket flap. Figure 13, shows the same shirt with the flap over his shirt pocket raised up. The presence of bloodspatter beneath the flap is evidence that the flap portion of his shirt was moving up and down at the time it was spattered with blood, a movement completely consistent with raising his arm up and down as in a beating.

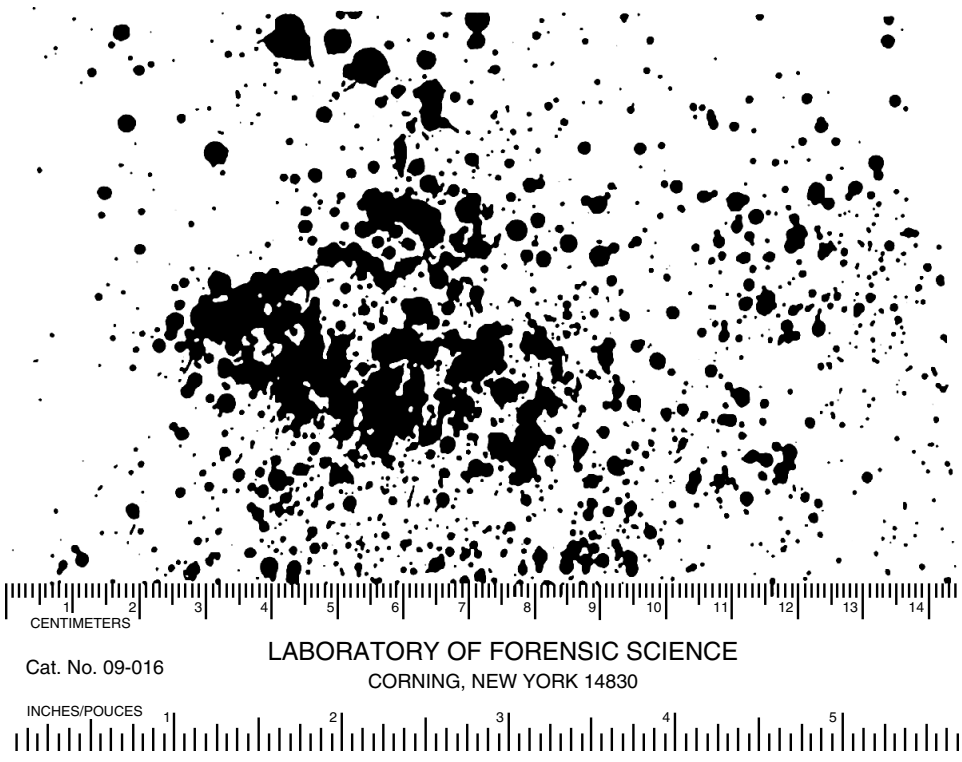


Figure 11 Typical medium-velocity impact spatter. Although the majority of these stains are over 1 mm in diameter, many are also much smaller

Another, perhaps more popular method of photographing the location of small bloodstains is to cut small, sharp triangles from masking tape and simply press them onto the garment so that their most acute angle points to the individual bloodstains. The blue jeans in Figure 14 show an example of this procedure. The triangular pieces of masking tape could be numbered for specific identification of each bloodstain, if desired. Surprisingly, masking tape adheres quite tenaciously. Garments that were marked in this manner had the tape still sticking to them several years after a trial even though they had undergone considerable handling.

Impact velocity associated with beatings and stabbings has been measured about 25 ft s^{-1}. Therefore, medium-velocity impact spatter is considered to be spatter that results from an impact to a blood source in this range. Extreme situations, of course, must always be considered. For example, a golf club may be swung at a velocity of over 75 ft s^{-1}. Fortunately, beatings with a golf club are rare but when such an uncommon event is encountered, it must be given proper consideration when interpreting bloodstain patterns that result. The highest beating velocities cannot approach the velocity of the slowest projectiles fired from handguns, rifles, or shotguns. For this reason, differentiation between bloodstain patterns that result from these two mechanisms is easily recognized by investigators who are familiar with evidence of this type.

Although very small bloodstains can only be produced when some form of energy is available, there are times when the mechanism that caused bloodspatter cannot be determined. Caution is required before forming a firm conclusion as to the only manner by which a pattern could be produced. It is usually better to use the words "consistent with" than it is to be more definite. This advice can be applied to many different bloodstain pattern types and is not restricted only to those that result from a medium-velocity impact.

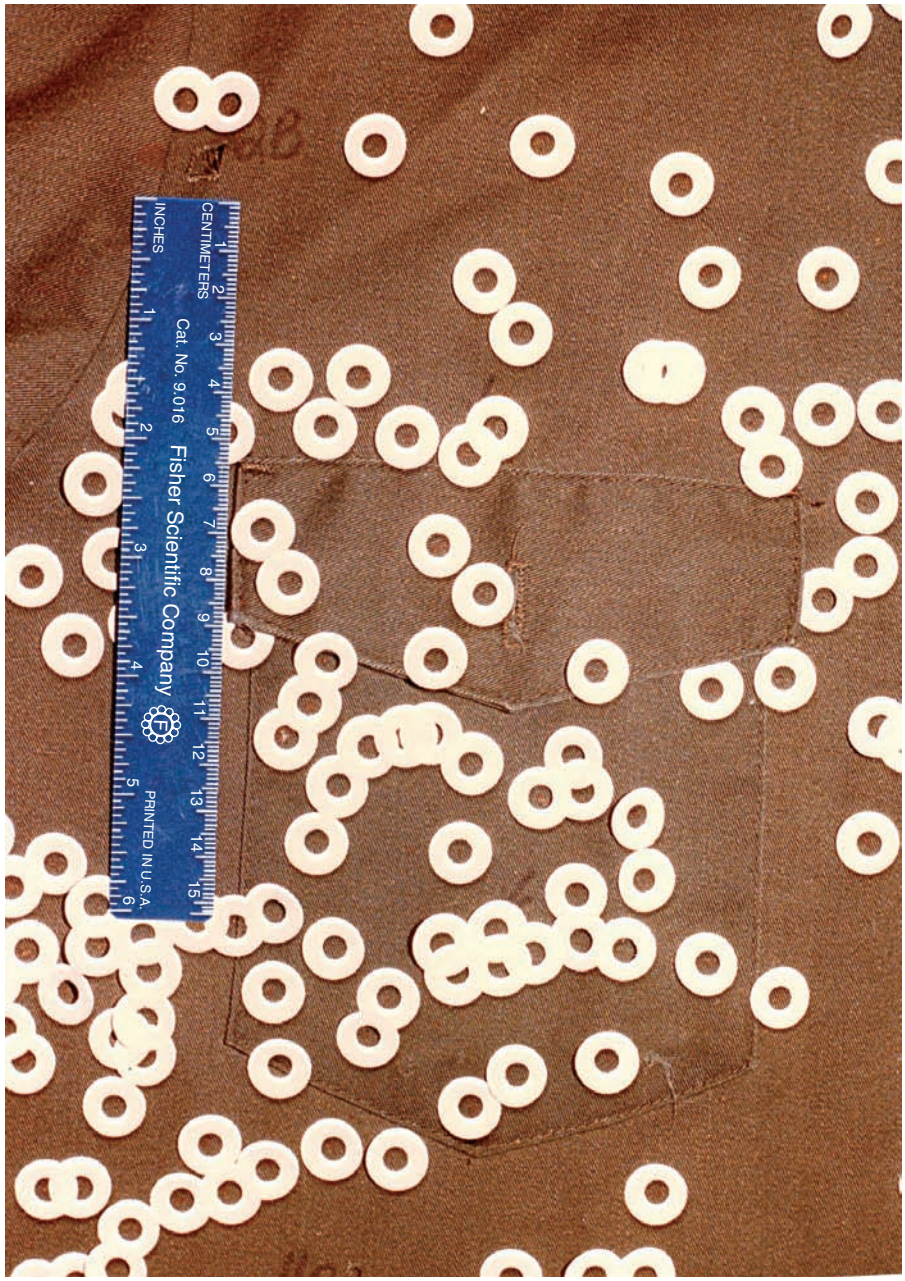


Figure 12 Three ring binder reinforcers used to identify the locations of each small bloodstain on the shirt front. Bloodspatter was present on both the outside and the inside of this pocket's flap

An important rule is that in most beatings there is little, if any, bloodspatter resulting from the first blow. The first blow in a beating does not usually strike blood. If it is severe enough, this blow only creates an open wound that will bleed. After blood comes to the surface of an injury any subsequent blows delivered to the same location will cause blood to spatter.

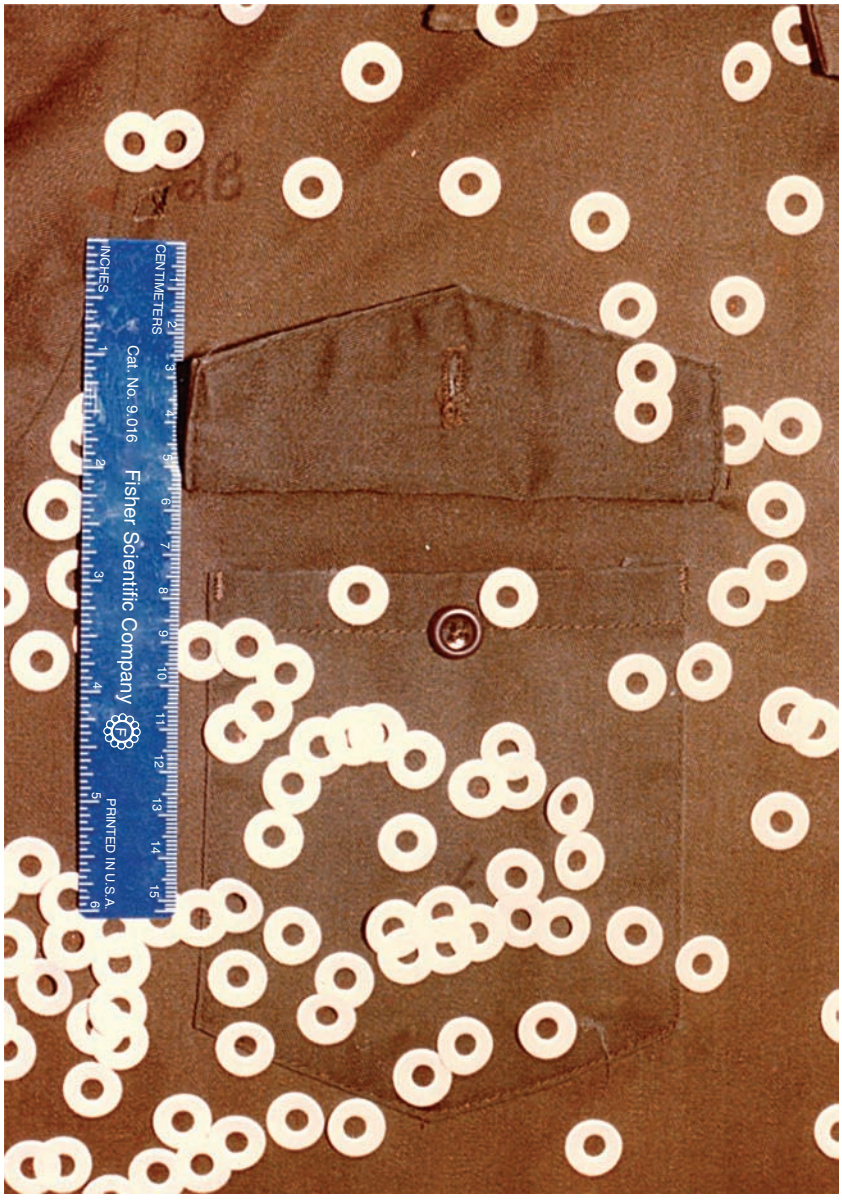


Figure 13 The same shirt that was shown in Figure 12 but the pocket flap has been raised to show that blood had also spattered under the flap and onto the front of the pocket beneath it. Therefore, this flap had to be up as well as down to allow blood to spatter onto both these locations. If the person wearing this shirt were raising and lowering their arm as they were beating someone, such a motion could produce this pattern

Blood Spattered from High-Velocity Impact

When considerable energy is applied to a liquid such as blood, the result is a proportionally greater dissociation of the liquid resulting in extremely small droplets. Stated quite simply, the more energy that is available, the more blood will spatter as its surface tension is ruptured by hydrostatic force. High energy allows much greater subdivision of liquids



Figure 14 Triangles of masking tape placed on the blue jeans to show the locations of small blood spots. Each piece of tape can be numbered so that a correlation between the results of chemical tests and the stains tested can be easily established

upon impact than can be accomplished during a beating. Thus, the higher the energy source, the smaller the blood drops. In almost all criminal cases wherein high-velocity impact spatter is produced, it is the result of one or more gunshots. It must be remembered, however, that almost any type of explosive force is capable of producing this type of pattern when a source of blood is available. The explosion of a hand grenade, high-speed machinery, or an aircraft propeller can also produce high-velocity impact spatter. Such cases are unusual but should caution the reader to remember that the term "*high-velocity impact spatter*" means just that and is not necessarily the result of gunshot.

In the usual case of a shooting where the projectile strikes exposed skin, the energy at impact is hydrostatically transmitted throughout much of the adjoining tissue. This results in the spattering of blood in a very fine, almost mist-like, spray. These atomized droplets of blood have a very high surface area and, therefore, cannot be projected very far in a horizontal direction. Usually, these very small droplets will be found within 3 or 4 ft of the impact point.

In addition to very small, mist-like droplets, several larger droplets will be produced as well. A typical

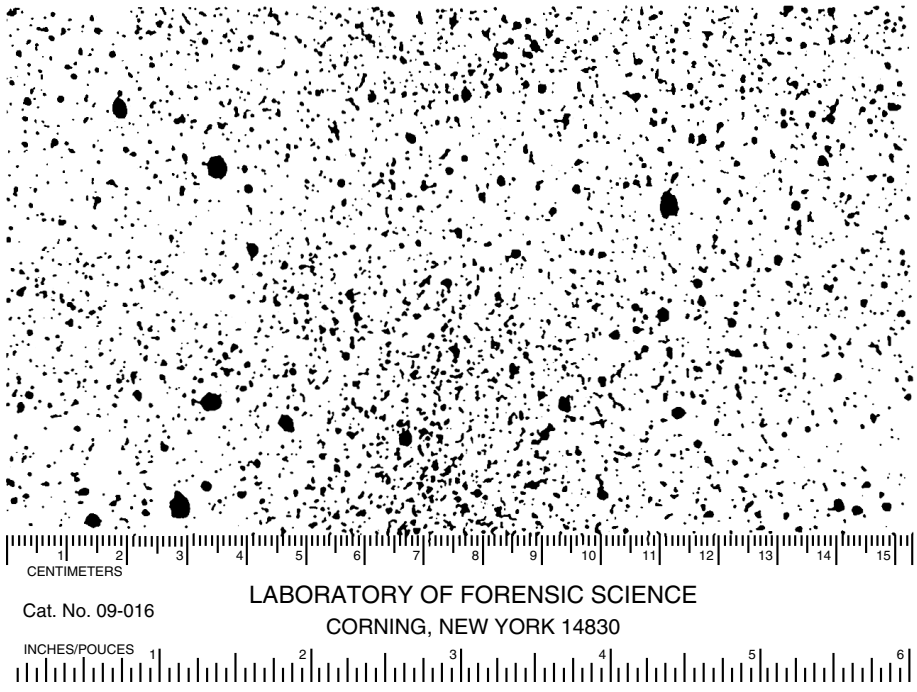


Figure 15 A high-velocity impact pattern resulting from a gunshot. Note that the diameter of the majority of blood spots is <1 mm but some much larger are also produced

spray pattern, characterized as high-velocity impact spatter, may be seen in Figure 15. Note that while the vast majority of these blood spots are well <1 mm in diameter, many larger ones are also produced. Interpretation of the bloodstain patterns of this type at a crime scene may be difficult when the body is not present. The investigator may believe that he is looking at medium-velocity impact spatter when he is actually looking at the larger droplet stains, which were projected beyond the limits of the characteristic mist that resulted from a gunshot impact. This is an unusual situation but it has been known to occur and will likely occur again.

The scene of a typical shooting showing extensive bloodspatter from high-velocity impact is shown in Figures 16 and 17. The weapon used in this case was a 30–30 rifle. It is evident in Figure 17 that void areas on the wall resulted when bloodspatter was intercepted by the brass tubing frame of the head of the bed. Such well-defined vertical lines demonstrate that the spatter came from a single impact and not what would likely have resulted had multiple blows been administered during a beating. The very small mist-like bloodstains that were on the wall cannot be resolved in this figure.

When very fine bloodstains are observed in a shooting incident, they should always be considered as most likely having originated from the impact of a bullet. In more than one case, the characteristic mist-like spray of blood was present inside the shirt cuff of a suspected suicide victim. Such evidence requires the open cuff area to be facing the victim's wound indicating a posture consistent with their holding a gun to their head.

It is important to examine a high-velocity impact spatter very closely to be able to resolve the very small blood spots that identify this type of pattern. However, it is frequently obvious that a gunshot wound has occurred from viewing scene photographs that were taken at a normal distance to show an overall area. The crime scene photograph shown in Figure 18, for example, is too distant to resolve the fine spray that can only be seen by getting very close to the ceiling. Nevertheless, when blood is projected upward onto the ceiling, such as shown in this figure, it is almost always the result of a gunshot that had



Figure 16 High-velocity impact spatter at the scene of a shooting. The victim was lying in bed with her head just to the right of the pillow when she was shot in the mouth with a 30–30 rifle

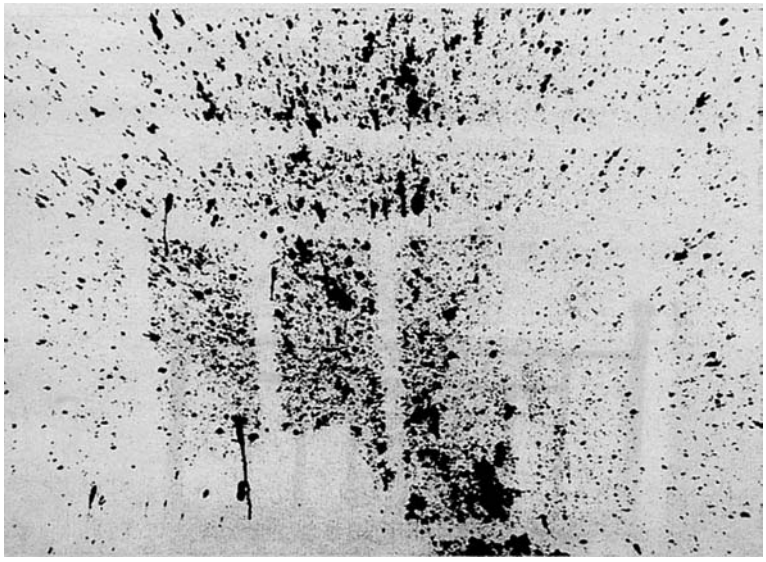


Figure 17 Close-up of the bloodstain pattern on the wall behind the bed shown in Figure 16. Void areas clearly define the outline of the brass headboard and the pillow. Such clear lines of demarcation establish the position of the victim when she was shot in her head

an upward trajectory. Such trajectory is most often associated with a suicide, as was the case shown in Figure 18.

When a high concentration of high-velocity impact spatter is found on a floor, furniture, or other surfaces, it will usually be directly below the point of impact that produced it. Clothing and shoes that are stained with high-velocity impact spatter must have been very close to the victim at the time of the shooting.



Figure 18 A high-velocity impact blood pattern on the ceiling of a bedroom. Although the very fine individual blood spots cannot be resolved in this photograph it is easily recognized from its overall appearance

It must be remembered that when blood is spattered as the result of a high-velocity impact, the fine mist produced does not travel very far in a horizontal direction. Air resistance acts against the projected droplets slowing them down so their distance traveled is inversely proportional to their cross sections. Thus, even the extensive spatter shown in Figure 19 did not result in any very fine, mist-like droplets reaching the walls or doors as they could only travel 5 ft or less. An examination of the walls and door in this figure suggests that considerable energy was available to have produced such extensive spatter.

It seems appropriate to discuss an apparent difference of opinion between two schools of thought regarding terminology of impact spatter. This author prefers the term *velocity*, rather than *energy* to describe the nature of an impact that results in bloodspatter. Thus, the terms *low-velocity impact spatter, medium-velocity impact spatter, and high-velocity impact spatter* are preferred over *low-energy impact spatter, medium-energy impact spatter, and high-energy impact spatter*. While these two physical terms are closely related, the justification for using *velocity* rather than energy may be demonstrated by a comparison of the difference in bloodspatter resulting from two events of equal *energy*. In the first example consider a bullet impacting a sponge that has been saturated with blood. The result is a very fine spray of blood, high-velocity *impact* spatter, which is easily recognized. In the second example, an elephant steps into a large pool of blood causing very large splashes of blood to appear around the elephant's foot. The "elephant splashes" could never be mistaken as being produced by gunshot and yet the same *energy* was available in both instances. Thus, the more appropriate descriptive term *velocity* is the better choice. Possibly, the term *flux*, which could be used to describe the *rate of flow of energy,* ought to be considered.

Bloodstain Patterns Resulting from Projected Blood

When large volumes of blood are projected onto a surface, a somewhat unique bloodstain pattern can result. The impact surface may be a wall, a floor or any other surface. Only under unusual circumstances could it be a ceiling. The most common situation in which this occurs is when an artery, usually a major artery has been severed or otherwise breached from gunshot or some form of blunt force injury. Systolic arterial pressure may cause blood to spurt or gush from the wound in large volume pulses. If blood is projected in this manner and strikes a surface that is relatively close to the injured victim, the resulting bloodstain pattern will fall into one of three basic classifications each of which may be easily recognized.



Figure 19 Extensive bloodspatter resulting from a shotgun discharge to the victim's face. When he was shot, the victim was lying face up on the floor. This photograph is an excellent example of how larger blood drops, produced from high-velocity impact, can travel great distances but mist-like smaller droplets cannot

Type I Arterial is characterized by large bloodstains, which have very elongated spines. The elongated spines indicate that an energy greater than that resulting from normal gravitational attraction was present.

An example of a Type I Arterial projected bloodstain pattern is shown in Figure 20. A young girl was stabbed in the neck and her left carotid artery was severed. She fell to her knees as she crawled away from her assailant and left a trail of large arterial splashes on a driveway. The trail of her blood extended over 60 ft from where she was stabbed to where she was found.

Type II Arterial is characterized by bloodstains which are, somewhat smaller than Type I. They have well-defined borders with few, if any, spines. Bloodstains of this type often have sufficient volume that blood run down a wall or other vertical surface. They usually result from a wound that somewhat restricts the clean projection of blood away from the victim's body or that is a few feet from the target upon which they land. An example of this is shown in Figure 21.

Type III Arterial is a difficult arterial pattern to recognize because it consists of small drops rather than the large splashed blood volumes, which identifies Types I and II Arterial. Smaller drops of blood are produced because the artery, which is breached or severed, is covered with overlaying tissue, which can restrict blood from exiting in the large discrete pulses typical of Types I and II Arterial. Nevertheless, one characteristic of Type III Arterial identifies it with certainty when it is present, the appearance of a series of individual heartbeats, which reflect the systolic mode of the heart. Such patterns look somewhat like one that could be produced by raising and lowering a garden hose nozzle, which was set to a coarse spray while slowly moving it to one side or the other. Figure 22 shows an excellent example of Type III Arterial that resulted from a self-inflicted gunshot would to the right temporal area.

Another example of Type III Arterial where blood spurted from beneath crushed tissue over an anterior temporal artery may be seen in Figure 23. These patterns resulted from spurts of blood onto a bathroom wall. Individual heartbeats are identified as the peak of each spurt in this figure.

Cast-Off Bloodstain Patterns

During a beating, blood usually does not accumulate on the surface of the victim's wound after the first blow. After blood covers that area, however, a "second" blow to the same area will produce spatter and often adheres to the beating instrument. As a



Figure 20 A Type I Arterial pattern produced by blood, which was projected from a carotid artery that had been severed by a stabbing. Well-defined pulses are evident



Figure 21 A Type II Arterial pattern that was produced when blood gushed from a wound to a wall at a greater distance than in Figure 20. Therefore, the large spurts separated into smaller drops. Two distinct pulses are evident



Figure 22 A Type III Arterial that was produced when blood from the victim's severed right temporal artery was projected on the carpet several feet from its origin. As her head turned slowly to her right side, blood created a zigzag pattern on the floor. This pattern is like a hospital strip chart recording of a cardiogram



Figure 23 A Type III Arterial spray pattern that was produced when the victim of a beating stood in front of his bathroom sink. Blood was sprayed from his crushed temporal artery to a wall indicating that he stood in that location for several seconds. Very small droplets were produced because of his skin overlying his wound

bloody hammer is raised and swung back by an assailant, the path of the hammer head is more up and back than it is in a curved motion. A study of the beating mechanism, using high-speed strobe photography confirmed this rather right angle course. As a result, when the direction of the hammer head is shifted from an upward, raising motion, to a backward, swinging arc, the angular momentum of the blood adhering to the swinging hammer is usually sufficient to overcome the surface tension of blood

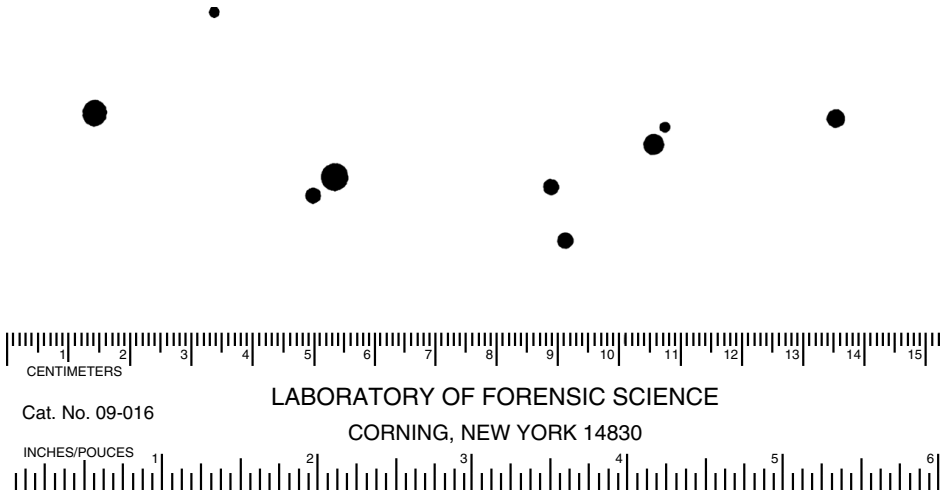


Figure 24 A cast-off bloodstain pattern on a ceiling directly over the point of impact. These spots are round because they struck the ceiling about 90°

allowing small drops to be released. These small drops usually strike the ceiling about 90° directly over the victim. A few bloodstains that were on a ceiling directly over the head of a person who was beaten on a bed are shown in Figure 24. The direction the instrument was being swung when this pattern was made cannot be determined from blood spots that are as nearly circular as those that are shown in this figure.

As the instrument continues to be swung backward, its movement accelerates and additional blood droplets will be cast off from its surface. If any of this blood reaches the ceiling, it will strike at ever increasing acute angles and the bloodstains will become more elongated. Figure 25 shows the continuation of the cast-off arc pattern shown in Figure 24, but at a distance some 22 in. further to the right. Directionality of these bloodstains is clearly from left to right. Observing a cast-off pattern such as this on a ceiling identifies the location where a beating had occurred. Although only the two ends of this pattern have been used as an illustration, the entire overhead pattern is in a "single file" straight line configuration.

Frequently, multiple cast-off patterns are present when a victim is struck repeatedly in the same general area so blood can accumulate on the weapon being used. If the victim was not mobile the minimum number of blows may be estimated by counting the individual cast-off blood trails. One blow should be added to the total because the first blow struck will not usually allow blood to adhere to the beating instrument in a sufficient volume to produce a cast-off pattern on the backswing. If the victim was mobile while being beaten, ceiling cast-off patterns can appear at more than one location and do not radiate out from a single origin. Likewise, if more than one person administers a beating, or if a single assailant moves about, more complex patterns can be produced.

When sufficient blood adheres to a beating instrument while it is being raised up, cast-off patterns can also be produced on the downward or beating swing. If such pattern is produced, it will most likely consist of two or more parallel lines whose individual blood drops will show a downward directionality. The reason for the "single file" bloodstain pattern on the upswing and the "double (or triple) file" bloodstain pattern on the downswing is simply that more blood is released, or flung from the instrument nearer the end of a swing when it is at its greatest velocity. The upward swing casts blood off while the instrument is accelerating, while the downward swing has already reached its maximum velocity. This applies to walls only as these patterns reverse for cast-off bloodstains on ceilings. When there is a considerable blood volume on an instrument, it may create a confusing "double file pattern" where there are bloodstains that appear to be traveling in two different directions. This only occurs when the victim is near a wall and the person doing the beating is facing that wall. Some

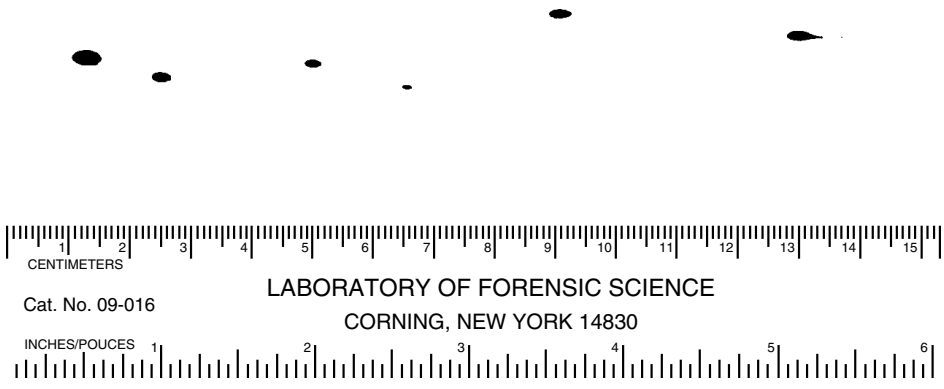


Figure 25 A continuation of the right side of the pattern shown in Figure 24. The sharp pointed ends on the right of these bloodstains show their direction of travel, left to right

large drops of blood are released as the instrument is raised and they do not travel upward very far before falling and they strike the wall in a downward direction. Many smaller drops are then released as the instrument continues to be swung upward and even though they are actually part of the same cast-off "single file" pattern, they were traveling in the opposite direction when they struck the wall.

Horizontal cast-off patterns are common when blows are delivered by horizontal rather than vertical movement. As a result of gravity, the end of such patterns will be seen to be falling off or downward toward the floor.

Expirated Bloodstain Patterns

If blood accumulates in the air passages of an injured person, it is frequently blown out through the mouth, the nose or a combination of both. Blood projected by such mechanism is called "*expirated*" *blood*, not "aspirated" blood. Aspiration is the sucking in action of a chest wound, not a blowing out. Blood that is projected blood by expiration produces bloodstains that appear very much like those which result from a medium or high-velocity impact. In contrast, however, droplets that are blown out of the nose or mouth can only be found in locations where such an origin is possible.

Blood spots on the fingers and hand shown in Figure 26 appear small enough to have been produced as a result of a medium-velocity impact. The actual origin of this blood, however, was the victim's mouth, as shown in Figure 27. If there is a question as to whether or not blood could have been expirated, rather than from impact, photographs should be taken of the victim's face and their oral and nasal cavities should be swabbed if it is not obvious that they contain blood.

Recognition of expirated bloodstain patterns is not difficult. Common sense dictates that if the victim had no blood in the nasal or oral orifices it would have been impossible for blood to have been expirated from either. When air bubbles are present in bloodstains this is evidence of expirated blood. However, it should be remembered that penetrating wounds to the chest can also produce bubbles in blood. In such cases, there may not be obvious blood accumulated in the mouth or nasal passage.

Expirated blood may also be blown out of wounds as well as from the nose or mouth. In a case in Pittsburgh, a man stabbed himself in the chest with a large buck knife that remained stuck in the wound. Several very small bloodstains were evident on the knife directly over his wound as a result of blood that was expirated up onto it.

Transfer Bloodstain Patterns

Human blood is frequently a medium of transfer, which allows geometric shapes of various kinds to be reproduced as a mirror image onto a surface. Bloody fingerprints, footprints, and footwear prints have been received in evidence countless times so an extension of this basic transfer mechanism to include fabrics, hammer heads, knives, clubs, tire irons and the like, is nothing unusual. A transfer pattern frequently



Figure 26 Bloodspatter on the victim's hand suggests that it was the result of medium-velocity impact; however, the actual origin of this expirated bloodstain pattern was the victim's mouth



Figure 27 Actual source of the bloodstains shown in Figure 26. Blood was expirated from the mouth and/or nose onto the victim's hand

encountered in homicide investigations is the hair swipe pattern. Two basic patterns of this type are quite distinct in appearance and both are easily identified. The first is produced when hair that is wet with blood makes a lateral swiping of a surface and leaves characteristic fine lines and a feather-edged reproduction of several individual hairs. A typical hair swipe pattern is shown in Figure 28.

When hair that is wet with blood comes into contact with a surface without lateral movement, the result is compression transfer rather than a swipe pattern. A young girl was killed when her assailant

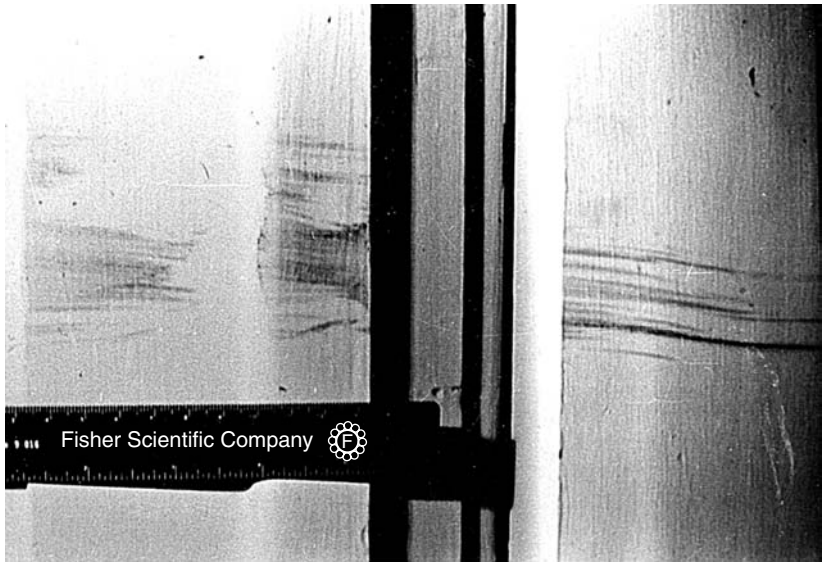


Figure 28 A typical hair swipe pattern that was caused when bloody hair latterly brushed across a door and doorway



Figure 29 A typical compression transfer pattern produced when blood soaked hair pressed against a surface without lateral motion. Fine line transfers show the impressions of individual hairs

stamped on her head after raping her. Considerable blood had soaked into her hair and when her head came into contact with a wall and baseboard a compression transfer was produced. That pattern is shown in Figure 29. A study of this figure discloses several points of contact where characteristic fine, single hair replicas have been left on the baseboard and the wall above it.

In most instances where a compression transfer pattern is produced, blood is deposited from a bloody object in a mirror image of the original. However, when a coarse fabric, such as corduroy or knit

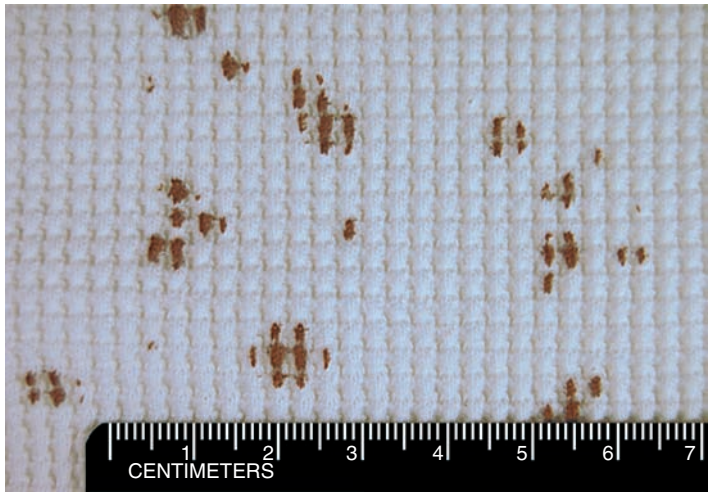


Figure 30 A typical bloodstain pattern resulting from compression transfer. Only the top of several strands of fibers are bloodstained

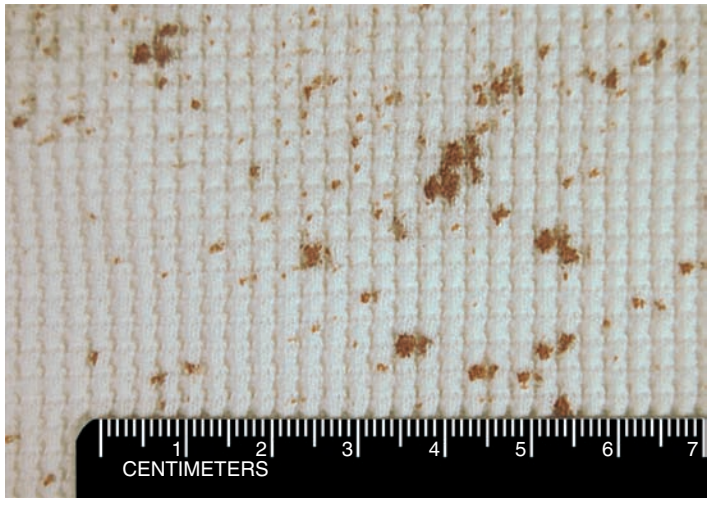


Figure 31 Appearance of small bloodstains that resulted from blood that was projected onto a sweater. Note their random distribution between and on the "ribs"

sweater, is involved, only the top of the fabric will become stained. The presence of blood on the "ribs" or upper weave portion of a fabric will establish the staining as a transfer mechanism because had the blood been projected onto the fabric it would be equally deposited between the "ribs" as well as on top of them.

A bloodstain compression transfer pattern may be seen in Figure 30, which resulted when a person who was wearing a sweater bent down over a fresh bloodstain pattern. Blood was transferred only to the top of the weave pattern. Projected blood droplets on fabric are not restricted to the top of a weave as they penetrate down into the lower portion or "valleys" as well.

Figure 31 shows bloodstains that were projected to the back of the same sweater shown in Figure 30. Because of their random distribution, many of these stains landed between, as well as on the top of, the

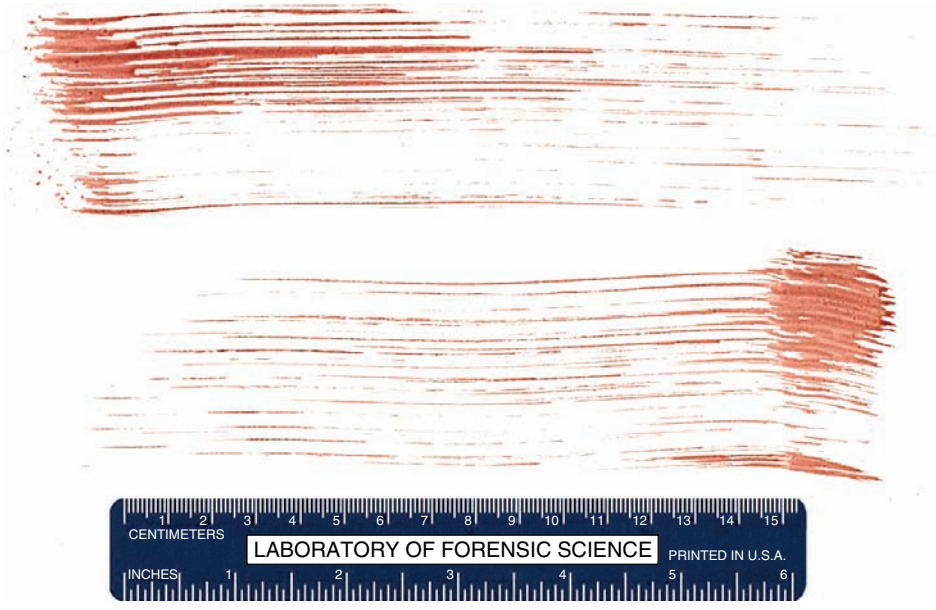


Figure 32 Two similar lateral swipe transfer patterns. Both were produced from a left-to-right lateral motion

weave pattern. This is a relatively simple observation that can be made with only a hand magnifying glass.

Some authors have written that it is possible to determine the direction of travel of the object that created a swipe transfer pattern. This may often be true with hair swipes; however, it is not possible to make this determination with 100% accuracy when a bloody fabric leaves the bloodstain. Figure 32 shows two swipe bloodstain patterns both of which were made by a left-to-right lateral motion although the lower one appears to be just the opposite. Obviously, caution must be exercised in forming a firm opinion when exceptions to general rules are known to exist.

Recovering and Documenting Bloodstain Pattern Evidence

Photographic Documentation

It is essential that good photographs be taken of all bloodstains at the scene of a crime. Close-up photographs should be taken of all bloodstain patterns. These should be taken normal (90°) to surface upon which the bloodstains are found. A ruler or some other scale of reference should be included in each photograph. It is always important to document which bloodstains are being photographed. One convenient way to achieve this is to use a felt marker to circle and number all bloodstains and/or any bloodstain patterns which are of interest. Overall photographs should be taken to show the geometric relationship of bloodstains to the general crime scene area as well as to each other. When photographing bloodstains at 90° it is important not to have a flash on the camera. The satin-like sheen of dried blood will cause serious flash back and the bloodstains may not show up.

The exact location of bloodstains, which will be used for calculating angles of impact, must be recorded in some systematic manner. The distance of a bloodstain from the floor or ceiling and from the nearest wall should be entered on a form that provides a blank space for the width of a blood stain, its length, and its width-to-length ratio followed by space for the impact angle after it has been calculated.

Distant or overall photographs do not usually require a reference scale because furniture and/or other objects are usually present, which allow the viewer a good size relationship. If the origin of bloodspatter is to be photographed, elastic string that has contrast with the background should be selected.

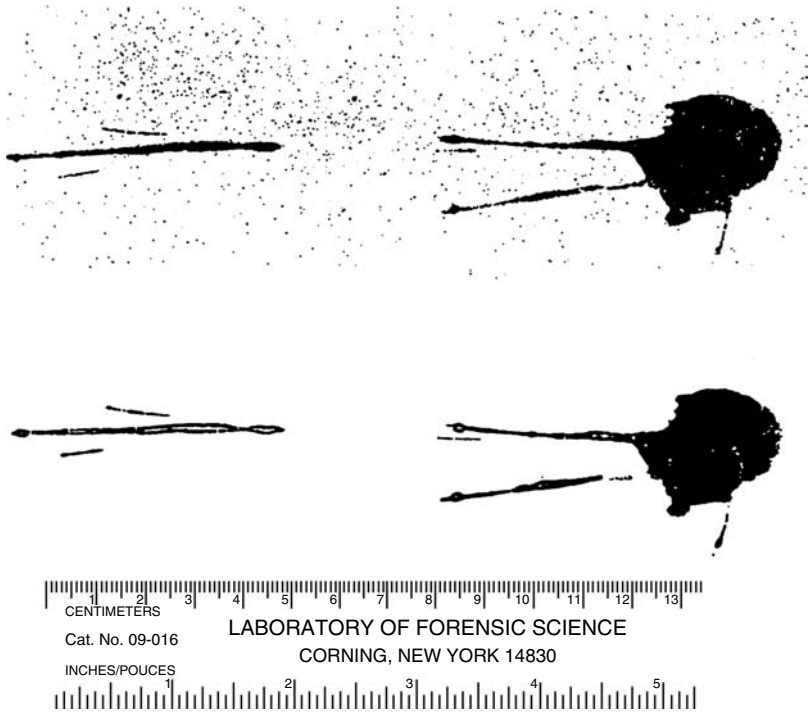


Figure 33 Bloodstains that were lifted from a kitchen floor tile. The top stain shows how the blood appeared before it was lifted. The bottom stain is the lift that was removed from the floor with clear tape

Lifting Bloodstain Pattern Evidence

In addition to the photographic recording of bloodstain patterns, another simple, and perhaps even better technique, should be considered; remove the bloodstain itself. This recovery method at a crime scene is apparently not often employed by investigators. The method is no more difficult than lifting a processed fingerprint and, in fact, employs the same material, transparent fingerprint tape.

After photographs have been taken if bloodstains are on a flat, hard, smooth surface, such as major kitchen appliances, an attempt should be made to lift some of the bloodstains to preserve them as evidence. One and one-half inch wide transparent fingerprint tape, which also can serve as a scale of reference, can be used. The procedure is very straightforward; one simply presses the transparent tape over the bloodstain applying considerable pressure to the stain area with a ballpoint pen or fingernail. The tape is then carefully peeled from the surface. The bloodstain is frequently almost completely removed. The lift should be placed on white plastic so, if for any reason, it can be lifted again in a laboratory to expose the blood. This is to suggest that this is not a method of choice for recovering blood samples for serology.

Although hard, smooth surfaces are ideal for lifting bloodstains at a crime scene, many unlikely surfaces often yield excellent results. Nothing is lost if the attempt to lift a bloodstain is unsuccessful so there is really no reason not to make the effort. Figure 33 shows how efficiently a bloodstain was removed from a kitchen floor tile. Blood shown in the top of this figure is how it appeared on the floor. Blood shown at the bottom of this figure is what was lifted using fingerprint lifting tape.

Blood that remained on the floor tile consisted of very fine lines that would fill the narrow voids in the streaks of the lift. Recovery was well over 99% effective. Many other more porous surfaces such as wallpaper often permit removal of the bloodstain; however, part of the wallpaper is frequently removed as well. In such cases, the size and shape of the bloodstain are still preserved so nothing is really lost.

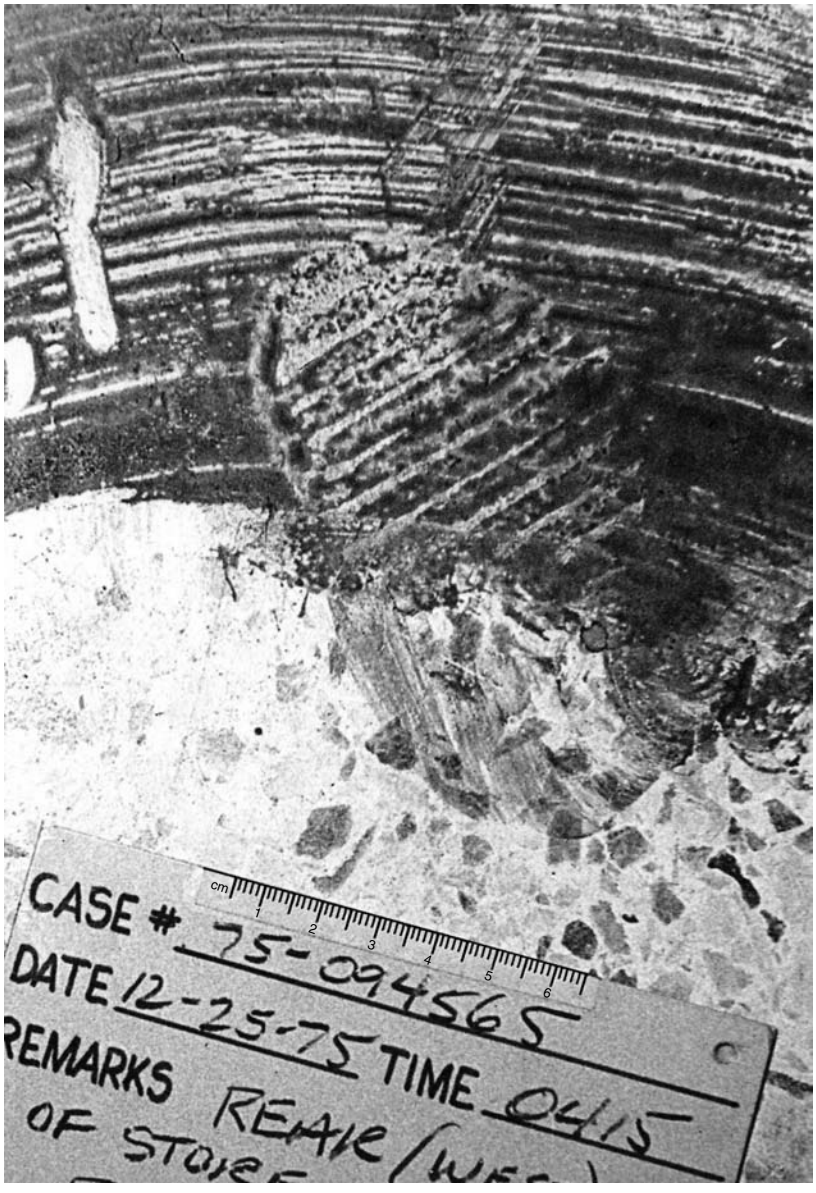


Figure 34 Large, thick smear of blood on the floor of a furniture store

Research When Necessary

From time to time, bloodstain patterns in a case will have some aspect that could be difficult for a jury to understand. When that happens, it may be advisable to conduct and document certain experiments. This was the situation in a Florida case where the elimination/identification was in dispute. There was a large smear of blood on the floor of a furniture store where four people had been murdered. Within the smear there was a void that looked very much like it could have been caused by the heel of a shoe. Its overall appearance and size matched the suspect's shoe; however, when the shoe and the photograph were sent to a federal laboratory for evaluation, the examiner reported that the pattern could not have been made by

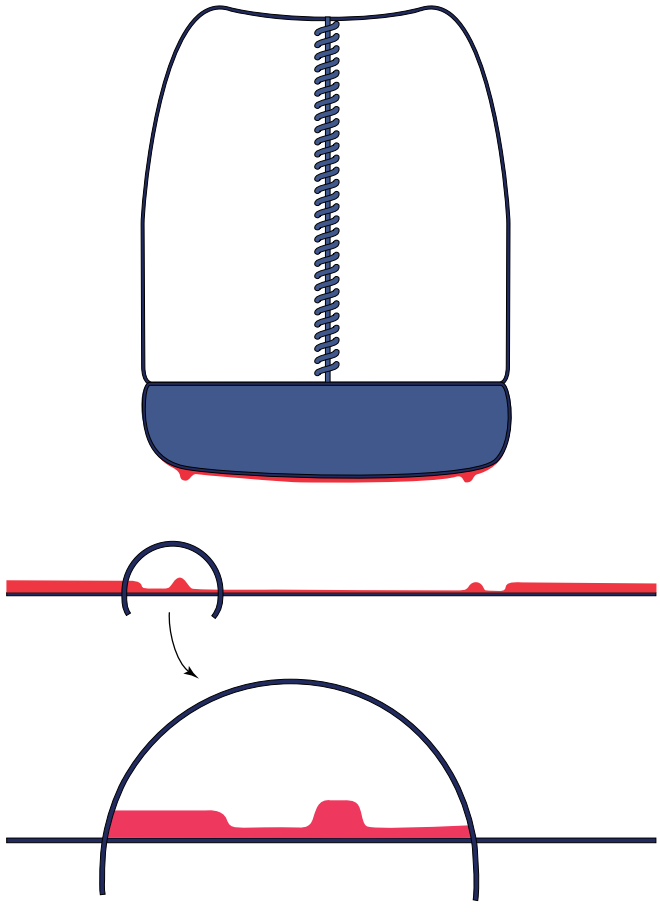


Figure 35 Blood on the heel after being lifted off the blood smear on the floor. The broken web of blood is on both the shoe heel and the floor

the suspect's shoe. The basis for concluding that the bloodstain and the shoe did not match was the presence of a ridge of blood, which was just inside the outer edge of the heel. The arrow points to this ridge in the large blood smear shown in Figure 34. The suspect's shoe was very smooth in the corresponding area and had no such ridge.

The prosecution had a problem. If this was not the defendant's heel print, then another person must have been at the scene of the murders. Actually, it should be obvious to any scientist why there was a "false ridge" in the bloodstain; it is called *capillary action* and is the result of the surface tension of blood. When the suspect stepped into the thick smear of blood almost all of the blood was forced to the sides of the shoe. Then, as the shoe was raised up, blood around its edges was drawn back under the shoe by capillary action.

As the shoe was raised up still further a "web" of blood formed and broke leaving the false ridge just within the outline of the heel and on the heel itself. A diagram of this phenomenon is shown in Figure 35.

To demonstrate to the jury how easily a false ridge of this type could be produced, a smear of blood was made on a tiled floor. The defendant's shoe was pressed down into the blood and raised back up. This was done several times using the evidence shoe and on every occasion a clear false ridge was produced. The result of four experiments may be seen in Figure 36. This experiment had been conducted at least 300 times at the Bloodstain

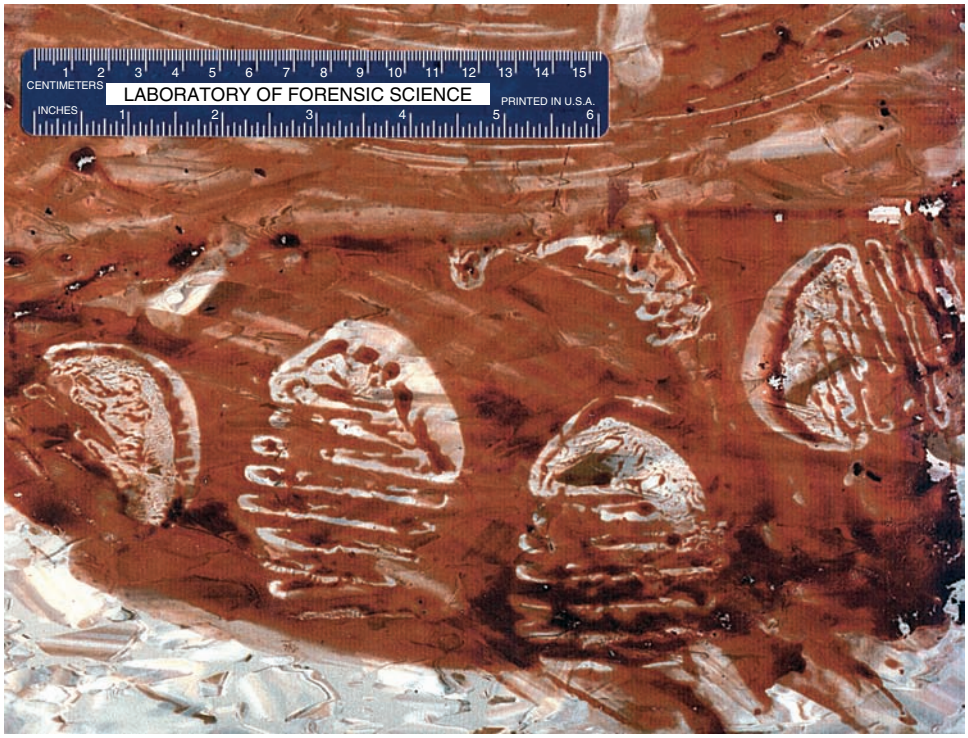


Figure 36 Blood ridges produced using a shoe that is absolutely smooth in the area of concern

Evidence Institute, using a shoe of the same type that was worn smooth in the same area and to the same degree as the evidence shoe. The results are always the same.

A young girl in Oregon claimed that bloodstains on the ceiling of an upstairs bedroom resulted from an injured older lady shaking her hair, which was wet with her own blood. The ceiling patterns were obvious cast-off patterns, which are characterized by their in-line and arc-like geometry. A series of experiments using live subjects, whose hair was saturated with human blood (this was before acquired immune deficiency syndrome (AIDS)), were conducted to determine how bloodstain patterns produced by blood cast off from hair would actually look. Several younger and more athletic girls with a wide variation in their hair lengths were the models. None of them could throw even one small drop of blood from their hair to strike a ceiling of the same height above a bed as in the actual case. It was proven that blood was cast off from a fire poker and not the victim's long hair.

Reconstruction of Events Using Models as Victims or Perpetrators

People were used as models for over 30 years to aid in a better understanding of what is and is not anatomically possible [15]. I have data on some 250 individuals in my computer who are listed by gender, height, weight, hair length, hair color, color of eyes, and age to pose in a reconstruction so they can be photographed and better show a jury what the evidence suggests rather than a word description.

Case Example #1

A young girl was found as shown in Figure 37. She had many bloodstains on her body and clothing and it was obvious that she had been put in the location and posture in which she was found after her blood had run on her legs. How it ran down both sides of her right leg was resolved by using a model.

In Figure 38, a technician is transcribing bloodstains that are shown in Figure 39 to a model's leg

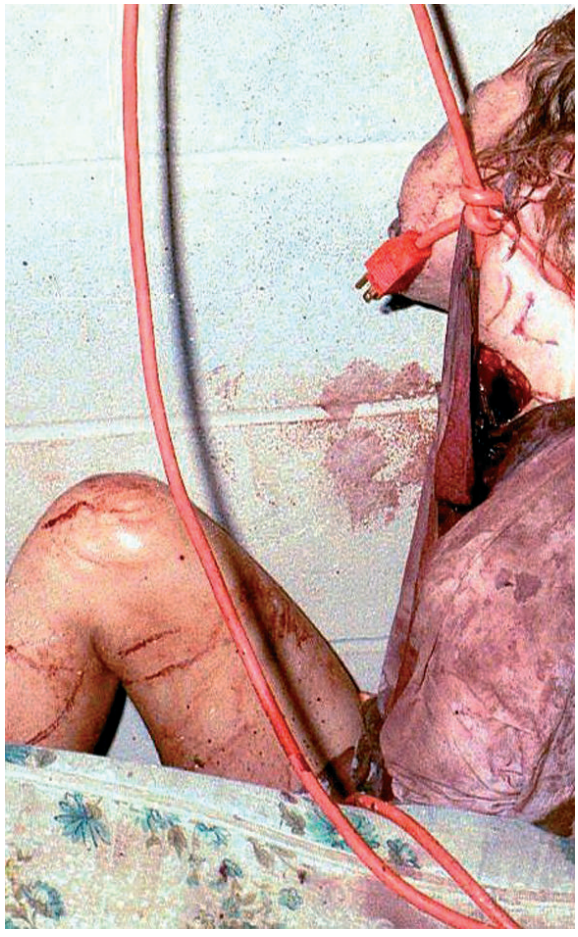


Figure 37 Victim as discovered and photographed at the crime scene

using a red felt marker. After this had been done, the model assumed a position similar to the position the victim was in when she was found as shown in Figure 37. Several photographs of the model were taken to show how her leg had to have been in a normal sitting posture when blood ran down her leg. Blood does not run sideways nor can it flow upward under normal circumstances.

The victim had a slit on her wrist, which was more consistent with a defense wound rather than one that had been self-inflicted. This cut was the only logical source of the blood, which had run down both the inside and the outside of her right thigh. Figure 39 shows how easily that cut could have caused the bloodstains on her leg if her arm was resting on her knee.

Case Example #2

In the OJ Simpson case, many bloodstains of various sizes and shapes were present on Nicole Brown Simpson's back but none of these were ever collected for analysis. Figure 40 shows the size and shape of many of these bloodstains. The bloodstain over her left shoulder was the size of an adult fingernail, yet it was left only to be washed away at the morgue.

The question as to the origin of these bloodstains was solved by having a model of the same physical build assume the posture in which Nicole Brown Simpson was found and photographed. Using seven photographs of Nicole's back taken at the scene the location, size, and shape were transcribed to a model's back using a red felt marker as shown in Figure 41.

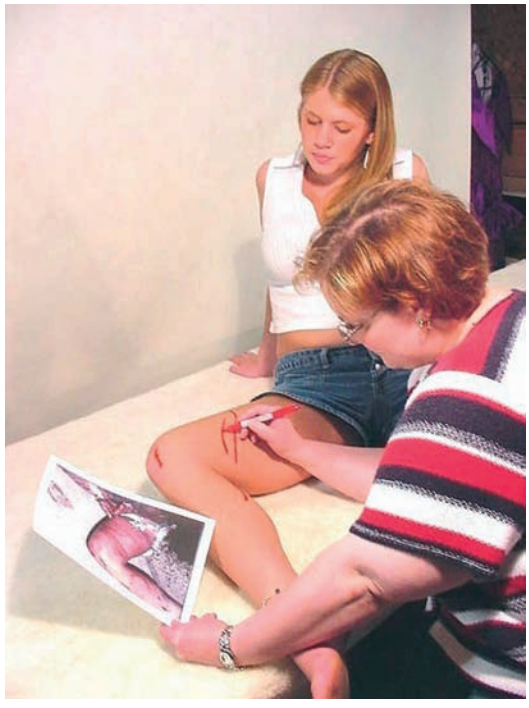


Figure 38 Drawing "bloodstains" on a model's leg with a red felt marker in the same size, shape, and location as it appears on the victim's leg in the crime scene photograph



Figure 39 "Bloodstains" on the model's leg in relationship to a cut on her wrist



Figure 40 Bloodstains on the back of Nicole Brown Simpson which were never collected

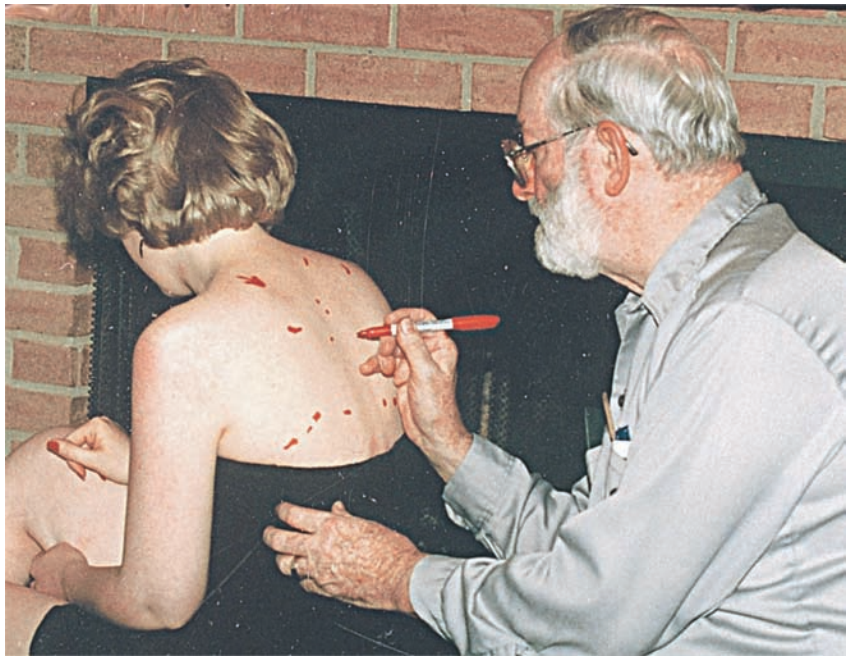


Figure 41 Drawing a bloodstain on the model's back with red felt markers in the same location, size, and shape as those of the back of Nicole Brown Simpson. Seven separate photographs were used for maximum accuracy



Figure 42 (a) Bloodstains on Nicole Brown Simpson's back at the crime scene. (b) "Blood" drawn on the back of a model with a red felt marker duplicating the same pattern of spattered blood shown in Figure 42(a)



Figure 43 Notice the concentration of blood that was spattered from the victim's head wounds to the snow over her right shoulder. The dark spots over her left shoulder are small stones and not bloodstains



Figure 44 This figure shows the only location where the perpetrator could have been standing relative to the victim to have had blood spattered over her right shoulder as seen in Figure 43

After all of the larger bloodstains were drawn on the model's back, she assumed the position of the victim. A comparison between the victim's position and that of the model may be seen in Figure 42 (a), (b). Using the "bloodstains" on the model's back, it was possible to estimate the location from whence the blood spattered. As it turned out there was no need for this information; however, it was available had it been wanted.

When comparing Figure 42(a) to Figure 42(b) it must be noted that they were taken from slightly different perspectives. Therefore, the bloodstains on the top of Nicole Brown Simpson's shoulder, which are very evident in Figure 40, do not show up well at all in Figure 42(a).

Case Example #3

In an Ontario case, the relative position of the man who beat a woman to death was important as a witness had given this information to the police. This was a relatively simple task as may be seen from the location of the bloodspatter pattern in the snow. The victim's wounds were all to her right temporal area. From the pattern of blood shown in Figure 43, it does not appear to have but a single origin. This is in agreement with the witness' statement that the victim was already incapacitated when the final beating took place.

The instrument used for this beating was an automobile jack. It had a long oval slot cut out of the metal and an impression of its outline was clearly evident on the victim's skin around her head wounds. Therefore, there were only two places the person administrating the beating could have been standing at the time. He was either standing beside her right side or over her left shoulder. For blood to have spattered in the location where it was found there was only one possible choice. He stood as shown in Figure 44, over her right shoulder. This provided corroborative evidence as that is what the eyewitness had reported.

Summary

When bloodstain pattern evidence is properly studied and evaluated, it should result in a more accurate reconstruction of events than otherwise would have been achieved. In conjunction with other evidence, bloodstain pattern interpretation can often be of great assistance to juries. Frequently, bloodstain patterns will be helpful in confirming conclusions that have been established from other evidence, such as a projectile's trajectory and *vice versa*.

Physical evidence is evidence that can be measured and analyzed. It is usually evidence of substance, but not always. When blood ages, it goes from

red to maroon to almost black, so, can the color of a bloodstain be useful in estimating its age? Yes and no. The most important factor is where the bloodstain is located and how it is viewed and/or photographed. If the bloodstain is on the floor or some opaque surface, its age can be estimated within a reasonable range of days. However, if it is on a curtain or a lamp shade and there is light behind it so the view is by transmitted rather than reflected illumination, it can appear dark red when, if on the floor, it would appear black. Thus, the age of the bloodstain is erroneously estimated as being much more recent that it actually is. Estimating the age of a bloodstain from its color in a photograph is especially difficult. There must be at least one other colored object in the photograph for comparison to be sure that the color rendition of that object is accurate. Still, a range must be assigned to estimate the age of the bloodstain.

Several years ago, the following was written:

> You can lead a jury to the truth, but you can't make them believe it. Physical evidence can not be intimidated. It does not forget. It doesn't get excited at the moment something is happening - like people do. It sits there and waits to be detected, preserved, evaluated, and explained. That is what physical evidence is all about. In the course of a trial, defense and prosecuting attorneys may lie, witnesses may lie, the defendant certainly may lie. Even the judge may lie. Only the evidence never lies [16].

Much of this glossary has been taken from nomenclature that was compiled by the Terminology Committee of the International Association of Bloodstain Pattern Analysts (IABPA). A suggested list of bloodstain terminology was proposed to that organization in 1996 during their annual meeting held in Albuquerque, New Mexico, with the following important qualification: "The committee's goal was to comprise a list of basic terms that are used and taught by the majority of bloodstain analyst [sic] in the field. It is offered only as a guideline, and is not designed to be all encompassing." It continued, "It is strongly recommended by this committee that Bloodstain Pattern Analysts be able to individually define terms as it [sic] pertains to their own use."

Terminology used and defined in this article is not limited to the terms on the IABPA listing. Of the 37 terms on the IABPA list 36 have been included in the above glossary. The term *impact site* is not included because it has the same meaning as does the lay term. Variations exist in many textbooks and other sources.

The terminology presented here represents significant agreement and consensus of opinion of many practitioners who have had extensive experience in this discipline. Interestingly, the great majority of what could be considered "bloodstain pattern" terms are the same ones coined in 1969 to have a working vocabulary during the law enforcement assistance administration (LEAA) original research into this discipline. At that time no one could have realized how important they would eventually become. Not all were good choices.

References

[1] Kirk, P.L. (1953). *Crime Investigation*, Interscience, New York, p. 3.

[2] MacDonell, H.L. (1971). *Flight Characteristics and Stain Patterns of Human Blood*, U.S. Department of Justice, L.E.A.A., N.I.L.E.C.J., Washington, DC, p. 77.

[3] MacDonell, H.L. (2000). *Segments of History, the Literature of Bloodstain Pattern Interpretation, Segment 00*, Literature Through the 1800's, Laboratory of Forensic Science, Elmira Heights, p. 47.

[4] Piotrowski, E. (1895). *Uber Entstehung, Form, Richtung und Ausbreitung der Blutspuren nach Hiebwunded des Kopfes*, Vircow, 95, Bl.1, Wein, p. 49.

[5] Gross, H. (1904). *Handbuch fur Untersuchungsrichter als System der Kriminalistik*, J. Schweitzed, Munchen, pp. 102–128.

[6] Orsos, F. (1938). *Die Bedeutung der Spurenkunde in der Gerichtlichen Medizin*, Orvoskepzes, Budapest, pp. 16–30.

[7] Balthazard, V., Piedelievre, R., Desoille, H. & Derobert, L. (1938). Etude des sang projete, *Annales Medecine legale De Criminologie Police Scientifique, Medecene Sociale, et Toxicologie* **19**(4), 265–323.

[8] Kirk, P.L. (1955). *Affidavit Regarding the State of Ohio v. Samuel Sheppard*, Court of Common Pleas, Criminal Branch, No. 64571, 26 April.

[9] MacDonell, H.L. (1996). Crime scene evidence – blood spatters and smears and other physical evidence, *Quinnipaic Health Law* **1**(1), 33–78.

[10] Edgerton, H. & Killian Jr, J.R. (1939). *Flash*, Charles T. Branford, Boston, p. 112.

[11] Hofmann, E.R. (1887). *Gerichtlichen Medicin*, Urban und Schwarzenberg, Wien, pp. 405–434.

[12] Underhill, H.C. (1923). *Criminal Evidence*, Bobbs-Merrill, Indianapolis, pp. 742–743.

[13] Walcher, K. (1939). *Gerichtlich-Medizinische und Kriminalistiche Blutunte-rsuchung*, J. Springer, Berlin und Heidelberg, pp. 15–33.

[14] Ziemke, E. (1914). Die untersuchung von blutspuren, in *Gerichtsarzliche und Polizeiarztliche Technik*, Th. Lochte, ed, J.F. Bergman, Wiesbaden, pp. 152–165.

[15] Anonymous (1978). *"Herbie's Angels" replace Watson*, Star-Gazette, Elmira, p. 3.
[16] Lewis, A.A. & MacDonell, H.L. (1984). *The Evidence Never Lies*, Holt, Rinehart, and Winston, New York, p. v.

Related Articles

Blunt Force Trauma
Crime Scene Investigation
Gunshot Wounds

HERBERT L. MACDONELL

Bloodstain Spatter Interpretation

see Bloodstain Pattern Interpretation

Blood-Alcohol Analysis ***see***

Alcohol: Analysis

Blood-Alcohol Concentration ***see***

Alcohol: Analysis

Blunt Force Trauma

Injuries to the Integument

Abrasions

Abrasions are superficial injuries to the skin characterized by traumatic removal, detachment, or destruction of the epidermis, mostly caused by friction. In so-called *tangential* or *brush abrasions* [1], a lateral rubbing action scrapes off the superficial layers of the skin (e.g., if the body slides across a rough surface; Figure 1) and leaves a denuded corium, which is initially covered with serosanguineous fluid. In fresh grazes, the direction of impact can often be determined by abraded epidermal shreds, which remain attached to the end of the scrape. Later, the tissue fluid dries out and forms a brownish scab. If the lesion does not reach the dermis, it heals within several days without scarring. Infliction just before or after death results in a leathery ("parchment-like") appearance with yellowish-brown discoloration.

Another type of abrasion is caused by a vertical impact to the skin (the so-called *pressure* or *crushing abrasion*). In such cases, the injuring object may be reflected by the shape of the skin injury, so that the patterned abrasion can be regarded as an imprint of the causative object (Figure 2; [2–6]). Photogrammetry may be used to obtain a three-dimensional image of an abrasion and to compare the data with an object that may have caused the injury [7].

When investigating dead bodies, it should be borne in mind that abrasions are easily overlooked as long as the epidermis-free areas have not changed their initially pink color due to drying [4]. The location of an abrasion always indicates the site of contact or impact. If the skin was affected before death, the resulting injury often consists of both an abrasion and an underlying bruise.

Fingernail scratches constitute a special group [1] and are particularly often found in sexual offenses and attacks to the neck; they can be seen both on the victims and the offenders, if the victims have defended themselves. Depending on the intensity of the abrading traumatization, the appearance of such lesions may range from superficial reddening of the skin via shallow excoriations to deep, bleeding skin lesions extending down to the corium. Terminal epidermal tags indicating the direction, from which the abrasion was caused, are easily lost and thus relatively seldom seen in surviving trauma victims. Typical features of fingernail scratches are parallel abrasions (if two or more fingernails were raked across the skin simultaneously) of roughly equal width along their course, whereas, e.g., handcuffs leave linear, much thinner, train tracklike marks on the wrists [4].

Under special circumstances, the form and location of abrasions may be indicative of a special



Figure 1 Back of a passenger who was flung out of the car in a traffic accident and suffered extensive excoriations with black particles from the asphalt surface rubbed into the skin on being dragged along the road



Figure 2 Excoriation on the left side of the neck from the safety belt (31-year-old car driver who ran away after having caused a traffic accident)

trauma; e.g., crescent-shaped fingernail marks and/or scratches in manual strangulation, linear excoriations with subsequent parchment-like drying in cases of hanging and ligature strangulation, seat belt marks in vehicle occupants after head-on collisions, human bite marks (typically, showing two semicircular bows with their concavities facing each other). In contact shots, the muzzle imprint is a patterned pressure abrasion reflecting the contours of the barrel end and other constructional parts situated near the muzzle (**Gunshot Wounds**).

Contusions

Contusions or bruises are extravasations of blood within the soft tissues originating from ruptured vessels as a result of blunt trauma (mostly caused by squeezing, but sometimes also by suction). In this context, only the contusions that are visible externally are considered. Textbooks usually differentiate between intradermal and subcutaneous bruises. In the former category, the hemorrhage is located directly under the epidermis, i.e., in the corium. This kind of superficial hematoma is usually sharply defined and red, whereas the more common bruises of the deeper subcutaneous layer have blurred edges and, at least initially, a bluish-purple color. Blood localized in the subcutis appears blue on the surface due to scattering processes in the dermis, as the blue wavelengths of the light are scattered (and thus reflected) to a greater extent than the red wavelengths [8].



Figure 3 Thoracic wall of a 37-year-old car driver with a patterned abrasion/contusion mark caused by the starlike Mercedes logo of the steering wheel



Figure 4 Intra- and subcutaneous hematoma in an infant who died from physical child abuse. The pattern of the intradermal bruises reflects the texture of the fabric covering the region of impact

Intradermal bruises may reflect the surface configuration of the impacting object (Figure 3). The skin squeezed into grooves will show intradermal bleeding, whereas the areas exposed to the elevated parts remain pale. Especially in falls from a height, in physical child abuse, or in traffic accidents, the texture of the clothing may produce a pattern of intradermal bruises corresponding to the weaving structure (Figure 4). Patterned extravasations of this type are also seen in tire tread marks, if an individual is run over by a wheel, and in bruises from vertical stamping with ribbed soles [1].

Subcutaneous bruises are usually nonpatterned. Nevertheless, there may be bruising of special significance. If the body is struck by a stick, a broomhandle, a pool cue, a rod, or another elongated instrument, every blow leaves a double "tramline" bruise consisting of two parallel, linear hematomas with an undamaged zone in between [1, 4, 6]. Human bites produce bruises forming two opposing arches (from the upper and lower jaw) with a pale center in between; often the hematomas are combined with abrasions and/or lacerations, which may sometimes reflect individual teeth characteristics (**Odontology**; [1, 2]).

Victims of blunt force violence often sustain contusions from self-defense, typically located on the ulnar aspects of the forearms and the back of the hands [9]. The upper arms may show groups of roundish bruises from fingertip pressure in cases of vigorous gripping [1]. A periorbital hematoma ("black eye") is induced by either direct indirectly (e.g., punch or kick; Figure 5) or indirectly (due to seepage of blood from a fractured orbital roof, a fractured nasal bone or from a neighboring scalp injury of the forehead; Figure 6).

In darkly pigmented individuals, contusions are sometimes invisible to the naked eye. In such cases, the use of infrared light sources can be helpful for diagnosing subcutaneous blood extravasations



Figure 5 Periorbital hematomas and excoriations in the victim of a robbery attacked with fist blows to the face. Photograph taken three days after the incident

[2]. Diaphanoscopy (also called *transillumination*) is another means to determine the presence of bruises, which otherwise would remain undetected [10].

In general, bruises are regarded as a *sign of vitality* indicating that the contusion was inflicted prior to death. During life, the blood from the ruptured vessels is forced into the soft tissue by active extravasation. Nevertheless, to a limited extent postmortem formation of contusions is possible due to passive ooze of blood. In the surviving victims, a deep bruise may not become apparent on the skin until several hours or even days later because of the slow percolation of free blood from the original site to superficial tissue layers.

In a living person, the contusion undergoes a temporal series of color changes usually proceeding from the periphery toward the center. Initially, most subcutaneous bruises appear purple-blue. As the hematoma resolves during the healing process, the hemoglobin released from the red blood cells is



Figure 6 (a) Surgically treated laceration above the left eyebrow and periorbital hematoma in a 28-year-old man who was found dead six days after sustaining the injury. The man had left the hospital against the physician's advice. (b) Purulent meningitis caused by pneumococci arising from the nasal sinuses due to a fronto-basal fracture of the skull with concomitant laceration of the dura mater. (c) Close-up view of the anterior cranial fossa with fronto-basal fracture line (arrow)

chemically degraded into other pigments such as hemosiderin, biliverdin, and bilirubin. The color changes – usually in the course of several days – to green and yellow before it finally disappears. However, the rate of change is quite variable and depends on numerous additional factors, above all the extent of the bruise.

The size of an intradermal or subcutaneous hematoma is not always indicative of the intensity of the force applied to the affected area. Elderly people or patients suffering from bleeding diathesis may get bruises from slight knocks or for other minor reasons. On the other hand, absence of an externally visible injury does not necessarily mean that there was no relevant trauma. Subcutaneous bruises of the surviving victims are often followed by gravity shifting of the hemorrhage leading to a secondary downward movement of the hematoma.

A special type of blunt injury to the soft tissues is frequently seen in pedestrians who were struck or run over by motor vehicles. Both the skin and the subcutaneous layer may be avulsed from the underlying fascia or bones by shearing forces so that a blood-filled pocket is formed, typically in combination with crush damage to the adjoining fatty tissue [6].

Lacerations

Lacerations are tears of the skin or of internal organs (see below). They may be caused by blows from blunt objects (such as a hammer, a whipped pistol [11], a rod, a toe-cap of heavy footwear, or a fist); other lacerations are produced by impact from vehicles or by a fall to the ground. Lacerations occur most commonly in body regions where the integument directly overlies a firm bony base acting as support (scalp, face, back of the hand, shins, and elbows). When the force acts on the skin, the subcutaneous tissue is squeezed between the injuring object and the bony platform so that the integument is compressed and crushed until it tears and splits sideways (Figures 7 and 8).

In contrast to incisions (**Wounds: Sharp Injury**), lacerations are characterized by abraded, bruised, and crushed wound margins. The edges of the tears are typically irregular and ragged with bridging tissue strands (vessels, nerves, and fibers) running from side to side. The wound slits may be linear (especially in blows with a narrow, edged instrument), Y-shaped, or starlike. Semicircular or crescent-shaped tears of the scalp are produced by blows with an edged instrument having a round contour such as the flat end of some hammers. If the impacting object hits the skin at an oblique angle, one of the edges will be ripped away resulting in unilateral undermining (undercutting and avulsion), which indicates the direction of the force. Sometimes foreign material from the causative instrument/surface is deposited in the depth of the wound slit. The abrasion surrounding the tear may correspond to the shape and dimensions of the impacting blunt-surfaced instrument or – in case of a fall to the ground – the area of contact (Figures 7 and 8).



Figure 7 Shaved occipital region of a 55-year-old man who fell down stairs with a blood-alcohol concentration of 320 mg per 100 ml. Slitlike laceration of the scalp surrounded by a reddish-brown excoriation



Figure 8 Parietal region of a 48-year-old man, whose scalp was shaved before autopsy, showing a stellate laceration with an eccentric abrasion (suicide by jumping from a height)

Head Injuries

The head is a common target in assaults with blunt objects; other frequent causes of head injuries are traffic accidents, falls from a height, and falls from a standing position. The area of impact usually reveals injuries of the scalp or the facial skin, but it has to be emphasized that severe and even lethal traumatization is not necessarily associated with scalp bruising, marked swelling, excoriation, and/or laceration. There may be no externally visible signs, especially in skin areas covered with cushioning hair and in cases of a fall onto a flat surface. An impact site on the vertex may suggest that the head sustained a blow, whereas in falls from standing positions the scalp injuries are expected at the level of the brim of the hat. According to autopsy studies, there are many exceptions to this "rule": only the "top of the head" lacerations (in the dorsofrontal and parietal regions) are quite uncommon in falls from a standing position [12].

Skull Fractures

These may involve the cranial vault, the base of the skull, and the facial skeleton. Though the presence of a skull fracture indicates severe traumatization, the fracture itself rarely threatens the victim's life. There are several types of skull fractures to be distinguished [1, 3–6].

Single or multiple *linear fractures* are caused either by a blow with an object having a broad, flat surface area, or by a fall on the head so that the skull is deformed (flattening/indenting at the point of impact and outward bending/bulging in the periphery). The fracture lines originate where the bone is bent outward and therefore is exposed to traction forces exceeding the limits of the bone's elasticity; from these extruded parts of the skull, the fractures extend not only toward the area of impact but also in the opposite direction. For this reason, either of the ends is often in congruity with the impact injury of the scalp (local hematoma with/without concomitant abrasion and/or laceration, Figure 9a). Several fracture lines may radiate outward from a central point of impact where the skull is often depressed and/or shattered to pieces forming a *spider's web* or mosaic pattern consisting of radiating and circular fractures. The sequence of skull injuries may be determined according to Puppe's rule: a later fracture does not cross a preexisting fracture line, but terminates when reaching an earlier one.

Before fusion of the cranial sutures (i.e., in children and young adults), a fracture may travel along the seam resulting in diastasis (*diastatic fractures)*. If a gaping fracture runs from one side of the cranial base to the other (mostly after lateral impact or side-to-side compression), this transverse type is called a *hinge fracture* (Figure 10) because of the independent movement of the front and rear halves of the skull base.

Longitudinal fractures of the base of the skull frequently occur due to a fall on the occiput; in such instances, the linear fractures typically run through the posterior fossa either ending near the foramen

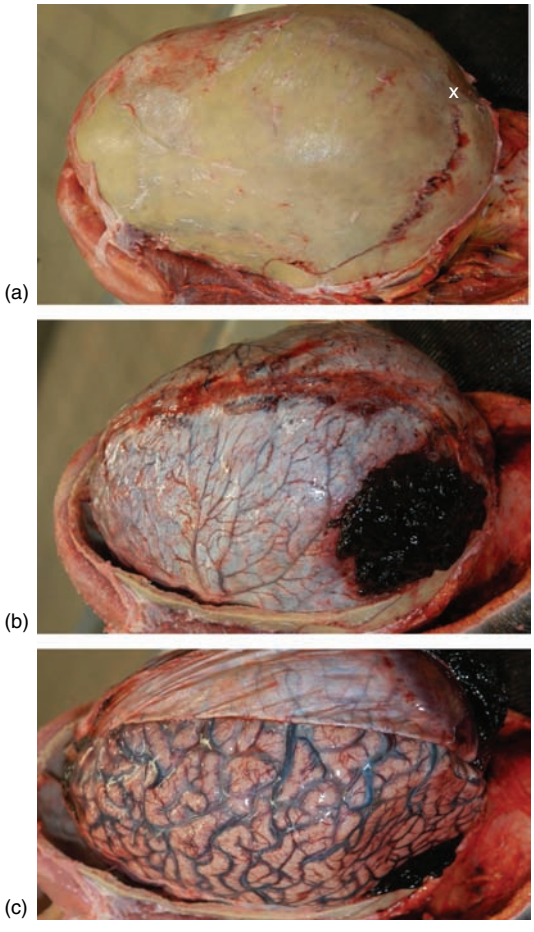


Figure 9 (a) Left half of the cranial vault with linear fracture of the skull. The dorsal end is in congruity with the area of impact (x). After a fall on the occiput, a 51-year-old man suffered from a headache, but did not seek medical treatment. He was found dead three days later. (b) Extradural (epidural) hemorrhage in the parieto-occipital area. The hematoma is located between the inner surface of the skull and the detached dura. (c) Concave impression of the brain surface after removal of the space-occupying epidural hematoma. The dura mater was reflected upward during autopsy



Figure 10 Transverse (side-to-side) "hinge fracture" dividing the skull into a front and a rear half

magnum or extending to the floor of the middle and anterior fossa. On the other hand, longitudinal fractures of the base can also be produced by impaction of the frontal region.

Depending on its course and location, a *base fracture* may be followed by several *clinical signs*: bleeding from the ear (in fractures of the temporal bone with concomitant hematotympanum and rupture of the eardrum); bleeding from the nose and mouth (in fractures involving paranasal sinuses, which provide a communication with the nasopharynx); periorbital hematoma (from fractures of the orbital roofs); leakage of cerebrospinal fluid coming out of the nose or the ear (if the *dura mater* is injured along the fracture); bacterial infection of the meninges (by spread from the nasal cavity, the paranasal sinuses and the middle ear, especially when the fracture is accompanied by a tear of the *dura mater*; cf. Figure 6).

Some special types of skull fractures can only be mentioned briefly. A *ring fracture* is located in the posterior fossa and encircles the foramen magnum. It mostly occurs in a fall from a height onto the victim's feet or buttocks, so that the cervical spine is driven into the skull. Another mechanism takes effect in high-speed head-on collisions: the frontal impact

abruptly decelerates the vehicle and the trunk of belt-restrained passengers, whereas the head continues to move in the previous direction due to inertia resulting in hyperextension/hyperflexion of the craniocervical junction. Ring fractures of the skull base are also seen in motorcyclists [13].

Bone impressions and *depressed fractures* are always localized at the point of impact where the head is struck with an object having a relatively small surface area such as a hammer, a protruding corner of a piece of furniture or another edged implement (Figure 11). The outline of a clean-cut defect in the outer table may reproduce the shape and size of an edged instrument, for instance, in cases of pistol-whipping. If only limited force is applied, the depressed fracture can be restricted either to the outer or, less often, to the inner table of the skull (the latter with inward displacement of the bone fragments). A depressed fracture from a blow striking the skull cap at an angle may be concentrically terraced.

The diaphanoscopic postmortem examination of blunt impact injuries to the head sometimes reveals nondiaphanous regions derived from intraossary hematomas. Their location coincides with the corresponding injuries of the scalp such as contusions and lacerations so that conclusions can be drawn as to the area of impact, even if, owing to postmortem changes, the scalp is no longer assessable [14].

Hole fractures from bullets perforating a flat bone of the skull are mostly roundish and clean cut at the site of entrance, but beveled out in a craterlike manner at the exit site (**Gunshot Wounds**).

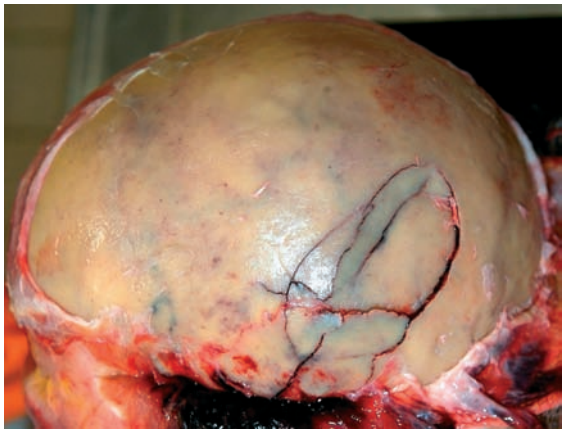


Figure 11 Depressed fracture of the left parieto-temporal region sustained in a fall down a staircase hitting the edge of a step

Blunt force applied to the occiput, mostly as a consequence of a fall on the back of the head, frequently causes independent fractures of the anterior cranial fossa such as cracks of the thin orbital roofs (*secondary fractures)* at the site of the contrecoup [3].

A fracture of the base of the skull involving an air sinus (e.g., in the sphenoid bone) provides a communication between the cranial cavity and the nasopharynx permitting the passage of blood into the airways. If the victim has lost consciousness, aspiration of blood into the lungs will be the consequence, possibly followed by fatal asphyxiation. The same complication can be seen after blunt traumatization of the visceral cranium.

Intracranial Hemorrhages

A space-occupying bleeding under/over/between the brain membranes is followed by local displacement of the brain and raised intracranial pressure with concomitant flattening of the cerebral hemispheres. Intracranial hematomas as well as traumatic brain swelling, which often accompanies head injuries, may result in transtentorial (uncal) herniation (in cases of supratentorial mass lesion) and/or herniation of the cerebellar tonsils which are forced into the foramen magnum leading to compression of the brainstem with secondary damage and failure of the medullary respiratory centers.

From the clinical and forensic point of view, the possible occurrence of a so-called *lucid* or *latent* interval has to be mentioned. After initial unconsciousness (due to cerebral concussion), there may be a symptomless period of several hours or even days before the victim becomes comatose again because of the increased hemorrhage leading to elevated intracranial pressure.

Epidural (extradural) hemorrhages are located between the skull and the underlying *dura mater*, which is stripped from the bone by bleeding from a torn vessel (Figure 9b, c). Epidural hematomas have a typical disk- or lens-shaped appearance, which can be demonstrated by means of pre- or postmortem computed tomography (CT) scans (**Radiology**; Figure 12; [15–17]). The most common site is the temporal and the adjacent parietal region where the branches of the middle meningeal artery are easily lacerated in the course of a transecting fracture

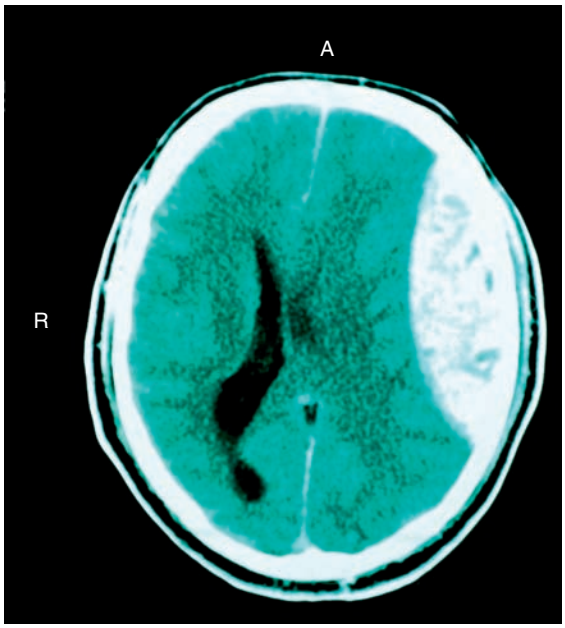


Figure 12 Computed tomography (CT) scan of a 39-year-old man showing a large left-sided extradural hematoma with concomitant shift of the midline to the right and brain swelling. The victim was hit by a fist blow to the face and fell on the back of the head. The extradural hematoma was associated with frontal contusions. The patient underwent neurosurgical craniotomy with removal of the space-occupying blood accumulation and survived with only slight neurological deficits

line. Since the well-adherent dura mater has to be avulsed from the bone, epidural hematomas originate more often from arterial bleeding than from venous bleeding (e.g., due to a torn dural sinus). In the great majority, an extradural hemorrhage is associated with a cranial fracture.

An antimortem traumatic extradural hematoma must not be confused with a postmortem epidural thermal hematoma [18]. The latter one is seen in severely burnt bodies with an (at least partially) charred skull. It is caused by a heat-induced shift of bloody fluid from the diploe and the venous sinuses. The thermal hematoma has a crumbly or fatty appearance and a brick-red or brown color (**Injury: Burns, Scalds, and Chemical**).

Subdural hematomas are intracranial bleedings located beneath the dura mater and above the arachnoid (Figure 13). Most often the hemorrhage arises

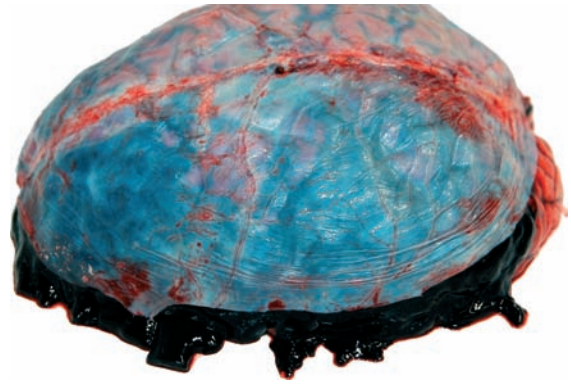
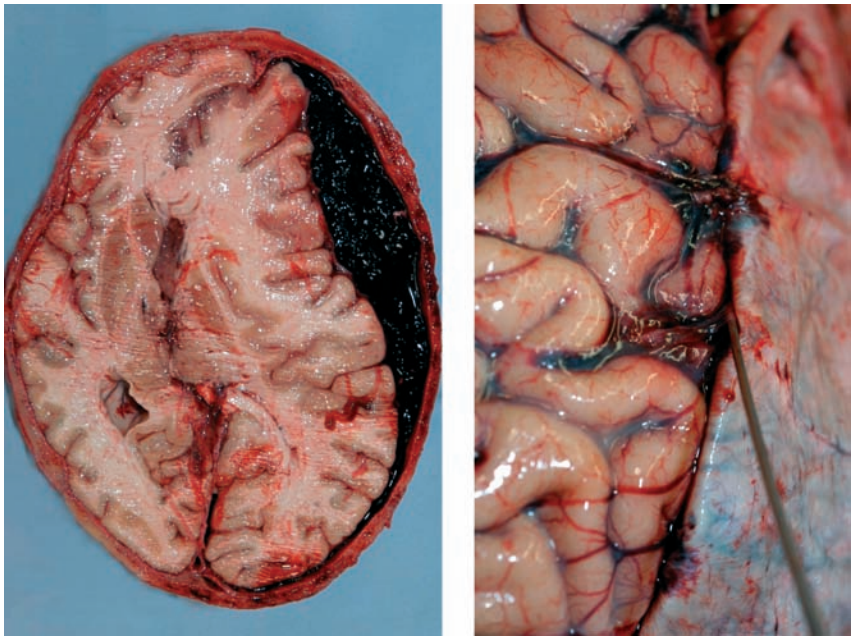


Figure 13 Aspect of the left cerebral hemisphere still covered with the dura which is discolored by an underlying hematoma arising from torn bridging veins

from the tearing of overstretched bridging veins that traverse the subdural space between the surface of the cerebral hemispheres and the superior sagittal sinus (Figure 14). Other possible sources of subdural bleeding are injuries to venous sinuses or to the cerebral parenchyma (such as cerebral contusions with concomitant laceration of the arachnoid). The subdural hemorrhage usually covers one cerebral hemisphere in a caplike manner from the parasagittal area via the lateral surface down to the basal fossas; on a horizontal section, it appears as a sickle-shaped accumulation of blood.

In contrast to epidural hematomas, subdural hemorrhages are often not associated with skull fractures; additional damage to the brain tissue may also be absent. A high percentage of subdural bleedings are caused by acceleration or deceleration of the head, for instance, not only in falls when the head impacts a hard surface but also in traffic accidents and physical child abuse (battered child and shaken baby syndrome, **Battered Child Syndrome**; **Shaken Baby Syndrome**). Apart from acute and subacute subdural hemorrhages, there are prolonged cases of hematoma formation and organization, mainly in elderly people and sometimes without a history of previous traumatization. Such chronic subdural hematomas typically consist of brown and gelatinous blood accumulations adherent to the meninges and sometimes covered with a tough membrane [1].

Traumatic subarachnoid bleeding may result from damage to the cortex such as brain contusion (e.g., contrecoup lesions), from penetrating injuries to



(a) (b)

Figure 14 (a) Acute subdural hemorrhage. The unilateral space-occupying hematoma has markedly shifted the midline of the brain to the opposite side. (b) Close-up view of torn bridging veins (source of the subdural bleeding). Probe inserted into a ruptured vessel

the brain and as a consequence of vessel tears within the subarachnoid space. An extensive hemorrhage in the basal cisterns and around the brainstem may arise from a laceration of an artery belonging to the circle of Willis or from another great vessel (such as a torn basilar and vertebral artery).

In sudden deaths without any preceding trauma, subarachnoid hemorrhage is most often due to a spontaneous rupture of a so-called berry aneurysm, which is typically located at bifurcations and branches of cerebral arteries belonging to the "circle of Willis" [1]. The blood extravasation usually fills the basal cisterns and then spreads laterally over the cerebral hemispheres. Sometimes the question arises whether a preexisting aneurysm was mechanically ruptured by a blunt head injury.

Cerebral Injuries

"Concussion of the brain" is a clinical diagnosis which means a disorder of cerebral function following immediately upon a (blunt) head injury. It is usually characterized by a transient loss of consciousness (initial coma) with subsequent amnesia from the actual moment of trauma; it is often combined with retrograde amnesia and vegetative signs such as nausea and vomiting. In mere concussions, the unconsciousness lasts only for a relatively short time (less than 1 h) and the brain tissue does not show any evidence of structural damage. Nevertheless even a simple cerebral concussion may be followed by the victim's death, if the head trauma is joined by interfering mechanisms (for instance, drowning or aspiration of gastric contents during unconsciousness).

Cerebral contusions are traumatic lesions of the brain frequently seen in the cortex and sometimes extending into the underlying white matter [19, 20]. Fresh contusion hemorrhages are mostly located on the crests of the gyri and composed of grouped streak-like or punctate blood extravasations. The cortical lesions are often covered with subarachnoid bleeding. In contrast to cerebral contusions, the term "laceration" means a major destruction of the anatomical context (for instance, mechanical separation of the tissue due to bone fragments or penetrating bullets).

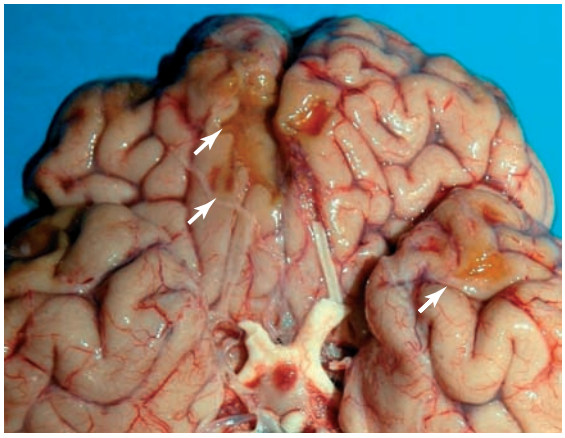


Figure 15 Undersurface of the brain in a 46-year-old alcohol addict who died from hypothermia. Autopsy accidentally revealed survived contusion hemorrhages with a yellowish-brown appearance and softening of the tissue (right frontal lobe and left temporal lobe, arrows)

In case of survival, the contusion hemorrhages are reabsorbed and assume a yellowish-brown appearance with softening and finally liquefaction of the affected areas (Figure 15).

Due to the injuring mechanism, most cerebral contusions occur in brain regions that are directly opposite to the point of impact. This contrecoup type of contusion is classically caused by a fall on the occiput, when the moving head is suddenly decelerated with the consequence that the inlying brain is damaged due to inertia. In falls on the back of the head, the contrecoup areas of the brain (poles and undersurfaces of the frontal and temporal lobes) are subjected to an ultrashort negative pressure ("cavitation") resulting in vessel ruptures and cortical hemorrhages (Figures 16 and 17). On the other hand, the so-called coup contusions arise at the area of impact due to the local deformation and compression of the brain. Even severe coup and contrecoup injuries are not necessarily associated with skull fractures. In victims with both the coup and contrecoup lesions, the degree of contrecoup damage is usually more pronounced. Fracture contusions are localized in topographical correspondence to fracture lines and/or depressed fractures.

Severe head traumas are often followed by posttraumatic epilepsy, especially in victims with penetrating head injuries, depressed skull fractures, dural tears, intracranial hematomas, and structural brain damage. Early posttraumatic seizures occur within the first week, whereas late posttraumatic epilepsy develops at any point thereafter [16, 20].



Figure 16 Contrecoup contusions on the undersurface of the frontal lobes and the right temporal pole with minor subarachnoid hemorrhage

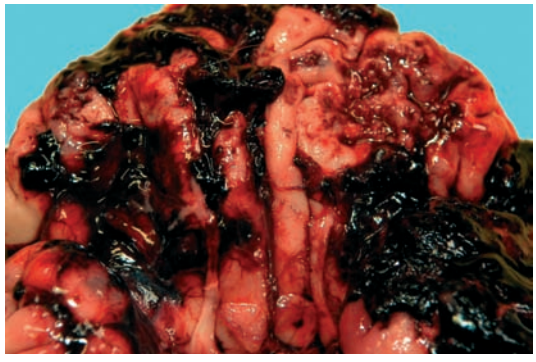


Figure 17 Contrecoup contusions on the undersurface of the frontal lobes (fall on the back of the head, survival time one day)

Diffuse axonal injury (DAI) and the accompanying tissue tear hemorrhages are considered a consequence of shear and tensile strains from sudden acceleration/deceleration or rotational movements of the head. Overstretching of the nerve fibers in the white matter leads to axonal injury varying from temporary dysfunction to anatomical transection, the latter being followed by microscopically visible club-shaped retraction balls on the axons. The sites of predilection include the corpus callosum, the parasagittal white matter, the superior peduncles, and the rostral brainstem. In the course of the repair process, microglial cells proliferate in the areas of axon damage. In victims of substantial head injuries, especially after traffic accidents, diffuse axonal injuries may be responsible for prolonged coma and a fatal outcome even in the absence of an intracranial mass lesion. Macroscopically, DAI is characterized by small focal hemorrhages and salient cerebral swelling [19, 20].

Cerebral edema is a frequent finding in significant head injuries. The formation of edema is due to an increase in the fluid content of the brain, predominantly in the white matter. Posttraumatic edema may be generalized (diffuse) or related to focal tissue damage (e.g., adjacent to an area of cerebral contusion or laceration). At autopsy, the weight of the brain is increased; the gyri are pale and flattened with shallow sulci in between. From the pathogenetic point of view, edema is attributed to a heightened vascular permeability which in turn may be worsened by additional hypoxia.

As with space-occupying bleedings such as subdural or epidural hematomas, brain swelling is a common cause of raised intracranial pressure. The enlarged volume of the edematous brain leads to a displacement of cerebral tissue downward through the midbrain opening resulting in grooving of the unci and/or hippocampal herniation. Expansion of the subtentorial brain leads to herniation of the cerebellar tonsils which are forced into the foramen magnum. Herniation with concomitant compression of the brainstem may be followed by secondary hemorrhages (localized in the midbrain and pons) and finally by lethal dysfunction of vital centers.

Injuries of the Chest

Nonpenetrating blunt force may damage the thoracic wall and/or the chest organs [1, 4, 5]. *Rib fractures* are caused either by direct or by indirect violence. In the first case, a localized force is applied and the underlying ribs are broken in the contact area (for instance, along the seat belt in head-on collisions); the other (indirect) type of rib fracture occurs away from the impact, mainly due to compression of the chest.

Rib fractures are frequently associated with *complications* that may be dangerous or even life threatening.

- If a victim sustains numerous fractures, the rib cage loses its rigidity so that the injured section of the chest wall will not participate in the expansion of the thorax during inspiration with the result of paradoxical respiration (*flail chest)* and concomitant hypoxia.
- Sharp, pointed ends of the rib fragments may penetrate the pleura and lacerate the lung and/or the intercostal blood vessels with consecutive bleeding into the chest cavity (*hemothorax).*
- A leak in the visceral pleura or a penetrating injury of the thoracic wall permits air to enter the pleural cavity (*pneumothorax)* so that the lung collapses, if it is not fixed to the chest wall by preexisting pleural adhesions. A valvelike leakage in the pleura leads to a so-called tension-pneumothorax caused by an increasing pressure of trapped air in the pleural cavity and followed by a complete collapse of the affected lung and a shift of the mediastinum to the opposite side.
- The presence of air bubbles in the subcutis (Figure 18) or in the mediastinum (*subcutaneous/mediastinal emphysema*) may derive from injuries of the trachea, the bronchi, the thoracic wall or the lungs by air entering the adjacent soft tissues.

Blunt force injuries to the *lung* are mainly encountered as contusions or lacerations. A contusion is typically caused by a substantial impact on the chest with consecutive inward bending of the thoracic cage. In young victims, contusions are not necessarily accompanied by fractures of the ribs or of the sternum because of the high pliability of the juvenile thoracic cage. From the morphological point of view, a contused lung shows bruising either as a subpleural suffusion (Figure 19) or as an intrapulmonary hemorrhage. Lacerations of the lung can result when a severe compressive or crushing force is applied to the chest so that the pulmonary tissue bursts or tears.



Figure 18 Subcutaneous emphysema of the face in a traffic accident victim (massive blunt thoracic trauma, severance of a large bronchus)

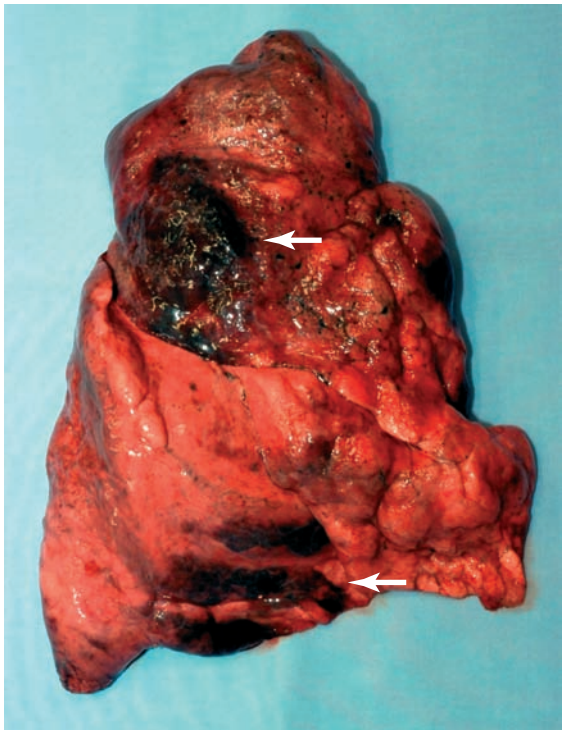


Figure 19 Right lung of a 44-year-old car driver who sustained a blunt trauma to the thorax with contusions of the pulmonary tissue (arrows)

Another possible mechanism is inward displacement of a fractured rib which impales the lung.

Blunt traumatization of the *heart* manifests itself as concussion, contusion, or myocardial rupture. In most cases, the force is directly applied to the anterior chest, which compresses or crushes the heart between the sternum and the vertebral column. Bruises of the cardiac wall may be localized in the subepicardial fatty tissue (sometimes in combination with posttraumatic coronary occlusion) or within the myocardium, which then appears dark red from interstitial hemorrhage. Lacerations of the heart are most often seen in the relatively thin right ventricle (Figure 20) or in the atria; they are less common in the left ventricle, the papillary muscles, the cardiac valves, the interatrial, and interventricular septum.

The risk of cardiac rupture is especially high during diastole, when the heart chambers are filled with blood and therefore easily burst on being exposed to a sudden compressive force. Such injuries usually have a fatal outcome either from massive blood loss and hemorrhagic shock (if the pericardial sac is torn and the blood pours into the pleural cavity) or from cardiac tamponade (blood accumulation in the pericardial sac resulting in insufficient filling of the

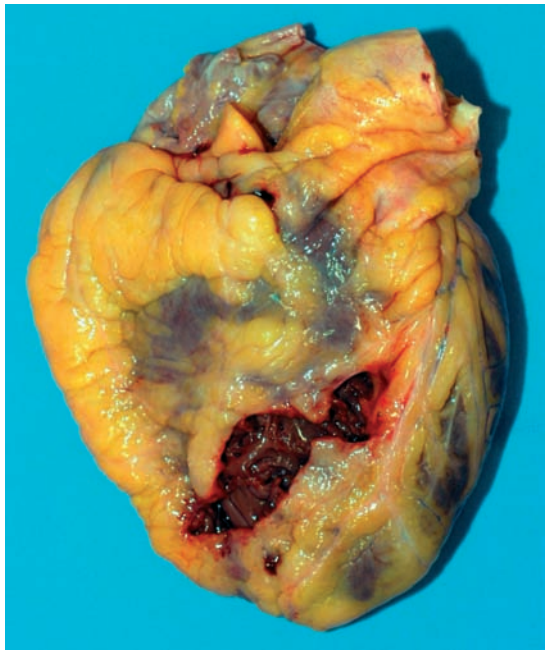


Figure 20 Anterior aspect of the heart of a 37-year-old motorcyclist who died after a head-on collision with a car. Traumatic laceration of the right ventricular wall

cardiac chambers and impaired forward circulation). In very rare cases, the traumatic heart rupture may be delayed, if the blunt chest trauma first led only to a tear of the inner myocardial layers, which later perforates into the pericardial sac [21].

Traumatic aortic ruptures typically occur in vehicular accidents and in falls from a height. The most important mechanism is sudden deceleration, possibly in combination with compression and/or shearing. Traction forces tear the aorta transversely at two sites of predisposition: in the descending part of its arcus (near the attachment of the ligamentum arteriosum; Figure 21) or immediately above the cusps of the aortic valve. Other locations (for instance, in association with a dislocated vertebral fracture) are rather rare. The laceration of the aorta may occur either as a complete or partial transection. In the latter case, the outer layers of the vascular wall are not damaged; the intimal tears are often multiple, semicircular, and parallel (so-called ladder-rung tears). If the trauma is survived at least for a short time, a parietal thrombosis or a posttraumatic aneurysm may follow as secondary complications.

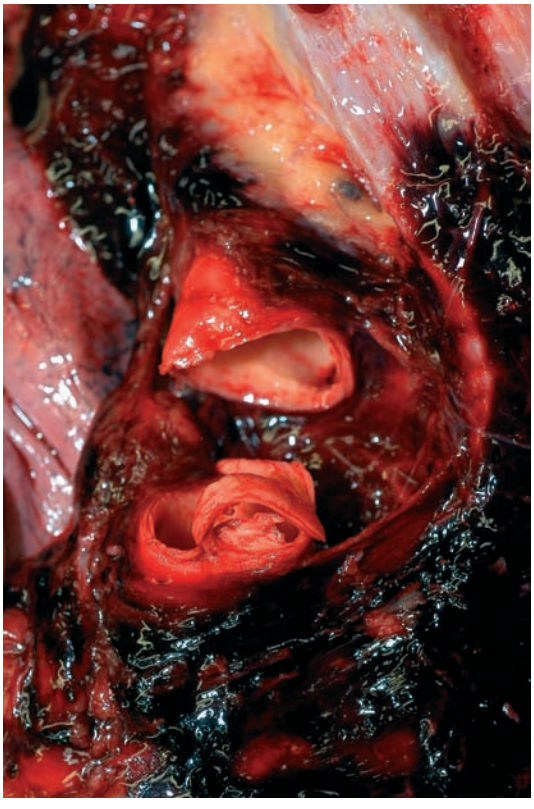


Figure 21 Traumatic severance of the aorta in the distal part of the arch (deceleration injury) surrounded by perivasal hematoma of the mediastinum

Abdominal Injuries

Blunt force injuries of the abdomen are frequently seen in traffic and work accidents, in child and spouse abuse, in other criminal assaults (with kicking, stamping, and punching), and also in suicidal falls from heights. The abdominal organs most vulnerable to blunt trauma are the solid liver and spleen on the one side and the mesentery on the other. Concomitant external signs of blunt traumatization such as contusions or abrasions are by no means obligatory.

Substantial injuries to the liver, the spleen and the mesentery always have to be regarded as life threatening and potentially fatal, especially in cases without rapid surgical treatment. The main reason is internal bleeding into the peritoneal cavity from lacerations. Ruptures of the liver and spleen can be classified as either transcapsular or subcapsular laceration. In the first case, both capsule and parenchyma are injured (Figures 22 and 23) so that the blood instantaneously pours into the peritoneal cavity. The second type of laceration is characterized by the initial formation of a subcapsular hematoma which expands continuously and may cause a delayed rupture when the covering capsule tears due to overstretching (mostly several hours or even days after the trauma).



Figure 22 Ruptures of the right hepatic lobe

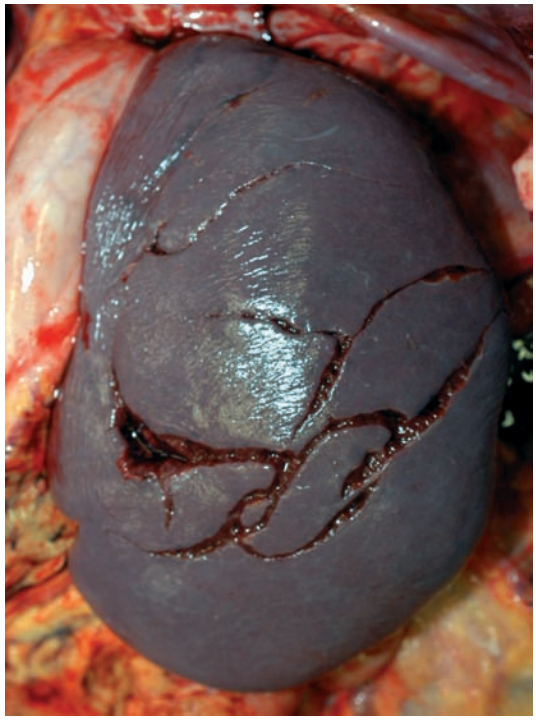


Figure 23 Laceration of the spleen

The stomach and the intestine are less susceptible to blunt traumatization than the parenchymatous abdominal organs. The hollow viscera are more likely to rupture, if they are filled with food or fluid. Another reason why the intestine or stomach might be prone to damage is squeezing of the organs between the indented abdominal wall and the lumbar vertebrae. Fatal outcomes from contusions or lacerations of the gastrointestinal tract are usually due to diffuse peritonitis.

Renal injuries are a relatively rare source of severe bleeding since the kidneys are deeply located behind the peritoneum. Nevertheless, they can be ruptured by a heavy impact to the loin (for instance, in traffic accidents or assaults).

Although the empty urinary bladder is placed within the pelvis, when filled it moves upward and is therefore exposed to blunt traumatization of the lower abdomen. Consequently, rupture of the empty bladder is expected to be extraperitoneal and accompanied by pelvic fractures, whereas a bladder distended with urine may rupture into the peritoneal cavity.

Blunt traumatization of a pregnant uterus is a possible cause of fetal death, mostly due to separation or rupture of the placenta.

Injuries to the Extremities

Apart from injuries to the skin and the subcutaneous layer (see above), other anatomical structures such as the muscles, the bones, and the joints may be involved in blunt force trauma. Extensive crushing of the soft tissues, the formation of blood-filled cavities, comminuted fractures, and severance of large vessels are frequent findings in victims of automobile-pedestrian accidents. Internal bleeding (from closed injuries) and external bleeding (from traumatic amputation, severe avulsive wounds and compound fractures) are important factors contributing to hemorrhagic shock. Another sequel to blunt trauma is pulmonary and systemic fat embolism (caused by globules of fat, usually subsequent to fractures or damage of fatty tissues). In cases of prolonged survival, intercurrent infection, sepsis-associated multiorgan failure and pulmonary thromboembolism (originating from posttraumatic venous thrombosis) are dangerous and often fatal complications of an originally nonlethal injury (e.g., fracture of the femur neck).

References

[1] Saukko, P. & Knight, B. (2004). *Knight's Forensic Pathology*, 3rd Edition, Arnold, London, pp. 136–221.

[2] Smock, W.S. (2000). Recognition of pattern injuries in domestic violence victims, in *Encyclopedia of Forensic Sciences*, J.A. Siegel, P.J. Saukko & G.C. Knupfer, eds, Academic Press, London, Vol. 1, pp. 384–391.

[3] Pollak, S. & Saukko, P.J. (2000). Blunt force, in *Encyclopedia of Forensic Sciences*, J.A. Siegel, P.J. Saukko & G.C. Knupfer, eds, Academic Press, London, Vol. 1, pp. 316–325.

[4] Spitz, W.U. (2006). Blunt force injury, in *Spitz and Fisher's Medicolegal Investigation of Death*, 4th Edition, W.U. Spitz, ed, Thomas, Springfield, pp. 460–531.

[5] DiMaio, V.J. & DiMaio, D. (2001). *Forensic Pathology*, 2nd Edition, CRC Press, Boca Raton, pp. 92–185.

[6] Pollak, S. & Saukko, P. (2003). *Atlas of Forensic Medicine (CD-ROM)*, Elsevier, Amsterdam.

[7] Brüschweiler, W., Braun, M., Fuchser, H.J. & Dirnhofer, R. (1997). Photogrammetric evaluation of injuries to skin, soft tissue and bones to determine the weapon – basic aspects, *Rechtsmedizin* **7**, 76–83.

[8] Bohnert, M., Baumgartner, R. & Pollak, S. (2000). Spectrophotometric evaluation of the colour of intra- and subcutaneous bruises, *International Journal of Legal Medicine* **113**, 343–348.
[9] Pollak, S. & Saukko, P. (2003). Defense wounds, in *Encyclopedia of Forensic Sciences*, J.A. Siegel, P.J. Saukko & G.C. Knupfer, eds, Academic Press, London, Vol. 1, pp. 374–378.
[10] Horisberger, B. & Krompecher, T. (1997). Forensic diaphanoscopy: how to investigate invisible subcutaneous hematomas on living subjects, *International Journal of Legal Medicine* **110**, 73–78.
[11] Pollak, S., Missliwetz, J. & Mortinger, H. (1989). Pistols and revolvers as striking tools, *Archiv für Kriminologie* **184**, 30–37.
[12] Maxeiner, H. & Ehrlich, E. (2000). Site, number and depth of wounds of the scalp in falls and blows – a contribution to the validity of the so-called hat brim rule, *Archiv für Kriminologie* **205**, 82–91.
[13] Maeda, H., Higuchi, T., Imura, M., Noguchi, K. & Yakota, M. (1993). Ring fracture of the base of the skull and atlanto-occipital avulsion due to anteroflexion on motorcycle riders in a head-on collision accident, *Medicine, Science and the Law* **33**, 266–269.
[14] Hellerich, U. & Pollak, S. (1991). Significance of intraosseous skull cap hematoma for reconstruction of skull injuries, *Beiträge zur Gerichtlichen Medizin* **49**, 33–37.
[15] Marks, P. (2005). Deaths: trauma, head and spine, in *Encyclopedia of Forensic and Legal Medicine*, J. Payne-James, R.W. Byard, T.S. Corey & C. Henderson, eds, Elsevier Science, Amsterdam, Vol. 2, pp. 75–81.
[16] Marks, P. (2005). Head trauma: pediatric and adult, clinical aspects, in *Encyclopedia of Forensic and Legal Medicine*, J. Payne-James, R.W. Byard, T.S. Corey & C. Henderson, eds, Elsevier Science, Amsterdam, Vol. 2, pp. 461–472.
[17] Thali, M.J. & Vock, P. (2003). Role and techniques in forensic imaging, in *Forensic Medicine: Clinical and Pathological Aspects*, J. Payne-James, A. Busuttil & W. Smock, eds, Greenwich Medical Media, London, pp. 731–745.
[18] Bohnert, M. (2004). Morphological findings in burned bodies, in *Forensic Pathology Reviews*, M. Tsokos, ed, Humana Press, Totowa, Vol. 1, pp. 3–27.
[19] Case, M. (2005). Head trauma: neuropathology in *Encyclopedia of Forensic and Legal Medicine*, J. Payne-James, R.W. Byard, T.S. Corey & C. Henderson, eds, Elsevier Science, Amsterdam, Vol. 2, pp. 472–480.
[20] Oehmichen, M., Auer, R.N. & König, H.G. (2006). *Forensic Neuropathology and Associated Neurology*, Springer, Berlin, pp. 97–270.
[21] Pollak, S. & Stellwag-Carion, C. (1991). Delayed cardiac rupture due to blunt chest trauma, *American Journal of Forensic Medicine and Pathology* **12**, 153–156.

Related Articles

Battered Child Syndrome
Gunshot Wounds
Injury: Burns, Scalds, and Chemical
Odontology
Radiology
Shaken Baby Syndrome
Wounds: Sharp Injury

STEFAN POLLAK AND PEKKA J. SAUKKO

Bomb Scene Management

Introduction

An explosion is the rapid production of gas from liquid or solid usually accompanied by the production of heat, light and sound. The reaction is very efficient and the end products usually bear no resemblance to the ingredients. Sometimes traces of unreacted explosive and associated bomb components can be recovered from the scene. Explosions can occur without criminal involvement for example in industrial accidents where gaseous fuels or organic dusts like flour form explosive mixtures with air and explode with devastating effects. Typically, boiling liquid and expanding vapor explosions (BLEVEs) represent a significant hazard to firemen extinguishing large gas tank fires. Explosives are used by criminals for a wide range of purposes, including intimidation. Bombs remain the favored tool of terrorists because of their ability to create chaotic scenes that are visually dramatic and newsworthy.

The role of all first responders is primarily to save life, then to preserve the scene. The focus then moves to the creation of order from chaos. The objective of an investigation is to identify the cause, clarify the facts, determine whether a criminal act has been committed, and if so identify perpetrators and bring the matter before the Court with the ultimate purpose of protecting society.

The investigators' questions will include the five "W's" (What, Where, When, Why, Who) and How, plus permutations thereof, for example; *What* was the target? *What* was the intended target? *What* was the bomb made from? And so on. If inductive reasoning suggests criminal involvement, deductive reasoning may be used to narrow the field of suspects. In most crimes there are usually a lot of people who may have a motive; however, fewer have both motive and opportunity, and very few have the motive, opportunity, and the means to commit the crime. Physical evidence, sometimes in the form of invisible traces that can be linked with both the scene and suspect can provide strong evidence of the suspect having the "means" to commit the crime (See Figure 1).

The forensic forward commander contributes to the investigation by coordinating the activities of those involved in the recording, collection, preservation, and evaluation of the potential evidence.



Figure 1 The suspect's triangle

The investigators may use physical evidence to identify a suspect; however, it should be appreciated that much of the power of science lies in exclusion. The role of the scientist changes from establishing facts to challenging the proposition that implicates the suspect. The jurors need to consider the robustness of the challenge when determining their findings.

Crime Scene Movements

A successful bombing on a civilian target results in pandemonium. Inevitably a bomb scene will have a number of professional groups transitioning through it and each has the potential to diminish or destroy evidence. Controlling this traffic at the earliest practical opportunity will help mitigate loss of valuable evidence (See Figure 2).

The forensic forward commander usually arrives at the scene some hours after the explosion. A great deal of change is likely to occur in that period; however, little can be done about this. The period between notification and arrival does provide a valuable window to mentally prepare oneself, collate information, scope the magnitude of the task, and consider possible resource implications such as how to sustain a lengthy or multifaceted investigation. It is important to record decisions and their rationale.

There are a number of actions that should be initiated in the immediate aftermath to protect the crime scene and to capture potential evidence that would otherwise be ephemeral. A good example of this was the actions of police in meeting victims of the first Bali bombings at the Australian airports. First hand accounts and the examination of their clothing yielded evidence, including the confirmed identification of trinitrotoluene (TNT) from those who were in Paddy's Bar at the time.

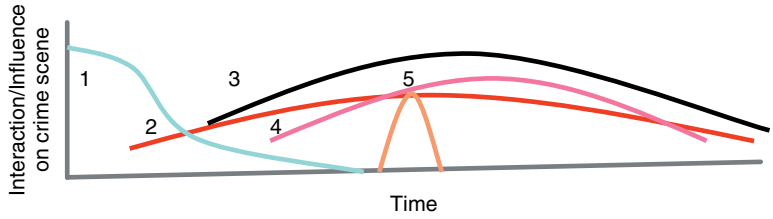


Figure 2 Depicting movements through the crime scene. 1, the nonambulatory victims and the general public; 2, first responders–emergency services–fire brigade–bomb technicians; 3, investigators, forensic forward commander, specialists, and scene recorders; 4. disaster victim identification – body recovery; 5. coroner and other administrators of governance

The main purpose for taking control of the scene is to protect evidence until it can be captured in a thorough examination. It is important to reduce the risk of contamination and ensure that the integrity of all the evidence is maintained from crime scene to courtroom.

There are a number of activities that can help reduce the loss of potential evidence.

- Only those persons with legitimate need should have access to the scene. Strategies may need to be implemented to discourage others. An entry log that records time and date, identity, intended purpose, and their potential availability as a witness should be created.
- The wearing of protective clothing such as overalls/Tyvek suits in the primary scene sends a strong message that the scene's integrity is of utmost importance to the forensic practitioners.
- Actions such as defining the locations for the inner and outer perimeters, and arranging for them to be marked with barrier tape and guarded will help control traffic.
- Under calm weather conditions explosive residues will not travel more than about 60 m from the epicenter, unless carried by primary fragmentation [1]. Buildings and landscape will define the shape of the cordoned areas, but as a general rule, the inner perimeter should be one and a half times (1.5×) the distance from the epicenter of the outermost piece of primary fragmentation. The outer perimeter is a convenience to keep the public and media out of the working area.
- Some of the main ingredients in improvised explosives are soluble in water. Therefore it may be important to initiate actions to protect against moisture damage, e.g., erecting tents over areas of high interest, and covering areas with plastic sheeting.
- All potential evidence should be collected as soon as possible, packaged in appropriate pristine containers at the scene, sealed, labeled, and kept cool. They should be conveyed to the laboratory at the earliest practical opportunity.
- An effective team has defined roles for its members. The size and complexity of a Bomb Scene Examination team may vary according to jurisdiction, the size of the incident, and the available resources. Typical team roles include a leader, a bomb technician, a photographer/recorder, search members, and an exhibits officer. A common observation of effective teams is that their members "think evidence and not just exhibits". In other words, experience and good judgement are particularly desirable attributes especially under challenging circumstances such as threatening inclement weather.

Coordination of the Various Forensic Disciplines

The forensic forward commander reports to the police forward commander and is responsible for

- coordinating the activities of all the various forensic specialists, including the disaster victim identification (DVI) teams; and
- working with local authorities in charge of utilities such as transport, health, electricity, gas, business councils, and the like.

Although DVI is important in its own right, it is also useful in the immediate investigation. The location of victims and the nature of their injuries can provide important information about the bombing. Similarly, victims and deceased persons often become repositories of primary fragmentation. Explosive residues were identified from the clothing of witnesses to the first Bali bombings, and shrapnel recovered from the wounds of a living victim at the second Bali bombing confirmed the presence of TNT by two independent techniques gas chromatograph with thermal energy analyser and negative chemical ionisation gas chromatograph with mass spectral analyser (GC-TEA and NCIGC-MS) (See Figure 3).

It is very important for the forensic forward commander to include the specialists such as pathologists in comprehensive briefings, as their ability to recognize potential evidence will be enhanced if they are aware of its relevance in the crime scene.

The pathologist plays a vital role in bomb scene investigations, not only in recovering evidence, identifying victims, determining the nature of injuries and cause of death, but also in identifying the perpetrator in suicide bombings. There is a natural tendency in major bombing cases to focus on whole bodies first in order to make rapid progress. This is a potential mistake because if suicide is involved, the perpetrator will be in small parts and usually mixed with the bodies of other victims. Decomposition and exposure



Figure 3 Metal objects recovered from wounds of living victim. (Note TNT was confirmed by GC-TEA and NCIGCMS)

to extraneous body fluids will make DNA analysis more difficult. If suicide is suspected, the investigator's priority is to have the small parts examined first so that a DNA profile can be collected and compared with the suspect's direct relatives.

Contamination Minimization Strategies

There are no chemical solutions that chemically degrade explosives. It has been stated by the Bureau of Alcohol, Tobacco and Firearms (ATF) that clothing, footwear, and tools cannot be effectively decontaminated following exposure to *bulk* explosives [2]. While it may not be possible to totally eliminate the potential for contamination, in most cases where the degree of exposure is relatively low the risk can be reduced to acceptably low levels by wearing protective clothing, washing, and dispersal.

Tools and footwear should be scrubbed with soap and water and rinsed with alcohol. Swabs should then be taken for testing. Ion mobility spectrometry is a useful technology for the presumptive screening of most organic explosives and nitrate, however quantifying inorganic traces and the peroxides is more problematic.

The bomb scene and mortuary are referred to as *primary scenes*, and any scene associated with a suspect is referred to as a *secondary scene*. It is critically important for primary and secondary scenes to be processed independently from each other to ensure trace matter has not been inadvertently transferred between the two.

Good primary scene hygiene involves

- defining epicenter (hot zone), inner perimeter (warm zone), and outer perimeter (cool zone);
- defining an access route that is unlikely to have been used by the perpetrator;
- locating the crime scene evidence collection post at the edge of the inner perimeter. In some cases it may also be beneficial to locate analytical support (mobile laboratory) here as well;
- wearing of clean protective clothing and the use of stepping plates;
- deploying a practiced team with defined roles such as
  - the team leader coordinating activities,
  - the searcher ("dirty" man crawling through the debris),
  - the collector ("clean" man taking notes and packaging samples in as hygienic manner as possible), and
  - the scene photographer and recorder.

Good secondary scene hygiene includes

- deploying a team and equipment that has not been in the primary scene, or which have been showered, scrubbed and rested;
- planning the entry to record, search and sequentially collect potential evidence in a manner that protects and minimizes the loss, and (in tropical climates) allows those in the scene to reduce their personal protection as quickly as practical;
- wearing clean pretested protective or disposable clothing, double gloves, and boot covers. This includes using clean and preswabbed sample collection equipment;
- using collection methods and packaging techniques that protect and preserve potential evidence.

If the main laboratories examination area for secondary scene samples is limited, and if the sizes of the items are large, such as vehicles, it may be worth considering choosing another location and setting up an improvised laboratory and examination area.

Searching for Evidence and Explosive Residues

The objective of the examination is to establish the facts, identify the cause, and if criminality is involved, identify the perpetrators. It is important to record the scene from the outset, and detail the recovery of potential evidence during the systematic search.

An unexploded bomb has the highest potential to yield evidence of the perpetrators identity, skill, and physical evidence of association. If the incident involves more than one bombing, the scene with least damage probably has a higher propensity to yield evidence than the scene where the bomb damage is greater. There is a natural tendency to be drawn to the site of greatest damage, and this can be a mistake.

The force of the explosion decreases exponentially with distance, so items in close proximity to the main charge will exhibit greater fragmentation than items further away. Sometimes components of the bomb are ejected or only partially consumed. Metal objects such as ends of detonators, leg wires, crimps, and plastic casings may survive smaller explosions. The recovery of nails, ball bearings, bolts, or other metal fragments from around the epicenter should arouse suspicion that "shrapnel" was part of the bomb.

The way the bomb was manufactured and set will affect the dispersal of postblast components and residues. It is always important to record where things are found and employing a planned search strategy such as a grid or a zone search can reveal useful insights, for example finding fragments of shattered fabric on opposing sides of the epicenter is evidence of it being part of the bomb. Similarly, the metal fragments found around the crater and in victims is evidence of shrapnel. This would suggest a higher degree of malice by the bomb maker (See Figures 4–7).

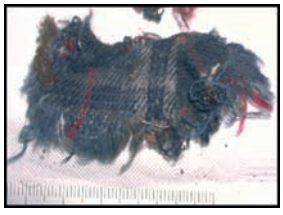


Figure 4 Tartan fabric



Figure 5 Fabric pieces

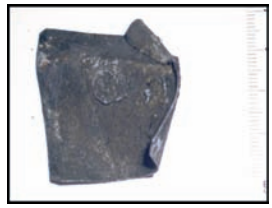


Figure 6 Metal fragment

It is important to collect "control" samples. There are two types, (i) equipment – to demonstrate tools and containers were not contaminated with relevant substances prior to use; and (ii) environmental – to quantify background levels of ions and traces of interest. Environmental controls should be collected from a site as near as practical to, but not associated with, the incident under investigation. The examiners should search for areas that were protected, such as soil from under the bitumen road adjacent to the crater was found to be most appropriate at the first Bali bombings 2002.

Explosives that are organic in nature such as TNT are more soluble in organic solvents than water. They tend to preferentially persist in moist or wet environments compared to water-soluble materials. Improvised explosives are often made by containing mixtures of strong oxidizers such as nitrate and chlorate salts with a fuel which may also be either organic (diesel, flour, or sugar) or inorganic (sulfur or aluminum dust). Inorganic materials such as the strong oxidizing salts are particularly soluble in water and prone to being washed away by a burst water main or the fire fighters' hoses. While the solvent and swabbing techniques used by forensic scientists vary, most would agree that small swatches soaked with isopropanol, such as MediPrep wipes commonly used in medicine, are convenient and effective.

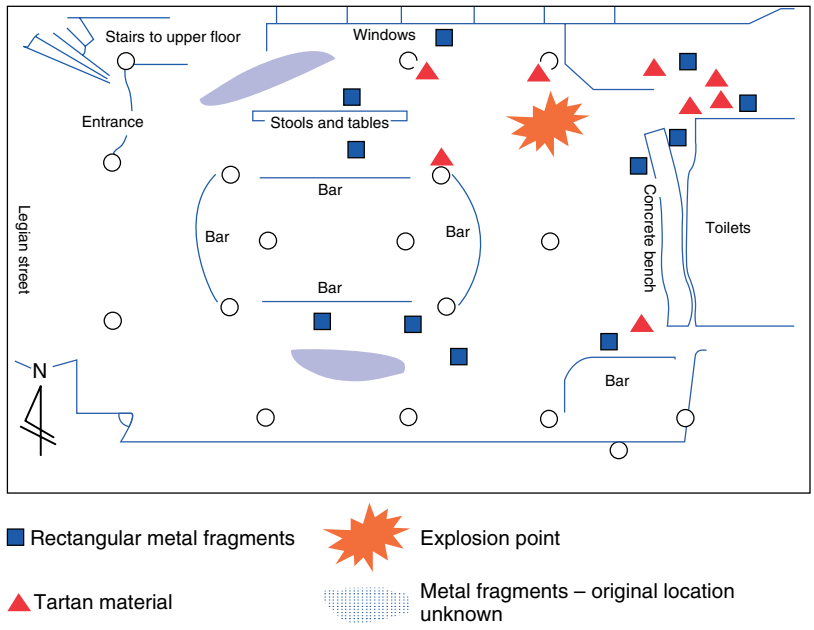


Figure 7 Sketch showing distribution of fabric and shrapnel around epicenter in Paddy's Bar – first Bali bombings 2002

There is no documented evidence of natural sources of chlorate in the environment. Intuition would suggest it is highly reactive and liable to attenuate and degrade in moist fuel rich environments. Logan [3] concluded that high humidity combined with city grime and soot was likely to cause the rapid disappearance of residues following the first Bali bombings in 2002. A review of the scientific literature adds further support in that, although effluent-chlorate anions have been found to be relatively stable in sea water they are nonpersistent in other environments due to natural biological processes.[a] Nitrates however, can and do persist in the environment and especially in soils, and especially those to which fertilizers have been applied. In these circumstances, before it can be argued that elevated levels were found in the vicinity of the bombing one must conduct environmental surveys to determine background.

Kelleher [1] observed that there is evidence to support the suggestion that explosives residue is derived from a thin outer layer of the charge and that the proportion of residue will decrease as both the charge size and the velocity of detonation increase. He also found that simple mathematical models indicate that residue not associated with fragments is concentrated within a limiting radius of approximately 60 m regardless of the charge size (excluding wind effects), however the distribution of explosives fragments and residue does not follow a simple inverse square distribution. High concentrations of residue are encountered near the blast seat and on primary fragmentation, but also in areas further from the blast seat than would be expected if the distribution followed a simple inverse square law.

Therefore, it may be necessary in a large bombing case to screen very large numbers of samples in order to find one that is rich in residues. In such cases, deploying scientists with portable instruments/mobile laboratory capability can be particularly useful and cost effective.

Mobile and Improvised Laboratories

Recent technological advances have seen the emergence of small instruments for use in the field. They are small, portable, "smart" (microprocessor control), easier to operate, more sensitive, and less expensive per sample to run. This means that scientists can offer better service from the incident scene and in recent years we have seen the development of mobile laboratories for field appreciation. A mobile laboratory provides a safe clean work area where

extraneous variables can be controlled. The high cost of transporting a mobile laboratory platform by air to an incident overseas becomes unattractive when an improvised laboratory that is almost as effective can be set up in a motel room near the incident. The most important thing about a mobile or improvised laboratory is that they enable work to be performed to a standard as the work performed in the main accredited laboratory. Practices and procedures, especially for cleaning and preparing examination areas for samples should be the same.

The technologies commonly used in these laboratories include microscopy, ion chromatography, infrared spectroscopy fourier transform infra-red spectroscopy (FTIR), ion mobility spectrometry, and a range of microchemical spot tests.

There are many advantages for deploying a mobile laboratory to the scene of a major incident. The mobile laboratory should be located between the inner perimeter and outer perimeter. Interfacing with the warm zone will facilitate the rapid screening of large numbers of samples, but without incurring the problems associated with decontaminating all the surfaces in the various instruments. Portable instruments should not be taken into the actual scene because decontamination becomes problematic.

The mobile laboratory preliminary findings are welcomed by investigators because of their timeliness and potential value to progress the investigation rapidly. It is important however, for investigators to understand what a *presumptive* test result means, and to appreciate the importance of treating this information in a tentative manner, similar to intelligence.

One of the main advantages of deploying scientists and portable instruments is to reduce the amount of irrelevant work that can so often clog up the main laboratory. A smaller number of samples can be subjected to far greater scrutiny and in depth analyses. Field samples that yield positive presumptive test results can be sorted and sent to different laboratories for independent testing. A higher degree of confidence can be placed in a conclusion by different scientists working in different laboratories and using different methods and procedures.

Circumstances will dictate the best site to operate the mobile laboratory as there can be advantages in setting it close to the work area, and advantages in being further away. When it is adjacent to the work area it will enable the rapid examination of a large number of samples, however, contamination is likely to become more of an issue. When it is positioned at a more distant location such as in a motel room or adjacent to the forward command post, it may be easier to access, communicate, and sustain, and contamination issues may be minimized; however, the throughput is likely to diminish.

In all cases, good science requires the laboratory furnishings and instruments to be cleaned, swabbed, and tested prior to use. Another good practice is to relocate the mobile laboratory when the investigation moves into the secondary phase (examining samples associated with suspects and secondary scenes).

Report Writing

The forensic sciences evolved under the influence of legal process. It is important for practitioners to understand the client they service and to appreciate the nuances between adversarial and inquisitorial law, especially if invited to offer a "second opinion" by investigators from another country.

The degree of confidence that can be placed in an instrumental result has been subject to deliberations by international experts from both the Technical Working Group for Fire and Explosive (TWGFEX) and the Forensic International Network of Explosive Examiners (FINEX). A guideline was published by experts sitting in the Laboratory Explosion Group of TWGFEX, and although it refers to *intact* explosives, there appears to be general agreement in the international scientific community that it is applicable for trace work. If applied, it means that scientists from around the world will be in agreement as to the degree of confidence that can be placed in the identification of an explosive based on instrumental analyses. Similarly, investigators must appreciate the limits and caveats associated with the scientist's report as some observations may be based on presumptive test as opposed to confirmatory tests derived from testing in the laboratory.

End Notes

[a.] Review of the Toxicity and Fate of Chlorate in the Effluent Discharge. Bell Bay Pulp Mill Project.

References

[1] Kelleher, J.D. (2002). Explosive residue: origin and distribution, *Forensic Science Communications* **4**(2).
[2] Thurman James, T. (2006). Appendix D bureau of Alcohol, Tobacco and Firearms (ATF) guidelines for the prevention of contamination of explosives evidence, in *Practical Bomb Scene Investigation*, Taylor & Francis.
[3] Logan, I. & Ferguson, B. (2004). *An investigation into the decomposition of chlorate in potassium chlorate explosive residues*. Supervised but unpublished undergraduate research for Canberra Institute of Technology.

DAVID ROYDS

Bomb-Pulse Dating

Introduction

Traditionally, radiocarbon dating has been considered as an archaeological tool rather than a forensic one (*see* **Radiocarbon Dating**). Radiocarbon or carbon-14 (^{14}C) is produced naturally in the atmosphere by cosmic ray interactions with nitrogen. Single carbon atoms in the atmosphere are chemically reactive and are quickly oxidized to carbon dioxide CO_2. The atmospheric concentration of natural ^{14}C with respect to all carbon has remained relatively stable at about 1.2 parts per trillion over the past several thousand years. With a radioactive half-life of 5730 years, the radioactive decay of ^{14}C is minimal within the time period of interest in medical forensic cases and applicable for samples over 300 years of age. Willard Libby was awarded the Nobel Prize in Chemistry in 1960 for the development of radiocarbon dating [1].

Atmospheric testing of nuclear weapons during the 1950s and early 1960s doubled the concentration of $^{14}C/C$ in the atmosphere (Figure 1) [2]. From the peak in 1963, the level of $^{14}CO_2$ has decreased with a mean life of about 16 years, not due to radioactive decay, but due to mixing with large marine and terrestrial carbon reservoirs. The ^{14}C has not actually disappeared, it has simply moved out of the atmosphere. The temporal variations of artificially high levels of atmospheric radiocarbon have been captured in organic material worldwide and thus offer an opportunity to determine a date of synthesis for biomolecules. Since radiocarbon is incorporated into all living things, this pulse is an isotopic chronometer of the past half-century.

The atmospheric $^{14}CO_2$ curve depicted in Figure 1 is a northern hemisphere annual growing season average. It is constructed using several different data sets that used tree rings, recent plant growth, and direct atmospheric sampling to provide carbon samples [5–8]. Since there were relatively few geographic sources of bomb-pulse ^{14}C, the upswing and the peak values of the curve do vary with location around the globe [7, 8]. However, since CO_2 is a gas, the pulse of $^{14}CO_2$ mixed in the atmosphere with all other CO_2 to produce a relatively homogeneous distribution of atmospheric $^{14}CO_2$ by the late 1960s [9].

The isotopic content of new plant growth reflects the atmospheric radiocarbon concentration. New leaves are produced in a matter of weeks while larger fruits and vegetables form over the period of a month or two. Herbivores lag the atmosphere slightly because their primary carbon source is on the order of months old. Omnivores and carnivores lag the atmosphere further because their carbon sources are another step removed.

Within organisms, tissues turn over at different rates so ^{14}C levels vary between tissues. The date of formation of a tissue or specific biomolecule can be estimated from the bomb-curve by considering these lags in incorporation and relating the ^{14}C concentration with the date. Using an annual average of the carbon intake over a growing season can account for much food chain lag and produce a usable curve (Figure 1). Caution must be exercised when dating an elevated sample since the pulse is double valued. Placing a sample on the ascending or descending side of the pulse can often be accomplished if other information is available.

Measurement of Radiocarbon Samples

Today most ^{14}C dating analyses are conducted using accelerator mass spectrometry (AMS), although there are still labs that use decay counting. AMS is much faster and generally more precise than decay counting since it measures differences in carbon atom mass and is not constrained to wait for atomic decay. AMS can also use smaller samples than decay counting, an important issue when analyzing evidence.

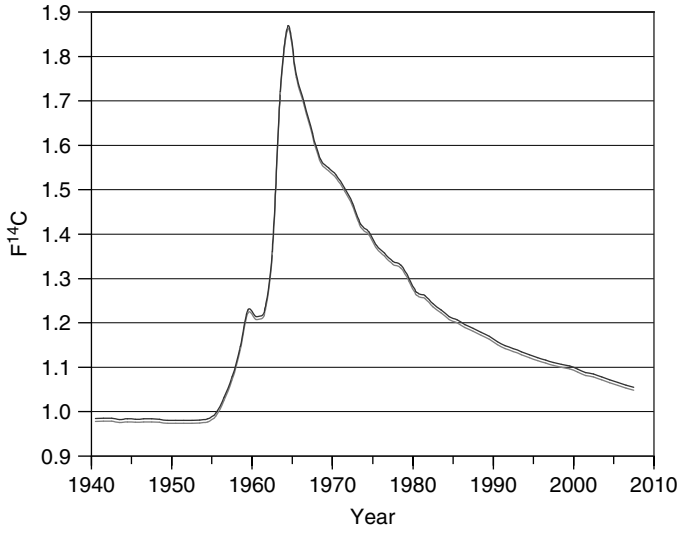


Figure 1 Northern hemisphere growing season average of atmospheric ^{14}C concentration in CO_2 from 1940–2007. The vertical axis uses the $F^{14}C$ nomenclature defined by Reimer, *et al.* (2006) specifically for bomb-pulse applications [3]. Other nomenclatures for expressing ^{14}C concentration can be found in the literature [4]

Sample preparation and measurement details vary among AMS facilities, depending on the type of sample to be analyzed and the design of spectrometer. Routine radiocarbon analyses using AMS are performed on samples containing about 1 mg carbon. Samples as small as 50 μg carbon can be analyzed at some laboratories, but measurement uncertainties are larger. Nearly all AMS facilities that perform high precision dating follow these general procedures to minimize contamination from outside sources of carbon and reduce measurement backgrounds. Samples are dried completely and then combusted with excess oxygen to produce CO_2. The CO_2 is purified to remove water vapor, nitrogen, oxides of nitrogen, and oxides of sulfur. It is then reduced to graphite or elemental carbon on metal catalyst, often cobalt or iron powder.

Primary standards, secondary standards, and backgrounds are similarly processed to produce graphite, which is the form of carbon analyzed by the majority of AMS systems [10]. Graphite is the preferred form of carbon because it can be made easily at high purity and produce intense negative ion currents. It is important to have consistent sample source material (e.g., all carbon graphite) because different molecules ionize with different efficiencies. A handful of gas accepting ion sources that take direct feed of CO_2 exist, but they are not typically used for high precision dating.

The precision of bomb-pulse dating depends on the ability to measure the ^{14}C concentration in a sample and the slope of the curve. It is relatively easy to achieve 0.5–0.8% precision when analyzing recent full-sized samples. As the slope of the pulse flattens, the uncertainty in ^{14}C analysis translates into a larger chronological uncertainty. When the slope was steep, the uncertainty was typically ±1 year. Since 2000, that same measurement precision yields a chronological uncertainty of ±2–4 years.

Soft Human Tissues Amenable to Dating

Most soft tissues exhibit relatively rapid carbon turnover. In a forensic context, the ^{14}C concentration in soft tissues can be used to determine an approximate date of an event, such as death of an unidentified body. Shortly after the start of the bomb-pulse [11], Broecker *et al.* [12] noted that bomb-curve radiocarbon concentrations in human tissues lag the atmosphere, relating to dietary issues and tissue turnover [13, 14]. They documented a lag time of about 1.1 years for blood and 1.8 years for lung tissue [12]. Libby *et al.* further noted that ^{14}C concentrations in human brain tissue can lag atmospheric levels by

only a few months [15]. Nydal *et al.* found good agreement in radiocarbon levels between samples of human blood and hair [16].

Most soft tissues decay quickly in the environment and are unsuitable for dating. Hair is a notable exception [17]. Hair tends to resist rapid decay and is often found with skeletal remains. Hair also records exposure to toxins and drugs and has been used to determine causes of death. Although individual strands of hair have little mass, the carbon content of hair is high and milligram-sized samples are relatively easy to attain. Although not widely used to determine date of death, bomb-pulse dating of hair has been used in several cases [18, 19].

Bone and Cartilage

Bone is among the preferred samples for traditional radiocarbon dating. Bone's ability to resist rapid decay while containing a relatively high concentration of carbon makes it a desirable material for traditional dating [20]. Traditional bone dating uses a collagen extraction to avoid potential complications with mineral exchange of carbonates in the environment. Collagen is a protein and is not affected by environmental carbonate exchange like the mineral component of bone.

Attempts to use bone and cartilage for bomb-pulse dating have not been very successful. Bone and cartilage do not lock carbon in an inert structure like hair. Bone and cartilage are alive and exhibit low but variable turnover, depending on activity and type of bone [21–29]. Furthermore, bone and collagen turnover varies with age. Older individuals tend to lose more bone than they replace during the bone recycling process. In general, bomb-pulse dating of bone can be used to determine if someone was alive during the period of the pulse, but cannot determine a date of birth or death. Cartilage has the same limitations as bone. In 1964, an autopsy study of the cartilage collagen of 70-year-old adults showed little increase of artificial radiocarbon despite living throughout the entire rise of the bomb-pulse from 1955 to 1964 [15].

Teeth

Although dental enamel is the hardest substance in the body, teeth are not routinely used in traditional radiocarbon dating due to fear of carbonate mineral exchange during centuries of burial. After being produced, there is no turnover of enamel, and the ^{14}C concentration reflects the level in the atmosphere at the time of enamel formation. Since teeth are formed at distinct, well-documented ages during childhood [29, 30], the ^{14}C concentration in dental enamel can be used to determine an approximate date of birth [31, 32]. The absence of bomb-pulse carbon from enamel places date of birth in the 1940s or earlier. The technique reports accurate determination of teeth of known age with precision ± 1.5 years, a significant improvement over previous techniques.

Processing of enamel samples is different than soft tissues because the carbon resides in a mineral matrix. Enamel samples must be dissolved in acid and the liberated CO_2 must be trapped for isotopic analysis.

The live part of the tooth, principally dentin, is like bone, with high collagen content and slow turnover and recycling of the carbon. It provides little information from ^{14}C other than whether an individual was alive during the pulse [27].

Efforts to develop a laboratory test for assessing age at death from teeth has focused on measurements of racemization of aspartic acid [33–38]. Aspartic acid slowly changes its structure over time at body temperature, but slows dramatically at cooler temperatures. Racemization analysis works best in cold climates in which ambient temperature is well below body temperature. The technique does not work for bodies that have been burned at high temperature. By measuring the ratio of one orientation (D-aspartic acid) to the other (L-aspartic acid) an approximate age can be determined. Reported precision varies widely among labs over the past 30 years [33–38]. Recent studies report precision of ± 1.5 years, similar to bomb-pulse dating. A single tooth can be split and analyzed for both aspartic acid and ^{14}C to get estimates of date of birth and age at death.

Documents

Radiocarbon dating is routinely used to date archaeological and art objects to the correct era. The purpose of the date is to detect forgeries. In theory, the same approach can be used to date documents originated during the bomb-pulse. Unfortunately, the carbon in paper is not from a specific year due to the relatively long lifetime of trees. The absence of bomb-pulse

carbon in paper does not necessarily indicate a document is more than 50 years old. The presence of bomb-pulse carbon does, however, place the document after the onset of the pulse.

Seeds, Pollen, and Agriculture

Seeds and pollen are used in traditional radiocarbon dating to place archaeological sites in a historical context. Since both are produced during a relatively short window of time in a single growing season, they are considered good archaeological dating samples. Because seeds and pollen are produced during limited times, they are potentially useful for placing forensic evidence in a historical context. The bomb-pulse has been used to date agricultural products in a couple of applications [39]. Vintages of recent wines can be confirmed. Interdicted contraband, such as illegal drugs, can be dated with the pulse to determine year of production. Products such as ivory from endangered species can also be dated, but customs services have not routinely used the technique.

Acknowledgments

This work was performed under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

References

[1] Libby, W.F., Anderson, E.C. & Arnold, J.R. (1949). Age determination by radiocarbon content – world-wide assay of natural radiocarbon, *Science* **109**, 227–228.

[2] Natural Resources Defense Council (2008). *Table of Known Nuclear Tests Worldwide: 1945–1969. 1970-1996*, http://www.nrdc.org/nuclear/nudb/datab15.asp.

[3] Reimer, P.J., Brown, T.A. & Reimer, R.W. (2004). Discussion: reporting and calibration of post-bomb ^{14}C data, *Radiocarbon* **46**, 1299–1304.

[4] Stuiver, M. & Polach, H.A. (1977). Discussion: reporting of ^{14}C data, *Radiocarbon* **19**, 355–363.

[5] Stuiver, M., Reimer, P.J., Bard, E., Beck, J.W., Burr, G.S., Hughen, K.A., Kromer, B., McCormac, G., Van der Plicht, J. & Spurk, M. (1998). INTCAL98 radiocarbon age calibration, 24000-0 cal BP, *Radiocarbon* **40**, 1041–1083.

[6] Stuiver, M., Reimer, P.J. & Baziunas, T.F. (1998). High-precision radiocarbon age calibration for terrestrial and marine samples, *Radiocarbon* **40**, 1127–1151.

[7] Levin, I. & Kromer, B. (2004). The tropospheric $^{14}CO_2$ level in mid latitudes of the northern hemisphere (1959–2003), *Radiocarbon* **46**, 1261–1272.

[8] Hua, Q. & Barbetti, M. (2004). Review of tropospheric bomb ^{14}C data for carbon cycle modeling and age calibration purposes, *Radiocarbon* **46**, 1273–1298.

[9] Ubelaker, D.H. & Buchholz, B.A. (2006). Complexities in the use of bomb-curve radiocarbon to determine time since death of human skeletal remains, *Forensic Science Communications* **8**, http://www2.fbi.gov/hq/lab/fsc/backissu/jan2006/research/2006_01_research01.htm.

[10] Brown, T.A. & Southon, J.R. (1997). Corrections for contamination background in AMS ^{14}C measurements, *Nuclear Instruments and Methods in Physics Research Section B-Beam Interactions With Materials and Atoms* **123**, 208–213.

[11] Rafter, T.A. & Fergusson, G.J. (1957). The atom bomb effect – recent increase of carbon-14 content of the atmosphere and biosphere, *Science* **126**, 557–558.

[12] Broecker, W.S., Schulert, A. & Olson, E.A. (1959). Bomb Carbon-14 in human beings, *Science* **130**, 331–332.

[13] Harkness, D.D. & Walton, A. (1969). Carbon-14 in the biosphere and humans, *Nature* **223**, 1216–1218.

[14] Harkness, D.D. & Walton, A. (1972). Further investigations of the transfer of bomb ^{14}C to man, *Nature* **240**, 302–303.

[15] Libby, W.F., Berger, R., Mead, J.F., Alexander, G.V. & Ross, J.F. (1964). Replacement rates for human tissue from atmospheric radiocarbon, *Science* **146**, 1170–1172.

[16] Nydal, R., Lövseth, K. & Syrstad, O. (1971). Bomb ^{14}C in the human population, *Nature* **232**, 418–421.

[17] Geyh, M.A. (2001). Bomb radiocarbon dating of animal tissues and hair, *Radiocarbon* **43**, 723–730.

[18] Wild, E.M., Arlamovsky, K.A., Golser, R., Kutschera, W., Priller, A., Puchegger, S., Rom, W., Steier, P. & Vycudilik, W. (2000). ^{14}C dating with the bomb peak: an application to forensic medicine, *Nuclear Instruments and Methods in Physics Research B* **172**, 944–950.

[19] Nakamura, T., Kojima, S., Ohta, T., Nishida, M., Rakowski, A., Ikeda, A., Oda, H. & Niu, E. (2007). Application of AMS C-14 measurements to criminal investigations, *Journal of Radioanalytical and Nuclear Chemistry* **272**, 327–332.

[20] Taylor, R.E., Suchey, J.M., Payen, L.A. & Slota, P.J. (1989). The use of radiocarbon (^{14}C) to identify human skeletal materials of forensic science interest, *Journal of Forensic Sciences* **34**, 1196–1205.

[21] Manolagas, S.C. & Jilka, R.L. (1995). Bone marrow, cytokines and bone remodeling, *The New England Journal of Medicine* **332**, 305–311.

[22] Jackson, S.H. & Heininger, J.A. (1975). Proline recycling during collagen metabolism as determined by concurrent $^{18}O_2$- and ^{3}H -labeling, *Biochimica Et Biophysica Acta* **381**, 359–367.

[23] Babraj, J.A., Smith, K., Cuthbertson, D.J., Rickhuss, P., Dorling, J.S. & Rennie, M.J. (2005). Human bone collagen synthesis is a rapid, nutritionally modulated

process, *Journal of Bone and Mineral Research* **20**, 930–937.
[24] Hedges, R.E.M., Clement, J.G., Thomas, C.D.L. & O'Connell, T.C. (2007). Collagen turnover in the femoral mid-shaft: modeled from anthropogenic radiocarbon tracer measurements, *American Journal of Physical Anthropology* **133**, 808–816.
[25] Moorrees, C.F.A., Fanning, E.A. & Hunt Jr, E.E. (1963). Formation and resorption of three deciduous teeth in children, *American Journal of Physical Anthropology* **21**, 205–213.
[26] Shin, J.Y., O'Connell, T., Black, S. & Hedges, R. (2004). Differentiating bone osteonal turnover rates by density fractionation; validation using the bomb ^{14}C atmospheric pulse, *Radiocarbon* **46**, 853–861.
[27] Ubelaker, D.H., Buchholz, B.A. & Stewart, J. (2006). Analysis of artificial radiocarbon in different skeletal and dental tissue types to evaluate date of death, *Journal of Forensic Sciences* **51**, 484–488.
[28] Parfitt, A.M. (2002). Misconceptions (2): turnover is always higher in cancellous than in cortical bone, *Bone* **30**, 807–809.
[29] Nolla, C.M. (1960). The development of permanent teeth, *Journal of Dentistry for Children* **27**, 254–266.
[30] Bolanos, M.V., Manrique, M.C., Bolanos, M.J. & Briones, M.T. (2000). Approaches to chronological age assessment based on dental calcification, *Forensic Science International* **110**, 97–106.
[31] Spalding, K.L., Buchholz, B.A., Druid, H., Bergman, L.E. & Frisén, J. (2005). Forensic medicine: age written in teeth by nuclear bomb tests, *Nature* **437**, 333–334.
[32] Cook, G.T., Dunbar, E., Black, S.M. & Xu, S. (2006). A preliminary assessment of age at death determination using the nuclear weapons testing C-14 activity of dentine and enamel, *Radiocarbon* **48**, 305–313.
[33] Helfman, P.M. & Bada, J.L. (1975). Aspartic acid racemization in tooth enamel from living humans, *Proceedings of the National Academy of Sciences* **72**, 2891–2894.
[34] Ohtani, S., Abe, I. & Yamamoto, T. (2005). An application of D- and L-aspartic acid mixtures as standard specimens for the chronological age estimation, *Journal of Forensic Sciences* **50**, 1298–1302.
[35] Ohtani, S. & Yamamoto, T. (2005). Strategy for the estimation of chronological age using the aspartic acid racemization method with special reference to coefficient of correlation between D/L ratios and ages, *Journal of Forensic Sciences* **50**, 1020–1027.
[36] Ohtani, S., Ito, R., Arany, S. & Yamamoto, T. (2005). Racemization in enamel among different types of teeth from the same individual, *International Journal of Legal Medicine* **119**, 66–69.
[37] Yekkala, R., Meers, C., Van Schepdael, A., Hoogmartens, J., Lambrichts, I. & Willems, G. (2006). Racemization of aspartic acid from human dentin in the estimation of chronological age, *Forensic Science International* **159**(Suppl 1), S89–S94.
[38] Waite, E.R., Collins, M.J., Ritz-Timme, S., Schutz, H.W., Cattaneo, C. & Borrman, H.I.M. (1999). A review of the methodological aspects of aspartic acid racemization analysis for use in forensic science, *Forensic Science International* **103**, 113–124.
[39] Zoppi, U., Skopec, Z., Skopec, J., Jones, G., Fink, D., Hua, Q., Jacobsen, G., Tuniz, C. & Williams, A. (2004). Forensic applications of C-14 bomb-pulse dating, *Nuclear Instruments and Methods in Physics Research B* **223**, 770–775.

BRUCE A. BUCHHOLZ

Botany

Botany is a branch of biology. The study of botany involves the scientific investigation of organisms classified as plants. What classifies as a "plant" is not as simple as it may appear. The naming and classification of organisms, *taxonomy*, is broadly a part of *systematics*. The main goal of systematics is to trace evolutionary history or *phylogeny*.

Classically, taxonomy involves some form of hierarchical system of classification that groups similar organisms. Unfortunately, an examination of a variety of botany books is likely to be confusing even to a general scientist, far less a "lay" reader, as there are numerous ways of grouping organisms and these have changed with time. Most systems since the late 1960s attempt to classify evolutionary branches in groups called *clades*. However, clades do not always align with taxonomic classification.

Plant Kingdom

In a classic introductory botany book that was being used in the 1960s, *Intermediate Botany* by Brimble [1], the Plant Kingdom was classified into five divisions as shown in Table1.

In 1969, Whittaker proposed a classification of five kingdoms in which fungi were separated from the Kingdom Plantae (see [2]). More recent research, including emerging DNA evidence, and based on the concept of "clades", would suggest a three-domain system in which Plantae and Fungi are included in a higher level of Eukarya.

Table 1 Classification of plant kingdom

Divisions	Examples
Thallophyta	Algae, Fungi, and Lichens
Bryophyta	Mosses and Liverworts
Pteridophyta	Ferns, Horsetails, and Clubmosses
Gymnosperms	Conifers, Cycads, Gnetophytes, and Ginkgo
Angiosperms	Monocotyledons and Dicotyledons

According to Brimble [1]

The principle goal in forensic botany is to accurately identify and assign a name. This in turn will lead to information about the plant bearing that name [3].

The lowest distinct group or population is a *species*. This is defined as the population whose members have the potential to breed with one another to produce fertile, viable offspring. The species is the lowest level of distinct *taxon*. The next level is called a *genus* (plural genera). Owing to the eighteenth-century Swedish botanist, Karl Linnaeus, who proposed the "binomial" system of nomenclature, all species are named by their genus and species name; for example, ryegrass is *Lolium perenne*.

To give some idea of the complexity facing the forensic botanist, there are in the order of 12 000 mosses, 12 000 ferns but over 250 000 flowering plants or angiosperms! To complicate matters further, there can also be significant variation within a species!

The study of plants within each major group is a specialization in its own right. Within each group, there are also many subspecializations. Along with this complexity, there are two other factors. Firstly, the knowledge base of specialists is often further narrowed by their area of application. For example, a botanist who works on cereal grains is unlikely to know much about "ferns". The second major problem facing the forensic botanist is that they will be dealing with small and incomplete examples, often a few fragments or seeds. Plant classification classically relies on having access to complete, flowering specimens.

Role of Forensic Botany

Although is has been suggested in several papers that forensic botany is a relatively recent endeavor, this is not the case. Indeed, before the significant developments of the 1960s in analytical instrumentation, the emergence of broader forensic biology, and, most recently, the emergence of DNA technology, forensic botany formed a significant part of forensic science. This was because much could be achieved with the use of microscopic techniques. There was also less emphasis on identification of suspects simply because, with the exception of fingerprints, it was not possible.

Paradoxically, the emergence of DNA, and its more recent application to forensic botany, has led to a *renewed* interest in forensic botany (see [4]).

As botany has such a wide coverage, it cannot be expected that even a botanist specializing in forensic applications can be an expert in all areas. Broadly, it can be expected that the forensic botanist will have expertise in the following areas:

- assessing the potential for forensic botany to contribute given case specifics;
- advising on appropriate sampling and scene examination;
- preliminary examination of material of possible botanical origin;
- more detailed examination depending on the specific expertise of examiner;
- advice on appropriate external experts; and
- working with nonforensic experts to ensure forensic standards and requirements are met.

The "bread and butter" for forensic botanists in many parts of the world remains the examination and identification of plant material alleged to be cannabis. However, the potential scope of forensic botany includes examination of the following:

- partially digested food such as stomach contents, vomit, and feces;
- wood and wood fragments;
- pollen and other spores;
- poisonous plants;
- diatoms;
- plant fibers used as textiles or in paper;
- plants, or plant parts, used in illicit drugs, including, but not limited to, cannabis;
- nonspecific plant parts or whole plants which may be present only as fragments or physical traces; and
- soils for organic, botanical traces.

Such examinations may assist in the following:

- locating a crime scene;
- determining the timing of an event;
- reconstructing events;
- drowning diagnosis;
- investigating war crimes;
- providing intelligence in drug investigations;
- establishing evidence of illegal importations; and
- establishing contact between individuals and scenes.

The above is not a complete list of possible areas where forensic botany may be able to assist. The forensic botanist needs to have a broad background aimed at recognizing key indicators that would identify the type of plant material. Although there is a shared body of general botanical knowledge, a local knowledge of the flora and an understanding of the local ecology are the essential components in a competent forensic analysis. Added to this is a knowledge of how plant materials may be altered due to a wide variety of environmental and other impacts. For example, what will a cereal grain, and its components parts, look like after being processed into a food product or after traveling through the human digestive tract?

Because of the limitations placed on identification owing to the incomplete nature of many forensic samples, examinations may only be comparative. Often, however, plants possess very characteristic features that are strongly indicative of identity. It is critical to understand that identity, although the gold standard, is only one aspect of the work of a forensic botanist. Prior to identification, the forensic botanist needs to consider criminalistic aspects such as the following:

- Is the nature of recovered material what would reasonably be expected, given what is known about the alleged case circumstances?
- Is the material recovered "consistent" with the transfer and persistence of botanical materials (little research has been done in this area)?
- Can any missing species or differences where the "known" sample has more species present than the recovered sample be explained?

The answers to these types of questions are rarely simple, in part, due to the dearth of research into criminalistics aspects of forensic botany.

It is beyond the scope of this article to consider *in detail* the broad range of materials that may be present to the forensic botanist. A few examples are presented to illustrate what can, and what cannot, be done. Finally, the current status of the potential contribution that DNA may make to forensic botany is considered.

Plants and Plant Fragments

Besides the more specialized categories of plant material, the forensic botanist may be confronted with an almost endless array of plant fragments, seeds, or even (but rarely) whole plants. This is because plants are to all intents, ubiquitous. Even the barest alleyway will have some plant material, whether it is a weed growing out of a crack in the road, moss or algae on a rock, or dried seeds blown in by the wind. Many seeds are specifically adapted to adhere to the surface of passing animals and this mechanism for dispersal also results in ready adherence to various fabrics. While such materials may be recovered during evidence collection and triage, only in those rare cases where the plant material may be critical, is it likely that it will be examined.

The task of a forensic botanist is to attempt to identify (usually) fragments or specific parts (such as seeds) from plants. This requires a generalist approach to recognize the broad grouping to which the recovered material belongs. If the botanist has some specialist knowledge of this group then it may be possible to identify to a more specific level. Where the scene of the incident is known and appropriate known samples are available, this is extremely useful. First, identification of the actual plants at the scene is invariably made easier and second, it is then possible to compare the "fragment" recovered from the case item with the relevant part of the known plant. In some cases, the reverse may apply where the botanist is able to suggest a type of ecological site based on the plant species present. Although the basics of plant examination are the same worldwide, knowledge of the local flora and habitat is *vital*.

The use of identification keys, with a few exceptions, is not especially useful to the forensic botanist. Demmelmeyer and Adam [5] discuss some specific examples where keys may be of value. More recent interactive keys do not require keying out in a particular order, which increases there potential value for application in a forensic context.

In summary, the forensic botanist needs to be first and foremost a generalist and in many cases will need to consult with a specialist in the type of plant present in the particular case.

From a crime scene perspective, it is most likely that the scene will be an outdoor setting, although indoor settings are encountered. The types of case in which botany are most likely to be encountered will be sexual assault and/or murder cases. In some cases, the body may be in a shallow grave or simply covered with plant debris. A rather specific type of outdoor scene is where a body is located partly or fully submerged in water or near water where drowning is suspected.

If it is not possible to have a forensic botanist attend the scene, the guiding rule should be to collect more, not less. Plant material should be placed in either cardboard boxes or in paper bags and preferably *not* in plastic bags. If the material is damp, it can be placed for a short term in plastic but must be dried as soon as possible. Finally, the possibility of a physical fit between the scene and suspect samples should be considered.

Plants from an outdoor scene will often be associated with soil.

Cannabis and Other Plants Used as Illicit Drugs

By far, the most common plant material examined in forensic laboratories is cannabis, *Cannabis sativa*. The identification of cannabis plant material is straightforward because the combination of microscopic features of cannabis was unique. Excellent descriptions of the microscopic features of cannabis are to be found in older pharmacy books, reflecting the fact that cannabis was commonly used as a medicinal drug (see [6]).

From a crime scene perspective, the forensic investigator is only likely to be involved in more serious cases involving cultivation. These may involve outdoor scenes, the use of glass houses and increasingly indoor settings with nonsoil-based hydroponic cultivation. This type of specialist assessment is not the normal role of the crime scene investigator, unless they have relevant competencies and experience. The competencies will include botanical training to identify male and female cannabis plants, to assess the state of maturity, and form an opinion (where appropriate) on likely yields from a particular crop situation. This is most accurately achieved when the crop is mature and can be weighed. In other situations, it is sometimes possible to broadly estimate a potential yield of an immature crop from an assessment of the site and growing conditions. Without being prescriptive, this assessment could include factors such as the number of plants present, potential of the site to nurture that number of plants, water and nutrient availability, time-of-year climatic factors, comparison with similar crops where the yield is already known, or for an indoor cultivation, lighting, and temperature conditions. Provided it is supported by appropriate data, it is also sometimes possible to predict a range of potential yield for a site.

Papaver somniferum is one of approximately 100 species of poppy and is generally considered to have two subspecies, i.e., subspecies *somniferum* and subspecies *setigerum*. Both subspecies are annual herbaceous plants. The wall of the immature seed capsule produces a milky latex, which contains various opiates, including morphine. When the capsule is scored, the latex is exuded and congeals on the outer wall surface. This is later collected and can be smoked or consumed with tea, or the morphine extracted from the raw opium gum and converted into heroin. The subspecies *somniferum* is legally cultivated in some countries, including Australia, for the supply of morphine to the pharmaceutical industry.

Whole plants in flower or with intact seed capsules are readily identified; however, detached capsules need to be analyzed chemically for the presence of morphine as they lack sufficient features to enable identification to species level.

Other plant species are used as sources of drugs with some plants being common in specific geographic areas.

Pollen Analysis

Pollen grains from the male reproductive cells of angiosperms and gymnosperms have highly resistant outer walls, which can be intact for very long periods of time under the right conditions. The study of pollen, *palynology*, is a specialist field of botany and requires specific training in the preparation of samples and in the identification of pollen grains. It is normal for there to be pollen present from a number of species and this is called a *pollen assemblage*.

Plant species present at a scene may be unique, rare, or quite common but the "assemblage" may be quite specific for a vegetation area, which may indicate a scene. Because it would be very difficult to avoid carrying away pollen adhering to shoes, clothing, tools, or a vehicle, these objects can be potentially linked back to a scene. Specialist knowledge is required about the species in a locality and how pollen from these species may be dispersed to properly interpret a pollen assemblage. For example, pollen from wind pollinated species (anemophilous) is borne by air currents and can travel considerable distances. Pollen from plants that require insect intervention tends to be heavier and displaced pollen will generally be located very close to the parent plant. Hence, interpretation of what is a meaningful difference in comparing two samples is not always simple.

The forensic literature contains a significant number of case reports that illustrate how forensic palynology can contribute to questions of interest to the investigator and, ultimately, in a court situation. Papers from two major international meetings on forensic palynology since 2000 have been published in special editions for the journal, *Forensic Science International*. In the lead editorial to the 2006 volume, Mildenhall *et al.* [7] present persuasive arguments for why forensic palynology is of value and how it can be used.

Wood

The need to examine wood or wood fragments can arise from timber or timber products used as weapons, where wooden doors or window frames are damaged during a break and enter, or from wood shavings that were used in older style safes where the shavings formed the "ballast". The latter was an important part of forensic biology until, at least, the 1970s.

Botanically wood is divided into two broad groups, *softwood* and *hardwood*. These can be distinguished by very obvious differences in the major cell types present. While the anatomy of wood is quite complex, it is well described for all commercial species. There are numerous physical and computer-based keys and programs for the identification of small samples or fragments of wood. Their use requires a detailed knowledge of the tissues and cellular features found in wood. This requires detailed knowledge and is quite a specialized field. The relevant expertise is not commonly found in today's forensic laboratory.

There are keys for the identification of larger wood samples using a hand lens. At present, the identification of a wood sample would often require the use of a scientist in a wood research institute. Scientists experienced in looking at wood or timber are very good at identifying timber by eye or with the aid of a hand lens. However, for forensic purposes identification must be based on a full examination using appropriate identification keys and confirmation through comparison with a verified reference specimen.

Identification usually requires information gained by examining three types of sections: a transverse cross section, tangential, and radial longitudinal sections. As wood ballast consists of very small fragments, often of 1 mm or less, examination involves preparing sections of these using a freezing microtome. Sections are suitably stained and their microscopic features recorded. With very small particles of wood, it may not be possible to obtain all three sections even with the aid of a freezing stage microtome.

In collecting the samples for examination, and/or comparison with a questioned forensic sample, it is important to think about what might be available for transfer and in what form. The collection of appropriate known samples, and later of possible reference samples, can require some specialist advice or involvement. For example, care should be taken to ensure that samples are representative. It is not unknown for a wooden window to have a repaired section comprising a different timber from that used in the original construction. This is especially so in older properties as a wider range of commercial timbers were used historically in building construction than is the case today.

The possibility of physical fits with wood should also be considered. Wood or timber may also be associated with other potential evidence, for example, paint, which may have higher probative value Fiddian [8] gives several examples of physical fit evidence.

As wood is laid down through an annual growth cycle, the annual growth rings in timber can be "read" and give a history of the life of the tree. There are cases reported in the literature in which growth ring patterns have been used to link stolen timber (a

major issue in North America) back to stumps at the location where the trees were harvested [9].

The most famous, or infamous, case in which growth ring analysis played a significant role was in the Lindbergh kidnapping. A home-made ladder, used to gain access to the Lindbergh home, was linked to a piece of wood found in the suspect's home, by matching the wood used to repair a broken rung in the ladder. A full account of this fascinating historical case can be found in [10].

Fiddian [8] also describes the use of growth ring analysis to link a sawn-off rifle or shotgun stocks with recovered weapons.

Food Residues

Food residues may typically present themselves to the forensic botanist in the form of stomach contents, vomit, and feces.

Stomach or gastric contents may be examined as an aid in establishing the time of death. This relies on identifying partly digested residues of foods (plant or animal) to assess when the deceased had last eaten. The analysis of vomit is essentially the same as for stomach contents except that the material to be recovered may be partially or wholly dried by the time it reaches the forensic laboratory. It is not uncommon for people to vomit during a violent assault. Fecal material will contain only the remnants of foodstuffs that are most resistant to the digestive process. Once again, it is quite common for people who are involved in a violent situation to defecate. It is also not unusual for burglars to defecate intentionally as a further "insult to injury".

It is beyond the scope of this article to consider in detail the types of information relied upon by the forensic botanist conducting this type of work. It requires an in-depth knowledge of cell types and their occurrence and appearance in plant foodstuffs. However, much of the useful information can only be found in texts that are invariably old and often out of print. Furthermore, information is generally not available in a systematic way aimed at, or suited for, forensic application. Often the key skill is to recognize the broad type of plant material and then go through a process of using reference works and preparing the known samples for comparison. By necessity, this places a heavy emphasis on an individual's experience and on what the individual has previously seen. Attempts to develop a more systematic analytical approach using manual- and computer-based keys have been made only in those areas where the materials have some commercial or academic interest, for example, with woods, seeds, or powdered vegetable drugs.

An in-depth knowledge of tissue and cell structures and how these are affected by food processing and by digestion are prerequisites for the analysis of food residue samples.

Plant Toxicology

Plants have been used for medicinal purposes, and as poisons, by humans since the ancient days. Many wild plants are poisonous when consumed by humans and exposure to these poisons is a common occurrence in everyday life, causing problems ranging from allergic reactions to gastrointestinal irritation and, at times, death. Fortunately, death by ingestion of toxic plants is relatively rare and occurs more often in young children where plant parts are eaten or chewed because of their attractiveness to the child. However, accidental overdoses of traditional remedies such as herbal medicines have caused deaths in various countries.

Rapid identification of poisonous plants that have been ingested can be paramount in saving a patient's life by providing immediate advice to the treating doctor as to what poisonous compound has been ingested, therefore enabling the administration of the correct antidote. The Royal Botanic Gardens, Kew, in the United Kingdom has developed an illustrated interactive key on CD-ROM to enable doctors to identify plants suspected of causing poisoning [11].

In some countries, certain plants have become popular suicidal agents and the incidence of fatal poisoning has increased. A famous homicide involved ricin, the poison derived from the seed of castor oil plant (*Ricinus communis*), which was most likely administered to the victim via a projectile fired from an umbrella tip. As in this case, where plant toxins have been used in homicides, they usually have been extracted from the parent plant and their identification is carried out by a toxicologist using chemical assays and tests [12]. However, if the "raw" form is used, a botanist may be required to identify the plant parts from a gastric content examination. Fiddian [8] describes several cases in which it was necessary to identify plants or preparations of plants.

In Europe, perhaps less so elsewhere, it is not uncommon for people to mistakenly collect poisonous mushrooms through misidentification and this has resulted in several fatalities.

Diatoms and Other Microalgae

Botanical evidence is not restricted to terrestrial habitats. Aquatic botanical evidence is highlighted in a case study where two young boys were attacked and bound while fishing in a pond, then dragged into the water. They managed to escape and the three suspects were apprehended soon after. The sediment-encrusted shoes of both the victims and the suspects were examined. Numerous species of plankton (or microalgae, i.e., unicellular algae, including various diatoms and chrysophytes) were recovered from the shoes and pond sediment. There were marked similarities in the compositions of plankton species in all the samples, which strongly indicated a common source [13].

Diatoms are the most common microalgae encountered as trace evidence. They occur in many habitats, including marine, freshwater, and terrestrial environments. There are various species that flourish in generally inhospitable environments, such as highly acidic habitats and thermal water bodies, where temperatures prevent the growth of most other life forms. Diatoms have extremely resilient silicous cell walls termed *frustules*, which are composed of two valves. They are classified into two groups based on their symmetry: centric (with radial symmetry) and pennate (with bilateral symmetry). The taxonomy of diatoms is based on frustule morphology, including shape and surface ornamentation [14].

A consequence of the highly resistant nature of frustules, along with the widespread distribution of diatoms in the environment and in manufactured products containing diatomaceous earth, has resulted in their periodic appearance as trace evidence in various situations.

Older types of safes are still widely used, consisting essentially of an inner vault placed inside a metal box, with insulation material (or safe packing) placed between the walls of the two boxes to give the safe fire-resistance. The common types of safe packing include diatomaceous earth and sawdust mixed with vermiculite mica, cement, or gypsum. To gain access to the inner vault, the safe breaker must pass through the packing, releasing it as a trace material. Deposits on the suspect's clothing may be examined by a botanist to ascertain the diversity and relative proportions of foreign diatom species present as well as to identify the wood fragments when present.

Diatoms are also found in other manufactured products, all of which have the potential to be a source of trace evidence. These include abrasive cleaning agents such as car polishes, some paints, and various brands of match heads.

However, the most common use of diatoms in forensics graph is in the diagnosis of drowning. During the process of drowning, plankton in the water are inhaled into the lungs and passed via the lung capillaries into the general arterial circulation. The planktons are thereby dispersed through the body to lodge in tissues such as bone marrow that may only be reached by circulation. This plankton almost invariably includes diatoms, with small pennate species being the most common type found in bone marrow [15].

There is still considerable debate over the interpretation of the presence of diatoms in drowning cases, primarily because it has been reported that diatoms have been recovered from the organs of persons who have not drowned and conversely not detected in persons known to have drowned. However, there would appear to be widespread acceptance that the presence of diatoms *can be* an indicator of drowning, with appropriate understanding of the local and specific case circumstances and appropriate sampling.

DNA Analysis

The use of DNA testing for plants has evolved and matured over the last decade and is well established in a wide range of botanical applications (see [16]). Coyle [4] also describes DNA testing for forensic applications. Although there are a small number of case reports on the use of DNA testing of plants in forensic situations, such testing is still not common. The DNA tests and interpretation of the results will usually be done outside the forensic laboratory. The interpretation of plant DNA is quite different from that of human DNA and needs to be clearly understood.

Thus far, most of the research effort on forensic applications of plant DNA has been with cannabis, to identify cannabis and to compare cannabis seizures. Howard *et al.* [17] have published the first validation

study for the use of short tandem repeat (STR)-based DNA analysis system for cannabis.

References

[1] Brimble, L.J.F. (1962). *Intermediate Botany*, 4th Edition, Macmillan, New York.

[2] Campbell, N.A. & Reece, J.B. (2002). *Biology*, 6th Edition, Benjamin Cummings, San Francisco.

[3] Cullen, J. (2006). *Practical Plant Identification*, Cambridge University Press.

[4] Coyle, H.M. (ed) (2005). Forensic botany, *Principles and Applications to Criminal Casework*, CRC Press, Boca Raton.

[5] Demmelmeyer, H. & Adam, J. (1995). Forensic investigation of soil and vegetable materials, *Forensic Sciences Review* **7**, 119–142.

[6] Jackson, B.P. & Snowden, D.W. (1968). *Powdered Vegetable Drugs*, Churchill, London.

[7] Mildenhall, D.G., Wiltshire, P.E.J. & Bryant, V.M. (2006). Forensic palynology: why do it and how it works, *Forensic Science International* **163**, 163–172.

[9] Jozsa, L.A. (1985). Contribution of tree-ring dating and wood structure analysis to the forensic science, *Canadian Society of Forensic Science Journal* **18**, 200.

[8] Fiddian, S.(2005). Botany in forensic science, in *Expert Evidence*, J. Freckelton & H. Selby, eds, Chapter 91, Thomson Lambrook.

[10] Graham, S.A. (1997). Anatomy of the Lindbergh kidnapping, *Journal of Forensic Sciences* **42**, 368–377.

[11] Dauncey, E.A., Rayner, T.G.J. & Shah-Smith, D.A. (2002). *Poisonous Plants and Fungi in Britain and Ireland*, 2nd Edition, CD-ROM, Royal Botanical Gardens, Kew.

[12] Moffat, A.C. (1980). Forensic pharmacognosy – poisoning with plants, *Journal of the Forensic Science Society* **20**, 103–110.

[13] Silver, P.A., Lord, W.D. & McCarthy, D.J. (1994). Forensic limnology: the use of freshwater algal community ecology to link suspects to an aquatic crime scene in Southern New England, *Journal of Forensic Sciences* **39**, 847.

[14] Bortolotti, F., Tagliaro, F. & Manetto, G. (2004). Objective diagnosis of drowning by the "diatom test" – a critical review, *Forensic Science Review* **16**, 136–148.

[15] Pollanen, M.S., Cheung, C. & Chiasson, D.A. (1997). The diagnostic value of the diatom test for drowning, 1. Utility: a retrospective analysis of 771 cases of drowning in Ontario, Canada, *Journal of Forensic Sciences* **42**(2), 281–285.

[16] Weising, K., Nybom, H., Wolff, K. & Kohl, G. (2005). *DNA Fingerprinting in Plants. Principles, Methods and Applications*, 2nd Edition, Taylor & Francis, Boca Raton.

[17] Howard, C., Gilmore, S., Robertson, J. & Peakall, R. (2008). Developmental validation of a Cannabis Sativa. STR multiplex system for forensic analysis, *Journal of Forensic Sciences* **53**(5), 1061–1067.

Further Reading

Robertson, J. & Grieve, W. (1999). *Forensic Examination of Fibres*, 2nd Edition, Taylor & Francis.

Stearn, W.J. (1992). *Botanical Latin*, 4th Edition, David & Charles.

JAMES ROBERTSON AND SUSAN FIDDIAN

Brain Abnormalities in Children

see Neuropsychological Assessment: Child

Brain and Deception Detection

see Deception: Detection of and Brain Imaging

Brain Imaging Test for Deception

see Deception: Detection of and Brain Imaging

Brainwashing ***see*** Deception: Truth Serum

Breath Alcohol Analysis ***see*** Alcohol: Analysis

Wiley Encyclopedia of
FORENSIC SCIENCE

Volume 2
C–E

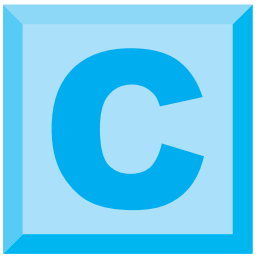
Editors-in-Chief

Allan Jamieson
The Forensic Institute, Glasgow, UK



Andre Moenssens
Forensics and Law Center, Columbia City, IN, USA





CAI ***see*** Case Assessment and Interpretation

Canadian Dangerous Offender ***see*** Dangerousness: Risk of

Canadian Long-Term Offender ***see*** Dangerousness: Risk of

Cannabis

Introduction

Cannabis, possibly one of the oldest plants cultivated by man, is thought to have originated on the plains of Central Asia. There is evidence that the Chinese cultivated it over 8000 years ago for fiber and oil. The Chinese recognized its medicinal potential, but it was seldom used as it "disturbed the equilibrium of the brain".

The plant was cultivated in the Indian subcontinent about 2000 BC where it was considered sacred and used in the performance of rituals. The ancient Hindus also hailed cannabis for its medicinal properties.

Its use became widespread in the Middle East, with initially little cultural opposition. Muhammad's teachings forbade the use of alcohol, but not cannabis derivatives. Muslim scholars and clerics debated the use of cannabis for centuries and finally, in the eighteenth century, decided that its use for pleasure was prohibited as it "beclouded" the mind in the same way as alcohol does. Its use for legitimate medicinal purposes was permitted.

Throughout the Middle Ages, cannabis was central to any herbalist's medicine store. However, the Holy Inquisition regarded those using it as sorcerers or witches and forced its use underground.

The word assassin has long been associated with the consumption of the cannabis derivative hashish. The story is based on reports by Marco Polo about a group of highly motivated, skilled, merciless political killers, who may have used cannabis as part of their religion. The use of cannabis, however, does not produce a violent mental state and an assassin under the influence of the drug is likely to be a liability.

The rapid increase in travel in the sixteenth century created an increased demand for hemp cloth and rope. King Henry VIII required all farmers to grow a one-fourth acre of cannabis for every 60 acres cultivated. Farmers were reluctant as the crop gave a poor financial return and impoverished the soil. Russia, the major producer of hemp, provided 90% of Britain's seventeenth century needs [1].

O'Shaughnessy, an Irish surgeon working in India, was impressed with the effectiveness of cannabis derivatives as a muscle relaxant, anticonvulsant, antiemetic, and analgesic. After experimenting on animals and humans, he introduced it to the British medicine in 1830, prescribing it for a wide range of illness. He recognized that the drug did not cure many of the diseases for which it was prescribed but enabled the patient to withstand the suffering. However, the cannabis derivatives were of variable potencies and it was difficult to replicate the desired effects. The advent of specific medicines such as aspirin and barbiturates eventually resulted in cannabis being removed from the British Pharmacopoeia [2].

Source of Cannabis and Cannabinoids

Cannabis (Figure 1) can be recognized at all stages of growth by its distinctive botanical characteristics. It has its own botanical family Cannabaceae and its closest relative is hops (*Humulus lupulus*). The appearance of individual plants is highly dependent on the habitat and the growing conditions. This variability led to the view that there were several different species but it is now widely accepted that only one cannabis species, *Cannabis sativa*, exists.

The cannabis plant is usually unisexual but occasionally both the male and female flowers occur on the same plant. The male plant matures before the female plant and dies off soon after flowering. The female plant is generally denser and leafier than the male plant and survives until the maturation of the fruit. The so-called seed is a very distinctive fruit consisting of a single seed in a pod.

The cannabis plant has a variety of glandular and nonglandular hairs that enable it to be identified microscopically even when finely fragmented. Glandular hairs secrete the resin that contains the psychoactive constituents of the plant and are found on all parts of the plant.

Cannabis contains at least 483 natural compounds of which 6 have been identified in the last decade [3]. While the majority of these are classified as terpenes, hydrocarbons, sugars, flavonoids, and miscellaneous compounds (413 in total), 70 are classified as cannabinoids. The cannabinoids typically contain 21 carbon atoms, although their analogs and transformation products, which may not have 21 carbons, are also included in the 70. The chemical structure of the cannabinoids appeared to be unique to cannabis until the recent discovery of cannabinoid biphenyls in liverwort [4].



Figure 1 Flowering heads of the cannabis plant

The best-known cannabinoid and the one considered responsible for most of the physiological effects experienced by cannabis users is delta-9-tetrahydrocannabinol (THC). This molecule contains two chiral centers and the naturally occurring molecule possesses the trans configuration, illustrated in Figure 2. THC is not only present in plant material but is also formed during the smoking process by decarboxylation of an acid precursor. The amount of THC in confiscated plant frequently exceeds the amount of the precursor [5]. In fresh hemp, the precursor levels may exceed those of THC [6].

THC is found on all parts of the cannabis plant, with the highest concentration being found in the female flowering heads (sinsemilla). The potency of the cannabis plant depends on the source of its seed as well as the growing conditions. Cannabis grown for fiber (hemp) may have negligible THC even on the flowering parts, whereas selected illicit cannabis,

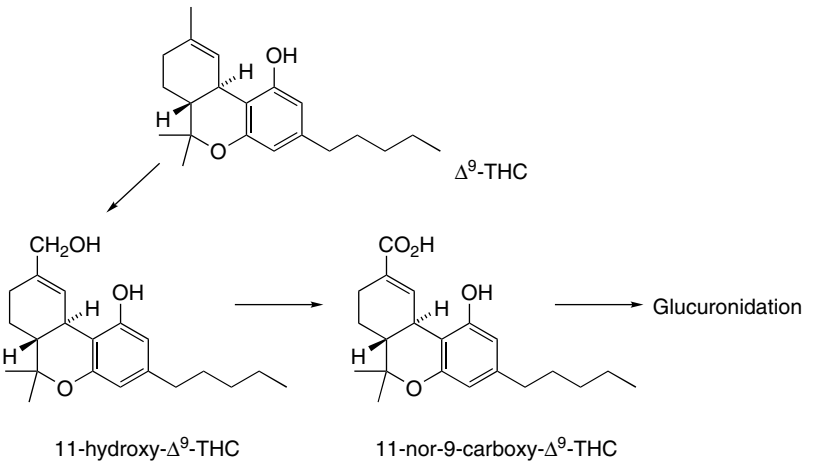


Figure 2 Principal metabolites of THC

grown hydroponically, may produce flowering heads containing more than 40% THC.

Hemp fiber has superior qualities compared to cotton and is currently used in the production of textiles, paper, fiberboard, molded car parts, insulation, and animal bedding. The cultivars grown for hemp production in countries that legislate against cannabis must contain less than 0.3% THC per dry weight of the plant, a level too low to produce psychoactive effects. Cannabis can produce four times as much paper per acre as pine trees, but the plant requires a high level of nutrients. Such requirements may make the plant useful for the removal of excess nitrogen or heavy metal contaminants from soil (bioremediation). Cannabis seeds are oil rich and were used as a human and animal food for thousands of years. The oil produced from the seeds has properties similar to linseed oil and has been used in paints, soaps, and cosmetics. Hemp seed oil does not contain THC and is used in health food and body-care products [7].

Medical use of Cannabis

Interest in the medicinal use of cannabis has recently been rekindled. In particular, cannabis has proved effective for some patients in controlling nausea and vomiting associated with chemotherapy. It has also been found useful as an appetite stimulant for those suffering from AIDS wasting syndrome. The use of cannabis for the treatment of chronic pain has yielded inconsistent results, and while smoking cannabis provides some relief of eye pressure in glaucoma sufferers, the relief is short lived. Studies are continuing on the effectiveness of cannabis for the control of muscle spasms associated with multiple sclerosis and epilepsy [8].

The mode of administration of medicinal cannabis is an issue. Smoking has associated health issues and oral administration produces inconsistent absorption. Aerosol delivery of synthetic cannabinoids offers potential advantages. A number of synthetic cannabinoids have been developed for research purposes. These include nabilone (nausea and vomiting) marketed in several European countries and the United States and rimonabant (obesity) in the United Kingdom [4].

Cannabis Abuse

It would be surprising if consumption of cannabis with its variable ratio of chemical constituents did not produce a wide range of inconsistent physiological effects. Indeed, this very complexity presents a significant scientific challenge in unraveling whether cannabinoids have any future as therapeutic agents. There is a fine line between the desirable medicinal effects and the possibly undesirable physiological effects.

The recreational use of cannabis in the Western world became widespread after World War II and is still increasing. The leaf and flowering parts of the plant are most widely used. However, the use of

cannabis resin (hashish) and cannabis oil (made by extracting the plant with a solvent) is also common.

The determination of recent cannabis usage is important in forensic cases. Drummer [9] has demonstrated that the incidence of THC in a range of Victorian (Australia) coronial case types ranged from 2.2% in natural deaths to 36% in heroin-related deaths. Many studies have found high rates of recent cannabis use by fatally injured drivers (up to 37% in a US study).

Mechanism of Action

It is now accepted that cannabis constituents exert their pharmacological effect by attaching to binding or receptor sites in mammalian tissues. Such sites have now been found in birds and aquatic species [4]. The discovery of these binding sites greatly increased our knowledge of the pharmacology of cannabis. Two types of cannabinoid receptors, CB_1 and CB_2, have been identified [10]. The former are present in the brain, spinal cord, and peripheral tissues, whereas the latter are found in immune and reproductive tissues.

The occurrence of these receptors led to the search for naturally occurring compounds that utilized the receptor sites. Endogenous cannabinoids were soon identified [11], with the neuromodulators, anandamide, and its structural relative, 2-arachidonoyl glycerol being the most significant. THC has some structural similarity to these compounds [2] and mimics their action at the receptors.

Pharmacological Effects of THC

Cannabis combines many of the properties of alcohol, tranquilizers, opiates, and hallucinogens and thus affects many behavioral and physiological functions. It influences mood, induces euphoria, and decreases anxiety. In sociable surroundings, it may increase sociability but anxious users may experience increased anxiety, paranoia, or panic. Cannabis has been linked with psychotic disorders. Acute misuse may produce a temporary psychotic state. Whether cannabis causes schizophrenia or accelerates its development in predisposed persons is still a matter of debate.

Cannabis affects perception with colors appearing brighter, music more vivid, and time and spatial perception distorted. Cannabis also affects an individual's reaction time, memory, and ability to concentrate and to study. It is not surprising therefore that the role of cannabis in traffic crashes has received considerable attention. Early studies did not show that cannabis users had an increased responsibility rate for accidents. More recent studies, however, have demonstrated the opposite [9]. This apparent contradiction appears to be due, at least in part, to the fact that the earlier studies used the long-lived and pharmacologically inactive THCCOOH as an indicator of use. The later studies measured THC, which is more relevant as it is indicative of more recent use. There is therefore increasing support for the position that recent cannabis usage, i.e. within about the last 3–4 hours, is a contributory factor in traffic crashes.

The use of cannabis causes an increase in the heart rate, thus placing people with a preexisting heart condition at risk. In view of the higher concentrations of mutagens in cannabis smoke compared with that in tobacco smoke, it is not surprising that the rates of some cancers, particularly oral cancers, are higher in cannabis smokers. The smoking process produces more than 2000 compounds by pyrolysis [2].

While cannabis abuse is a major concern, it should be noted that, unlike other recreational drugs, no deaths have been attributed solely to a cannabis overdose.

Analysis in Biological Specimens

The analysis of cannabinoids in biological specimens is generally performed as a two-stage process. Low-cost immunoassay screening tests are conducted to determine which specimens may contain cannabinoids. Specimens that give a positive response are subjected to further testing to ensure that the screening result was a true positive and to quantify the cannabinoid concentrations in the sample.

Gas chromatography-mass spectrometry (GC-MS) is the most common confirmatory procedure for quantifying THC in blood and oral fluid and its carboxy metabolite in urine. These methods typically involve a chemical derivatization of the analytes to produce acceptable chromatographic peaks [2]. The limits of detection of $1–2\,\text{ng ml}^{-1}$ of THC in blood that are typically achievable using low-cost instrumentation are sufficient for most applications.

Liquid chromatography tandem mass spectrometry (LC-MS/MS) is increasingly being used for quantifying cannabinoids in biological specimens [12, 13]. This technique has some significant advantages over GC-MS as it eliminates the need for derivatization and can achieve lower limits of detection. LC-MS/MS instrumentation, however, is more expensive than a bench-top GC-MS.

Routes of Administration and Pharmacokinetics

Smoking is by far the most common means of administration of cannabis with THC being detectable in the blood within seconds of commencement of smoking [14]. The bioavailability of the drug is reported to range from 18 to 50% [15], reflecting variability in smoking styles.

Several studies have established that THC is absorbed through passive exposure to cannabis smoke. Concentrations of both THC in oral fluid and metabolites in urine following such exposure are much lower than from active smoking [16].

Oral intake of cannabis, often in the form of cookies, results in much slower absorption, with peak blood THC concentrations being achieved in 1–5 h [17]. The degradation of THC by stomach acids and first-pass metabolism reduces its bioavailability to between 4 and 20%. Peak concentrations of THC in plasma after oral intake are about a third of those achieved after smoking an equivalent amount of THC.

THC is very lipophilic (fat loving) and after absorption is rapidly distributed from the blood to other storage depots in the body. Initially, THC distributes to highly vascular tissues including the brain and then redistributes to fat deposits. This accumulation in fat and the subsequent slow release accounts for the fact that THC has a long terminal half-life of more than 4 days.

Figure 2 shows the major metabolites in man. While the 11-hydroxy-THC (11-OH-THC) has pharmacological activity, the carboxy metabolite (THC-COOH) is inactive.

The typically rapid increase and decline in plasma THC concentrations and corresponding changes in metabolite levels are illustrated in Figure 3 [14]. The rate of decline in THC varies between individuals and is also dependent on smoking technique, frequency of use, and cigarette potency.

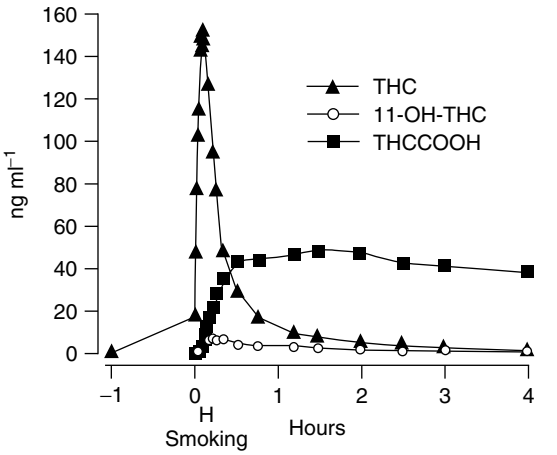


Figure 3 Mean plasma levels of THC and THCCOOH during and after smoking a single 3.55% THC marijuana cigarette [Reproduced from the Journal of Analytical Toxicology by permission of Preston Publications, A Division of Preston Industries, Inc.]

THC is sequestered in the oral cavity during the smoking process, with high concentrations being indicative of recent smoking [18]. Unlike many drugs, there is very little transfer of THC from blood to oral fluid [19].

Most of the drug is eliminated in feces (over 65%) or urine [20]. Minor amounts are excreted in hair, nails, and sweat. The window of detection for cannabis metabolites in urine is dependent on the quantity of drug consumed and the sensitivity of the analytical method employed. Low-to-moderate doses may be detected in urine for up to 12 days, but in heavy users it can be detected for up to 77 days after the last use [9]. Urine concentrations of THCCOOH are useful in determining usage but are of no use in establishing impairment.

Regulatory Aspects and Offenses

The first regulatory control efforts to control the use of cannabis were promulgated in the early twentieth century. The arrival of the drug with Mexican immigrants first brought the drug to prominence in the United States and this rapidly led to moves to control its use. Some observers consider that the "demonization" of the drug by the US authorities at that time has colored attitudes worldwide ever since [7].

Three major international drug control treaties have been developed under the auspices of the United Nations Office for Drug Control and Crime Prevention and ratified by most of the UN members, although many have attached reservations to their treaty signatures. While most countries have thus acknowledged the need for regulations to minimize the harm the use of drug has on society, the means to achieve this end are hotly debated in most democratic societies. Proponents of liberalization policies argue that "... the social and economic costs of cannabis prohibition outweigh any benefits ..." [21]. Prohibitionists argue with similar conviction that rates of cannabis use would be higher if it were legal. Numerous expert inquiries have been set up around the world but few seem to have had a significant impact on the thinking of the government.

Many countries approach cannabis control from the enforcement angle, with users and dealers being liable to prosecution and imprisonment. In the United States, cannabis is listed as a Schedule I substance along with heroin. Courts in a few countries, including Singapore (500 g) and Malaysia (200 g), may impose the death penalty for possession of cannabis.

The Netherlands, by contrast, distinguished itself from other Western countries in 1976 by decriminalizing the possession or sale of small quantities of cannabis. Differences in ethnic, cultural, and socioeconomic factors make it difficult to evaluate which approach is more successful at harm minimization. Certainly the United States has a higher prevalence of cannabis abuse than any of the 32 Western and Central European countries detailed in a 2006 report [22], whereas The Netherlands is ranked in the middle. More important, perhaps, is the fact that The Netherlands has a significantly lower prevalence of opiate, cocaine, and amphetamine abuse than the United States. The Netherlands has a higher prevalence than the United States in only one of the five drug classes documented in the report, namely, ecstasy.

Several countries have introduced legislation to control the use of cannabis by motor vehicle drivers. Studies have demonstrated that the use of cannabis especially in conjunction with alcohol significantly increases the risk of road crashes. The correlation between THC blood concentrations and impairment, however, is not strong. It is only possible to conclude that a person may have been impaired at the time of crash if they have high-blood THC levels indicative of recent use. Several countries (e.g., Germany) have addressed this by specifying blood THC concentrations above which it is illegal to drive.

Oral fluid testing for THC is increasingly being considered as a viable option for roadside testing of drivers, in part, because sample collection is relatively noninvasive. Successful trials in Victoria have led to legislation in other Australian states [23]. Improvements in technology are needed to enable the window of detection to be extended beyond about 2 h and to speed up the test procedure.

Conclusions

The use of Cannabis has had a long and colorful history. The complex chemical mixture in cannabis produces a wide variety of behavioral and physiological effects after smoking or oral ingestion. While some of these effects may produce pleasurable experiences, there are also significant adverse health consequences for some users. Potential medical applications of cannabinoids are emerging but more research is required.

Abuse of the drug has increased rapidly since the end of World War II and this has led to legislation to control its use. Whether harm minimization can be achieved more effectively by cannabis prohibition or decriminalization is still a matter of strong disagreement. The continuing increase in rates of abuse, however, points to the need for better policies.

Oral fluid testing for the use of cannabis looks set to have a significant role in improving road safety.

References

[1] Booth, M. (2004). *Cannabis: A History*, Bantam, London.

[2] Huestis, M.A. (2002). Cannabis (Marijuana) – effects on human behavior and performance, *Forensic Science Review* **14**, 15–60.

[3] ElSohly, M.A. & Slade, D. (2005). Chemical constituents of marijuana: the complex mixture of natural cannabinoids, *Life Sciences* **78**, 539–548.

[4] Grotenhermen, F. (2006). Cannabinoids and the endocannabinoid system, *Cannabinoids* **1**, 10–14.

[5] ElSohly, M.A. & Jones, A.B. (1995). Drug testing in the workplace: could a positive test for one of the mandated drugs be for reasons other than illicit use of the drug? *Journal of Analytical Toxicology* **19**, 450–457.

[6] Dussy, F.E., Hamberg, C., Luginbuhl, M., Schwerzmann, T. & Briellmann, T.A. (2005). Isolation of Δ^9-THCA-A from hemp and analytical aspects concerning

the determination of Δ^9-THC in cannabis products, *Forensic Science International* **149**, 3–10.
[7] Iversen, L.L. (2000). *The Science of Marijuana*, Oxford University Press, New York.
[8] Robson, P. (2005). Human studies of cannabinoids and medicinal cannabis, *Handbook of Experimental Pharmacology* **168**, 719–756.
[9] Drummer, O.H. (2001). *The Forensic Pharmacology of Drugs of Abuse*, Oxford University Press, New York.
[10] Pertwee, R.G. (1999). Cannabinoid receptors and their ligands in brain and other tissues, in *Marihuana and Medicine*, G.A. Nahas, K.M. Sutin, D.J. Harvey & S. Agurell, eds, Humana Press, Totowa, pp. 177–187.
[11] Pertwee, R.G. (1998). Advances in cannabinoid receptor pharmacology, in *Cannabis: The Genus Cannabis*, D.T. Brown, ed, Harwood Academic Publishers, Amsterdam, pp. 125–174.
[12] Dickson, S., Park, A., Nolan, S., Kenworthy, S., Nicholson, C., Midgley, J., Pinfold, R. & Hampton, S. (2007). The recovery of illicit drugs from oral fluid sampling devices, *Forensic Science International* **165**, 78–84.
[13] Teixeira, H.M., Verstraete, A., Proenca, P., Corte-Real, F., Monsanto, P.V. & Vieira, D.N. (2006). Validated method for the determination of Δ^9-tetrahydrocannabinol (THC), 11-hydroxy-Δ^9-THC and 11-nor-9-carboxy-Δ^9-THC in oral fluid, urine and blood using solid-phase extraction and liquid chromatography-mass spectrometry with electrospray ionization, in *Programme and Abstracts The International Association of Forensic toxicologists (TIAFT) 44th International Meeting*, M.Z. Karlovsek, ed, Ljubljana, p. 208.
[14] Huestis, M.A., Henningfield, J.E. & Cone, E.J. (1992). Blood cannabinoids. I. Absorption of THC and formation of 11-OH-THC and THCCOOH during and after smoking marijuana, *Journal of Analytical Toxicology* **16**, 276–282.
[15] Agurell, S., Halldin, M., Lindgren, J.E., Ohlsson, A., Widman, M., Gillespie, H. & Hollister, L. (1986). Pharmacokinetics and metabolism of delta-tetrahydrocannabinol and other cannabinoids with emphasis on man, *Pharmacology Reviews* **38**, 21–43.
[16] Niedbala, R.S., Kardos, K.W., Fritch, D.F., Kunsman, K.P., Blum, K.A., Newland, G.A., Waga, J., Kurtz L., Bronsgeest, M. & Cone, E.J. (2005). Passive cannabis smoke exposure and oral fluid testing. II. Two studies of extreme cannabis smoke exposure in a motor vehicle, *Journal of Analytical Toxicology* **29**, 607–615.
[17] Ohlsson, A., Lindgren, J.E., Wahlen, A., Agurell, S., Hollister, L.E. & Gillespie, H.K. (1980). Plasma delta-9-tetrahydrocannabinol concentrations and clinical effects after oral and intravenous administration and smoking, *Clinical Pharmacology and Therapeutics* **28**, 409–416.
[18] Niedbala, R.S., Kardos, K.W., Fritch, D.F., Kardos, S., Fries, T. & Waga, J. (2001). Detection of marijuana use by oral fluid and urine analysis following single-dose administration of smoked and oral marijuana, *Journal of Analytical Toxicology* **25**, 289–303.
[19] Huestis, M.A. & Cone, E.J. (2004). Relationship of Δ^9-tetrahydrocannabinol concentrations in oral fluid and plasma after controlled administration of smoked cannabis, *Journal of Analytical Toxicology* **28**, 394–399.
[20] Wall, M.E., Sadler, B.M., Brine, D., Taylor, H. & Perez-Reyes, M. (1983). Metabolism, disposition and kinetics of delta-9-tetrahydrocannabinol in men and women, *Clinical Pharmacology and Therapeutics* **34**, 352–363.
[21] Hall, W. (2007). A cautious case for cannabis depenalization, in *Pot politics – Marijuana and The Costs of Prohibition*, M. Earleywine, ed, Oxford University Press, New York.
[22] Chawla, S. & le Pichon, T. (*The World Drug Report 2006*, 2006). The Office for Drug Control and Crime, United Nations Publication.
[23] Drummer, O.H. (2006). Drug testing in oral fluid, *Clinical Biochemistry Reviews* **27**, 147–159.

STUART DICKSON AND HELEN POULSEN

Capacity Assessment

Introduction

The right to choose is a fundamental societal value, but in certain situations, an individual's ability to make personal choices may be questioned. In such situations, individual autonomy must be weighed against another societal interest – that of protecting the individual. Competency as a legal term, defined by Black's Law Dictionary as "the mental ability to understand problems and make decisions," is closely related to the clinical issue of capacity [1]. Although mental health professionals may make an assessment of capacity, the determination of competency is reserved for the judicial system, and competency requirements may vary from one jurisdiction to another. The common fundamental principles of competency are the focus of this article.

Assessment of competence is based on four elements, specifically with regard to having the requisite mental skills for the acquisition, comprehension, and processing of relevant information in order to make a decision [2]. Essentially, an individual must understand the basic facts of a given situation. Once a factual understanding is attained, the individual must appreciate the implications of the situation and its relation to himself or herself. On that foundation,

the individual must be able to utilize this understanding to rationally assess the risks and benefits of the possible choices. Finally, a competent individual must be able to communicate that choice. For example, a person who articulates a reasoned choice but subsequently changes his or her decision too frequently to allow implementation of a particular choice might be considered incompetent. In order to be considered competent, the choice made by an individual need not be the course of action that most people would prefer. Rather, the question rests on whether the process of choosing was based on the reasoned application of adequate factual understanding.

Capacity may be significantly impacted by preexisting cognitive limitations or an underlying psychiatric disorder. However, the presence of a mental illness alone is insufficient to indicate a lack of capacity. A context-specific functional impairment must also be present. Just as symptoms of mental illness may improve or worsen, capacity similarly may change over time and may require reassessment as circumstances change. Clarifying the type and extent of impairment resulting from mental illness is necessary in determining capacity. There are numerous assessment tools, including the MacArthur Competence Assessment Tools [3] to aid in this task. However, there is no standard tool to evaluate capacity, and evaluations typically include minimum clinical interview and review of available records.

Competency issues arise in both civil and criminal arenas, and in general, the court conducts such evaluations in an effort to safeguard the individual's rights. As the determination of competency for specific issues depends upon skills and abilities specific for each issue, an individual may have competence to decide one issue but not another. As such, a capacity evaluation must address a specific question. In civil settings, such areas include wills, contracts, treatment, and research participation. For criminal settings, competence may be questioned in relation to trial procedures and waiving various rights, including Miranda rights (*see* **Capacity to Waive Miranda Rights**), the right to privilege, to have legal counsel, and to posing an insanity defense. Competency issues may also arise in the sentencing phase and may become particularly contentious in capital cases.

Civil Law

Testamentary Capacity

The determination of testamentary capacity requires that the testator understand the following issues: (i) that he or she is making a will and that the will specifies how his or her estate will be distributed after death, (ii) the extent of his or her bounty, or possessions to be distributed, and (iii) which individuals would be considered the natural heirs. From a legal perspective, competence is presumed until challenged, and the burden of proof rests with the person challenging the will [2, 4, 5].

Testamentary capacity may be questioned, particularly when a preexisting will is significantly altered or if natural heirs are excluded. Legal challenges typically arise after the death of the testator. This creates particular challenges in assessing testamentary capacity, and assessments become retrospective and limited to witness statements, medical records, and other documentation. In some instances, attorneys anticipating potential challenges to a will may have arranged for a contemporaneous evaluation of the testator at the time the will was written. Additionally, some testators may leave a durable record by having the evaluation videotaped. An example of illness contributing to a lack of testamentary capacity is dementia resulting in increased confusion, reduced orientation, or in delusional beliefs that may impact a testator's ability to understand the basic elements of making a will.

The question of undue influence may arise if there are suspicions that the testator was coerced or manipulated into altering a will. The presence of undue influence may void a will even if the testator otherwise has capacity. The presence of the following may suggest undue influence: (i) the relationship between the testator and the beneficiary allowed for the beneficiary to exert control over the will, (ii) "unnatural provisions" in the will, and that (iii) the will did not accurately represent the testator's wishes [5]. A testator may be more susceptible to undue influence as a result of medical o-psychiatric illness such as those described earlier. Should a testator be available for examination, the involvement of another individual besides the testator or his or her attorney may indicate undue influence, particularly if that individual makes the appointment, brings the testator to the appointment, or answers questions for the testator. Additionally, uncertainty on the part of

the testator regarding details of the will may indicate undue influence [6].

In assessing testamentary capacity, an evaluator should also consider the ethical implications. A competent will ensures the appropriate distribution of property and may be of great importance to potential beneficiaries. However, finding a testator incompetent after death negates the last wishes of the deceased with regard to the disposition of their estate [2].

Contractual Capacity

A contract is, by definition, "an agreement between two or more parties creating obligations that are enforceable or otherwise recognizable at law" [1]. Contracts, both implied and written, are essential to society. Competence to contract, as in testamentary competence, is presumed, but unlike writing a will, entering into a contract is a potentially adversarial process. Contractual competence involves three essential elements: (i) the individual understands the various aspects of the transaction, (ii) the relationship of those aspects is appreciated, and (iii) the individual is able to make a rational judgment based on this understanding. If an individual lacks the requisite understanding due to mental illness or defect, a contract may be nullified [7]. In some jurisdictions, if an individual would not have entered a contract "but for" mental illness, the contract is invalid [2].

In determining whether to void a contract, courts consider both the issues of whether incompetence is proven and whether the status quo can be restored. In instances where the status quo cannot be restored, consideration is given to the fairness of the contract, including the possibilities of undue influence or fraud [2].

Capacity to Consent to or Refuse Treatment

In clinical settings, capacity concerns may arise in relation to consent for treatment as well as the converse, to refuse treatment, be that treatment psychiatric or medical/surgical in nature. Justice Benjamin Cardozo described the concept of consent in 1914 when he wrote, "Every human being of adult years and sound mind has a right to determine what shall be done with his own body" [8]. In the present day, patient autonomy remains an important concern [9]. Consent must be informed, voluntary, and competent. (*see* **Treatment, Right to Refuse: Mental Health**; **Treatment, Mandated: Mental Health**).

With informed consent, a competent individual must understand: (i) the nature of the medical condition, (ii) the nature of the proposed treatment, (iii) the risk and benefits of the proposed treatment as well as no treatment, and (iv) possible alternatives to the proposed treatment [2, 10]. However, as in the case of many capacity evaluations, there is no one clinical assessment tool to evaluate the capacity to consent to treatment. As a result, the assessment entails obtaining clinical history in addition to considering cognitive functioning, overall day-to-day functioning, and mental state [11, 12]. A patient's understanding of these basic topics requires a fundamental level of communication from the treatment provider [12]. Prior medical knowledge is not necessary as long as there is an ability to learn, retain, and process information in a rational manner. A key point underlying consent issues is that an individual need not make the "right" or expected decision but a reasoned choice. Even with life-saving treatment, a competent individual must be allowed to refuse treatment if he or she so chooses.

Competent individuals may establish advance directives or a living will outlining their treatment preferences in the event they become incompetent at a later time. Additionally, a proxy may be designated to provide informed consent should the individual become incompetent [13]. In the absence of previously designated decision makers, families may be utilized as substitute decision makes, with preference given to spouses and close relations. Otherwise, the matter becomes a judicial issue, and a court hearing may be requested. In many jurisdictions, a specific "mental health court" addresses treatment issues related to psychiatric illnesses. In this venue, both the patient and the treatment provider have opportunity to present their arguments before a judge.

Medical emergencies, when treatment decisions regarding life-threatening conditions must be made in the moment, constitute an exception to the requirement for consent [10]. For instance, informed consent is not required prior to treating an unconscious patient who requires surgery to stop severe internal bleeding. Similarly, in psychiatric settings, emergency treatment to prevent injury to oneself or to others may be given without patient consent.

Another exception to informed consent is patient waiver. A patient may choose not to be informed; in

this situation, the patient must have the capacity to understand the consequences of not being informed. More rarely used is therapeutic privilege – when a clinician determines that a discussion of risks and benefits of proposed treatment would harm the patient or cause distress to the extent that an informed decision could not be made. A substitute decision-maker would instead be informed and provide consent [10].

An area related to capacity to accept or refuse a specific treatment is that of civil commitment. Inpatient psychiatric treatment often occurs in locked settings where patients cannot simply come and go without agreement from the treatment provider. Civil commitment, therefore, entails a certain curtailment of personal liberty. Further, besides the loss of liberty, courts have noted civil commitment may be associated with social stigma. Patients may agree to inpatient treatment, in which case they are considered voluntary patients. In order to have to capacity to consent to voluntary inpatient psychiatric treatment, an individual must have a mental illness that may benefit from treatment, must understand that treatment occurs in a mental health setting, and must consent free of coercion [2, 10]. In instances where an individual does not agree with inpatient treatment, he or she may be admitted involuntarily but may also challenge the admission in a court hearing. Often, evidence that the individual is a danger to him or herself or to others as a result of their mental illness is required.

Capacity to Consent to Research

Clinical research is essential in increasing our understanding of medical illnesses and in furthering treatment advances. Given the experimental nature of research, informed consent includes understanding of (i) the nature and purpose of research, (ii) the risks and benefits, (iii) alternative treatments to the research protocol, (iv) limits of confidentiality, and (v) compensation [14]. In the area of mental health, the issue of competence to consent to research becomes more complicated, as the very cause of the symptoms targeted in research may also render a participant incompetent [15]. The standards for competency are (i) a factual understanding of relevant issues, (ii) an ability to manipulate that information in a meaningful manner, (iii) an ability to appreciate the nature of the situation, and (iv) the ability to make and articulate a choice regarding participation in research.

Capacity to Waive Privilege

Privilege is an individual's right to protect confidential communications occurring in the context of relationships seen to have societal value, such as attorney–client, doctor–patient, husband–wife, and priest–penitent relationships [16]. However, various exceptions exist, including patient-initiated litigation involving issues of mental health or mental status, court-ordered evaluations, civil commitment, and treatment refusal proceedings, instances in which a patient poses an emergent danger, and situations of child protection or child custody [2]. As privileged information may be damaging, when an individual waives privilege outside these exceptions, capacity to do so becomes important. The individual should be aware of what information may be revealed and what potential impact that information may have on the legal question (*see* **Capacity to Waive Miranda Rights**).

Criminal Law

Capacity to Proceed with Trial

The intricacies of courtroom trial procedures require an ability to understand rules and roles beyond average day-to-day functioning (*see* **Capacity to Stand Trial**). In this adversarial setting, the defendant's ability to interact with their attorney and aid in their own defense is crucial to the fairness of a trial. Just as cognitive impairment or symptoms of mental illness may interfere with competence to make informed treatment decisions, they may also interfere with competence to proceed with trial. For example, mental retardation may prevent a clear, factual understanding of the circumstances of the instant offense or of court procedure. A depressed defendant may be unmotivated to assist in their defense. Schizophrenic defendants with conceptual disorganization may have an impaired ability to understand, reason, and appreciate legal procedures [17].

Historically, competence to stand trial was ensured through duress. In 17th century England, trial could not proceed until a defendant entered a plea. To encourage the defendant to plea, heavy rocks were

placed on the defendant's chest until a plea was given or the defendant was suffocated [2]. After the use of this method was abolished, defendants found incompetent to plead by a jury were merely imprisoned until they could voice a plea. In modern times, a finding of incompetence entails a court mandated psychiatric evaluation of capacity that is used to inform the determination.

In the United States, the standard for competence was established by the 1960 Supreme Court case of *Dusky v. United States* [18]. The defendant, along with two codefendants, was accused of abducting a teenaged girl and transporting her across state lines. The codefendants were accused of rape, and Dusky was accused of attempted rape. Prior to trial, Dusky was detained at a federal medical center for evaluation and found to have a psychotic illness complicated by alcoholism. The evaluator noted that although Dusky understood his charges and basic criminal procedure, he was "unable properly to assist" in his defense due to confusion, suspiciousness, and "an inability to interpret reality from unreality." The trial court, however, maintained that Dusky was oriented and able to assist counsel. After Dusky's conviction, the case was appealed and eventually reversed by the Supreme Court. As articulated by the Court's holding, the proper test for competence is "whether the accused has sufficient present ability to consult with his lawyer with a reasonable degree of rational understanding and whether he has a rational as well as a factual understanding of the proceedings against him."

In the Insanity Defense Reform Act of 1984, the United States federal standard stipulated that incompetence requires that "present mental disease or defect renders the defendant unable to understand the nature and consequences of the proceedings against him or to assist in his defense" [16].

Correct assessment of capacity is essential, as failure to recognize impaired ability may render a court proceeding unfair. Inaccurately finding a defendant incompetent, on the other hand, may unnecessarily delay court proceedings, possibly lengthening the duration of confinement or causing involuntary commitment and treatment [17].

Essentially, a competent defendant understands the nature of the charges as well as court proceedings and is able to assist his or her attorney in his or her own defense. Various assessment tools have been assembled in an effort to delineate the specific skills necessary to fulfill these general requirements [19]. The McGarry Scale articulates 13 different functional abilities: (i) ability to appraise available legal defenses, (ii) level of unmanageable behavior, (iii) quality of relating to attorney, (iv) ability to plan legal strategy, (v) ability to appraise the roles of various participants in the courtroom proceedings, (vi) understanding of court procedure, (vii) appreciation of the charges, (viii) appreciation of the range and nature of possible penalties, (ix) ability to appraise the likely outcomes, (x) capacity to disclose to the attorney available pertinent facts surrounding the defense, (xi) capacity to challenge prosecution witnesses realistically, (xii) capacity to testify relevantly, and (xiii) manifestation of self-serving versus self-defeating motivation [2]. The MacArthur Studies also addressed adjudicative competence, with development of the MacArthur structured assessment of the competencies of criminal defendants (MacSAC-CD). Research utilizing the MacSAC-CD indicated that the presence of psychotic symptoms regardless of diagnosis was associated with a determination of incompetence; a diagnosis of schizophrenia versus other diagnoses was similarly more strongly associated with incompetence [20]. Despite this association, no one diagnosis or symptom necessarily equates the lack of competence to proceed with trial, and just as with other competencies, the lack of competence to proceed with trial does not necessarily indicate a lack of competence in other areas.

Should a defendant be incompetent to proceed with trial, the trial process must be halted, at least temporarily. Incompetent defendants may be committed to long-term psychiatric facilities to restore capacity. However, restoration may not always be possible, particularly if the primary barrier may not be lifted with treatment. Examples may include stable conditions such as mental retardation, progressive illnesses such as dementia, or illnesses with severe, treatment-resistant symptoms. In these situations, a defendant could face commitment for a duration exceeding the possible sentence for a guilty plea. The case of *Jackson v. Indiana* in 1972 involved a mentally retarded and deaf-mute man arrested for two robberies, the value of which totaled nine dollars. He was found incompetent to stand trial and committed for restoration of competency. His attorney appealed on the basis that Jackson was unlikely to become competent given his deficiencies; commitment was tantamount

to a life sentence despite lack of a guilty verdict. The Supreme Court agreed, stating, "due process requires that the nature and duration of commitment bear some reasonable relation to the purpose for which the individual is committed" [21].

Capacity to Waive Legal Counsel

Besides the broader topic of capacity to proceed with trial, more specific capacity issues focus on a defendant's capacity to waive certain rights, such as the right to have legal representation. As discussed earlier, courtroom procedure involves various intricacies with rules and roles not commonly associated with average daily tasks, and many defendants rely on an attorney trained and experienced in legal problems to guide and advise them. Consequently, the decision to waive the right to counsel potentially carries significant consequences, and the ability to make such a decision may become called into question. There are no formal standards in specifically assessing capacity to waive counsel, but this may be assessed as part of an overall capacity to proceed with trial evaluation. Basic considerations are that the defendant understands the role of an attorney and has a rational understanding of the benefits of having an attorney as well as the risks of representing oneself *pro se*.

The US Supreme Court, in the 1993 ruling on *Godinez v. Moran*, held that a defendant's choice to represent himself or herself is contingent on the competence to choose self-representation, not on his or her actual qualifications for self-representation [22]. This position is not undisputed, and an opposing view is that the defendant's functional ability to self-represent is an essential consideration in establishing competence to waive counsel [19.]

Capacity to Waive the Insanity Defense

Despite the perception that the insanity defense is overused and a way to "get off easy," the insanity defense is successful in less than one in two thousand felony cases [2]. In addition, cases arise in which the insanity defense is recommended by defense counsel, but the defendant does not agree. Whether a criminal defendant with a plausible insanity defense has the right to waive it has been a point of contention. Defendants may decline an insanity defense for an assortment of reasons – they may not believe they have a mental illness, may wish to avoid the stigma of being an insanity acquittee, or may be concerned that they may be committed to a psychiatric institution for a longer period than the sentence for being found guilty.

In some jurisdictions, the insanity defense can be imposed against a defendant's wishes, even if the defendant is considered otherwise competent. The rationale is that a defendant's choice, even a competent one, is outweighed by a greater societal interest in the "just determination of the charge against the defendant." An alternative view is that the insanity defense, as a specific legal defense, is encompassed within general evaluations of competency to proceed with trial. According to this interpretation, a competent defendant ought not to have the insanity defense imposed upon them against their will [23].

Capacity for Sentencing

The purpose of sentencing guilty defendants is threefold: (i) to punish the perpetrator, (ii) to act as a deterrent for future offenses, and (iii) to convey a sense of justice. During the sentencing phase, a defendant may introduce evidence that may potentially mitigate the sentence or refute aggravating circumstances that may result in a more severe penalty. For these reasons, capacity remains an important issue. A defendant lacking capacity for sentencing may be unable to make the court aware of important mitigating factors or may even be unable to appreciate the meaning of the sentence itself [24]. In most jurisdictions, competency for sentencing is determined by the same process as competency to stand trial.

Capacity for Execution

Capital punishment is a contentious ethical issue in itself; adding to the conflict are those cases in which the capacity of an inmate to be executed in uncertain. The traditional arguments favoring the death penalty – punishment, justice, and deterrence – come into question when the inmate is incompetent. Historically, in English common law, the execution of an insane person was thought to be "savage and inhuman." Reasons provided spanned the unfairness of trying a defendant unable to make his defense to the lack of example to others [16].

The case of *Ford v. Wainwright* [25] involved the murder conviction of Alvin Bernard Ford in 1974. During trial, there was no question of incompetence,

and he was sentenced to death. Beginning in 1982, he began to demonstrate gradual behavioral changes. Over time, he became increasingly delusional. Two psychiatrists for the defense evaluated Ford, who concluded that he was genuinely psychotic. His attorneys then requested a determination of competency; three state psychiatrists concluded that Ford was able to understand his situation at the time, the nature of the death sentence, and why he had been so sentenced. On appeal to the Supreme Court, the Court ruled that under the Eighth Amendment prohibition of cruel and unusual punishment, execution of an insane prisoner was also prohibited, and therefore Ford had the right to a judicial hearing to determine competency. The Court provided six reasons for ruling that an inmate must be competent prior to execution: (i) an incompetent person might be unable to provide last minute information useful for vindication, (ii) insanity is already a punishment, (iii) an incompetent person cannot make peace with God, (iv) execution of an incompetent person does not deter others, (v) execution of an incompetent person is extreme inhumanity and cruelty, and (vi) retribution cannot be exacted from an incompetent person [25].

The specific criteria for determining capacity to be executed vary by jurisdiction. For *Ford*, the Florida standard stipulated that those to be executed have the capacity to understand the nature of the death penalty and the rationale for its imposition. The Court deemed this standard sufficient and is followed by most states. The alternative standard for competency to be executed is whether the inmate, due to mental disease or defect, is unable to understand the nature of the proceedings, the accusation, the purpose of punishment, and the punishment itself. This standard made additional note that the inmate should be able to convey factual information that may change the verdict or sentence to defense counsel [16].

The Supreme Court, regarding capital punishment and the mentally retarded, has reached separate decisions. In the 1989 case of *Penry v. Lynaugh*, the Court ruled that mental retardation does not equate incompetence to be executed, but may be used as a mitigating factor during sentencing [26]. However, in 2000, the Court in *Atkins v. Virginia* that execution of the mentally retarded constituted cruel and unusual punishment and therefore violated the Eighth Amendment [27] (*see* **Mental Retardation: Death Penalty**; **Death Penalty and Age**).

Significant clinical-ethical concerns arise with relation to determinations by mental health professionals of capacity to be executed, particularly if the evaluator supports a finding of competency, thereby furthering execution. Although the capacity evaluation in itself does not constitute active participation in an execution, some psychiatrists view any involvement in the death penalty to be unethical. Alternatively, such evaluations may spare some inmates from execution. If a prisoner is found incompetent, however, the death penalty is typically not vacated. In such instances, another ethical dilemma arises regarding the question of treating an individual to restore competency for execution. The opinion of the American Medical Association is that in this situation, unless the death sentence is commuted, treatment should only be given for the purposes of relieving extreme suffering [28].

References

[1] Garner, B.A. (ed) (2004). *Black's Law Dictionary*, 8th Edition, Thomson West, St. Paul.

[2] Gutheil, T.G. & Appelbaum, P.S. (2000). *Clinical Handbook of Psychiatry and the Law*, 3rd Edition, Lippinocott Williams & Wilkins, Philadelphia.

[3] Dunn, L.B., Nowrangi, M.A., Palmer, B.W., Jeste, D.V. & Saks, E.R. (2006). Assessing decisional capacity for clinical research or treatment: a review of instruments, *The American Journal of Psychiatry* **163**(8), 1323–1334.

[4] Jacoby, R. & Steer, P. (2007). How to assess capacity to make a will, *BMJ* **335**(7611), 155–157.

[5] Shulman, K.I., Cohen, C.A., Kirsh, F.C., Hull, I.M. & Champine, P.R. (2007). Assessment of testamentary capacity and vulnerability to undue influence, *The American Journal of Psychiatry* **164**(5), 722–727.

[6] Ciccone, J.R. (2003). Civil competencies, in *Principles and Practice of Forensic Psychiatry*, 2nd Edition, R. Rosner, ed, Arnold, London, pp. 309–312.

[7] Sprehe, D.J. (2003). Geriatric psychiatry and the law, in *Principles and Practice of Forensic Psychiatry*, 2nd Edition, R. Rosner, ed, Arnold, London, pp. 651–660.

[8] Schloendorff v. Society of New York Hospital, 211 N.Y. 125 (1914).

[9] Doyal, L. & Sheather, J. (2005). Mental health legislation should respect decision making capacity, *BMJ* **331**, 1467–1468.

[10] Schwartz, H.J. & Mack, D.M. (2003). Informed consent and competency, in *Principles and Practice of Forensic Psychiatry*, 2nd Edition, R. Rosner, ed, Arnold, London, pp. 98–106.

[11] Lai, J.M. & Karlawish, J. (2007). Assessing the capacity to make everyday decisions: a guide for clinicians and an agenda for future research, *The American Journal of Geriatric Psychiatry* **15**(2), 101–111.

[12] Terry, P.B. (2007). Informed consent in clinical medicine, *Chest* **131**, 563–568.
[13] Nys, H., Welie, S., Garanis-Papadatos, T. & Ploumpidis, D. (2004). Patient capacity in mental health care: legal overview, *Health Care Analysis* **12**(4), 329–337.
[14] Appelbaum, P.S. & Roth, L.H. (1982). Competency to consent to research: a psychiatric overview, *Archives of General Psychiatry* **39**, 951–958.
[15] Appelbaum, P.S., Grisso, T., Frank, E., O'Donnell, S. & Kupfer, D.J. (1999). Competence of depressed patients to consent to research, *The American Journal of Psychiatry* **156**(9), 1380–1384.
[16] Reisner, R., Slobogin, C. & Rai, A. (2004). *Law and the Mental Health System: Civil and Criminal Aspects*, 4th Edition, Thomson West, St. Paul.
[17] Hoge, S.K., Poythress, N., Bonnie, R.J., Monahan, J., Eisenberg, M. & Feucht-Haviar, T. (1997). The MacArthur adjudicative competence study: diagnosis, psychopathology, and competence related abilities, *Behavioral Science and the Law* **15**, 329–345.
[18] Dusky v. United States, 362 U.S. 402, 80 S. Ct., 788 (1960).
[19] Cruise, K.R. & Rogers, R. (1998). An analysis of competency to stand trial: an integration of case law and clinical knowledge, *Behavioral Science and the Law* **16**, 35–50.
[20] Poythress, N.G., Bonnie, R.J., Monahan, J., Otto, R. & Hoge, S.K. (2002). Research issues in adjudicative competence, in *Adjudicative Competence, the MacArthur Studies*, Kluwer Academic/Plenum Publishers, New York, pp. 91–110.
[21] Jackson v. Indiana, 406 U.S. 715 (1972).
[22] Godinez v. Moran, 113 U.S. 2680 (1993).
[23] Miller, R. (2002). *Hendricks v. People*: forcing the insanity defense on an unwilling defendant, *The Journal of the American Academy of Psychiatry and the Law* **30**, 295–297.
[24] Perlin, M.L. (2003). Beyond Dusky and Godinez: competency before and after trial, *Behavioral Science and the Law* **21**, 297–310.
[25] Ford v. Wainwright, 106 S. Ct. 2595 (1986).
[26] Penry v. Lynaugh 109 S. Ct 2934 (1989).
[27] Atkins v. Virginia 260 Va. 375; 534 S.E. 2d 312 (2000).
[28] AMA (2000). *Current Opinions of the Council on Ethical and Judicial Affairs*, www.ama-assn.org.

LI-WEN G. LEE

Capacity Consent to Treatment

see Capacity Assessment

Capacity for Guardianship ***see*** Guardianships of Adults

Capacity for Independent Living

Questions about independent living capacity are typically raised by family members, healthcare providers, lawyers, and/or concerned others when an individual is exhibiting or is perceived to exhibit deficits in independent functioning. The issue may become especially urgent when behavior is unsafe, such as when one lacks the cognitive capacity to understand that refusing to eat, bathe, or use a mobility support (e.g., cane) may pose a serious medical risk. It is vital to note that concerns about decisional capacity should not be based on judgments about whether an individual has made "acceptable" choices, but rather on whether one has the cognitive capacity to adequately reason through such choices. Although questions about independent living capacity may be relevant for any adult exhibiting deficits in independent functioning, concerns are more frequent in specific populations, such as individuals with neurological problems, developmental disabilities, psychiatric issues, and substance abuse, and are frequently encountered in adults aged 60 and older with these diagnoses. Since concerns about independent living capacity in older adults will likely increase as this population doubles in the next 25 years [1], this article includes a focus on capacity in older adults, although discussion points are generally broad enough to be tailored to other cases. An invaluable supplement to this information is found in the publication, *Assessment of older adults with diminished capacity: A handbook for lawyers* [2], which provides an excellent overview of practical, ethical, and theoretical considerations regarding decisional capacity, as well as worksheets that can be used when meeting with clients.

Myths about Decisional Capacity

Before defining the concepts of independent living and its assessment, a discussion of some global myths

regarding decisional capacity assessment is helpful [3]. First, decisions about capacity are not "all or nothing" phenomena, and clinical judgments may reflect this by including recommendations for assistance in one area of functioning but not another (e.g., assistance in medication management but not financial management). Such specification of the need for assistance in specific domains, but not in others, is reflected by some states through limited guardianship. Second, a diagnosis in and of itself does not infer decisional capacity. For example, an individual with Alzheimer's disease or schizophrenia may have compromised memory or reality testing, but intact daily functioning and decisional abilities. Conversely, an individual with no cognitive or psychiatric impairment may exhibit impaired decision making. Thus, "it is important that functional performance be defined, measured, and interpreted separately from cognitive performance" [4].

Preserving Autonomy

Several research studies suggest that if older adults lose control of their daily decisions or environment, quality of life can become significantly compromised and mortality rates may increase (see [5] for a review). Clinical observations support these findings, and suggest that some individuals exhibit significant psychological reactions if decisional rights are constrained or revoked. Thus, most clinicians are extremely mindful about the potential impact of decisional capacity evaluations. Most often, there is an attempt to strike a balance between safety and autonomy when making recommendations, with a focus on improving quality of life regardless of decisional ability.

Requirements for Independent Living

Defining Incapacity

The Uniform Guardianship and Protective Proceedings Act (UGPPA) defines an incapacitated individual as one who is unable to receive and evaluate information or make or communicate decisions to such an extent that the individual lacks the ability to meet essential requirements for physical health, safety, or self-care, even with appropriate technological assistance [6]. While incapacity statutes in three states have incorporated the 1997 Act (Colorado, Minnesota, Hawaii), other states have utilized information from the 1982 version of the Act, or from other state-specific sources [7].

What Skills Are Necessary for Independent Living?

Although there is absolute definition of the functional tasks that comprise independent living capacity, two key sources – clinical definitions of independent living capacity and the *Handbook for Judges* [8] – provide helpful guidance.

Clinical definitions of independent living capacity commonly include the constructs of "activities of daily living" (ADLs) and "instrumental activities of daily living" (IADLs). Although triggers for concern about independent living capacity are varied, common scenarios include compromised performance of ADLs or IADLs. While there is no authoritative or complete list of ADLs or IADLs, ADLs often refer to basic self-care skills such as dressing, eating, toileting, transferring, or moving from one sitting position to another, walking and mobility, or bathing [7]. IADLs are more cognitively complex than ADLs and typically include management of finances and medications, medical decision making, transportation use, and management of household functions such as shopping for food and other necessities, preparing meals, doing housework, and doing laundry [4]. Given that IADLs are more cognitively complex than ADLs, IADLs may be compromised before ADLs in individuals with cognitive difficulties. The *Handbook for Judges* also provides a list of general tasks associated with independent living (see Table 1), and provides a supplement including a detailed checklist ([8], p. 33).

Cognitive Functioning

Questions about independent living capacity often relate to concerns about cognitive impairment. Although age-appropriate memory loss occurs throughout the lifespan, starting in the mid-20s [9], if cognitive decline becomes impaired beyond levels seen in normal aging and begins to interfere with daily functioning (e.g., medication management, financial management), a diagnosis of dementia may be considered if the condition leading to cognitive impairment is not reversible or temporary. Although

Table 1 Task associated with independent living[a]

Domain	Description
Care of self	Maintain adequate hygiene, bathing, dressing, toileting, dental Prepare meals and eat for adequate nutrition
	Identify abuse or neglect and protect self from harm
Financial	Protect and spend small amounts of cash Manage and use checks
	Give gifts and donations Make or modify will Buy or sell real property
	Deposit, withdraw, dispose, invest monetary assets
	Establish and use credit
	Pay, settle, prosecute, or contest any claim
	Enter into a contract, commitment, or lease arrangement
	Continue or participate in the operation of a business
	Employ persons to advise or assist him/her
	Resist exploitation, coercion, undue influence
Medical	Give/withhold medical consent
	Admit self to health facility
	Choose and direct caregivers
	Make or change an advance directive
	Manage medications
	Contact help if ill or in medical emergency
Home and community life	Choose/establish abode
	Maintain reasonably safe and clean shelter
	Be left alone without danger
	Drive or use public transportation
	Make and communicate choices about roommates
	Initiate and follow a schedule of daily and leisure activities
	Travel
	Establish and maintain personal relationships with friends, relatives, coworkers
	Determine degree of participation in religious activities
	Use telephone
	Use mail
	Avoid environmental dangers and obtain emergency help
Civil or legal	Retain legal counsel
	Vote
	Make decisions about legal documents

[a] Judicial Determination of Capacity of Older Adults in Guardianship Proceedings: A Handbook for Judges [8]

age is the greatest risk factor for dementia, dementia is not a natural consequence of aging. For example, in North America, only 0.8–1.6% of persons, ages 65–74, are diagnosed with dementia. Although the prevalence of dementia doubles every 5 years after age 74 (impacting 7–8% of persons 75–84 years old and 18–32% of persons over 85; [10]), most older adults are not diagnosed with dementia. Incidentally, the term *dementia* is often used interchangeably with "Alzheimer's disease", especially in the media. To clarify, Alzheimer's disease is the most common of several subtypes of dementia – describing about 50–70% of all cases – and refers to the cellular pathology that is presumably present, though confirmative diagnosis is not possible without autopsy. It is important to note that the increasing older adult population may lead to a projected 70% increase in the prevalence of Alzheimer's disease in the next 25 years [11].

Cognitive functioning may be compromised for brief periods of time because of acute stressors (e.g., delirium, seizures, medication issues, alcohol and drug use, and psychiatric issues, among others), it may be mildly impaired with the possibility of

worsening over time (mild cognitive impairment), and/or it may be permanently impaired owing to other conditions (e.g., irreversible dementia, neurodegenerative disorders, traumatic head injury). Thus, characterization of cognitive impairment as temporary, mild, or permanent and progressive plays a vital role in decisional capacity assessment.

Assessment Domains and Instruments

When decisional capacity is unclear or potentially compromised, a clinical evaluation can be a helpful tool in assessing and documenting capacity, and may assist in identifying the least restrictive alternatives to guardianship. Clinicians including psychologists and physicians often evaluate decisional capacity, and different assessment methods are often utilized. Psychologists are most likely to use psychometric tests to assess cognitive functioning and independent living, though this is not always the case. Typically, the most important criterion in selecting a clinician is his or her level of experience in evaluating decisional capacity [2].

Theoretical and Legal Components

Although there is no gold standard regarding the components of a typical capacity assessment, many clinicians find it helpful to structure the assessment by integrating theoretical principles, legal considerations, and clinical models. Theoretical information from Grisso's conceptual model [12] provides an important guide for the assessment process by integrating legal and clinical considerations into five domains – functional, causal, interactive, judgmental, and dispositional (*see also* **Capacity to Consent to Medical Treatment**).

Although legal requirements for assessment content may be state specific, suggested recommendations are included in the *Uniform Probate Code*/UPC Section 5–306 [13], and the *Handbook for Judges* [2]. Assessment outcomes from the UPC include (i) a description of the nature, type, and extent of the respondent's specific cognitive and functional limitations; (ii) an evaluation of the respondent's mental and physical condition and, if appropriate, educational potential, adaptive behavior, and social skills; (iii) a prognosis for improvement and a recommendation as to the appropriate treatment or habilitation plan; and (iv) the date of any assessment or examination upon which the report is based. The *Handbook for Judges* [8] recommends that a comprehensive evaluation cover "six pillars" of capacity, including the medical condition, cognitive functioning, everyday functioning, values and preferences, risk and level of supervision needed (including social support), and means to enhance capacity at the hearing and later.

Clinical Components

After a clinician has considered theoretical and legal issues, clinical models are often used to guide the actual practice of conducting the assessment. A publication by the Veterans Administration (VA) – *Assessment of Competency and Capacity of the Older Adult: A Practice Guideline for Psychologists* [14] – is particularly helpful in this regard. This model is discussed in the article entitled "Capacity to Consent to Medical Treatment" by Edelstein and Gould, and will be referred to here in the context of assessing independent living capacity. The VA model recommends the following five steps in conducting a capacity assessment.

Clarification of the Referral Question. The more specific the legal question, the better the clinician is able to structure a helpful assessment. For example, a common referral might include the following information: "Mrs X wants to live at home, but her family is concerned that her health has deteriorated due to poor hygiene and medication noncompliance. She refuses any assistance and wants to be left alone. Please assess her cognitive functioning and her capacity for making decisions about where she will live, her ability to manage medications, her self-care, and other common issues that might interfere with her ability to live independently." Since the construct of independent living is so multifaceted (recall the discussion of several ADLs and IADLs), referral questions need to be as specific as possible. On the other hand, referral questions do not need to include an exhaustive list of all problems. Rather, the question should serve as a starting point for a clinician who may uncover additional areas of concern during the evaluation.

Informed Consent. The nature of the evaluation as well as the purpose, risks, and benefits are explained

to the client. The client is asked to paraphrase the information provided, and is asked whether they consent to the evaluation. Consent or lack of consent is documented.

Conduct Evaluation. The typical assessment consists of four components:

Clinical Interview. An interview is conducted with the client to psychiatric history, medical history, social history, mental status, behavior, culture, religion, personality, perception of referral issues, and preferences for support (who would they like to help them, if necessary?). Given the subjectivity inherent in decision making, assessment of values and preferences is integral to this process (see *Handbook for Judges* p. 84 for a list of questions associated with independent living values [8]). The clinician will also note his or her behavioral observations of the client during the interview and testing.

Cognitive Testing. Cognitive tests are administered to assess functioning across multiple areas/"domains" of cognitive skills including executive functioning (reasoning, judgment, insight, mental flexibility), memory, attention, language, and visuospatial abilities, among others (see *Handbook for Judges* pp. 50–52 for details [8]). In essence, the process of cognitive testing is analogous to putting the brain on a "road test". Whereas neuroimaging tools (e.g., computed tomography (CT) scan or magnetic resonance imaging (MRI)) provide a picture of brain structure, cognitive testing is the only tool that can demonstrate how the brain actually *functions* in the outside world. Most cognitive tests are designed to be highly reliable and valid indicators of performance. Incidentally, when independent living capacity is compromised, the cognitive domain that is often most impacted is that of executive functioning [4].

Functional Testing. Whereas cognitive testing provides data about several different cognitive skills, functional tests allow for assessment of skills *specific* to independent living. This process may require the client to demonstrate specific behaviors associated with independent living (e.g., "show me who you would call in an emergency"), and demonstrate reasoning regarding potential situations ("if someone knocked on your door at 3 a.m., what would you do?"). Several psychological tests and interviews have been specially created as functional tests, including tests assessing ADL/IADLs, and instruments designed to assess an individual's need for guardianship (guardianship instruments). Further categorization includes tests and structured interviews designed for clients with dementia or without dementia (see [7], for a complete list of independent living functional assessment instruments).

Informant Interview. An interview with collateral sources including family members and/or healthcare providers about the client's history and behavior is often vital to assessing the validity of other assessment information and resources available to the client.

Integration of Findings. Results from the first three steps are consolidated into a comprehensive clinical report that addresses the referral question. Several sample reports are included in the *Handbook for Lawyers* (Appendix 2, [2]). A report may be entered as evidence.

Follow up. A repeat evaluation is helpful in evaluating the effectiveness of previously recommended interventions and supports and to assess whether there have been any changes in functioning since the previous evaluation. If repeat testing is judged to be helpful by the clinician, the duration of time until retesting is typically specified by the clinician conducting the initial evaluation.

Decision Making about Capacity

As described in the *Handbook for Lawyers* [2], lawyers may make efforts to "enhance capacity" for clients who appear to have minor or momentary difficulties understanding a particular transaction or issue. However, if a client has potentially significant difficulties with capacity and a capacity assessment has been requested, decision making by the clinician requires exceptionally careful integration of assessment information. As previously discussed, decision making about capacity is not an all-or-nothing phenomenon and, even in light of objective test findings, conclusions typically involve a weighing of information that cannot but involve some measure of subjective clinical judgment. Given these considerations, the following questions may be

helpful in guiding clinical decision making after the assessment has been completed:

1. Is there evidence of incapacity in independent living that cannot be explained by values, culture, historic decision making, etc.?
2. If there is evidence of incapacity, what is the presumed cause (e.g., medical, cognitive, or psychiatric condition; lack of reasoning ability in medically and cognitively intact individual)?
3. Is the problem temporary or permanent? Are there variables that impact capacity temporarily (e.g., medications, fatigue, history of substance use, bereavement, sensory problems)?
4. What are the specific functional domains in which incapacity or lack of capacity is demonstrated? What functional domains and skills appear to be intact? (see Table 1; see p. 67 in the *Handbook for Judges* [8])
5. Is there any immediate risk of harm? What level of supervision, if any, is needed?
6. What social supports can the individual be provided with that would allow maximum liberty and independence?
7. What are the least restrictive alternatives to guardianship? (see pp. 62–65 in *Handbook for Judges* for a detailed list of least restrictive alternatives in each functional area, developed by Joan O'Sullivan; [8]). Can the client be provided with "appropriate technological assistance" [6] to enhance capacity?

Conclusion

Given the rapidly growing population of older adults and the increasing awareness of the least restrictive alternatives to guardianship, issues regarding the assessment of capacity for independent living are likely to become more common for lawyers and clinicians. Most lawyers and clinicians share the common goal of maximizing an individual's independent functioning, liberty, and choice. In addition, clinicians are specially trained to make data-based judgments about decisional capacity, and to provide recommendations that balance safety and autonomy and maximize quality of life. An understanding of current theories, instruments, and decisional models, along with the use of medicolegal resources (e.g., the *Handbook for Lawyers* and *Handbook for Judges* [2, 8]), may lead to enhanced decision making.

References

[1] U.S. Bureau of the Census (2000). *Population Projections of the United States by age, sex, race and Hispanic origin: 1995–2050, Current Population Reports*, U.S. Bureau of the Census, pp. P25–1130.

[2] American Bar Association and American Psychological Association Assessment of Capacity in Older Adults Project Working Group (2005). *Assessment of Older Adults with Diminished Capacity: A Handbook for Lawyers*, American Bar Association and American Psychological Association.

[3] National Center for Ethics in Health Care (2002). *Ten Myths About Decision-Making Capacity*, The National Ethics Committee of the Veterans Health Administration.

[4] Marson, D. & Hebert, K.R. (2006). Functional assessment, in *Geriatric Neuropsychology: Assessment and Intervention*, D.K. Attix & K.A. Welsh-Bohmer, eds, Guilford Press, New York.

[5] Pilisuk, M., Montgomery, M.B., Parks, S.H. & Acredolo, C. (1993). Locus of control, life stress, and social networks: gender differences in the health status of the elderly, *Sex Roles: A Journal of Research* **28**, 147–166.

[6] National Conference of Commissioners on Uniform State Laws (1997). *Uniform Guardianship and Protective Proceedings Act*, http://www.law.upenn.edu/bll/ulc/fnact99/1990s/ugppa97.htm (accessed Oct 2007).

[7] Moye, J. & Braun, M. (2007). Assessment of medical consent capacity and independent living, in *Changes in Decision-Making Capacity in Older Adults: Assessment and Intervention*, S.H Qualls & M. Smyer, eds, John Wiley & Sons, New York, pp. 205–236.

[8] American Bar Association and American Psychological Association Assessment of Capacity in Older Adults Project Working Group (2006). *Judicial Determination of Capacity of Older Adults in Guardianship Proceedings: A Handbook for Judges*, American Bar Association and American Psychological Association.

[9] Park, D.C., Lautenschlager, G., Hedden, T., Davidson, N., Smith, A.D. & Smith, P. (2002). Models of visuospatial and verbal memory across the adult life span, *Psychology and Aging* **17**, 299–320.

[10] U.S. Preventive Services Task Force (1996). *Guide to Clinical Preventive Services*, 2nd Edition, U.S. Department of Health and Human Services, Washington, DC.

[11] Alzheimer's Association (2006). *Fact Sheet: Growth of Alzheimer's Disease Through 2025*; Retrieved August 2, 2007 from http://www.alz.org.

[12] Grisso, T. (2003). *Evaluating Competences*, 2nd Edition, Plenum, New York.

[13] National Conference of Commissioners on Uniform State Laws (2004). *Uniform Probate Code*. Retrieved October 26, 2007 from http://www.law.upenn.edu/bll/archives/ulc/upc/final2005.htm.

[14] Department of Veterans Affairs (1997). *Clinical Assessment for Competency Determination: A Practice Guideline For Psychologists*, Department of Veterans Affairs, National Center for Cost Containment, Washington, DC.

Further Reading

UPC http://www.law.upenn.edu/bll/archives/ulc/upc/final2005.pdf, p. 457.

MICHELLE BRAUN

Capacity for Parenting: Assessment of *see* Parenting: Assessment of Capacity

Capacity to be a Parent: Determination of *see* Parenting: Assessment of Capacity

Capacity to be Executed *see* Capacity Assessment

Capacity to be Sentenced *see* Capacity Assessment

Capacity to Consent to Medical Treatment

All adults are considered legally competent to consent to treatment unless determined otherwise by a court of law. However, it is not uncommon for the competence of older adults (over 65 years) to be questioned more often than that of younger adults. This trend is unlikely to change, as the older adult population is expected to double by 2020 and more than triple by 2050 [1]. This apparent increase in suspected incompetence is probably due to a variety of factors including aging myths and biases, an increasing prevalence of chronic diseases, and normal age-related changes in selective cognitive abilities. Among the myths about older adults are their characterization as sick and disabled, likely to live in nursing homes, no longer desirous of sexual activities, and often physically tired [2]. Age-related cognitive deficits include slowed information processing, slowed reaction time, and impaired memory [3], although interindividual differences in deficits abound. Most older adults have at least one chronic disease and many have multiple conditions [4]. This latter fact undoubtedly contributes to an increased likelihood of facing medical decisions, as does the increasing prevalence of degenerative diseases (e.g., Alzheimer's disease [5]).

The focus of this article is the assessment of capacity to consent to treatment, with particular attention to the assessment of medical decisional capacity in older adults. Consent is a communication process between a patient and his or her health-care provider, which ultimately leads to the patient's agreement to health services [6]. We use the term *capacity* rather than competency, as many states have moved away from the use of the terms *competence* and *incompetence*. The shift to capacity tends to avoid the categorical conceptualization of competence or incompetence [7], and permits a focus on specific decisional and behavioral strengths and weaknesses. Thus, the scope of decision making of individuals or their surrogates may be limited, as determined by the nature of the medical condition, and the potential consequences of decisions, that is, the risk to benefit ratio.

We begin with a brief history of the doctrine of consent to treatment, followed by discussions of the current state of consent capacity and assessment to determine capacity to make medical decisions.

History of Doctrine of Consent to Treatment

The doctrine of consent to treatment, or informed consent, dates back at least to the writings of early Greek and Byzantine authors [8]. It has its roots in several disciplines, including health, law, social sciences, and philosophy [9]. In his *Laws*, Plato notes that doctors who treated primarily free men, as opposed to slaves, requested information from patients and the patients' friends, and informed patients about their illnesses before seeking consent for treatment. Patients were differentiated on the basis of their autonomy, that is, slaves versus free men. Consent was obtained only from free men, as slaves were not afforded the right to self determination.

The history of consent to treatment in the United States dates back to *Slater v. Baker and Stapleton* [10] in 1767 [11]. In *Slater*, the court stated that "It is reasonable that a patient should be told what is about to be done to him, that he may take courage and put himself in such a situation as to enable him to undergo the operation" [12]. However, informed consent was not extensively discussed in the American medical literature until the late 1950s and early 1960s [9]. The expression "informed consent" was coined in 1957 case law [13]. Martin Salgo, who became paralyzed following treatment, sued his physicians for failing to disclose risks and alternate treatments. Until the late 1950s, "the justification of practices of disclosure and consent seeking were strictly governed by what we shall call a beneficence model rather than an autonomy model of the physician's responsibility for the patient" [9]. The beneficence model emphasizes that behavior of the physician should be in the best interest of the patient, whereas the autonomy model emphasizes the physician's responsibility to respect the autonomy of the patient through information disclosure and the seeking of consent. The Salgo case involved the clashing of these two models.

The doctrine of informed consent "originally developed and flourished under the battery theory of liability" [9]. Battery is intentional touching that is unwanted or legally unpermitted, and which is actionable even if no physical injury results. Individuals have the right to choose whether they are to be touched or treated. The 1914 case of *Schloendorff v. Society of New York Hospitals* [14] nicely illustrates this position. Justice Cardozo argued that "Every human being of adult years and sound mind has a right to determine what shall be done with his own body; and a surgeon who performs an operation without his patient's consent commits an assault" [9].

Consent Capacity Today

The prefatory note of the Uniform Health-Care Decisions Act of 1993 [15] states that "... the Act acknowledges the right of a competent individual to decide all aspects of his or her own health care in all circumstances, including the right to decline health care or to direct that health care be discontinued, even if death ensues. An individual's instructions may extend to any and all health-care decisions that might arise and, unless limited by the principal, an agent has authority to make all health-care decisions which the individual could have made. The Act recognizes and validates an individual's authority to define the scope of an instruction or agency as broadly or as narrowly as the individual chooses".

Consent capacity, or health-care decision-making capacity, is defined by state statute. In most states the definition can be found within the advance directives laws [16]. The definitions tend to be similar to that of the Uniform Health-Care Decisions Act of 1993 [16], which defines capacity as "an individual's ability to understand the significant benefits, risks, and alternatives to proposed health care and to make and communicate a health-care decision".

In the 1970s, Meisel, Roth, and Lidz [12] examined relevant scholarly and judicial works on consent to treatment and identified three elements of a valid health-care decision: provision of information (knowledge), competency, and understanding. One precondition (voluntariness) and one consequence (consent or refusal) were also specified. Voluntariness requires that the treatment decision "be free from coercion and from unfair persuasions and inducements" [12]. The authors also argued that there must be social support (e.g., from caregivers) for the individual's right to self determination. The provision of information, the knowledge component, has evolved in recent years. The initial standard came out of the reasonable practitioner clause in *Natanson v. Kline* [17]. Prior to this case, there was no formal requirement of informed consent. Following *Canterbury v. Spence* in 1972 [18], this standard has been replaced in many states by the materiality of information standard, whereby informed consent requires information that an ordinary person would want to know,

thereby shifting the focus from the practitioner to the patient. This information required includes risks, discomforts, side effects of the proposed treatment, alternative treatments (including risks, discomforts, and side effects), and the consequences of not receiving the treatment.

There are four conditions under which nondisclosure or an abbreviated disclosure is permitted. First, if there is an emergency and time does not permit an adequate disclosure, or consent cannot be obtained without jeopardizing the health of the patient. Second, nondisclosure is acceptable if the disclosure of information would disturb the patient to the point the patient is unable to make a rational decision, or if such disclosure poses a serious physical or psychological threat. In this case, the practitioner can withhold the information and invoke "therapeutic privilege". Therapeutic privilege permits the practitioner to be exempt from informed-consent requirements. Third, consent is not required if the patient is incompetent. In this case, a surrogate decision maker would be incorporated into the decision-making process. Fourth, there can be a waiver of medical privilege by the patient. The patient can, for example, say to a physician that he or she trusts the physician to do the right thing and requests that the physician proceed.

Assessment of Medical Decision-Making Capacity

At the present time, there is no gold standard for the assessment of the capacity to consent to or refuse treatment, that is, to make a medical decision. Assessment should address the behavioral, cognitive, and functional abilities of the individual [11, 16]. Assessment has typically involved clinical interviews, reviews of medical records, and often the administration of a mental status examination [19], which can fall short of contemporary recommendations [11, 16, 20]. A recent study of guardianship case files across three states revealed frequent poor documentation of functional abilities when capacity was assessed, and evidence that some older adults are having guardians appointed on the basis of assessment documentation of only a few sentences [21]. The approach to assessment is gradually changing, as functional assessment instruments and interview guides have been developed for the assessment of capacity over the past decade. The instruments constructed for the assessment of functional abilities are intended to be used in conjunction with information obtained through other procedures (e.g., diagnostic interviews, tests of cognitive abilities) [20]. The functional assessment instruments and the associated research were recently reviewed by Moye and colleagues [20] with an eye to clinical implications and research needs. It is important to understand that these are all first-generation instruments, which can undoubtedly be improved through further development of their reliability and validity as well as through the incorporation of additional important features (e.g., values assessment).

Although guidelines and recommendations have occasionally been published to guide the assessment of medical decision-making capacity (e.g., [11, 22]), there remain differences of opinion and confusion regarding how decisional capacity should be operationalized and assessed by clinicians [20, 23]. On a more positive note, recommendations for the interpretation of the documentation of such evaluations have recently been published to guide lawyers [16] and judges [24]. Nevertheless, approaches to assessment remain varied.

Conceptual Model

Grisso [25, 26] has offered an excellent and widely cited conceptual model to guide the assessment process. It enables one to integrate clinical practice standards, law, and clinical research [16]. The most recent version of the model [26] comprises five components that define legal competencies: functional, causal, interactive, judgmental, and dispositional components.

Functional Component. The functional (cognitive and behavioral) component consists of functional abilities that enable one to accomplish specific decisional tasks. These abilities also encompass the knowledge, understanding, and beliefs that are needed in order to complete a task or an action. Grisso [25] argues that functional abilities are central to the determination of a patient's competence. Roth, Meisel, and Lisz [27] identified several tests or elements of competency that comprise the functional abilities. These were later refined by Grisso and Appelbaum [28]. They include the expression of a choice, understanding, appreciation, and reasoning. Demonstration of the ability to express or communicate a choice and maintain it with reasonable stability

over time is the simplest test [28]. The demonstration of the remaining three abilities would be moot if the choice cannot be expressed by the patient.

The demonstration of the ability to understand information relevant to treatment decision making, the second test of functional abilities, requires comprehension of the important information about one's medical condition, treatment options, and potential outcomes.

Demonstration of one's ability to appreciate the relevance of information to one's own situation is the third test of functional abilities. This includes the appreciation of the probable consequences of each treatment option for oneself. This complicated ability is difficult to assess, and oftentimes courts have found patients to be incompetent if they appear to understand information about their condition, but have not appreciated the relevance of the information to their own situation [28] For example, a patient may understand that his doctors believe that he is ill, but the patient may deny being ill despite objective evidence suggesting otherwise. This would suggest a lack of appreciation for the information.

Demonstration of the ability to reason with relevant information is the fourth test of functional abilities. This is typically conceptualized as the ability to rationally manipulate information [28], which would include treatment related information regarding risks and benefits, and one's values and preferences.

As one might expect, each of the abilities discussed above can be influenced or compromised by deficits in cognitive abilities. This fact is particularly important for consideration when one is assessing these functional abilities with older adults. Several cognitive abilities undergo age-related changes, including, for example, memory and attention [3]. In addition, information tends to be processed more slowly as we age [3]. Such deficits can impair demonstrations of capacity, particularly with regard to understanding, reasoning, and appreciation. Not only should one take these factors into consideration, but one should also be prepared to augment the performance of older adults by reducing the cognitive demands of the assessment process. For example, information can be presented in written, in addition to oral, form to reduce the demand on attention and auditory memory. Information can be presented at a somewhat slower speed to accommodate the slower information processing of older adults. Information can be broken into small units to reduce the memory load. Other accommodations can be made to address age-related visual and auditory sensory deficits. For example, information can be printed in a large font on nonglare paper. Extraneous background noise can be minimized. Lighting can be bright without glaring.

Causal Component. The causal component is the clinical or disabling condition that is likely the cause of incapacity. The effects can be transient (e.g., delirium) or long term (e.g., Alzheimer's disease). An important question is whether the problem will persist, worsen, or improve.

Interactive Component. The interactive component takes into account the specific demands of the treatment situation and how it is relevant to or interacts with the individual's functional deficits, that is, person–context interactions. In different contexts or environments, certain functional abilities may be needed for an individual to be competent to make a medical decision. It is possible that any incongruence between a patient's abilities and the demands of the situation could be addressed by either teaching the patient new decisional skills if possible or reducing the demands in the patient's life situation.

Judgmental and Dispositional Components. A judgment must be made regarding the congruency between the patient's abilities and the situation demands and supports. The judgment could be conceptualized as a matter of template matching, where the template comprises the cognitive and behavioral demands of a medical decision-making task and the decision-making context, and a match is sought between the template and the cognitive and behavioral abilities of the individual, plus any environmental supports (i.e., physical, social). To the extent that there is a mismatch, there may be diminished decision-making capacity. Lastly, it is possible that any incongruence between a patient's abilities and the demands of the situation could be addressed by either teaching the patient new decisional skills if possible or reducing the demands in a person's life situation [26]. The dispositional component constitutes the consequences of the judgment as prescribed by law [26].

This model takes into account many different aspects of a competency decision and highlights the important aspects to assess. The model helps to give shape to the construct of legal competency by structuring the important components.

Assessment Guidelines and Instruments

Two good guides for assessing medical decision-making capacity are Grisso and Appelbaum's [28] guide for physicians and other health professionals, and *Assessment of Competency and Capacity of the Older Adult: A Practice Guideline for Psychologists* published by the Veterans Administration [22]. The Grisso and Appelbaum book includes a very thorough and thoughtful discussion of relevant concepts and issues, and also contains an array of questions that can be used to ascertain medical decision-making capacity, and a copy of the MacArthur Competence and Assessment Tool-Treatment (MacCAT-T) manual. The Veterans Administration guide focuses on older adults, whereas the Grisso and Appelbaum book and the MacArthur assessment instrument are relevant for adults of all ages.

The Veterans Administration Guideline also remains a very reasonable guide for mental health practitioners who are preparing for the assessment of an individual's medical decision-making capacity. The Guidelines describe five key steps in the assessment of competency (capacity): The first step includes review of the consultation request and clarification regarding the decisional capacity in question. The second step involves consideration of informed consent and an ethical, appropriate, and valid assessment. Step three is the assessment process, which may include interviews with the patient, family, and health-care informants to gain an appreciation of the patient's values. This is followed by performance-based assessment of cognitive functioning, assessment of mental health factors, and performance-based assessment of specific decisional capacities. The fourth step includes a synthesis of data from the interviews and assessment, and a report of the findings. The results are discussed with the patient and relevant family members. Finally, the reader is encouraged to perform a follow-up evaluation that considers the impact of recommended interventions and assessment of changes in functioning. These guidelines include a very helpful algorithm for assessing decisional capacity. The questions of voluntariness and knowledge, though not formally addressed, also must be determined.

A variety of performance-based assessment instruments have been developed to aid in the assessment of medical decision-making capacity. These are not designed as stand-alone "capacimeters", but rather as one source of information that is to be used in a comprehensive assessment of capacity, which would also include cognitive and diagnostic assessment. A review of those instruments is beyond the scope of this article. The interested reader is referred to Grisso [28] and Moye *et al.* [20] for reviews of relevant forensic instruments. A listing of these and other related instruments also can be found in a handbook created for lawyers by the American Bar Association Commission on Law and Aging and the American Psychological Association [16]. Following their review of capacity assessment instruments, Moye *et al.* recommended the MacCAT-T [28] and other Grisso and Appelbaum instruments for psychiatrically disturbed populations. The Capacity to Consent to Treatment Instrument [19] is recommended for demented populations. Finally, the Hopemont Capacity Assessment Instrument [29, 30] is recommended for impaired adults in long-term-care settings.

The practice of assessment of capacity to consent to treatment has become more sophisticated and arguably more valid in recent years with the introduction of assessment instruments created explicitly to aid in the assessment of capacity. Nevertheless, judgment of practitioners remains the gold standard for capacity determination, and that judgment is moderately reliable at best. Available guides and guidelines can facilitate more thorough, and hopefully more reliable, assessments. The further development and refinement of existing assessment instruments will undoubtedly improve the psychometric properties of this first generation of instruments and enable us to assess capacity to consent with greater reliability, validity, and confidence.

References

[1] U. S. Census Bureau (2004). *Global Population Profile: 2002. International Population Reports*, U.S. Agency for International Development.

[2] Palmore, E. (1999). *Ageism: Negative and positive*, Springer, New York.

[3] Park, D.C. (2000). The basic mechanisms accounting for age-related decline in cognitive function, in *Cognitive Aging: A Primer*, D.C. Park & N. Schwarz, eds, Taylor & Francis, Philadelphia, pp. 3–21.

[4] Federal Interagency Forum on Aging-Related Statistics (2006). *Older Americans Update 2006: Key Indicators of Well-Being*, Retrieved from http://agingstats.gov/agingstatsdotnet/main_site/default.aspx (accessed on, 2007).

[5] Alzheimer's Association (2007). *Alzheimer's Disease Facts and Figures 2007*, Retrieved from Alzheimer's

Association. http://www.alz.org/national/documents/Report_2007FactsAndFigures.pdf (accessed on, 2007).
[6] Rosovsky, F.A. (1999). *Cumulative Supplement: Consent to Treatment*, 2nd Edition, Aspen Publishers, Gaithersburgh.
[7] Sabatino, C. (1996). Competency: refining our legal fictions, in *Older Adults' Decision Making and the Law*, M. Smyer, K.W. Schaie & M.B. Kapp, eds, Springer, New York, pp. 1–28.
[8] Dalla-Vorgia, P., Lascaratos, J., Skiadas, P. & Garanis-Papadatos, T. (2001). Is consent in medicine a concept only of modern times? *Journal of Medical Ethics* **27**, 59–61.
[9] Faden, R.R. & Beauchamp, T.L. (1986). *A History and Theory of Informed Consent*, Oxford University Press, New York.
[10] Slater v. Baker & Stapleton, 95 Eng. Rep. 860 (KB. 1767).
[11] Grisso, T. & Appelbaum, P.S. (1998). *Assessing competence to Consent to Treatment*, Oxford University Press, New York.
[12] Meisel, A., Roth, L.H. & Lidz, C.W. (1977). Toward a model of the legal doctrine of informed consent, *American Journal of Psychiatry* **134**, 285–289.
[13] Salgo v. Leland Stanford Jr. Univ. Bd. Of Trustees, 317 P.2d 170 (Cal. Ct. App. 1957).
[14] Schloendorff v. Society of New York Hospital, 211 N.Y. 125, 105 N.E. 92 (1914).
[15] Uniform Health-Care Decisions Act (1993). Retrieved from http://www.law.upenn.edu/bll/archives/ulc/fnact99/1990s/uhcda93.pef (accessed on, 2007).
[16] American Bar Association Commission on Law and Aging & American Psychological Association (2005). *Assessment of Older Adults with Diminished Capacity: A Handbook for Lawyers*, American Bar Association and American Psychological Association, Washington, DC.
[17] Natanson v. Kline, 350 P.2d 1093 (Kan. 1960).
[18] Canterbury v. Spence, 464 F.2d 772 (D.C. Cir., 1972).
[19] Marson, D., Ingram, K., Cody, H. & Harrell, L. (1995). Assessing the competency of patients with Alzheimer's disease under different legal standards, *Archives of Neurology* **52**, 949–954.
[20] Moye, J., Gurrera, R., Karel, M., Edelstein, B. & O'Connell, C. (2006). Empirical advances in the assessment of the capacity to consent to medical treatment: clinical implications and research needs, *Clinical Psychology Review* **26**, 1054–1077.
[21] Moye, J., Wood, E., Edelstein, B., Wood, S., Bower, E., Harrison, J. & Armesto, J. (2007). Statutory reform is associated with improved court practice: results of a tri-state comparison, *Behavioral Sciences and the Law* **25**, 425–436.
[22] Department of Veterans Affairs (1997). *Clinical Assessment for Competency Determination: A Practice Guideline For Psychologists*, Department of Veterans Affairs, National Center for Cost Containment, Washington, DC.
[23] Ganzini, L., Volicere, L., Nelson, W. & Derse, A. (2003). Pitfalls in the assessment of decision-making capacity, *Psychosomatics* **44**, 237–243.
[24] American Bar Association Commission on Law and Aging, American Psychological Association, & National College of Probate Judges (2006). *Judicial Determination of Capacity of Older Adults in Guardianship Proceedings: A Handbook for Judges*, American Bar Association and American Psychological Association, Washington, DC.
[25] Grisso, T. (1986). *Evaluating Competencies: Forensic Assessments and Instruments*, Plenum, New York.
[26] Grisso, T. (2003). *Evaluating Competencies: Forensic Assessments and Instruments*, 2nd Edition, Kluwer Academic, New York.
[27] Roth, L.H., Meisel, C.A. & Lidz, C.A. (1997). Tests of competency to consent to treatment, *Canadian Journal of Psychiatry* **134**, 279–284.
[28] Grisso, T. & Appelbaum, P.S. (1998). *MacArthur Competence Assessment Tool for Treatment (MacCAT-T)*, Professional Resource Press, Sarasota.
[29] Edelstein, B., Nygren, M., Northrop, L., Staats, N. & Pool, D. (1993). Assessment of capacity to make financial and medical decisions, *Paper Presented at Meeting of the American Psychological Association*, Toronto.
[30] Edelstein, B. (1999). *Hopemont Capacity Assessment Interview Manual and Scoring Guide*, Morgantown, West Virginia University.

BARRY A. EDELSTEIN AND CHRISTINE E. GOULD

Capacity to Consent to Research

see Capacity Assessment

Capacity to Enter into a Contract

see Capacity Assessment

Capacity to Make a Will ***see***

Capacity Assessment

Capacity to Stand Trial

For a person living with a mental illness who has been charged with a criminal offense, the road ahead can be quite lengthy and complicated. Courts refer only a fraction of cases for an evaluation to assess a person's ability to proceed to trial, which is not surprising, given the presumption that criminal defendants are competent to stand trial [1]. If unable to participate meaningfully in the proceedings, a person is deprived of due process [2]. While approximately 95% of felony cases in the United States are resolved by a plea bargain agreement, there is currently no data that would tell us how many potentially incompetent people negotiate their way through a plea bargain agreement. As between 15 and 20% of prison and jail populations consist of individuals living with a serious mental illness, depending upon how mental illness is defined, one wonders how many are being processed or evaluated properly.

What is Competency to Stand Trial?

In the United States, the legal standard for competency to stand trial varies by state but cannot fall below federal standards as articulated in *Dusky v. United States* [3]. Dusky incorporates a two-prong test of whether a defendant has "sufficient present ability to consult with his lawyer with a reasonable degree of rational understanding and whether he has a rational as well as factual understanding of the proceeding against him" (p. 781). Ultimately, the trier of fact (usually a judge) is relied upon to interpret what capacities are sufficient to meet the legal threshold of trial competency. Competency is specific to a defendant's capacities at the time of trial, which is different from a defendant's mental condition at the time of an offense (e.g., insanity).

It is important to note that Dusky does not demand or specify any particular type of disorder, only that the person lacks the capacity to proceed with his case. The reasons a defendant is typically found to be legally incompetent to stand trial are because of a mental disorder, mental retardation, or significant medical conditions (such as brain injuries resulting in cognitive damage, delirium caused by intoxication, or an acute medical condition) that would directly impact his ability to understand the proceedings or assist in his defense. In some cases, the condition is permanent and untreatable, such as in cases of moderate or severe mental retardation or irreversible cognitive damage caused by a traumatic brain injury. In other cases, treatment of psychiatric or medical illnesses and/or education may be utilized to successfully restore a defendant's competency to proceed.

There is no gold standard that defines what makes a defendant competent. Ultimately, the trier of fact will determine which factors are relevant to trial competency, given the requirements referenced by Dusky [3] are ill-defined. For example, expectations of having the capacity to testify relevantly when a defendant intends to enter a guilty plea or accept a plea bargain offer may not be as significant in some courtrooms, while in other jurisdictions, the standard is held regardless of a defendant's legal strategy. In *Godinez v. Moran*, the US Supreme Court held that the competency standard for pleading guilty or waiving the right to counsel was the same as for standing trial [4]. Importantly, the court found that actual decisions must also be made "knowingly" and "voluntarily" [4]. The distinction between capacity and actuality is important to bear in mind. While the focus of our inquiry into a defendant's competency to stand trial will assesses his mental "capacity" or ability to make decisions about his case, it is the court's role to determine that the defendant's actual decision is made knowingly and voluntarily. The mental health consultant may ask how a defendant has made general decisions in the past to inform his opinion about the defendant's current capacities. The court must determine, generally by conducting a colloquy (or entering into a discussion), whether a defendant does "actually" understand the significance and consequences of his decision specific to his case and whether or not that decision is uncoerced.

The Legal Process of Referral, Evaluation, and Restoration of Competency to Stand Trial

The process of referring a person for an evaluation to assess competency can occur at any time prior to adjudication. The arresting officer(s), jail staff, prosecution or defense counsel, or even a family member

might initially raise concerns about a defendant's capacity to proceed, although the evaluation process will not become formal until a judicial order is signed authorizing the examination procedure. Ethical standards insist that the defendant have an attorney assigned to him. When we receive referrals where there is no identified defense attorney, we will contact the court to request an attorney be assigned before we begin our competency evaluation.

Many jurisdictions employ outpatient clinics where an evaluee might be transported from the correctional facility to a clinic for evaluation. There are fewer jurisdictions where evaluations are conducted at a hospital (e.g., a state psychiatric hospital in the United States). Many jurisdictions employ in-custody evaluations wherein the evaluator conducts the examination where the defendant is detained, whether that is in a correctional setting or a civil commitment facility. The time allotted for an evaluation to be completed varies by jurisdiction. Evaluations can also occur out of custody for defendants who are not incarcerated or who have been released on their personal recognizance.

A defendant's case is processed through the legal system based upon the category of his legal charge. For those charged with a misdemeanor or more minor offense, the potential sentencing range is generally shorter than a felony or more serious offense, which can dictate the timelines for court hearings. Persons charged with felonies may be treated differently than those charged with misdemeanors, and we encourage evaluators to clarify the statutory guidelines within their jurisdiction regarding these potential differences.

Statutory permission for time allotted to restore a defendant's competency to stand trial also varies between jurisdictions and may depend upon the crime charged or other factors (e.g., violence history). In some jurisdictions, when a person is charged with a misdemeanor crime and is found to be incompetent, the charge may be dismissed and the defendant may be released or sometimes referred for civil commitment (*see* **Civil Commitment**). Other jurisdictions may have guidelines for misdemeanant competency restoration treatment, although often for a shorter duration than for felonies. A person charged with a felony may be committed for competency restoration, again based upon the jurisdiction's guidelines, for several weeks or months. However, the indefinite commitment of a criminal defendant, because of continued legal incompetency, would violate his due process [5]. Therefore, if a felony defendant's competency to stand trial is not restored after a reasonable amount of time (pursuant to guidelines of his jurisdiction), the charges may be dismissed and he may (or may not) be referred for civil commitment. Alternatively, if competency restoration treatment is considered to be successful, the defendant will proceed to court to address the charges as any other defendant.

In *Sell v. United States* [6], the US Supreme Court addressed whether or not psychotropic medications should be administered involuntarily to treat mental symptoms so a defendant can proceed to trial (*see* **Treatment, Right to: Mental Health**). The Supreme Court identified four findings that must be satisfied prior to the authorization of involuntary treatment with psychotropic medications: (i) that a court must find important governmental interests are at stake (e.g., prosecution of the crime); (ii) that involuntary medications will significantly further those concomitant state interests because administration of the drugs is substantially likely to render the defendant competent to stand trial and administration of the drugs is substantially unlikely to have side effects that will interfere significantly with the defendant's ability to assist counsel in conducting a trial defense; (iii) any alternative, less intrusive treatments are unlikely to achieve substantially the same results; and (iv) administration of the drugs is medically appropriate, i.e., in the patient's best medical interest, in light of his medical condition [6] (p. 12–14). If these conditions are met, the court can exert judicial authorization for involuntary administration of psychotropic medications. If these conditions are not met, and the defendant is nonetheless referred for competency restoration treatment, the treatment may be limited by a defendant's refusal to accept psychotropic medications, and competency may not be restored. However, following the dicta of Sell, some jurisdictions employ either civil commitment (when the defendant presents a danger to himself or others or is gravely disabled) or a policy-driven medication override process (cf. *Washington v. Harper* [7]) as an alternative mechanism for involuntary administration of medications. The purpose of administering these medications would not be for the purposes of competency restoration treatment but for medical necessity, such as when a person presents a danger to himself or others.

The Role of the Mental Health Consultant in Evaluations of Competency to Stand Trial

Competency to stand trial is a legal standard, not a medical or mental health concept. Mental health professionals who consult with the court about a defendant's competency are in the position of needing to explain how mental health issues factor in to a defendant's ability to proceed with his case. The issue of whether or not a person is competent to stand trial cannot be answered simply or dichotomously. For example, a head-injured defendant might have adequate knowledge about court proceedings, be able to cooperate with his attorney, and have an interest in resolving his case expediently. However, he also might be unable to retain information across meetings with his attorney or demonstrate impulse control problems that would substantially interfere with his capacity to assist in his defense or appear in court appropriately, despite his interest in doing so. Similarly, a person with a psychotic disorder may have no other discernable impairment that would preclude him from comprehending his legal situation beyond a deeply ingrained delusional thought that there is a larger significance to his incarceration that will only be revealed to him after he enters a guilty plea. As the ultimate decision is legal, mental health professionals are faced with the challenge of explaining or educating the court about a defendant's mental symptoms that impact his capacity to understand, participate, and make decisions in his case.

As mental health consultants, we are skilled in establishing a person's cognitive, emotional, and interpersonal dynamics. We have been trained to assess a defendant's general overall capacities, such as his level of cognitive and intellectual functioning, his level of rationality or decision-making capabilities, his personality constructs, and how he interacts with others. We are also trained to recognize potential psychiatric symptoms or belief systems that could interfere with a defendant's ability to attend to his basic daily needs. Our role as court consultants is therefore to explain a defendant's capacities in relation to the relevant legal concepts. We are not asked to define legal terms, as this is reserved for the court and attorneys. We do need to be informed of the legal question, as this will guide our evaluation to focus on the relevant legal requirements. It is not up to the mental health consultant to translate our understanding of a defendant's capacities into legal terminology, as this determination is reserved for the judge or jury. Our role is to supply the court with information about the sorts of things a defendant can do that will demonstrate the kinds of abilities that are relevant to the legal question.

A mental health consultant will address three primary capacities that are relevant to a defendant's competency to stand trial: general capacity, decisional capacity, and particular capacities that are unique to his case. We can examine a defendant's general cognitive capacity in terms of his basic level of knowledge and about court-related matters. If a defendant presents with symptoms that would interfere with his general capacities, there is little reason to go further. For example, active auditory hallucinations might prevent someone from utilizing his intrinsic ability to attend to a court hearing or listen to his attorney, apart from his otherwise intact cognitive abilities. It is helpful to understand a defendant's baseline level of psychiatric and intellectual functioning, attention, memory, degree of rationality, and his ability to work with or interact with others to provide the context for competency-related capacities. Within this context, we are in a better position to inform the court whether a defendant's inability to trust his attorney is based upon a pessimistic view of the legal system or a paranoid delusional perception. We are also able to assist in establishing whether or not a presented claim of psychosis is legitimate and how it may or may not bear upon a defendant's case. A link must be established between a potential symptom or condition and a trial-relevant incapacity. It is not enough to say that a person with a cognitive or psychotic disorder is incompetent without relating how his condition affects his capacities.

We also examine a defendant's capacity to make decisions and whether he can apply these capacities to inform a legal decision. The manner in which people make decisions is often not based upon rational or syllogistic methods. Rather, decisions are more frequently made from emotional, reactive, automatic, or sometimes delusional states, particularly when a person is under duress or facing serious criminal charges for the first time. Consider how one may have decided on what to order the last time he/she went to a restaurant. They might have based their decision upon trying to stick to their diet or breaking their diet because they were tempted by the daily special. Understanding the process of what goes into

making general or everyday decisions will inform us of how a person will approach more critical decisions. When evaluating a defendant's decisional capacity, we need to consider how he will make decisions, in general, which can then be applied to the specific decisions he is facing in his legal case. Is the defendant capable of weighing risks and benefits of his decision? Is he able to project himself into the future after having made the decision to inform his choice? Is his factual knowledge of the case intact? Is his decision based upon a rational understanding of his available legal options? Is his decision stable over time? It is important to recognize that competency to stand trial is based upon a "capacity" to make reasonable decisions, not whether the decisions were reasonable. People are free to make unreasonable decisions, although they must have the ability to do otherwise. For example, a defendant might wish to take his case to trial despite considerable evidence against him that is certain to result in a guilty finding because he wants to have his day in court. Another defendant might decide to accept a plea bargain offer to avoid trial for emotional reasons or so he can get out of jail sooner, even when he has salient information that might hold up well at trial.

Finally, as mental health consultants, we examine a defendant's capacities as relevant to his specific case. Each case has its own unique particularities that may interact with a defendant's situation or mental condition. An individual living with a mental disorder may have a well-encapsulated psychotic delusion that does not bear upon his specific legal case and be otherwise quite capable of resolving his case as long as the case does not overlap with the delusion. For example, a defendant may harbor grandiose delusional ideation in regards to his own professional affiliations, such as believing that he has acquired a number of doctoral degrees in several different fields, and be charged with stealing food, which he does not dispute. In negotiating his way through his court hearing, the issue of his delusion about his numerous doctoral degrees may never come up, because it is not relevant to the specifics of his case. Alternatively, it may interfere with his competency to stand trial if he believes that his professional affiliations entitle him to take food without needing to pay for it. We have also seen cases where an individual may have a delusional notion about a specific entity, perhaps an electric or telephone company, which prompted him to commit arson or some other crime of property damage. He may not demonstrate any other aspects of psychosis, and once the act was committed, he no longer feels at risk. However, when faced with the legal consequences, he insists that he had no alternative but to act in the manner he did. This individual, who may meet the legal criteria for insanity at the time of the crime, might reject a mental defense because he does not believe that he was demonstrating exculpable symptomatology at all and rather insists that his attorney prove his delusional reasons. The factors a defendant weighs when considering various plea alternatives, such as whether or not to waive an insanity defense, must not only be well informed, but free of delusional ideation. A defendant may present with otherwise intact general capacities in regards to his understanding of his charges and trial process and can engage appropriately with his attorney, but for a particular element that is specific to his case that calls his decisional capacities into question. A defendant for whom an insanity defense would be a viable course of action may not wish to proceed in this manner, given the potential for a lengthier commitment to a mental institution than he would have received if imprisoned. The potential influence of existent mental symptoms in formulating this decision would need to be carefully considered, particularly if the defendant lacks insight into his mental illness or expects his attorney to present his delusional beliefs about the case. This is precisely why evaluations of a defendant's capacity to proceed must go deeper than simply reviewing a defendant's familiarity of legal terminology, his charges and the potential consequences he is facing.

Evaluation of Competency to Stand Trial

Evaluating an individual's competency to stand trial can be fairly straightforward – such as in the case of a floridly psychotic individual – or extremely complex and subtle, when considering a defendant's particular decisional capacities that are relevant to his case. A review of a defendant's knowledge about legal concepts specific to standing trial is generally included in an evaluation of his capacities, but what actually occurs in the courtroom may differ. For example, the sorts of questions a judge may ask during a colloquy may go beyond questions posed in the evaluation, such as a defendant's understanding of a specific plea agreement or conditions of release

that were not articulated at the time the evaluation was conducted.

Elements to consider when assessing a defendant's knowledge or understanding of his case include the nature and seriousness of the charge, the outcome he could face if convicted including the type of punishment whether incarceration (e.g., prison or jail) or out of custody (e.g., fines or probationary requirements), the length of sentence (commitment or community probationary requirements), the impact of a conviction (e.g., restrictions imposed by felony convictions, confinement, or permanent criminal history) or the impact of a finding of legal insanity (e.g., hospitalization, conditional release conditions, or possible incarceration in some jurisdictions following a finding of guilty but mentally ill). A defendant should be familiar with his available legal defenses and the ramifications of those choices (e.g., guilty, not guilty, no contest or Alford plea, not guilty by reason of insanity, plea bargain), as well as what evidence exists against him and what witnesses might testify to. A defendant should be familiar with the roles and function of courtroom participants, such as his attorney, the prosecutor, the judge, jury, and witnesses; and the adversarial nature of a trial, particularly the opportunity to testify and the potential risk involved in cross-examination. Most of these elements speak to the first two levels of evaluating a defendant's general competency-related capacities.

However, competency to stand trial involves more than simple knowledge of legal concepts. In working with defense counsel, a defendant's ability or capacity to cooperate, supply relevant information, manage disagreements, participate in trial strategy to the degree possible, and to understand the concept of privilege are important to consider, particularly within the context of a defendant's general capacities as outlined above. A defendant's capacity to then apply these concepts to his specific case, weigh his available legal options and communicate his choices to his attorney is what will inform a judicial determination about a defendant's competency to proceed.

An evaluation of a defendant's competency to stand trial generally includes but is not limited to a review of court documents (court order requesting the competency evaluation) and discovery information (charging documents, probable cause statement, police reports, and witness statements), interview(s) with the defendant, consultation with collateral sources (family members, medical, and mental health providers) and sometimes psychological assessments that may focus on cognitive abilities, psychiatric symptomatology or specific competency-related capacities.

The opportunity for a neutral party familiar with mental health issues to be available to observe the initial interactions of a defendant in court would greatly inform a referral for a competency evaluation. Ideally, evaluations of competency to stand trial will have defense counsel present or at least available for consultation, so that potential concerns can be raised, and the relationship between counsel and client can be observed and assessed. A defendant has a legal right to have counsel present during an evaluation; however, if the attorney cannot be available or if a defendant elects to proceed on a *pro se* basis (e.g., represent himself), this option may not be possible.

Sometimes interviews are conducted in jails under conditions a defendant might feel humiliated or offended by (e.g., through the pass-through of a jail cell door or while physically restrained) that would establish a less than neutral context for an attorney or an evaluator. Every effort should be taken to treat a defendant with respect and courtesy, given this can often set the stage for the level of cooperation and disclosure during the interview, and even have a bearing on how a defendant will present in court or impact his relationship with defense counsel. During a competency evaluation, the limitations of confidentiality should be carefully explained so that a defendant can decide whether or not to participate and the advantages or disadvantages of not cooperating with an evaluation.

It is important to consider cultural differences and level of experience with the criminal justice system in the jurisdiction a defendant is involved in. Simple lack of exposure, culturally based traditions or actions accepted by one culture that differ from the jurisdiction where the defendant is criminally charged, would not necessarily support a finding of legal incompetency if the defendant has the capacity to learn the rules of the legal system or the expectations of the society he is involved with. This may require the assistance of a court-certified interpreter or a consultation with a cultural expert to establish that an initial lack of understanding does not preclude the ultimate ability to participate.

Most often, a report is generated consolidating all of the information and supplied to the court and others as directed by the court order (e.g., defense,

prosecutor, jail mental health staff, or others), unless the evaluation is retained by defense and protected under attorney–client privilege. Reports that are detailed, objective, and include collateral information that supports the opinion posited, are generally sufficient to stand alone. However, if the issues are less clear, expert testimony may also be required to allow parties to clarify the proffered opinion or to gather additional information.

Report Writing and Testimony

The conclusions following an evaluation of a defendant's capacity to stand trial are generally communicated to the court in a report (*see also* **Report Writing for Courts**) and/or testimony. Some jurisdictions require that the evaluation includes opinions regarding a defendant's current mental condition in addition to the opinion regarding his capacity to stand trial. Reports generally include the referral question, the legal charges pending, an opinion regarding the defendant's mental condition at the time of the evaluation, and the defendant's current capacities in relation to trial. A more thorough analysis will also include developmental/biographical, clinical, medical, and criminal history information supplied by the defendant and from a variety of collateral records including medical, correctional, mental health, sometimes academic facilities, and consultations with family members, medical and mental health treatment providers, community corrections officers, defense counsel, and other sources that are familiar with the defendant. The more comprehensive the report is, the less likelihood that a mental health consultant will be called upon to testify to clarify his opinions. However, in cases of opposing expert opinions or when a defendant's mental status may be different in court from the time the examination was conducted, court testimony may be required to clarify or elaborate on the proffered opinion. When testifying, a consultant is generally asked about his professional qualifications to establish his expertise in the field, the process he utilized to evaluate the defendant including the sources of information that substantiated his opinion, and his opinion regarding the defendant's trial-relevant capacities. It should be emphasized that it is the court's responsibility to make the ultimate opinion regarding a defendant's competency to stand trial, and as mental health consultants we are responsible for describing a defendant's trial capacities.

Recommendations to the Court

Mental health professionals aid the court in understanding how a defendant's mental symptoms could impact his understanding of his case, but also dispel some potential misinformation about mental illness. A defendant is not incompetent to stand trial simply because he is diagnosed with schizophrenia or bipolar disorder. Mental health consultants serve to explain to the court how the potential symptomatology could specifically impact a defendant's trial capacity. For example, a defendant diagnosed with schizophrenia exhibiting disorganized speech would have difficulty communicating effectively with counsel. A defendant, who is in a manic episode with pressured, uninterruptible speech, may have similar difficulty communicating with counsel or testifying relevantly, whereas another defendant with schizophrenia, whose symptoms are less severe or are well controlled with psychotropic medications, may still carry the diagnosis of schizophrenia but be able to participate in a rational manner.

An often overlooked aspect of a defendant's competency to stand trial is the stability of his mental status across time and situations, which is why the ultimate decision about legal competency is reserved for the trier of fact. Differences between settings of where a defendant might consult with counsel or be evaluated for his legal competency (e.g., a quiet or less-distracting atmosphere) and a courtroom (particularly a presiding court where numerous cases are processed quickly over a short period of time) must be taken into consideration when assessing the stability of a defendant's mental status. Further, the intensity of a defendant's mental symptoms can wax and wane with time and may be less related to the behavioral environment, necessitating the mental health consultant to educate the court about the nature and course of expected mental symptoms and the defendant's specific limitations and capacities.

A mental health professional involved in assessing competency is in a good position to inform the court about conditions that may be conducive or counterproductive to a defendant's court appearance or his ability to work with his attorney. An evaluator will sometimes be faced with conflicts that exist between the defendant and his attorney wherein the difficulty communicating is not based upon a mental illness as much as characterological issues, misunderstanding, or even limited opportunities for consultation.

Depending upon the time frame of a defendant's arrest, arraignment, and opportunity to consult with counsel, it might be that a temporary condition such as intoxication or a delirium interfered with initial communication abilities. Sometimes the context and manner of the initial contact can dictate how the working relationship will follow. Additionally, sometimes, defense attorneys will change from the time of arraignment to the start of a trial, which has the potential of adding confusion and sometimes frustration for a mentally ill defendant.

The defendant's behavior in court and while consulting with his attorney can vary and it is helpful to be able to delineate which elements might be related to characterological aspects that are more volitional, as opposed to symptoms of mental or medical impairment. Collateral information from a variety of sources aid in establishing a pattern of symptomatic presentation related to a mental illness versus personality dynamics. It is recognized that nervousness or anxiety about a court appearance is customary, and only when it reaches the level of significantly interfering with a defendant's comprehension and ability to attend to his attorney or the proceedings will it require intervention.

As variations in a defendant's mental symptoms could range significantly across time and situations, the tolerance level of the court's time schedule may impact a defendant's capacity to participate, follow court proceedings or cooperate with counsel. It is essential to apprise the court of a defendant's potential limitations (e.g., attention span, ability to comprehend or process verbally presented material, and potential sensory disabilities), so that the court may decide how much time it will allow or expect from defense counsel to spend with the defendant on his case.

There has been a recent development of Mental Health Courts in the United States (*see* **Mental Health Courts**). By 2005, there were approximately 125 operational courts in 36 states [8]. All require voluntary participation by the defendant. Mental Health Courts serve as a diversion into judicially supervised community-based mental health treatment programs where charges are deferred or dismissed pending successful completion of court-ordered mental health treatment. Initially, defendants charged with felony or violent crimes were not admitted; however, a second generation of mental health courts [9] that accept felony cases is beginning to emerge.

Other Related Trial Competencies

The legal standards for competency to stand trial are the same as for competency to be sentenced, to plead guilty, to waive an insanity defense, to waive privilege, and to waive the right to counsel, as each speaks to relevant decisional capacities that are specific to a defendant's particular case. Competency to proceed on a *pro se* basis (self-representation) is every defendant's right as long as the waiver of counsel is made voluntarily and intelligently [10], although at a minimum, the defendant "must have the capacity to understand and weigh the risks and benefits of self-representation, including the risk that he could be erroneously convicted due to deficiencies of his self-representation" [11] (p. 300). Although the legal standards are the same, the decisional expectations are not and vary depending upon context [11].

While everyone who is arrested has the right to remain silent [12], a person's competency to confess or to supply information is generally not assessed at the time a statement is given. The defendant's rights are generally read to him and he may be asked to initial or sign a document signifying that he had his rights read to him, but inquiry into the extent of a defendant's comprehension of these rights is generally not pursued. Assessing a defendant's competency to confess requires the evaluator to conduct an analysis of the defendant's mental status at the time he was questioned, akin to an evaluation of mental state at the time of the alleged offense (e.g., insanity). While it may be helpful to review the legal terminology utilized in Miranda [12] (*see also* **Head Injury: Neuropsychological Assessment**) when questioning a defendant about his understanding of these terms, it is also essential to consult with collateral sources about a defendant's general cognitive and intellectual functioning and behavioral accounts of the defendant at the time he received his rights particularly if mental health symptoms were present.

The legal standard for competency to be executed similarly parallels the Dusky [3] standard of possessing a rational understanding and an ability to cooperate with counsel. There may be ethical considerations of restoring a person's competency to be executed, such as in the case where a competent defendant's

mental condition deteriorates while awaiting execution, particularly if this requires involuntary administration of psychotropic medications. It was subsequently determined that the due process clause in the US Constitution would permit involuntary treatment of serious mental symptoms if the inmate presents a danger to himself or others [7]. It is unconstitutional to execute an insane person, per the Eighth Amendment in the US Constitution [13], and as found in the US Supreme Court case of *Ford v. Wainwright* [14].

In the case of *Singleton v. Norris* [15], after finding that the Sell [6] criteria was met, the Eighth Circuit court of appeals concluded that due process would not be violated by involuntary treatment as the mandatory medication scheme that was valid under the stay of execution would still be valid when an execution date was set. This matter continues to be debated, particularly in cases where an inmate may have a factual awareness of the reason for his execution but lacks appreciation that he is being executed for retribution because of a delusional belief [16].

Acknowledgment

The author wishes to extend her deepest appreciation to her husband, Bruce Gage, M.D. but for whom, without his support and suggestions, this article would have never seen the light of day.

References

[1] Cooper v. Oklahoma, 116 S. Ct. 1373 (1996).
[2] Drope v. Missouri, 420 U.S. 162, 95 S. Ct. 896 (1975).
[3] Dusky v. United States, 362 U.S. 402, 80 S. Ct. 788 (1960).
[4] Godinez v. Moran, 113 S. Ct. 2680 (1993).
[5] Jackson v. Indiana, 406 U.S. 715, 32 (1972).
[6] Sell v. United States, 539 U.S. (2003).
[7] Washington v. Harper, 494 U.S. 210 (1990).
[8] Mental Health Courts, A National Snapshot http://consensusproject.org/mhcourts_archive/national-snapshot (2005).
[9] Redlich, A.D., Steadman, H.J., Monahan, J., Petrila, J. & Griffin, P.A. (2005). The second generation of mental health courts, *Psychology, Public Policy and Law* **11**(4), 527–538.
[10] Faretta v. California, 422 U.S. 806 (1975).
[11] Bonnie, R.J. (1992). The competence of criminal defendants: a theoretical reformulation, *Behavioral Sciences and the Law* **10**, 291–316.
[12] Miranda v. Arizona, 384 U.S. 436 (1966).
[13] Mount, S. United States Constitution - Amendment 8 US Constitution.net 10 May 2008. http://www.usconstitution.net/xconst_Am8.html (26 Dec 2008).
[14] Ford v. Wainright, 106 S. Ct. 2595 (1986).
[15] Singleton v. Norris, 319 F.3d 1018 (8th Cir., 2003).
[16] Panetti v. Dretke, 448 F.3d 815 (5th Cir., 2006).

Related Articles

Delusions

Hallucinations

INDRA A. FINCH

Capacity to Waive Insanity Defense *see* Capacity Assessment

Capacity to Waive Legal Counsel *see* Capacity Assessment

Capacity to Waive Miranda Rights

In *Miranda v. Arizona* [1, 2], the US Supreme Court ruled that any statement arising from a custodial interrogation of a suspect would be presumed to be involuntary and not admissible unless the suspect is informed of his rights to remain silent, to avoid self-incrimination, to obtain legal counsel before and during police questioning, and to obtain free legal counsel if indigent.

Litigation concerning a defendant's capacity to confess has been increasing in the criminal and juvenile courts. A confession or incriminating statement given by a suspect can greatly influence the final court

decision of a defendant's guilt or innocence. Self-incriminating statements to law enforcement, even in the absence of other incriminating evidence, often produce guilty verdicts [3]. Mental health professionals have assisted the court in assessing a defendant's capacity to have waived *Miranda* rights at the time of the interrogation. Like the evaluation to assess a defendant's legal insanity, the *Miranda*-focused evaluation involves a retrospective analysis of the defendant's mental state at the time of the police questioning. The clinician evaluates current capacity, to gain insight into the defendant's present abilities, and then extrapolates the analysis to the point in the past when the defendant had actually waived the rights [4]. In 1966, the US Supreme Court ruled in *Miranda v. Arizona* that a suspect's statement is presumed involuntary and inadmissible in court if law enforcement does not provide four warnings. These warnings include (i) the right to remain silent, (ii) that any statement can be used against the suspect in the future court proceedings, (iii) the right to the presence of an attorney before and during the interrogation, and (iv) an attorney provided free of charge if the suspect is unable to pay for one. Some jurisdictions have added a fifth warning, whereby the suspect is informed that the rights can be invoked at any time. The 1967 case, *In re. Gault* [5] applied the Miranda protections to juvenile suspects.

The terms *competency to confess* or *competency to waive Miranda rights* have been used interchangeably with capacity to waive *Miranda* rights. While competency generally refers to the legal determination by the court, capacity has to do with the individual's ability to waive the rights at the time of the interrogation.

Mental health testimony is generally presented at a pretrial suppression hearing. A forensic evaluation must be functionally based; that is, integrated with the legally relevant criteria [6, 7]. The astute clinician should break down the assessment into a determination of factors relevant to a knowing waiver, an intelligent waiver, and a voluntary waiver.

Knowing, Intelligent, and Voluntary Waiver of Rights

A mental heath professional can offer an opinion whether a waiver of *Miranda* was made knowingly and intelligently. A knowing waiver is the individual's understanding or comprehension of the rights in addition to the manner in which the rights were administered [8]. The rights can be read by law enforcement or the suspect might be given a "waiver form" to read and sign. With the latter, an assessment of reading comprehension would be important. Written *Miranda* waiver forms vary greatly in their length and complexity, from a 2.8 grade level to postcollege [9]. The readability of the waiver form can be analyzed by the Flesch–Kincaid readability formula, easily calculated with word processing programs. Comprehension can also be affected by how the rights were presented, such as a police officer handing the suspect a form with their rights and asking the suspect to sign it without explanation.

An intelligent waiver of rights involves a decision-making capacity, an appreciation of the rights based upon knowledge of the legal process. For example, a defendant may understand one has the right not to speak to the police but cannot make an intelligent use of the right to silence because of a faulty belief that invoking the right to silence would be perceived by the trier of fact as the actions of a guilty person. It must be noted that in the United Kingdom, unlike in the United States, a defendant who elects not to speak to law enforcement can have that right to silence be weighed negatively by the trier of fact. It can be assumed that the suspect has something to hide.

The mental health professional should be cautious in directly opining on the third prong of a valid waiver, the voluntariness component. In the 1986 US Supreme Court case, *Colorado v. Connelly*, the court ruled that a confession "compelled" by hallucinatory and delusional thinking was voluntarily given because the police did not take advantage of Connelly's mental state to extract the confession. There was no showing of undo influence. The court ruled that there must be a "link" between the *Miranda* waiver and inappropriate police conduct [10]. It is not a clinical issue whether law enforcement overstepped their bounds in extracting a *Miranda* waiver. Nevertheless, clinicians can assess relevant psychological factors that make an individual more susceptible to the effects of police conduct. Valuable information can be given to the court regarding such factors as the defendant's interrogative suggestibility, compliance, intelligence, anxiety, memory, and the effects of drug intoxication and sleep deprivation.

The mental health profession understands that a judicial determination of the validity of a *Miranda* waiver is dependent not only on the evaluation results

but also on the *totality of circumstances* surrounding the *Miranda* waiver. This includes the nature of the interrogation process and a defendant's individual characteristics, which may increase or reduce comprehension of the rights [11]. In that respect, there is no one IQ cutoff or a diagnostic threshold, which would automatically render a defendant incompetent to have waived rights during the interrogation.

Protocol for Miranda Capacity Evalutaion

Frumkin [4] proposes a protocol for conducting a capacity to waive *Miranda* rights evaluation. The importance of reviewing relevant third-party data cannot be overstated. The referring attorney should collect for the clinician a copy of the *Miranda* waiver form or the card from which *Miranda* was read. The mental heath professional should also obtain from the attorney, school records, work records, psychological and medical records, and any and all records which would help one interpret evaluation findings in a historical context. If it exists, it is important to review the audio or videotape of the interrogation and/or depositions of law enforcement. Although it is important for the psychologist to obtain an arrest history of the defendant, research has shown [8] no simple correlation between an arrest history and *Miranda* comprehension and appreciation.

After a review of third-party data, a thorough clinical interview is conducted including a psychosocial history and mental status examination. An essential part of the evaluation consists of obtaining the defendant's step-by-step version of what transpired with first contact with law enforcement, up through the *Miranda* waiver and subsequent statement. The goal is to get the defendant to describe how the rights were administered by law enforcement. This is needed to assess the defendant's ability to recite the rights, spontaneously from memory, and also whether the rights can be remembered immediately after they are read by the examiner or by the defendant himself (depending upon the manner in which the rights were administered by the police). It also allows an assessment of the defendant's reading and listening abilities.

Intelligence testing, such as administration of the Wechsler scales are important not only to obtain IQ scores but also to obtain scaled scores from subtests which are relevant to the types of cognitive capacities need to understand and appreciate the abstract concepts contained in *Miranda.* Reading comprehension testing is also warranted if the defendant was asked to read the rights by law enforcement. Personality testing can be useful in examining psychological variables that could impact upon an individual's ability to comprehend and process information as well as for issues relevant to the voluntariness of the *Miranda* waiver. As with any forensic evaluation, the clinician should assess the minimization or exaggeration of cognitive and/or psychopathological symptoms, (*see* **Malingering: Forensic Evaluations**).

Specialized Tests

Thomas Grisso, as part of a large federally funded research project in the late 1970s, developed four tests aimed at assisting a clinician in evaluating a defendant's capacity to make a knowing and intelligent waiver of *Miranda* rights at the time of the police questioning [8]. These tests help assess the current *Miranda* understanding and appreciation. It is up to the mental health professional to integrate this data with the defendant's history, what transpired during the police questioning from both the law enforcement and defendant perspectives, behavioral observations, and intelligence and personality testing data, to make a retrospective analysis of what the defendant would have understood and appreciated at the time of the police questioning.

The four tests are the Comprehension of Miranda Rights (CMR), the Comprehension of Miranda Rights-Recognition (CMR-R), the Comprehension of Miranda Vocabulary (CMV), and the Function of Rights in Interrogation (FRI). The CMR helps assess an individual's current understanding of the *Miranda* rights (knowing waiver). The subject is shown each of the four *Miranda* rights. After the examiner reads the right, the subject must state what the right means in his or her own words. The CMR-R is a less verbal means of helping to assess an understanding of the rights. The subject is read and shown each *Miranda* right and is then given comparison statements. The subject must state whether the comparison statement means the same or different from the *Miranda* right. The CMV requires the subject to define six presented words often found in *Miranda* warnings. The usefulness of this test is limited if few of the six words are actually contained in the *Miranda* version given by the police. The FRI measures factors

relevant to an intelligent waiver of the rights, the ability to appreciate the significance of the *Miranda* rights based upon a subject's knowledge of the legal system. It is composed of three subscales, Nature of Interrogation, Right to Counsel, and the Right to Silence. They are thought to be independent variables and scores do not correlate well with each other [8].

Scoring for the Grisso tests, as they are commonly referred, is from a test manual. The manual was derived by incorporating the work of a national panel of judges, lawyers, and legal scholars. They decided what types of responses indicated a full understanding or appreciation of the right (two-point response), what type of responses indicated partial understanding of the right (one-point response), and what type of responses indicated no understanding (zero-point response).

These tests have normative data for both adults and juveniles. Scores can be compared to individuals based on age and IQ. As Grisso and others have emphasized, the tests are to be used as an aid to assist in evaluating capacity to waive *Miranda* rights. The tests do not provide data relevant to the voluntariness of a waiver. They only help assess current understanding and appreciation, not capacity at the time of the interrogation. The tests are frequently misused by clinicians who put undo weight on the scores or the percentiles and who also do not take into consideration the limitations of the tests themselves [4, 12, 13].

Research regarding Miranda Understanding and Appreciation

Research has shown [8, 14] a strong relationship between intelligence and capacity to understand and appreciate the *Miranda* warnings. In fact, with most individuals with IQs in the borderline to mental retardation range, the *Miranda* warnings are "words without meaning" [15, 16]. There is also a relationship with age. Juveniles aged 14 and younger do not understand or appreciate the rights as well as older juveniles and adults. In fact, 23% of adults and 55% of juveniles do not understand at all at least one of the four *Miranda* rights [8].

Owing to the continued maturational growth of juveniles and evidence that during adolescence, performance improves in areas such as memory, attention, reasoning, and executive functioning, many juveniles have not yet developed the cognitive abilities to make an intelligent waiver [14]. Suspects need to have enough cognitive capacity to enable comprehension and reasoning abilities to be used to make rational legal choices.

The clinician needs to be aware of the research concerning mental disorders and *Miranda* comprehension. Mental illness, *per se*, does not automatically equate with incapacity to understand and appreciate the rights. A person with a psychosis can still have the capacity to waive their rights, as long as their mental disorder does not affect their cognitive abilities to make a knowing or an intelligent waiver. Research suggests that psychosis in conjunction with low intelligence may be a good predictor of incapacity to waive the Canadian rights [17]. Psychosis by itself was not a good predictor of one's lack of ability to comprehend the meaning and effect of the rights. No link was found between depression, anxiety, or behavior problems and the capacity to waive *Miranda* rights [14].

Voluntariness Component and Interrogative Suggestibility

The clinician can provide useful information to the court on psychological characteristics, which make a defendant more likely to be misled by police or to change responses under pressure compared to other people. This, as well as other psychological variables (such as compliance, submissiveness, coping skills, and impulse control) may also be the relevant factors the court may wish to consider when determining the voluntariness of a *Miranda* waiver. Personality tests, such as the Minnesota Multiphasic Personality Inventory-2 (MMPI-2), Personality Assessment Inventory (PAI), and the 16 Personality Factor (16 PF) may provide useful information regarding a defendant's psychological functioning and how that functioning interacts with the demands of the interrogation process (*see* **Psychological Testing**).

Interrogative suggestibility (*see* **Interrogative Suggestibility**) is defined as the extent to which an individual comes to accept messages or information communicated during formal questioning, essentially believing as true the information given. The Gudjonsson Suggestibility Scales (GSS) [18] is a specialized psychological test that can be used to identify those most susceptible to give into leading

questions and shift to different responses under pressure. The examinee thinks that the GSS is a memory test. A narrative story is read aloud and the subject is asked 20 questions about the story, 15 of the questions are misleading. The test quantifies how much the person "yields" to the leading questions, both after being asked the leading questions, and later when firmly told that errors were made in answering questions. Shift is also calculated, which is how many times the subject shifts from one response, right or wrong, to a different response. A total suggestibility score is calculated, which is the sum of the first Yield and the Shift. The ultimate determination of a suspect's capacity to waive *Miranda* rights at the time of the police questioning falls in the domain of the court. Forensic mental health professionals can provide valuable data to the court to assist in the determination whether a defendant made a knowing, intelligent, and voluntary waiver of rights.

References

[1] Miranda v. Arizona, 384 U.S. 436 (1966).

[2] Walker, J.K. (1993). A comparative discussion of the privilege against self-incrimination. *New York Law School Journal of International and Comparative Law*, **14**, 1–38.

[3] Kassin, S. & Neumann, K. (1997). On the power of confession evidence: an experimental test of the fundamental difference hypothesis, *Law and Human Behavior* **21**(5), 469–484.

[4] Frumkin, I.B. (2008). Psychological evaluations in *Miranda* waiver and confession cases, in *Clinical Neuropsychology in the Criminal Forensic Setting*, R. Denny & R. Sullivan, eds, Guilford Publications, New York, pp. 135–175.

[5] In re Gault, 387 U.S. 1 (1967).

[6] Grisso, T. (2003). *Evaluating Competencies: Forensic Assessments and Instruments*, 2nd Edition, Plenum, New York.

[7] Heilbrun, K. (2001). *Principles of Forensic Mental Health Assessment*, Kluwer Academic Press, New York.

[8] Grisso, T. (1998). *Instruments for Understanding and Appreciating of Miranda Rights*, Professional Resource Press, Sarasota.

[9] Rogers, R., Harrison, K.S., Shuman, D.W., Sewell, K.W. & Hazelwood, L.L. (2007). An analysis of *Miranda* warning and waivers: comprehension and coverage, *Law and Human Behavior* **31**, 177–192.

[10] Colorado v. Connelly, 479 U.S. 157 (1986).

[11] Fare v. Michael, 442 U.S. 707 (1979).

[12] Rogers, R., Jordan, M. & Harrison, K. (2004). A critical review of published competency to confess measures, *Law and Human Behavior* **28**, 707–718.

[13] Grisso, T. (2004). Reply to "A critical review of published competency to confess measures", *Law and Human Behavior* **28**, 719–724.

[14] Viljoen, J. & Roesch, R. (2005). Competence to waive interrogation rights and adjudicative competence in adolescent defendants: cognitive development, attorney contact and psychological symptoms, *Law and Human Behavior* **29**, 723–742.

[15] Fulero, S. & Everington, C. (1995). Assessing competency to waive *Miranda* rights in defendants with mental retardation, *Law and Human Behavior* **19**, 533–543.

[16] Cloud, M., Shepherd, G., Berkoff, A. & Shur, J. (2002). Words without meaning: the constitution, confessions, and mentally retarded suspects, *The University of Chicago Law Review* **69**, 495–624.

[17] Viljoen, J., Roesch, R. & Zapf, P. (2002). An examination of the relationship between competency to stand trial, competency to waive interrogation rights, and psychopathology, *Law and Human Behavior* **26**, 481–506.

[18] Gudjonsson, G. (1997). *The Gudjonsson Suggestibility Scales Manual*. Psychology Press,20 Hove, U.K.

I. Bruce Frumkin and Gretchen M. Lamendola

Capacity to Waive Privilege *see* Capacity Assessment

Capital Punishment Intelligence *see* Mental Retardation: Death Penalty

Carbon-14 Bomb Pulse Dating *see* Bomb-Pulse Dating

Cardiac and Natural Causes of Sudden Death

Definitions

In the discussion that follows we have adopted the terminology used by the World Health Organization, which defines "sudden unexpected death" as that occurring within 24 h of the onset of symptoms, where natural disease or functional disorder is recognized as a cause of death. The term "sudden cardiac death" has traditionally been reserved for deaths that occur within 1 h of symptom onset. Worldwide, such deaths are almost always related, in one way or another, to the existence of ischemic heart disease. More recently, new subdivisions have been added to the original classifications. The diagnosis of "sudden unexpected death in epilepsy" is widely accepted as a distinct entity. In 1993, the US Food and Drug Administration (FDA) first developed the generally accepted criteria for the diagnosis of sudden unexpected death during epilepsy (SUDEP). To be classified as SUDEP, the patient must have a proven history of recurrent seizure but must be otherwise healthy at the time of death, with death occurring within minutes, but in the absence of seizure activity. Finally, and perhaps most important, no cause of death is apparent at autopsy. Lastly, the term sudden infant death syndrome (SIDS) is now reserved for death where no cause of death is apparent. If a cause is apparent, death is then classified as being a result of sudden unexpected infant death syndrome (SUID) or unexplained infant. Such deaths are usually cardiovascular or respiratory in nature.

Sudden deaths (excluding traumatic deaths) can be divided into four general groups:

1. *physiological or death due to genetic defect*;
2. *pathological death* due to diseases (including malnutrition and starvation), which can be slow or sudden;
3. *accidental or violent death* due to trauma, such as a wound, or poisoning, which, unless homicidal, is also accidental (for example, an addict dying from clostridia cardiotoxicity);
4. *unexplained death* when all findings are insufficient to explain its cause.

Historical Considerations

The study of sudden, unexpected deaths goes back many centuries. Leonardo da Vinci described sudden, unexpected death as *la dolce morte* (the "sweet death"), because the very speed with which it occurs tends to shorten the grieving period. Sudden death can occur in conjunction with many different diseases. Perhaps the best known example occurred in 480 B.C., with the sudden cardiac death of the soldier Philippides. He had run from Maratona to Athens in order to announce victory against the Persians. Another example occurred in 1705 when so many people died during an epidemic of sudden death in Rome that Pope Clemente XI ordered his "archiatra" (doctor in chief), named Giovanni Maria Lancisi, to undertake an autopsy investigation. The results were published under the title *De subitaneis mortibus* ("the gross anatomical findings"). He found cardiac involvement in most of the decedents. The victims described in early reports of sudden death were generally men living a life of luxury. However, this disease process soon came to involve those not so well-off [1]. Even today the pathophysiology of this disorder remains poorly understood, although there is definite progress to report (see below). Today most instances of sudden unexpected death (SUD), autopsied or not, are really cases of sudden cardiac death (SCD), a consequence of heart disease, especially coronary atherosclerosis. Obviously, this does not apply to individuals under 40, where deaths from coronary artery disease are uncommon, but deaths from cardiomyopathy are fairly frequent.

This chapter is devoted to the analysis of "any *death which is rapid (without prodrome), unexpected and/ or unforeseen, that occurs in apparently healthy people, or in ill patients during a benign phase of their disease"*. A negative history does not necessarily exclude the previous occurrence of a silent infarct, and the finding of a myocardial monofocal fibrosis, involving 10% or more of the left ventricular mass, should be considered as an acceptable sign of an old, undiagnosed myocardial infarct. It is increasingly recognized that the existence of such a scar predisposes to sudden death because of recurrent ventricular tachycardia arising from the boundaries between the scar and normal tissue [2].

For many years it was commonly held that SCD was the result of ischemia induced by the presence

of critically stenotic lesions. In fact, stenotic lesions are present less than 30% of the time [3].

Here, we review the most common natural causes of sudden, unexpected death. In the United States, more than 325 000 cases of fatal myocardial infarction occur annually [4], and most of these deaths are sudden. We also know that in North America the annual incidence of out-of-hospital cardiac arrest is approximately 0.55 per 1000 population; this implies that about 164 600 episodes of out-of-hospital cardiac arrest occur annually in the United States. The incidence in other industrialized countries is probably comparable if, for no other reason than that, the incidence of coronary artery disease varies little from one industrialized country to another. Compared with coronary artery disease, the incidence of all other causes of SCD is negligible, except in the young who do not, as a rule, suffer from coronary disease. In fact, the investigation of sudden death in the young and apparently healthy poses a difficult challenge for death investigators and is the focus of this article.

In 2001, Virmani *et al.* systematically analyzed the pathologic changes observed in a large consecutive group of young SUD victims (31–40 years) [5]. The findings are illustrated in Table 1. Nearly a decade has passed since this list was originally compiled. If it were compiled again today there would be substantial differences; the frequency of cardiomyopathy, inflammatory myocardial diseases, and ion channel disorders would be much greater [6]. The increased availability of genetic testing has substantially increased our awareness of many disorders and increased the frequency with which they are diagnosed. At the same time, some of the diagnostic categories, such as "idiopathic left ventricular hypertrophy (LVH)", probably would now be eliminated entirely because there etiology is now understood; the genetics of hypertrophic cardiomyopathy (HCM) has been largely characterized, and the explosive increase in stimulant drug abuse, with the powerful ability of these drugs to induce LVH (itself a powerful predictor for sudden arrhythmic death), is increasingly recognized [7].

Table 1 Pathologic changes in young SUD victims $n = 229$ (all aged 31–34 years)

1.	Atherosclerosis, 60%
2.	No finding, 9%
3.	Hypertensive LVH, 6%
4.	Idiopathic left ventricular hypertrophy, 4%
5.	Dilated cardiomyopathy, 4%
6.	Hypertrophic cardiomyopathy, 3%
7.	Myocarditis, 3%
8.	Sarcoidosis, 2%
9.	Aortic dissection, 2%
10.	Endocarditis, 1%
11.	Floppy mitral valve, 1%
12.	Coronary artery tunneling, 1%
13.	Right ventricular dysplasia, 1%
14.	Rheumatic mitral stenosis, 1%
15.	Anomalous coronary artery, 0.5%
16.	Coronary artery dissection, 0.4%
17.	Congenital heart disease, 0.2%
18.	Lipomatous hypertrophy, atrial septum, 0.2%

Specific Medical Disorders

Coronary Artery Disease (60%)

Table 2 shows the *percentage distribution*, the degree of maximal stenosis, and the frequency of occlusive thrombi in those with and without evidence (monofocal scarring) of previous infarction (SUD chr) compared with noncardiac patients dead from other diseases (NCA) and patients dying accidental deaths (AD).

These data clearly demonstrate that SCD from acute myocardial infarction (AMI), as the first indicator of coronary heart disease, may occur in absence of a critical stenosis (SUD 35%; AMI 11%). It is also apparent that "chronic" cases had more severe and more extensive coronary atherosclerosis, but a lower frequency of occlusive thrombus (15 vs. 40%) in AMI. The high frequency of severe coronary atherosclerosis (similar to that seen in SUD) in the control no cause apparent (NCA) and in accidental death (AD) group is also noteworthy [1].

Intracoronary perfusion with radio opaque material will demonstrate an extensive compensatory collateral system, formed by intramyocardial giant, capillarylike vessels, joining branches of the same artery (homocoronary) or branches of different arteries (intercoronary), in normal and diseased hearts. At the level of the atherosclerotic plaque, proximal and distal secondary branches may be seen communicating with a neovascularized atherosclerotic intimal (innermost) and residual lumen. The collateral system was identical in SUD, AMI, and controls. The enlargement of normal anastomotic vessels was proportional to the degree and number of critical stenosis [1].

Table 2 Percentage distribution of maximal lumen reduction and number of atherosclerotic coronary arteries with severe (> 70%) lumen/diameter sterosis

Degree of coronary stenosis %	SUD 1st 133 cases	SUD chr 75 cases	AMI 1st 145 cases	AMI chr 55 cases	NCA 100 cases	AD 97 cases
0	8	–	2	–	7	8
<50	13	–	2	2	10	21
50–69	13	7	7	–	17	32
70	16	11	21	14	11	19
80	29	18	31	20	24	14
90	20	64	37	64	31	6
Severe >70% in						
1	30	17	42	29	26	23
2	25	35	34	40	18	13
≥ 3 vessels	10	41	13	29	22	3
Occlusivethrombus	8	28	41	40	–	–

The lack of symptoms or signs of ischemia in this group of decedents clearly demonstrates the importance of collateral connections. In chronic patients, the higher frequency of critical, multiple stenotic lesions is not an indication of more severe coronary disease but, rather, evidence that the atherosclerotic process simultaneously involves the entire coronary tree. Studies by several investigators suggest that in patients with acute myocardial infarction, all three major coronary arteries are widely diseased and have multiple yellow though nondisrupted plaques, suggesting that AMI may represent a pan-coronary process of vulnerable plaques in all stages of development throughout the process [8].

However, because of the established compensatory flow from the collaterals, aggravation of atherosclerotic process does not imply clinically evident dysfunction; the presence of these collaterals raises doubts about the significance of occluding thrombi. When present, they are found only in old plaques in vessels containing a critical lumen reduction. This observation strongly suggests that existing collateral flow will nullify the effects of thrombotic occlusion, which in turn implies that the thrombus may not be the primary cause of sudden death or infarct, but rather a secondary phenomenon [1].

Natural History of Coronary Atherosclerotic Plaque in Humans

We have performed systematic, quantitative, histological examination of diseased coronary arteries of SCD victims, quantitating intimal thickness increases and lumenal reduction, and made morphometric measurements of the length, shape (concentric, semilunar), and intimal/media thickness of coronary plaques in large groups of sudden death victims: 100 acute myocardial infarcts, 208 decedents with sudden/unexpected death cases, 50 individuals with chronic coronary heart disease patients, and 97 healthy controls. This included examination of coronary sections taken from the origin of left main trunk, left anterior descending branch, left circumflex branch, right coronary artery, posterior descending branch and the middle tract of anterior descending, and marginal and posterior tracts of right coronary artery [1]. Of 3640 coronary sections, 1519 without and 1315 with a lumen reduction less than 70%, and an intimal physiological thickness less than 300 or 600 μm, respectively, we never observed fatty streaks, lipoprotein–cholesterol subendothelial infiltration, monocytes, macrophages, foam cells, platelet aggregates, fibrin-platelet thrombi, inflammatory elements, or intimal fissures.

SCD with Normal Heart (9%)

Many of these cases are a consequence of genetic aberration and, more likely than not, these particular mutations act to disrupt normal myocardial repolarization, resulting in the family of diseases known as “long QT syndromes” or “LQTS”. The QT interval represents the total duration of ventricular systole. It is controlled by the sequential opening and closing of various sodium and potassium channels during cardiac depolarization and repolarization [9]. Under normal circumstances, the QT interval should be less

than half as long as the time elapsed from one "r" wave to the next. If it is longer, then repolarization is also prolonged and the chances increase that an "r" wave will fall on a "T" wave and trigger a fatal arrhythmia. More than 80% of LQT1, 2, 3 deaths occur during exercise or after being startled, chiefly because as the heart rate increases, the r–r interval becomes shorter.

The incidence of LQTS defects is estimated at 1 : 5000, but if acquired forms of LQTS are included, the true incidence of LQTS may be closer to 1 in 1000 [10]. At least eight different genes are involved, though disorders such as idiopathic ventricular fibrillation [11], progressive cardiac conduction disease [12], and the only recently recognized short QT [13] syndrome are all very rare. There are two types of defects, depending on whether alpha or beta channel subunits of the channel are involved [14].

By far the most interesting of the mutations involves the gene for the human ether-a-go-go-related gene (HERG) potassium channel, located on Chromosome 7 and the *MiRP1* gene located on Chromosome 21. These interact to form I_{rk}, the inward rapid potassium current. If there is a defect in HERG, the resultant syndrome is called *LQT2*; if on *Mir*, it is called *LQT6*. LQT3 is due to mutations in the sodium cardiac channel SCN5A, and the resulting syndrome is commonly referred to as *Brugada syndrome*. The LQT4 syndrome is due to a mutation in the ankyrin-B protein [15]. LQT7 is the result of yet another defect in the alpha subunit of the I_{Kr} channel. No matter what the defect is, the result is the same: prolonged repolarization. Testing of patients with suspect LQTS syndromes is available commercially in the United States, but the cost is still prohibitively high [16].

QT prolongation leads to the occurrence of a lethal type of ventricular tachycardia known as *torsades des pointes* (*TdP*). The blockade of any potassium channel can precipitate torsades des pointes (TdP). Most, if not all, clinical cases of TdP are drug-induced, secondary to abnormalities of the (HERG) ion channel, the one that conducts the rapid component of the delayed rectifier KC current I_{Kr} [17].

LQT3 is a rare syndrome, accounting for less than 5% of all LQTS. It is the result of a "gain of function" mutation that causes sodium channels to remain open. Asian men suffering from the disorder known as *bangungut* (literally, "sleep death") die during nightmares [18]. Alternatively, there may be a "loss of function", and the sodium channels stay closed and the result is Brugada syndrome. LQT3 is the only "channelopathy" associated with morphological changes. Biopsy samples obtained from 18 patients who were believed to have this syndrome were found to contain detectable SCN5A mutations in only 4 patients, but structural alterations were seen in 14: right ventricular micro aneurysms in 7 patients and 4 of these also had micro aneurysms on the left. Right ventricular cardiomyopathy was present in one patient; cardiomyopathic changes were noted in three additional cases [19]. The findings reinforce the relatively recent notion that there may also be acquired forms of Brugada syndrome.

The ryanodyine receptor (Ry) is the single largest gene in the human body, with 102 exons. Death from this mutation occurs as a result of stress-induced bidirectional ventricular tachycardia, which is also called *catecholaminergic polymorphic ventricular tachycardia* (CPVT). The arrhythmia occurs because of abnormal calcium concentration variations within the individual myocytes. In this disorder, Ry is inherently unstable and prone to unpredictable failure [20]. When that occurs, the cell is flooded with excessive amounts of calcium producing electrical instability. When the Ry mutation occurs in skeletal muscle, it causes malignant hyperthermia (anesthesia-related death) and central core disease. In cardiac muscle, RYR2 mutations cause CPVT, and still other mutations in the same gene can lead to the occurrence of right ventricular dysplasia (see below) [21].

The importance of these disorders in the process of death investigation cannot be overemphasized. In a postmortem study published in 2005, LQTS genetic testing was performed in a cohort of 49 SCD cases, with an average age at death 14.2 ± 10.9 years. A combination of genetic techniques was used to identify all eight genes implicated in the pathogenesis of either LQTS (LQT1 to LQT6) or multisystem disorders involving either QT or QU prolongation. Ten LQTS-associated mutations (four novel) were discovered in the 10 SUD cases, with 20% occurring in women with an average age at death 18.0 ± 11.8 years). The activities at the time of SUD included sleep (five), exertion (two), auditory arousal (one), and undetermined (two). Sudden death was the sentinel event in two-thirds of the cases [22].

Hypertensive Left Ventricular Hypertrophy (6%)

LVH is associated with a marked increase in premature ventricular contractions and complex arrhythmias, as well as the occurrences of SUD [23, 24]. This increase is independent of all other risk factors, including even atherosclerosis. In hypertensives and athletes, LVH represents a physiologic response to increased workload, though in the hypertensives, it is a counterproductive one. Although not apparent to the naked eye, LVH is associated with decreased coronary artery reserve, decreased capillary density, and increased diffusion distance (the distance that oxygen and nutrients must traverse to reach the enlarged cardiomyocytes). As a consequence, there is relative ischemia at all times, even in the absence of atherosclerotic narrowing or increased workload. This situation predisposes to electrical instability and an increased incidence of SUD. The process of myocardial remodeling also increases the incidence of SUD. Many important myocardial changes can only be seen at the microscopic level (see below), and all of them favor the occurrence of arrhythmias. At the macroscopic level, the heart assumes a spheroid shape, leading to repolarization dispersion ("QT prolongation"), which also carries with it increased risk of SUD, probably by the same mechanisms involved in other forms of long QT. Even modest degrees of hypertrophy are associated with increased risk for SUD; therefore, it is extremely important that the heart be weighed and normalized to standard tables, such as the Mayo Clinic Nomogram [25].

Idiopathic Left Ventricular Hypertrophy (4%)

This term is outdated and should really be abandoned. If the cause of hypertrophy is not apparent, then there are really only two possible explanations: HCM (*see* section "Hypertrophic Cardiomyopathy (HCM)") and drug-induced hypertrophy. In industrialized countries, the latter is almost always a consequence of substance abuse. There is evidence that most abused stimulant drugs (cocaine, methamphetamine, and [3, 4] methylenedioxymethamphetamine (MDMA)) initiate a process of myocardial remodeling that ultimately results in LVH (Figure 1). The process involves individual myocytes and the collagen network that supports them. The process is required so that the heart can adapt to new working conditions. Part of the remodeling process is initiated by the occurrence of cell death and the activation of fetal genes, whether from ischemia, toxin, or programmed cell death (apoptosis). No matter what initiates the process, myocardial collagen content increases, as does the severity of ventricular fibrosis [26, 27].

The pattern of fibrosis resulting from stimulant abuse resembles that seen in pheochromocytoma or any other form of catecholamine toxicity, such as

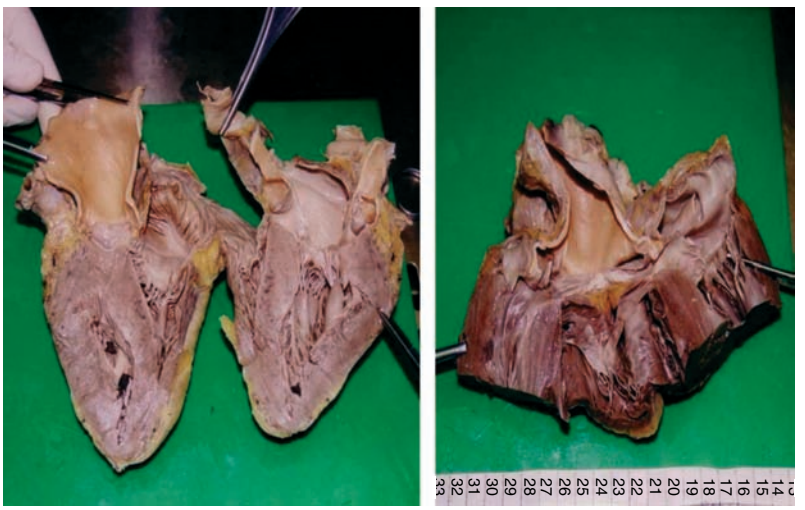


Figure 1 Left ventricular hypertrophy on "long axis" and "short axis" section

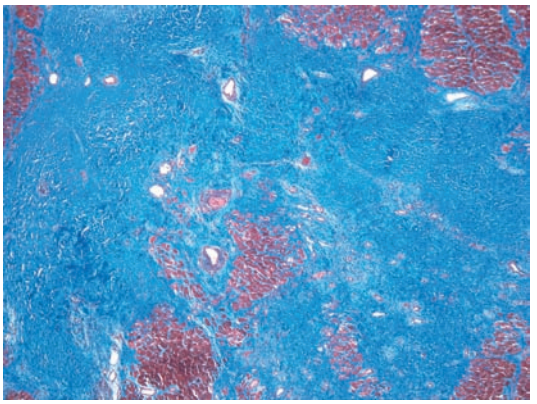


Figure 2 Dense and compact fibrosis as repair process of an ischemic necrosis (Trichromic stain)

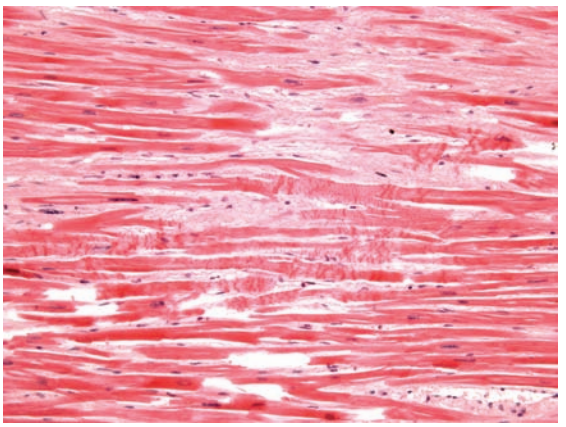


Figure 3 Microfocal fibrosis as result of catecholamine contraction based necrosis

methamphetamine or cocaine abuse. Damage to the heart is microfocal in nature and can be easily distinguished from the type associated with healing ischemic necrosis (Figure 2). When lesions are due to catecholamine toxicity, individual damaged cells are seen surrounded by normal myocytes, and damage cannot be related to the territory served by any one blood vessel (Figure 3) [28].

Hypertrophy in drug users is due to direct activation of the gene for calmodulin kinase II [29]. Other changes occur at the same time; the predominant form of myosin in the heart shifts to a slow isoform, and there is increased production of atrial naturetic factor. The rennin–angiotensin system is activated, and calcium levels within the endoplasmic reticulum decrease [30], generally reducing the cardiac output. Additional changes occur in the number and function of β-receptors (to compensate for the decreased output) while, at the same time, potassium channels controlling myocardial depolarization begin to function less efficiently [31, 32]. All these processes favor the occurrence of QT dispersion.

Dilated Cardiomyopathy (4%)

Cardiomyopathies are usually classified by their appearance (i.e., dilated, hypertrophic, and restrictive). Many etiologies are recognized, but all seem to act via identical common pathways. Dilated cardiomyopathy is by far the most common form and it is frequently responsible for heart failure. It is estimated that 400 000 cases occur in the United States every year [4]. Familial cardiomyopathy, mainly inherited as an autosomal dominant, was once considered rare, but is now diagnosed with increasing frequency. Cardiac enlargement and systolic dysfunction of one or both ventricles leading to congestive failure are the main clinical findings. Early phases of the disease are difficult to recognize; ventricular dilatation and pump dysfunction may remain asymptomatic for years and, in fact, severe pump dysfunction, with an adverse prognosis identical to that associated with dilatation, may occur with only minimal dilatation. Ventricular arrhythmias, including ventricular tachycardia, are observed in 20–60% of patients [33]. Systemic or pulmonary embolism has been reported in 3–18% of patients who have very severe forms of the disease. Echocardiography has shown that left ventricular thrombi are present in 8–60% of patients [34].

The most common pattern seen at autopsy is four-heart chambers dilatation with an increased heart weight (>500 g), despite a normal or reduced cardiac wall thickness. Microscopic examination disclosed abnormal myocytes with attenuation, edema with cloudy swelling, loss of myofibrils (colliquative myocytolysis), and often nuclear atypism. These changes are most evident in the left ventricular wall. Minimal lymphocytic infiltration with minor interstitial myocardial fibrosis is often present.

Hypertrophic Cardiomyopathy (HCM) (3%)

This disease is inherited as an autosomal dominant affecting 1 in 500 people (or about 0.2% of the general population). As such, it is probably the second

most common cause of SUD after atherosclerosis. Most never display any symptoms. Death is particularly likely in competitive athletes because their disease actually leads, albeit transiently, to improved cardiac performance with supernormal ventricular function and ejection fractions of 65–80%. HCM mutations have been identified in 14 different genes and account for more than 500 possible mutations. Most of the mutations are single nucleotide substitutions found within coding exons, intron–exon junctions, or promoter regions of the genes for β-myosin heavy chain, cardiac troponin T, or myosin binding protein-C. Those most likely to die suddenly are those with clusters of mutations in the β-myosin heavy chain gene (*MYH7*), a mutation associated with marked hypertrophy and poor prognosis [35, 36].

Not all forms of HCM result in obvious cardiac hypertrophy. In two-thirds of the cases, there is asymmetric hypertrophy of the interventricular septum, but in nearly a quarter of the cases the entire ventricle is involved. In a very small proportion of patients (approximately 10%), hypertrophy is confined to the apex. The latter pattern is much more common in Japan than in the West [37]. Traditionally, it has been presumed that when sudden death occurred in patients with HCM, it was, somehow, a consequence of myocyte disarray and fibrosis (see Figure 4). It is now clear that adrenergic stress of any sort can produce the same type of disarray and that the presence of disarray is a nonspecific finding [38]. There exists a strong association between hypertrophy and the occurrence of lethal arrhythmias due to QT dispersion [39].

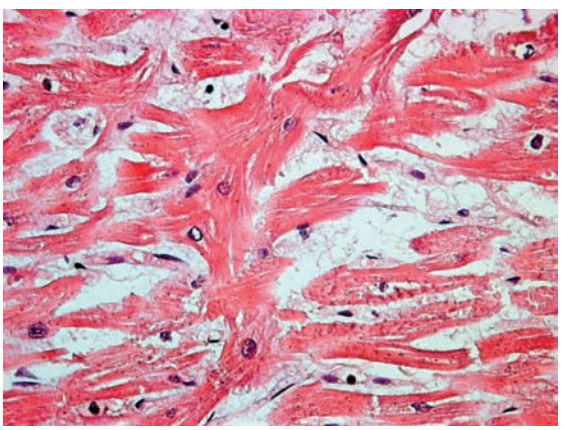


Figure 4 Myocardial disarray with interstitial fibrosis [Reproduced from Ref. 28. © American Medical Association, 1988.]

One type of HCM mutation involves α-tropomyosin and it results in little if any hypertrophy. Nonetheless, it is associated with a high incidence of SUD [40]. The incidence of this mutation is not really known since few medical examiners have the facilities to do DNA screening, leaving the final cause of death undetermined. A prototype chip, capable of detecting mutations in HCM genes, has already been developed [41], and as the technology becomes more widely available the true prevalence of this disorder should become more apparent.

Myocarditis (3%)

The type of inflammatory cell reaction elicited by the causal agent distinguishes the different forms of myocarditis. Unfortunately, the causative agent can only be demonstrated by DNA resequencing, and sometimes not even then. Acquired immune deficiency syndrome (AIDS) and iatrogenic (transplanted heart) immunodeficiency syndrome are conditions in which opportunistic infections occur frequently, and microorganisms can be seen in the myocardium. However, when the latter are not visible, a specific chemotropic cellular response may indirectly suggest the etiology of the underlying disease.

Myocarditis can be *primary* or *secondary:* it may begin within the myocardium or in other organ's systems. The etiology of myocarditis may be due to identified or unidentified causes (the term unidentified is preferable to idiopathic, especially since adequately equipped laboratories are usually capable of identifying the responsible virus) [42]. The area of involvement, either in interstitial or intercellular, can also help distinguish different types of myocarditis. In our experience (AIDS, Chagas', drug hypersensitivity), SUD is most often associated with the presence of extensive lymphocytic myocarditis. The actual cause of cardiac arrest in most instances remains a matter of speculation. For example, we have frequently observed the coexistence of myocarditis and catecholamine necrosis, suggesting that several different mechanisms were operating at the same time, including destruction (or irritation) of the intramyocardial nerves, which alters sympathetic control and induces arrhythmias.

A causal link between sudden death and limited foci of a few inflammatory cells has never

been established [43]. In the rare instances of diffuse myocarditis, cardiac arrest may be due to myocardial insufficiency consequent to massive interstitial exudation (e.g., acute rejection of a transplanted heart) or other mechanisms involving adrenergic nerves resulting in ventricular fibrillation (e.g., sudden death in Chagas' disease in which diffuse lymphocytic myocarditis and catecholamine necrosis coexist) [44].

It has only recently been discovered that patients thought to be suffering from acute coronary syndrome (ACS) may actually be suffering from a form of myocarditis where the virus exhibits an endothelial tropism. ACS is generally diagnosed by angiography, but angiographic studies of patients with florid symptoms are occasionally negative. When endomyocardial biopsies from such patients are analyzed by nested polymerase chain reaction/reverse transcriptase–polymerase chain reaction, viruses, especially Parvo viruses (PVB19 is not related to the agent infecting dogs) are often detected, even in the absence of the histological changes conforming to the requirements of the Dallas criteria [43, 45, 46]. The true extent of this problem is not known for two reasons: (i) no local medical examiner has the training to perform DNA testing and, even if they did, they do not have the equipment and (ii) more often than not the Dallas Criteria are not applied correctly. In a retrospective review of a large case series, the incidence of myocarditis based upon death certificates was found to be slightly less than 5%, with a higher incidence in males than in females. Just how much weight should be placed in these findings is a matter of important debate because histopathological reanalysis of the same listed with myocarditis in the death certificates demonstrated that only 32% of cases actually fulfilled the Dallas criteria, suggesting that the true incidence of myocarditis may only amount to a fraction of official estimates [47].

Sarcoidosis (2%)

Sarcoidosis is a multisystem disorder of unknown etiology. It is characterized by the presence of noncaseating epithelioid cell granulomas. Many infectious and environmental and genetic factors have been implicated, but none definitely identified [48]. The prognosis and clinical manifestations depend on the location and extend of the granulomatous infiltrates. At autopsy, cardiac involvement has been identified in 20–30% of sarcoid victims. When the heart is involved, a wide range of clinical manifestations are possible (conduction disorder, ventricular arrhythmias, atrial arrhythmias, pericarditis, valvular dysfunction, and congestive heart failure) [49]. However, it is unusual for patients with sarcoidosis to present with isolated cardiac involvement. In one autopsy study of sarcoid patients, cardiac involvement proved to be the cause of death in 37% of the patients [50]. In these cases, SUD is the result of ventricular tachyarrhythmia or conduction block in 25–65% of incidents. Consequently, cardiac involvement is associated with a much poorer prognosis, and the mortality rate may exceed 40% at five years and 55% within 10 years. When the disease is limited to the lungs, the survival rate is much higher.

Aortic Dissection (2%)

The new classification of aortic dissections reported by the Task Force on Aortic Dissection and by the European Society of cardiology [51] is basically a subdivision of the original classifications of Stanford and De Bakey [5]. It distinguishes five classes of aortic dissections: class 5 includes traumatic and iatrogenic aortic dissections (Figure 5).

Endocarditis (1%)

Valvular and endocardial diseases are not common causes of SUD. The frequency with which either disorder is diagnosed varies with each pathologist's practice and with the composition of the population examined [52]. Infective endocarditis involves the valvular endocardium and is associated with infected thrombotic vegetations. Any microorganism or bloodborne parasite has the potential to cause infective endocarditis. In practice, Gram-positive cocci cause myocarditis much more often than Gram-negative organisms and bacteria much more often than fungi. Usually, one organism induces the disease, but sometimes more than one is responsible (superinfection). Nosocomial infections may cause infective endocarditis. Generally, the process involves a single, native, left heart valve (usually the mitral), but multivalvular infections of left heart valves or of both right-sided and left-sided can occur under certain conditions.

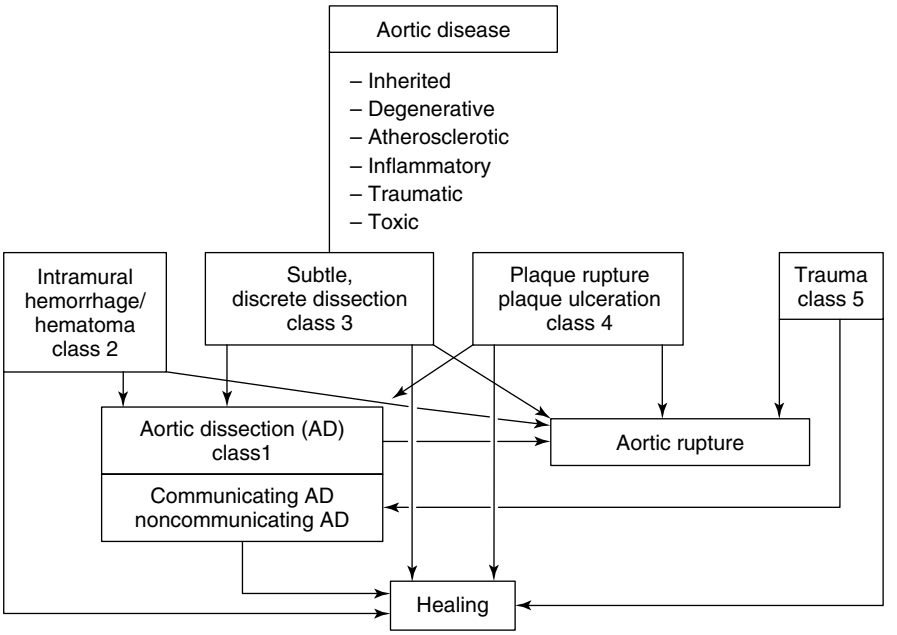


Figure 5 Schematic illustration of different aortic disease aetiologies that can result in aortic dissection including progression or regression of the disease. (Task Force of Aortic Dissection and by the European Society of cardiology [51])

Vegetations attach to the atrial aspect of an atrioventricular leaflet or to the ventricular surface of a semilunar cusp. In both instances, the distribution of the vegetation is related to the lines of leaflet/cusp closure on their flow surfaces. Valvular infection can remain localized to one cusp/leaflet or spread to a contiguous one ("kissing" lesions) or even to the chordae tendineae or papillary muscles. When that occurs, there is a good chance of papillary muscle rupture.

Sometimes, vegetations occur distant from their usual sites. If that is the case, a pathological cause will be usually evident. Microscopically, vegetations are composed of intermingled thrombus, inflammatory cells, and microorganisms; the latter sometimes form colonies. The organisms often, but not always, can be visualized in a Gram or Gomori methanamine silver stain [1].

Both the vegetations and adjacent heart valve display an inflammatory reaction that may be acute or chronic. When an infection has been present for a protracted period, healing changes may be obvious, and portions of vegetations may calcify. Often the valvular pathology found at autopsy provides a plausible explanation for a fatal arrhythmia, for example, the occurrence of myocarditis or myocardial infarction, or annular abscess with damage to the atrioventricular node [53]. In other instances, the cause of death may be impossible to determine.

Floppy Mitral Valve (MVPS) (1%)

In autopsy series the incidence of MVPS ranges from 4 to 5%. Clinical data hint at an incidence of about 2.5%, but even apparently benign cases of MVPS, in young adults, might result in sudden, unexpected death. Such cases are generally not included in hospital-based studies, which may lead to a serious underestimation of the fatal risk associated with the disease [54], and MVPS as a cause of sudden death in the young is almost certainly underdiagnosed.

Gross morphology varies with the etiology of valve prolapse, as will the mechanism of sudden death. When prolapse develops suddenly, for example, following chordal rupture or rupture of the body of an infarcted papillary muscle, then the leaflets themselves may be normal. In these cases, acute pulmonary edema may be the cause of death. In cases of myxomatous degeneration, the prolapsed valve is usually gray-white and thickened, and bulges with

a hooded appearance toward the left atrium. The bulging is best appreciated if the affected valve is viewed from the left atrial aspect before opening the left ventricle.

The endocardium on the surface of the leaflet is usually thickened by fibromuscular tissue, and tiny thrombi may be attached to both leaflet and chordae. Linear thrombi can also be formed at the base of a prolapsed leaflet in the angle between it and the adjacent atrial wall. If the thrombus becomes organized, the angle will become obliterated by fibrous tissue. Cerebral ischemic events prevalent in younger individuals were related to prolapsed mitral valves and such thrombi, but subsequent clinical studies became controversial.

In the late 1990s the first study was published, describing dysplasia of the atrioventricular nodal artery. Careful dissection demonstrated dysplasia of the artery in 18 of 24 hearts with mitral valve prolapse but in only 4 of 16 controls hearts ($p = 0.003$). Since then additional reports have been published and it appears that arterial dysplasia in mitral valve prolapse may contribute to sudden cardiac death mediated by ventricular fibrosis [55].

Coronary Artery Tunneling (Coronary Bridging) (1%)

Myocardial bridging (MB) is a congenital disorder involving a segment of a major epicardial coronary artery that runs intramural (tunneled artery) through the myocardium. The incidence in various reported autopsy series is very wide, ranging from 5.4 to 85.7%. The prevalence also varies substantially in different reports, with a much higher rate of detection at autopsy than with angiography (5–12%). The most frequent site of occurrence is the proximal half of the anterior descending branch of the left coronary artery (60%) [56].

Muscular bridging of right coronary arteries is very rare and more variable both in form and location (2.8–11.4%). The presence of tunneling is usually of no clinical significance, but myocardial ischemia, infarction, ventricular arrhythmia, and sudden death have all been reported, frequently in association with HCM and coronary atherosclerosis.

The clinical significance of MB is controversial and still very much debated. When it occurs, there is systolic compression of the tunneled segment; however, the compression remains clinically silent in the vast majority of cases. Nonetheless, there have been episodic reports of bridging-related episodic angina, tachycardia-induced ischemia, myocardial ischemia, sudden death during strenuous exercise, and malignant arrhythmia.

Angiographically, MB of a coronary artery is demonstrated by rhythmical occlusion of a segment of the artery during every systole that disappears during diastole (milking effect). If patients do become symptomatic, it is probably explained by the fact that during tachycardia diastole shortens more than systole. The literature contains reports of ischemic ECG changes, increased lactate production, and myocardial perfusion defects visible with thallium scintigraphy. An increase in sympathetic drive during stress or exercise facilitates ischemia in these individuals because increased contractility aggravates systolic and diastolic compression [56]. Recently, it has been suggested that exercise increases QT interval duration impairing coronary blood flow to the area perfused by the bridged artery, resulting in ventricular repolarization abnormalities that may increase the risk of sudden death caused by malignant arrhythmias. Repetitive ischemic episodes, even if not symptomatic, may cause patchy myocardial necrosis and fibrosis. The long-term prognosis of individuals with isolated myocardial bridges appears to be excellent, but in some cases they may cause ventricular tachyarrhythmias and SUD.

Right Ventricular Dysplasia (1%)

Arrhythmogenic right ventricular disease (ARVD) is a major cause of sudden death in the young, and it seems to be especially common among athletes. It is a type of cardiomyopathy characterized by the loss of cardiomyocytes with fibrofatty replacement (Figure 6). Three patterns of disease have been identified. In the classic form, there is isolated right ventricular disease with significant functional impairment. Disease may or may not extend to the left ventricle. Alternatively, involvement with the left ventricle may predominate, with early and prominent LV manifestations, such as sudden cardiac death, but relatively mild right-sided disease. In the third form, substantial biventricular involvement is observed. Arrhythmias are associated with all three forms [57, 58].

Both autosomal-dominant and recessive forms exist (the latter is known as *Naxos disease*, where

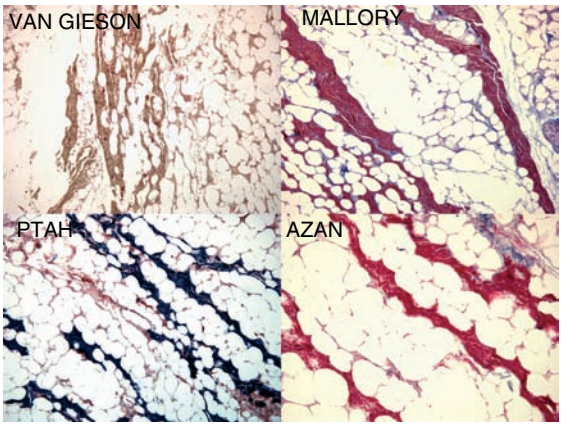


Figure 6 Histological aspect of ARVD: loss of myocytes with fibrofatty replacement (Trichromic stains)

the cardiac changes are accompanied by skin disorders) [59]. The abnormal genes responsible for this condition primarily encode desmosomal proteins, but mutations in the cardiac ryanodine receptor (RYR2) also occur (see above). The abnormal proteins produced are all components of the desmosomal intercellular junction complex responsible for maintaining tissue integrity and. The same proteins also seem to be involved in cell signaling [60].

Rheumatic Mitral Stenosis (1%)

This disorder affects women more than men and, in most instances, only the mitral valve is involved. Only occasionally there is an involvement of both the mitral and aortic valves. Tricuspid involvement is quite uncommon. Gross changes observed at autopsy are the result of organized, minute endocardial vegetations located along the lines of closure. They are formed during acute attacks that induce inflammation in various portions of the valve. This resultant scars distort the valve components. Contributing to the destruction is the deposition and organization of minute thrombi that result from the altered hemodynamics of the distorted valve. Thus, valve damage is cumulative.

The process results in shortened and thickened cusps and leaflets and this causes valvular regurgitation; fusion of the adjacent sides of cusps/leaflets at commissural areas leads to luminal stenosis, the principal change observed in tricuspid stenosis. The result is a diaphragmlike orifice with only minor associated chordal changes; fusion, thickening, or shortening of the chordae tendineae of the mitral valve result in leaflet tethering and valve regurgitation; alternatively, the chordae fuse to cause a mitral subvalvular stenosis.

Thus, a rheumatic valve may be purely stenotic, purely regurgitant, or exhibit both functional abnormalities. The histological features of a distorted valve may vary, but generally, cusps/leaflets are diffusely fibrosed, often with obliteration of their original architecture, and show increased vascularization (small, thick-walled vessels); focal lymphocytic infiltrates are often found, usually near the cusp/leaflet base. Aschoff's nodules, *per se*, are not found in cases that have smouldered for many years, but they are, on rare occasions, observed in the valves of people with active rheumatic or in areas where stenosis developed rapidly.

The secondary morphologic changes most often encountered include valvular, annular commissural, and chordal calcification. These complications are less likely when valvular stenosis developed rapidly. The deposition of ectopic calcium further distorts valve components. If they ulcerate, they may embolize their contents, causing hemolysis, or act as a site of thrombus deposition with subsequent embolization or infection. Occasionally, cardiac fibroelastoma (CPF), otherwise known as "*Lambl's excrescences*", the most common cardiac tumor, may be encountered on a valve distorted by rheumatic fever. Even though the tumor is histologically benign, it can cause life-threatening complications. The tumor is characterized by multiple papillary fronds formed by a central fibroblastic core and is surrounded by a

myxomatous layer and endothelial cells. These may embolize, leading to heart attack and stroke [61].

Anomalous Coronary Artery (0.5%)

Congenital coronary artery anomalies have a statistical incidence of 0.3–0.8% and account for 0.1–2% of all cases of congenital heart disease worldwide. At autopsy anomalous origin of the right coronary artery from the left valsalva sinus is the least common form of the anomaly observed (0.026%). In angiographic studies, the incidence of anomalous origin of right coronary from the left sinus valsalva depends on the population being studied. Although uncommon, this anomaly is frequently associated with SUD [62].

Most coronary anomalies do not result in signs, symptoms, or complications, and usually are discovered as incidental findings at the time of autopsy. Congenital anomalies of the coronary arteries present great difficulties in diagnosis because patients may be absolutely asymptomatic, though rarely the underlying abnormality can manifest itself with syncopal (fainting) episodes or even florid heart failure, which explains why the diagnosis of abnormal coronary artery origin is usually made only at autopsy examination [63].

Ischemia is usually precipitated by strenuous, prolonged effort, and this explains why a basal ECG or even a stress test ECG may be negative. Syncopal episodes are the only prodromal (early, nonspecific) symptoms. Repetitive ischemic episodes may cause patchy myocardial necrosis and fibrosis as well as ventricular hypertrophy, which eventually can elicit arrhythmias because of the malignant combination of acute and chronic substrates. This may explain why sudden death, associated with an anomalous origin of a coronary artery from the wrong sinus, may occur in adults, even though the anomaly has been present since birth.

Anomalous origin of the left coronary artery arising from the pulmonary artery is a rare but serious congenital anomaly, commonly referred to as *Bland–White–Garland syndrome*. It is quite rare and is said to account for approximately 0.25–0.5% of all congenital heart disease. The genetics is neither known nor have any risk factors been identified. Infants with the disease appear normal at birth and then display a typical downhill course. Occasionally, there may be no symptoms until the teenage years, and this is when SUD may occur, especially during exercise. Because the coronary artery is compressed by the pulmonary artery, there may be ischemia and fibrosis, sometimes extensive, is found at autopsy [64]. Fortunately, the diagnosis is easily made with newer imaging techniques.

Coronary Artery Dissection (0.4%)

Spontaneous coronary artery dissection is a rare cause of SUD. Until quite recently, it was thought to be almost exclusively a disease of peripartum period. However, a connection with drug abuse and even exercise is now recognized. More than 50 cases related to stimulant abuse have now been reported. It is speculated that increased arterial blood pressure from cocaine's inotropic and chronotropic effects combined with its direct vasoconstrictive effect leads to increased shear forces on the coronary endothelium [65]. Elevated wall stress may cause an intimal tear and the subsequent dissection of the coronary artery. If the lumen becomes occluded because of the dissection, infarction can result. Presumably the same mechanisms are operative in cases of dissection associated with intense exercise, even in the absence of coronary artery disease [66].

Congenital Heart Disease (0.2%)

In recent decades, because of dramatic improvement in surgical techniques, the types of congenital heart disease associated with SUD have changed significantly. Improved methods of medical control, more aggressive surgery, early surgical correction rather than palliation, implantation of defibrillators, and radiofrequency ablation of arrhythmic foci all have helped to improve survival rates. Severe cardiac malformations are likely to cause early death, and some of these cases will be sudden. Mortality is highest in newborns, especially in the first few days of life, and is due to the ductus arteriosus closing in cases of ductus-dependent malformations, such as ductus hypoplastic left heart syndrome, critical aortic stenosis, complete transposition of great vessels, and aortic coarctation. Mortality and the incidence of sudden death (SD) decrease with survival into early infancy. Cardiac malformations in the newborn are becoming rare in some countries, because diagnoses are now being made in early fetal life. Sometimes termination of the pregnancy follows these diagnoses. Most newborns with malformations diagnosed during

pregnancy are treated at birth and often receive early surgical palliation or corrective surgery. It should be remembered, however, that deaths in congenital heart disease subjects at any age are far more likely to be expected and protracted rather than sudden.

Lipomatous Hypertrophy of the Interatrial Septum (0.2%)

This is another rare disease characterized by the excessive deposition of fatty tissue within the interatrial septum. The reported prevalence in autopsy reports is approximately 1%, although echocardiographic studies of living patients suggest a much higher incidence. It is generally considered a disorder of the elderly and the obese; however, the prevalence of arrhythmias in these patients has been reported to be as high as 40%, and there are rare reports of SUD. Histologic examination shows that this fatty deposition is composed of mature fat with varying quantities of fetal (brown) fat, inflammation, fibrosis, and entrapment of myocardial fibers. The fatty accumulation often assumes a dumbbell shape, easily recognized in positron emission tomography (PET) scans of the living [67].

References

[1] Fineschi, V., Baroldi, G. & Silver, D.M. (2006). *Pathology of the Heart and Sudden death in Forensic Medicine*, CRC Press, Boca Raton, FL.

[2] Stevenson, W.G. & Soejima, K. (2007). Catheter ablation for ventricular tachycardia, *Circulation* **115**(21), 2750–2760.

[3] Barodi, G. & Silver, M.D. (2004). *The etiopathogenesis of coronary heart disease: a heretical theory based on morphology*, 2nd ed, Landes Bioscience, Georgetown, Texas.

[4] Rosamond, W., Flegal, K., Go, A., Greenlund, K., Haase, N., Ho, M., Howard, V., Kissela, B., Kittner, S., Lloyd-Jones, D., McDermott, M., Meigs, J., Moy, C., Nichol, G., O'Donnell, C.J., Roger, V., Rumsfeld, J., Sorlie, P., Steinberger, J., Thom, T. Wasserthiel-Smoller, S. & Hong, Y. (2007). American Heart Association statistical update: heart disease and stroke statistics – 2007 update, *Circulation* **115**, e89–e94.

[5] Virmani, R., Burke, A.P. & Farb, A. (2001). Sudden cardiac death, *Cardiovascular Pathology* **10**(6), 275–282.

[6] Ladich, E., Virmani, R. & Burke, A. (2006). Sudden cardiac death not related to coronary atherosclerosis, *Toxicologic Pathology* **34**(1), 52–57.

[7] Karch, S.B., Green, G.S. & Young, S. (1995). Myocardial hypertrophy and coronary artery disease in male cocaine users, *Journal of Forensic Sciences* **40**(4), 591–595.

[8] Asakura, M., Ueda, Y., Yamaguchi, O., Adachi, T., Hirayama, A., Hori, M. & Kodama, K. (2001). Extensive development of vulnerable plaques as a pan-coronary process in patients with myocardial infarction: an angioscopic study, *Journal of the American College of Cardiology* **37**(5), 1284–1288.

[9] Al-Khatib, S.M., LaPointe, N.M., Kramer, J.M. & Califf, R.M. (2003). What clinicians should know about the QT interval, *JAMA: The Journal of the American Medical Association* **289**(16), 2120–2127.

[10] Roden, D.M. (2006). Long QT syndrome: reduced repolarization reserve and the genetic link, *Journal of Internal Medicine* **259**(1), 59–69.

[11] Noda, T., Shimizu, W., Taguchi, A., Aiba, T., Satomi, K., Suyama, K., Kurita, T., Aihara, N., & Kamakura, S. (2005). Malignant entity of idiopathic ventricular fibrillation and polymorphic ventricular tachycardia initiated by premature extrasystoles originating from the right ventricular outflow tract, *Journal of the American College of Cardiology* **46**(7), 1288–1294.

[12] Fernandez, P., Moolman-Smook, J., Brink, P. & Corfield, V. (2005). A gene locus for progressive familial heart block type II (PFHBII) maps to chromosome 1q32.2-q32.3, *Human Genetics* **118**(1), 133–137.

[13] Cerrone, M., Noujaim, S. & Jalife, J. (2006). The short QT syndrome as a paradigm to understand the role of potassium channels in ventricular fibrillation, *Journal of Internal Medicine* **259**(1), 24–38.

[14] Vincent, G.M. (1998). The molecular genetics of the long QT syndrome: genes causing fainting and sudden death, *Annual Review of Medicine* **49**, 263–274.

[15] Mohler, P.J., Scouarnec, S.Le., Denjoy, I., Lowe, J.S., Guicheney, P., Caron, L., Driskell, I.M., Schott, J.J., Norris, K., Leenhardt, A., Kim, R.B., Escande, D. & Roden, D.M. (2007). Defining the cellular phenotype of "ankyrin-B syndrome" variants: human ANK2 variants associated with clinical phenotypes display a spectrum of activities in cardiomyocytes, *Circulation* **115**(4), 432–441.

[16] Napolitano, C., Priori, S.G., Schwartz, P.J., Bloise, R., Ronchetti, E, Nastoli, J., Bottelli, G., Cerrone, M. & Leonardi, S. (2005). Genetic testing in the long QT syndrome: development and validation of an efficient approach to genotyping in clinical practice, *JAMA : The Journal of the American Medical Association* **294**(23), 2975–2980.

[17] Sanguinetti, M.C. & Tristani-Firouzi, M. (2006). hERG potassium channels and cardiac arrhythmia, *Nature* **440**(7083), 463–469.

[18] Nademanee K., Veerakul, G., Nimmannit, S., Chaowakul, V., Bhuripanyo, K., Likittanasombat, K., Tunsanga, K., Kuasirikul, S., Malasit, P., Tansupasawadikul, S. & Tatsanavivat, P (1997). Arrhythmogenic marker for the sudden unexplained death syndrome in Thai men, *Circulation* **96**, 2595–2600.

[19] Frustaci, A., Priori, S.G., Pieroni, M., Chimenti, C., Napolitano, C., Rivolta, I., Sanna, T., Bellocci, F. & Russo, M.A. (2005). Cardiac histological substrate in patients with clinical phenotype of Brugada syndrome, *Circulation* **112**(24), 3680–3687.

[20] George, C.H., Jundi, H., Thomas, N.L., Fry, D.L., & Lai, F.A. (2007). Ryanodine receptors and ventricular arrhythmias: emerging trends in mutations, mechanisms and therapies, *Journal of Molecular and Cellular Cardiology* **42**(1), 34–50.

[21] Chimenti, C., Pieroni, M., Maseri, A., & Frustaci, A. (2004). Histologic findings in patients with clinical and instrumental diagnosis of sporadic arrhytmogenic right ventricular dysplasia, *Journal of the American College of Cardiology* **43**(12), 2305–2313.

[22] Tester, D.J. & Ackerman, M.J. (2007). Postmortem long QT syndrome genetic testing for sudden unexplained death in the young, *Journal of the American College of Cardiology* **49**(2), 240–246.

[23] Wachtell, K., Okin, P.M., Olsen, M.H., Dahlöf, B., Devereux, R.B., Ibsen, H., Kjeldsen, S.E., Lindholm, L.H., Nieminen, M.S. & Thygesen, K. (2007). Regression of electrocardiographic left ventricular hypertrophy during antihypertensive therapy and reduction in sudden cardiac death: the LIFE Study, *Circulation* **116**(7), 700–705.

[24] Kahan, T. & Bergfeldt, L. (2005). Left ventricular hypertrophy in hypertension: its arrhythmogenic potential, *Heart* **91**(2), 250–256.

[25] Kitzman, D.W., Scholz, D.G., Hagen, P.T., Ilstrup, D.M. & Edwards, W.D. (1988). Age-related changes in normal human hearts during the first 10 decades of life. Part II (Maturity): A quantitative anatomic study of 765 specimens from subjects 20 to 99 years old, *Mayo Clinic Proceedings* **63**(2), 137–146.

[26] Abbate, A., Bussani, R., Amin, M.S., Vetrovec, G.W. & Baldi, A. (2006). Acute myocardial infarction and heart failure: role of apoptosis, *The International Journal of Biochemistry and Cell Biology* **38**(11), 1834–1840.

[27] Lafontant, P.J. & Field, L.J. (2006). The cardiomyocyte cell cycle, *Novartis Foundation Symposium* **274**, 196–207; discussion 208–13, 272–6.

[28] Karch, S.B. & Billingham, M.E. (1988). The pathology and etiology of cocaine-induced heart disease, *Archives of Pathology and Laboratory Medicine* **112**(3), 225–230.

[29] Henning, R.J. & Cuevas, J. (2006). Cocaine activates calcium/calmodulin kinase II and causes cardiomyocyte hypertrophy, *Journal of Cardiovascular Pharmacology* **48**(1), 802–813.

[30] Henning, R.J., Silva, J., Reddy, V., Kamat, S., Morgan, M.B., Li, Y.X. & Chiou, S. (2000). Cocaine increases beta-myosin heavy-chain protein expression in cardiac myocytes, *Journal of Cardiovascular Pharmacology and Therapeutics* **5**(4), 313–322.

[31] Swynghedauw, B. (1999). Molecular mechanisms of myocardial remodeling, *Physiological Reviews* **79**(1), 215–262.

[32] Furukawa, T. & Kurokawa, J. (2006). Potassium channel remodeling in cardiac hypertrophy, *Journal of Molecular and Cellular Cardiology* **41**(5), 753–761.

[33] Ellinor, P.T., Sasse-Klaassen, S., Probst, S., Gerull, B., Shin, J.T., Toeppel, A., Heuser, A., Michely, B., Yoerger, D.M., Song, B.S., Pilz, B., Krings, G., Coplin, B., Lange, P.E., Dec, G.W., Hennies, H.C., Thierfelder, L. & MacRae, C.A. (2006). A novel locus for dilated cardiomyopathy, diffuse myocardial fibrosis, and sudden death on chromosome 10q25-26, *Journal of the American College of Cardiology* **48**(1), 106–111.

[34] Gavazzi, A., De Maria, R., Renosto, G., Moro, A., Borgia, M., Caroli, A., Castelli, G., Ciaccheri, M., Pavan, D., De Vita, C., *et al.* (1993). The spectrum of left ventricular size in dilated cardiomyopathy: clinical correlates and prognostic implications. SPIC (Italian Multicenter Cardiomyopathy Study) Group, *American Heart Journal*, **125**(2 Pt 1), 410–422.

[35] Brito, D., Richard, P., Isnard, R., Pipa, J., Komajda, M. & Madeira, H. (2003). Familial hypertrophic cardiomyopathy: the same mutation, different prognosis. Comparison of two families with a long follow-up, *Revista Portuguesa de Cardiologia* **22**(12), 1445–1461.

[36] Osio, A., Tan, L., Chen, S.N., Lombardi, R., Nagueh, S.F., Shete, S., Roberts, R., Willerson, J.T. & Marian, A.J. (2007). Myozenin 2 is a novel gene for human hypertrophic cardiomyopathy, *Circulation Research* **100**(6), 766–768.

[37] Hughes, S.E. & McKenna, W.J. (2005). New insights into the pathology of inherited cardiomyopathy, *Heart* **91**(2), 257–264.

[38] Fineschi, V., Silver, M.D., Karch, S.B., Parolini, M., Turillazzi, E., Pomara, C. & Baroldi, G. (2005). Myocardial disarray: an architectural disorganization linked with adrenergic stress? *International Journal of Cardiology* **99**(2), 277–282.

[39] Nomura, M., Nakaya, Y., Kishi, F., Kondo, Y., Yukinaka, M., Saito, K. & Ito, S. (2007). *Clinical Application of QT Dispersion on Ventricular Hypertrophy using Magnetocardiogram*.

[40] Watkins, H., McKenna, W.J., Thierfelder, L., Suk, H.J., Anan, R., O'Donoghue, A., Spirito, P., Matsumori, A., Moravec, C.S. & Seidman, J.G., *et al.* (1995). Mutations in the genes for cardiac troponin T and alpha-tropomyosin in hypertrophic cardiomyopathy, *The New England Journal of Medicine* **332**(16), 1058–1064.

[41] Fokstuen, S., Lyle, R., Munoz, A., Gehrig, C., Lerch, R., Perrot, A., Osterziel, K.J., Geier, C., Beghetti, M., Mach, F., Sztajzel, J., Sigwart, U., Antonarakis, S.E. & Blouin, J.L. (2008). A DNA resequencing array for pathogenic mutation detection in hypertrophic cardiomyopathy, *Human Mutation* **29**(6), 879–885.

[42] Cioc, A.M. & Nuovo, G.J. (2002). Histologic and in situ viral findings in the myocardium in cases of sudden, unexpected death, *Modern Pathology* **15**(9), 914–922.

[43] Pomara, C., Villani, A., D'Errico, S., Riezzo, I., Turillazzi, E. & Fineschi, V. (2006). Acute myocarditis

mimicking acute myocardial infarction: a clinical nightmare with forensic implications, *International Journal of Cardiology* **112**(1), 119–121.

[44] Baroldi, G., Oliveira, S.J. & Silver, M.D. (1997). Sudden and unexpected death in clinically "silent" Chagas' disease. A hypothesis, *International Journal of Cardiology* **58**(3), 263–268.

[45] Kühl, U., Pauschinger, M. Bock, T., Klingel, K., Schwimmbeck, C.P., Seeberg, B., Krautwurm, L., Poller, W., Schultheiss, H.P. & Kandolf, R. (2003). Parvovirus B19 infection mimicking acute myocardial infarction, *Circulation* **108**(8), 945–950.

[46] Fineschi, V., Neri, M., Riezzo, I. & Turillazzi, E. (2004). Sudden cardiac death due to hypersensitivity myocarditis during clozapine treatment, *International Journal of Legal Medicine* **118**(5), 307–309.

[47] Kytö, V., Saraste, A., Voipio-Pulkki, L.M. & Saukko, P. (2007). Incidence of fatal myocarditis: a population-based study in Finland, *American Journal of Epidemiology* **165**(5), 570–574.

[48] Doughan, A.R. & Williams, B.R. (2006). Cardiac sarcoidosis, *Heart* **92**(2), 282–288.

[49] Mori, M., Hanon, S. & Rachko, M. (2007). Case report: cardiac sarcoidosis presenting with ventricular arrhythmias: case report and review of the literature, *International Journal of Cardiology* **120**(2), e21–e23.

[50] Perry, A. & Vuitch, F. (1995). Causes of death in patients with sarcoidosis. A morphologic study of 38 autopsies with clinicopathologic correlations, *Archives of Pathology and Laboratory Medicine* **119**(2), 167–172.

[51] Erbel, R., Alfonso, F., Boileau, C., Dirsch, O., Eber, B., Haverich, A., Rakowski, H., Struyven, J., Radegran, K., Sechtem, U., Taylor, J., Zollikofer, C., Klein, W.W., Mulder, B. & Providencia, L.A. (2001). Task Force on Aortic Dissection, European Society of Cardiology. Diagnosis and management of aortic dissection, *European Heart Journal* **22**(18), 1642–1681.

[52] Tleyjeh, I.M., Abdel-Latif, A., Rahbi, H., Scott, C.G., Bailey, K.R., Steckelberg, J.M., Wilson, W.R. & Baddour, L.M. (2007). A systematic review of population-based studies of infective endocarditis, *Chest* **132**(3), 1025–1035.

[53] Bashore, T.M., Cabell, C. & Flower, V. Jr (2006). Update on infective endocarditis, *Current Problems in Cardiology* **31**(4), 274–352.

[54] Anders, S., Said, S., Schulz, F. & Püschel, K. (2006). Mitral valve prolapse syndrome as cause of sudden death in young adults, *Forensic Science International* **171**(2–3), 127–130.

[55] Burke, A.P., Farb, A., Tang, A., Smialek, J. & Virmani, R. (1997). Fibromuscular dysplasia of small coronary arteries and fibrosis in the basilar ventricular septum in mitral valve prolapse, *American Heart Journal* **134**(2 Pt 1), 282–291.

[56] Riezzo, I., Monciotti, F., Pomara, C. & Fineschi, V. (2007). Myocardial bridging of the right coronary artery and emotional stress: a fatal link? *International Journal of Cardiology* **115**(3), e99–101.

[57] Kiès, P., Bootsma, M., Bax, J. & Schalij, M.J. & van der Wall, E.E. (2006). Arrhythmogenic right ventricular dysplasia/cardiomyopathy: screening, diagnosis, and treatment, *Heart Rhythm* **3**(2), 225–234.

[58] Moric-Janiszewska, E. & Markiewicz-Loskot, G. (2007). Review on the genetics of arrhythmogenic right ventricular dysplasia, *Europace* **9**(5), 259–266.

[59] Protonotarios, N. & Tsatsopoulou, A. (2005). Naxos disease, *Indian Pacing and Electrophysiology Journal* **5**(2), 76–80.

[60] Lahtinen, A.M., Lehtonen, A., Kaartinen, M., Toivonen, L., Swan, H., Widén, E., Lehtonen, E., Lehto, V.P. & Kontula, K. (2007). Plakophilin-2 missense mutations in arrhythmogenic right ventricular cardiomyopathy, *International Journal of Cardiology* **126**(1), 92–100.

[61] Fumero, A., Villalba, L., Kemeny, J.L. & Camilleri, L. (2003). Coexisting chordal papillary fibroelastoma and leaflet Lambl's tumour of the mitral valve, *Interactive Cardiovascular and Thoracic Surgery* **2**(3), 382–384.

[62] Eckart, R.E., Jones, S.O. 4th, Shry, E.A., Garrett, P.D. & Scoville, S.L. (2006). Sudden death associated with anomalous coronary origin and obstructive coronary disease in the young, *Cardiology in Review* **14**(4), 161–163.

[63] Earls, J.P. (2006). Coronary artery anomalies, *Techniques in Vascular and Interventional Radiology* **9**(4), 210–217.

[64] Gumina, R.J., Callahan, M.J. & Murphy, J.G. (2007). Bland-White-Garland syndrome: not just a pediatric coronary anomaly, *The Journal of Invasive Cardiology* **19**(1), E9.

[65] Steinhauer, J.R. & Caulfield, J.B. (2001). Spontaneous coronary artery dissection associated with cocaine use: a case report and brief review, *Cardiovascular Pathology* **10**(3), 141–145.

[66] Kalaga, R.V., Malik, A. & Thompson, P.D. (2007). Exercise-related spontaneous coronary artery dissection: case report and literature review, *Medicine and Science in Sports and Exercise* **39**(8), 1218–1220.

[67] Kuester, L.B., Fischma, A.J., Fan, C.M., Halpern, E.F. & Aquino, S.L. (2005). Tomography/CT18F fluorodeoxyglucose positron emission septum: prevalence and features on fusion lipomatous hypertrophy of the interatrial, *Chest* **128**, 3888–3893.

Related Articles

Autopsy

Postmortem Toxicology: Interpretation

Seizures: Behavioral

Toxicology: Initial Testing

VITTORIO FINESCHI AND IRENE RIEZZO

Cartridge: Identification *see* Firearms: Bullet and Cartridge Case Identification

Case Assessment and Interpretation

Introduction

This article describes a fundamental piece of work that is transforming the practice of forensic science and its application to criminal justice. This relatively new approach, described as case assessment and interpretation (CAI), has had a significant impact upon the forensic science profession, challenging the basis of expert opinion given in past times. It has also led to changes in the application of forensic science beyond the profession and into the working environment of police forces, prosecuting authorities, and the courts. However, some of the changes in thinking and practice implied by CAI have not yet been implemented fully. Much more needs to be done, including changes to the way in which the adversarial system integrates scientific evidence into the process of criminal justice, to maximize the utilization of forensic science through effective application of CAI.

The development of CAI was driven not only by serious questions raised about the quality of expert opinion but also by the growing requirement to manage limited forensic science resources in the most appropriate and cost-effective way. It has led to further thinking on how forensic science could continue to develop in the future.

Quality of Expert Opinion

Over the past 30 years, there has been a history of dubious or misleading expert opinion being adduced in courts of law. Setting to one side the issue of the technical competence of experts, a key cause of unsafe opinions has been the unsatisfactory way in which experts have arrived at, and communicated, opinions. It is our view that an understanding of the nature of expert opinion is necessary for judges and juries to make informed decisions based on the evidence given by experts. We also suggest that an understanding of the nature of expert opinion will only be achieved when all parties in the criminal justice system agree on the nature of the role of the expert. By "role", we do not mean what the expert should do in terms of examination but rather how expert opinion fits with, and is incorporated into, the criminal justice process.

There may well be differing views on the expert's role and, in the English and Welsh jurisdiction, only limited guidance has been given by the courts on that role. The Appeal Court of England and Wales [1] considered two cases in which experts had obtained DNA profiles from samples of semen recovered from the complainant in each of the two cases. The experts reported matches with the defendants in each of the two cases and, in answer to questions in court, had given a view on the likely origin of samples of semen. The Appeal Court judgment made it clear that, in cases where the issue is the origin of a questioned sample of semen, the forensic scientist should not give an answer to a question of the form "How likely is it that the semen originated from Mr X?". That question was solely for the jury to try to answer as, unlike the expert, they would be in possession of all the available evidence, not just the scientific evidence of a DNA match. The role of a scientist who was called to give expert evidence in court was to help the jury to reach that decision. In principle, the guidance from the Appeal Court can be applied not just to the issue of the source of a recovered sample of DNA but to any scientific evidence that addresses the source of any recovered material. Instances of dubious expert evidence being adduced in court include many examples where this principle has been overlooked and has lead to questionable opinion being given. The application to other forensic disciplines is desirable and, in a recent ruling in the case against Barry George, their lordships explicitly mention CAI as a fundamental tool to evaluate the contribution of forensic evidence in the context of the case [2].

Cost-Effective Forensic Science

It is probably a safe assumption to say that the vast majority of law enforcement agencies worldwide

have limits on their financial resources and on the number of personnel available for crime investigation and criminal trial proceedings. It is also probably the case that, in most countries, police and court services are centrally funded from the public purse and are required to operate in efficient, effective, and economic ways. As forensic science services are usually provided as a subsidiary of the police, local authority or court service, similar requirements will be placed on them. In contrast, England and Wales have moved to a commercialized forensic science sector where the police, prosecution, and defense pay directly for forensic science services. Whatever the arrangements are, law enforcement agencies will almost certainly have legal, political, and social requirements to investigate crime and to bring prosecutions. A major issue for these agencies and their suppliers is how to balance the requirement to provide "value for money" while, at the same time, finding and bringing criminals to justice and exonerating the innocent. Furthermore, whatever be the form of customer–supplier relationship, decisions have to be made, both strategically and within individual cases, on the course of action that should be taken. In forensic science, these decisions are made routinely by scientists, police, and lawyers when they are considering, for example, which items should be examined in individual cases. How these professionals make such decisions and whether they have any formal guidance to do so are key considerations.

The "CAI" model is intended to guide decision-making and to provide a framework for safe, robust opinions. While the guidance was developed in the English and Welsh jurisdiction, we believe the basic principles are just as appropriate to any other jurisdiction.

History

The Development of Interpretation

The manner in which forensic scientists interpreted their findings was largely unchanged for most of the twentieth century, primarily relying on personal experience and personal opinion. In the sense that it was usually unclear to anyone else how the expert arrived at an opinion, such opinion could be hard to understand and to challenge. In addition to this lack of transparency, there was, and still is, a fundamental issue with the way in which forensic science was perceived and used by all parties in the Criminal Justice System.

It has been recognized for some time that forensic science, while having a key role to play in many cases, does not fit well with the adversarial system of justice in the United Kingdom.

This lack of fit can be explained in a number of ways but the following two differences are perhaps the most fundamental and most revealing:

1. Scientists, through their professional training, seek the truth impartially and objectively. In contrast, the adversarial system is more a "contest" during which the side that puts forward the more persuasive argument usually wins. It is not uncommon in a trial involving evidence of opinion for each side to bring more than one forensic expert to support their proposition. Although recent guidelines in the English/Welsh criminal jurisdiction have allowed for pretrial resolution of differences between experts, in practice this does not always take place. Scientists are commissioned by either prosecution or defense and are given a specific issue to address. Often the scientist working under instructions from the prosecution is unsighted of the defense until they, the prosecution-commissioned scientist, are giving their evidence. Occasionally, they may only become aware of the defense after their court appearance. Making a balanced assessment of the scientific findings in these situations is difficult if not impossible – the forensic scientist can address either only the prosecution proposition or an assumed, possibly inappropriate, alternative on behalf of the defense.
2. Scientists are comfortable with uncertainty and with thinking in probabilistic terms. In contrast, the prosecution seek certainty in order to prove their case "beyond reasonable doubt" and may seek to present forensic science as clear and incontrovertible evidence. The jury also may have difficulty in understanding scientific evidence that provides anything less than certainty. The expectations of forensic science, heightened by extensive media coverage, can, and has in the past, lead to pressure being put on the scientist to overstate his or her evidence.

Overstating the value of evidence has been a key issue in a number of miscarriages of justice in which forensic evidence has been discredited. High-profile

examples in the United Kingdom include R v Preece, R v Ward, R v Power, Hill and others, and R v Clark. It could be argued that in these cases, the expert was either biased, whether knowingly or not, or incompetent. Bias is perhaps difficult to avoid (*see* **Interpretation: Observer Effects**), particularly when trying to operate in a system not designed to determine the truth and in which, in an adversarial jurisdiction, experts are commissioned by one side or the other. In comparison, competence can be defined as having attained a particular level of technical ability in the analytical methods used. As such, technical competence can be, and is, tested to assure quality of results. However, we suggest competence requires more than just a certain level of technical ability. Competence in forensic science requires, in addition to technical skills and knowledge, a particular level of interpretative skill. We would argue that application of a robust interpretational framework would help to improve competence and to counteract bias. The question is, do we, as an international forensic science community, have such a framework?

Beginning in the 1980s, a small number of scientists within the forensic science profession began arguing that a more robust, more transparent way of interpreting findings was provided by a likelihood ratio (LR) approach within a Bayesian paradigm (reference [3] was, perhaps, the first example in the forensic science literature). The logical approach provided by the LR took many years to gain understanding and acceptance by the general forensic science community but is now, arguably, close to being accepted as "main-stream" practice. We return to this in the section Underpinning Philosophy.

The Development of a Competitive Market Place for Forensic Science

During the 1980s and 1990s, reviews of various services provided by UK Government Departments had identified improvements to help increase efficiency, effectiveness, and economy. A main political theme of the time was the introduction of practices and experiences from the private sector that were perceived as offering more business-like ways of working. The provision of forensic science services was not excluded from this review and various reports all dealt with ways of improving the delivery of forensic science [4].

Emerging from one such review was a proposal for direct payment by the police and prosecution for the services provided by forensic science organizations. The aim of the new arrangement was to achieve a more direct relationship between the ever-increasing demand for forensic science services and the resources available to deliver forensic science. If "customers" believed they really needed more forensic science services, and these services were seen as worth paying for, then forensic science providers could increase their level of resources, in terms of personnel, equipment, and range of services. On the other hand, if customers decided that the service was not worthwhile, the resources of forensic science should decrease.

A corollary of direct charging was that it gave police forces freedom to purchase forensic science from other providers in the private sector rather than from a central government agency.

The Forensic Science market place in England and Wales was thus born.

"Value for Money" and Bayes Theorem

The introduction of direct charging provided a stimulus to examine further how scientists operated within cases, particularly the way in which scientists made decisions on which items to examine and which tests to employ (Case Assessment), and the way in which they interpreted their findings (Interpretation). It was an opportunity to explore how scientists thought about a case and to understand and capture, in a formal way, how the more effective, more efficient scientists operated in practice. From this work, a model was developed to provide a framework intended to guide scientists in making good decisions [5–8]. The model was based on the logic of Bayesian thinking, particularly focusing on evaluation of LRs, and was developed and refined through workshops with practitioners across all main-stream disciplines of forensic science. Meetings were also held with a sample of police and prosecution personnel to explore implications in the criminal justice system. We return to this topic in the section The CAI Model.

Underpinning Philosophy

Dealing with Uncertainty – Bayes Theorem

How do we make good decisions when faced with a series of options? How do we choose which course of action will give best benefit?

We make these sorts of decision every day of our lives. What we eventually decide to do in a particular situation will depend on many factors including our experience, our instincts, our habits, our finances, the available information, and the constraints of rules, regulations and laws. However, underpinning all of this is the influence of the way in which we think – our style of thinking.

For many everyday situations, most of us are probably guided by our own instinctive feeling on which will be the best course of action – we do not think too long or too hard about the problem. For slightly more difficult problems, we check out our thinking with a trusted friend or colleague, or we carry out some research to obtain further information. For particularly difficult or profound problems, we may well carry out a long analysis of the situation and think through the pros and cons of the various options: we will try to take a rational decision based on data, knowledge, and understanding.

In the forensic field, it can be argued that scientists have tended to make decisions in an instinctive, heuristic way – "This course of action has worked in the past for me and therefore I will apply it in this particular case". For some, this can be an effective way of working, particularly if the scientist is good at analyzing their experiences, at learning from those experiences, and at being able to recall, at appropriate times, the lessons learned. However, we have all seen scientists whose instincts are not the best – their proposed course of action would lead to inefficient, ineffective, or even incorrect outcomes. In fact, we have all, at some time, probably experienced this situation ourselves! But even with the best of instincts, sometimes instinctive judgments let us down – better decisions could have been taken if we had had more background information and a clearer understanding of the requirements and likely outcomes.

The elegance of Bayes theorem is, for some, its simplicity in showing how new information can be incorporated in a logical, robust manner to help change a personal view on the probability that an uncertain event actually occurred. There are numerous works that explain Bayes theorem in detail – we refer readers to a relevant article (**Evidence Interpretation: a Logical Approach**) in this encyclopaedia and to other references that are written in a forensic context [9–11]. We do not intend to reiterate here an explanation of Bayes theorem, but we reproduce some formulae simply to help the reader understand this article. We base our notation on that provided in reference [12].

A classic example where Bayesian thinking can be applied to good effect is in the field of medical diagnoses. Consider the situation in which a person who has just been told he had tested positive for HIV. What is the probability that he actually has the disease, given he has a positive test result? For this example, we use the following notation:

Let H_1 be the proposition that he has the disease.
Let H_2 be the proposition that he does not have the disease.

Note that these two propositions are mutually exclusive (they both cannot be true at the same time) and that they are exhaustive (there are no other propositions to be considered).

Let R be the positive test result.
Let I be any other background information that may have a bearing on the probabilities.

One other symbol is the vertical conditioning bar | which can be read as "given". The probability assignment will be based, or conditioned, on some assumption(s) or piece (s) of information.

We will use what is known as the *odds form of Bayes theorem* to evaluate the probability that the person has HIV. Bayes theorem can be written in various forms but the odds form is most relevant to our medical example and to judicial proceedings. The odds form can be written as follows (formula 1.16 in [12]):

$$\frac{\Pr(H_1|R,I)}{\Pr(H_2|R,I)} = \frac{\Pr(R|H_1,I)}{\Pr(R|H_2,I)} \times \frac{\Pr(H_1|I)}{\Pr(H_2|I)} \qquad (1)$$

It is called the *odds form* because we are dealing with the ratio of two probabilities – the probabilities of each of two mutually exclusive, exhaustive propositions, H_1 and H_2. The ratio of two such probabilities is known as *odds*.

The left-hand side of the equation is the odds that he has the disease, conditional on the test result. It is the result of the multiplication of two expressions. The first of these two, $\Pr(H_1|I)/\Pr(H_2|I)$, is again a ratio of two probabilities but, contrary to the left-hand side of the formula, the probabilities are not conditional on the test result – they are derived from information obtained before the result of the test is known.

They are therefore known as the *prior probabilities*. The second expression, $\Pr(R|H_1, I)/\Pr(R|H_2, I)$, is known as the *likelihood ratio* (*LR*) and is the ratio of two probabilities:

1. the probability that the test would be positive, given that *he has* the disease (H_1) and
2. the probability that the test would be positive, given that *he does not* have the disease (H_2).

It is called a *likelihood ratio* because, simply, it is the ratio of two likelihoods (or probabilities). It can be seen as a measure of the weight, or impact, of the test result in helping to decide whether or not he has the disease.

Considering each of the terms, what data would inform the probabilities? The prior probabilities would be informed by the relative frequency of occurrence of the disease in the population of which the man is a member. The numerator of the LR would be informed by knowledge of the sensitivity of the test – the number of times a positive test result would be seen in samples from people who are known to have HIV. The denominator would be informed by knowledge of the specificity of the test – the number of times a positive result would be seen in samples from people who are known not to have the disease.

If robust estimates of the prior probabilities and the two probabilities of the LR are not available then a false inference could be made from a positive test result.

Bayes Theorem is eminently suited to problems such as the diagnosis of disease but we are more concerned with judicial proceedings. Instead of H_1 and H_2 to represent our propositions, we use H_p and H_d to represent the propositions of the prosecution and defense respectively. Instead of R, we use E to represent the test result. Formula 1 then becomes

$$\frac{\Pr(H_p|E, I)}{\Pr(H_d|E, I)} = \frac{\Pr(E|H_p, I)}{\Pr(E|H_d, I)} \times \frac{\Pr(H_p|I)}{\Pr(H_d|I)}$$

We referred earlier to the role of the forensic scientist in court. The odds form of Bayes theorem helps to clarify the scientist's and the jury's (in some jurisdictions, the judge's) role. The jury/judge concern themselves with the *prior* and *posterior probabilities*; the scientist concerns herself with an evaluation and communication of the *LR*, $\Pr(E|H_p, I)/\Pr(E|H_d, I)$. That role for the scientist has been described elsewhere as that of an "*evaluator*" [13]. This view of the respective roles fits well with the guidance offered by the Appeal Court in the Barry George case [2].

A major concern among some scientists, however, is the feeling that you cannot apply Bayes unless you have hard data to inform *likelihoods* (the probabilities for the scientific findings – the test results and observations). A discussion of this issue is beyond the scope of this article and we refer readers to one reference for further discussion [14]. Suffice it to say here that we believe it is helpful to consider likelihoods not as a precise number but more as a reflection of the scientist's belief that she would obtain a particular test result or observation given the truth of each of the competing propositions. This belief would be informed not only by whatever "hard" data there were but also by the scientist's own knowledge and experience. In that sense, likelihoods are subjective and are open to test and challenge.

Full Form of Bayes Theorem. We have seen how the odds form of Bayes theorem can be applied in judicial proceedings but, in investigations prior to a suspect being apprehended, there will be no propositions put forward by prosecution or defense. Indeed, the police investigator may have no idea of what has happened. Recent work has explored this distinction between investigative and judicial situations [13] and other work has demonstrated how Bayesian models may be applied to investigations [15–16]. What is clear is that the odds form of Bayes is not suited to investigative problems but that the fuller form has much potential. In "*investigative*" mode, the scientist's role may well be to provide possible explanations (propositions) for findings and, if feasible, posterior probabilities for these explanations. The scientist's role in investigations, in contrast to the role in a judicial context, may well include consideration of *prior probabilities* for explanations as well as *likelihoods* for the findings given the truth of the explanations.

The Nature of Opinion

Published work mentioned earlier on the nature of opinions [13] attempts not only to categorize opinions according to Bayes theorem, but also categorizes opinions in terms of a "hierarchy of issues" that a scientist may address. The scientist is encouraged to decide in advance, and in consultation with the clients, the level of issue to be addressed. The

scientist formulates the specific issues within the case, beginning with "offense" level, proceeding down through "activity" level and "source" level to "subsource" level. There are very few instances, if any, in which the scientist, as an evaluator, should address the "offense" level and experience has shown that, while scientists may feel more comfortable with "subsource" and "source" level issues, addressing the higher "activity" level provides the court with a much more valuable, more appropriate service. "Activity" level is the level that we would encourage all practitioners to consider. If, after consideration, the scientist does not feel able to address "activity", then the limitation of "source" and "subsource" opinions, and the possibly misleading impressions that may be given, must be made clear to the client and court. For further information on the topic, we refer the reader to the earlier references [6, 13] and to some specific case examples [17].

The categorization into "investigative" or "evaluative" opinion, together with the use of the hierarchy of issues, has helped to describe the different mode of operation of forensic scientists in the different phases of a criminal investigation and subsequent trial. We believe this will bring a new clarity and a common understanding of the contribution of forensic science to the Criminal Justice System.

The CAI Model

Many practitioners will have encountered situations in which police investigators submit large numbers of items, for example a wardrobe of clothes, with a request that the laboratory "do a full forensic examination". This type of request may have been due to a variety of reasons including:

police officers not being aware of the potential evidence types and their value;
police officers' desire to "cover all angles";
prosecution's desire to "cover all angles";
scientists' demands to see everything in a case;
local procedures; and
the seriousness of a case.

All these reasons may be justifiable in any one particular case. However, before any scientific work is started, serious consideration of the underlying reasons for doing the work must be made by all parties involved.

First, even if all items were to be examined, which tests should be used? Second, can the considerable expenditure of valuable resources, time and funding be justified?

The answer to the first question has usually been within the domain of the scientist. Police officers have been, and some still are, largely unaware of the specific scientific tests that are available let alone their applicability or usefulness. Some officers may have an awareness of, for example, DNA profiling but will not be able to specify the type of DNA-profiling tests that can be done. Police officers have tended to leave the choice of tests to the scientist, believing that the scientist will know best what to do. Scientists, on the other hand, tend to commission only those tests that are available in-house within their own organizations, and possibly just within their own building! They would tend to select those tests that appear to have served them well in the past, or that have been recommended and used by more experienced colleagues, or that are written down in set procedures for their organization. Their choice of test may have been determined more by technical considerations than by the needs of the case.

The second question – "can the expenditure be justified?" – is one that came to the forefront of people's thinking in the early 1990s with the move to Agency status of the main forensic science provider in England and Wales, the Forensic Science Service (FSS). As a result of major changes in the way the FSS and its customer police forces interacted, both customers and providers focused more on "value-for-money" forensic services. Police forces were required to obtain best value from their providers. Furthermore, the FSS had a vested interest in operating in an efficient, effective, and economic manner – it did not intend to waste resources and was required to "balance its books". In the early phases of this new relationship, attempts were made to monitor measures of performance such as the length of time to complete a case, the cost of a case, the effectiveness of the outcomes, the numbers of items examined, and the number of tests completed. Most of these measures are relatively straightforward to compile, but one measure that has been difficult to track is the effectiveness of the scientific outcome. It may seem a straightforward task to do so, but no robust, easy-to-apply system has yet been developed to assess the effectiveness of the scientific outcome. Attempts have been made [18–20] in specific areas but there remain

large gaps in knowledge and understanding on this aspect of performance.

The CAI model, as mentioned earlier (Section "Value for Money" and Bayes Theorem), was developed to provide a framework that would help scientists devise cost-effective examination strategies for the client [5]. The underpinning, logical structure of the framework was Bayes Theorem and, particularly, use of the LR to assess weight of evidence. Through workshops on case studies with groups of leading practitioners, the model was applied to the majority of case types within forensic science and was refined as more was learnt about the functioning of the model in different situations. The model comprised a series of key stages that followed the natural course of a case. For each stage, there was a series of questions that the scientist could ask to help develop the examination strategy and interpret the outcomes of tests. These questions can be captured in the form of an aide-memoir to help prompt the scientist – an example is shown in the Appendix. Such aide-memoirs have been found to be useful in the training of new practitioners as well as in offering a permanent, auditable record of the decision-making process within a case.

The key stages of the model are as follows:

define the customer requirement;
assess how forensic science can help;
agree on a case examination strategy;
carry out examination;
interpret the results; and
communicate the test results and opinions.

The emphasis of the model is in understanding the issues in a case and devising the examination strategy before any substantive work is carried out. This approach contrasts with earlier practice in which little attention may have been given to these aspects. Furthermore, the model forces the scientist to articulate her expectation before any material is recovered, hence avoiding any *post hoc* rationalization once the material has been found and analyzed.

The model encourages practitioners to communicate with the customer and to understand the following:

1. what is at issue or is uncertain;
2. the date by which the scientific information is required; and
3. the financial budget of the customer.

Having taken these aspects into consideration, the scientist can devise an appropriate strategy, consult again with the customer and gain her approval, or otherwise, for the examination to go ahead. During the communication between scientist and customer, the scientist can take the opportunity of educating the customer on likely outcomes and their evidential value, thereby managing the customer's expectation. Only when approval to proceed has been gained, should the scientist commence substantive work.

The model is not a rigid procedure. Rather it is a guide, a set of useful questions that can be asked. Neither is it a straight-through process to be followed from start to finish with no diversion. Real-life cases tend to be more complex, and the model allows for "looping back", or revisiting stages, in the light of new information, changing circumstances or surprising findings.

Some practitioners would argue that the model reflects just what they had been doing for many years, and we would agree with that view. However, development of the model helped to explore and capture the elements of good forensic science practice and, for the first time, to record formally that practice.

Case Example – a "Simple" Transfer Case

Case Circumstances

A vehicle used in the commission of an armed robbery has been found abandoned. Tape lifts were taken from the driver's seat and submitted along with the outer clothing – a sweater (item S) and a pair of jeans (item J) – taken from a suspect. The police request simply a "comparison of fibers".

Stage 1: Define the Customer Requirement. The model encourages the scientist to check, at the very beginning of a case, whether she has enough information to help her understand the key issues in the case. If the issues are "evaluative", has sufficient information on the key elements of the circumstances been provided to (i) help formulate propositions and alternatives, and (ii) inform probabilities for obtaining findings? If the issues are "investigative", the scientist needs to understand what would be the key investigative questions to address.

In this case, the scientist would probably want to know the following, for example:

- Do the police suspect this man of being the driver at the relevant time?
- What is the prospective charge against him?
- What strength of evidence will be required for charging, if not already charged, or for prosecution?
- How long had the car been in the possession of the robber/s?
- What did the lawful driver wear while driving the car?
- When were the suspect's items of clothing and the tape lifts taken?
- What happened in the intervening period?
- What does the suspect allege as an alternative?
- By when do the police need any information?
- What is their budget?

This list is not necessarily exhaustive and there may well be other queries that the scientist would have. However, obtaining answers to this list would help the scientist begin to formulate an examination strategy. We will assume the following answers:

Do the police suspect this man of being the driver at the relevant time **– yes**
What is the prospective charge – **unlawful taking of a motor vehicle**
What strength of evidence is required – **suspect has been charged but there is very little other evidence against the suspect and so strong evidence would be required to proceed with the charge; if strong evidence is not forthcoming, the charge will be dropped**
How long had the car been in the possession of the robber/s – **about 1 hour**
What did the lawful driver wear while driving the car – **unknown**
When were the suspect's items of clothing and the tape lifts taken **– both about 1 hour after car abandoned**
What happened in the intervening period – **suspect: unknown activity; car: no-one else had access to the car**
What does the suspect allege as an alternative – **he says he was not the driver and has never been in that vehicle**
By when do the police need any information – **suspect bailed to reappear in three weeks time**
What is their budget – **maximum of £3000 (~€4500)**

This additional information helps the scientist to understand the issues in the case and to take into account the timeliness and budgetary requirements of the police.

Specifying the key issues in a case may not be an easy task and requires significant care, analytical skills and effective, two-way communication. The first consideration is whether the issue is "evaluative" or "investigative". As there is a suspect, and he is giving an alternative explanation, the issue would seem to be "evaluative" and, considering the circumstances of the case, the key issue could be phrased as follows:

Is the suspect the person who drove the vehicle at the time of the robbery?

Following the guidance we discussed earlier, this issue can be classified as "activity" level – the level that is the more valuable level from a court's perspective.

Stage 2: Assess How Forensic Science Can Help. As the issue has been identified as "evaluative", this stage of the model requires the scientist to consider the LRs of all possible outcomes of the various potential examinations. One way of evaluating LRs is through the development of tables based on pairs of relevant propositions that come directly from the specified issue. Following our earlier notation, we will let H_p represents the prosecution proposition and H_d the defense proposition. In our case example, the following, competing propositions have been generated from the specific, activity-level issue:

H_p – The suspect is the person who drove the vehicle at the time of the robbery.
H_d – The suspect is not that person; someone else drove the vehicle.

Provided these propositions are mutually exclusive and exhaustive in the context of the case, the scientist can attempt to evaluate LRs for the different examinations. In this case, there is a choice of examining

both items;
just one of the items; or
neither of the items.

We will assume for the moment that there is only one examination technique available. We will also assume that the scientist has carried out a short preview examination to assess how well the garments

shed fibers and to determine the fibre type and color. On the basis of the scientist's knowledge of transfer, persistence and detection of the garment's fibers and on their background levels on vehicle seats, let us assume the scientist has developed the LR tables as in Tables 1 and 2 for the two garments.

There is the potential for much discussion amongst practitioners around the issue of how a scientist assigns the probabilities. However, as mentioned earlier, we leave that topic for discussion elsewhere. Assuming the tables are a fair reflection of probabilities, it can be seen that the LRs in Table 1 are spread more widely and are of greater magnitude generally than those in Table 2. Therefore, whatever the outcomes of the two examinations, examination for fibers like S will be the more informative, the more probative – the LR for each outcome in Table 1 "pushes" one way or the other the prior probability of the suspect having driven the car. The LRs in Table 2 either do not change that prior probability or change it only by a small amount.

Furthermore, considering the likelihoods for obtaining the outcomes, there is a greater chance of obtaining useful outcomes that point the "right" way in Table 1 rather than in Table 2. From Table 1, we can see that, if the police really have the "right" person, i.e., the conditioning is on H_p being true, then there is a combined 98% probability (0.05 plus 0.25 plus 0.68) of obtaining an outcome (LRs of 5, ~30, and 340 respectively) supporting this proposition, i.e., the results point the "right" way. There is, however, a 2% probability that, even though the police have the "right" man, the scientist will obtain an outcome (LR of 1/49) that supports the defense proposition, i.e., the result points the "wrong" way. On the other hand, if the police have the wrong man, i.e., the conditioning is on H_D being true, then there is a 98% probability that the outcome (1/49) would point that way, thus supporting the defense proposition.

If the same analysis is carried out for Table 2, we can see that, if the suspect is the "right" man, there is only a 2% probability of obtaining an outcome (*LR* of 4) that would support significantly that proposition.

This method of comparing the benefits of different strategies can be applied not only where there is a choice of items from one suspect but to any situation in which there is a choice to be made. Examples include cases in which there is a choice of several tests that can be applied to the same item, cases in which there is a choice of evidence types and cases in which there is potential for two-way transfer.

Table 1 This describes the probabilities for all the potential outcomes of examination of the seat tape lifts for fibers matching those of garment S

Outcomes from examining tape lifts for fibers matching S (E_S)	Pr [E_S|H_p,I]	Pr [E_S|H_d,I]	LR
1. No fibers matching S	0.02	0.98	1/49
2. Few fibers (1–5) matching S	0.05	0.01	5
3. Moderate number of fibers (6–20) matching S	0.25	0.008	~30
4. Many fibers (>20) matching S	0.68	0.002	340
	1	**1**	

Table 2 This describes the probabilities for all the potential outcomes of examination of the seat tape lifts for fibers matching those of garment J

Outcomes from examining tape lifts for fibers matching J (E_J)	Pr [E_J|H_p,I]	Pr [E_J|H_d,I]	LR
1. No fibers matching J	0.2	0.4	1/2
2. Few fibers matching J (1–5)	0.3	0.35	~1
3. Moderate number of fibers matching J (6–20)	0.3	0.2	1.5
4. Many fibers matching J (>20)	0.2	0.05	4
	1	**1**	

We must point out that the two tables have been evaluated as two independent tests. To evaluate the potential strength of carrying out both tests, we would need to reappraise the results of the second test based on the outcome of the first.

Of course, there are not only benefits, but also costs and limitations. Let us assume that the examination in Table 1 would cost £4000 and would be completed within two weeks while the examination in Table 2 would cost £2000 and would be completed in three days. Both examinations meet the timeliness requirement of the police but only examination 2 is within their target budget. However, examination 1 is the only one of the two examinations that has a potential of providing significant evidence – the outcome of "many" fibers matching in Table 1 provides the largest LR in the two tables and would be reported with a verbal qualifier of "moderately strong" evidence, using one published verbal scale [8]. None of the other outcomes would be as strong. So, while examination 2 would meet the cost and speed requirements of the police, it does not have the potential to provide the strength of evidence that the police say they need.

The information on benefits, costs, and limitations of the various strategies can form the basis of a dialogue between the forensic science provider and the police/prosecution/defense/court customer.

Stage 3: Agree A Case Examination Strategy. Depending on the nature of the relationship between customer and provider, and on the seriousness or complexity of the case, there may be a requirement on the provider to contact the customer to discuss and agree a case examination strategy. The CAI model captures this contact as a formal step to emphasize that the final decision on which tests on which items to commission rests with the customer.

Stages 4–6 Carry out Examination; Interpret the Results; and Communicate the Test Results and Opinions. These stages are what may be viewed as the traditional work of a forensic scientist. The main difference in the CAI model from traditional procedures is the recognition that opinion should be communicated in the form of an LR, if offering an "evaluative" opinion, or as a list of explanations or posterior probabilities if "investigative".

We will assume in this case that, notwithstanding the cost and potential strength of evidence, examination of the seat tape lifts for fibers like garment S was commissioned. Let us also assume that the result of the examination was 12 matching fibers, outcome 3 in Table 1. There will be different ways of communicating an opinion based on this outcome and its associated LR. One way could be as follows:

In my view, the finding of such a number of matching fibers on the seat tape lifts is approximately 30 times more likely if the suspect, rather than someone else, was the person who drove the vehicle at the time of the robbery.

This could be followed by a further paragraph expressing the (log of the) magnitude of the LR. The following example uses again a published verbal scale [8]:

The findings provide moderate support for the view that it was the suspect, rather than some other person, who drove the vehicle at the time of the robbery.

An "Investigative" Case

As described earlier, cases requiring so-called "investigative" opinion may well not involve a suspect. There will not therefore be an alternative proposition and, if there is uncertainty about the details of the crime, there may not be a police/prosecution proposition about what happened at the scene. An LR table, as shown in the first case example of an "evaluative" issue, cannot be developed. The "investigative" forensic scientist is therefore in a different position from the "evaluative" forensic scientist and will operate with a different mind-set and therefore in a different role. While this role will still, neccessarily, involve identifying the issues, it will focus on those areas of forensic science that will provide the most useful information to help move the investigation forward. Usefulness could be based on various measures including the speed of provision of the information, the potential of the proposed examination to reduce uncertainty for the police investigator, and the cost of the examination. The aim of the "investigative" scientist would be to provide useful explanations for the observations made at the scene or on the victim. If possible, the scientist would also rank these explanations in order of their probability.

In terms of application of the CAI model, the "investigative" scientist would consider and recommend the examination that would have the best chance of providing the most useful information within the time and budgetary limits of the police.

Wider Implications

For CAI to be applied effectively, not only must the scientists have the ability to apply this approach, but also each *organization* involved in the forensic process has to play their part. This means each organization having a common understanding of the scientist's role and having supportive procedures in place. These procedures can be categorized under the following headings: **communications and awareness; quality; and organizational support**.

Communications and Awareness

A key element of CAI is that of establishing the customer's requirements (as set out in The CAI Model section). This requires effective and efficient communication and discussion with the customer – police, prosecuting team (i.e., police and prosecuting authority) or defense.

Once the customer's requirements have been established, and the scientist has assessed the case, the scientist needs to communicate to the customer whether or not forensic science can assist. If it can, what are the chances of a helpful outcome (one way or the other); and what is the best examination strategy. At this stage, an estimate of the charges and the time scales for the examination can be agreed. Once the examination is completed, then further communication is required to explain the result or to carry out more work if needed.

To enable all this dialogue to take place in the first place requires collaboration and excellent communication on a case-by-case basis between the forensic science provider and the agencies involved. In England and Wales, in order to facilitate these communications, National and Local Tripartite Protocols have been agreed among the police, forensic science providers, and the Crown Prosecution Service. These protocols make clear the responsibilities of each of the agencies and set out processes that will support the required communications. The identification, at the earliest stage, of the key issues to be addressed is a key element of these protocols.

All parties in the forensic process should be aware of the needs of each of the others. Police and lawyers should be aware of what forensic science can provide and also its limitations. Lawyers and judges should understand how the forensic scientist is interpreting the evidence and how their approach fits with the rest of the nonscientific evidence and either augments or diminishes it. Forensic scientists should also develop their understanding of their customers' investigative and legal requirements.

Quality

In order to apply CAI effectively, forensic scientists need a good understanding of the Bayesian approach and an ability to assess probabilities based upon relevant data that have been documented and, preferably, published. They also need to understand and apply the concept of the hierarchy of issues [6, 7] and be comfortable with the difference between "investigative" and "evaluative" opinion [13]. Every practicing forensic scientist should be able to demonstrate competence in these skills to managers, external assessors and the courts. Provider organizations should set standards against which performance can be audited and validated as part of their quality management. Accreditation bodies should also assess performance against the principles underpinning CAI although as yet there are no international standards for forensic science.

Organizational Support

Apart from training and continuous professional development, any organization that provides forensic science should support its practitioners. Management should ensure that protocols and systems are in place to enable practitioners to communicate with personnel in other agencies. They should build CAI into their internal processes.

Organizations should ensure that sufficient reliable data are available to practitioners to enable them to assess probabilities. This requirement is ongoing and should shape future research programs. The application of CAI should also shape the future service provision by the forensic science provider. At a corporate level, customer communication and awareness programmes should also include CAI elements.

Further Consequences

Source- and Activity-Level Evidence

As we saw earlier, a very useful concept that emerged from the early work on CAI was that of the Hierarchy of Issues and the idea of subsource, source and

activity level propositions. This has provoked further thinking on how forensic evidence should be interpreted and presented, and has highlighted the potential overvaluing of some evidence if it is evaluated solely at source or, even worse, subsource.

Staged Reporting

The application of CAI provides the potential for a staged approach to forensic examinations – an approach sought by other partners in the Criminal Justice System.

Forensic science can provide limited information at early stages in the criminal justice process in order that the prosecution team can ask the right questions of the suspect or even to assist in putting charges to a suspect. Depending on the response of the suspect at each stage of the process, more detailed forensic work can be carried out to address specific issues as they arise – the suspect's response will frame the alternative proposition considered by the scientist and, hence, may influence the magnitude of the LR. An admission of guilt or a "guilty plea" at any stage in the investigation could eliminate the need for lengthy and costly forensic examinations thus reducing the time and cost of forensic intervention.

Factual versus Evaluative Reporting

A further corollary of CAI and staged reporting is the provision of factual reports that do not include evaluation or opinion but simply state the results of the forensic examination. Examples of this are as follows:

> "DNA from the blood on the murder weapon matched the DNA from the suspect."
> "A large amount of semen was found on the internal vaginal swabs."
> "Twenty-five fragments of freshly-broken glass, falling into two distinct groups or types, were recovered from the jacket'."
> "The white powder in possession of the suspect was found to be cocaine."

Factual reports should only be provided when the police and prosecuting team have clear understanding of what is being provided and would usually be part of a staged approach to the forensic investigation.

One key benefit of factual reporting is a result of interpretation of the findings being limited to the analytical or "source/subsource" level. This may bring some savings in cost and time if, for example, there is a "guilty" plea or the forensic science evidence is not disputed. However, it must always be understood by the police and lawyers that such reports do not involve higher level interpretation or expert opinion in its true sense and that there has been no assessment of the weight or value of the findings.

Development of a commercial market place and the procurement of forensic services

Although a market place for forensic science has been established in England and Wales, it is currently unregulated (see later) and is still in its early stages of development. Indeed forensic science itself continues to develop as new technology stretches the limits of forensic knowledge and interpretation. Arguably, the development and use of CAI has demonstrated the true value of forensic science as an interpretative discipline that can be applied to any technology that provides results or observations. Note that we are drawing a distinction between the *provision* of test results and the *interpretation* of test results within the CAI framework. The authors believe that when working as "evaluators", forensic scientists should interpret results at the highest level possible. To do anything less could leave the interpretation to those less able to do so, with dangerous consequences and potential miscarriages of justice of the kind referred to earlier in this article.

One of the consequences of the development of what is currently an unregulated market place is the move by the police to demand, as paying customers, test results on items preselected by them. This has two further consequences. First, there is the potential for forensic science to be applied only from a prosecution viewpoint. Second, test results may not be interpreted in the most robust and effective way, even if there is any interpretation at all – the judge and jurors, alone, will be left to make sense of the results or to advance from subsource/source issues to activity issues, perhaps under the conflicting persuasion of the prosecuting and defending lawyers.

Regulation of forensic science

Although there is no regulation of the market place in England and Wales, a government initiative to develop a regulatory framework for forensic science

practice is well underway. A Forensic Science Regulator has been appointed and will oversee the end-to-end forensic process from scene/suspect examination through laboratory work to the presentation of evidence in court. This should include the way in which tests are commissioned, how to avoid bias, and the delivery of interpretation within, we believe, the CAI framework in all cases where possible.

Strong representation from forensic providers is keeping the regulation of expert opinion firmly on the agenda so that standards will be maintained and overseen by the Regulator.

Clarification of Roles in the investigative and evaluative stages of Crime Investigation

The development of CAI has led to questions being asked about the role of forensic scientists, crime scene examiners, and police and lawyers in crime investigation. It has also lead to the concept of investigative and evaluative forensic science discussed earlier. It is right to say that our understanding of these "phases" of forensic science is still work in progress. However, what is emerging is that once an individual is called upon to provide expert opinion, whatever their role or job description may be or whatever organization they may belong to, there are benefits from operating within the CAI framework.

Changes to the Adversarial System of Justice

It is within the adversarial system itself that some of the most fundamental changes need to take place to enable forensic scientists to apply CAI effectively. It is beyond the scope of this article to expand in detail on this subject although many of the main points have been discussed earlier.

Perhaps the central issue for the adversarial system to address is the recognition of forensic science as a tool for enabling truth to be ascertained rather than a tool for prosecution or defense to win their case. Many changes to the process have been recommended from various reviews in England and Wales – few, if any, have been implemented fully. It is maybe the case that a full understanding of the true nature of forensic science is lacking by the key players. Indeed, before CAI and its underlying principles were developed, the forensic science profession itself was unclear. Now, however, with the growing appreciation of the power of CAI and a confidence that that brings, perhaps the profession itself can, with a stronger resolve, make its voice heard and bring about the much-needed changes.

Appendix : A step-by-step checklist for Case Assessment and Interpretation (CAI)

1. Define the customer requirement
   - What are the case circumstances – timings, witness statements, allegations, medical and other specialist evidence?
   - What are the uncertainties – with what aspects of the case does the customer need help?
   - Is there a suspect? Has he/she been charged? What is the charge?
   - What strength of evidence is required – for charging or for prosecution?
   - What are the customer's deadlines – statutory requirements; other needs?
   - What is the budget?
   - Do you have sufficient information to help identify the key issues in the case? If not, try to obtain it.
2. Assess how forensic science can help
   Based on your understanding of the case circumstances:
   - Are the issues investigative and/or evaluative?
   - What level in the hierarchy of issues (sub-source/source/activity/offense) will you be addressing?
   - What type of opinion (explanation/posterior probability/ single likelihood/likelihood ratio) will you be offering?
   - If investigative, develop list of possible explanations with, if possible, realistic prior probabilities. Considering the tests available and the likelihoods of the different outcomes of the tests, decide which tests on which items have the best chance of providing the most effective information to help direct the investigation.
   - If evaluative, specify a relevant pair of propositions based on prosecution and defense positions. Develop likelihood ratio tables for the different items/tests, previewing items if necessary. Identify the relative cost-effectiveness of the different examinations.

3. Agree on a case examination strategy
   - Contact customer and explain rationale and manage customers' expectations
   - Discuss options
   - Agree purpose of examination and the examination strategy
4. Carry out case examination strategy
   - Do the tests
5. Interpret the results
   - If investigative, revisit your list of explanations. If providing posterior probabilities, incorporate the likelihoods for the findings with the prior probabilities, and revise list in order of posterior probabilities. If not offering posterior probabilities, then simply revise list and ensure all possible, feasible explanations are given.
   - If evaluative, compare results with previous assessment and refine the likelihood ratio.
6. Communicate
   - Check your "conclusion" agrees with your "purpose" as agreed in the case strategy. Convey to the customer your findings and conclusion in the most appropriate way.
   - Ensure the strengths and limitations of your opinion are clearly set out and understood by the customer.

Acknowledgments

We wish to thank all our colleagues, both within the profession of forensic science and in the wider scientific, police and legal community, for their help and support over many years during the development of the ideas presented in this article. In particular, our colleagues in the initial and subsequent CAI development teams have been a source of inspiration and support – our special thanks go to Ian Evett, Jim Lambert, Roger Cook, Christophe Champod, Stella Jones, and Gareth Booth.

References

[1] R v Alan James Doheny, R v Gary Adams. Court of Appeal–Criminal Division [1996] EWCA Crim 728 (31st July, 1996).

[2] R. v. Barry George, Court of Appeal – Criminal Division [2007] EWCA Crim 2722 (15th of November 2007).

[3] Evett, I.W. (1983). What is the probability that this blood came from that person? A meaningful question? *Journal of the Forensic Science Society* **23**, 35–39.

[4] Fraser, J. (2007). The application of forensic science to criminal investigation, in *Handbook of Criminal Investigation*, T. Newburn, T. Williamson & A. Wright, eds, Willan.

[5] Cook, R., Evett, I.W., Jackson, G., Jones, P.J. & Lambert, J.A. (1998). A model for case assessment and interpretation, *Science & Justice* **38**(3), 151–156.

[6] Cook, R., Evett, I.W., Jackson, G., Jones, P.J. & Lambert, J.A. (1998). A hierarchy of propositions: deciding which level to address in casework, *Science & Justice* **38**(4), 231–239.

[7] Evett, I.W., Jackson, G. & Lambert, J.A. (2000). More on the hierarchy of propositions: exploring the distinction between explanations and propositions, *Science & Justice* **40**(1), 3–10.

[8] Evett, I.W., Jackson, G., Lambert, J.A. & McCrossan, S. (2000). The impact of the principles of evidence interpretation on the structure and content of statements, *Science & Justice* **40**(4), 233–239.

[9] Robertson, B. & Vignaux, G.A.R. (1995). *Interpreting Evidence: Evaluating Forensic Science Evidence in the Courtroom*, John Wiley & Sons.

[10] Aitken, C.G.C. & Taroni, F. (2005). *Statistics and the Evaluation of Evidence for Forensic Scientists*, John Wiley & Sons.

[11] Lucy, D. (2005). *Introduction to Statistics for Forensic Scientists*, John Wiley & Sons.

[12] Taroni, F., Aitken, C.G.C., Garbolino, P. & Biedermann, A. (2006). *Bayesian Networks and Probabilistic Inference in Forensic Science*, John Wiley & Sons.

[13] Jackson, G., Jones, S., Booth, G., Champod, C. & Evett, I.W. (2006). The nature of forensic science opinion – a possible framework to guide thinking and practice in investigations and in court proceedings, *Science & Justice* **46**, 33–44.

[14] Taroni, F., Aitken, C.G.C. & Garbolino, P. (2001). De Finetti's subjectivism, the assessment of probabilities, and the evaluation of evidence: a commentary for forensic scientists, *Science & Justice* **41**, 145–150.

[15] Biedermann, A., Taroni, F., Delemont, O., Semadeni, C. & Davison, A.C. (2005). The evaluation of evidence in the forensic investigation of fire incidents (part I): an approach using Bayesian networks, *Forensic Science International* **147**, 49–57.

[16] Biedermann, A., Taroni, F., Delemont, O., Semadeni, C. & Davison, A.C. (2005). The evaluation of evidence in the forensic investigation of fire incidents (part II): practical examples of the use of Bayesian networks, *Forensic Science International* **147**, 59–69.

[17] Evett, I.W., Gill, P.D., Jackson, G., Whitaker, J. & Champod, C. (2002). Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks, *Journal of Forensic Sciences* **47**(3), 520–530.

[18] Burrows, J. & Tarling, R. (2004). Measuring the impact of forensic science in detecting burglary and autocrime offences, *Science & Justice* **44**(4), 217–222.

[19] Bond, J.W. (2007). Value of DNA evidence in detecting crime, *Journal of Forensic Sciences* **52**(1), 128–136.
[20] Williams, R. (2008). *The Management of Crime Scene Examination in Relation to the Investigation of Burglary and Vehicle Crime*, http://www.homeoffice.gov.uk/rds/pdfs04/rdsolr2404.pdf (accessed 10-01-08).

GRAHAM JACKSON AND PHILIP J. JONES

Case Congregations *see* Behavioral Science Evidence

Casting of Evidential Marks *see* Evidence Collection and Preservation: Casting

CDR *see* Firearm Discharge Residue: Analysis of

Ceiling Principle: DNA

Introduction

The National Research Council (NRC) committee was assembled in 1990 to assess the general applicability and suitability of using DNA technology in forensic science. The NRC Report [1] provided recommendations in six areas related to the DNA typing process: technical issues, statistical interpretation, laboratory standards, databanks and privacy, legal issues, and societal and ethical issues.

Assessing the probability of observing a particular multilocus genotype is a pursuit that is only attempted in forensic science. At the time of the NRC deliberations, this was already a controversial issue and a fundamental reason for the committee's existence. However, the NRC recommendations in the area of statistical interpretation actually provoked further controversy, rather than allaying it. Arguably the most controversial recommendations were those known as the *ceiling principle* and *interim ceiling principle*.

Much of the early debate surrounding DNA evidence interpretation involved the need to account for population substructure in statistical estimates. Owing to population-level differences in allele frequencies, statistical estimates can vary depending upon the source of population data. In a criminal case it is not usually known what database is the most appropriate to apply as this depends on, firstly, being able to assess the community of persons that had opportunity to commit the crime, and secondly, to have access to a representative database of this community.

At its simplest, it was the belief in population heterogeneity and its potentially discriminating effect on statistical evidence that led to the justification for the ceiling principle. It was believed that the ceiling principle would yield conservative estimates, even for a substructured population, provided the allele frequencies used in the calculation exceed the allele frequencies in any of the population subgroups. To apply the ceiling principle, the upper bound for the frequency of each allele is required, irrespective of the population of origin. The largest frequency in any of these populations, or a fixed figure of 5%, whichever is larger, should be taken as the ceiling frequency. The ceiling frequencies are then multiplied using the product rule to obtain the genotype frequency. This product of maximal frequency values serves as an upper bound for the product of all other unknown frequencies. To determine the frequencies themselves, the NRC recommended that random samples of 100 persons be drawn from each of 15–20 populations that each represents a relatively homogeneous genetic group.

Both the population genetic and statistical basis of the ceiling principle recommendation and the study proposed by the NRC were subsequently criticized in the literature (see for example [2–6]). It also caused conjecture in court by undermining the use of population-specific estimates [7, 8] or provoking significant controversy that it failed the general acceptance test and was ruled inadmissible [9]. In the

second report produced by the NRC [10], the ceiling method was described as unnecessarily conservative and was abandoned in favor of recommended methods to obtain population-specific frequency estimates that account for population subdivision.

References

[1] National Research Council (1992). *National Research Council Report: DNA Technology in Forensic Science*, National Academy Press, Washington, DC.

[2] Weir, B.S. (1992). Population genetics in the forensic DNA debate, *Proceedings of the National Academy of Science, USA* **89**, 11654–11659.

[3] Weir, B.S. (1993). Forensic population genetics and the NRC, *American Journal of Human Genetics* **52**, 437–439.

[4] Devlin, B., Risch, N. & Roeder, K. (1994). Comments on the statistical aspects of the NRC's report on DNA typing, *Journal of Forensic Sciences* **39**(1), 28–40.

[5] Cohen, J.E. (1992). The ceiling principle is not always conservative in assigning genotype frequencies for forensic DNA testing, *American Journal of Human Genetics* **51**(5), 1165–1168.

[6] Slimowitz, J.R. & Cohen, J.E. (1993). Violations of the ceiling principle: exact conditions and statistical evidence, *American Journal of Human Genetics* **53**, 314–323.

[7] State v. Sivri, 646 A.2d 169 (Conn. 1994).

[8] State v. Carter, 246 Neb. 953, 524 N.W.2d 763 (1994).

[9] People v. Wallace, 14 Cal. App. 4th 651, 17 Cal. Rptr. 2d 721 (1993).

[10] National Research Council (1996). *National Research Council Report: The Evaluation of Forensic DNA Evidence*, National Academy Press, Washington, DC.

Related Articles

Interpreting Expert Opinions: History of

SIMON J. WALSH

Cellular Telephone Networks *see* GSM Analysis and PDAs

Certification: Criminalistics *see* Training and Certification (in Criminalistics)

Certification: Fire Investigator *see* Fire Investigator: Standardization, Accreditation, and Certification

Chain of Custody *see* Chain of Possession of Tangible Evidence

Chain of Possession of Tangible Evidence

Laying the Foundation for Admission of Physical Items

The chain of possession of tangible evidence, sometimes also referred to as the *chain of custody*, must be proved before such evidence, or testimony based thereon, can be admitted in court. This requirement exists especially in adversary systems of justice, though courts in civil law countries adhere to a similar, though more relaxed, rule. Today, chain of possession evidence is often still referred to as *proof of the chain of custody*, but the older term is somewhat disfavored because of the possible confusion its use creates with custody of a defendant in a criminal case. As in the decisional case law, we use the two terms here interchangeably.

Proof of the chain of possession requires evidence (i) that the item itself is connected to the controversy

being litigated, and (ii) that the physical evidence has not been tampered with or altered. Thus, the party seeking to introduce evidence of the analysis of physical evidence must show to the court that the physical item itself was, in fact, derived or taken from the particular person or place alleged. The process of satisfying this preliminary evidential rule is also referred to as *laying the foundation* for its admissibility.

The requirement of showing that proper proof of custody was maintained finds its origin in the need to prevent the courtroom use of evidence that may have been improperly altered, contaminated, or substituted between the time of its original discovery and trial.

Laying the foundation for proof of integrity is of particular importance in criminal cases. To satisfy the rule, the proponent of the evidence must call, to the stand, each person who has had possession of the physical item from the moment it was first obtained or discovered to the time of trial. In civil cases, when a break in the chain of possession has occurred, that fact is often said to affect only the weight to be accorded the evidence, and not preclude its admissibility. In criminal cases, by contrast, the unbroken chain of possession must be established with some measure of certainty. Where there exists a serious break in the chain of possession, its admission may be deemed prejudicial error requiring reversal of a conviction.

When dealing with forensic examinations by experts, the chronicle of proof of custody includes testimony of the initial possessor of the specimen or object, of its transportation to and receipt at the laboratory, the method of storage at the laboratory prior to analysis, and retention, whenever feasible, of either the original item or the unused portion of the specimen or object after its examination and analysis, all the way up to the time of the production of relevant testimony at the trial. This is a laborious process that may necessitate the presence of many witnesses who have little to offer the court beyond the testimony needed to establish the chain of possession. For that reason, such testimony is often dispensed with through a stipulation entered into by the counsel, who may agree to limit the testimony on possession to that of the person who originally obtained the item and the expert who examined it.

Where physical evidence has been transmitted through the mail by a law enforcement agency to a laboratory facility, it is obviously impossible to know which postal employees handled the specimen during shipment. For that reason, courts also dispense with the requirement that all persons who may possibly have been in contact with evidence be called as *witnesses*.

Identification and Marking of Evidence

In a criminal case where proof of possession is important, initial acquisition testimony must reflect that the investigator obtained an item, identified it, and placed it in the appropriate sealed container after marking it for identification. Tangible items that are not consumed in the analysis are marked for identification by the analyst as well, and secured until the time of trial. Customs vary in law enforcement and laboratory facilities as to the exact procedures and documentary requirements personnel must follow when dealing with the possession of evidence.

Defendant's guilt in a criminal case must be established by proof beyond a reasonable doubt. However, the law does not require that proof of a chain of possession be established by such a high quantum of evidence. If an unbroken chain of possession cannot be established, the court determines whether the break affects the possible validity of the expert's findings. If the broken link has no affect on the expert analysis, courts usually decide that the evidence can be used. Practicalities of proof do not require a party offering physical evidence to negate the remotest possibility of substitution or alteration. All that needs to be established is a reasonable certainty that there has been no wrongdoing. The courts do not require proof of the impossible.

When it comes to solid objects, such as guns or ammunition, a break in the chain of possession of some hand-to-hand transfers that cannot be accounted for by documentary evidence will not likely affect the item's admissibility, since missing evidence of a transfer of the weapon or ammunition is not likely to affect the expert's examination of striation evidence.

When Evidence Is Not Available

There are instances when the actual item itself is no longer available, or has been inadvertently destroyed, and all that is left is the expert's report of the examination. When destruction is inadvertent, courts are unlikely to be overly sympathetic to

the defendant's argument that he has been denied due process because he was unable to engage his own expert. While there are court cases where the unaccounted-for absence of the physical item itself has resulted in the expert's analysis being excluded from the trial, such cases occur infrequently.

The possibility of alteration, contamination, or destruction is far greater when dealing with extremely minute quantities of suspected narcotics, biological evidence such as a blood stain or semen stain, and other trace evidence. Marking a container may not be sufficient proof of lack of tampering where an unsealed container may have been handled by unknown third persons.

For example, assume that an undercover agent purchased a foil packet said to contain PCP (phencyclidine), placed the evidence in his wallet, and thereupon adjourned to a party where drugs were used, sleeping thereafter in a dormitory-like setting where his trousers were accessible to other persons while he slept. This is clearly a situation where courts may likely consider that a fatal break in the proof has occurred and the evidence will be denied admissibility.

When a substance, such as contraband narcotics, is tested in a forensic laboratory and is thereafter altered or consumed during testing, or destroyed after testing pursuant to law, judges are likely to hold that there is no break in the chain of custody requiring exclusion of expert testimony based thereon. The same result will occur when items analyzed have been returned to a rightful owner who was wrongfully deprived of his property so that these items cannot be exhibited in court. It must be remembered, however, that even where a link in the custodial chain is deemed not to affect admissibility of opinion evidence, the physical absence of the item at trial may well affect the jury's belief in the expert's testimony.

Related Articles

Cross-Examination of Experts

Demonstrative Evidence

Direct Examination of Experts

Expert Opinion: United States

ANDRE MOENSSENS

Chemical, Biological, Radiological, and Nuclear Investigations

What is CBRN?

CBRN is an acronym for chemical, biological, radiological, and nuclear and is a term usually used to describe the threatened, intended, or actual use of harmful CBRN material in warfare or against civilians or economic targets. The malicious use of CBRN agents against individuals, human populations, animals, plants, and infrastructure is an enduring and increasing concern.

Malicious Use of CBRN Agents in History

Some CBRN agents developed or used in state-sponsored military programs have been described as weapons of mass destruction (WMD) and their use is not new. Persian, Greek, and Roman literature refer to the use of chemical and biological weapons for military purposes pre-400 BC. Over the two millennia chemical and biological agents have been maintained as a prominent military strategy and were used in both the First and Second World Wars. In the final stages of the Second World War, nuclear weapons quickly became the weapon of choice for mass destruction. Since then, WMDs have served primarily as military deterrents.

The Geneva Protocol of 1925 banned the use of chemical weapons in warfare. A number of nations, including the United States and the former Soviet Union, continued to develop and stockpile significant quantities and varieties of these types of weapons well into the twentieth century. The Biological and Toxins Weapons Convention (BWC) of 1972 and the Chemical Weapons Convention (CWC) of 1997 banned the production and stockpiling of biological and chemical agents for nonpeaceful purposes. There is no corresponding convention related to nuclear weapons that would unilaterally ban production and stockpiling of nuclear material. Instead, the Nuclear Non-Proliferation Treaty, which entered into force in 1970, aims at preventing the spread of nuclear weapons and weapon technology from the original five nuclear weapons states (United

States, Russia (formerly Union of Soviet Socialist Republics), United Kingdom, France, and China) to nonnuclear weapon states.

A number of signatories to the BWC and the CWC continue to work with these agents for research purposes to describe the geographic distribution and variation, or as part of their defensive military programs for the development of medical countermeasures, or more controversially, for risk analysis [1]. The exact nature and extent of most historic state WMD programs remains classified, though clues can be gained from the testimonies of defectors such as Ken Alibek who was involved in the biowarfare program of the former Soviet Union that continued after they signed the BWC in 1972 [2].

The former Iraqi regime under Saddam Hussein developed and used chemical weapons against Iran and its own citizens in the 1980s. Following the first Gulf War in 1991, Iraq's alleged continued development and stockpiling of both chemical and biological weapons instigated the invasion by Western coalition forces in 2003. The extent, nature, and current status of Iraq's WMD program has not been fully understood.

CBRN terrorism is the use of these agents by individuals or groups motivated by religious, political, ecological, or other ideological objectives [3]. With the occurrence of the 2001 terrorist attacks on the World Trade Center and the "Amerithrax" incidents, governments have refocused attention on CBRN agents as the interest in such weapons grows within terrorist organizations. There has been significant media discussion of al-Qa'ida's plans to acquire and use cyanide, mustard, anthrax, and botulism, and in 2004 Jordanian officials disrupted an al-Qa'ida plot to release toxic chemicals in Jordan against the United States and Jordanian security interests. In 1984, a large outbreak of salmonellosis was investigated in Oregon, USA, and was subsequently found to be intentionally caused by a religious group in an attempt to influence local elections [4]. More recently, the Aum Shinrikyo cult used sarin gas against Japanese civilians in 1994 and 1995 [5].

The use of these agents for murder, extortion, or revenge is a criminal offence. A notable example are the "Amerithrax" attacks involving *Bacillus anthracis* spores sent through the post in October 2001 [6]. Five people subsequently died from the disease anthrax. Forensic and microbiological investigation revealed that the *B. anthracis* used was the Ames laboratory strain [7]. This has led to media speculation that the agent was sourced from a US laboratory and the perpetrator likely to be an experienced laboratory worker with access to hazardous biological agents. Other examples include the documented cases of theft and black market trading of highly enriched uranium or plutonium from the former Soviet Union [8].

Inadvertent release of hazardous CBRN agents may result in injury, death, or significant economic damage. Such events would also be subject to police investigation. For the forensic investigator, the issues surrounding the hazards of the scene and the handling of evidence would be the same as for a terrorist event.

Chemical, Biological, Radiological, and Nuclear Agents

Related entries provide further detail on investigations involving specific CBRN agents (*see* **Biological Agents**; **Chemical Warfare Agents**; **Nuclear Forensics**). CBRN are described here in brief.

Chemical warfare (CW) agents include nerve, blister, blood, and pulmonary (choking) agents [9, 10]. Traditional CW agents such as VX, sarin, and tabun interfere with the normal transmission of neurotransmitter signals within the body and at high doses cause convulsions, loss of consciousness, and death within minutes. Many commercially available pesticides can have similar effects as nerve agents. Blister agents tend to be debilitating rather than lethal, causing effects immediately (Lewisite) or many hours (sulfur mustard) following exposure. Typical symptoms include severe pain, blistering of the skin, and burning of the eyes. Blood agents, including hydrogen cyanide and cyanogen chloride, are absorbed into the blood through inhalation and act by preventing cells from using oxygen. Symptoms may not appear for several hours following a low-dose exposure and include nausea, weakness, and anxiety, with loss of consciousness and death following exposure to high concentrations. Cyanide is found in many natural products and is used for many industrial purposes. The pulmonary (choking) agents, including chlorine gas and phosgene, are so called because the alveoli in the lungs are burned by the strong acids that are produced as the gas comes in contact with water inside the lung tissue. The damage induces edema, resulting in "dry land drowning". Pulmonary agents induce symptoms 2–24 h following exposure.

The biological agents that can be used as weapons are either living organisms (pathogens, including bacteria and viruses) or are nonliving toxins [9, 11]. Bacteria and viruses work by infecting the host, replicating and inducing disease through direct damage to cells or by producing toxins that interfere with normal cellular function. The range of diseases caused by biological agents is enormous and requires the consideration of numerous variables when contemplating how to respond to potential biological releases. Depending on the route of exposure, different disease forms may be produced, with symptoms appearing hours (toxins) to weeks or even years (pathogens) following infection. Further, unlike chemicals, radionuclides, and toxins that are self-limiting, many of the living biological agents cause contagious diseases that can spread from host to host and prolong the disease outbreak. In addition, host susceptibility and immunity are important factors in determining whether exposure to a biological agent will cause disease. Many of the important biological agents are zoonotic pathogens, meaning they can infect both humans and animals, and others are spread by vectors, including fleas, ticks, and mosquitoes, making their natural "reservoirs" difficult to eradicate. Biological agents such as *B. anthracis* (causing anthrax) and *Francisella tularensis* (causing tularemia) do not normally spread from host to host. *Yersinia pestis* (causing plague) can easily spread from person to person when the disease manifests as the pneumonic form, but is rarely spread when manifested as the bubonic form. The *Variola* virus (causing smallpox) has been globally eradicated and as a result population immunity is low to nonexistent. Biological toxins including botulinum (causing botulism, produced by the bacterium *Clostridium botulinum*) and ricin (derived from *Ricinus communis*, the castor bean plant) can cause lethality at extremely low doses.

Many naturally occurring elements are radioactive or have radioactive isotopes that emit ionizing radiation in the form of alpha, beta, or neutron particles and gamma rays [12, 13]. The effect of radiation exposure is dependent upon the type of radiation, dose, and the route of exposure. External exposure occurs when the subject is exposed to radiation emitted from a source (radioisotope) that is deposited either on the surface of a person or is in close proximity to a person. Internal exposure occurs when a radioisotope is inhaled or ingested, or gets inside the body through wounds, cuts, or injection. Highly penetrating, long-range radiations like gamma or neutrons are the main concern for external exposure, whereas weakly penetrating, short-range radiations like alpha and beta particles are the main concern for internal exposure. Many radioisotopes emit more than one form of radiation. For example, radioactive isotopes cobalt-60 and cesium-137 decay by emitting beta particles, accompanied by gamma radiation. Uranium-235, uranium-238, radium-226, americium-241, and polonium-210 decay by emitting alpha particles and accompanying gamma radiation. Low doses of radiation may not cause immediate symptoms but may increase the likelihood of cancers later in life and increase susceptibility to infection by pathogens. High radiation doses cause severe illness and often lead to death.

The malevolent use of radioactive materials by terrorists might be in the form of an improvised nuclear device (IND), a radiological dispersion device (RDD), such as a "dirty bomb", or a radiological emission device (RED).

Nonstate Acquisition, Production, and Use of CBRN Agents

CBRN agents have distinct properties and capacity to cause profound effects on individuals, society, and the environment, making them attractive weapons to terrorists and criminals. Their effects can be seen immediately or years after release, producing a spectrum of illnesses, from rapid lethality, transient incapacitation to lifelong, chronic disease. They can be targeted to individuals (assassination and murder) or to populations (mass casualties and economic disruption). Because many of these agents occur infrequently or are unseen in nature, medical and veterinary practitioners and environmental protection agencies may have limited experience in recognizing them in the event of a deliberate release, limiting treatment and further enhancing the spread/contamination by the agent, and reducing the capability of law enforcement to locate and adequately examine a crime scene.

To date, all known nonstate-sponsored CBRN attacks have been committed by cults, sects, or individuals using crude preparations and/or delivery methods. In spite of the apparent attraction of CBRN terrorism, there have been no confirmed CBRN attacks by known terrorist groups such as al-Qa'ida, which questions the feasibility of this attack strategy [14]. Developing and deploying CBRN material

requires access to source material and the appropriate technical skills to produce harmful quantities of agent while minimizing the likelihood of the operator being inadvertently harmed by the agent. In spite of this, publicly available anarchist and terrorist literature describe methods for the production of some harmful chemical and biological agents and some source materials can be purchased over-the-counter or occur naturally in the environment. For example, radioactive sources that could be used for malevolent purposes can be purchased or stolen from medical institutions or industrial sites.

Recognizing a CBRN Scene

Indicators of a CBRN terrorist attack include mass casualties with symptoms typical of exposure to a CBRN agent and claims of responsibility. The covert release of CBRN agents, particularly biological and radiological agents, may only be recognized as a result of affected people presenting for medical treatment. Misdiagnosis and delayed diagnosis, including failure to consider an "unusual" cause, such as an act of terrorism, reduce the likelihood that a terrorist event will be detected and investigated. Locating the site of the release and the source of the agent used may require a combined effort by law enforcement investigators and public health epidemiologists, who are expert at tracing the source of disease outbreaks. The collaborative interaction between law enforcement and public health officials following the biological attacks in the United States in October 2001 involving *B. anthracis* was an important measure in the response to this act of bioterrorism [15]. Chemical weapons are more likely to produce immediate effects, so identifying the crime scene and the type of weapon used is often more straightforward.

Efforts by law enforcement agencies to identify and interdict planning processes for CBRN terrorism is a key pillar in preventing CBRN attacks. In these cases, law enforcement agencies are required to recognize the indicators of a CBRN production/preparation facility (clandestine laboratory), as well as other indicators of planning, such as recruitment of scientists from universities or legitimate laboratories and acquisition of laboratory equipment and terrorist training literature/recipes. Crime scene investigators need to be aware of the indicators of clandestine laboratories involved in the preparation of materials other than illicit drugs. Key indicators include dedicated equipment (centrifuges, glassware, and autoclaves), unusual laboratory chemicals (agar, culture media, and chemical precursors), and possible dispersion devices (explosives, gas cylinders, and pressurized sprayers). Relatively simple "back yard" laboratories might allow the preparation of some simple biological and chemical agents, while more complex laboratories, requiring commercial laboratory equipment, would be required to prepare most biological and CW agents, particularly if large-scale or high-purity preparations are the desired end product. It is technically infeasible to manufacture radiological material in a clandestine laboratory as such preparations require access to a nuclear reactor. Nevertheless, the assembly of an RDD using illicitly acquired radioactive material and conventional explosives might be achieved in a clandestine laboratory. Specific laboratory equipment for safe handling of radioactive materials will most likely include lead bricks for shielding during storage/preparation and radiation detectors.

Responding to a CBRN-Contaminated Crime Scene

The greatest concern to the investigator is the hazards of the CBRN environment. Law enforcement agencies with technical expertise in CBRN agents use information gathered from the crime scene to determine the likely impact and risks to crime scene examiners and the general public [16]. The need for personal protective equipment (PPE) can severely limit the capability of crime scene investigators to gather appropriate and adequate evidence. If CBRN agents have been released in a public place, there will be pressure to rapidly decontaminate the site to limit the potential to have further impact on the environment and to restore normal operations. The greatest challenge presented to investigators following a covert act of CBRN terrorism may be identifying and containing the crime scene. Further, there may be multiple crime scenes, including where the agent was manufactured and the site(s) of release.

A number of field-deployable detectors and screening kits exist that can rapidly identify the presence of explosive, radiological, chemical, and biological agents at a CBRN scene, though the accuracy and sensitivity of these kits for specific

agents is highly variable. With that said, detectors and kits are constantly evolving and improving [17–19].

Safety within the CBRN Crime Scene

Operating in a CBRN crime scene is hazardous and requires the use of PPE. Four types of PPE are usually described, with "level A" offering the most protection to "level D" offering the least. In selecting the most appropriate PPE to use, investigators must weigh up the degree of protection over flexibility and utility. For example, level A suits, which are the fully encapsulated airtight chemical resistant suits, can only be used for approximately 20 min at a time. Fatigue and heat are the major risk factors to the user. Safe operation requires extensive training and practice.

In addition to the use of appropriate PPE, particular care must be taken with procedures to avoid contamination outside the "hot zone". In the analysis of a crime scene involving CBRN or other hazardous material, the crime scene is divided into three zones: the "hot zone" is the hazardous region of the crime scene where evidence is collected, the "warm zone" is the boundary between the hot and cold zone where decontamination takes place, and the "cold zone" marks the outer perimeter of the incident site (the safe zone). Material and persons should not leave the warm zone unless they are fully decontaminated. If there is potential that the agent has been widely dispersed, the distance between the cold zone and the hot zone may be significant. For investigators using highly restrictive PPE, such as a level A fully encapsulated suit, it may not be possible to perform effective examination of an extensive hot zone following a CBRN incident. In these cases, it may require multiple entries into the hot zone to collect adequate evidence. In some cases, it may be more practical to use robotics to survey and collect evidence from within the hot zone.

Collecting Evidence (Sampling)

The CBRN agent itself will need to be examined as evidence, as well as other material, including equipment, computers, and documents. In addition, the crime scene will need to be examined for standard evidence, including DNA, fingerprint, hair, fiber, and handwriting.

When collecting evidence within the hot zone, crime scene investigators should work in pairs, using the "clean man/dirty man" approach where one person handles the contaminated items (the "dirty man"), whereas the other (the "clean man") only handles unused items and is responsible for stand off recording, including photography [20]. The clean man/dirty man approach reduces the likelihood of cross-contamination between samples and also enhances the safety of the operators within the hot zone. A number of factors need to be considered in the decision making that takes place around the collection of evidence from within the hot zone and the significance of different factors will vary from case to case [21]. In all the cases, a strategy of approach to the crime scene should be developed prior to entry to maximize effective use of the limited time within the hot zone. Crime scene examiners need to prioritize the types of items to collect, where to sample within the scene, how much CBRN material to collect and determine the types of vessels that are suitable for the storage and/or transportation of CBRN samples. As CBRN agents may not be visible to the naked eye, a collection approach that maps the crime scene may be required. A number of field detectors for chemical and radiological agents will allow the scene examiner to precisely determine the location of the agent for sampling, though this is not the case for biological agents. Further, adequate samples will need to be collected from within and around the crime scene to determine the background levels. These data will also be important for site remediation and for declaring that the previously contaminated site is clean.

An effective forensic investigation relies on ensuring the integrity of the evidence from collection to analysis and through to trial. For CBRN investigations, integrity is considered on multiple levels, including maintaining the physical integrity of the CBRN agent, as well as avoiding compromise through inadvertent contamination and the breaking of chain of custody. Many CBRN agents are unstable, so sampling, packaging, and transportation should be performed in a manner that best maintains the integrity of the material, while minimizing risk of accidental exposure to investigators and laboratory staff. Some standards exist, including from the American Society for Testing and Materials (ASTM) International [22]. Following a covert release, critical evidence, particularly clinical samples from the affected individuals, may have been collected and

analyzed without law enforcement involvement or adherence to chain-of-custody principles. Furthermore, detailed analysis of esoteric CBRN agents may not be within the realm of analyses performed in police forensic laboratories, which will require evidence to be referred to medical, military, research, or even private industry laboratories which may not have the same accreditation standards as standard forensic laboratories.

Evidence Storage

The rarity and nature of CBRN criminal and terrorism events means that investigations tend to be extensive and protracted. This may necessitate long-term storage of the agents and contaminated items, including biological specimens, computers, and vehicles. Many CBRN agents, or the solvents in which they are collected, can react with standard evidence storage containers (notably plastic bottles or bags), potentially resulting in the loss of evidence and harm to people and the environment. In addition, some plastics in storage bags and tubes can absorb chemical and biological agents, thus reducing the yield of agent for analysis or altering its properties. As most CBRN agents are unstable, they will degrade over time, losing their viability or potency and even changing their chemical properties. In some cases, it is possible to treat the agents to minimize loss (such as freezing, drying, or immobilizing in a matrix), but all of these treatments will alter the properties of the agent from its original condition.

Most standard forensic laboratories and law enforcement agencies do not have appropriate facilities for safe, long-term storage of CBRN agents and contaminated items [23]. These items may need to be stored in facilities dedicated to the analysis or research of specific material, where capacity to store large or multiple items may be lacking. Consideration must be given to maintaining security, evidence integrity, and chain-of-custody for storage of CBRN items in facilities that are not managed by law enforcement agencies.

Finally, once evidence has been processed and is not longer needed, specific arrangements must be made for the safe destruction and/or disposal of CBRN agents and contaminated items. The crucial factor that must be considered at the time of destruction is whether the item will be required for criminal proceedings. It may be the case that the destruction of the contaminated evidence may be in breach of the local, state, federal, or international legal requirement to provide this to the legal defense lawyers in the case of a criminal trial. Due care and attention must be taken when disposing of the contaminated evidence with the legal ramifications considered prior to this action.

Impact of CBRN Agents on Standard Forensic Evidence

The CBRN agents themselves, particularly radiological and chemical agents, can damage or destroy critical evidence, including biological evidence (DNA, hair, and fingerprints), as well as documents, computer data, and other reactive material. Standard forensic analysis for DNA, fingerprints, and handwriting from evidence gathered from a CBRN crime scene will be unsafe if evidence is not handled with physical protection or decontaminated. Most forensic laboratories do not have facilities that allow physical examination of contaminated evidence, so decontamination is often the only option. Chemical, heat, and irradiation methods used for decontamination can significantly alter surface materials as well as affect the integrity of latent fingerprints, DNA, and the characteristics of writing inks and dyes [24–31].

Forensic Analysis of CBRN for Intelligence

Specific disciplines are required for the direct examination of CBRN agents and their production/preparation processes, providing information that can be used for the attribution of CBRN crime or to exonerate the innocent. As with any forensic field, the application and interpretation of data is critical in excluding a particular source as the origin of the sample, as well as attributing the sample to a particular source. This is particularly important in discriminating malicious use of CBRN agents from natural causes.

"Microbial forensics" is the scientific discipline dedicated to analyzing evidence from an actual or threatened bioterrorism act, biocrime, or inadvertent microorganism/toxin release for attribution purposes, as well as to exonerate the innocent [32]. The forensic examination of microbial agents will be most effective if there is sufficient basic scientific information concerning microbial genetics, evolution, physiology,

and ecology of the agent [33]. Apart from providing greater confidence in the data for attribution purposes, understanding the "microbial background" of relevant geographic areas is critical for determining the source of the biological agent for consequence management and for remediation.

The analysis of chemical reagents, chemical degradation products, and equipments found at the scene can determine the types of agents being manufactured and the preparatory method. Chemical trace analyses, including isotopic ratios and mass spectrometry, are not only useful for chemical tracing but can support microbial forensics to determine the origin of source material (such as toxic plants) and the origins of growth media.

"Nuclear forensics" is the branch of science that seeks to examine the nature, use, and origin of intercepted illicit nuclear or radioactive materials. The analysis seeks to produce a characteristic energetic "signature" for the material as evidence for attribution. It can be applied to investigations involving the malevolent use of nuclear materials and is instrumental in determining adherence to international safeguards for illicit trafficking. Nuclear forensic analyses aim to provide legally admissible evidence that could lead to prosecution of the offenders involved with trafficking of the illicit material. Thus, it assists law enforcement agencies in the fight against illicit trafficking in nuclear and radioactive material in order to minimize and/or eliminate the likelihood of such materials being used by terrorists.

Acknowledgment

The information in this chapter has been obtained from open-source reports and publications

References

[1] Dana, A.S. (2007). *The National Biodefense Analysis and Countermeasures Centre: Issues for Congress*, RL32891, Congressional Research Service Report for Congress.

[2] Alibek, K. & Handelman, S. (1999). *Biohazard: The Chilling True Story of the Largest Covert Biological Weapons Program in the World – Told from the Inside by the Man who Ran it*, Dell Publishing, New York.

[3] Carus, W.S. (2002). *Bioterrorism and Biocrimes. The Illicit Use of Biological Agents Since 1900*, Fredonia Books, Amsterdam.

[4] Török, T.J., Tauxe, R.V., Wise, R.P., Livengood, J.R., Sokolow, R., Mauvais, S., Birkness, K.A., Skeels, M.R., Horan, J.M. & Foster, L.R. (1997). A large community outbreak of salmonellosis caused by intentional contamination of restaurant salad bars, *The Journal of the American Medical Association* **278**, 389–395.

[5] Vale, A. (2005). What lessons can we learn from the Japanese sarin attacks? *Przeglad Lekaski* **62**, 528–532.

[6] CDC (2001). Update: Investigation of anthrax associated with intentional exposure and interim public health guidelines, *Centers for Disease Control and Prevention, Morbidity and Mortality Weekly Report (MMWR) Weekly* **50**, 889–893.

[7] Hoffmaster, A.R., Fitzgerald, C.C., Ribot, E., Mayer, L.W. & Popovic, T. (2002). Molecular subtyping of *Bacillus anthracis* and the 2001 bioterrorism-associated anthrax outbreak, United States, *Emerging Infectious Diseases* **8**, 1111–1116.

[8] Ewell, E.S. (1998). *NIS nuclear smuggling since 1995: a lull in significant cases? The Nonproliferation Review* **5**(3), 119–125.

[9] Sidell, F.R., Patrick, W.C. & Dashiell, T.R. (2003). *Jane's Chem-Bio Handbook*, Jane's Information Group, Surrey.

[10] Romano, J.A., Lukey, B.J. & Salem, H. (2008). *Chemical Warfare Agents: Chemistry, Pharmacology, Toxicology and Therapeutics*, CRC Press, Boca Raton.

[11] Anderson, B., Friedman, H. & Bendinelli, M. (2006). *Microorganisms and Bioterrorism*, Springer, New York.

[12] Loveland, W., Morrissey, D.J. & Seaborg, G.T. (2006). *Modern Nuclear Chemistry*, Wiley-Interscience, New Jersey.

[13] Byrnes, M.E., King, D.A. & Tierno, P.M. (2003). *Nuclear, Chemical and Biological Terrorism. Emergency Response and Public Protection*, CRC Press, Boca Raton.

[14] Quillen, C. (2007). Three explanations for al-Qaeda's lack of a CBRN attack, *Terrorism Monitor* **5**, 7–9.

[15] Butler, J.C., Cohen, M.L., Friedman, C.R., Scripp, R.M. & Watz, C.G. (2002). Collaboration between public health and law enforcement: new paradigms and partnerships for bioterrorism planning and response, *Emerging Infectious Diseases* **8**, 1152–1156.

[16] Morrish, B.C. & Wegner, E. (2008). The Australian CBRN Data Centre, *Microbiology Australia* **29**, 70–71.

[17] Hoile, R., Yuen, M., James, G. & Gilbert, G.L. (2007). Evaluation of the rapid analyte measurement platform (RAMP) for the detection of *Bacillus anthracis* at a crime scene, *Forensic Science International* **171**, 1–4.

[18] Lim, D.V., Simpson, J.M., Kearns, E.A. & Kramer, M.F. (2005). Current and developing technologies for monitoring agents of bioterrorism and biowarfare, *Clinical Microbiology Reviews* **18**, 583–607.

[19] Sun, Y. & Ong, K.Y. (2005). *Detection technologies for chemical warfare agents and toxic vapors*, CRC Press, Boca Raton.

[20] Douglas, J. (2006). Beecher Forensic application of microbial culture analysis to identify mail intentionally

contaminated with *Bacillus anthracis* spores, *Applied and Environmental Microbiology* **72**, 5304–5310.
[21] Budowle, B., Schutzer, S.E., Burans, J.P., Beecher, D.J., Cebula, T.A., Chakraborty, R., Cobb, W.T., Fletcher, J., Hale, M.L., Harris, R.B., Heitkamp, M.A., Keller, F.P., Kuske, C., Leclerc, J.E., Marrone, B.L., McKenna, T.S., Morse, S.A., Rodriguez, L.L., Valentine, N.B. & Yadev, J. (2006). Quality sample collection, handling, and preservation for an effective microbial forensics program, *Applied and Environmental Microbiology* **7**, 6431–6438.
[22] ASTM (2006). *Standard Practices for Bulk Sample Collection and Swab Sample Collection of Visible Powders Suspected of being Biological Agents from Nonporous Surfaces*, ASTM International Standard, pp. E2458–E2406.
[23] Roffey, P., Norman, K. & Royds, D. (2008). A mobile laboratory for real time analysis during forensic operations, *Microbiology Australia* **29**, 91–94.
[24] Wilkinson, D.A., Hancock, J., Lecavalier, P. & McDiarmid, C. (2005). The recovery of fingerprint evidence from crime scenes contaminated with chemical warfare agents, *Journal of Forensic Identification* **55**, 326–361.
[25] Hoile, R., Walsh, S.J. & Roux, C. (2007). Bioterrorism: processing contaminated evidence, the effects of formaldehyde gas on the recovery of latent fingerprints, *Journal of Forensic Sciences* **52**, 1097–1102.
[26] Solazzo, C., Erhardt, D., Marte, F., von Endt, D. & Tumosa, C.S. (2004). Effects of chemical and biological warfare remediation agents on the material of museum objects, *Applied Physics A* **79**, 247–252.
[27] Ramotowski, R.S. & Regen, E.M. (2005). The effect of electron beam irradiation on forensic evidence. 1. Latent print recovery on porous and non-porous surfaces, *Journal of Forensic Sciences* **50**, 298–306.
[28] Ramotowski, R.S. & Regen, E.M. (2007). Effect of electron beam irradiation on forensic evidence. 2. Analysis of writing inks on porous surfaces, *Journal of Forensic Sciences* **52**, 604–609.
[29] Tumosa, C.S., Erhardt, D. & Solazzo, C. (2002). The effect on ballpoint pen and marker inks of chemical and electron beam remediation techniques for biological warfare agents, *Mid-Atlantic Association of Forensic Scientists (MAAFS) Newsletter* **30**, 5–8.
[30] Solazzo, C., Tumosa, C.S. & Erhardt, D. (2004). The effect of electron beam irradiation on ballpoint pen and marker inks, *Mid-Atlantic Association of Forensic Scientists (MAAFS) Newsletter* **32**, 13–16.
[31] Wilkinson, D.A., Sweet, D. & Fairly, D. (2007). Recovery of DNA from exhibits contaminated with chemical warfare agents: a preliminary study of the effects of decontamination agents and chemical warfare agents on DNA, *Canadian Society of Forensic Science Journal* **40**, 15–22.
[32] Budowle, B., Schutzer, S.E., Einseln, A., Kelley, L.C., Walsh, A.C., Smith, J.A., Marrone, B.L., Robertson, J. & Campos, J. (2003). Public health. Building microbial forensics as a response to bioterrorism, *Science* **301**, 1852–1853.
[33] Budowle, B., Beaudry, J.A., Barnaby, N.G., Giusti, A.M., Bannan, J.D. & Keim, P. (2007). Role of law enforcement response and microbial forensics in investigation of bioterrorism, *Croatian Medical Journal* **48**, 437–449.

Bronwyn C. Morrish, Paul E. Roffey, Victoria C. Gillman, Keith W. Norman, George J. Koperski, Serena F. Abbondante and Eric Wenger

Chemical Castration *see* Sex Offenders: Treatment of

Chemical Injury *see* Injury: Burns, Scalds, and Chemical

Chemical Warfare Agents

Introduction

Toxic industrial chemicals (TICs) are also referred to as *toxic industrial materials* (*TIMs*). TICs are the hazardous industrial chemicals that are manufactured, stored, transported, and used throughout the world. TICs include chemicals in the gas, liquid, or solid state. They can be chemical hazards including but not wholly biological hazards such as carcinogens, reproductive hazards and, corrosives or physical hazards such as being flammable, explosive, or reactive. In addition to their commercial uses, TICs have been used as stimulants for chemical weapons, where they can aid in determining the effectiveness of respirators, detectors, and other defensive measures. Concern arises periodically, however, that TICs could be used

directly by individuals or groups for chemical terrorism or could be diverted by nations into use in the production of chemical weapons [1].

A TIM is any substance that, when supplied in a given quantity, produces a toxic effect [2]. TICs and TIMs are produced in substantial amounts for legitimate purposes. A classic example is chlorine gas, which is used in multiple industrial applications such as water purification, sanitization, disinfection, bleaching, and in the manufacture of plastics.

Chemical warfare agents (CWAs) are a subset of TICs. There are no legitimate nonoffensive uses for CWAs. The specific chemical compounds of concern are detailed in the schedules issued by the Organization for the Prohibition of Chemical Weapons (OPCW). The chemical weapons convention (CWC) developed three lists of chemicals that are monitored and regulated under the CWC. The Schedules 1 and 2 chemicals as outlined in the CWC are listed in Tables 1 and 2, respectively.

The following criteria were used to determine whether a toxic chemical or precursor should be listed in Schedule 1:

1. It has been developed, produced, stockpiled, or used as a chemical weapon as defined in Article II of the CWC.
2. It poses otherwise a high risk to the object and purpose of the CWC by virtue of its high potential for use in activities prohibited under the CWC because one or more of the following conditions are met:
   a. It possesses a chemical structure closely related to that of other toxic chemicals listed in Schedule 1, and has, or can be expected to have, comparable properties.
   b. It possesses such lethal or incapacitating toxicity as well as other properties that would enable it to be used as a chemical weapon.
   c. It may be used as a precursor in the final single technological stage of production of a toxic chemical listed in Schedule 1, regardless of whether this stage takes place in facilities, in binary munitions, or elsewhere.
3. It has little or no use for purposes not prohibited under the CWC.

The following criteria were used to determine whether a toxic chemical or a precursor should be listed in Schedule 2:

- It poses a significant risk to the object and purpose of the CWC because it possesses such lethal or incapacitating toxicity as well as other properties that could enable it to be used as a chemical weapon.
- It may be used as a precursor in one of the chemical reactions at the final stage of formation of a chemical listed in Schedule 1 or Schedule 2, part A.
- It poses a significant risk to the object and purpose of the CWC by virtue of its importance in the production of a chemical listed in Schedule 1 or Schedule 2, part A.
- It is not produced in large commercial quantities for purposes not prohibited under this the CWC.

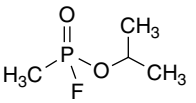
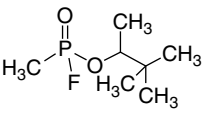
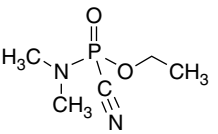
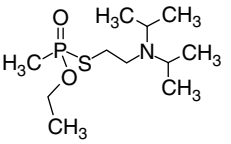
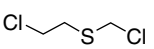
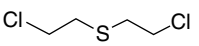
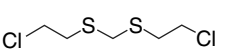
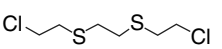
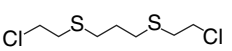
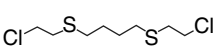
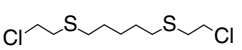
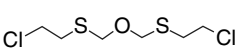
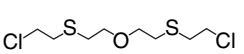
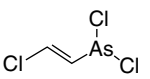
A toxic chemical is described by the OPCW as "Any chemical which through its chemical action on life processes can cause death, temporary incapacitation or permanent harm to humans or animals. This includes all such chemicals, regardless of their origin or of their method of production, and regardless of whether they are produced in facilities, in munitions or elsewhere" [3]. Chemical warfare (CW) should therefore be considered as combat that utilizes the toxic properties of chemical substances to kill, injure, or incapacitate an enemy.

This article provides an overview of CWAs, their detection in the field, and the laboratory-based methods required for the forensic analysis of a suspected CWA. It does not cover the effect of CWAs on traditional evidence or decontamination of traditional evidence. CWAs are organized into several categories according to the manner in which they affect the human body, which also affect the types of analysis used to identify the material. CWAs are also classified according to their persistence, a measure of the length of time that a chemical agent remains effective after dissemination. The names and number of categories vary slightly from source to source, but in general the basic classes of CWAs are identified in Table 3.

Analysis for CW Agents

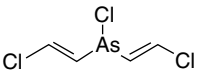
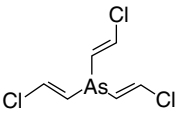
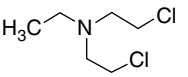
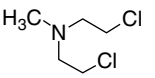
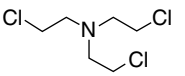
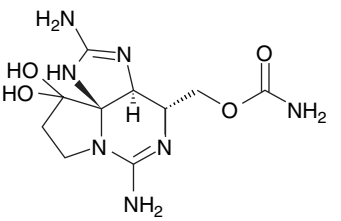
CWAs can be analyzed by numerous methods each of which has varying degrees of success. There are a number of issues that need to be taken into consideration when analyzing CWAs, from the chemistry

Table 1 Schedule 1 chemicals

A.	Toxic chemicals		CAS registry number
1.	*O*-Alkyl (≤C10, incl. cycloalkyl) alkyl (Me, Et, *n*-Pr or *i*-Pr)-phosphonofluoridates,		
	e.g., Sarin: *O*-isopropyl methylphosphonofluoridate		107-44-8
	Soman: *O*-pinacolyl methylphosphonofluoridate		96-64-0
2.	*O*-Alkyl (<C10, incl. cycloalkyl) *N*, *N*-dialkyl (Me, Et, *n*-Pr or *i*-Pr) phosphoramidocyanidates		
	e.g., Tabun: *O*-ethyl *N*, *N*-dimethyl phosphoramidocyanidate		77-81-6
3.	*O*-Alkyl (H or <C10, incl. cycloalkyl) *S*-2-dialkyl(Me, Et, *n*-Pr or *i*-Pr)-aminoethyl alkyl(Me, Et, *n*-Pr or *i*-Pr) phosphonothiolates and corresponding alkylated or protonated salts		
	e.g., VX: *O*-ethyl *S*-2-diisopropylaminoethyl methyl phosphonothiolate		50782-69-9
4.	*Sulfur mustards*		
	2-Chloroethylchloromethylsulfide		2625-76-5
	Mustard gas: *bis*(2-chloroethyl)sulfide		505-60-2
	Bis(2-chloroethylthio)methane		63869-13-6
	Sesquimustard: 1,2-*bis*(2-chloroethylthio)ethane		3563-36-8
	1,3-*Bis*(2-chloroethylthio)-*n*-propane		63905-10-2
	1,4-*Bis*(2-chloroethylthio)-*n*-butane		142868-93-7
	1,5-*Bis*(2-chloroethylthio)-*n*-pentane		142868-94-8
	Bis(2-chloroethylthiomethyl)ether		63918-90-1
	O-Mustard: *bis*(2-chloroethylthioethyl)ether		63918-89-8
5.	*Lewisites*		
	Lewisite 1: 2-chlorovinyldichloroarsine		541-25-3

(*continued overleaf*)

Table 1 (*continued*)

A.	Toxic chemicals		CAS registry number
	Lewisite 2: *bis*(2-chlorovinyl)chloroarsine		40334-69-8
	Lewisite 3: *tris*(2-chlorovinyl)arsine		40334-70-1
6.	*Nitrogen mustards* HN1: *bis*(2-chloroethyl)ethylamine		538-07-8
	HN2: *bis*(2-chloroethyl)methylamine		51-75-2
	HN3: *tris*(2-chloroethyl)amine		555-77-1
7.	Saxitoxin		35523-89-8
8.	Ricin		9009-86-3

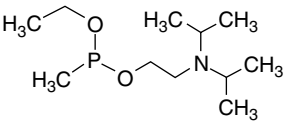
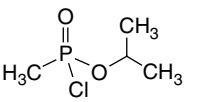
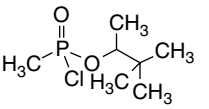
B.	Precursors		CAS registry number
9.	Alkyl (Me, Et, *n*-Pr or *i*-Pr) phosphonyldifluorides, e.g., DF: methylphosphonyldifluoride		676-99-3
10.	*O*-Alkyl (H or <C10, incl. cycloalkyl) *O*-2-dialkyl(Me, Et, *n*-Pr or *i*-Pr)-aminoethyl alkyl(Me, Et, *n*-Pr or *i*-Pr) phosphonites and corresponding alkylated or protonated salts e.g., QL: *O*-ethyl *O*-2-diisopropylaminoethyl methylphosphonite		57856-11-8

Table 1 (*continued*)

A.	Toxic chemicals		CAS registry number
11.	Chlorosarin: *O*-isopropyl methylphosphonochloridate		1445-76-7
12.	Chlorosoman: *O*-pinacolyl methylphosphonochloridate		7040-57-5

of the specific agent to the matrix that the agent is to be analyzed in. Analysis of an unknown chemical sample can yield significant information from the type of CWA used, the method of manufacture utilized to produce the CWA, as well as the presence or absence of degradation products. The synthetic method used can be identified through the presence of route-specific reaction markers. This information can give an indication of the sophistication of the manufacturing process, the level of purity of the final material, and some detail on precursors used. In turn, this provides information about the synthetic laboratory, the equipment, and the knowledge and skills of the chemist(s) involved.

Identification of degradation products provides an estimate of the time since the chemical agent had been manufactured and the length of time since its dissemination. Identification of the CWA, its precursors, reaction markers, and degradation products is crucial in an investigation to establish links between the agent(s), manufacture method, location of manufacture and, if possible, the person, persons, or group responsible. Rapid identification of the agent and other toxic compounds assists in developing a successful decontamination procedure for victims and evidence.

The persistence of an agent at the site of release is dependant upon many factors including the type of agent used, the location of the release, the climatic conditions, the mode of dissemination, and whether decontamination procedures have commenced. With this in mind, the presence of the CWA in its original form cannot always be expected. Degradation products may be present; however, it is likely that these will be compromised and unlikely to be in pure form.

The analysis of an unknown sample, either in a laboratory or at a crime scene, should take into consideration preparation of the unknown sample prior to analysis, which will then assist in deciding what chromatographic or spectrometric technique should be utilized. In general, chromatographic techniques are applied within specific detectors and certain spectrometric techniques are used to identify a number of chemical candidates.

Presumptive Analysis – Detectors for CWAs

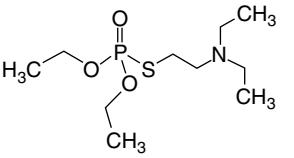
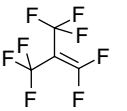
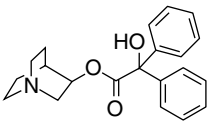
There are a number of handheld electronic detectors and colorimetric indicators that are utilized for presumptive identification of CWAs, TICs, and TIMs. Detector responses are used to establish what procedures are to be used in order to respond to a specific chemical incident.

Some examples of detection principles utilized in the identification equipment are

- colorimetry
- electrochemical analysis
- ion mobility spectrometry (IMS)
- flame photometry
- photoionization
- enzyme-based techniques.

Described below are examples of these types of technologies and a brief overview of the chemistry behind the detection methodology [4]. One of the greatest limitations of all of the methods described below is the limited capability to decontaminate equipment. Each technique or piece of equipment should be decontaminated in such a way that this

Table 2 Schedule 2 chemicals

A.	Toxic chemicals		CAS registry number
1.	Amiton: *O*, *O*-diethyl *S*-[2-(diethylamino)ethyl] phosphorothiolate and corresponding alkylated or protonated salts		78-53-5
2.	PFIB: 1,1,3,3,3-pentafluoro-2-(trifluoromethyl)-1-propene		382-21-8
3.	BZ: 3-quinuclidinyl benzilate (*)		6581-06-2

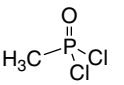
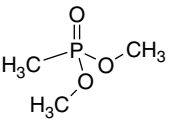
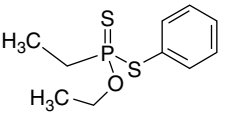
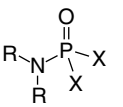
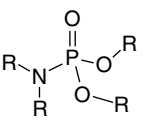
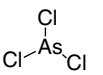
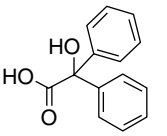
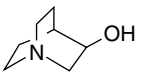
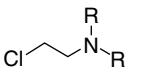
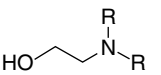
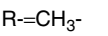
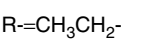
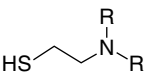
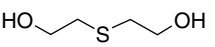
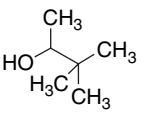
B.	Precursors		CAS registry number
4.	Chemicals, except for those listed in Schedule 1, containing a phosphorus atom to which is bonded one methyl, ethyl, or propyl (normal or iso) group but not further carbon atoms, e.g., Methylphosphonyl dichloride		676-97-1
	Dimethyl methylphosphonate		756-79-6
	Exemption: Fonofos: *O*-ethyl *S*-phenyl ethylphosphonothiolothionate		944-22-9
5.	*N*, *N*-Dialkyl (Me, Et, *n*-Pr, or *i*-Pr) phosphoramidic dihalides		
6.	Dialkyl (Me, Et, *n*-Pr or *i*-Pr) *N*, *N*-dialkyl(Me, Et, *n*-Pr, or *i*-Pr)-phosphoramidates		
7.	Arsenic trichloride		7784-34-1
8.	2,2-Diphenyl-2-hydroxyacetic acid		76-93-7

Table 2 (*continued*)

A.	Toxic chemicals		CAS registry number
9.	Quinuclidin-3-ol		1619-34-7
10.	*N*, *N*-Dialkyl (Me, Et, *n*-Pr, or *i*-Pr) aminoethyl-2-chlorides and corresponding protonated salts		
11.	*N*, *N*-Dialkyl (Me, Et, *n*-Pr, or *i*-Pr) aminoethane-2-ols and corresponding protonated salts		
	Exemptions: *N*, *N*-dimethylaminoethanol and corresponding protonated salts		108-01-0
	N, *N*-Diethylaminoethanol and corresponding protonated salts		100-37-8
12.	*N*, *N*-Dialkyl (Me, Et, *n*-Pr, or *i*-Pr) aminoethane-2-thiols and corresponding protonated salts		
13.	Thiodiglycol: *bis*(2-hydroxyethyl)sulfide		111-48-8
14.	Pinacolyl alcohol: 3,3-dimethylbutan-2-ol		464-07-3

action will not compromise the detector. For example, the full submersion of a detector that draws air into the detector may have a disastrous effect on the detector.

Colorimetric Detectors

Colorimetric detectors are based upon a color change either in the visible or ultraviolet spectrum, which occurs in the presence of the target agent. These colorimetric indicators are the fastest, cheapest, lightest, and easiest type of detector. A number of Western military forces employ two types of paper colorimetric measurements, commonly referred to as *M8 and M9 detection paper*.

The main limitation of the identification papers is that they have to be in intimate contact with a liquid sample to work. Neither M8 nor M9 paper detects CWAs in the vapor phase. These papers are highly subject to false positives from common household materials including pesticides, smoke, petrol, strong bleaches, acid, and common solvents.

The M8 paper, as seen in Figure 1, detects and also differentiates between V-type nerve agents (e.g., VX), G-type nerve agents (e.g., sarin and soman), and H-type vesicants (e.g., sulfur mustard). The beige paper contains two dyes and an acid–base (pH) indicator. Together, these change to yellow when in contact with liquid sarin, green when in contact with liquid VX, and red when in contact with a liquid mustard agent. The principal difference between M8 and M9 paper is that M9 paper shows only a generic red spot when there is a positive indication of any CWA [5].

Electrochemical Detection

Electrochemical or chemiresistor detectors work on the principle that an electrical current changes when a CWA is present. The most common mechanism for an electrochemical gas sensor is a conducting wire or filament that is coated with a reactive material that oxidizes rapidly when it comes into contact with a chemical such as a CWA. The oxidation of the surface material exposes the conducting wire to air and the electrical resistance increases substantially.

Table 3 CW agents and their physical properties, which include the mode and rate of action

Properties of chemical warfare agents									
	Chemical and physical properties (at 20 °C)								
Agent class	Chemical agent name; formula	Symbol	CAS number	Appearance	Odor	Rate of detoxification	Eye and skin toxicity	Routes	Physiological action	Rate of action
Choking	Chlorine; Cl_2	CL	7782-50-5	Greenish yellow gas or amber liquid	Pungent, suffocating, bleach-like odor	–	–	Inhalation; skin absorption (liquid, vapor); ingestion	Respiratory poison; bronchitis and accumulation of fluid in the lungs; liquid causes burns and blisters	Rapid
	Phosgene; $COCl_2$	CG	75-44-5	Colorless gas (foglike in initial concentration – becomes colorless as it spreads)	Newly mown hay, highly toxic suffocating odor	Not detoxified; cumulative	Mild eye irritation	Inhalation	Damages and floods lungs	Delayed; immediate irritation in high concentration; low concentration no ill effects for 3 h or more
Blister (vesicants)	Phosgene oxime; $Cl_2C=NOH$	CX	1794-86-1	Colorless solid or liquid	Intense, penetrating, disagreeable and violently irritating odor	–	Violently irritating to eyes; low concentration cause tearing, inflammation and temporary blindness; liquid on skin is corrosive	Inhalation; skin absorption (liquid, vapor); Ingestion	Violently irritates mucous membrane of eyes, nose, and lungs	Rapid; immediate effects on contact

Sulfur mustard; $(ClCH_2CH_2)_2S$	HD	505-60-2	Colorless when pure, but normally a yellow to amber oily liquid; thickened agent has viscosity similar to honey	Garlic- or mustardlike odor; color and/or odor may vary for thickened agent	Very low; repeated exposures of HD are cumulative in their effects	Eyes very susceptible to low concentration; skin less so	Inhalation; skin absorption (liquid, vapor); ingestion	Blisters; destroys tissues; injures blood vessels	Delayed – 4–6 to 24 h to days (up to 12 days in rare cases)
Nitrogen mustard; $CH_3CH_2N(CH_2CH_2Cl)_2$	HN-1	538-07-8	Oily colorless to pale yellow liquid; salts are solid	Faint fishy or musty odor	Not detoxified; cumulative	Eyes susceptible to low concentration; less toxic to skin, incapacitation requires higher concentration	Inhalation; skin absorption (liquid, vapor); ingestion	Blisters; affect respiratory tract; destroys tissues; injures blood vessels	Delayed action – 12 h or longer
Nitrogen mustard; $CH_3N(CH_2CH_2Cl)_2$	HN-2	51-75-2	Pale amber to yellow oily liquid; salts are solid	Fruity odor in high concentrations. Low concentrations fishy odor similar to "soft soap"	Not detoxified	Toxic to eyes; blisters skin (greatest blistering power of nitrogen mustards in vapor form)	Inhalation; skin absorption (liquid, vapor); ingestion	Similar to HD; bronchopneumonia may occur after 24 hours	Skin effects delayed; 12 h or longer
Nitrogen mustard; $(ClCH_2CH_2)_3N$	HN-3	555-77-1	Colorless to pale yellow liquid which darkens during extended storage; salts are solid	Butter almond odor	Not detoxified; cumulative	Eyes very susceptible; skin less so	Inhalation; skin absorption (liquid, vapor); ingestion	Similar to HN-2	Serious effects same as for HD; minor effects sooner
Lewisite; $ClCH{=}CHAsCl_2$	L; L-1	541-25-3	Colorless to brownish liquid; thickened agent has viscosity similar to honey	Geranium-like odor; no odor when pure; thickened agent odor may vary from unthickened	Body does not detoxify L	Even limited concentrations causes extreme irritation of eyes; Intimate contact exposure causes severe corneal damage; same blistering action on skin as HD	Inhalation; skin absorption (liquid, vapor); ingestion	Similar to HD plus may cause systemic poisoning	Rapid; more rapid than N-mustards

(*continued overleaf*)

Table 3 (*continued*)

Properties of chemical warfare agents									
	Chemical and physical properties (at 20 °C)								
Agent class	Chemical agent name; formula	Symbol	CAS number	Appearance	Odor	Rate of detoxification	Eye and skin toxicity	Routes	Physiological action	Rate of action
Blood	Hydrogen cyanide; HCN	AC	74-90-8	Colorless liquid that is highly volatile	Bitter almonds or peachy kernels odor; not detectable by all individuals, even at lethal concentrations	Rapid; 0.017 mg $kg\,min^{-1}$	Moderate	Inhalation; skin absorption (liquid, vapor); ingestion	Interferes with use of oxygen by body tissues; accelerates rate of breathing	Very rapid; incapacitation within 1–2 min; death can occur in 15 min after receiving lethal dose
	Cyanogen chloride; CNCl	CK	506-77-4	Colorless gas and liquid (liquid below 12.8 °C)	Pungent biting pepperish odor; odor can go unnoticed because of discom-fort/irritation to mucous membranes	0.02–0.1 mg $kg\,min^{-1}$	Low lachrymatory; highly irritating to eyes, upper respiratory tract, and lungs; can cause dry land drowning	Inhalation; skin absorption (liquid, vapor); ingestion	Chokes, irritates, causes slow breathing rate	Immediate intense irritation; systemic effect believed to be due to CK conversion to AC in the body
	Arsine; AsH_3	SA	7784-42-1	Colorless gas	Mild garlic like odor	–	None	Inhalation	Damages blood, liver, and kidneys	Delayed action – 2 h to as much as 11 days
Nerve	Tabun; $C_2H_5OP(O)(CN)N(CH_3)_2$	GA	77-81-6	Clear, colorless to brown liquid	None when pure; impurities may give a faintly fruity odor	Slight but definite	Very high; much greater through eyes than skin but liquid penetrates skin easily	Inhalation; skin absorption (liquid, vapor); ingestion	Cessation of breath and death may follow	Very rapid
	Sarin; $CH_3P(O)(F)OCH(CH_3)_2$	GB	107-44-8	Clear, colorless liquid	None	Low; essentially cumulative	Very high; much greater through eyes than skin but liquid and vapor penetrates skin easily; lethal dose is 1.7 g per person	Inhalation; skin absorption (liquid, vapor); ingestion	Cessation of breath and death may follow	Very rapid; death usually occurs within 15 min after absorption of fatal dose

Vomiting	Soman; $CH_3P(O)(F)OCH(CH_3)C(CH_3)_3$	GD	96-64-0	Clear, colorless to brown liquid; thickened agent has viscosity similar to honey and may vary in color	Slight camphor odor and gives off a colorless vapor	Low; essentially cumulative	Very high; much greater through eyes than skin but liquid and vapor penetrates skin easily; lethal dose is 0.35 g per person on bare skin	Inhalation; skin absorption (liquid, vapor); Ingestion	Cessation of breath and death may follow	Very rapid; death usually occurs within 15 min after absorption of fatal dose
	VX; $CH_3P(O)(OCH_2CH_3)SCH_2CH_2N-[CH(CH_3)_2]_2$	VX	50782-69-9	Clear, colorless to amber oily liquid; similar in appearance to motor oil; thickened agent has viscosity similar to honey	None	Low; essentially cumulative	Extremely toxic by skin and eye absorption; does not injure but penetrates rapidly; about 100× potent as GB	Inhalation; skin absorption (liquid, vapor); ingestion	Produces casualties when inhaled or absorbed	Very rapid; death usually occurs within 15 min after absorption of fatal dose
Riot control (tear)	Adamsite; $C_6H_4(NH)(AsCl)C_6H_4$	DM	578-94-9	Light yellow to green crystals	No pronounced odor; vapors are irritating	Quite rapid in small amounts; incapacitating amounts lose their effects after about 30 min	Irritating; relatively nontoxic	Inhalation; ingestion	Like cold symptoms plus headache, vomiting, nausea	Very high; requires only about 1 min to temporarily incapacitate at a concentration of 22 mg m^{-3}
	Chloroacetophenone (mace); $C_6H_5C(O)CH_2Cl$	CN	532-27-4	Colorless to gray crystalline solid	Sharp irritating floral odor, similar to apple blossoms	Rapid; effects disappear in minutes; high concentration may cause skin irritation but usually disappears within a few hours	Temporary severe eye irritation; mild skin irritation	Inhalation	Lachrymatory; irritates respiratory tract	Instantaneous

(*continued overleaf*)

Table 3 (*continued*)

Agent class	Properties of chemical warfare agents									
	Chemical and physical properties (at 20 °C)									
	Chemical agent name; formula	Symbol	CAS number	Appearance	Odor	Rate of detoxification	Eye and skin toxicity	Routes	Physiological action	Rate of action
Incapacitating	2-Chloro-benzylidene malono-nitrile; $ClC_6H_4CH{=}C(CN)_2$	CS	2698-41-1	White crystalline solid; may also appear as a 1% solution in trioctylphosphite	Burnt to create colorless gas with pungent pepperlike odor	Quite rapid; incapacitating dosage loses effect in 5–10 min	Highly irritating; nontoxic	Inhalation; ingestion	Highly irritating; but nontoxic	Very rapid (max. effects in 20–60 s)
	Pepper spray (Oleoresin Capsicum); Capsaicin: $(CH_3)_2CHCH{=}CH(CH_2)_4CONHCH_2C_6H_3$-4-(OH)-3-($OCH_3$)	OC	8023-77-6 (404-86-4; capsaicin)	Varies depending on manufacturer; typically amber to light red appearance; mixture may include dye (capsaicin: orange red liquid or dark red solid)	Slight ethereal odor	–	Highly irritating	Inhalation; local eye/skin impacts	Highly irritating	Instantaneous
	Buzz; $(C_6H_5)_2C(OH)CO_2C_7H_{12}N.HCl$	BZ	6581-06-2	White crystalline solid	Odorless	From and ICt50 dose severe effects last 36 h; mild effects last 45 h	–	Inhalation; skin absorption (sld); ingestion	Fast heartbeat, dizziness, vomiting, dry mouth, blurred vision, stupor, and increasing random activity	Delayed action – 1–4 h



Figure 1 Example of M8 paper (http://www.saferamerica.com/productDetail.asp?categoryID=19&productID=365) [Reproduced with permission from Safer America.]

The change in current or increase in temperature is indicative of CWA exposure.

Other electrochemical detectors employ chemically selective membranes, which allow only certain chemical types to pass (CWAs), those that are required to complete a circuit. A change in the current or resistance of the solution on the detector side of the membrane indicates the presence of a CWA.

A newer type of chemiresistor instrument uses a quartz or silicon substrate that is coated with a conducting polymer. The degree of current change is dependent on the chemistry of the absorbing agent. The polymers provide limited specificity, so that the various classes of CWAs can be differentiated. The response time for electrochemical sensors is generally very fast, for example less than a minute, and often the measurement can be completed in seconds. This type of detector is limited to the detection of chemical vapors, and as such the detector response will be affected by the chemical properties of the agent in the environment. Figure 2 shows an example of an electrochemical detector.



Figure 2 Combination ion mobility spectrometry of ion mobility spectrometry and electrochemical detectors CHEMPRO 100 manufactured by the Finnish company, Environics Oy (http://www.environics.fi/) [Reproduced with permission from Environics Oy.]

Ion Mobility Spectrometry

Material deposited on a surface can be analyzed using IMS. The sample is rapidly thermally desorbed and the resultant vapors passed, via the inlet liner, into the ionization chamber. In the ionization chamber, a radioactive source generates ions from the sample. The ionized species are then electrically trapped at a gating grid so that all ions enter a drift tube at the same time. Inside the tube, the ions move through the drift tube toward the collector electrode under the influence of the electric field and against a drift flow. Included are small amounts of the internal calibrant 4-nitrobenzonitrile for explosives or nicotinamine for illicit drugs and CWAs. The ions travel at different speeds and arrive at the collector electrode with characteristic drift times. The drift times are inversely proportional to the characteristic reduced mobilities of the ions. CWAs, TICs, TIMs, illicit drugs, and explosives can be identified by analyzing directly the ratio of drift time to that of the internal calibrant. Figure 3 contains a schematic of the important features of an IMS.

The Ionscan 500DT manufactured by Smith Detection (http://www.smithsdetection.com/eng/1526.php) is an example of the IMS instrument, which is designed for fast and effective presumptive test. The instrument is extremely sensitive, with detection limits in the order of nanograms for detecting CWAs, TICs, TIMs and illicit drugs and down to picograms for explosives. A "positive" result is generated when an ion reaches the detector with a "drift time" similar to that of one of the standards. This generates a presumptive positive for the presence of CWAs, TICs, TIMs, illicit drugs, or explosives.

IMS is relatively specific; however, there are still compounds that can produce a false positive reading. A positive IMS result is strong indication that a

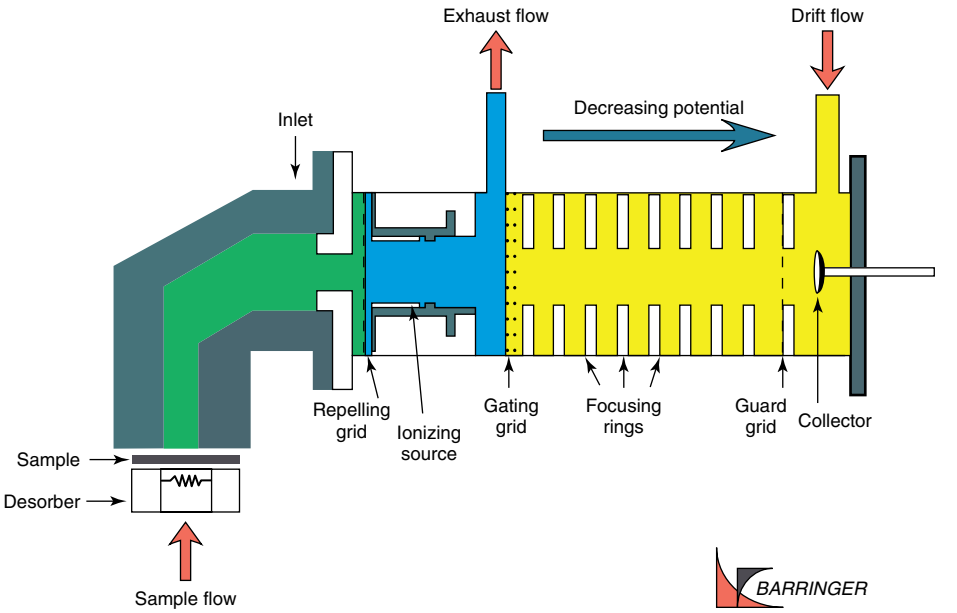


Figure 3 Schematic of the important features of an IMS



Figure 4 Bruker RAID-1 portable IMS system manufactured by Bruker Daltonics Inc. (http://www.bdal.de/cbrn-detection/chemical-detection/raid-1.html) [Reproduced with permission from Bruker Daltonics Inc.]



Figure 5 CAM portable IMS system, manufactured by Smiths Detection (http://www.smithsdetection.com/eng/CAM.php) [Reproduced with permission from Smiths Detection.]

particular target chemical is present; however, independent analysis by gas chromatography/mass spectroscopy (GC/MS) and/or gas chromatography/mass spectroscopy/dual flame photometric detection (GC/MS/DFPD) is required for confirmation. Figures 4 and 5 show examples of portable IMS systems used for CWA detection.

The downside of these technologies is that the detectors can be overwhelmed by high concentrations of the agent, then requiring several minutes of exposure to clean air in order to clear the agent from the detector so it can be returned to use.

Flame Photometry Detectors

In flame photometric detection (FPD) a small sample is ignited in a hydrogen flame, and the characteristic emission spectrum produced serves as a fingerprint for the atoms in the compounds analyzed. A quantitative reading of the amount of a certain element, such as phosphorous or sulfur, in a sample can be detected. Optical filters can be selected for specificity of a target material. A light-detecting element, typically a photodiode, recognizes patterns that correspond to CWAs, TICs, and TIMs. FPD detectors identify compounds in a vaporous state and as such are affected by the chemical properties of the agent in question, i.e., the vapor pressure of an agent at a specific temperature. To overcome this, scrapings from a surface can be taken and heated with the vapors collected and analyzed. An FPD detector can also



Figure 6 AP4C portable FPD system manufactured by Proengin (http://www.proengin.com/index.php?option=com_content&task=view&id=25&Itemid=32) [Reproduced with permission from Proengin.]

be combined in a laboratory setting with a GC to improve complex mixture separation. An example of a portable FPD unit is shown in Figure 6. FPDs are less influenced by high concentrations of an agent.

Photoionization Detectors

Photoionization detectors (PIDs) use ultraviolet light to ionize vapor or gas samples. A detector measures the amount of ions on the basis of a change in electrical current. PID systems are highly quantitative when compared with a calibrated, known sample and provide excellent sensitivity in such situations. Popular PID systems have very limited specificity, are highly subject to false positives in unknown or mixed environments, and are costly pieces of equipment. Nonetheless, for applications such as leak testing, PIDs are appropriate.

The limitation of the PID is that the compounds of interest must have an ionization potential less than the voltage of the lamp. If the ionization potential is greater than the voltage of the lamp, they will not be detected. PIDs can also be susceptible to failure and quenching in high humidity; for this reason a water vapor filter is normally fitted when in use. An example of a portable, handheld PID is shown in Figure 7.

Enzyme-Based Techniques

Enzyme or immunoassays approaches have been utilized for military and commercial CW detectors. Some enzyme-based CW detection systems exploit the tendency of organophosphate nerve agents to bind to acetylcholinesterase as a detection technique. Enzyme-linked immunoassays have been developed, much like the biological warfare agent counterpart,



Figure 7 Handheld PID MiniRAE 2000 manufactured by RAE Systems Inc. (http://www.raesystems.com/products/MiniRAE2000_986.html) [Reproduced with permission from RAE Systems Inc.]

with specificity for G-type nerve agents. Other systems exploit the natural enzyme that catalytically hydrolyzes or breaks down in the presence of water. Organophosphorous hydrolase can be incorporated into handheld assays or tickets. A pH sensitive probe reacts to change in acidity due to the hydrolysis of G-type nerve agents. The response can be as simple as a colorimetric pH indicator changing from red to blue, or a potentiometric electrode.

One of the limitations for the enzyme-based screening is the time taken for the assay to produce a result. The response time can be critical in cases where the agent can compromise the personal protective equipment being used by the first responder. A further limitation of the enzymatic assay is that it generates a result for a single source or location. This requires that another detector is utilized to detect the point source of contamination prior to using the enzymatic detection system. An example of an enzyme-based identification system is shown in Figure 8.

Presumptive testing is used by first responders and forensic members in the field in order to safely respond to an incident. Presumptive screening of a scene can also assist in determining the possible point of release of an agent, thereby increasing the likelihood of obtaining good-quality samples for analysis. Conformational testing is required to be



Figure 8 Nerve agent vapor detector enzymatic indication system manufactured by Anachemia Canada, Inc. (http://www.anachemia.com/engnew/frame/product4.html) [Reproduced with permission from Anachemia Canada, Inc.]

undertaken in a laboratory to confirm the presence and use of CWAs.

Laboratory-Based Analysis

For definitive analysis of CWAs, a range of sophisticated laboratory-based instrumental techniques have been developed and standardized for laboratory confirmation. The aim of this analysis is to confirm through scientific processes the presence and type of CWAs or TICs through a more comprehensive understanding of the manufacturing process and degradation products produced. The complete unequivocal analysis of the samples can take considerable time. Mesilaakso [6] described an analytical strategy for the analysis and forensic identification of CWAs.

The analytical strategy displayed below should not be considered all-encompassing. Additional instrumental methods can be utilized for the identification of the process used to manufacture the agent. Potentially, the location of manufacture of the agent, or precursors, can also be determined. One such method is stable isotope ratio mass spectrometry (IRMS). IRMS has been utilized for profiling illicit drugs, and to a lesser extent explosives, to identify the location of manufacture or production.

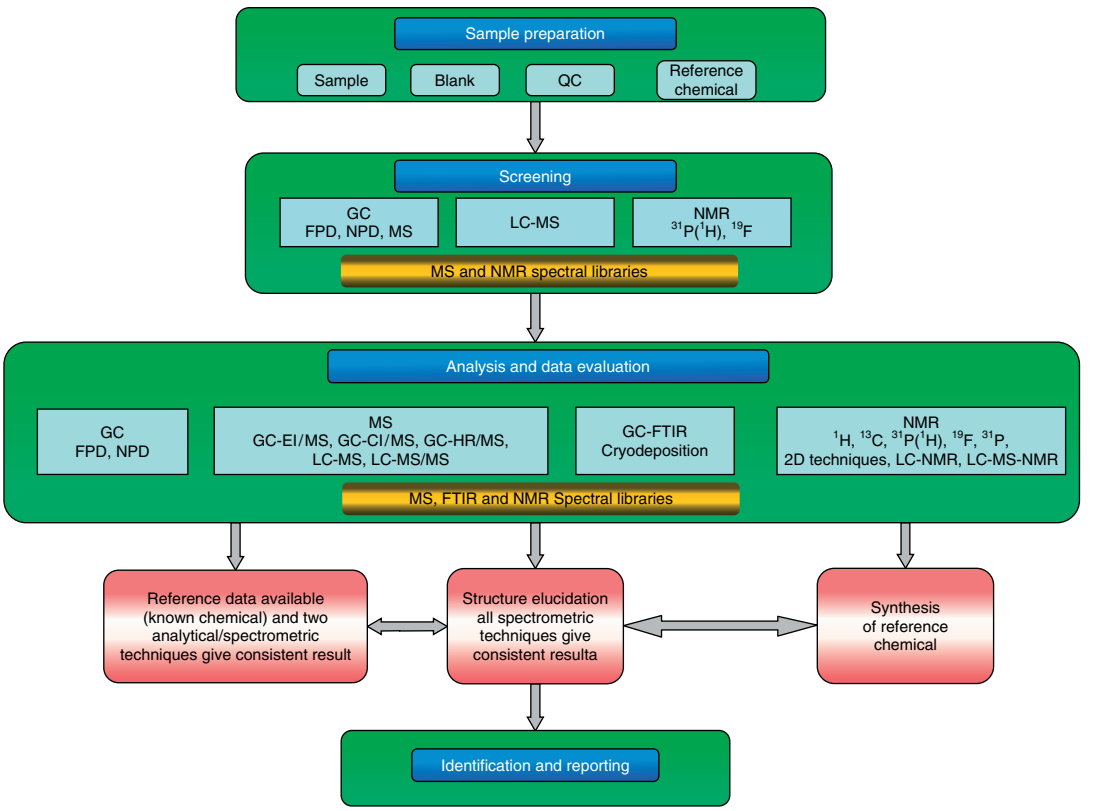


Figure 9 Analytical strategy for the analysis of CWAs

It is possible that this method could be used in the identification of CWAs and other TICs and TIMs. Figure 9 is a simplified version of these procedures.

Sample Collection

The collection of good samples is one of the most critical parts of a successful analysis. The location of the samples and their movements should be detailed and recorded throughout the entire sample collection, transportation, and analysis stages of the investigation. Having centralized laboratories capable of testing and analyzing CWAs is imperative. These laboratories should be equipped to track the continuity of the sample and maintain the chain of custody. The chain of custody for the evidence in a case relating to the use of a CWA will be closely examined and could jeopardize the case because of technicalities.

Sample Preparation

The accurate analysis of CWC-related chemicals depends on the collection of high-quality samples and well-planned and effective sample preparation procedures that are suited to the methods of analysis. The preparation of samples for analysis requires a thorough understanding of the behavior of the various types of chemicals in different sample matrices, both before and during the sample preparation, and an awareness of the limitations of the chosen instrumental method of analysis. Samples and prepared subsamples need to be appropriately labeled, and preparation procedures should follow endorsed standard procedures. Continuity of all samples and subsamples must be established

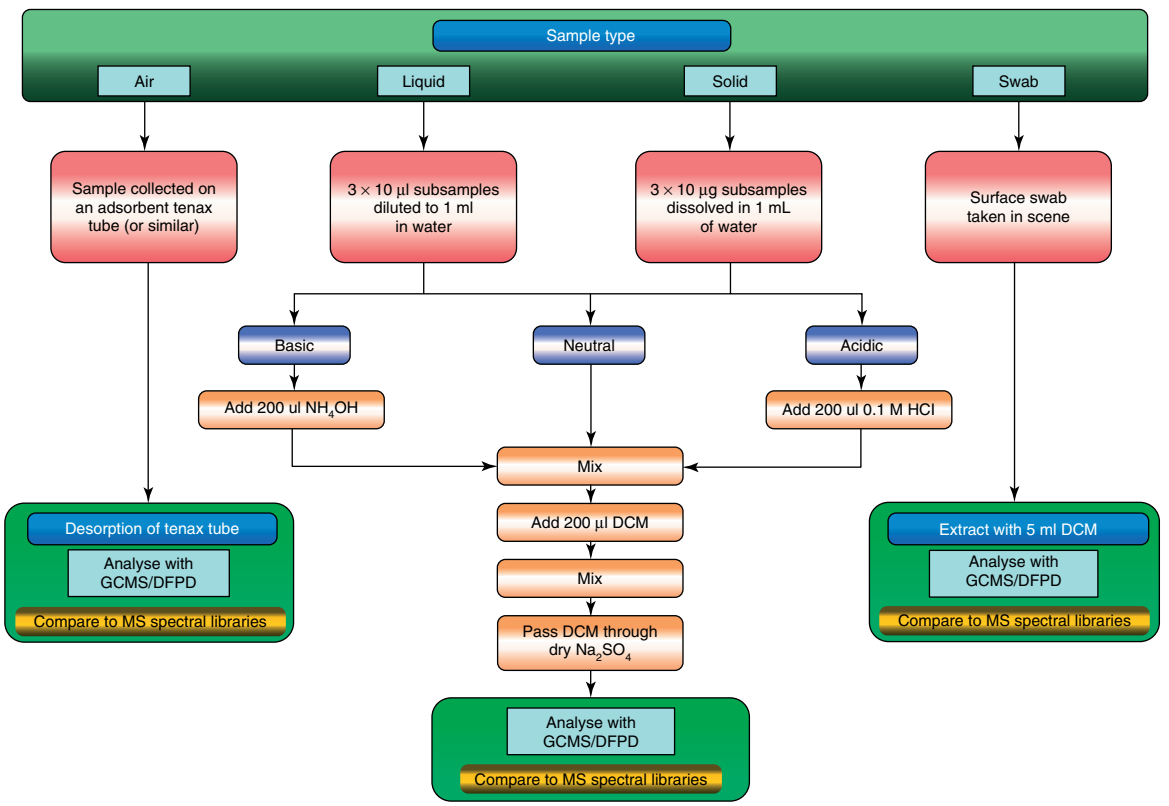


Figure 10 Example of a possible sample preparation protocol

and maintained during the analytical aspect of the investigation.

There are specific methods to be utilized for different sample types, such as air, aqueous liquid samples, soil samples, wipe samples, activated charcoal samples, concrete samples, paint, rubber, and other polymeric samples. Figure 10 is an example of possible sample preparation steps that could be utilized in the case of a suspected CWA, which are aimed at producing a presumptive identification only.

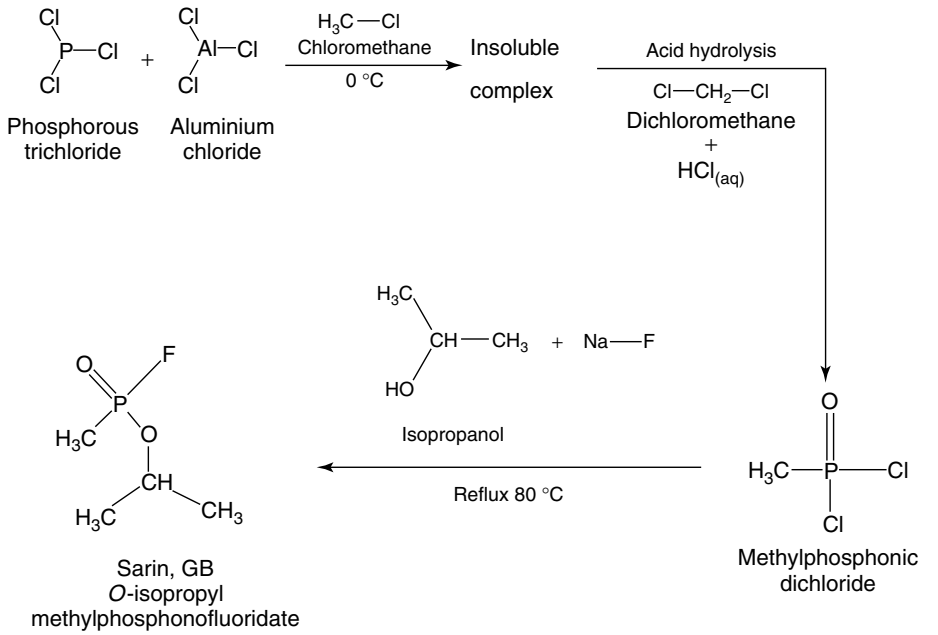


Figure 11 An example of a synthesis method for the nerve agent sarin (GB) [7]

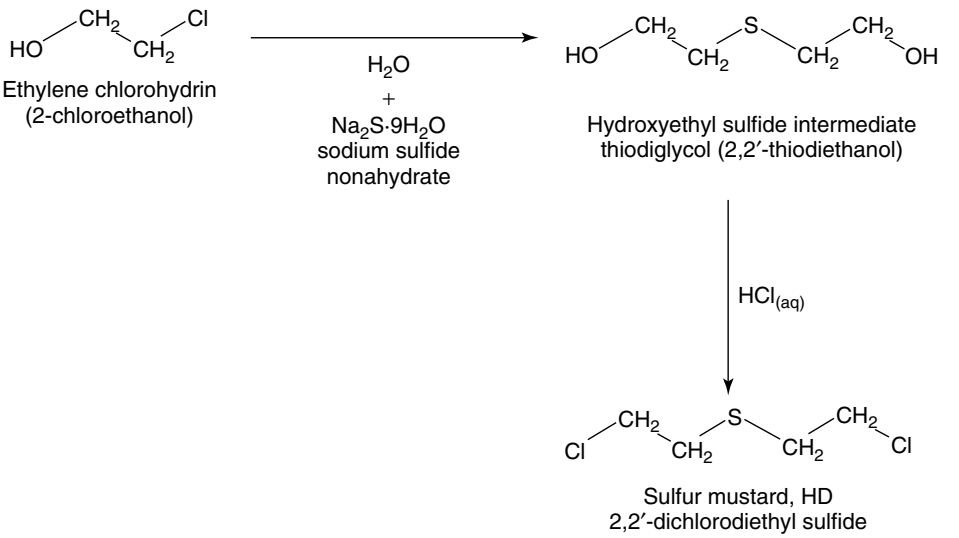


Figure 12 An example of a synthesis method for blister agent sulfur mustard (HD) [7]

Each sample matrix presents a unique set of challenges that must be met in order to maximize the recovery of the agents while minimizing matrix interference. With such difficult sample matrices, a substantial amount data analysis will be required to identify the compound(s) of interest. A common tool that has been used for deconvolution of complex data files is automated mass spectral deconvolution and identification system (AMDIS) created by the National Institute of Standards and Technology (NIST). This data analysis package enables complex mixtures that cause coeluting chromatographic peaks to be resolved, at least to an extent, into their component parts. This does not mean that a less rigorous chromatographic method is acceptable, but in cases of complex mixtures there is a greater probability that the component compounds could be identified.

Once all the analysis and data interpretation have been completed, a report will be issued. This may require the presentation of all the analytical data to a court, or at least made available to the defense team. With this in mind, scientists working in this area must be prepared for the court and should fully understand the procedures, instruments, and data analysis packages used to prepare the final report, as they may have to explain them to the judiciary.

Synthesis of CWAs

During the analysis of a particular CWA, reaction intermediates and markers from the synthetic process utilized may be identified. Examination of these reaction markers will assist in establishing the synthetic method utilized, which, in turn, will assist in the investigation of a CW-related incident. The absence of reaction markers is a critical finding, as this would indicate that a competent scientist produced the material of interest and knows that there is a requirement to purify the final product.

As an example of possible synthesis processes for CWAs, sarin and sulfur mustard have been chosen, and one process for each of these CWAs is detailed below. As can be seen from these synthetic pathways, the overall production of these two CWAs is not overly technical or challenging for an individual who

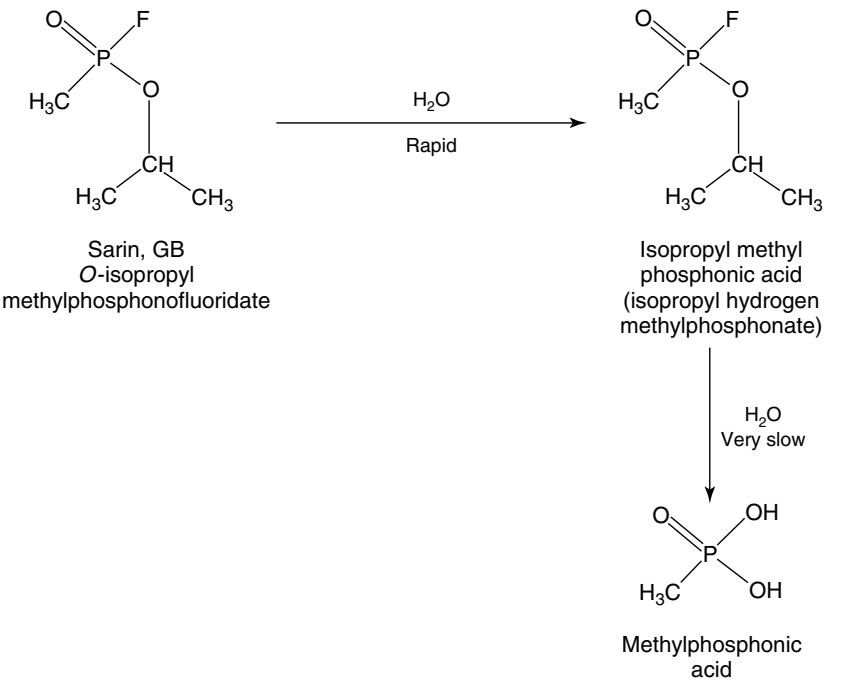


Figure 13 An example of a sarin (GB) hydrolysis [8]

has basic undergraduate chemistry training or experience. Having access to suitable protective equipment or decontamination equipment is not really required. Figures 11 and 12 are examples of possible synthetic routes for sarin and mustard.

Degradation of CWAs

When a CWA is exposed to the atmosphere, it will commence degradation to form more stable compounds. Some of the degradation products

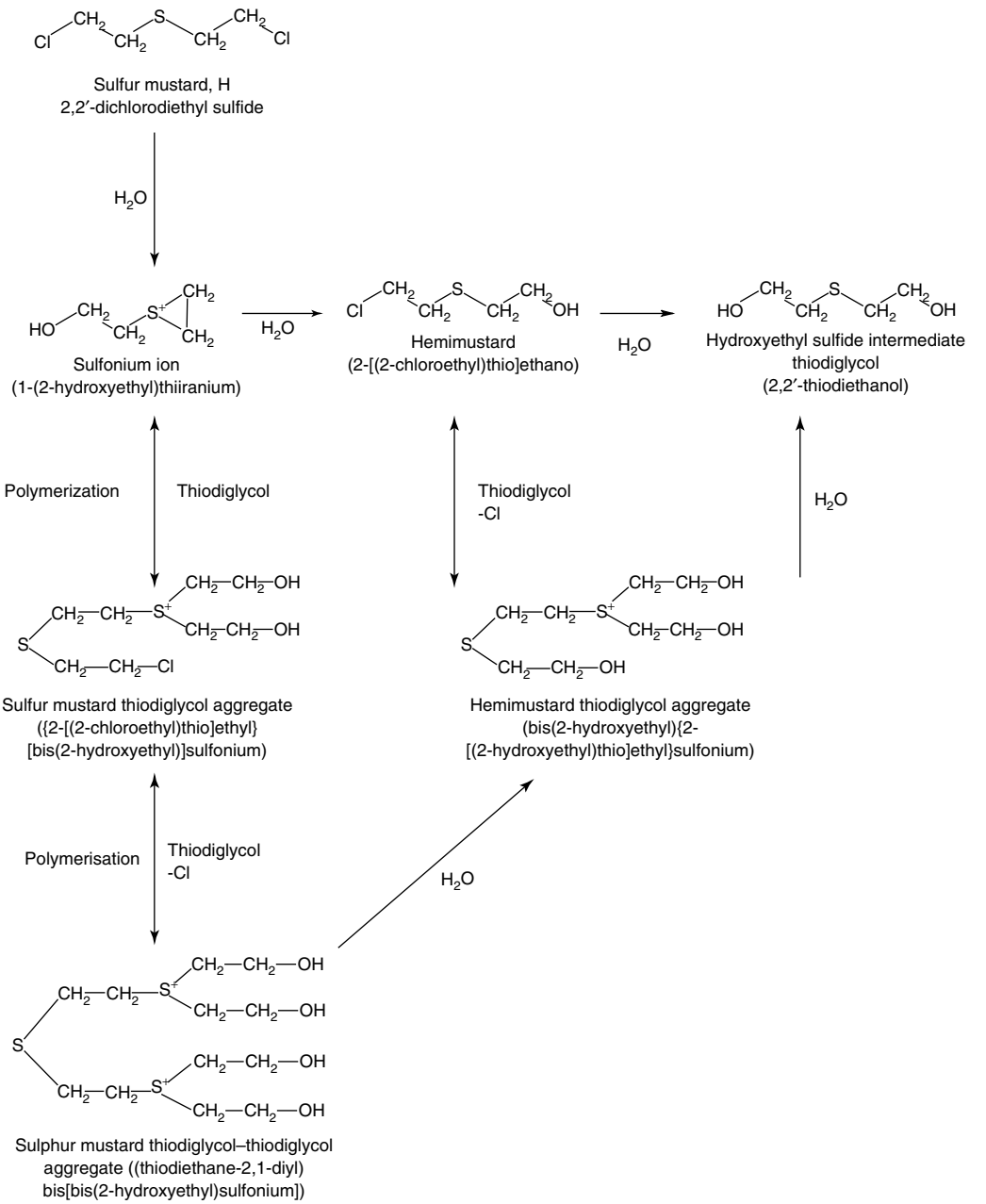


Figure 14 An example of sulfur mustard (HD) hydrolysis [8]

formed will have a degree of toxicity; however, as the degradation pathway progresses, the toxicity of the final product decreases. Examples of the degradation (hydrolysis) in water for sarin and sulfur mustard are shown in Figures 13 and 14, respectively.

Sarin hydrolyzes rapidly to isopropyl methyl phosphonic acid and then slowly to an inert, stable product, methylphosphonic acid. A publication by the Stockholm International Peace Research Institute (SIPRI) detailed the hydrolysis of sarin in water and reported the half-lives of sarin in water at different temperatures and pHs (Table 4) [9]. In this report, it has also been identified that the half-life of sarin was greatly affected by the salt concentration of the water (Table 5). It is possible that, if the environmental conditions are known, the ratio of the hydrolysis products and sarin will give an indication of the time since manufacture for the agent.

Table 4 Half-life of sarin in water (h) [9]

	pH				
Temperature (°C)	6.5	7.0	7.5	8.0	9.0
0	8300	2650	830	265	26.5
10	1870	591	187	59	6.0
20	461	146	46	15	1.5
25	237	75	24	7.5	0.8
30	125	39	12.5	4	0.4

Reproduced from Ref. 9. © Taylor and Francis Group, 1985

Table 5 Rate of hydrolysis of sarin due to salinity [9]

Water condition		Half-life of sarin (h)
Sea water	pH = 8.0, temperature = 25°C	0.4
Distilled water	pH = 7.9, temperature = 25°C	7.5

Reproduced from Ref. 9. © Taylor and Francis Group, 1985

Table 6 Hydrolysis rates and half-lives for sulfur mustard [9]

Water condition	Observations
Saturated solution of sulfur mustard (temperature = 20°C)	110 min for 99% hydrolysis
Saturated solution of sulfur mustard (temperature = 50°C)	4 min for 99% hydrolysis
Distilled water (temperature = 25°C)	Half-life 8.5 min
Salt water (Temperature = 25°C)	Half-life 60 min

Reproduced from Ref. 9. © Taylor and Francis Group, 1985

Table 6 details the hydrolysis rates and half-lives for sulfur mustard under various water conditions. The hydrolysis rates and half-lives vary considerably and appear to be shorter in fresh water; this is commonly not the case, as the hydrolysis of mustard is surface controlled [9]. Although the hydrolysis rate for sulfur mustard is rapid, during hydrolysis sulfur mustard has the capacity to form polymeric thiodiglycol structures. These thiodiglycol polymers have the capability of producing a stable impervious "heel" that protects the bulk of the chemical agent. Trapp [9] reports the formation of this protective phase or "heel" in water where there is no turbulence. Trapp [9] attributes this fact as the reason for injuries when Japanese chemical weapons were recovered from ocean dump sites in 1962 and 1970 [9]. It should be noted that as thiodiglycol is both a precursor and a degradation product, care must be taken when interpreting the analytical results obtained from the analysis of a sulfur mustard sample.

(The information in the above section has been compiled from open source reports and publications.)

References

[1] Garrett, B. & Hart, J. (2007). *V Historical Dictionary of Nuclear, Biological and Chemical Warfare (Historical Dictionaries of War, Revolution, and Civil Unrest)*, The Scarecrow Press.

[2] Wax, P. (2004). *Toxic Warfare: Looking Beyond Conventional Chemical Weapons Banner Samaritan Medical Center*, American College of Medical Toxicology, Phoenix.

[3] (2008). Organisation for the Prohibition of Chemical Weapons Website, http://www.opcw.org/.

[4] Kosal, M. (2003). *The Basics of Chemical and Biological Weapons Detectors*. PhD Thesis, James Martin Center For Nonproliferation Studies, November 24, http://cns.miis.edu/pubs/week/031124.htm).

[5] Yin, S. & Kwok, Y. (2005). *Detection Technologies for Chemical Warfare Agents and Toxic Vapors*, CRC Press.

[6] Mesilaakso, M. (2005). *Chemical Weapons Convention Chemical Analysis Sample Collection, Preparation and Analytical Methods*, John Wiley & Sons.

[7] Ledgard, J. (2003). *The Preparatory Manual of Chemical Warfare Agents*, The Paranoid Publications Group.
[8] Munro, N., Talmage, S., Griffin, G., Waters, L., Watson, A., King, J. & Hauschild, V. (1999). The sources, fate, and toxicity of chemical warfare agent degradation products, *Environmental Health Perspectives* **107**(12), 933–974.
[9] Trapp, R. (1985). *The Detoxification and Natural Degradation of Chemical Warfare Agents, SIPRI*, Taylor & Francis.

Further Reading

Compton, J. (1988). Military chemical and biological agents, in *Chemical and Toxicological Properties*, The Telford Press, Caldwell, pp. 11–12.
Organisation for the Prohibition of Chemical Weapons, (2005). *Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction*, Technical Secretariat of the Organisation for the Prohibition of Chemical Weapons.
Dishovsky, C., Pivovarov, A. & Benschop, H. (2006). *Medical Treatment of Intoxications and Decontamination of Chemical Agents in the Area of Terrorist Attack"*, Springer.
Kurata, H. (1980). Lessons learned from the destruction of the chemical weapons of the Japanese imperial forces, in *Chemical Weapons: Destruction and Conversion, Stockholm International Peace Research Institute*, Taylor and Francis, London, pp. 77–93.
Marrs, T., Maynard, R. & Siddell, F. (2007). *Chemical Warfare Agents – Toxicology and Treatment*, Second Edition, John Wiley & Sons.
Romano, J., Lukey, B. & Salem, H. (2007). *Chemical Warfare Agents – Chemistry, Pharmacology, Toxicology and Therapeutics*, CRC Press.
Stockholm International Peace Research Institute (1971). The problem of chemical and biological warfare, *A Study of the Historical Technical Military, Legal, and Political Aspects of CBW and Possible Disarmament Measures*, *The Rise of CB Weapons*, Humanities Press, New York, Vol. 1, pp. 119 ff., 305.
Theobald, N. & Ruhl, N.-P. (1994). Chemical warfare agent munitions in the baltic sea, ***Deutsche Hydrographische Zeitschrlft*** **46**, 121–131.

Related Articles

Biological Agents

Bomb Scene Management

Chemical, Biological, Radiological, and Nuclear Investigations

Crime Scene Investigation

Crime Scene Management

Drug Analysis

Drug Profiling

Explosions: Scene Investigation

Microscopy: FTIR

Nuclear Forensics

Poisons: Detection of Naturally Occurring Poisons

KEITH W. NORMAN AND VICTORIA C. GILLMAN

Child Abuse *see* Battered Child Syndrome

Child Custody *see* Parental Alienation

Child Custody Visitation Evaluations *see* Visitation Rights

Child Neuropsychological Assessment *see* Neuropsychological Assessment: Child

Child Sexual Abuse

Introduction

The forensic aspects of the sexual abuse of children is one part of the very broad topic of child maltreatment. The forensic investigation of child sexual abuse requires the cooperation of a diverse group of professionals, including emergency room staff, pediatricians and other primary care physicians, law enforcement, social service personnel, and a variety of mental health professionals. Psychiatrists, psychologists, and other mental health professionals may evaluate children for forensic or legal purposes, either in private practice or as part of an interdisciplinary team. Forensic evaluators may assist the court in determining what happened to the child and make recommendations regarding placement or treatment. Evaluators may be asked to assess the credibility of a child who allegedly was sexually abused. In civil law suits, mental health professionals may testify about the cause, nature, and extent of the child's psychological injuries.

Definitions

Sexual abuse of children refers to sexual behavior between a child and an adult or between two children, when one of them is significantly older or when coercion is used. The perpetrator and the victim may be of the same sex or the opposite sex. Sexual behaviors may include touching of breasts, buttocks, and genitals, whether or not the perpetrator or victim are undressed. They may also include exhibitionism, fellatio, cunnilingus, and penetration of the vagina or anus with sexual organs or objects. Sexual abuse may involve behavior over an extended period of time or it may be an isolated incident. Developmental factors must be considered in assessing whether sexual activities between two children are abusive or normative. In addition to touching, sexual abuse also refers to *sexual exploitation of children*, for instance, engaging in prostitution of minors or activities related to pornography depicting minors.

The *Diagnostic and Statistical Manual of Mental Disorders,* Fourth Edition, Text Revision (*DSM-IV-TR*) [1] addresses the classification of child maltreatment in the article, "Other Conditions That May Be a Focus of Clinical Attention," and in the section, "Problems Related to Abuse or Neglect." There are separate diagnoses for physical abuse of a child, sexual abuse of a child, and neglect of a child. Also, there are separate numerical codes depending on whether the focus of attention is on the perpetrator of the abuse or the victim of the abuse. These conditions and problems are listed on Axis I of the multiaxial assessment of *DSM-IV-TR.*

In federal law, child sexual abuse means the employment, use, persuasion, inducement, enticement, or coercion of any child to engage in, or assist any other person to engage in, any sexually explicit conduct (or simulation of such conduct for the purpose of producing a visual depiction of such conduct) or the rape (and in cases of caretaker or interfamilial relationships, statutory rape), molestation, prostitution, incest with children, or other form of sexual exploitation of children. A variety of legal definitions and guidelines regarding child sexual abuse exists at the state level, so clinicians should be aware of the terminology and criteria used in their own locale.

Epidemiology

Each year the Children's Bureau, an agency within the Department of Health and Human Services, collects data on child maltreatment. The results are published in an annual document called *Child Maltreatment* [2]. The agency estimated that during 2005, about 3.3 million allegations of child abuse and neglect comprising 6.0 million victims were reported to child protective services. Of the allegations that were screened in and investigated, about 25% were considered substantiated or indicated and about 55% were considered unsubstantiated. The substantiated or indicated cases were distributed as follows: neglect, 63%; physical abuse, 17%; sexual abuse, 9%; psychological abuse, 7%; and medical neglect, 2%.

The data were analyzed for patterns of maltreatment by the sex and age of victims. Rates of many types of maltreatment were similar for male and female children, but the sexual abuse rate for female children was higher than that for male children. Examining the age distribution of victims, the 0–3 age group had the highest victimization rate, and the rate of victimization declined as the age of the victims increased. For example, the abuse rate for infants (age 0–3) was 16.5 per 1000, while the rate for children

(age 4–7) was 13.5 per 1000. The highest rate of sexual abuse, however, was in the young adolescent (age 12–15) group, with 17.3% of all victims suffering sexual abuse *versus* only 8.9% in the 4–7 age group. All these figures are approximations because the actual amount of abuse is unclear.

Perpetrators and Patterns of Child Sexual Abuse

There is enormous diversity among those who perpetrate child abuse [3]. Perpetrators come from diverse backgrounds in terms of *age, sex, mental capacity*, and *socioeconomic status*. Family members are often perpetrators of sexual abuse. Other times, the perpetrator may be a person entrusted with the care and well-being of a child such as relatives and baby-sitters. Both men and women are perpetrators of sexual abuse. Sometimes sex offenders themselves are children or adolescents – for example, older siblings abusing younger siblings or adolescents who rape peers. Finally, one must include mentally ill or especially mentally retarded persons who may be the perpetrators of child sexual abuse without a clear understanding of their own behavior. Perpetrators may commit child sexual abuse only once in their life or they may serially abuse scores or hundreds of children. Some perpetrators may be attracted to children of a particular gender, age, and physical appearance; others may prey on children at large in a rather indiscriminate manner.

Intrafamilial Pattern of Abuse

Incest may be strictly defined as sexual relations between close blood relatives, for example, between a child and the father, uncle, or sibling. Incest may be defined more broadly to include sexual intercourse between a child and a stepparent or stepsibling. Although father–daughter abuse is the most common form of incest, it may also involve father and son, mother and daughter, and mother and son. Healthy, self-confident children refuse the intrusions either directly (via oppositionality and temper tantrums) or indirectly (through silence and distancing maneuvers) or by adopting any strategy that causes the offender to refrain.

Intrafamilial sexual abuse – as well as other sexual abuse that occurs over a period of time – frequently evolves through five phases [4]:

- **Engagement phase**

The perpetrator induces the child into a special relationship.

- **Sexual interaction phase**

The sexual behaviors progress from less to more intrusive forms of abuse.

- **Secrecy phase**

The perpetrator threatens the victim not to tell.

- **Disclosure phase**

The abuse is discovered accidentally (when another person walks in the room and sees it), as a result of the child's reporting it to a responsible adult, or because the child is brought for medical attention and an alert clinician asks the right questions.

- **Suppression phase**

The child may often retract statements of the disclosure because of family pressure or because of the child's own mental processes. That is, the child may perceive that violent or intrusive attention is synonymous with interest or affection.

Close Associate Pattern of Abuse

The perpetrator of sexual abuse may be a person who is not a parent or relative, but is entrusted with the care and well-being of a child. This group of perpetrators may include family friends, neighbors, baby-sitters, scout leaders, clergymen, and other religious leaders. In this circumstance, the pedophilic perpetrator typically grooms the child over a period of time. He or she gains the friendship of the child through enjoyable activities and gifts, introduces sexual activities that may seem innocent and even pleasurable, and progresses to more intrusive activities. This type of sexual abuse has a pattern that is similar to the five phases of intrafamilial sexual abuse.

School is another setting where the close associate pattern of sexual abuse may take place, with the perpetrators in this case being teachers, coaches, counselors, or principals. Both boys and girls may be victims. While this type of abuse may take the form of coerced sexual activities, often, especially for adolescents, it may appear to be a consensual relationship between an underage student and a teacher or other staff. Of note, corporal punishment by school personnel still occurs in some states and

localities. This form of punishment, which usually consists of adult's hitting the child's buttocks with a wooden paddle, may be perceived as a sexual assault by some children.

Stranger Sexual Abuse

Children may be abducted and sexually abused by strangers, which often involves more violence than abuse by a family member or close associate. A perpetrator may observe a playground and identify a child who is not closely supervised. In this way, a pedophile may molest hundreds of children before he is apprehended. For each child victim, this is usually a single, isolated experience. A solo sex ring is a form of child sexual abuse that involves one adult perpetrator and multiple child victims, who may know about each other's sexual activities with the perpetrator.

With the proliferation of computers and the Internet, cyberspace has become a fertile ground for pedophiles to perpetrate their crimes on children [5]. This may include pedophiles' producing and trading child pornography over the Internet; interacting with minors through chat rooms, often with the goal of luring them into sexual activities; and outright trafficking of children for sexual purposes.

Clinical Features of Child Sexual Abuse Victims

Sexually abused children manifest a variety of emotional, behavioral, and somatic reactions [6, 7]. These psychological symptoms are neither specific nor pathognomonic since the same symptoms may occur without any history of abuse. The psychological symptoms manifested by abused children can be organized into clinical patterns. Although it may be helpful to note whether a particular individual falls into one of these patterns, this in itself is not diagnostic of sexual abuse.

- **Anxiety symptoms**

This includes fearfulness, phobias, insomnia, nightmares that directly portray the abuse, somatic complaints, and posttraumatic stress disorder.

- **Dissociative reactions and hysterical symptoms**

The child may exhibit periods of amnesia, daydreaming, trancelike states, hysterical seizures, and symptoms of dissociative identity disorder.

- **Depression**

This may be manifested by low self-esteem and suicidal and self-mutilative behaviors.

- **Disturbances in sexual behaviors**

Some sexual behaviors are particularly suggestive of abuse, such as masturbating with an object, imitating intercourse, and inserting objects into the vagina or anus. In contrast to these overly sexualized behaviors, the child may avoid sexual stimuli through phobias and inhibitions.

- **Somatic complaints**

This includes enuresis, encopresis, anal and vaginal itching, anorexia, bulimia, obesity, headache, and stomachache.

Approximately one-third of sexually abused children have no apparent symptoms. On the other hand, the following factors tend to be associated with more severe symptoms in the victims of sexual abuse: greater frequency and duration of abuse, sexual abuse that involved force or penetration, and sexual abuse perpetrated by the child's father or stepfather. Other factors associated with poorer prognosis are the child's perception of being less believed, family dysfunction, and lack of maternal support. Of note, multiple investigatory interviews appear to increase symptoms.

Evaluation Process

A forensic evaluation emphasizes collecting accurate and complete data to determine – as objectively as possible – what happened to the child. The data collected in a forensic evaluation must be preserved in a reliable manner through audiotape, videotape, or detailed notes. The results of the forensic evaluation are organized into a report that will be used in the court and read by attorneys, a judge, and others. From the psychiatric and psychological perspective, the interview is usually the primary source of information and the physical examination is secondary. In practice, children who may have been sexually abused are interviewed first and later given a physical examination and other tests.

In addition to interviewing the child, the evaluator obtains a history from the parents (separately, in most cases) and other pertinent informants. In the evaluation of suspected sexual abuse, the examiner should consider the possibility that the parents or other informants are not telling the truth. For example, the mother may wish to avoid the discovery of father–daughter incest by blaming the child's genital injury on another child or a stranger. In another scenario, the mother may fabricate an allegation of incest when the child had never been abused at all, for example, to gain advantage in a child custody dispute. The first version protects a father who is guilty; the second version implicates a father who is innocent.

The examiner should determine how the allegation originally arose and what subsequent statements were made [8]. Determine the emotional tone of the first disclosure (e.g., whether it arose in the context of a high level of suspicion of abuse). Determine the sequence of previous examinations, the techniques used, and what was reported. Try to determine whether the previous interviews were likely to have distorted the child's recollections. If possible, review transcripts, audiotapes, and videotapes of earlier interviews. Seek a history of sexual overstimulation (lax attitudes toward nudity and sexual activity in the home), prior abuse, or other traumas. Consider other stressors that could account for the child's symptoms. The examiner should also ask about exposure to other possible male and female perpetrators.

A psychosocial history should be collected and organized that includes

- symptoms and behavioral changes that may be related to the alleged abuse;
- confounding variables such as psychiatric disorder or cognitive impairment that may need to be considered;
- the family's attitude toward discipline, sex, and modesty;
- developmental history from birth through periods of possible trauma to the present;
- family history, such as earlier abuse of the parents, substance abuse by the parents, spouse abuse, and psychiatric disorder in the parents; and
- underlying motivation and possible psychopathology of the adults involved.

Collateral information may be useful in verifying allegations of sexual abuse, and thus the evaluator should consider requesting pertinent information from the following, after obtaining authorizations: protective services, school personnel, other caregivers (e.g., baby-sitters), other family members (e.g., siblings), the pediatrician, and police reports.

The Child Interview

Several structured and semi-structured interview protocols have been introduced that are designed to maximize the amount of accurate information and minimize mistaken or false information provided by children. These approaches include the cognitive interview, which encourages witnesses to search their memories in various ways such as recalling events forward and then backward. The interview protocol developed at the National Institute of Child Health and Human Development (NICHD) includes a series of phases and makes use of detailed interview scripts [9]. The stepwise interview (outlined below) [10] is a funnel approach that starts with open-ended questions and, if necessary, moves to more specific questions.

In interviewing children who may have been abused, it should be possible to follow a standard protocol and also be flexible, considerate, and generally supportive. As when seeing any patient, the evaluator must size up the situation and use techniques that are likely to help the youngster become comfortable and communicative. One victim might need a favorite object (e.g., a teddy bear or a toy truck); another might need to have a particular person included in the interview. Some children are comfortable talking; others prefer to draw pictures. The child might make important comments while chatting during the break time instead of during the structured interview.

Interview Process

The interviewer of the allegedly sexually abused child should remember the following principles [8]:

- audiotape or videotape the interview, if possible;
- use a minimum number of interviews (perhaps two or three), as multiple interviews may encourage confabulation;
- avoid repetitive questions, either/or questions, and multiple questions and try to avoid leading and suggestive questions;

- use restatement, that is, repeat the child's account back to the child (which allows the interviewer to see if the child is consistent and ensures that the interviewer understands the child's report);
- conduct the examination without the parent present (if the child is very young, consider having a family member in the room);
- use an examination technique that is appropriate to the child's age and developmental level, such as drawings and play re-enactment; and
- determine the child's terms for body parts and sexual acts; do not educate or provide new terms.

Interview Content

The interview should not take the form of an interrogation. Note the child's affect while discussing these topics and be tactful in helping the child manage anxiety. Young children may not be able to report all of the relevant information. The examiner should explore the following:

- whether the child was told to report or not report anything;
- who the alleged perpetrator was;
- what the alleged perpetrator did;
- where it happened;
- when it started and when it ended;
- number of times the abuse occurred;
- how the child was initially engaged and how the abuse progressed over time;
- how the alleged perpetrator induced the child to maintain secrecy;
- whether the child is aware of specific injuries or physical symptoms associated with the abuse;
- whether any photography or videotaping took place?

The Stepwise Interview

The usual clinical interview may need some modification for evaluating a child who may have been abused. The stepwise interview [10], which is primarily intended for forensic evaluations, consists of the following components.

- **Build rapport**

Build rapport and informally observe the child's behavior, social skills, and cognitive abilities.

- **Ask the child to describe two specific past events**

This step assesses the child's memory and models the form of the interview for the child. Ask nonleading, open-ended questions in the pattern that will occur through the rest of the interview.

- **Establish the need to tell the truth**

Reach an agreement with the child that in this interview only the truth will be discussed, not "pretend" or imagination. Explain to the child that it is fine not to know the answer to a question. It is fine to correct the interviewer.

- **Introduce the topic of concern**

Start with general questions such as, "Do you know why you are talking with me today?" Proceed, if necessary, to more specific questions such as, "Has anything happened to you?" Drawings may help initiate disclosure.

- **Elicit a free narrative**

Once the topic of abuse has been introduced, the interviewer encourages the child to describe each event from the beginning without leaving out any details. If abuse has occurred over a period of time, the interviewer may ask for a description of the general pattern and then for an account of particular episodes.

- **Pose general questions**

The interviewer may ask general questions to elicit further details. These questions should not be leading and should be phrased in such a way that the child realizes an inability to recall or lack of knowledge is acceptable.

- **Pose specific questions if necessary**

Asking specific questions in a careful way may yield helpful clarification. For example, the interviewer may follow-up on inconsistencies in a gentle, nonthreatening manner. Avoid repetitive questions or appearing to reward particular answers in any way.

- **Use interview aids if necessary**

Anatomical dolls may be useful in understanding exactly what sort of abusive activity occurred. The dolls are not used to diagnose child abuse, only to clarify what happened.

- **Conclude the interview**

Toward the end of the interview, the interviewer may ask a few leading questions about irrelevant issues (e.g., "You came here by taxi, didn't you?"). If the child demonstrates susceptibility to the suggestions, the interviewer must verify that the information

obtained earlier was not contaminated. Finally, the child is thanked for participating in the interview. However, the child should not be complimented or rewarded, since that may communicate that the child has given "the correct answer" to the interviewer.

Physical Examination

A forensic medical examination is usually performed by a pediatrician, preferably one with special expertise in child sexual abuse. The examination consists of a general physical examination with a focus on the genital and anal areas. A comprehensive examination must direct attention to all aspects of the victim's health status, including maturity, developmental assessments, and a review of systems [11–13].

After sexual abuse, labial and perineal tissues may manifest abrasions, contusions, lacerations, petechiae, and edema. Examination may reveal bloodied tissue, genitorectal tears, and infected secretions. On the other hand, scarring after sexual abuse may be minimal or not evident if vaginal penetration has been gentle, consensual, or gradual. Vaginal and rectal tissues rapidly heal after injury and abrasions can be repaired within several days. Thus, the lack of physical findings may indicate either a return to normalcy after rapid healing of minor lesions or abuse perpetrated in a manner that left no material evidence. In other words, a normal genital examination does not mean the child was not sexually abused. Most victims of child sexual abuse have a normal or nonspecific genital examination.

Serious injuries to the penis – such as scaldings and partial amputation – produce noticeable scars. The more common traumas – hematomas, abrasions, and lacerations – are superficial injuries and heal rapidly without scarring.

Anal findings include dilation with loss of anal–rectal tone, flattened rectal rugae, lacerations, and wedge-shaped and linear scars. Acute trauma may result in swelling and spasm secondary to submucosal hemorrhage. Most rectal injuries, massive or minor, heal quickly and nearly completely, leaving little evidence of abuse. Rectal irritants – such as pinworms, lichen sclerosis, atopic dermatitis, and rare hematological disorders – may be present in a child, and should be considered as part of the differential diagnosis of anorectal lesions.

Special Tests and Laboratory Examination

When sexual abuse is suspected, the victim requires evaluation for sexually transmitted diseases (STDs) [14]. If the examination occurs within 72 h of the alleged abuse, the clinician should consider cultures of the genital, anal, and oral mucosa. The following STDs are essentially diagnostic of child sexual abuse unless they were acquired perinatally: gonorrhea, syphilis, and *Chlamydia trachomatis*. Positive serology for the human immunodeficiency virus (HIV) is diagnostic of child sexual abuse unless it was acquired perinatally or by blood transfusion. The following STDs raise suspicions of child sexual abuse and should ordinarily be reported to protective services: herpes simplex virus in a genital location unless there is a clear history of autoinoculation; *Trichomonas vaginalis*; and condylomata acuminata or anogenital warts. However, infants and young children may acquire anal and genital warts perinatally. Or course, pregnancy is diagnostic of child sexual abuse, as is the presence of semen or sperm acid phosphatase.

Determining the Validitiy of a Case of Sexual Abuse

The psychiatric evaluation of youngsters who may have been sexually abused involves assessing the child's credibility. Although generally children tell the truth when they talk about abuse, sometimes children make false denials (saying they were not abused, when actually they were) or false allegations (saying they were abused, when actually they were not) [15].

Possible Explanations of Denials of Abuse

A false denial or retraction may occur for several reasons. The child may have been pressured by the perpetrator or family members to recant the allegation. The pressure may consist of bribery, mockery, or threats of injury. The child may be protecting a parent or other family member, even without external coercion. The child might be frightened or distressed by the investigation and decide to withdraw participation. For instance, an interviewer could induce a false denial by asking overly challenging questions. A child may be inhibited by shame or guilt; the child

may mistakenly assume responsibility for what happened. Finally, the child may have "accommodated" to the abuse, consciously or unconsciously, instead of objecting to it.

Possible Explanations of Allegations of Abuse

A false allegation of abuse may also occur for several reasons [16]. For example, sometimes a false allegation arises in the mind of a parent or another adult and is imposed on the child, and this may happen in several ways. The parent may have misinterpreted an innocent remark, a neutral piece of behavior, or a benign physical condition as evidence of abuse and induced the child to endorse this interpretation. The parent and child may share a *folie à deux* or the child may simply give in and agree with a delusional parent. A parent may have fabricated the story and induced the child to collude in presenting it to the authorities. Also, previous interviewers may have asked leading or suggestive questions. An interviewer who believes abuse occurred may unwittingly shape a child's responses until the child validates the interviewer's assumptions.

A false allegation of abuse may also occur through unconscious or nonpurposeful mental mechanisms in the child. For example, a young child may confuse fantasy with reality. Although rare, older children and adolescents may experience delusions about sexual activities in the context of a psychotic illness. False allegations occur through the mental mechanisms of misinterpretation (when the child misunderstood what happened and later reported it inaccurately) and miscommunication (when the child misunderstood an adult's question and the adult may later misinterpreted or took the child's statement out of context). Also, through the process of confabulation, the child may fill gaps in his memory with whatever information makes sense to him at the time.

A false allegation of abuse may occur through conscious or purposeful mental mechanisms in the child such as lying. Some children engage in pseudologia or pathological lying, the enthusiastic proliferation of falsehoods that may have no obvious motivation. Through the mechanism of innocent lying, younger children may make false statements without appreciating their serious implications because that seems to be the best way to handle the situation they are in. Through deliberate lying, children may choose to avoid or distort the truth for some personal advantage. This happens more with older children and adolescents.

Finally, through the mechanism of perpetrator substitution, the child may have actually been sexually abused and exhibits symptoms consistent with abuse but identifies the wrong person as the perpetrator, resulting in a false allegation. The child may do this to protect the actual offender or the child may displace the memories and accompanying affects onto another individual.

Prognosis After Child Sexual Abuse

Neurobiological Consequences of Child Sexual Abuse

Both severe physical abuse and repeated sexual abuse may cause changes in the child's developing brain that persist into adulthood [17–19]. Adult survivors of abuse are more likely to have abnormalities of their electroencephalograms (EEGs), which indicate limbic irritability. They are more likely to have a variety of changes in the structure of the brain as visualized by magnetic resonance imaging (MRI), including decreased volumes of the anterior cingular cortex, cerebellum, hippocampus, and amygdala. Decreases in the midsagittal area of the corpus callosum have been found in both children and adults with trauma histories. These neurobiological effects of persistent child maltreatment probably mediate the behavioral and psychological symptoms that follow abuse, such as increased aggressiveness, heightened autonomic arousal, depression, and memory problems.

Psychosocial Sequellae of Child Sexual Abuse

In a forensic context, psychiatrists and psychologists are sometimes asked to predict or estimate the prognosis of a child who has been sexually abused. In conducting such an assessment, the evaluator should rely as much as possible on research, not one's personal opinion or speculation [20]. Children who have been sexually abused are more likely as adults to have problems with anxiety, depression, suicidality, substance abuse, and chronic medical problems compared with children who were not

sexually abused [21–23]. These long-term effects of sexual abuse are not inevitable, however. The psychological effects of sexual abuse and other traumas may be affected by the child's genetic makeup as well as by the amount and nature of support from the nonabusive persons in the child's life [24].

Chronic severe abuse can lead to a syndrome known as *complex posttraumatic stress disorder* (*PTSD*) or disorder of extreme stress not otherwise specified (DESNOS). This disorder, which has been proposed for inclusion in the *DSM-V*, is characterized by alterations in the following: ability to modulate emotions; identity and sense of self; ongoing consciousness and memory; relations with the perpetrator and others; physical and medical status; and systems of meaning. The core of this disorder is a breakdown in the ability to regulate internal states.

References

[1] American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition, Text Revision, American Psychiatric Association, Arlington.

[2] U.S. Department of Health and Human Services, Administration on Children, Youth and Families (2007). *Child Maltreatment, 2005*. U.S. Government Printing Office, Washington, DC.

[3] Chaffin, M., Letourneau, E. & Silovsky, J. (2002). Adults, adolescents, and children who sexually abuse children: a developmental perspective, in *The APSAC Handbook on Child Maltreatment*, 2nd Edition, J.E.B. Myers, L. Berliner, J. Briere, C.T. Hendrix, C. Jenny & T.A. Reid, eds, Sage Publications, Thousand Oaks, pp. 205–232.

[4] Sgroi, S. (1982). *Handbook of Clinical Intervention in Child Sexual Abuse*, Lexington Books, Lexington.

[5] Deirmenjian, J.M. (2002). Pedophilia on the Internet, *Journal of Forensic Science* **47**, 1090–1092.

[6] Friedrich, W.N., Fisher, J.L., Dittner, C.A., Acton, R., Berliner, L., Butler, J. Damon, L., Davies, W.H., Gray, A., Wright, J. (2001). Child sexual behavior inventory: normative, psychiatric, and sexual abuse comparisons, *Child Maltreatment* **6**, 37–49.

[7] Putnam, F.W. (2003). Ten-year research update review: child sexual abuse, *Journal of the American Academy of Child and Adolescent Psychiatry* **42**, 269–278.

[8] American Academy of Child and Adolescent Psychiatry (1997). Practice parameters for the forensic evaluation of children and adolescents who may have been physically or sexually abused, *Journal of the American Academy of Child Psychiatry* **36**, 423.

[9] Poole, D. & Lamb, M. (1998). *Investigative Interviews of Children: A Guide for Helping Professionals*, American Psychological Association, Washington, DC.

[10] Yuille, J., Hunter, R., Joffe, R. & Zaparniuk, J. (1993). Interviewing children in sexual abuse cases, in *Child Victims, Child Witnesses: Understanding and Improving Testimony*, Guilford, New York.

[11] American Academy of Pediatrics, Committee on Child Abuse and Neglect (1999). Guidelines for the evaluation of sexual abuse of children, *Pediatrics* **103**, 186.

[12] Heger, A., Emans, S. & Muram, D. (2000). *Evaluation of the Sexually Abused Child*, 2nd Edition, Oxford University Press, New York.

[13] Kini, N. & Lazoritz, S. (1998). Evaluation for possible physical or sexual abuse, *Pediatric Clinics of North America* **45**, 205–219.

[14] Hammerschlag, M.R. (1998). Sexually transmitted diseases in sexually abused children: medical and legal implications, *Sexually Transmitted Infections* **74**, 167–174.

[15] Ney, T. (1995). *True and False Allegations of Child Sexual Abuse: Assessment and Case Management*, Brunner/Mazel, New York.

[16] Bernet, W. (1993). False statements and the differential diagnosis of abuse allegations, *Journal of American Academy of Child and Adolescent Psychiatry* **32**, 903–910.

[17] Teicher, M.H., Tomoda, A. & Andersen, S.L. (2006). Neurobiological consequences of early stress and childhood maltreatment: are results from human and animal studies comparable? *Annals of the New York Academy of Sciences* **1071**, 313–323.

[18] van der Kolk, B.A. (2003). The neurobiology of childhood trauma and abuse, *Child and Adolescent Psychiatric Clinics of North America* **12**, 293–317, ix.

[19] van der Kolk, B.A. (2006). Clinical implications of neuroscience research in PTSD, *Annals of the New York Academy of Sciences* **1071**, 277–293.

[20] Bernet, W. & Corwin, D. (2006). An evidence-based approach for estimating present and future damages from child sexual abuse, *The Journal of the American Academy of Psychiatry and the Law* **34**, 224–230.

[21] Widom, C.S. (1999). Posttraumatic stress disorder in abused and neglected children grown up, *The American Journal of Psychiatry* **156**, 1223–1229.

[22] Widom, C.S., Raphael, K.G. & DuMont, K.A. (2004). The case for prospective longitudinal studies in child maltreatment research: commentary on Dube, Williamson, Thompson, Felitti, and Anda (2004), *Child Abuse and Neglect* **28**, 715–722.

[23] Fergusson, D.M., Horwood, L.J. & Lynskey, M.T. (1996). Childhood sexual abuse and psychiatric disorder in young adulthood: II. Psychiatric outcomes of childhood sexual abuse, *Journal of the American Academy of Child and Adolescent Psychiatry* **35**, 1365–1374.

[24] Caspi, A., Sugden, K., Moffitt, T.E., Taylor, A., Craig, I.W., Harrington, H. McClay, J., Mill, J., Martin, J., Braithwaite, A., Poulton, R. (2003). Influence of life

stress on depression: moderation by a polymorphism in the 5-HTT gene, *Science* **301**, 386–389.

Related Articles

Battered Child Syndrome
Child Sexual Abuse Accommodation

WILLIAM BERNET AND NEELANJAN RAY

Child Sexual Abuse Accommodation

Introduction

The child sexual abuse accommodation syndrome (CSAAS) has no specific definition. It is a clinical concept that was first introduced by Ronald C. Summit, M.D [1] in 1983. Dr Summit utilized the syndrome to explain some of the atypical behaviors of children who have been the victims of sexual abuse (*see* **Child Sexual Abuse**; **Assault: Sexually Motivated**). He postulated that children who have been sexually abused often present with a number of peculiar characteristic features. These peculiar features might account for the conflicting details of the alleged abuse as presented by these children. These peculiar characteristics frequently evoke skepticism in adults about the actuality of the abuse upon eventual disclosure. This in turn results in further stigmatization of the child who is often accused of "lying, manipulating, or imagining by parents, courts, or clinicians."

Dr Summit explained that the conceptual basis of the syndrome arose from his attempt to foster better understanding of sexually abused children by clinicians. He stated that his main motive for the syndrome was to encourage clinicians to be more empathetic and supportive of abused children. He asserted that better understanding of children who have been sexually abused would lead to improved clinical interventions to help them to recover from the psychological impact of abuse.

In his original article, Dr Summit reported that the majority of the theoretical basis for the syndrome was drawn from the studies of female victims of sexual abuse. This was attributed to the fact that females were more likely to be sexually abused than males by adult male caretakers. Moreover, males were less likely to admit to being victims of sexual abuse. However, he indicated that the syndrome can be present in male victims since the response of children to the experience of being sexually abused is the same irrespective of their sex.

Over the years, the CSAAS has been a source of great dispute within the medicolegal community. The term syndrome is most commonly applied during the prosecution of alleged perpetrators of child sexual abuse. There have been attempts by several child sexual abuse "experts" to utilize the syndrome as the theoretical proof for the occurrence of sexual abuse in the absence of more incriminating evidence. This has led to intense dissent by the opposing school of thought which believes that there is no uniform set of characteristics that identifies the children who have been the victims of sexual abuse. The controversy regarding the authenticity of the syndrome has resulted in some skepticism amongst courts and Juries. The overall result has been the lack of uniformity in the verdict of legal battles concerning the applicability of the syndrome.

Clinical Characteristics

Dr Summit described five categories of behaviors that can be observed in the children who have been sexually abused that might lead to alienation upon disclosure of their abuse. These five categories of behaviors represent the discernible phases of the CSAAS. He characterized the first two sets of characteristics as "preconditions [1]" that are required for sexual abuse to occur. The remaining three categories of behaviors are usually observed after the abuse has been established. They represent "sequential contingencies [1]" that are required for the abuse to continue over a prolonged period of time without disclosure by the child. These categories of behaviors also explain the child's hesitation to reveal the details of the abuse when it is finally discovered by others. The five categories of the CSAAS described by Dr Summit include the following:

- secrecy;
- helplessness;
- entrapment and accommodation;

- delayed, conflicted and unconvincing disclosure;
- retraction.

Secrecy

Dr Summit theorized that the initial goal of an adult who is engaged in the sexual abuse of a child is to ensure that the child does not spontaneously report the abuse to other trusted adults. This often involves an initiation period during which the perpetrator attempts to create a common emotional link between the child victim and the perpetrator.

The establishment of a common shared secret, the integrity of which is maintained by various psychological techniques, is the initial weapon of the perpetrator. Some of the psychological techniques utilized by the perpetrator include threats, intimidation, blackmail, and rationalization. The perpetrator ensures that the abusive acts occur only when he is alone with the child. The perpetrator also ingrains a belief in the child that the shared act is a secret that "must never be shared with anyone else [1]".

According to Dr Summit, the combination of the perpetrators' psychological manipulation and the child's inherent vulnerabilities results in significant internal conflict for the child. The child's predicament is also compounded by the fact that she has little understanding of the concept of sex. Furthermore, any attempt by the child to disclose the abuse is often welcomed by "adult disbelief and suspicion of the child's motives [2]". As a result, the child becomes afraid of being blamed for the abuse. She might also worry that other trusted adults would not be capable of protecting her from possible retaliation by the perpetrator. Finally, the child develops a sense of isolation and helplessness. This in turn leads to chronic "self-blame [1]" with the resultant effect of a lifetime of "self-imposed exile from intimacy, trust, and self-validation [1]".

Helplessness

Dr Summit asserted that contrary to the expectations of adults, sexually abused children are usually physically incapable of resisting their abusers despite being opposed to the abuse. This is due to the tremendous amount of psychological power that most abusers have over the child. This usually precludes the need for physical force. Dr Summit emphasized that an imbalance of power exists between the perpetrator and the abused child. Most abusers are individuals who occupy important positions of trust and control in the child's life. He cited studies that suggested that a child was more likely to be sexually abused by close relatives such as their fathers, stepfathers, and other important caretakers. Although it was easy for the child to resist "strangers, weirdos, and kidnappers [1]", it is often not so easy to resist important caretakers within "an authoritarian relationship [1]." As such, sexually abused children often quietly succumb to the abuse without putting up any resistance.

Dr Summit challenged the common assumption by adults that a child's failure to resist or report sexual abuse suggests that the child was a consenting partaker in the abusive relationship. Especially erroneous is the suggestion that the abused child might have initiated the abuse by some seductive and sexually provocative behavior. He emphasized that the perpetrating adult was solely responsible for "any clandestine sexual activity" with a trusting child who is usually helpless in resisting the abuse. This helplessness is created by the psychological position of power that the perpetrator often occupies in the child's life.

Entrapment and Accommodation

Dr Summit hypothesized that child sexual abuse is often repetitive and occurs over a prolonged period of time. The perpetrator often experience some initial regret, guilt, or fear of discovery that leads to a resolution to stop the abuse. However, the "forbidden quality of the experience [1]" and the ease with which the abuse is achieved leads to a compulsion to repeat the acts. If the child does not receive help or protection from the abuser, she finally alters her mental state so as to continue to survive the act while maintaining her emotional well-being.

Dr Summit explained that the child accommodates to the abuse by learning to accept the abuse. This is achieved by mentally shifting the burden of blame from the perpetrator to the child. The child experiences significant internal conflict due to the contradiction between the perpetrator's image as a loving and protective caretaker versus the unfathomable nature of their acts of betrayal. Being unable to deal with the pain, the child succumbs to the mental torture and turns the blame on herself. She accepts the responsibility for the acts and makes attempts to continue to be "good" so as to ensure that she continues to

receive the abusing adult's love and protection. This she achieves by continuing to yield to the perpetrators sexual demands as well as by maintaining the code of secrecy and silence. The battled for internal control is often evident in the behavior of abused childs, which are usually characterized by "self-punishment, self-mutilation, selective restructuring of reality and multiple personalities [1]." The child then manifests these conflicts by an increased potential for suicidal behavior, sexual promiscuity, and frequent attempts to escape from the pain of abuse by running away.

Dr Summit postulated that the psychological reaction of male victims is often different from that of female victims of child sexual abuse. Unlike female victims, male victims usually externalize their internal conflict by becoming outwardly aggressive, violent, and antisocial. The male victim is more likely to view his abuse as mutually beneficial, which might lead to a continued wish for an "idealized relationship" with the perpetrating adult. This might ultimately lead the male victims to engage in their own acts of sexual molestation of other children or adults.

Delayed, Conflicted, and Unconvincing Disclosure

Dr Summit surmised that reporting sexual abuse is "the exception rather than the norm [1]." Most cases involving young children are reported either following incidental discovery by a third party or as a result of "sensitive outreach and community education [1, 3]" by child protective agencies. This usually happens when another concerned adult such as a schoolteacher or a neighbor observe physical or behavioral changes in the child that leads to suspicion. Upon questioning, the child might then hesitantly reveals the details of the abuse to investigating authorities. Dr Summit asserted that in adolescents, disclosure might occur following a period of intense conflict between the child and her family. As the adolescent victim matures and seeks outside relationships, the perpetrator (who is most often the father) becomes possessive and imposes strict rules aimed at "protecting" the child from "the dangers of outside peer involvement [1]." The child responds to this attempt to limit her freedom by becoming delinquent. This in turn leads to harsh punishment from the perpetrator. Anger and resentment then develops in the child, due to the inability to gain freedom. Disclosure usually occurs following an episode of severe punishment by the perpetrator. However, because of accommodation to the child's maladaptive pattern of delinquent behavior prior to disclosure, her credibility is questioned. The child's timing of disclosure after several years of abuse is often interpreted as retaliation for reasonable discipline issued by the perpetrator. This notion is reinforced by the lack of corroboration of the child's story by other family members.

Dr Summit described an alternative pattern of accommodation involving adolescents who divert their anger into positive avenues. Rather than becoming delinquent, these girls are usually successful both in their academic and personal lives. They are usually high achievers who are very popular with teachers and peers. When they finally decide, due to internal conflict, to disclose the details of their abuse, their story is often greeted with disbelief. The rationale is that it is impossible for a child to adjust so well after such a sordid experience.

Retraction

Dr Summit asserted that a child often reverses her initial reports of sexual abuse once the abuse has been reported to investigative and prosecuting authorities. At the initial point of disclosure, the child is not aware that her report has very grave consequences for her family. However, as the investigation proceeds and prosecutors and family members pressurize the child, she begins to question the wisdom in pursuing a course that could lead to the end of her family unit, as she has always known it. In an attempt to protect her family, the child assumes the role of a martyr and retracts her initial report. She then concocts an explanation that she had made up the story because she was angry about being punished by the perpetrator. In earnest, the child assumes the role of the villain while letting the perpetrator of the hook. Dr Summit cautioned, "Unless there is special support for the child and immediate intervention to force responsibility on the father, the girl will follow the 'normal' course and retract her complaint."

Historical Controversy Involving the CSAAS

Following the publication of Dr Summit's article in 1983, the CSAAS has been embroiled in legal controversy. At one point, the controversy was so intense that Dr Summit wrote in 1992, "The CSAAS is a clinical observation that has become both elevated as

gospel and denounced as dangerous pseudoscience [4].” This controversy arose from an initial misunderstanding of the scope and purpose of the CSAAS.

After Dr Summit published his epic article in 1983, there was some excitement in the legal community. This excitement was greatest amongst prosecutors and other authorities involved in the prosecution of perpetrators of child sexual abuse. The syndrome was misunderstood and considered to be analogs to the battered child syndrome (*see* **Battered Child Syndrome**), which described the classic physical and radiological findings associated with the children who have been the victims of prolonged, severe physical abuse. Since Kempe and his colleagues [5] expounded the features of the battered child syndrome in 1969, it has been a useful tool in the successful prosecution of child physical abuse cases. Hence the initial fantasy was that the CSAAS would achieve similar status in sexual abuse cases. However, the point missed was that the CSAAS was developed as a tool to assist clinicians to better understand the unusual presentation of some children who have been the victims of sexual abuse and not as a diagnostic tool. Unlike the battered child syndrome, the CSAAS does not present a set of physical or psychological characteristics that can help to confirm that a child has been sexually abused. Rather, it represents a set of observations that can help adults to better understand certain unusual aspects of child sexual abuse such as delayed disclosure and retraction after initially reporting details of the abuse.

The rash of expert witness testimony utilizing the CSAAS as proof of child sexual abuse resulted in skepticism amongst jurist and jurors. The fact that the syndrome was not widely understood and accepted within the psychiatric community did not help its reputation. The result was that there were many prolonged legal battles in various jurisdictions, which were settled in courts of higher authority in many states. Although no legal arguments involving the syndrome has ever been addressed by the United States Supreme Court to date, it has been the center of contention in numerous cases addressed by some state supreme courts.

Legal Applications of the CSAAS

In the United States, the CSAAS has been utilized mainly in the criminal prosecution of child sexual abuse cases. There is no indication that it has ever been the central issue in any civil disputes to date.

Judicial opinion on the admissibility of the CSAAS varies depending on jurisdiction and its legal application. The two most common standards for admitting expert testimony in the United States are the Frye test [6] and the Federal Rules of Evidence (FRE) 702 (*see* **Expert Opinion: United States**; ***Frye v. United States***; **Evidence: Rules of**; **Alcohol: Behavioral and Medical Effects**) [7]. The Frye test requires that the scientific basis for expert testimony be generally acceptable within the relevant field. On the other hand, the FRE 702 allows expert testimony if it will “assist the trier of fact to understand the evidence or to determine a fact at issue”. It outlines four requirements for admissibility of expert testimony based on the case of *Daubert v‘ Merrell Dow Pharmaceuticals, Inc* [8] (Daubert standard) (*see* ***Daubert v. Merrell Dow Pharmaceuticals***) namely

1. The scientific theory or technique has been tested.
2. The scientific theory or technique has been subjected to peer review and publication.
3. The scientific theory or technique has known or potential error rates and is subject to standards controlling its existence and maintenance.
4. The scientific theory or technique has general acceptance in the relevant scientific field (the Frye test).

Most of the legal analysis of the admissibility of the CSAAS has been related to its admissibility under the Daubert standards.

Admissibility to Support a Claim of Sexual Abuse

Many high courts in the United States have rejected expert testimony that utilizes the CSAAS to support a child’s claim of being sexually abused. In *State v. Foret* [9] a man was convicted of attempted molestation of a juvenile under his control, based on expert testimony that utilized the CSAAS to support the truthfulness of the victim’s claims. The conviction was affirmed by the Louisiana Court of Appeals but rejected by the Louisiana Supreme Court following a Daubert analysis. The Louisiana Supreme Court reasoned that the CSAAS was based on psychodynamic theories that are difficult to test. The court also noted that the CSAAS was not developed as

a diagnostic test for the detection of sexual abuse and that it was not generally accepted within the field of clinical psychology and psychodynamic theory from which it was developed. Similarly, the New Jersey Supreme Court rejected expert testimony that presented the CSAAS as substantive evidence that child sexual abuse occurred in *State v. J.Q* [10]. In that case, which involved the allegations by two sisters that their father sexually abused them, the court pointed out that the CSAAS was not diagnostic of child sexual abuse. Rather, the court suggested that expert testimony would have been admissible if had it had been used to demonstrate that delayed reporting and later recantation of an allegation of sexual abuse was not inconsistent with a history of such abuse.

Finally, in *Commonwealth v. Dunkle* [11], the Pennsylvania Supreme Court ruled that the CSAAS was not admissible for any legal purpose because it had not gained general acceptance in the particular field in which it belonged and because its symptoms were not specific to children who have been sexually abused.

Admissibility for Rehabilitation of Victim Credibility

Legal opinion varies on the admissibility of the CSAAS for the rehabilitation of the credibility of a victim of child sexual abuse. The majority view is that it is admissible for this purpose. The Delaware Supreme Court in *Wheat v. State* [12] addressed admissibility of the CSAAS to rehabilitate the credibility of a child victim of sexual abuse. In that case, a man was accused of sexually abusing his 10-year-old stepdaughter. The victim later recanted the abuse following pressure from the perpetrator's ex-wife, but insisted during trial that she was abused. Expert witness testimony was used by the prosecution to demonstrate that it was not unusual for children who have been the victims of sexual abuse to recant their stories under pressure from family members. The perpetrator appealed the jury's guilty verdict on grounds that expert testimony on the CSAAS should not have been admissible. Although the Delaware Supreme Court reversed the conviction, it ruled that expert testimony on the CSAAS to assist in rehabilitating the victim's credibility was admissible as it assisted the jury to better understand the behavioral characteristics of sexually abused child especially as it relates to the recantation of their initial claims of abuse.

On the other hand, the Pennsylvania Supreme Court rejected the use of the CSAAS in the rehabilitation of victim credibility in *Commonwealth v. Dunkle*. It ruled that victim credibility was the sole responsibility of the trier of fact and that was outside the domain of expert witness testimony.

Summary

The CSAAS was conceptualized to help mental health practitioners to better understand the abnormal behavioral characteristics of children who have been sexually abused. It has been misunderstood by some as diagnostic of child sexual abuse. This has led to its misuse in the prosecution of child sexual abuse cases. The overall effect is that it is now viewed with skepticism by some legal authorities. When properly applied, it can be used to rehabilitate the credibility of a child sexual abuse victim. It can also be used to assist Juries to better understand the contradictions in the victim's behavior. The general consensus is that it is not admissible as proof that allegations of abuse are true.

References

[1] Summit, R.C. (1983). The child sexual abuse accommodation syndrome, *Child Abuse and Neglect* **7**, 177–193.

[2] Garrison, A.H. (1998). Child sexual abuse accommodation syndrome: issues of admissibility in criminal trials, *Institute of Psychological Therapies (IPT)* **10**, 1–19.

[3] Children's Law Office (a program of the University of Southern California School of Law) (1996). Children's law report of the University of Southern California law school, *Children's Law Report* **1**, 4.

[4] Summit, R.C. (1992). Abuse of the child sexual abuse accommodation syndrome, *Journal of Child Sexual Abuse* **153**, 153.

[5] Kempe, H.C., Silverman, F.N., Steele, B.F., Droegemueller, W. & Silver, H.K. (1962). The battered-child syndrome, *Journal of the American Medical Association* **181**(1), 105–112.

[6] Frye v. United States, 54 App. D. C. 46, 47, 293 F. 1013, 1014 (1923).

[7] The United States Federal Rules of Evidence, (1975).

[8] Daubert v. Merrell Dow Pharmaceuticals (92–102), 509 U.S. 579 (1993).

[9] State v. Foret, 628 So.2d 1116 (La., 1993).

[10] State v. J.Q, 130 N.J. 554, 617 A.2d 1196, (N.J. 1993).

[11] Commonwealth v. Dunkle, 602 A.2d 830, 837 (Pa. 1992).

[12] Wheat v. State, 527 A.2d 269275 n.5 (Del. 1987).

Related Articles

Syndromes: Psychological

SYLVESTER SMARTY

Childbirth Depression *see* Postpartum Psychosis

Children: as Defendants

Introduction

In the past, numerous legal systems, including the United States', have recognized that juvenile offenders should be treated differently than adult offenders, particularly with regard to criminal responsibility.[a] Despite a lack of systematic research to support this notion, these legal systems recognized that because of juveniles' developmental immaturity, they were inherently more impulsive, exhibited worse judgment, and were therefore less blameworthy than adults who committed similar acts. Through the concept of "penal proportionality", they were subsequently subject to less harsh penalties. Society's withholding of certain privileges until youth reached certain ages provided additional implicit support for this line of reasoning and the eventual establishment of a separate juvenile justice system to both adjudicate and rehabilitate juveniles.

The first juvenile court in the United States was established in 1899 in Chicago by the Illinois legislature and Progressive reformers. These reformers viewed juveniles as having diminished (or absent) culpability and more malleable, "unformed character" than adult offenders. Therefore, the juvenile courts initially focused primarily on rehabilitation rather than the other penal objectives, namely incapacitation (i.e., the protection of society), deterrence (both specific (that offender), and general (other potential offenders)), and retribution. *Parens patriae* (literally "parent of the state") rather than *police power* was the operative doctrine. Proceedings were typically less adversarial and more informal than in adult criminal court and tended to focus on the nature of the offender rather than the offense itself. Prosecutors, defense attorneys, and judges often worked in a collegial manner in order to serve the best interests of the youth. Because the court fashioned itself strictly rehabilitative and proceedings were considered civil rather than criminal in nature, due process protections were generally not considered important.

In their first 65 or so years of existence in the United States, juvenile courts generally operated unfettered by constitutional mandates that applied to adult criminal court proceedings. Judges had a great deal of discretion with regard to almost every aspect of delinquency proceedings (including the ultimate disposition of youths adjudicated delinquent) and there were some abuses of power. Because of these abuses, the lack of due process protections in juvenile court proceedings was challenged in a number of court cases in the 1960s. The U.S. Supreme Court's ruling in *In re Gault* (1967) extended due process protections to delinquency proceedings [1]. The subsequent (and necessarily altered) version of the juvenile justice system probably acknowledged more accountability in youths than its prior incarnation. Nonetheless, the special character and mission of the juvenile justice system were, for the most part, preserved.

However, over the past 20 years, there has been a fundamental change in the way juvenile defendants are viewed by and subsequently have interacted with the legal system. What happened and why? In the late 1980s, there was a marked increase in the rate of crimes committed by juveniles, particularly homicide. Over time this led to a change in the public perception of minors involved in the juvenile justice system. These youths began to be seen as budding psychopaths who needed adult punishment rather than troubled children in need of rehabilitation. Catchy slogans such as "adult time for adult crime" and "old enough to do the crime, old enough to do the time" were promulgated.

Subsequently, a great deal of public pressure was put on state and federal legislators to use juvenile courts (and adult criminal courts) as mechanisms for social control (i.e., protecting the public) rather than rehabilitation of juvenile offenders. Juveniles began

to be transferred (also known as *waived*) to adult criminal court at much higher rates. There they faced adult sentences potentially as harsh as life without parole or death (at least until the 2005 US Supreme Court decision in *Roper v. Simmons*) [2]. Those lucky enough to remain under juvenile court jurisdiction faced much stiffer sanctions there (e.g., in California, incarceration up to age 25). In short, juveniles were forced to negotiate an increasingly unforgiving juvenile justice system or an adult criminal justice system that was not designed or prepared to deal with them. The latter did not have experience with the unique challenges that psychosocial immaturity and "unformed character" presented with regard to the adjudicative process (which encompasses all proceedings from initial arrest through sentencing) and criminal responsibility.

Although the proverbial pendulum may have started to swing back over the past decade because of a variety of factors [3], this sort of thinking is still evident in a great many jurisdictions.

There has been a positive, albeit unforeseen result of society's shift to a more punitive stance toward juvenile offenders: renewed attention has been brought to juveniles' developmental immaturity and its subsequent potential effect on their abilities to understand and participate meaningfully in the adjudicative process (a general definition of adjudicative competence). Concerns about juveniles' adjudicative competence certainly existed previously (particularly after the *Gault* decision). However, because of the less adversarial nature of juvenile proceedings and the less severe penalties faced by juveniles in the past, the potential effects of their developmental immaturity on the fundamental fairness of the adjudicative process were not thoroughly explored and quantified until quite recently. Recent studies have raised serious doubts about many juvenile defendants' abilities to understand pretrial and trial proceedings and to participate meaningfully in their defense, particularly those who are young and have lower IQs [4].

With this background in mind, we can now consider what light the existing, mostly recent, literature can shine on the topic "children as defendants" by considering the following questions:

- What are the characteristics of almost all juveniles that make their negotiating the current legal systems (both juvenile courts and adult criminal courts) potentially difficult?
- What are additional characteristics of juveniles likely to become involved with the justice system that may make their navigating the same legal systems even more daunting?
- How do (or should) these characteristics or limitations affect different components of the adjudicative process (including disposition/sentencing)?

Adjudicative Competence and Culpability: Limiting Factors

The juvenile adjudicative process can be started and completed in either juvenile court or adult criminal court and movement between the two venues is not uncommon (e.g., a juvenile initially under juvenile court jurisdiction can be transferred to adult criminal court or *vice versa*) (*see also* **Juvenile Justice: Transfer to Adult**). Typically, those minors transferred to (or "directly filed on" in) adult criminal court are 13- to 17-year-old and have been charged with serious crimes. If convicted, they face potentially lengthy sentences. "Amenability to treatment" or "rehabilitative potential" is usually cited as the most important criterion in determining "juvenile fitness" (i.e., whether a minor remains under juvenile court jurisdiction). However, issues of culpability are implicit in the decision to try a minor in adult criminal court (i.e., if a minor is facing adult sanctions, he must be as responsible for his actions as an adult).

Numerous features of juveniles (some of which are listed below) may influence their adjudicative competence and culpability. Obviously, given the potential jeopardy juveniles face in adult criminal court, it is extremely important to consider these factors in order to protect these defendants' various trial-related rights (e.g., 6th and 14th Amendment rights, right to be competent to stand trial, etc.) and to ensure that the punishment meted out is consistent with the blameworthiness of the perpetrator. However, these factors also need to be considered even in minors who remain under juvenile court jurisdiction. Although these individuals may be facing less serious sanctions, the aforementioned features are, in most cases, more descriptive of these younger juveniles and/or have a larger impact on their adjudicative competence and culpability.

General Developmental Immaturity

Society has long recognized that juveniles are generally more impulsive, more susceptible to peer pressure, more prone to immediate gratification, and more likely to engage in risky behavior than adults (*see also* **Juvenile Justice: Adolescent Development**). For this reason, individuals are generally precluded from participating in certain activities until they reach certain ages (e.g., driving, entering into contracts, serving in the military, and consuming alcohol).

Recently, researchers have begun to define more precisely particular domains of "impairment", quantify the degree of these impairments by age, and examine how these impairments may affect the adjudicative process and potentially impact criminal responsibility. Most adolescents manifest deficits in the following areas:

1. **Risk appraisal:**

Adolescents tend to discount and undervalue risk.

2. **Time perspective:**

Adolescents tend to care more about short-term consequences than long-term consequences and are less "future-oriented" than adults.

3. **Peer influence:**

Adolescents are much more likely than adults to be subject to and succumb to peer influence.

4. **Abstract thinking:**

Children and adolescents' perceptions and decisions may be based on overly concrete ideas; they may view rights as discretionary or conditional as opposed to automatic and inalienable.

5. **Perceived autonomy:**

Children and adolescents' lack of perceived autonomy can manifest itself as passivity, inattention, or compliance with authority.

6. **Stability of "character":**

Coherent integration of various elements of identity does not occur until early adulthood under the best of circumstances; adolescent antisocial behavior is extremely common and only a small percentage of adolescents continue this behavior into adulthood [5].

With regard to the above, Fried and Reppucci [6] demonstrated that resistance to peer influence, temporal perspective, and risk appraisal all had nadirs at approximately age 15–16. Interestingly, this is approximately the age at which the arrest rate peaks for both violent and nonviolent crime.

Neuroimaging has also proved a powerful tool in providing an anatomical explanation for at least a portion of these impairments. For example, Gogtay *et al.* [7] showed that central nervous system pruning in children and adolescents (a proxy measure of neuroanatomical maturity) proceeded in a parietal-to-frontal direction and that pruning continued until at least age 21 (this was the oldest individual examined). He also noted that the dorsolateral prefrontal cortex (sometimes referred to as the "superego" of the brain) did not *start* to prune until the end of adolescence. Bartzokis *et al.* [8] found that myelin volumes in the frontal and temporal lobes (another measure of neuroanatomical maturity) did not peak until age 45–50.

Mental Disorders

Youths in the juvenile justice system evidence much higher rates of mental disorders than youths in the general population. Approximately 66% of juvenile pretrial detainees or delinquents meet the Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition (DSM-IV) criteria for a mental disorder (including substance abuse/dependence or conduct disorder), though not all of these individuals require treatment [9].

In their meta-analysis studies involving juvenile justice populations, Otto *et al.* [10] noted high prevalence rates of mental disorders (see Table 1). Similarly, in a more recent study, Teplin *et al.* (see Table 2) noted relatively high rates of specific mental disorders in youth in Cook County, Illinois juvenile detention facilities.

Table 1 Prevalence of mental disorders amoung youth in the juvenile justice system

Conduct disorder	50–90%
ADHD	19–46%
Substance abuse	25–50%
Personality disorders	2–17%
Mental retardation	7–15%
Learning disorders	17–53%
Mood disorders	32–78%
Anxiety disorders	6–41%
Psychoses and autism	1–6%

Table 2 Six-month prevalence and odds ratios (ORs) of *DSM-III-R* diagnoses by sex with and without diagnosis-specific impairment criteria[a]

Disorder	Male, % (95% CI) ($n = 1170$)		Female, % (95% CI) ($n = 656$)		Female–Male ORs (95% CI)	
	Diagnosis	Diagnosis with impairment	Diagnosis	Diagnosis with impairment	Diagnosis	Diagnosis with impairment
Any of the listed disorders	66.3 (61.6–70.7)	63.3 (58.6–67.8)	73.8 (70.1–77.1)	71.2 (67.5–74.7)	1.43 (1.09–1.88)	1.43 (1.10–1.87)
Any except conduct disorder	60.9 (56.2–65.5)	59.7 (54.9–64.3)	70.0 (66.2–73.5)	68.2 (64.4–71.8)	1.49 (1.15–1.94)	1.45 (1.12–1.88)
Any affective disorder	18.7 (15.2–22.8)	16.1 (12.8–20.0)	27.6 (23.6–32.0)	22.9 (19.0–27.2)	1.66 (1.20–2.29)	1.55 (1.09–2.20)
Major depressive episode	13.0 (10.0–16.6)	11.0 (8.3–14.5)	21.6 (17.8–25.9)	18.9 (15.2–23.2)	1.85 (1.27–2.70)	1.88 (1.25–2.82)
Dysthymia	12.2 (9.3–15.8)	9.9 (7.3–13.2)	15.8 (13.1–18.8)	12.5 (10.2–15.3)	1.34 (0.93–1.95)	1.31 (0.87–1.96)
Manic episode	2.2 (1.1–4.3)	2.0 (1.0–4.1)	1.8 (1.0–3.2)	1.2 (0.6–2.4)	0.81 (0.33–1.99)	0.58 (0.21–1.63)
Psychotic disorders	1.0 (0.4–2.6)	–	1.0 (0.5–2.1)	–	0.98 (0.30–3.25)	–
Any anxiety disorder	21.3 (17.6–25.6)	20.7 (17.0–24.9)	30.8 (27.2–34.6)	28.9 (25.5–32.7)	1.64 (1.22–2.20)	1.56 (1.16–2.10)
Panic disorder	0.3 (0.1–0.6)	0.1 (0.0–0.4)	1.5 (0.8–2.7)	1.0 (0.5–2.0)	5.65 (2.04–15.65)	8.13 (2.01–32.85)
Separation anxiety disorder	12.9 (9.9–16.5)	10.8 (8.1–14.2)	18.6 (15.7–21.9)	16.3 (13.6–19.4)	1.55 (1.08–2.21)	1.61 (1.10–2.34)
Overanxious disorder	6.7 (4.6–9.7)	5.9 (4.0–8.7)	12.3 (9.9–15.1)	11.5 (9.2–14.2)	1.95 (1.23–3.10)	2.06 (1.27–3.35)
Generalized anxiety disorder	7.1 (4.9–10.2)	6.4 (4.3–9.4)	7.3 (5.6–9.6)	6.8 (5.1–9.0)	1.03 (0.63–1.69)	1.07 (0.64–1.79)
Obsessive-compulsive disorder	8.3 (6.1–11.3)	–	10.6 (8.4–13.2)	–	1.31 (0.86–2.00)	–
Attention-deficit/hyperactivity disorder[b]	16.6 (13.3–20.5)	11.2 (8.5–14.6)	21.4 (18.4–24.8)	16.4 (13.7–19.5)	1.37 (0.99–1.89)	1.55 (1.07–2.25)
Any disruptive behavior disorder	41.4 (36.8–46.2)	31.4 (27.2–36.0)	45.6 (41.4–49.8)	38.0 (33.9–42.2)	1.19 (0.92–1.53)	1.33 (1.02–1.75)
Oppositional-defiant disorder	14.5 (11.4–18.2)	12.6 (9.8–16.2)	17.5 (14.7–20.6)	15.1 (12.5–18.1)	1.25 (0.89–1.76)	1.23 (0.86–1.76)
Conduct disorder	37.8 (33.3–42.6)	24.3 (20.5–28.5)	40.6 (36.5–44.8)	28.5 (24.6–32.8)	1.12 (0.86–1.46)	1.24 (0.92–1.67)
Any substance use disorder	50.7 (45.9–55.5)	–	46.8 (42.6–51.1)	–	0.86 (0.66–1.11)	–
Alcohol use disorder	25.9 (21.9–30.4)	–	26.5 (22.6–30.9)	–	1.03 (0.76–1.40)	–
Marijuana use disorder	44.8 (40.1–49.6)	–	40.5 (36.8–44.4)	–	0.84 (0.65–1.08)	–
Other substance use disorder	2.4 (1.7–3.4)	–	6.9 (4.1–11.4)	–	3.00 (1.57–5.74)	–
Both alcohol and other drug use disorders	20.7 (17.0–24.9)	–	20.9 (18.0–24.2)	–	1.01 (0.75–1.38)	–

[a] CI indicates confidence interval. Ellipses indicate that diagnosis and diagnosis with impairment are identical because the diagnosis criteria for psychotic dis obsessive-compulsive disorder, and substance use disorders include impairment

[b] Attention-deficit/hyperactivity disorder is without the criterion of onset before age 7 years because caretaker information is not available and self-resymptoms before age 7 years unreliable

Reproduced from Ref. 9. © American Medical Association, 2002

Many youths with these mental disorders will meet minimum thresholds for adjudicative competence and criminal responsibility. However, these disorders frequently impact both the effectiveness of these youths' participation in the trial process and their culpability, particularly when considered in conjunction with their baseline developmental immaturity.

Generally, mental disorders have been thought to have less impact on the adjudicative process than developmental immaturity. Many scholars postulate that this is because psychotic disorders are relatively rare in minors (with a mean age of onset in the late teens or early 20s), whereas developmental immaturity is almost ubiquitous in juveniles. In a later section, we will examine how different mental disorders may impact specific stages of the adjudicative process.

Substance Use Disorders

Youths in the juvenile justice system have rates of substance use disorders (i.e., abuse or dependence) approximately 25 times higher than youths in the general population (50 *vs.* 2%) [9, 11]. Comorbid psychiatric disorders are the rule rather than the exception. As with adults, acute intoxication, withdrawal phenomena, and substance-induced psychiatric disorders (which may include psychotic disorders) can affect various stages and aspects of the adjudicative process. However, because youths' brains are still developing and immature, substance use may impact their capabilities even more than adults.

Effects related to substance use may also impact juveniles' culpability. Although acute intoxication generally disqualifies a defendant from utilizing the insanity defense, it may be germane to questions of mitigation or diminished capacity/actuality (i.e., did the defendant form the intent required to be convicted of the crime? (e.g., "premeditation" for first-degree murder)). Youths who demonstrate "settled psychosis" (i.e., psychotic symptoms that persist after acute intoxication) could potentially utilize the insanity defense.

Intelligence

For a variety of reasons, the average IQ of youths involved in the legal system is at least one standard deviation (i.e., 15 points) below that of youths in the general population. Most studies have found the mean IQ of juvenile justice populations to be 80–85. In a study of juvenile adjudicative competence, Ficke *et al.* [12] found that their 9- to 16-year-old detained study participants had an average IQ of 73. According to Otto's data, the rates of mental retardation (i.e., an IQ $\leq$ 70 and particular deficits in adaptive functioning) are approximately three to eight times higher than the rates in the general population.

Obviously, sub-average intelligence can have a significant impact on both the adjudicative process and culpability. The latter was recognized explicitly in the 2002 U.S. Supreme Court decision in *Atkins v. Virginia* [13], which banned the execution of the mentally retarded on Eighth Amendment grounds.

Other Factors

In addition to higher rates of substance use disorders and other mental disorders, youths in the juvenile justice system also have higher rates of head trauma, prenatal exposure to drugs and alcohol, attachment disorders arising from problems in infancy, dysfunctional and chaotic families and neighborhoods, and exposure to violence and abuse. It has not been systematically established whether or not these factors independently and appreciably impact either these individuals' ability to participate meaningfully in the adjudicative process or their culpability. However, it is certainly reasonable to suspect that these factors may exacerbate or even cause deficits in these domains.

Effects of Selected Factors on Different Aspects of the Adjudicative Process and Culpability

Arrest and Interrogation

Because of their developmental immaturity and lack of experience, youths are likely to view their rights (e.g., the "right to remain silent") as conditional and discretionary rather than automatic and inalienable. They are also more likely to trust and obey authority figures. These observations are not surprising. Juveniles' abilities to think abstractly may not be fully developed (e.g., they may be unable to conceptualize a "right") and children are frequently socialized both to "tell the truth" and believe that "the police are your friend." This trust and lack of sophistication

may not adversely impact these youths in their everyday lives and may in fact serve them well. However, these same characteristics can become liabilities in dealing with a police interrogation, particularly if the youth is charged with a serious crime.

In the previously cited MacArthur study, when youths were faced with a hypothetical interrogation situation, approximately 50% of 11- to 13-year-olds was their best interrogation choice ("talk/admit" rather than "talk/deny" or "remain silent") [4]. In the same study, only 15% of 18- to 24-year-olds thought this was the best option. Similarly, Grisso [14] found that only 40% of adolescents recognized that they would not have to make statements about their offense even if a judge ordered them to do so.

Other Pretrial Proceedings and Trial

Age (a proxy for developmental maturity), intelligence, and mental disorders can all significantly impact a juvenile's understanding, reasoning, and appreciation related to pretrial proceedings and the trial itself. The MacArthur study found that age and IQ were the variables most closely tied to scores on a standardized competence assessment instrument (the MacArthur Competence Assessment Tool-Criminal Adjudication (MacCAT-CA)). Approximately 30% of 11- to 13-year-olds and 19% of 14- to 15-year-olds scored below the "clinically significant impairment" threshold on the MacCAT-CA and would likely have been found incompetent. Over 50% of 11- to 13-year-olds in the "detained" sample with IQs between 60 and 74 scored below this threshold [4].

In addition to age, intelligence, and psychotic disorders, numerous other untreated or inadequately treated mental disorders in juveniles can also significantly impact pretrial and trial proceedings. For example:

Attention-Deficit/Hyperactivity Disorder (ADHD):

Juveniles may not be able to adequately attend to conversations with their attorneys or to court proceedings. Their courtroom demeanor may also be adversely affected.

Social Anxiety Disorder:

Juveniles may not even agree to meet with their attorneys and may have extreme difficulty tolerating open court.

Depression:

Juveniles may not have the motivation to defend themselves. In the past (i.e., before *Roper v. Simmons*), suicidality had been a factor in some death penalty cases.

Pervasive Developmental Disorders (e.g., Autistic Disorder, Asperger's Disorder):

Juveniles may appear to lack remorse or empathy for the alleged victim (can also be factor at sentencing/disposition).

Culpability

All of the aforementioned factors may diminish juveniles' culpability and subsequent deserved punishment. Researchers are beginning to uncover evidence of the neurobiological basis of this developmental immaturity and are gathering evidence about adolescents' impaired decision making in hypothetical criminal situations. In response to both this growing database and recently rediscovered common sense notions about the psychosocial immaturity of juveniles, courts and legislatures are beginning to reassess how minors interface with and are treated by the juvenile justice and adult criminal justice systems.

A variety of individuals can consider any or all of these factors at various junctures in the adjudicative process: the police when they initially detain the youth; the district attorney when he or she decides whether to file charges in juvenile or adult criminal court; the judge when determining whether to divert the juvenile's case to "drug court" or "mental health court;" the trier of fact when they decide about issues related to diminished capacity/actuality (more germane for substance use issues); and the same trier of fact when considering mitigating factors at sentencing. Hopefully, research and evidence rather than politics and fear will drive these individuals' decision processes.

Minors in Adult Criminal Court: Final Thoughts

It is important to note that some degree of adolescent antisocial behavior is normative. Arrest rates for violent and nonviolent crime peak around the age of 16 or 17, decrease quickly and linearly until age 30, and then continue to decrease each year thereafter,

albeit more slowly [5]. What does this mean? Adolescent antisocial behavior generally does not persist into adulthood and is, by definition, "adolescent-limited". As contingencies change and neurological maturation progresses, the vast majority of adolescents are able to desist from their criminal behavior. However, there is a small subset of individuals who engage in "life-course-persistent" antisocial behavior.

Incarcerating individuals for lengthy periods imposes a tremendous cost on society, both directly (e.g., cost to house inmate) and indirectly (e.g., institutionalization, loss of employment opportunities because of criminal record). These costs may be justified in order to protect society or serve other legitimate penological interests. However, it is questionable whether indiscriminately incarcerating minors for extended periods serves these penological interests. Is severe, inflexible punishment (i.e., retribution) a legitimate penological objective if the actor is less blameworthy (or, in extreme cases, not culpable at all)? Is incapacitation necessary if the antisocial behavior is likely to cease even without specific interventions?

Essentially, we must learn to be more selective in waiving juveniles to adult criminal court. How can this be accomplished? We need to become more sophisticated in our understanding of the genesis and maintenance of juvenile antisocial behavior. We need to develop and employ instruments that can more accurately predict which juveniles are likely to continue offending as adults (particularly violently offending). We need to examine the validity and predictive value of the concept of juvenile psychopathy. Research in this area is ongoing and will hopefully lead to more just legal outcomes for minors, help protect the general public and preserve the dignity and integrity of the legal process.

End Notes

[a.] For sake of ease, the terms *child, adolescent, youth, juvenile*, and *minor* are used interchangeably in this article. However, the reader should note that each has a slightly different definition and connotation. The terms *criminal responsibility* and *culpability* are also used interchangeably with the same caveats.

References

[1] *In re Gault*, 371 U.S. 1 (1967).

[2] *Roper v. Simmons,* 543 U.S. 551 (2005).

[3] Grisso, T. (2007). Progress and perils in the juvenile justice and mental health movement, *Journal of the American Academy of Psychiatry and the Law* **35**(2), 158–167.

[4] Grisso, T., Steinberg, L. & Woolard, J., Cauffman, E., Scott, E., Graham, S., Lexcen, F., Reppucci, N.D. & Schwartz, R. (2003). Juveniles' competence to stand trial: a comparison of adolescents' and adults' capacities as trial defendants, *Law and Human Behavior* **27**(4), 333–363.

[5] Moffitt, T. (1993). Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy, *Psychological Review* **100**(4), 674–701.

[6] Fried, C. & Reppucci, N. (2001). Criminal decision making: the development of adolescent judgment, criminal responsibility, and culpability, *Law and Human Behavior* **25**(1), 45–61.

[7] Gogtay, N., Giedd, J. & Lusk, L., Hayashi, K.M., Greenstein, D., Vaituzis, A.C., Nugent, T.F. 3rd, Herman, D.H., Clasen, L.S., Toga, A.W., Rapoport, J.L. & Thompson, P.M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood, *Proceedings of the National Academy of Sciences USA* **101**(21), 8174–8179.

[8] Bartzokis, G., Beckson, M. & Lu, P.H., Nuechterlein, K.H., Edwards, N. & Mintz, J. (2001). Age-related changes in frontal and temporal lobe volumes in men: a magnetic resonance imaging study, *Archives of General Psychiatry* **58**(5), 561–565.

[9] Teplin, L., Abram, K. & McClelland, G., Dulcan, M.K. & Mericle, A.A. (2002). Psychiatric disorders in youth in juvenile detention, *Archives of General Psychiatry* **59**(12), 1133–1143.

[10] Otto, R., Greenstein, J., Johnson, M. & Friedman, R. (1992). Prevalence of mental disorders among youth in the juvenile justice system, in *Responding to the Mental Health Needs of Youth in the Juvenile System*, J Cocozza, ed, National Coalition for the Mentally Ill in the Criminal Justice System, Seattle, WA, pp. 7–48.

[11] Shaffer, D., Fisher, P. Dulcan, M.K., Davies, M., Piacentini, J., Schwab-Stone, M.E., Lahey, B.B., Bourdon, K., Jensen, P.S., Bird, H.R., Canino, G. & Regier, D.A. (1996). The NIMH diagnostic interview schedule for children version 2.3 (DISC-2.3): description, acceptability, prevalence rates, and performance in the MECA study. Methods for the epidemiology of child and adolescent mental disorders study, *Journal of the American Academy of Child and Adolescent Psychiatry* **35**(7), 865–877.

[12] Ficke, S., Hart, K. & Deardorff, P. (2006). The performance of incarcerated juveniles on the MacCAT-CA, *Journal of the American Academy of Psychiatry and the Law* **34**(3), 360–373.

[13] *Atkins v. Virginia,* 536 U.S. 304 (2002).
[14] Grisso, T. (1981). *Juveniles' Waiver of Rights: Legal and Psychological Competence*, Plenum, New York.

Related Articles

Juvenile Justice: Adolescent Development

CHRISTOPHER THOMPSON AND JOSEPH KENAN

Children: as Witnesses

Each year, thousands of children participate as witnesses in the legal system. Although they may bear witness to many types of crimes, in the criminal courts they are most often involved as alleged victims of child sexual abuse (CSA). In fact, despite constituting only 10% of the 3 million annual reports of child maltreatment to child protective services, CSA cases account for the majority of criminal trials in which children testify. Child victim testimony is often the main evidence in CSA cases. Thus, child witness accuracy is of great concern to psychologists, lawyers, judges, and society as a whole. As is true for eyewitness testimony generally, inaccurate accounts by children can lead to false convictions. Yet if accurate accounts are disregarded because of the witness's age, guilty perpetrators could be free to commit future crimes. Of course, both undiscovered cases of actual abuse and convictions based on false accusations are serious injustices. It is therefore important to understand, from a scientific perspective, the accuracy of child eyewitness testimony.

Research on Children's Testimony

Research conducted at the beginning of the 20th century led to the conclusion that children were particularly unreliable, highly suggestible witnesses. However, much of the early research lacked ecological validity and was arguably not fully applicable to actual legal cases involving children. Eighty years later, with an increase in CSA reports (including high profile convictions of day care providers who were accused by preschoolers of ritualistic sexual abuse but later exonerated), psychological interest in children's eyewitness accuracy and suggestibility drastically increased (*see* **Children: Suggestibility of**). Some researchers were most concerned about the need to obtain reports of child abuse from actual child victims who needed protection but were reluctant to disclose abuse, whereas other researchers were most concerned about controversial cases resting upon possibly false accusations fanned by forensic interviewers' suggestive and widely criticized interview techniques. Scientific research on children's eyewitness testimony has clarified many of the issues of concern surrounding children's memory and suggestibility, although some topics remain subject to scientific debate. Important recent research that is more directly applicable to cases involving child witnesses can help ensure that accurate testimony is provided in true cases as well as false cases of child abuse.

Modern researchers have discovered a host of factors that influence child witness accuracy, including child age and individual differences, social circumstances, interview characteristics, and so forth. An important goal is to understand these factors, and use this understanding to aid the legal system by, for example, developing interview techniques that encourage accurate eyewitness reports from children.

Researchers have examined children's eyewitness testimony abilities by questioning children about documented traumatic and nontraumatic events. Studies reveal that younger children, especially preschoolers, are typically more suggestible than older children and adults; that is, on average, younger children are more likely to incorporate into their own reports incorrect information contained in an interviewer's questions.

Eyewitness memory can be affected by children's level of stress at the time of the event or at the time of recall. Children are particularly prone to error, for example, when they are young (e.g., 3- to 4-years-olds); exposed to misleading questions asked by a biased, coercive, or intimidating interviewer; positively reinforced for incorrect responses; asked to identify strangers from target-absent photo lineups; and asked about peripheral details of events or about events that happened in the distant past. An issue of scientific and applied interest is whether false suggestions actually change children's memory or simply their reports of an event (i.e., social compliance). Both might be possible. In documented controversial multivictim CSA cases that were arguably false,

anecdotal evidence reveals that some alleged victims recanted their claims years later, recalling that they were pressured to make accusations and lied rather than having false memories, but others maintained their allegations. Experimental research has shown that when questioned in extremely suggestive ways, some children appear to incorporate the misleading and inaccurate information into their memory, although many do not (e.g., when confronted with the truth about an event).

Child Forensic Interviews

Researchers have been successful in translating research findings into useful methods for interviewing suspected child abuse victims. In particular, research has led to the creation of developmentally appropriate, research-based forensic interview protocols. The acceptance of such protocols by social service and law enforcement agencies has made highly suggestive interviews the exception, not the norm, in actual forensic investigations of child abuse. Forensic interviews, often one of the earliest interactions between the alleged abuse victim and the legal system, are used to determine whether the child has been abused and to gather information about the circumstances surrounding the abuse. Information gained from this interview is critically important in determining case disposition, from taking no action in unsubstantiated cases, to removing children from possible harm in the home and/or pursuing a case to trial in substantiated cases. There are a number of forensic interview protocols, including the Step Wise Interview, the National Institute of Child Health and Human Development (NICHD) child interview protocol, and others, but all share basic elements such as recommending that children be questioned with developmentally appropriate language and with open-ended, nonleading questions to the extent possible. Many use a phased approach, starting with introductory phases – such as introduction and rapport building, a developmental assessment (including learning the child's names for different body parts), a discussion of interview guidelines of truth versus lie – before moving to open-ended prompting designed to obtain a child's narrative description of the events under investigation, and then more specific or cued recall follow-up questions, if necessary. The use of props during the interview, such as anatomically detailed dolls, anatomical drawings and diagrams, dollhouses, and puppets is usually discouraged altogether or at least until the end of an interview that has not yet yielded a disclosure. Even then, their use is recommended in conjunction with nonleading questions.

Generally, child forensic interviewers are trained to remain objective and use open-ended questions to improve child witness accuracy. For example, forensic investigators allow children to recall information freely in response to open-ended, nonleading questions (e.g., "What happened while you were in the house?"); then follow up with more specific or focused questions about the details children mentioned. Forensic interviewers use socially supportive interviewing techniques, which include building rapport with the child and providing emotional warmth and support throughout the interview while remaining neutral. Research demonstrates that children questioned in a socially supportive manner are more resistant to misleading questions than those questioned in an intimidating way. This is a particularly useful finding to the field of forensic interviewing, because unlike a child's age, cognitive development, or temperament, the interviewer's social support is easily controlled.

Some interview techniques are controversial; for example, multiple interviews of children over time rather than once. In actual cases, however, children sometimes disclose after several interviews. Research illustrates that multiple, highly misleading and coercive interviews can erode children's accuracy over time, although they do not necessarily do so. Multiple interviews conducted in a recommended manner can result in increased accuracy.

Accommodating Children's Testimony in the Forensic Context and Courtroom

Child advocacy centers (CACs) can aid children who find themselves involved in the legal context. CACs are "one-stop shops" where children who are victims of alleged maltreatment can often receive services from social workers, police officers, prosecutors, and physicians all at one location. Forensic interviews are conducted at these locations by trained interviewers, and the interview is observed by authorities (e.g., police and prosecutors) from behind a one-way mirror to ensure that all necessary questions are asked. In some jurisdictions, interviews are videotaped, and

these tapes can be used under certain circumstances at preliminary hearings or trials. CACs potentially decrease the number of times children need to be formally interviewed. By centralizing the resources and investigation, CACs aim to reduce child stress and provide necessary services as fast as possible.

Many legal cases involving children are resolved through confessions, settlements, and plea bargains. Only a relatively small percentage of cases go to trial. For those cases that do continue to trial, a number of hearings will be held, some of which may involve the children testifying. Testifying in court can be stressful for adults, not to mention for vulnerable children. Facing the accused perpetrator, experiencing harsh and confusing cross-examination designed to discredit their accounts, and having to provide testimony about personal and possibly embarrassing events in open court are just some of the factors contributing to children's fears and anxiety about the courtroom. Studies have shown that testifying multiple times is associated with emotional distress in both the short and long term, especially in severe intrafamilial CSA cases. Furthermore, although testifying can be helpful for some children, for others it may cause a delay in emotional recovery. There is also concern about how the stress caused by courtroom testimony will affect children's memory reports and accuracy. Several procedural reforms have been studied as potential ways to make testifying less stressful for children while bolstering the accuracy of their memory reports, with the additional goal of promoting fair trials for defendants.

One way to reduce many children's fears and anxiety is to allow them to present testimony in a manner that does not require them to face the defendant. Several methods to avoid face-to-face confrontation have been studied. Special legal exceptions sometimes allow adults who recount a child's out-of-court statements to testify in place of an alleged child victim. In light of Crawford v. Washington, however, the child must to be available for cross-examination at trial. Such testimony is considered hearsay. Videotaped forensic interviews of children are also considered hearsay evidence; presenting the videotape at trial can reduce anxiety in children by preventing them from having to testify in open court. However, at least in the United States, in light of the Crawford v. Washington decision, if the hearsay statements are made to an "authority", they are deemed "testimonial" and cannot be entered as evidence in criminal court trials unless the child also appears. In contrast, in many other countries, such hearsay is regularly admitted without the need for children to take the stand.

Child testimony via closed-circuit television (CCTV) has also been proposed to protect children from seeing the defendant face-to-face. This technique can reduce children's anxiety as well as increase their memory report accuracy. The US Supreme Court ruled that this method can be used at trial after a case-by-case determination by the judge, at least in certain CSA cases. Using CCTV for children is quite limited in the United States, although it is more common in several other countries (e.g., the United Kingdom).

Jurors' Reactions to Child Witnesses

From a legal perspective, even if children's memory reports are sometimes flawed, as long as fact finders (e.g., jurors) can reach the truth, justice is served. From the initial police investigation and forensic interview to the courtroom, adults must decide whether children's reports are accurate. Research shows that many adults have a difficult time making this judgment. This is not surprising, because adults are often poor detectors of lies and truth, accuracy and inaccuracy. Researchers have studied factors aside from actual accuracy that influence adults' perceptions of children's credibility, with the goal of understanding how jurors make decisions in cases involving child witnesses. This research reveals that victim, juror, and case characteristics all influence jurors' judgments. For example, in CSA cases, jurors in mock trial studies often perceive younger children and children with mild intellectual disabilities as more credible than older children and children without disabilities, because young and disabled children are believed to be particularly honest, sexually naïve, and lacking in the cognitive abilities thought to be necessary for fabricating false charges. If highly suggestive interviewing of young children is demonstrated, however, younger children's credibility may fall below that of older children. Juror characteristics, such as gender, also influence perceptions of children's accuracy. Compared to men, women on average have more empathy for child victims, react more negatively to CSA, and believe children's reports more often. In turn, women tend to

be more punitive in their case judgments than men. Furthermore, when made known to fact finders, certain defendant characteristics, such as having a record of past offenses, can also increase the likelihood of convictions.

Courtroom factors also influence jurors' judgments. For example, courtroom innovations such as allowing children to testify via CCTV rather than in open court can lower jurors' belief in children's testimony. Children may appear more distant or perhaps more relaxed when testifying in this alternative manner, in the latter case, violating jurors' expectations that actual child victims will be upset during testimony. Expert testimony from psychologists can be useful in educating jurors about such issues.

Future Directions in Child Witness Research

The newest research directions include work focused on understanding individual differences in children's eyewitness memory and their responsiveness to various forensic interview techniques. For example, research has shown that some children may benefit more from socially supportive interviewing techniques than others. Most likely to benefit are children who have a low level of social support from friends and family, are insecurely attached to their caregiver, and are highly reactive physiologically to stressful situations. Research has also begun to use sophisticated neuroscience methods to understand the physiology of memory and suggestibility. In addition, researchers are currently addressing controversies regarding the ways in which children initially disclose CSA, and whether children delay and recant accusations. That is, although the media often focus on false reports of child abuse, a larger problem likely is that most abused children never disclose that they were sexually abused. Evidence of nondisclosure comes, for example, from research on nondisclosure in cases where there is incontrovertible evidence of abuse, such as the child having a sexually transmitted disease. Due to lack of disclosure, most cases of CSA are never investigated by authorities. Children fail to disclose, for example, because they are embarrassed, afraid of losing their family, afraid of retribution, or because they do not even understand that the abuse is wrong. Further, some children who disclose may later recant, especially children who are younger, who have experienced intrafamilial abuse, and who have unsupportive caregivers.

Because individual differences exist across children in disclosure and memory, it is essential that the legal system is flexible in providing a supportive environment in which children can testify fully and accurately, based on children's abilities and needs. The court must carefully guard these young witnesses while still protecting the rights of those accused.

Further Reading

Bottoms B.L., Kovera M.B. & McAuliff B.M. (eds) (2002). *Children, Social Science, and the Law*, Cambridge, New York.

Bruck M., Ceci S.J. & Principe G. (2006). The child and the Law, in *Child Psychology in Practice*, Renninger K.A. & Sigel I.E. volume editors. volume 5, in *Handbook of Child Psychology*, 6th Edition, W. Damon & R. Lerner general editors, John Wiley & Sons, New York, pp. 776–816.

Eisen M., Quas J.A. & Goodman G.S. (eds) (2002). *Memory and Suggestibility in the Forensic Interview*, Erlbaum, Mahwah, NJ.

Goodman G.S., Taub E.P., Jones D.P., England P., Port L.K., Rudy L. & Prado L. (1992). Testifying in criminal court. *Monographs of the Society for Research in Child Development* **57**(229), 1–152.

Myers J.E.B. (1997). *Evidence in Child Abuse and Neglect Cases*, 3rd Edition, John Wiley & Sons, New York.

Pipe M.E., Lamb M.E., Orbach Y. & Cederborg A.C. (2007). *Child Sexual Abuse: Disclosure, Delay, and Denial*, Erlbaum, Mahwah, NJ.

Lamb, M.E., Hershkowitz, I., Orbach, Y. & Esplin, P.W. (2008). *Tell Me What Happened: Structured Investigative Interviews of Child Victims and Witnesses*, Wiley, New York.

SYLVIA PERRY, CATHERINE YORK, BETTE L. BOTTOMS, STEPHANIE BLOCK, AND GAIL S. GOODMAN

Children: Cross-Examination of

see Cross-Examination: Impact on Testimony

Children: Suggestibility of

Within the field of psychology and law, the term *children's suggestibility* typically refers to the extent to which internal and external factors can inadvertently or intentionally induce children to adopt and incorporate misinformation into their memory reports. Concerns about children's suggestibility are common in legal cases that rely heavily on children's eyewitness memory, such as in prosecutions of child sexual abuse and in custody cases involving allegations of child maltreatment. (*see* **Children: as Witnesses**).

Such concerns are especially prominent when there is no evidence aside from the child's statements and when the child may have been questioned in a leading manner (e.g., by a parent, clinician, or forensic interviewer). Although researchers debate the extent to which suggestion actually alters memory (e.g., having an existing memory overridden by a new one, or creating a brand new but false memory for an event that never took place), it is clear that individuals' eyewitness memory reports are susceptible to adverse effects of social influence. Either way (from memory change or social influence), a false report can result.

Both adults' and children's reports about their eyewitness memories can be altered or affected through the use of suggestive techniques, but young children (e.g., preschoolers) are, on average, especially vulnerable to the effects of misinformation and suggestion. However, children in general tend to be less suggestible about negative taboo events than positive or neutral events, although it is still possible to mislead at least a subset of children about highly negative incidents, including ones that are implausible. Children's increased resistance to suggestion about negative events is especially relevant in a forensic context because, in the legal system, children are typically asked to provide eyewitness reports about negative experiences.

The length of time between an event and the first interview, as well as the number of subsequent interviews, may influence children's suggestibility and thus the accuracy of their reports. When memory is still fresh, resistance to suggestion is increased, albeit not eliminated. In terms of accuracy, it is best for children to be interviewed as quickly as possible after the event by a trained neutral interviewer. However, if children have been severely traumatized, the optimal delay may be less clear. Although repeated interviewing can increase the amount of information provided, it can also increase suggestibility if, for example, children's memory is weak and multiple misleading influences are placed on the children.

Suggestibility and Interview Techniques

To reduce suggestibility in the forensic setting, it is important to conduct appropriate interviews with children. Although children may fall victim to suggestive influences in the home (e.g., by emotionally motivated or well-intentioned caregivers) or in other contexts outside of a legal setting, clinicians, lawyers, and forensic interviewers should attempt to obtain accurate information by reducing misleading questions in child interviews. Researchers have examined interview techniques that contribute to children's suggestibility, and specifically to children's false reports. Researchers have also evaluated ways to question children about events in non or less-suggestive manners in hopes of developing optimal interview strategies for child witnesses.

In regard to interview techniques that contribute to children's suggestibility, several examples are provided here: First, false information presented before or after a target memory event can lead to a child having more difficulty remembering the actual (i.e., true) details. If an interviewer presents false information to a child after a target event, the child might incorporate some of this false postevent information into his/her memory report of that target event. One explanation for this phenomenon relies on the notion of "source monitoring error." The child misattributes the source of the false information to actual experience and then incorporates the false information into his/her memory report. Adults are susceptible to source monitoring errors as well. When it occurs, child and adult witnesses might not be consciously aware of this type of memory error.

A second example of an interview technique that can contribute to children's suggestibility concerns indirect parental or peer pressure. Children can feel pressure to conform to or agree with authority figures as well as with peers. As a result, children can be susceptible to an interviewer's implication

of parental and peer statements. If a child witness is told by the interviewer that his/her parent said something happened or that all of the child's classmates said it happened, the child witness might try to conform his/her answers to the interviewer's suggestions.

A third method of potentially influencing a child's memory report in a negative way is referred to as *stereotype induction*. This involves, for example, telling a child that the accused is a "bad" person. A parent or interviewer might tell the child that someone was arrested, and "only very bad people get arrested," which may imply to the child that the person must have done horrible things. Giving children negative stereotypes about individuals can influence the children's memory reports in a way that is consistent with the stereotype.

Turning now to ways to question children in less-suggestive manners, although currently the United States does not have generally accepted, national interview guidelines, there are best practice guidelines that have emerged from decades of research in the field of children's suggestibility. Open-ended, free recall questions are considered the least suggestive and have been empirically shown to result in the most accurate information. Thus best practice guidelines typically preference such questions. However, in the forensic context, young children often fail to provide lengthy narrative responses to these types of questions. According to best practice guidelines, after free recall, more specific cued recall questions can be asked followed, if necessary, by more direct questions (the "funnel approach"). It is considered best to avoid yes/no questions, especially with young children. It is also recommended that all interviews be videotaped so that fact finders and experts can evaluate the extent to which interviews are potentially suggestive. Some countries have national interview guidelines. For instance, Israel uses the National Institute of Child Health and Human Development (NICHD) protocol which relies on establishing rapport with the child, practicing answers to (innocuous) free recall and open-ended questions, and introducing the target event in a neutral, nonsuggestive way; the interviewer waits until the end of the interview to use specific, option-posing questions, if such questions are deemed necessary.

Another widely used interview technique that attempts to reduce adverse effects of suggestion while promoting accurate and complete reports is the revised cognitive interview (CI). Originally developed for adults, this interview also has a rapport building phase and then relies on having the interviewee engage in "mental reconstruction" of the target event. Next, the witness is asked to free recall the event from beginning to end, and then in reverse order from the end to the beginning (change order). Last the interviewee is asked to recall the event from someone else's perspective (change perspective). However, concerns have been raised about young children's ability to perform the last two steps (change order and change perspective). Consequently, revised CI interviews for children have often excluded either or both of these steps. Both the NICHD and revised CI protocols are useful in reducing suggestibility in interviews with children aged four years and older.

Props, dolls, and human figure drawings can help children by serving as external retrieval cues, but such aids can also be suggestive in nature. Critics of such retrieval tools are concerned that dolls and props themselves are misleading, especially if they are used to suggest inaccurate information. Research on these aids indicates that overall they enhance children's reports by increasing the amount of accurate information provided, but that they also increase the number of errors in memory reports depending in part upon the child's age. Younger children, especially, might be too cognitively immature to understand the dual representation required to effectively use such retrieval tools. Of note, critics have argued that, use of anatomically detailed dolls may heighten suggestibility by prompting the child to demonstrate sexual acts with the dolls; these acts may be mistakenly interpreted as evidence of child sexual abuse.

Individual Differences

Individual traits of children may predict increased susceptibility to suggestion. It should be emphasized, however, that individual differences among children are often inconsistently associated with suggestibility, and the predictive power of such differences is typically too weak for legal purposes (i.e., many children would be misclassified). That said, language ability has been shown to be associated with suggestibility, perhaps especially in younger children. Children with more developed language skills tend to be more resistant to suggestive questions. With better language

skills, children may be more capable of understanding the questions and of understanding instructions that they do not always have to agree with the interviewer. In contrast, children with mental retardation are more highly susceptible to suggestive questioning, although even here, the research literature is mixed.

Parental attachment may influence how a child thinks and talks about negative events. Parents who are more avoidant in their attachment style may be less likely to talk with their children about negative life experiences. Research has shown that parental avoidant attachment predicts errors in children's memory reports of stressful events. Moreover, for child victims of sexual abuse, emotional support from parents after the child discloses the abuse is associated with more accurate reports (years later in adulthood) compared to that of children whose parents were more avoidant and attempted to minimize or ignore their child's disclosure.

Children's inhibition skills (e.g., ability to ignore irrelevant stimuli) are also associated with their memory reports. Children with deficient inhibition skills evince weaker overall memory and more inaccuracies in the information they provide to both direct and misleading questions. Thus, such children are more likely to be susceptible to suggestive techniques.

Trauma and Suggestibility

Several laboratory and field studies have examined children's memory of and suggestibility about events that are negative and even traumatic. Results from such studies have been mixed in terms of whether children are more suggestible about stressful than neutral events, or less so. Some theories of stress and memory suggest that trauma victims might repress traumatic events and in doing so have no memory for the traumatic experiences, only to have such memories reawakened later; other theorists contend that such memories have a high likelihood of being false. Some research indicates that trauma victims are especially attentive to trauma-related stimuli and remember their trauma experiences quite well, rather than repressing memory of them. When suggestibility has been studied in actual child abuse victims, their performance is generally comparable to that of nonabused control children, although some forms of psychopathology may be associated with greater error.

Conclusions

Overall, suggestive techniques are of most concern when they lead to false reports against innocent people. There is evidence that such techniques, in the extreme, can lead to false memories of entire events that never occurred. It is important to note that the inaccuracies resulting from suggestibility would likely be inaccurate details that are intermingled with an accurate memory report (e.g., mistaking the color of a car in an accident that actually occurred). However, exposure to multiply suggestive techniques over a long period of time could potentially lead vulnerable children (e.g., certain preschoolers) to create false memories of entire events that never occurred. Fortunately, children, even young ones, can be resistant to suggestive techniques, perhaps especially when they are reporting about personally significant negative experiences.

When child witnesses are involved in the legal system, their memory reports may be relied upon to prosecute those who commit crime. It is important to reduce, as much as possible, children's suggestibility errors, and false memory reports to ensure that innocent defendants are not wrongfully convicted. At the same time, children's suggestibility should not be overstated. Children often can resist false suggestions and maintain accuracy in the face of misinformation and the use of suggestive techniques. However, even then, the child's credibility as a witness may be damaged as a result of concerns about children's suggestibility.

Further Reading

Brainerd, C. & Reyna, V. (2005). *The Science of False Memory*, Oxford University Press, New York.

Bruck, M., Ceci, S.J. & Principe, G. (2006). The child and the law, in K.A. Renninger & I.E. Sigel (Vol. Eds.). W. Damon & R. Lerner (Gen. Eds.). *Handbook of Child Psychology*, 6th Edition, *Child Psychology in Practice*, Vol. 5, Wiley, New York.

Goodman, G.S. (2006). Children's eyewitness memory: a modern history and contemporary commentary, *Journal of Social Issues* **62**, 811–832.

Quas, J.A., Malloy, L.C., Melinder, A., Goodman, G.S., D'Mello, M. & Schaaf, J. (2007). Developmental differences in the effects of repeated interviews and interviewer bias on young children's event memory and false reports, *Developmental Psychology* **43**(4), 823–837.

Orbach, Y., Hershkowitz, I., Lamb, M.E., Sternberg, K.J., Esplin, P.W. & Horowitz, D. (2000). Assessing the value of

structured protocols for forensic interviews of alleged child abuse victims, *Child Abuse & Neglect* **24**, 733–752.
Paz-Alonso, P.M., Larson, R., Castelli, P., Alley, D. & Goodman, G.S. (2009). Memory development: emotion, stress, and trauma, in *The Development of Memory in Childhood*, 2nd Edition, M.L. Courage & N. Cowan, eds, Psychology Press, Hove, pp. 197–240.

STEPHANIE BLOCK, DAISY A. SEGOVIA AND GAIL S. GOODMAN

Civil Commitment

Introduction

Mental illness can have a profound effect on one's thought process. Delusions (*see* **Delusions**) and hallucinations (*see* **Hallucinations**) can create frightening situations from which a person may feel the only escape is to strike out. Depression and catatonia can produce such devastation that a person can no longer feed or clean himself. Such altered realities can also impair one's ability to make rational decisions. Thus, civil commitment as we currently know was born out of the need to involuntarily hospitalize mentally ill patients who are assessed to be a danger to themselves or others while maintaining respect for their civil rights.

In 1403, the world's oldest psychiatric hospital, the Bethlem Royal Hospital of London (a.k.a. Bedlam), first opened its doors. Over three centuries later, the first psychiatric asylum in the United States was established. The number of institutions throughout the industrialized world continued to expand until, in the 1950s, over 550 000 individuals in the United States, and 150 000 in England were maintained in public facilities due to their mental illness [1, 2]. Because of the advent of psychotropic medications, the promise of community mental health centers and funds supposedly allocated for this care, the number of institutionalized patients was declined to about 70 000 and 30 000, respectively, in the 1980s and 1990s [1, 2].

During this same period, an increasing awareness developed that all citizens, including those suffering from mental illness, should receive equal protection under the law. Previously, civil commitment laws were created to make the process as simple and convenient as possible [3]. On the basis of their status, patients subject to civil commitment may have lost other civil liberties, including the right to marry, manage their property, or carry a driver's license [4, 5]. Often, civil commitment merely required the presence of mental illness and a need for treatment. These criteria were fueled by the idea of *parens patriae*, which refers to a government's obligation to care for its citizens. Historically, the other primary reason for civil commitment is supported by a country's police powers. The government has an obligation to protect its citizens and doing so may require the detention of a mentally ill individual who, due to his illness, is perceived to present a danger to others.

Over the past several decades, we have seen a shift in the role of civil commitment. This may again be altered on the basis of the public's perception of the relationship between mental illness and violence, especially following national tragedies such as the Virginia Tech massacre, in which a mentally ill individual killed 32 people and wounded 25 before killing himself [6].

Inpatient Civil Commitment

The Criteria

Though countries, and even individual states, have established their own unique procedures for civil commitment, there are several themes that are common to most. The first is the requirement of mental illness. There does exist, however, some discrepancy in how mental illness is defined. Some countries or states specifically exclude diagnoses such as substance use disorders, dementia, mental retardation, or personality disorders [3]. In the 1992 US Supreme Court case *Foucha v. Louisiana*, the court decided that it was unconstitutional to hold a not guilty by reason of insanity (NGRI) acquittee (*see* **Insanity: Defense**; **Interrogation**) in the hospital based on his risk of violence secondary only to his antisocial traits [7]. However, England and Scotland do permit the detention of patients whose dangerousness stems from the existence of a personality disorder [8, 9].

When patients are civilly committed, they are placed in a hospital setting. Without the presence of mental illness (which presumably can be ameliorated with treatment), hospitals would simply serve as an

alternative to incarceration. In 1975, the US Supreme Court, in *O'Connor v. Donaldson,* held that "a State cannot constitutionally confine *without more* a nondangerous individual who is capable of surviving safely in freedom by himself or with the help of willing and responsible family members or friends." [10] (emphasis added) In this decision, "more" is commonly construed as a reference to treatment. In the United Kingdom, based on the 1985 court case *Regina v. Hallstrom* [11], the focus of civil commitment rests on the intention to treat patients' mental illness [9].

In addition to the presence of mental illness, there are often three main potential criteria for civil commitment. The first is the risk of danger to self. This may be manifested in form of a suicide attempt, an overt suicidal threat, or behavior that indicates that one intends to harm oneself (e.g., writing a suicide note or stocking up on medication known to be lethal in overdose), which may occur despite a patient's denial of suicidal thoughts. The second criterion is the risk of danger to others, similar to the first criterion in that both actions and thoughts allow for a patient to be civilly committed. The final criterion for civil commitment often refers to a patient's gravely disabled condition or inability to care for self. This alludes to the patient's inability to provide for his own basic needs, such as food, shelter, clothing, medical care, and personal security [12] secondary to mental illness. On the basis of this assessment and prediction, the clinician is given both the power and responsibility of placing a person, at least temporarily, in an inpatient facility potentially against his will.

The Process

Once a mental health care provider has made the decision to civilly commit a patient, there exists a standardized procedure to which the clinician must adhere. Though the criteria are often the same, the logistics of civil commitment vary between countries and states.

The United States. In the United States, there exists a delicate balance between psychiatry and the law during the commitment process. All states allow for emergent commitment of patients meeting the aforementioned criteria with little to no involvement of the court system. Therefore, commitment is primarily based on the assessment of the clinician and occurs swiftly. At this time, the patient is provided with the potential duration of the confinement and the point in time when access to counsel and a hearing is granted [3]. This access must occur following an often brief and well-circumscribed period, which can range from 48 h [13] to 15 days (with the requirement of a second opinion by another psychiatrist) [14].

The second step in this process involves long-term detention and does require judicial approval. Patients who require only brief stabilization may be discharged before their cases reach court. On the basis of the holding in the Supreme Court case, *Addington v. Texas* [15], the decision to civilly commit must be founded on clear and convincing evidence, which is an intermediate level of proof between preponderance of the evidence (civil) and beyond a reasonable doubt (criminal). The landmark high water mark for the rights of civil committees is based on the 1972 court case *Lessard v. Schmidt*; the patient is eligible for all the same constitutionally protected due process rights afforded to criminal defendants [16]. Among different states, the initial commitment often extends from 90 to 180 days and may be renewed if necessary after the expiration of that period [3].

The United Kingdom. Civil commitment in the United Kingdom can be enacted to ensure both patients' health and the safety of those around them and is based on an intention to treat patients' mental illness [11]. In England and Wales, a patient can be involuntarily hospitalized, assessed, and treated for a period up to 28 days. If appropriate, after this evaluation period, a patient may be detained for a (renewable) period up to six months based on the consensus of two medical opinions. In Scotland, detention beyond the 28 days evaluation period must be sanctioned by a Mental Health Review Tribunal, which consists of a lawyer, a psychiatrist, and a third person with experience in handling mental health issues. In Northern Ireland, the initial evaluation period is only seven days, and this may be extended for another seven days if necessary.

Other Nations. In 1997, Applebaum [17] authored a review of the state of civil commitment worldwide. In the decade following the change from treatment to dangerousness-based commitment in the US, several other nations followed suit (including Austria,

Belgium, Germany, Israel, the Netherlands, Northern Ireland, Russia, Taiwan, and Ontario, Canada). However, many nations, including Sweden, Finland, France, Greece, and Italy, have opted for laws more similar to the British model that is based on the need for treatment. Applebaum noted that there is data to support the idea that, regardless of the laws, both courts and doctors will follow their moral compasses and civilly commit those patients that need psychiatric intervention. This appears to be an international phenomenon.

The Dual Role Dilemma. Mental health professionals are frequently called upon to serve as both the patients' treatment providers and their forensic evaluators in regard to their eligibility for civil commitment. Potential conflicts that may arise include a disruption in the therapeutic alliance and inappropriate access to patients' information. In other words, patients may not have chosen to reveal certain information to someone acting solely as a forensic evaluator. Therefore, patients may feel that their own psychiatrists testified "against" them [18].

Outpatient Civil Commitment

After deinstitutionalization created a push for increased psychiatric care in the community, civil commitment was no longer restricted to the inpatient setting. In the United States, in 1966, the court of appeals for the District of Columbia heard the case of *Lake v. Cameron* and decided that a nondangerous, civilly committed person should not have to be confined to an inpatient facility if less restrictive alternatives were available [19]. In 1999, the state of New York again brought attention to outpatient civil commitment when it enacted Kendra's Law, named after a woman who was pushed onto the subway tracks by a schizophrenic man [20]. To be eligible for outpatient civil commitment according to this law, an adult must be mentally ill and unlikely to survive safely in the community without supervision but unlikely to comply with treatment and have a history of treatment noncompliance and violent acts [20]. Many states have adopted similar laws, but enforcement occurs only in the counties opting to do so (California, Ohio, etc.).

Interventions for those people who are committed on an outpatient basis include options such as monitored medication use, required participation in individual or group therapy, mandatory presentation to day hospital programs, or residence in certain supervised living environments [21]. The United Kingdom also emphasizes the use of the least restrictive alternative, and jurisdictions have rules and resources in place to allow for outpatient civil commitment [9].

Other Factors Affecting Civil Commitment

Incompetence

Competence refers to one's capacity to function in a particular domain at a particular time. There are many specific types of competence with various standards. On the basis of evidence often collected by mental health professionals, the ultimate decision on competence is usually made by a judicial body (*see* **Capacity Assessment**).

Most cases of civil commitment involve the hospitalization of a patient who does not want such an intervention. Because the safety of the patient or others may be in serious jeopardy, it might be immaterial if the person is felt to be capable of making the decision in this case. Competency does become an issue, however, if the person meets commitment criteria and requests a voluntary admission but appears incapable of making that decision. This scenario renders civil commitment as the only viable option. On the basis of the 1990 US Supreme Court case *Zinermon v. Burch*, if a patient is suspected or known to be incompetent regarding his ability to sign himself in to the hospital, he cannot voluntarily do so. The Court noted that allowing an incompetent individual to voluntarily hospitalize himself when he is unable to consent to admission is a deprivation of his civil liberties [22].

In the case of incompetent patients, the presence of a guardian alters the procedure for hospitalization. In general, a guardian is a court-appointed individual who maintains control over other's property or person. Some guardians, referred to as *conservators* or *guardians* of the estate, only make decisions related to their wards' estates. For those persons deemed by the court to be globally incompetent, general guardians (or guardians of the person) are responsible for all of their decision making. This, of course, includes the decision to sign into a hospital or accept and refuse treatment.

Guardians may make decisions for their wards based on two different models. In the *best interest* standard, guardians base their decisions on what they judge would be in the best interest of their wards. Alternatively, in the "substituted judgment"

standard, guardians base their decisions on how their wards might have acted in the same situation if they were competent. With the second model, the guardians' task is simplified greatly if their wards (while competent) draft advanced directives outlining their wishes if the occasion arises in which they become incompetent [23]. However, it is important to recognize that these advance directives (and even the decision of the guardian) can be overridden by the civil commitment laws, which serve to protect both the patient and the community [24].

Forced Medications

As noted above, the capacity to determine the need for hospitalization and the capacity to make decisions about medical treatment are often two distinct types of competency. In 1980, the Utah legislature (United States) codified that, in order to civilly commit a patient, hospitalization must be the least restrictive alternative and the patient is incompetent to consent to treatment [25], and therefore, a civilly committed patient has no right to refuse medication [26].

Special Cases Related to Civil Commitment

Minors

Merely on the basis of their age, minors are generally considered incompetent to make legal decisions, whether or not they meet the criteria of a given competency. Parents (or the state) are recognized as their legal guardians and are therefore often responsible for their decision making (*see* **Parental Rights and Prerogatives**). On the basis of a 1967 case (*In re Gault*), the US Supreme Court stated that due process is required for juveniles [27]. Twelve years later, *Parham v. J.L. and J.R.* came before the court. This was a class action suit seeking to place civilly committed children in less restrictive environments. However, the Supreme Court declined to grant children the same procedural due process rights granted to adults who faced the possibility of civil commitment [28]. This decision served as an acknowledgment of parental authority over the civil liberties of minors.

Insanity Acquittees

Much as it did for minors, the US Supreme Court made a special exception for those found NGRI. In 1983, they heard the case of *Jones v. U.S.*, in which a man presented a successful insanity defense for shoplifting, was hospitalized and then requested to be released after he was hospitalized for the length of time equivalent to what his sentence would have been if he had been convicted. The Court rejected his argument, deciding instead that hospitalization is not a sentence but includes treatment, the NGRI acquittee may be confined indefinitely, and the burden of proof to prove lack of dangerousness rests on the acquittee [29].

Inmates

Over the past several decades, the US Supreme Court has heard major cases regarding the rights of mentally ill prisoners in regards to civil commitment. The outcome of *Baxstrom v. Herald*, decided in 1966, stated that the equal protection clause of the 14th amendment was violated if an inmate was civilly committed following a period of incarceration without review [30]. This decision lead to the movement of 969 "criminally insane" prisoners from the custody of the New York Department of Corrections to state civil hospitals; this was informally known as *Operation Baxstrom* [31]. Interestingly, only 26 of those people returned to forensic psychiatric hospitals over the next four years [32]. The second of these cases was *Vitek v. Jones* [33]. This decision stated that the rights of due process are applicable when inmates are transferred from the facility in which they are incarcerated to a psychiatric hospital for treatment. More specifically, they have the right to notice the transfer, an adversary hearing, and the availability of legal counsel [33].

Canada enacted a postconviction commitment law in 1997. In lieu of sentencing offenders that the court considers dangerous, this law allows for their indefinite confinement in federal or state institutions. This law applies to various crimes that can cause significant harm to the victim; however, 90% of those committed under this law are sex offenders [34].

Sexual Offenders

Sex offenders pose a particular challenge in regard to civil commitment, given that they are typically difficult to treat and may pose a great risk to others if they reoffend. This issue was addressed by the US Washington State Supreme Court in 1993. According to *In re Young and Cunningham*, the court decided

that the state law allowing the civil commitment of sexually violent predators with mental abnormalities or personality disorders was constitutional in some cases. Since the commitment was civil (not criminal), it did not violate a defendant's double jeopardy or ex post facto rights [35]. This issue made its way to the US Supreme Court in 1997 in *Kansas v. Hendricks*. The outcome was similar, when a 5-4 Court decided that the state of Kansas' civil commitment of sexual offenders did not violate substantive due process and the double jeopardy/ex post facto clauses. Furthermore, they noted that failure to offer treatment does not indicate that the law is merely punitive [36]. Some argue that this decision overextended the reach of civil commitment laws [37, 38] and places psychiatrists in the difficult situation of using psychiatric services for social control. One author went on to state that other selected classes of people, notably substance abusers, may be in jeopardy of serious consequences from this decision [38].

Abuses of Civil Commitment

In the nineteenth century, the case of Mrs E.W.P. Packard came to light when she published a text chronicling her difficulties. In 1839, she married a Presbyterian minister with whom she had six children. After she began to explore other religions, her husband had admitted her to an Illinois State Hospital for the Insane in 1860, where she was diagnosed with "monomania" and "moral insanity". Though she was released in 1863 and ultimately declared to be sane, she discovered she had no legal rights for her children or her property. Because of her experience, she dedicated her life to lobbying for asylum reform and women's rights [39].

This obvious violation of civil rights occurred over a century ago, but abuses of the power of civil commitment have occurred far more recently. When Khrushchev rose to power in Communist Russia, he moved political dissidents from labor camps to psychiatric hospitals in effort to make these revolutionaries appear better treated with the added benefit of making them seem "crazy" and unworthy of support [40, 41]. Prior to the fall of this regime, Western psychiatrists had the opportunity to directly examine the state of psychiatry in the Soviet Union and confirmed the presence of such activities [42]. In China, those who openly challenge the government are labeled as mentally ill and placed in special forensic hospitals known as *Ankang*. Only recently, China has become more open in allowing outsiders to investigate their forensic facilities [43].

Civil commitment exists as a delicate balance between psychiatry and the law, weighing a need to protect the patient and his community with his rights as a citizen. Commonly, patients are eligible for civil commitment if: (i) they are mentally ill and (ii) they are at risk of harming themselves or others. Each jurisdiction may go about civil commitment in a slightly different way, and a variety of laws have emerged. These laws are modified in special cases, such as those involving minors, insanity acquittees, inmates, and sex offenders.

References

[1] Coleman, N.M. & Gilbert, L. (1979). *Stalking the Least Restrictive Alternative: Litigative and Non-Litigative Strategies for the Indigent Mentally Disabled*. [S.l. : s.n.], c1979. KF337.5.M46 C64.

[2] Porter, R. (2002). *Madness: A Brief History*, Oxford University Press, Oxford.

[3] Melton, G.B., Petrila, J., Poythress, N.G. & Slobogin, C. (eds) (1997). Civil commitment, *Psychological Evaluations for the Courts: A Handbook for Mental Health Professionals and Lawyers*, 2nd Edition, The Guilford Press, New York, pp. 297–336.

[4] Sales, B.D., Powell, D.M. & Van Duizend, R. (1982). *Disabled Persons and the Law: State Legislative Issues*, Plenum Press, New York.

[5] Ennis, B. & Siegel, L. (1973). *The Rights of Mental Patients: The Basic ACLU Guide to Mental Patient's Rights*, Avon, New York.

[6] Shuman, D.W. (2007). School shootings: a word of caution, *Psychiatric Times* **24**(7) Available from http://www.psychiatrictimes.com/display/article/10168/54656 (last accessed 10/8/08).

[7] Foucha v. Louisiana, 112 United States 1780 (1992).

[8] Felthous, A.R. & Sass, H. (2000). Introduction to this issue: international perspectives on psychopathic disorders, *Behavioral Sciences and the Law* **18**(5), 557–565.

[9] Fennell, P. & Goldstein, R.L. (2006). The application of civil commitment law and practices to a case of delusional disorder: a cross-national comparison of legal approaches in the United States and the United Kingdom, *Behavioral Sciences and the Law* **24**(3), 385–406.

[10] O'Connor v. Donaldson, 422 U.S. 563 (1975).

[11] Regina v. Hallstrom, *ex parte* W, R. v. Gardner, *ex parte* L. (1985). 1 2 All.E.R. 306.

[12] Beis, E.B. (1984). *State Involuntary Commitment Statutes. Mental Health and the Law*. Aspen Publications, Rockville.

[13] Va. Code Ann. § 37.1–67.1.

[14] N.Y. Mental Hygiene Law § 9.39.

[15] Addington v. Texas, 441 U.S. 418 (1979).
[16] Lessard v. Schmidt, 349 F. Supp, 1078 (1972).
[17] Applebaum, P.S. (1997). International perspective on involuntary commitment, *Journal of the American Academy of Psychiatry and the Law* **25**(2), 135–147.
[18] West, S.G. & Friedman, S.H. (2007). To be or not to be: psychiatric clinician and forensic evaluator, *Psychiatric Times* **24**(6), 50–51.
[19] Lake v. Cameron, 364 F.2d 657 (1966).
[20] New York State, Office of Mental Health [online], (2007). Available from http://www.omh.state.ny.us/omhweb/Kendra_web/Ksummary.htm (last accessed 7/17/07).
[21] McCafferty, G. & Dooley, J. (1990). Involuntary outpatient commitment: an update, *Mental and Physical Disability Law Reporter* **14**, 277–287.
[22] Zinermon v. Burch, 494 U.S. 113, 110 S. Ct. 975 (1990).
[23] Melton, G.B., Petrila, J., Poythress, N.G. & Slobogin, C. (eds) (1997). Civil competencies, *Psychological Evaluations for the Courts: A Handbook for Mental Health Professionals and Lawyers*, 2nd Edition, The Guilford Press, New York, pp. 297–336.
[24] Swanson, J.W., McCrary, S.V., Swartz, M.S., Elbogen, E.B. & Van Dorn, R.A. (2006). Superseding psychiatric advance directives: ethical and legal considerations, *Journal of the American Academy of Psychiatry and the Law* **34**(3), 385–394.
[25] Appelbaum, P.S. (1981). A.E. & R.R.: Utah's compromise on the right to refuse treatment, *Hospital & Community Psychiatry* **32**(3), 167–168.
[26] A.E. & R.R. v. Mitchell, No. C78-466 (D. Utah, June 16, 1980).
[27] In re Gault, 87 U.S. 1428 (1967).
[28] Parharm v. J.R. and J.L., 442 U.S. 584 (1979).
[29] Jones v. U.S., 463 U.S. 354, 103 S. Ct. 3043 (1983).
[30] Baxstrom v. Harold, 383 U.S. 107, 86 S. Ct. 760 (1966).
[31] Hunt, R.C. & Wiley, D. (1968). Operation Baxstrom after one year, *American Journal of Psychiatry* **124**, 974–978.
[32] Steadman, H.J. (1973). Follow-up on Baxstrom patients returned to hospitals for the criminally insane, *American Journal of Psychiatry* **130**, 317–319.
[33] Vitek v. Jones, 445 U.S. 480, 100 S. Ct. 1254 (1980).
[34] Matson, S. & Lieb, R. (1997). *Sexual Predator Commitment Laws*, Document No. 97-10-1102, Institute for Public Policy, Washington, DC, pp. 1–15.
[35] In re Young v Cunningham, 857 P. 2d 989 (1993).
[36] Kansas v. Hendricks, 521 U.S. 346, 117 S. Ct. 2072 (1997).
[37] La Fond, J.Q. (2000). The future of involuntary civil commitment in the U.S.A. after Kansas v. Hendricks, *Behavioral Sciences and the Law* **18**(2–3), 153–167.
[38] Krongard, M.L. (2002). A population at risk: civil commitment of substance abusers after Kansas v. Hendricks, *California Law Review* **90**(1), 111–163.
[39] Carlisle, L.V. (2005). Elizabeth Packard and boundaries of gender, religion, and sanity in nineteenth-century, *America Dissertation Abstracts International Section A: Humanities and Social Sciences* **65**(9-A), 3533.
[40] Bloch, S. & Reddaway, P. (1971). *Psychiatric Terror: How Soviet Psychiatry is Used to Suppress Dissent*, Basic Books, New York.
[41] Applebaum, P.S. (1981). Law and psychiatry, Soviet style, *Hospital and Community Psychiatry* **32**, 601–602.
[42] Bonnie, R. (1990). Soviet psychiatry and human rights: reflections on the report of the United States delegation, *Law, Medicine and Health* **18**, 123–131.
[43] Munro, R. (2005). *A Question of Criminal Madness: Judicial Psychiatry and Political Dissent in the People's Republic of China*, PhD Thesis, University of London (SOAS Law Department).

Related Articles

Therapeutic Jurisprudence

SARA G. WEST AND SUSAN HATTERS-FRIEDMAN

Civil Discovery and Disclosure Rules in the United States *see* Discovery in the United States: Civil Cases

Civil Law Systems of Justice

The legal systems that exist in countries outside the English-speaking world are frequently referred to as *civil law systems of jurisprudence*. They are also referred to as *inquisitorial systems* or *investigative systems*. Civil law systems are based alternatively on the French Napoleonic Code, or on the Roman law from which the Napoleonic Code was in turn derived. Civil law systems stand in contrast to the adversary or common law systems, which developed predominantly in the English-speaking world.

Civil law systems are based on a search for truth that requires a judicial official, called an *examining magistrate* (or *juge d'instruction*) to investigate whether a crime has been committed. At the conclusion of this investigation, which may in some

celebrated or notorious cases be years in duration, a file or *dossier* will have been compiled. This *dossier* will contain all of the relevant statements and evidence that have been collected and examined. The *dossier* will also include the magistrate's conclusions about the strength of the evidence and recommendations for further action. When the investigation results in finding sufficient cause to establish that a crime has occurred, and that a specific person has committed that crime, the magistrate's file will recommend that the individual be placed on trial.

Related Articles

Adversary Systems of Justice

Expert Opinion in Court: a Comparison of Approaches

ANDRE MOENSSENS

Civil Rights Violations: Mental Health Support *see* Disaster Mental Health

CNBR *see* Chemical, Biological, Radiological, and Nuclear Investigations

Cocaine

History

Cocaine is a psychotropic drug prepared from coca that is found in the plant, *Erythroxylon coca*, which grows principally in the northern South American Andes and to a lesser extent in India, Africa, and Java. The plant, which may reach 3 m, favors higher elevations (up to 1800 m) because at lower elevations (below 500 m), the alkaloid content is significantly diminished due to more rapid growth. Coca leaves may first be harvested about 2 years after the time of planting. Then, depending on the altitude, the leaves may be harvested up to three times a year. The leaves are dried and converted into a coca paste, which is eventually used to produce cocaine hydrochloride (cocaine HCl). The yields from 100 kg of coca leaves are about 1 kg and 800 g, respectively, for coca paste and cocaine HCl.

In the mid-nineteenth century Carl Wöhler, the chemist who synthesized urea, had coca leaves imported to Germany and presented them to his graduate student, Albert Niemann, to analyze. Niemann was the first to successfully isolate cocaine from the coca plant. From the 1860s until the turn of the century, cocaine appeared in various elixirs and tonics purported to have "magic" properties. One of the more famous preparations was Vin Mariani, a mixture of wine and cocaine that received celebrity endorsements from Pope Leo XIII, US President McKinley, Thomas Edison, Auguste Rodin, and Jules Verne. The original Coca-Cola® beverage contained a coca-based syrup supplemented with caffeine and was marketed as a tonic and headache remedy. In 1884, the year Sigmund Freud popularized the drug in his famous treatise "Über Coca", Carl Koller became the first physician to use cocaine as a topical anesthetic in ophthalmological surgery. As cocaine abuse began to be viewed as a problem, its sales became restricted. Legitimate medical use is almost exclusively limited to topical administration as a local anesthetic in ear, nose, and throat surgery (as the HCl salt in 10–20% solutions) and in ophthalmological procedures (as a 1–4% solution). Cocaine is one of the most common illicit drugs of abuse. It is estimated that over 34 million Americans aged 12 or over have tried the drug, with about one million current users [1].

Cocaine Abuse

Cocaine is sold on the street in the form of the hydrochloride salt and crack. The salt form varies considerably in purity and is typically diluted with a variety of cutting agents. The cocaine powder supplied by dealers is often clumpy and first needs to be chopped. This is usually achieved using a mirror with a razor blade, after which the cocaine is arranged into thin lines about 30–60 mm long and 2 mm wide resulting in an average dose of 25 mg and then

snorted through a straw or "tooter". Alternatively, cocaine may be snorted from a "coke spoon" or "bullet", a device in which a vial containing cocaine may be inverted over a closed chamber into which the cocaine falls and may then be rotated for convenient snorting. A single long fingernail may serve as a natural "coke spoon".

Crack is a free base form of cocaine that produces a characteristic crackling sound when smoked. It is prepared by adding baking soda to aqueous cocaine HCl and heating it to remove the water. After heating, the mixture is cooled and filtered, and the free base cocaine precipitates into small pellets or "rocks." These "rocks" can then be smoked in a crack pipe. Crack pipes range from elaborate glass pipes to a soda can with a hole. A typical "rock" weighs about 20 mg and costs US$10–20 (2008), making this form of cocaine much less expensive, per unit, than the salt form that is typically sold for around US$100 per gram.

Pharmacology and Pharmacodynamics

Mechanism of Action and Effects

Cocaine (methylbenzoylecgonine) is an ester of benzoic acid and the amino alcohol, methylecgonine, and is chemically, but not pharmacologically, related to atropine. Cocaine is used medicinally as a topical local anesthetic but is structurally different from other local anesthetics by virtue of its tropine moiety (Figure 1). However, as other local anesthetics, cocaine consists of a hydrophobic and hydrophilic regions and contains ester linkages that allow the body to hydrolyze and deactivate the drug.

As a local anesthetic, cocaine's most important mechanism of action lies in its ability to block sodium channel conductance and thereby to increase the threshold required to generate an action potential. Cocaine has additional actions that make it unique from other local anesthetics, including its ability to block reuptake of the neurotransmitters norepinephrine, dopamine (DA), and serotonin. Norepinephrine (noradrenaline) is responsible for the classic adrenergic effects observed with cocaine use including mydriasis, vasoconstriction, hypertension, tachycardia, and tachypnea. However, the euphoric effects of the drug are mediated by DA, the neurotransmitter that stimulates the reward centers in the brain. After an acute dose of cocaine, DA concentrations in the brain become elevated, as cocaine prevents the reuptake of DA into the presynaptic neuron. This results in increased extracellular DA concentrations resulting in chronic stimulation of the DA receptor in the postsynaptic neuron. During this period, the user experiences the desirable behavioral effects of cocaine including intense euphoria, psychic energy, heightened sexual excitement, and elevation of mood. When DA is depleted, the central stimulatory effects (rush) are followed by depression (crash). Therefore, chronic cocaine users repeatedly administer cocaine (binge) to maintain increased synaptic levels of DA. The positive reinforcement of the rush *versus* the negative reinforcement of the crash is the principal reason for the development of cocaine abuse.

The time course for the high in humans parallels that of the cocaine concentration in the reward centers of the brain. For equivalent plasma cocaine concentrations, smoked cocaine induced significantly greater self-reports of "high" than intranasal cocaine and showed a trend for greater effect than intravenous cocaine [2]. This study demonstrates the importance of speed of cocaine delivery into the brain on its reinforcing effects and how crack cocaine can be even more dangerous than the salt.

Toxicity

There appears to be a good correlation between peak plasma cocaine concentrations and peak pharmacologic and behavioral effects [3]. Because the "high" is related to plasma concentration, not the dose, it is easy to understand how self-administration of toxic amounts of cocaine occurs during a binge.

The most common clinical manifestations following acute cocaine intoxication include profound central nervous system (CNS) stimulation with psychosis and repeated grand-mal convulsions, ventricular tachyarrhythmias and respiratory dysfunction with Cheyne-Stokes breathing, and ultimately respiratory paralysis. Other symptoms include mydriasis, hypertension leading to hypotension, and small-muscle twitching. Cocaine's ability to increase muscular activity and vasoconstriction may produce extreme hyperthermia.

A significant number of cases presented to emergency departments have been associated with a cocaine-induced psychosis, now commonly called

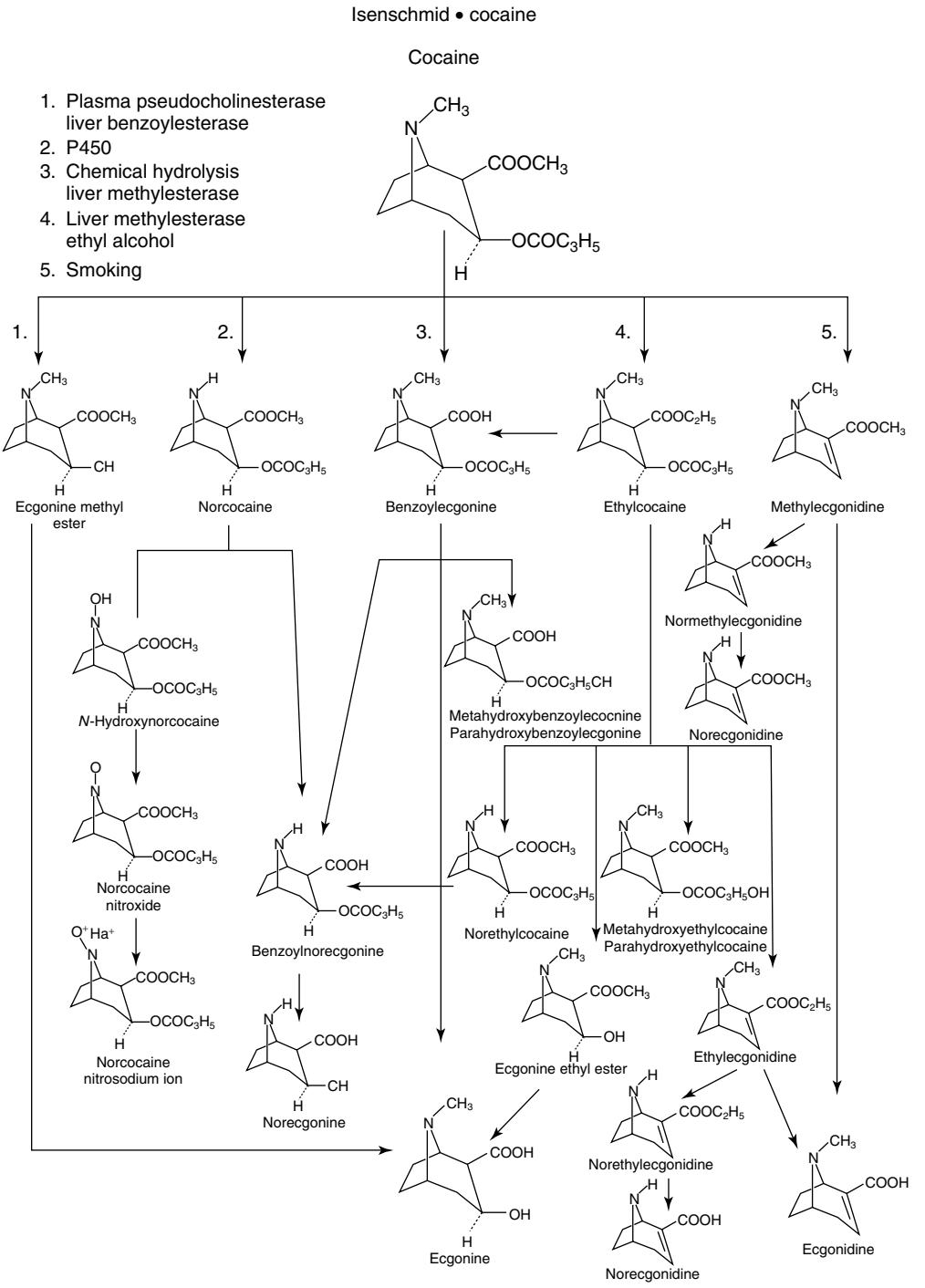


Figure 1 Pathways in the metabolism of cocaine. Cocaine, in *Principles of Forensic Toxicology*, B. Levine, ed, AACC Press, Washington DC, p. 221) [Reproduced with permission from AACC Press. © 1999.]

cocaine-induced excited or *agitated delirium*. This syndrome is characterized by severe hyperthermia (104–108 °F), extreme agitation and delirium, respiratory arrest, and sudden death. These individuals exhibit bizarre and violent behavior and extreme strength and are frequently seen running around – often naked – shouting, fighting, breaking things, and causing injury to themselves and/or others. Law enforcement is frequently called to respond to this type of situation and attempt to restrain the individual. In some cases, stress from restraint may result in catecholamine surges on an already sensitized myocardium, resulting in arrhythmias and possible death.

Cocaine can injure cerebral arteries and an acute hypertensive episode following a single dose in a chronic user can cause the vessels to rupture. Even though acute tolerance to the cardiovascular effects of cocaine has been shown to occur, during a cocaine "binge" the chronic cocaine user may self-administer more cocaine than the cardiovascular system can tolerate.

Pharmacokinetics

Absorption. Cocaine may be self-administered intranasally, by smoking, intravenously, and orally. Cocaine is usually not administered orally because low bioavailability results in reduced euphoric effects due to inefficient delivery to the brain. The intravenous route of administration, sometimes called *mainlining*, is the only route that consistently produces 100% drug bioavailability. Despite varied bioavailability of cocaine when administered intranasally or smoked, the convenience of these two routes of administration, and the latter's rapid, intense onset of effects, makes them the most commonly used.

Metabolism. Cocaine is metabolized primarily to benzoylecgonine and ecgonine methyl ester by different mechanisms (Figure 1). The mechanisms of cocaine metabolism are not straightforward due to complex *in vitro* and *in vivo* reactions.

Stability studies support the long-held findings that cocaine is metabolized to ecgonine methyl ester via enzymatic hydrolysis by pseudocholinesterase and to benzoylecgonine via spontaneous hydrolysis at physiological and alkaline pH [4]. However, cocaine can also be metabolized by liver esterases. Liver methylesterase catalyzes the conversion of cocaine to benzoylecgonine and the transesterification of cocaine to ethylcocaine (also known as *cocaethylene*), if cocaine is used with ethanol [5]. In the absence of ethanol, this enzyme hydrolyzes cocaine exclusively to benzoylecgonine. In the presence of both cocaine and ethanol, transesterification of cocaine to ethylcocaine occurs about 3.5 times faster than hydrolysis to benzoylecgonine [6]. As ethylcocaine is an active drug, it can contribute to cocaine toxicity. A separate and distinct human liver esterase, benzoylesterase, was found to catalyze the conversion of cocaine to ecgonine methyl ester but not ethylcocaine [5]. Both benzoylecgonine and ecgonine methyl ester are further metabolized to ecgonine.

Minor metabolites of cocaine include norcocaine, an *N*-demethyl metabolite of cocaine produced by liver cytochrome P-450 that received considerable study because of its conversion into a hepatotoxic metabolite [7]. In humans, however, reports of hepatotoxicity attributed to cocaine use are rare. Methylecgonidine (anhydroecgonine methyl ester) has been identified as a unique cocaine metabolite in postmortem blood and urine after smoking cocaine [8]. Related metabolites, ecgonidine, and norecgonidine methyl ester have been identified in urine specimens [9]. Other metabolites of simultaneous cocaine and ethanol users include ethylecgonidine and ecgonine ethyl ester. Hydroxylated metabolites of cocaine are found primarily in the urine [10].

Plasma Concentrations. Many pharmacokinetic studies have been performed with cocaine. Most studies performed were limited to single-dose studies with measurement of only the parent drug. Although considerable interindividual variation between subjects has been reported, several observations could be made. When bioequivalent doses of cocaine were administered by the intravenous and smoked routes, similar absorption and elimination curves for cocaine were obtained [3, 11]. When two different doses were given by the intravenous, smoked, or oral routes, intrasubject data indicated that mean peak plasma cocaine concentrations were dose related [11, 12]. When cocaine was administered intranasally, the dose and mean peak plasma concentration showed a poor correlation due to dose-dependent bioavailability by that route [13] with delayed peak plasma concentrations [3]. Similar delays in peak cocaine concentrations were observed oral cocaine administration [12].

On the basis of a review of the literature, typical peak plasma cocaine concentrations in most single-dose pharmacokinetic studies by the smoked (up to 100 mg), intravenous (up to 64 mg), and intranasal (up to 100 mg) routes averaged between 0.2 and 0.4 mg l^{-1} [14]. Of course, when cocaine is abused, it is rarely administered as a single dose. In the limited controlled studies that have examined the pharmacokinetics of cocaine following multiple doses, cocaine concentrations of up to 1–2 mg l^{-1} have been measured after smoked, intravenous, and oral administration without adverse effects [14].

After intravenous and smoked administration, cocaine follows first-order elimination using both one- and two-compartment models. On the basis of the literature, the average half-life for cocaine, for both routes of administration, was about 60 min and was usually not dose dependent [14]. After intranasal administration, cocaine pharmacokinetics have generally been described by one- or two-compartment models with first-order input. On the basis of the published data, the average absorption half-life for cocaine was 12 min and the average elimination half-life 84 min [14]. Dose-dependent kinetics has been reported in a few studies, which follows high intravenous and intranasal doses [13, 15].

Benzoylecgonine appears in the plasma within 15–30 min following cocaine administration by the intravenous, smoked, and intranasal routes of administration [3, 16, 17] On the basis of a single-dose pharmacokinetic study with sampling times sufficiently extended to study the pharmacokinetics of benzoylecgonine in plasma, the average formation half-lives for intravenous, smoked, and intranasal routes of administration were 34, 29, and 112 min, respectively [3]. Peak plasma benzoylecgonine concentrations usually occurred within 90 min after smoked and intravenous cocaine administration and were about half the peak cocaine concentration. After intranasal cocaine administration, peak benzoylecgonine concentrations were not reached until 3 h and were about twice that of cocaine and remained elevated for the next 5 h. This was consistent with the delayed absorption after intranasal administration during which time significant metabolism of cocaine may occur. In all routes of administration, the rate of benzoylecgonine elimination was slow compared to its rate of formation, accounting for its accumulation in plasma while cocaine concentrations were decreasing. The elimination half-lives for benzoylecgonine were 347, 324, and 213 min, respectively, after intravenous, smoked, and intranasal routes of administration.

The appearance of ecgonine methyl ester in blood after the administration of cocaine is not well understood. In some studies, only trace amounts of ecgonine methyl ester were detected in the blood [3, 16]. However, in other studies, ecgonine methyl ester could be detected in the blood at significant concentrations, although generally at much lower concentrations than benzoylecgonine [12, 18–20]. Factors that may play a role in the appearance of ecgonine methyl ester in blood include cocaine stability, route of administration, chronicity of use, and enzymatic reactions not yet identified. Some of these factors have been reviewed [21], but further research is still needed.

Excretion. Cocaine and its metabolites are excreted into the urine almost exclusively by simple filtration. In a study where urine specimens were collected for 3 days after six subjects were administered a single dose of cocaine by intravenous (25 mg), intranasal (32 mg), and smoked (42 mg) routes of administration, the mean percentage of the total dose recovered over 3 days was 45, 57, and 25%, respectively [10]. Cocaine concentrations declined to below detectable limits within 24 h (detection limit = 0.001 mg l^{-1}). Benzoylecgonine and ecgonine methyl ester were frequently detected at low concentrations (<0.025 mg l^{-1}) through 72 h.

Interpretation of Results

Blood

To attempt interpretation of cocaine and metabolite concentrations in blood, a large number of factors must be considered. A search of the literature suggests that cocaine concentrations of <0.30 mg l^{-1} are generally considered clinically therapeutic [14]. However, therapeutic, toxic, and lethal cocaine concentrations clearly overlap as observed by clinical and postmortem studies. Tolerance or reverse tolerance can play a significant role in the poor correlation observed. In a study of 130 patients presenting to an emergency room with acute cocaine toxicity, the mean plasma cocaine concentration was 0.34 mg l^{-1} and ranged from undetectable to

3.92 mg l^{-1}. The median cocaine concentration in these patients was only 0.07 mg l^{-1}. However, the mean and median benzoylecgonine concentrations in these patients were 1.57 and 1.06 mg l^{-1}, respectively. There was no correlation between the cocaine and metabolite concentrations in these patients and their clinical state or their outcome. The degree of symptoms of toxicity – most notably hyperthermia, heart rate, and psychosis – did provide a better predictor of a patients' outcome [22].

Other complicating issues include underlying pathology and the instability of cocaine in blood. In unpreserved blood, studies have shown that cocaine is rapidly hydrolyzed by plasma butyrylcholinesterase almost exclusively to ecgonine methyl ester [23]. The addition of sufficient sodium fluoride, while inhibiting enzymatic hydrolysis of cocaine to ecgonine methyl ester, does not prevent spontaneous hydrolysis of cocaine to benzoylecgonine. The rate of hydrolysis of both esters has been shown to be temperature and pH dependent, with higher temperatures and pH increasing the rate of hydrolysis.

In postmortem cases, additional complications are the competing factors of postmortem redistribution and release with cocaine hydrolysis [24]. Studies conducted with blood collected from the same cadavers at different times did not show predictable changes in cocaine concentrations. This leads to the conclusion that the usefulness of postmortem cocaine and metabolite concentrations after death may be limited especially since therapeutic, toxic, and fatal cocaine concentrations overlap. Studies conducted by other investigators have found similar results [25].

Urine

Urine is the most commonly utilized specimen for detecting cocaine use and can easily be analyzed for the presence of cocaine and metabolites, typically benzoylecgonine. There are few interpretive concerns for positive results, once alternative medical explanations are ruled out, since drug concentrations cannot be correlated to impairment. Generally, benzoylecgonine may be detected for 2–4 days using a 300 ng ml^{-1} cutoff concentration depending on dose, frequency of use, urinary pH, clearance, and hydration. Prolonged positive immunoassay results from 5 to 10 days have been reported after compulsive cocaine use with continuous positive results for up to 16 days [26, 27].

Other Specimens

Cocaine and metabolites may be detectable in hair for considerably longer periods of time than that in urine. For this reason, analysis of hair may be useful in determining a past history of cocaine use. Cocaine is present at higher concentrations than benzoylecgonine and ecgonine methyl ester in hair. Many important issues related to hair testing are still being studied including environmental contamination, washing techniques, sex or racial bias, specimen adulteration, quality-control procedures, proficiency testing, and the establishment of cutoff concentrations. Hair testing may be useful in postmortem cases where no other specimen is available and drug history is desired.

Saliva (oral fluid) may be a useful specimen for the detection of recent cocaine use in clinical studies, workplace testing, and roadside sobriety tests. Generally, a good correlation between saliva and plasma cocaine concentrations has been observed [28]. It offers the distinct advantage of collection by direct observation without any invasive procedure. It is important to develop standardized collection protocols for the collection of saliva because there is a risk for collecting a contaminated specimen from the oral cavity, especially shortly after cocaine administration. The saliva-to-plasma-cocaine ratio can also be affected by saliva pH and saliva flow rate after stimulation.

Methods of Analysis

For most purposes, analysis of cocaine and metabolites involves an initial test (screen) to rule out the negatives, and a second test to confirm positive initial findings. Immunoassays are commonly used for screening purposes because they are readily amenable to large-batch analysis, are relatively sensitive, and require little or no sample preparation. Because most immunoassays are targeted to detect benzoylecgonine, they are particularly well suited for screening urine specimens. There are several types of immunoassays on the market. Depending on the immunoassay selected, analysis of blood and tissue homogenates may be performed either directly or after protein precipitation and/or solvent extraction.

Confirmatory testing is usually accomplished using a chromatographic separation technique, followed by detection using a variety of available

detectors. Before cocaine and its metabolites can be analyzed using chromatographic techniques, the drugs must first be isolated from the biological matrix. This may be accomplished either by liquid–liquid or solid-phase extraction procedures for which there are many available procedures. For forensic purposes, chromatographic separation and detection is usually performed using gas chromatography-mass spectrometry or liquid chromatography-mass spectrometry. This ensures that there is adequate sensitivity and selectivity in the confirmation test. More recently, tandem mass spectrometry has been utilized using both chromatographic techniques. Depending on the techniques used, some metabolites, such as benzoylecgonine, require derivatization before they can be analyzed.

References

[1] National Survey on Drug Use and Health (2004). http://www.samhsa.gov.

[2] Volkow, N.D., Wang, G.J., Fischman, M.W., Foltin, R., Fowler, J.S., Franceschi, D., Franceschi, M., Logan, J., Gatley, S.J., Wong, C., Ding, Y.S., Hitzemann, R. & Pappas, N. (2000). Effects of route of administration on cocaine induced dopamine transporter blockade in the human brain, *Life Sciences* **67**, 1507.

[3] Cone, E.J. (1995). Pharmacokinetics and pharmacodynamics of cocaine, *Journal of Analytical Toxicology* **19**, 459.

[4] Inaba, T., Stewart, D.J. & Kalow, W. (1978). Metabolism of cocaine in man, *Clinical Pharmacology and Therapeutics* **23**, 547.

[5] Dean, R.A., Christian, C.D., Sample, R.H.B. & Bosron, W.F. (1991). Human liver cocaine esterases: ethanol-mediated formation of ethylcocaine, *The FASEB Journal* **5**, 2735.

[6] Brzezinski, M.R., Abraham, T.L., Stone, C.L., Dean, R.A. & Bosron, W.F. (1994). Purification and characterization of a human liver cocaine carboxylesterase that catalyzes the production of benzoylecgonine and the formation of ethylcocaine from ethanol and cocaine, *Biochemical Pharmacology* **48**, 1747.

[7] Kloss, M.A., Rosen, G. & Rauckman, E.J. (1984). Cocaine-mediated hepatotoxicity, *Biochemical Pharmacology* **33**, 16.

[8] Jenkins, A.J. & Goldberger, B.A. (1997). Identification of unique cocaine metabolites and smoking by-products in postmortem blood and urine specimens, *Journal of Forensic Sciences* **42**, 824.

[9] Shimomura, E.T., Hodge, G.D. & Paul, B.D. (2001). Examination of postmortem fluids and tissues for the presence of methylecgonidine, ecgonidine, cocaine, and benzoylecgonine using sold-phase extraction and gas chromatography-mass spectrometry, *Clinical Chemistry* **47**, 1040.

[10] Cone, E.J., Tsadik, A., Oyler, J. & Darwin, W.D. (1998). Cocaine metabolism and urinary excretion after different routes of administration, *Therapeutic Drug Monitoring* **20**, 556.

[11] Isenschmid, D.S., Fischman, M.W., Foltin, R.W. & Caplan, Y.H. (1992). Concentration of cocaine and metabolites in plasma of humans following intravenous administration and smoking of cocaine, *Journal of Analytical Toxicology* **16**, 311.

[12] Jufer, R.A., Walsh, S.L. & Cone, E.J. (1998). Cocaine and metabolite concentrations in plasma during repeated oral administration: development of a human laboratory model of chronic cocaine use, *Journal of Analytical Toxicology* **22**, 435.

[13] Javaid, J.I., Musa, M.N., Fischman, M., Schuster, C.R. & Davis, J.M. (1983). Kinetic of cocaine in humans after intravenous and intranasal administration, *Biopharmaceutics and Drug Disposition* **4**, 9.

[14] Isenschmid, D.S. (2002). Cocaine – effects on human performance and behavior, *Forensic Science Review* **14**, 64.

[15] Ambre, J., Ruo, T., Nelson, J. & Belknap, B. (1988). Urinary excretion of cocaine, benzoylecgonine, and ecgonine methyl ester in humans, *Journal of Analytical Toxicology* **12**, 301.

[16] Isenschmid, D.S., Levine, B.S. & Caplan, Y.H. (1992). The role of ecgonine methyl ester in the interpretation of cocaine concentrations in postmortem blood, *Journal of Analytical Toxicology* **16**, 319.

[17] Jeffcoat, A.R., Perez-Reyes, M., Hill, J.M., Sadler, B.M. & Cook, C.E. (1989). Cocaine disposition in humans after intravenous injection, nasal insufflation (snorting), or smoking, *Drug Metabolism and Disposition* **17**, 153.

[18] Brogan, W.C., Lange, R.A., Glamann, D.B. & Hillis, R.D. (1992). Recurrent coronary vasoconstriction caused by intranasal cocaine: possible role for metabolites, *Annals of Internal Medicine* **116**, 557.

[19] Saady, J.J., Bowman, E.R. & Aceto, M.D. (1995). Cocaine, ecgonine methyl ester, and benzoylecgonine plasma profiles in rhesus monkeys, *Journal of Analytical Toxicology* **19**, 571.

[20] Blaho, K., Logan, B., Winbery, S., Park, L. & Schwilke, E. (2000). Blood cocaine and metabolite concentrations, clinical findings, and outcome of patients presenting to an ED, *The American Journal of Emergency Medicine* **18**, 593.

[21] Warner, A. & Norman, A.B. (2000). Mechanisms of cocaine hydrolysis and metabolism in vitro and in vivo: a clarification, *Therapeutic Drug Monitoring* **22**, 266.

[22] Logan, B.K. (1998). Considerations when trying to determine the role of cocaine in death. *Presentation – California Association of Toxicologists Quarterly Meeting*, San Francisco, CA, February, 1998.

[23] Isenschmid, D.S., Levine, B.S. & Caplan, Y.H. (1989). A comprehensive study of the stability of cocaine and its metabolites, *Journal of Analytical Toxicology* **13**, 250.
[24] Logan, B.K., Smirnow, D. & Gullberg, R.G. (1997). Lack of predictable site-dependent differences and time-dependent changes in postmortem concentrations of cocaine, benzoylecgonine, and ethylcocaine in humans, *Journal of Analytical Toxicology* **20**, 23.
[25] Hearn, W.L., Keran, E.E., Wei, H.W. & Hime, G.H. (1991). Site dependent postmortem changes in blood cocaine concentrations, *Journal of Forensic Sciences* **36**, 673.
[26] Cone, E.J. & Weddington Jr, W.W. (1989). Prolonged occurrence of cocaine in human saliva and urine after chronic use, *Journal of Analytical Toxicology* **13**, 65.
[27] Burke, W.M., Ravi, N.V., Dhopesh, V., Vandegrift, B. & Manny, I. (1990). Prolonged presence of metabolite in urine after compulsive cocaine use, *The Journal of Clinical Psychiatry* **51**, 145.
[28] Cone, E.J., Kumor, K., Thompson, L.K. & Sherer, M. (1988). Correlation of saliva cocaine levels with plasma levels and with pharmacologic effects after intravenous cocaine administration in human subjects, *Journal of Analytical Toxicology* **12**, 200.

Further Reading

Karch, S.B. (1998). *A Brief History of Cocaine*, CRC Press, Boca Raton.
Karch, S.B. (2007). *Drug Abused Handbook*, 2nd Edition, CRC Press, Boca Raton.
Levine, B. (ed) (2006). *Principles of Forensic Toxicology*, Revised and Updated 2nd Edition, AACC Press, Washington, DC.

DANIEL S. ISENSCHMID

Code of Conduct for Expert Witnesses *see* Ethics: Codes of Conduct for Expert Witnesses

Coercion *see* Compulsion

Cofiler/CofilerPlus

The adoption of 13 core short tandem repeat (STR) loci for combined offender DNA index system (CODIS) entry in the United States prompted the creation of STR multiplex systems that encompassed

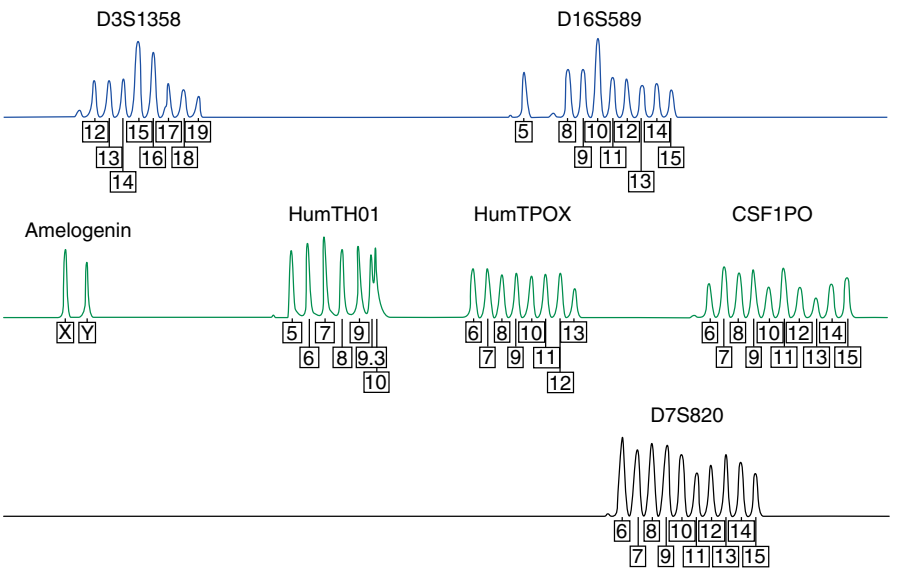


Figure 1 The seven loci amplified using the AmpF*l*STR® COfiler™ Multiplex System

the core loci. The AmpF*l*STR® COfiler™ STR multiplex system (Applied Biosystems, Foster City, CA) is designed to compliment the widely used nine-locus AmpF*l*STR® Profiler Plus™ STR multiplex system (Applied Biosystems, Foster City, CA). COfiler™ amplifies six tetranucleotide microsatellite loci and the Amelogenin gender marker. When both Profiler Plus™ and COfiler™ are applied the full set of 13 CODIS loci are typed. In addition, the two additional loci in the COfiler™ system (D3S1358 and D7S820) and the amelogenin locus overlap with the Profiler Plus™ system and are able to function as an internal quality control for the results of each test. The loci typed in the COfiler™ system are displayed in the image of the COfiler ladder (Figure 1).

Related Articles

Databases

SIMON J. WALSH

Cognitive Behavioral Therapy ***see*** Behavioral Science Evidence, Sex Offenders: Treatment of

Cognitive Functioning ***see*** Mental Retardation

Cognitive Interview ***see*** Elderly in Court

Collection of Evidence ***see*** Packaging and Transport

Command Hallucinations ***see*** Hallucinations

Communicating Evidence in Court ***see*** Statistical Evidence in Court

Communicative Arson ***see*** Firesetting

Community Mental Health ***see*** Disaster Mental Health

Competency ***see*** Capacity Assessment

Competency to Stand Trial ***see*** Capacity to Stand Trial

Compulsion

Introduction

According to the *Oxford English Dictionary* [1], compulsion is an "action, or an act, of compelling, or the condition of being compelled; constraint, obligation, coercion," and an "insistent impulse to behave in a certain way, contrary to one's conscious intentions or standards". Compulsion is thus defined by the presence of a force, either external or internal, that tends to impose behavior upon an individual.

External force may include coercion by others, as invoked in the affirmative defense of duress. A reasonable person confronted with a stranger brandishing a gun and instructed to steal a candy bar would likely perform the criminal act rather than risk losing his life. External force also operates in compulsory legal enforcement of treatment such as involuntary hospitalization for mental illness (*see* **Civil Commitment**) or in "leveraged" [2] forms of outpatient treatment such as outpatient commitment or court-mandated treatment as a condition of probation, though Bonnie and Monahan argue that the "coercive" aspect of these measures has been overemphasized [3].

External force also includes legal directives which an individual specifies while competent and which are later applied at a time when he is no longer competent. This is the case in *psychiatric advance directives*, where the individual creates a statement indicating his preferences for treatment at a future time when he no longer has the capacity to make a rational decision. In this way, the person places himself under a binding compulsion that he has formulated while lucid.

Internal forces leading the individual to act are comprised of complex phenomena within a wide range of psychopathology (*see* **Psychopathology: Terms and Trends**, and are also comprised of motivations and drives on a continuum with normal. Internal compulsion is thus more difficult to prove as a matter of fact, and is likely to require expert evaluation by a mental health professional. Whereas the presence of external compulsion is demonstrated through physical evidence and witness testimony according to an objective standard (a reasonable person would feel compelled), internally compelling forces are subjective, in that they are specific to the individual, arising from particular circumstances and beliefs that may be delusional or otherwise the product of psychopathology.

Involuntariness

The mental health field, and particularly the medical specialty of psychiatry, is distinguished by the fact that it is common to treat patients on an involuntary basis. The ability to assess the need for civil commitment is an essential skill of the psychiatrist. Psychiatrists hold significant power to limit the liberty of individuals with mental illness, and the ethics of the use of compulsory measures is necessarily an ongoing subject of reflection in the field.

Whether in the treatment setting or in a forensic evaluation, the assessment of an individual's ability to control his behavior is among the most difficult tasks for a mental health professional. Authors writing in the psychoanalytic tradition have defined compulsion as "a form of behavior to which the subject is obliged by an internal constraint [4]." Clinically, compulsion may be accompanied by anxious struggle against the impulses, with anxiety increasing the more the individual refrains from the compulsive behavior.

A related concept is the "repetition compulsion," in which the person is seized by an "ungovernable" tendency to put himself into "distressing situations, thereby repeating an old experience", [4] which he does not consciously recall. Freud began developing this concept prior to World War I, and formulated it explicitly in 1920 in his work *Beyond the Pleasure Principle* [5]. On the basis of his observation that soldiers returning from war had nightmares of their traumatic experiences, Freud concluded that mental functioning was not entirely guided by wish fulfillment (pleasure), but that the psyche included repetition of painful, unpleasant experiences that emerged as though forced upon the individual against his bidding. Psychoanalysts have not reached a consensus regarding explanations for this phenomenon, but agree that it exists [4].

Other accounts of compulsion as an involuntary phenomenon include biological explanations of temperament as a quantitative effect [6, 7] of the person's amount of instinctual drive. Social conventions, such as peer pressure in adolescents, or values such as cleanliness and purity in religious cultures, may become internalized by an individual and act as coercive forces within him. Whether it is attributed to the

force of unconscious meanings or to the compelling nature of social values or biological factors, the drive to act is experienced by the individual as surpassing his conscious control. In the study of pathological conditions such as addiction, controversy exists regarding the degree of control that the person ultimately has.

A Phenomenon of Mind and Behavior

The concept of internal compulsion is present in the definition of a wide variety of psychopathological conditions. A classic description of pathological compulsion is that of a person with obsessive-compulsive disorder who is unable to stop washing his hands, and does not know why. However, compulsion is also experienced by individuals with eating disorders, tics, pathological gambling, compulsive shoplifting, fire-setting, and uncontrolled buying. There is some evidence that these disorders are related to obsessive-compulsive disorder at the biological and psychological level [8]. Dermatological conditions may give rise to and result from compulsive behavior, such as hair pulling, itching, and scratching [8]. Individuals with personality disorders, such as borderline personality disorder, may experience the compulsion to self-injury. In some personality disorders as well as attention-deficit hyperactivity disorder, bipolar disorder, and posttraumatic stress disorder, the individual may manifest impulsive behavior, driven by a compulsion that may or may not be associated with a specific idea.

Impulsive Behavior

A definition of impulsivity as "a predisposition toward rapid, unplanned reactions to internal or external stimuli without regard to the negative consequences of these reactions to the impulsive individual or to others [9]," crosses many diagnostic categories. Impaired impulse control is included in the DSM-IV [10] symptom criteria of disorders such as antisocial personality disorder, borderline personality disorder, substance use disorders, attention-deficit hyperactivity disorder, and bipolar disorder [9]. One study suggests that impulse-control disorders were present in over 30% of psychiatric inpatients [11]. These disorders may be a common complication of other major psychiatric disorders, requiring specific treatment focused on these impulse-control symptoms [11].

There is some evidence that entering an altered state of consciousness ("dissociation") and a decreased ability to express feelings through words ("alexithymia") are associated with impulsive behavior, particularly self-mutilation in patients with a history of sexual abuse [12]. Impulsive behavior has also been studied in the context of violence and aggression (*see* **Aggression**), where the distinction between unplanned (impulsive) and planned aggressive behavior has been conceptualized as a basis for different treatment and different patterns of risk for future violence. "*Threat-control override*" [13] symptoms such as believing that thoughts are being inserted into one's mind, or voices that instruct the individual to perform wrongful acts, may result in actions that the person finds difficult to resist [14].

Addiction

Addictive behavior (*see* **Addictions**) is often invoked as a paradigm of compulsion, since nowhere is the drive to act more apparent than in the repeated intake of harmful substances, despite knowledge of the consequences. However, controversy exists as to the mechanism of the repeated behavior and the degree to which individuals have control and therefore responsibility for acting on their addictive impulses [15].

The choice to act on the compulsion to use illegal drugs has been viewed as a balance between pleasure-seeking tendencies and rational control. *Reward-related learning* is one theory that emphasizes neurobiological mechanisms that underlie pleasure-seeking behavior, which entail the creation of patterns that become reinforced even when the behavior no longer results in pleasurable outcomes [16]. *Impaired inhibitory control* seeks to explain repeated compulsive behavior as an abnormality of the brain circuits that regulate inhibition. In this model, a failure to inhibit compulsive behavior is a stronger factor than the strength of the drive for pleasure or the compulsion to act. Decreased ability to refrain from acting on impulses is thereby related to an inability to delay action to obtain satisfactions that are only possible at a later time [17]. This way of viewing the individual's choices as a balance between two systems governing motivation and control has been subject to criticism, because it does not take into account the impact of emotions in guiding decisions [18]. Mood and emotional state may tip the balance in a transient

manner, beyond the intensity of desire or the capacity to control or inhibit impulses.

Economic theories entail the construction of models for determining thresholds for choosing a particular action. In this perspective, the magnitude of the reward and the time to obtain the reward are factored into a system of relative values. For example, if the reward of graduating from college takes four years and requires abstinence from illegal drugs, whereas the reward of obtaining a small amount of pleasure through intoxication only takes 10 minutes to achieve, the individual may choose the more rapid and immediate satisfaction, even though the reward is smaller. In the theory of *delay discounting* or *hyperbolic discounting* the value of the long-term reward is diminished or "discounted" relative to the immediacy of the short-term satisfaction [19]. One way of treating this problematic is *choice bundling*, in which rewards are grouped and become sufficiently large in the person's mind as to override the desirability of short-term satisfaction of an impulse [19]. If multiple positive long-term outcomes are reinforced as a combination, their heightened value may make the immediate gratification no longer worthwhile. In this theoretical framework, the individual's self is viewed as a "population" with competing interests that require bargaining to arrive at a decision. Alliances or coalitions between interests and goals that share the aim of avoiding adverse short-term outcomes such as compulsively abusing illegal drugs may override the anticipated satisfaction that the individual seeks in using them [19].

Ego depletion is a concept in which the ability to control one's behavior is considered to be a resource that can become exhausted ("depleted") under conditions that require either strenuous or prolonged effort [20]. In psychoanalytic theory, the ego is the part of the psyche that is responsible for regulating drives in accordance with social norms. Depletion of the ego is thereby understood as a process in which the individual encounters increasing difficulty in regulating his choices. This was initially demonstrated through psychological experiments in which the ability to sustain concentration on a task was decreased when the task was immediately preceded by another strenuous exercise requiring self-control [20]. Further elaboration of the theory through experiments has led psychologists to distinguish between *autonomous* and *controlled* regulation of choice. These latter experiments suggest that choices that are made autonomously or spontaneously in accordance with the subject's personal desires are less fatiguing than controlled choices that are made under conditions where the subject is told that he must follow a particular rule that is imposed [21]. Although this distinction is logical, one unanswered question that is relevant for treatment is how self-regulation that is initially made under controlled or imposed conditions may become, over time, more spontaneous for the individual. Clinically, persons who attempt to change their patterns of behavior require intensive effort when first modifying their habits, and may gradually integrate these changes in a more personalized and automatic manner, particularly when they remove themselves from an environment that may contribute to depletion of self-control [22].

Subjective Experience of Coercion

Individuals placed under the compulsion of externally imposed measures react in a wide variety of ways. Prison inmates may adapt to their forced surroundings by embracing their identity as criminals among other criminals, in a process known as *prisonization* [23, 24]. When faced with conditions that are beyond their control, individuals may attempt to gain control by complying in a manner that they believe to be of their own choosing, such as volunteering to perform work while in prison or maintaining good behavior to obtain early release. A common reaction to imposed conditions that are perceived as coercive is that of rebellion. Although disobeying the parameters of confinement is counterproductive for the individual and paradoxically results in increased enforcement, a person may experience satisfaction in transgressing the limits of imposed constraint.

Court-mandated mental health and substance abuse treatment presents particular conceptual difficulties. Mental health treatment and substance abuse programs are most effective when the patient is motivated for treatment. External compulsion to adhere to treatment appears to run contrary to this principle. Many patients are compliant with the requirement for treatment only for the time that it is mandated by the Court. However, others initially comply owing to external compulsion and discover that treatment is of value to them. In these latter cases, an injunction for treatment serves to introduce the individual over a significant period of time to the treatment, which he is then able to make his own.

An individual's degree of awareness of internal compulsion may depend on what form of psychopathology is present. Persons who experience an internally compelling force such as addiction or impulsivity sometimes develop techniques, either spontaneously or learned through treatment, to oppose the compulsion. These individuals may describe intense psychological discomfort in their awareness of a struggle to resist acting on the compulsion. Persons who fail to refrain from acting on their impulses may experience distress and shame at not following their own rules and values.

Applications in Legal Proceedings: Does Free Will Matter?

A fundamental principle of law is that of the reasonable person's choice in action [25]. This is most clearly present in assessing *mens rea* in determinations of responsibility in criminal proceedings. At the opposite end of the spectrum, strict liability assigns responsibility even in the absence of awareness of wrongdoing. The controversial area of unconscious motivation implies a choice without awareness, and is generally excluded from legal proceedings, though formulations of some factors below the level of the individual's conscious awareness may contribute to criminal mitigation.

Although the topic of compulsion seems to imply an understanding of the problem of free will, and the question of whether actions are freely willed or not has been explored extensively by numerous authors in different disciplines, ranging from philosophy and sociology to neuroscience, the concept of free will has only limited relevance in legal adjudications. Pockett outlines three main positions with regards to free will. The *compatibilist* view posits that "in the absence of external (and arguably also internal) compulsion, acts are said to be freely willed [26]." The *libertarian* view requires that action be initiated consciously to be free. The *incompatibilist* view, between compatibilist and libertarian views, relies upon the determinism of physical laws and therefore implies that either there is no place for conscious initiation of action, or that if there is a role for consciousness, it is subsumed to unconscious processes. Pockett reviews experiments that suggest that actions are initiated prior to the individual's awareness of action, and proposes that, although legal decisions currently rely upon the compatibilist view (absence of compulsion indicates choice and therefore responsibility), scholars may wish in the future to examine whether a more accurate definition of intent should include unconscious motive. Kroeber presents the historical background of this debate regarding physical determinism and concludes that although there may be physical (and biological) laws governing behavior, civil and human laws cannot be understood in these terms. He suggests that individuals under the law are presumed to be aware of their actions incompletely but to the best of their ability, and are therefore responsible [27].

Morse presents an overview of the debates regarding free will in the legal context [25], and concludes that the concept of free will has no place in forensic mental health assessment. He points out that legal decisions do not require an evaluation of the underlying cause of behavior, including whether or not the action was a product of free will. He argues that the compatibilist position is the relevant measure of volition, in that in the absence of compulsion, an individual is deemed responsible. The law assumes that individuals are rational subjects, and this rationality or lack thereof is appreciated at the level of individual behavioral differences independent of the metaphysical question of what type of cause underlies the behavior and whether or not the action was freely chosen in the sense of being initiated at a conscious level [25].

Conclusion

Compulsion is an abstract concept that pertains to diminished voluntary control over one's actions, owing to a force that may be externally imposed or which arises internally in the context of pathological mental conditions. Psychiatric disorders characterized by reduced impulse control or addictive patterns may contribute to behavior that brings the individual to the attention of the legal system, through criminal charges or civil suit. Although many different theories exist to explain a person's reduced control over behavior, further empirical research is required to describe the phenomenon fully and to synthesize the findings of disparate approaches. Whether external or internal, compulsion may induce suffering in the individual under its sway. In legal proceedings, compulsion is assessed on the basis of the individual's ability to exercise reason and the presence or absence of force, rather than on the basis of metaphysical

considerations of free will or the ultimate biological, social, or chemical causes determining action.

References

[1] Oxford English Dictionary Online (2008). http://dictionary.oed.com/entrance.dtl. Last accessed January 21, 2008.

[2] Monahan, J., Redlich, A.D., Swanson, J., Robbins, P.C., Appelbaum, P.S., Petrila, J., Steadman, H.J., Swartz, M., Angell, B. & McNiel, D.E. (2005). Use of leverage to improve adherence to psychiatric treatment in the community, *Psychiatric Services* **56**, 37–44.

[3] Bonnie, R.J. & Monahan, J. (2005). From coercion to contract: reframing the debate on mandated community treatment for people with mental disorders, *Law and Human Behavior* **29**(4), 485–503.

[4] Laplanche, J. & Pontalis, J.B. (1973). *The Language of Psychoanalysis*, The Hogarth Press and the Institute of Psychoanalysis, London, Compulsion, Compulsive 77–78, Compulsion to Repeat (Repetition Compulsion), pp. 78–80.

[5] Freud, S. (1920). Beyond the Pleasure Principle, group Psychology and other words, in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. J. Strachey, Hogarth Press and institute of psychio-Annaly, London Vol. 18.

[6] Lara, D.R., Pinto, O., Akiskal, K. & Akiskal, H.S. (2006). Toward an integrative model of the spectrum of mood, behavioral and personality disorders based on fear and anger traits: I. Clinical implications, *Journal of Affective Disorders* **94**, 67–87.

[7] Lara, D.R. & Akiskal, H.S. (2006). Toward an integrative model of the spectrum of mood, behavioral and personality disorders based on fear and anger traits: II. Implications for Neurobiology, Genetics and Psychopharmacological treatment, *Journal of Affective Disorders* **94**, 89–103.

[8] Del'Osso, B., Altamura, A.C., Allen, A., Marazziti, D. & Hollander, E. (2006). Epidemiological and clinical updates on impulse control disorders: a critical review, *European Archives of Psychiatry and Clinical Neuroscience* **256**, 464–475.

[9] Moeller, F.G., Barratt, E.S., Dougherty, D.M., Schmitz, J.M. & Swann, A.C. (2001). Psychiatric aspects of impulsivity, *American Journal of Psychiatry* **158**, 1783–1793.

[10] American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition Text revision, American Psychiatric Press, Washington, DC.

[11] Grant, J.E., Levine, L., Kim, D. & Pontenza, M.N. (2005). Impulse control disorders in adult psychiatric inpatients, *American Journal of Psychiatry* **162**, 2184–2188.

[12] Zlotnick, C., Shea, M.T., Pearlstein, T., Simpson, E., Costello, E. & Begin, A. (1996). The relationship between dissociative symptoms, alexithymia, impulsivity, sexual abuse, and self-mutilation, *Comprehensive Psychiatry* **37**(1), 12–16.

[13] Link, B.G., Stueve, A. & Phelan, J. (1998). Psychotic symptoms and violent behaviors: probing the components of "Threat/Control-Override" symptoms, *Social Psychiatry and Psychiatric Epidemiology* **33**, S55–S60.

[14] Braham, L.G., Trower, P. & Birchwood, M. (2004). Acting on command hallucinations and dangerous behavior: a critique of the major findings in the last decade, *Clinical Psychology Review* **24**, 513–528.

[15] See special issue of the American Journal of Bioethics 2007 volume 7, which includes a series of commentaries on the lead article, Hyman, S.E. (2007). The Neurobiology of addiction: implications for voluntary control of behavior, *American Journal of Bioethics* **7**, 8–11.

[16] Hyman, S.E., Maleka, R.C. & Nestler, E.J. (2006). Neural mechanisms of addiction: the role of reward-related learning and memory, *Annual Review of Neuroscience* **29**, 565–598.

[17] Lubman, D.I., Yücel, M. & Pantelis, C. (2004). Addiction, a condition of compulsive behaviour ? Neuroimaging and neuropsychological evidence of inhibitory dysregulation, *Addiction* **99**, 1491–1502.

[18] Charland, L.C. (2007). Affective neuroscience and addiction, *American Journal of Bioethics* **7**, 20–21.

[19] Monterosso, J. & Ainslie, G. (2007). The behavioral economics of will in recovery from addiction, *Drug and Alcohol Dependence* **90S**, S100–S111.

[20] Baumeister, R.F., Bratslavsky, E., Muraven, M. & Tice, D.M. (1998). Ego depletion: is the active self a limited resource? *Journal of Personality and Social Psychology* **74**, 1252–1265.

[21] Moller, A.C., Deci, E.L. & Ryan, R.M. (2006). Choice and ego-depletion: the moderating role of autonomy, *Personality and Social Psychology Bulletin* **32**, 1024–1036.

[22] Levy, N. (2007). The social: a missing term in the debate over addiction and voluntary control, *American Journal of Bioethics* **7**(1), 35–36.

[23] Gillespie, W. (2002). *Prisonization. Individual and Institutional Factors Affecting Inmate Conduct*, LFB Scholarly Publishing, LLC, New York.

[24] Walters, G.D. (2003). Changes in criminal thinking and identity in novice and experienced inmates. Prisonization revisited, *Criminal Justice and Behavior* **30**, 399–421.

[25] Morse, S.J. (2007). The Non-problem of Free Will in Forensic Psychiatry and Psychology, *Behavioral Sciences and the Law* **25**, 203–220.

[26] Pockett, S. (2007). The concept of free will: philosophy, neuroscience and the law, *Behavioral Sciences and the Law* **25**, 281–293, page 292.

[27] Kroeber, H.L. (2007). The Historical Debate on Brain and Legal Responsibility – Revisited, *Behavioral Sciences and the Law* **25**, 251–261.

SUZANNE YANG

Compulsion Defense

Introduction

The law acknowledges that circumstances exist which may reduce or even exculpate an individual's behavior that would otherwise be considered criminal (*see also* **Insanity: Defense**). The fundamental principle that individuals may have genuine excuses for their acts, otherwise unacceptable in society, has its precedent in early social norms as well as in the history of criminal law. Nonetheless, few instances in the law elicit as much ambivalence as when individuals present themselves free from criminal responsibility due to unusual circumstances or states of mind. The most extreme example of the incapacity to be held criminally culpable, such as mental illness in the insanity defense, may be easier to intuitively accept as a legal standard because of the obvious deviations from common human experience.

The defenses of excuse, including coercion and compulsion, however, offer complex legal challenges that stem, in part, from their uncertain acceptance by society for psychological and scientific reasons. The circumstances of the criminal acts in compulsion defenses are such that we can readily imagine our own emotional and behavioral responses were we to face the same the terrible choices as the unfortunate defendant.

Society's mixed feelings regarding such defenses have perhaps contributed to the frequent confusion around their legal definitions and, as a result, the landscape of legal doctrine regarding such defenses is sometimes unclear. With advances in neuroscientific research in recent decades towards understanding the foundations of mental functioning and resulting human behaviors, there has also been heightened attention to its potential role in law rulings. Consequently, the legal principles of intention, choice (free will) and consciousness have been more closely scrutinized through the lens of neuroscience fueling an ongoing debate about its impact and in legal proceedings.

Definition of Compulsion Defenses

The word "compulsion" is defined in the Merriam-Webster Law Dictionary as "an act of compelling (as by threat or intimidation), specifically, 'coercion"' or "the state of being compelled, specifically 'Duress"'. A "compulsion defense" as a legal concept is vague in its definition because it is not a uniformly accepted term of art. Rather, the category of "compulsion defense" encompasses, according to some experts, distinct subbranches including defenses of "duress", "necessity", and "self defense" (the last of which will not be addressed in this review.) Here, the specific legal terms "duress" and "necessity" will be briefly discussed with respect to their distinctive uses as well as the relationship to the other related legal concepts of "excuses" and "justifications".

Duress

"Duress" is a defense accessible in the federal court system and also in several states. It excuses criminal conduct where "the actor was under an unlawful threat of imminent death or serious bodily injury, which threat caused the actor to engage in conduct violating the literal terms of the criminal law." Otherwise unlawful conduct caused by such duress will excuse the crime in question "unless that crime consists of intentionally killing an innocent third person" [1]. The historical rule of the common-law defense of duress does not apply to a charge of murder and thus far no federal courts have ruled on its admissibility as a defense to felony murder.

A defense of "duress" derives from the proof that a defendant was subjected to undue pressure or unusual circumstances outside of the defendant's control. An individual who commits an otherwise illegal act as a result of a threat of death or serious injury may affirmatively assert a defense of duress or coercion, thereby excusing him from criminal liability. A defendant exercising a defense of duress must bear the burden of proof being a preponderance of evidence. If the defendant is successful, some courts may give the prosecution the burden of persuading the jury that the defendant did not meet the elements of the defense.

Due to the heterogeneous nature of the legal criteria within statutory law and the variations among jurisdictions, the following elements represents a synthesis of common elements for the definition of duress: (i) an immediate threat of death or serious bodily injury to the defendant (The defendant must not merely perceive a threat, but must be actually faced with such a threat), (ii) a well-grounded

fear that the threat would be carried out; and (iii) no reasonable opportunity to avoid the threatened harm, (iv) any reasonable person in the defendant's position would have also committed the otherwise criminal act.

In practice, raising the defense of duress is seldom permitted during criminal prosecutions and, even where allowed to be presented to the fact finder, even less often successful in its outcome. In part this is due to the lack of agreement with respect to a fundamental legal theory or argument that would allow society to permit individuals to engage in unlawful behaviors. Also, the unclear nature of the law regarding these defenses reflects society's mixed feelings about the moral and emotional issues that underlie the issues of excuse and justification for otherwise criminal behavior. As reflection of this uncertainty, there has been sporadic attention in legal research for these issues except in cases of felony murder or where the issue of developmental dysfunction (e.g. Mental Retardation) is involved.

Necessity

The defense of "necessity" is distinguished from the defense of duress with respect to its legal criteria, although both constructs are seldom utilized in their essential forms. Commonly understood, the "necessity" defense is thought to pertain to threats from nature and its environment, while the duress defense has been relegated to circumstances in which the threats are considered man-made.

The necessity defense is raised when a defendant is faced with a naturally caused set of circumstances before being faced with a "choice of two evils". The necessity defense is often understood to be considered a "justification" defense based on the assumption that the individual chooses the lesser of two harms. The theory of "harms reduction" or "balance of harms" proposes that the harm caused by the individual's illegal act is less than were the actor to obey the law. An example would include an individual fleeing from an oncoming tornado and breaking into another residence's basement cellar for self-protection.

Conceptually related legal concepts include "justification" and "excuse" defenses. In such defenses, which also have been variably interpreted by legal scholars, the defendant has fulfilled the "mens rea" component of the criminal act, but has offered a conceivable argument as to why he should not receive punishment. In distinguishing a "justification" defense, the proposed argument is that an unlawful act was committed for the greater social interest. In contrast, an "excuse" defense asserts that specific circumstances prevent the individual from making a voluntary or free choice. A common example would include the insanity defense that is grounded in the individual's mental incapacity to make a reasoned decision.

In further clarification, one might consider the defense of "justification" to rest upon the nature of the criminal acts and the "excuse" defenses to depend upon the nature of the individuals who commit them. While there is no categorical distinction between the theories of "excuse" and "justification" when attempting to apply these constructs, duress is more commonly thought of as an "excuse". Some point to the absence of the underlying theory of "lesser of two harms" when clarifying duress from necessity. In further contrast between, the burden of the former lies with the defendant's affirmation while the burden of a justification defense would more expectedly rest upon the responsibility of the prosecution to prove its absence. In practice, these distinctions are artificially drawn and more often defenses are a combination of the two.

The Question of Neuroscientific Research on Issues of Criminal Responsibility

An analysis of criminal defenses based on internal [e.g., biological] and environmental variables that may potentially exculpate an individual would be ultimately unsatisfying without a brief mention of scientific research that has begun to challenge the underlying principles of our legal processes and systems. Advances in our understanding of the brain and its functioning have been increasingly focused on some of the most subtle and complex issues of the human experience, particularly, the empirical description of human consciousness and the notion of "free will".

These issues, once reserved for debate within the fields of philosophy and theology, are now receiving widespread attention for their potentially profound effects on legislation and case law. At the heart of one particular debate is the presumption of an individual's voluntary or "free" decision-making capacity that is a fundamental principle in the adjudication of criminal

responsibility. Scientific research has progressively attempted to break down and analyze the neurophysiological mechanisms behind such assumptions of "free will" and conscious intent, thereby raising questions about the nature of the law's authority.

Central to this neuroscientific discussion of "free will" was a series of experiments, developed by Benjamin Libet in 1983, in which attempted to trace the temporal relationships between an individual's conscious intent and the associated voluntary movement. Fundamental to Libet's work, was the previous determination of the "readiness potential" (RP), an electrophysiological measurement that, during an individual's decisional sequence, occurred hundreds of milliseconds before the recording of a voluntary movement.

In his experiments, Libet connected a series of electroencephalogram (EEG) electrodes to the scalp and mastoid areas of human subjects as well as an electromyography (EMG) electrode attached to their forearms. When the subjects were asked to move their arms, an EMG potential was recorded and a trigger signal was transmitted to the EEG recorder. Subjects were asked to watch a rotating spot of light on a cathode ray oscilloscope and were subsequently instructed to report on the position of the light at the moment of their being aware of wanting to move. This moment of first awareness was labeled "W" while the actual moment of movement was labeled "M". Approximately 40 trials of the EEG sweeps from the EMG recorder were then averaged.

Interestingly, the findings included averaged recordings of RPs occurring approximately 400 ms prior to "W" [the moment of first awareness] which itself preceded the measurements of "M" by approximately 200 ms. Libet's work demonstrated that measurable brain activity preceded not only voluntary movement but the individual's conscious awareness of that intent to move. Libet eventually hypothesized a "veto" function which would allow the conscious cessation of the unconscious decision in between moments "W" and "M".

Libet's work has been the source of extensive controversy; for its reliability, validity and its implications in legal standing. Using the ramifications of Libet's own work, critics have emphasized the confounding nature of other variables such other unaccounted events that affected his experiment. This would theoretically include the notion that the subjects' experience from the experiment's inception [e.g., when the subjects were first informed of the nature and details regarding the experiment's instructions] had created corresponding neural correlates to that experience and may have also effected the measurements of brain activity during recording.

Nonetheless, subsequent research in England [2] seemed to corroborate some of Libet's work, neuroanatomically mapping activity to three localized brain structures including the dorsal prefrontal cortex, the intraparietal sulcus and the presupplementary motor area. This and other similar research has subsequently underscored the issue that intentionality may truly commence before the individual's conscious awareness and, strangely, that the individual's attention to the preceding "absence" of awareness of the thought may actually catalyze the action itself. Alternatively understood, brain activity initiates a decision-making process before the conscious mind becomes aware of its intentions.

For some, these and similarly replicated results have reignited the age-old debate over the existence of "free will", a presumption upon which our modern legal system is commonly thought to rest. Instead, the implications of neuroscientific explorations have lead some, such as Daniel Wegner, to postulate that human behavior is derived from "determined and mechanistic processes". Such a position asserts that human behavior is more complex and comprehensively accounted for by the nuances of biological and neurophysiological processes that exist outside the human consciousness than by the reliance on the "incomplete" experience of our conscious will.

Some thinkers convincingly argue that these cellular and biological processes are shaped by previous environmental and emotional experiences which become part of the brain's "hard-wiring". These neurobiological representations then serve as activated cognitive substrates during future decision-making that are then triggered by similar environment stimuli of stressful situations or hard decisions.

Such research has raised the issue of the increasingly complex interplay between consciousness and unconsciousness during a decision-making processes. Do individuals freely make conscious and rational decisions for themselves and what are the forces (e.g., internal, environmental) that may shape the individual's rational choices and conscious behaviors?

Another neurobiological perspective on the process of decision-making is found in the research of Antonio Damasio who has explored the neural

basis of emotions and their inextricable relationship with the frontal lobe, the center of conscious control of human behavior. Damasio's work argues that emotions, as biologically represented in the limbic system, have profound influence over the regulation of an individual's decision-making processes. Damasio argues in his "Somatic Marker Hypothesis" that emotions arise from an integration of "somatic markers", direct signals from the body that are cognitively represented in the brain (particularly the ventromedial prefrontal cortex) and are recruited for, among other higher functions, the regulation of "rational" decision-making processes.

The above examples of neuroscientific research and the profound issues of morality they have been proposed to address are not offered to represent an established or cumulative body of evidence. Rather, these explorations are meant to underscore the scientific community's rapidly growing efforts at revealing the complex nature of the intersection between biology and human behavior. The interface of neuroscience and the law proves to be an exciting yet far from determined relationship.

References

[1] LaFave, W.R. (2003). *Substantive Criminal Law*, 2nd Edition, Thomson/West, St. Paul.

[2] Lau, H.C., Rogers, R.D., Haggard, P. & Passingham, R.E. (2004a). Attention to intention, *Science* **303**(5661), 1208–1210.

Further Reading

Dressler, J. (1989). Exegesis of the law of duress: justifying the excuse and searching for its proper limits, *Southern California Law Review* **62**, 1331.

Dunn, B., Dalgleish, T. & Lawrence, A. (2006). The somatic marker hypothesis: a critical evaluation, *Neuroscience and Biobehavioral Reviews* **30**, 239–271.

Kawohl, W. & Habermeyer, E. (2007). Free will: reconciling German civil law with libet's neurophysiologiocal studies on the readiness potential, *Behavioral Science and the Law* **25**(2), 309–320.

Morse, S. (1985). Excusing the crazy: the insanity defense reconsidered, *Southern California Law Review* **58**, 777.

Westen, P. & Mangiafico, J. (2003). The criminal defense of duress: a justification, not an excuse – and why it matters, *Buffalo Criminal Law Review* **6**, 833.

New York State Legislature (2007). New york cls penal 40.00 part one. provisions. title c. defenses. article 40. *Other Defenses Involving Lack of Culpability*.

ADAM N. RAFF

Computer: Animations as Evidence *see* Computer Animation and Simulation Evidence

Computer Animation and Simulation Evidence

Introduction

The presence and implementation of multimedia devices in the courtroom has become an increasingly common sight, in recent years.[a] Especially true in today's world, multimedia technology involves the extensive use of computer-generated visual products. Visual representations of key testimony and evidence both simplify and expand the ability of attorneys and experts to educate juries about complex information. Computer animations and simulations are two highly effective examples of visual techniques that break down dense information into manageable and compact segments. This technology served as one of the first types of computer-driven and created evidence introduced into the American legal system.

One significant reason for the relatively early use of computer animations and simulations is the immediate impact such evidence has on jurors. Additionally, animation and simulation was inexpensive compared to other multimedia options available 20 years ago [1].[b] While the humble beginnings of computer-generated evidence in the courtroom were mainly restricted to airline and vehicular accident scene reconstruction, the present technology has litigants employing animations and simulations in all manner of civil and criminal trials. Software is now interactive allowing experts to make adjustments during testimony and advanced enough that complex processes such as chemical reactions and machinery operations can be represented with relative ease.[c] To better understand why animation and simulation can be so important to a party's case, this article provides a brief explanation of how these technologies are viewed in the context of the courtroom,

including their influence on crime and accident scene processing.

Animation vs. Simulation

While computer animations and simulations may appear interchangeable on the surface, the construction methods and intended purposes of each differ. Computer animations are created in a two-step process that results in moving images much like those seen in animated movies. First, three-dimensional representations of all objects/locations to be included in the animations must be documented and then the images must be designed to move in a way that accurately represents the scene or information one is trying to show [2]. At this point, the still frames are strung together and played back in a set sequence; this results in a moving picture. Computer simulations, on the other hand, use both animation and mathematical computation to produce an illustrated presentation of possible outcomes predicted by a computer based on certain preset parameters [3].[d]

For instance, in an accident reconstruction simulation, facts about the accident are entered into the computer, which are then generated into probable outcomes based on assumptions the computer is programmed to apply. Specifically, alternative conclusions can be hypothesized, such as the wearing of a seatbelt, weather conditions, time of day, etc. Once a simulation is produced, it is converted into a graphic animation that can be stored on a CD or computer hard drive.[e]

Current Technology

As the use of computer-generated evidence has increased over the past 15 years, the methods used to collect the necessary information to formulate the recreations have advanced as well. The most up-to-date technology developed for crime and accident scene reconstruction involves the use of a device called a *total station*, which utilizes forensic mapping technology to coordinate and collate the pertinent data [4].

The total station is composed of four parts: a theodolite, an EDMI, an optical prism, and a data collector. A theodolite is an instrument that measures angles on horizontal and vertical axes. An EDMI (electronic distance measuring instrument) measures distance via infrared diode. The optical prism reflects the light emitted by the EDMI, and the data collector captures the measurements generated by the EDMI and theodolite and adds graphic attributes. These measurements are then fed into a forensic mapping program that converts the data into a diagram of evidence locations through a polar coordinate system.

When an investigator wants to use this system to record an accident or crime scene, an investigator will hold a pole, with the optical prism attached to the top, over the evidence to be documented. The infrared diode on the instrument is focused on the prism and reflected back, which allows the data collector in the machine to record the distance, angle, and elevation of the evidence. The operator presses a button to record the evidence and assigns it a code. He can then move on to the next location. Once all the sites at the scene are recorded, the information can be downloaded to a laptop on-site and later converted to an animation or simulation for use in court.[f]

Case Examples

In addition to the technical differences between animations and simulations, there is also a legal distinction applied by courts that determines the evidentiary function the moving images will have in the case presentation. If the images are merely presenting a moving illustration of testimony, then it is considered demonstrative evidence and referred to as an *animation*. The standard of admission for animation demonstrative evidence is fairly low requiring only that the evidence be relevant, accurate, and the probative value outweighs the possibility of prejudice [5]. On the other hand, if the images are meant to convey evidence to a jury, it is called a *simulation* and treated as substantive evidence subject to admission standards under *Daubert* or *Frye* [6].

While computer animation made its start in civil litigation, the first reported use of animation in an appellate criminal case was *People v. McHugh* [7]. The defendant, charged with second-degree manslaughter and driving while intoxicated, attempted to enter, into evidence, a computer recreation of his version of the events. The state argued that the evidence should be evaluated by the court via a pretrial *Frye* hearing as scientific evidence to ascertain if the mathematical formulas and techniques used in the computer program were reliable and generally accepted in the scientific community. The court,

however, did not view the recreation as scientific evidence and compared the presentation to a hand or mechanically drawn diagram. As an aid comparable to a drawing, the method of production is not the controlling factor in determinations of admissibility. Additionally, the court described computers as "mechanical tools – receiving information and acting on instructions at lightning speed" and held that the evidence was admissible provided it was relevant, accurate, aided the jury in understanding testimony and the expert was properly qualified.

Somewhat surprisingly, there are still courts that have not specifically decided if this type of evidence is admissible, despite the infiltration of computers into myriad aspects of daily life. A 2006 decision by the Pennsylvania Supreme Court, *Commonwealth v. Serge*, presented this question as an issue of first impression [8]. A defendant appealed the commonwealth's use of a computer-generated animation that depicted the state's version of the events that led to the murder of the defendant's wife. The information produced in the animation was based on physical and forensic evidence and was viewed by the trial court as a demonstrative aid for the jury to better understand the purported sequence of events. As a preliminary matter, this court noted that computer-generated evidence is just another mode of expressing information traditionally done through drawings or diagrams and as such, should not be held to a different or higher admissibility standard than other types of demonstrative evidence. Further, the court acknowledged that while this evidence may be more persuasive to a jury, that fact alone is not sufficient to warrant its exclusion. Thus, the court evaluated the evidence under the demonstrative evidence standard for admission: authentication, relevance, and probative value. The crucial prong for purposes of this case was whether the probative value of the computer-generated evidence outweighed its prejudicial effect. The court determined that the content of this presentation was not inflammatory beyond the extent inherent to a depiction of murder. The court specifically pointed to the following factors in not finding prejudice: (i) the absence of sound, (ii) no facial expressions, (iii) lack of life–life motion, (iv) no transition between scenes to create a story line, and (v) no images of blood or other injury. Therefore, the trial court properly allowed the admittance of the recreation animation.

While the admissibility requirements for computer animation are not complex, there are cases in which animations are not admitted into evidence. In *Dunkle v. State*, the state introduced four computer animations reenacting a murder based on statements made by the defendant and witnesses and physical evidence found at the scene [9]. The different scenarios presented to the jury were intended to dispute the defendant's version as inconsistent with the evidence. The animations placed the victim and gun in the same place but varied the location and actions of the defendant. The court questioned the accuracy and relevancy of the asserted sequences of events because the state had little physical evidence to go on in forming its version and the inconsistent statements of the defendant about the crime made it almost impossible to determine what story the state was arguing against. Consequently, it held that the admissions were improper and their admission was not harmless error. Further, the court acknowledged the strong impression and air of truth digital recreations give to juries and found the prejudice these animations would create due to the unsupported assertions that outweighed any relevancy or probative value they could have. Upon retrial, this court also required the trial court to issue instructions to the jury explaining the limited purpose of the reenactments as constructed illustrations of an expert's testimony and not substantive of the events that actually occurred.

By contrast, an example of the much more stringent admission requirements for computer simulations is illustrated in *State v. Sipin* [10]. In this case, the admissibility of a computer simulation was tested under the *Frye* standard for novel scientific evidence. The defendant was charged with vehicular homicide, and the state wanted to use an accident reconstruction simulation to show that the defendant was driving at the time of the accident using a simulation software program called *PC-Crash*. Specifically, the computations input into the program were supposed to produce the body movements of the multiple vehicle occupants during a multivehicle accident. The defendant objected to its introduction asserting that the state's expert had not validated the accuracy of the program for the predictive outcomes formed.

An accident reconstructionist expert produced two validation studies on the accuracy of the predictive simulations in car accidents. The first assessed the program's ability to predict accident-related information such as tire marks and speed against real world

data. The second looked at the body movement of a pedestrian hit outside of a vehicle during an accident. No validation study was produced that examined its accuracy in ascertaining the movement of multiple bodies within a car during an accident, and this program had never been used in a Washington court to predict interior movement in a comparable event. The expert supported his use of the program by arguing that the kinetic principles used to predict body movement inside and outside a vehicle do not differ and stated that there was not a debate in the scientific community about the ability of the program to calculate interior movement.

In determining whether scientific evidence meets the *Frye* standard of admissibility, the proponent must show that the science is generally accepted as valid among members of the relevant scientific community. Further, the application of the science at issue must be proper and accurate. In the case at bar, the court held that the application of the PC-Crash program to assess the body movement of interior occupants during a multivehicle crash was not proper. Manuals from the manufacturer explicitly warned of limitations the program had in calculating interior movement in accidents occurring at speeds in an excess of 40 kilometres per hour. Further, a multibody software option developed by the manufacturer to predict movement during accidents had preset physical parameters that constrained the predictive capabilities of the program. Also, a distributor of PC-Crash and a software programmer who worked on accident reconstruction programs both testified that PC-Crash was not sophisticated enough to produce accurate predictions for accidents of this type. Finally, the defendant produced three papers that argued against the accuracy of the program. These things in combination indicated to the court that there was not a consensus in the scientific community on the accuracy of PC-Crash and, consequently, it held that the simulation was inadmissible.

End Notes

[a.] In fact, some organizations are actively promoting and supporting the integration of technology into the courtroom, hoping to enhance the ability of the legal system to achieve justice. See the Center for Legal and Court Technology at William & Mary Law School. http://www.legaltechcenter.net

[b.] However, the cost of computer-generated evidence can easily add up to tens of thousands of dollars depending on the complexity of the reconstruction or process being depicted.

[c.] See Animators at Law (www.animators.com) for examples of animation representing biological processes, machine operation, and accident recreation.

[d.] Put simply, animations recreate real past events based on known facts, and simulations combine facts from past events with assumptions about various ways the events could/should have unfolded.

[e.] Under the category of animation, there is another type of computer-generated motion image known as *recreation animation*. This differs from basic animation in that it relies on scientific data to form the moving representation and is not limited to reproducing eyewitness testimony. The information put in is culled from facts determined after the occurrence of the illustrated event. Thus, some assumptions are made about the sequence of events and impact of forces, but it does not create predictive outcomes found in computer simulations.

[f.] Numerous software programs exist that can work in conjunction with total stations to create 3-D images. Some of these include Crime Zone, iWitness, and MapScenes.

References

[1] Breaux, S. (2002). Forensic animation, *Brief* **31**, 26–34.
[2] Joye, M. (1998). Computer animations, *Trial* **11**, 47–52.
[3] Galves, F. (2000). Where the not-so-wild things are: computers in the courtroom, the federal rules of evidence, and the need for institutional reform and more judicial acceptance, *Harvard Journal of Law and Technology* **13**, 161–300.
[4] Spraggs, D. (2004). The next dimension: detectives and crime scene investigators are using 3-D tech to bring crime scenes to life, *Police Magazine* **28**, 11.
[5] Lorraine v. Markel American Ins. Co. 241.F.R.D. 534 (D.Md., 2007); *Mintun v. Wyoming*, 966 P.2d 954 (Wyo., 1998); *Tillus v. Moses*, 748 So.2d 874 (Ala., 1999); *Harris v. State*, 13 P.3d 489 (Okla. Crim. App., 2000); *State v. Cauley*, 32 P.3d 602 (Colo. App., 2001).
[6] Morande at 1070–1071. See, *Frye v. United States*, 293 F. 1013 (D.C. Cir., 1923); *Daubert v. Merrell Dow Pharmaceutical*, 509 U.S. 579 (1993).
[7] 124 Misc.2d 559, 476 N.Y.S.2d 721 (N.Y. Supp. 1984).
[8] 586 Pa. 674, 896 A.2d 1170 (Pa. 2006).
[9] 139 P.3d 228 (Okla.Crim.App., 2007). For other cases, where animations were excluded, see *Sommervold v.*

Grevlos, 518 N.W.2d 733 (S.D. 1994); *Missouri v. Star*, 998 S.W.2d 61 (Mo. App., 1999); *Kane v. Triborough Bridge and Tunnel Authority*, 8 A.D.3d 239, 778 N.Y.S.2d 52 (N.Y.A.D. 2 Dept., 2004); *Smith v. Kansas City Southern Ry. Co.*, 846 So.2d 980 (La.App. 3 Cir. 2003); *Clark v. Cantrell*, 339 S.C. 369, 529 S.E.2d 528 (S.C. 2000).

[10] 130 Wash.App. 403, 123 P.3d 862 (Wash.App.Div. 1, 2005).

Further Reading

Beskind, D.H., Bocchino, A.J., Rothschild, F.D. & Siemer, D.C. *Effective Use of (2001). Courtroom Technology: A Judge's Guide to Pretrial and Trial*, Federal Judicial Center.

Albrecht, A.W. (2003). Laying a proper foundation for computer-generated demonstrative evidence, *Illinois Bar Journal* **90**, 261.

Baer, L.G. & Riley, C.A. (1999). Technology in the courtroom: computerized exhibits and how to present them, *Defence Counsel Journal* **66**, 176.

Bennett, R.B., Leibman, J.H. & Fetter Jr, R.E. (1999). Seeing is believing: or is it? An empirical study of computer simulations as evidence, *Wake Forrest Law Reviews* **34**, 257.

Borelli, M. (1996). The computer as advocate: an approach to computer-generated displays in the courtroom, *Indiana Law Journal* **71**, 439.

Breaux, S.P. (2003). Is forensic animation right for your case? *Trial* **39**, 66.

Butera, K.D. (1998). Seeing is believing: a practitioner's guide to the admissibility of demonstrative computer evidence, *Cleveland State Law Reviews* **46**, 511.

Campbell, R.P. (2003). Admissibility of test films, videotapes, and computer-generated technical evidence, *Brief* **32**, 13.

Carbine, J.E. & McLain, L. (1999). Proposed model rules governing the admissibility of computer-generated evidence, *Santa Clara Computer and High Technology Law Journal* **15**, 1.

Carney, B. & Feigenson, N. (2004). Visual persuasions in the Michael Skadel trial: enhancing advocacy through interactive media presentations, *Criminal Justice* **19**, 14.

Collins, J.M. (2003). Review of digital evidence and computer crime, *Journal of Forensic Science* **48**(2), 475.

Feigenson N. & Dunn M.A. (2003). *New Visual Technologies in Court: Directions for Research*. Law and Human Behavior, February 2003, pp. 109–126.

Fiedler, B.S. (2003/2004). Are your eyes deceiving you?: the evidentiary crisis regarding the admissibility of computer-generated evidence, *New York Law School Law Reviews* **48**, 295.

Fulcher, K.L. (1996). The jury as witness forensic computer animation transports jurors to the scene of the crime or automobile accident, *University of Dayton Law Reviews* **22**, 55.

Joseph, G.P. (1996). Virtual reality evidence, *Boston University of Journal Scientific and Technical* **2**, 12.

Joseph, G.P. (2003). A simplified approach to computer-generated evidence and animations, *ALI-ABA* **SJ048**, 147.

Joseph G.P. (2007). *A Simplified Approach to Computer-Generated Evidence and Animation*. American Law Institute-American Bar Association Course of Study March 2007, pp. 217–234.

Lederer, F.I. (2004). Courtroom technology: for trial lawyers, the future is now, *Criminal Justice* **19**, 14.

Malone, A. (2003). Affordable high tech trials, *Annals Association of Trial Lawyers of America* **2**, 2219.

Nations H.L. & Nations C.L. (2007). *The Rules of Digital Evidence*. American Law Institute-American Bar Association Course of Study July 2007, pp. 501–555.

Powell, C.E. (1996). Computer generated visual evidence: does daubert make a difference? *Georgia State University Law Reviews* **12**, 577.

Powers J.K. (2004). *Computer-Animated Accident Reconstruction*. ATLA Annual Convention Reference Materials 2004, pp. 1211–1218.

Sileo, C. (2006). Ruling on computer evidence animates Pennsylvania high court trial, *Trial* **8**, 74.

Spencer E.L. & Krumholz D.J. (2007). *Use and Misuse of Technical Data: Telling the Scientific Story to Scientific Virgins*. American Law Institute-American Bar Association Course of Study June 2007, pp. 131–138.

Webster, C. (2005). *Animation: The Mechanics of Motion*, Elsevier, Oxford.

Related Articles

Demonstrative Evidence

Expert Opinion in Court: a Comparison of Approaches

Expert Opinion: United States

Reconstruction: Accident

Reconstruction: Three Dimensional

LEEANNE FRAZIER

Computerized Fingerprint Storage and Retrieval (AFIS) *see* Automated Fingerprint Identification System

Computers

Introduction

Computer forensics is a fast growing area. In businesses and in law enforcement, the number of people involved is growing rapidly. Computer forensics can be described as "the use of specialized techniques for recovery, authentication, and analysis of electronic data with a view to present evidence in a court of law".[a] Within law enforcement, some of the objectives are to find information that might contribute to an insight into what has happened in an investigation, and who was responsible. Important issues are that the data on the seized computers are not altered and that a clear chain of evidence is documented.

Traditionally, a distinction is made between computer forensics (hard drives), mobile phone forensics, and video. Nowadays, mobile phones and personal digital assistants like the IPAQ or Blackberry are like computers with a hard drive and therefore are dealt with in mostly the same way. Standard software mostly used within law enforcement for these products are "Encase", "Ultimate Toolkit", XRY, and Tulp 2G (developed at the Netherlands Forensic Institute).

Digital video is still a different field because, although they are basically like computers with a hard drive, the problem with proprietary systems is so great that different techniques need to be used.

Multimedia

The popularity and ubiquity of mobile devices is immense and growing in an exponential way. There are systems and capabilities that would have been more akin to a computer in the past. Hand-palm computers and phones but also MP3-players, iPods, PS3, and even navigation systems are used for multimedia storage. The possibility to connect anything with everything and going online adds to three trends:

1. Data storage on an ever-growing number of different digital devices;
2. Growing volume of storage (especially due to the increase in photo and video shooting); and
3. Online storage is a network attached storage (NAS) where computers can access storage at file level over a local area network (LAN), a wide area network (WAN), or the Internet.

Storage is of two main forms: primary storage (local random-access disk storage and secondary storage (offline, sequential-access storage, e.g., magnetic tape).

There is a division to be made on the different types of storage media. Mostly, they are divided into three categories: magnetic, optical, and electronic, with each a total different technology.

1. Magnetic storage media
   - magnetic tape
   - floppy disks, Zip drive, JAZ drive, etc.
   - hard disk
2. Optical media
   - CD-R
   - CD-RW
   - DVD
   - DVD-RW
   - DVD+RW
   - DVD-RAM
   - Blu-Ray
3. Electronic storage media, which is used in digital cameras – video cameras and various types of MP3 players and portable gaming devices. With some of these devices "card readers" may be obtained as an accessory
   - Compact Flash (CF)
   - Ultra CF
   - SmartMedia
   - Memory Stick
   - MultiMediaCard (MMC)
   - xD Picture Card
   - PocketZIP
   - Microdrive
   - USB thumb drives
   - SIM cards (in mobile phones)
   - Portable navigation systems (e.g., TomTom wherein the newer systems have 20 GB hard drive mounted with 12 GB of the available memory).

There are also specialist gaming devices, wherein memory sticks are used as extendible memory, PSP, Playstation 3 and newer models of the Sony Ericsson mobile phones.

Hard Disks

For the majority of systems, tools such as FTK (http://www.accessdata.com/) and Encase (http://www.

guidancesoftware.com/) or the Linux variants of dd, for example, rdd, http://sourceforge.net/projects/rdd provide the necessary means to image and analyze hard drives. Special care has to be taken with redundant array of independent disks (RAID) systems, the technology where large amounts of DATA are being stored and which need to be secure and quickly accessible. Each array is seen by the computer system as a single hard disk even if there may be 2, 3, 4 or more physical hard disks in the system. RAID systems mostly use level 0–5 or a combination of them and mostly use SCSI disks as a base but there are controllers which use IDE/SATA. Special care has to be taken with acquiring and analyzing RAID systems, for example, drive order, block size, and direction of rotation. With networks, the aim here is to capture information (data) that moves over a network and make sense in a forensic way. A much-used forensic tool here is that from EnCase Enterprise, among other tools, which do in-depth analysis and network monitoring and detecting tools.

Systems such as Apple are taking an increasing share of the market due to popularity of their products, e.g., Ipod, Imac, Macbook (Pro) and the newly arrived iPhone. More Apple computers are coming to homes and businesses with a lot of them replacing Unix desktops and with the possibility for running multiple operating systems, it is getting increasingly popular and used more and more as a platform. This means that more such systems are involved in forensic examinations.

Securing Data from a Macintosh Computer

As most documentation has been written about Windows machines, this document concentrates around the newest technologies concerning the Macintosh and will give a brief explanation of the use of FireWire Target Disk Mode (TDM). TDM is a technology, which enables Macintosh computers to act as an external, Firewire hard disk. The advantage is that a Macintosh research computer could investigate and secure the data of another Macintosh computer in a forensically safe way.

Investigators should be very aware that the data is probably written onto the system under investigation when it includes installed FAT or NTFS partitions. These partitions are installed with BootCamp, for example, which is a standard tool of Mac OS X. This is caused by the FireWire connection, which is not write blocked in any way.

It should be noted that EnCase (versions 6.X) includes a write-blocker option within the Tools menu, but experts (e.g., the New York Forensic Department of the New York State Police and the FBI) have not (scientifically) endorsed that tool yet. Until now, there are no so-called "White Papers publications" on this method of write blocking. It is obvious that trusting on changing registry settings is venturing on thin ice. For this reason, it is *not* recommended to use TDM in combination with a Windows research computer.

A useful tool to secure a Macintosh computer is to use a LiveCD facility. There are various CDs available of which the EnCase Linen CD is a good example. For example, with this CD a Macintosh computer can be secured (in server mode) via a network connection (cross cable). Another Unix-oriented boot CD is Ubuntu, which makes copy of the hard disk via DD (*disk dump*, a Unix program to copy and convert files.)

It is very important when booting a system with a boot CD (to avoid writing actions on the Macintosh computer) that the computer is switched on with the CD placed in it (or insert it simultaneously during computer start-up) with the "option" key (i.e. Alt key) pressed. Within a few seconds, the Mac will show a list of bootable devices. You will then select your boot CD and start DD (a command line utility) to make a copy of the system that is under examination. Mac forensic users could also use, among others, the Macquisition Boot DVD. When used properly and observing the procedures correctly, the Macintosh research computer will not access the filesystem or other available data.

For a digital forensic examiner with a Macintosh research computer, TDM is an extremely useful tool. Its procedure is as follows.

Check whether the Mac system under investigation is switched off. When it concerns an Apple laptop, connect the AC power supply.
Boot the system while pressing the "Option" key. This may have the following outcome: a list of bootable devices or you will see a prompt for the FirmWare password. In the latter case, it is impossible to use TDM.

Use a Fire Wire cable to connect the research computer with the system under investigation. If it concerns a Mac, it is not necessary to switch off your own research system.
Switch on the system under investigation and immediately press the "T" key and keep it pressed until the FireWire icon appears. The hard disk of the system under investigation will now be available on your research computer (preferably a Mac) and will appear by using the Explorer (PC) or it will appear on your desktop (Mac).
When you have finished your investigation you will drag the icon of the system under investigation from your research Mac to the waste paper basket, select "put away" when Mac OS9 is concerned or "eject" from the file menu under Mac OS X.
Switch off the system under investigation by pressing the power switch.
Remove the Fire Wire cable from the systems.

Important Notes

On the Mac research computer->Disk Arbitration should be switched off on the research system. Disk Arbitration is a daemon (background process) in OS X, which mounts filesystems automatically. For example, mounting your USB memory stick and after this it will appear on your desktop.

Disk Arbitration will mount the available volumes in a read/write mode, which is not desirable in a forensic investigation. This could be avoided by switching off Disk Arbitration on the Mac.

Another important addition (made by Apple) further explains as follows:

Fire Wire Target Disk Mode functions on internal ATA drives only. TDM only connects with the master ATA drive on the Ultra ATA bus. TDM does not connect with a Slave ATA, ATAPI, or SCSI drives available on the system. This means that it is impossible to connect to a system in which more drives have been installed of the aforementioned types. If it is known that there are two or more hard disks in the system, it is probably better to use the Live CD method to switch on and acquire the system (read the part on the EnCase LinEn CD and Macquisition Boot DVD).

Note BootCamp:
If a system is switched on by pressing the "Option" key and a drive appears with "Windows" on it, then BootCamp has been installed.

Note EnCase Unallocated clusters:
If Unallocated clusters are only seen after acquiring a Mac hard disk with the help of EnCase, then "H+" within these clusters should be sought. "H+" is at the start of a sector and marks the beginning of a partition. After this, by going back two positions in EnCase this can be marked as a partition and EnCase will now display the data/volumes.

Note EnCase partitions:
EnCase will often show partitions without any information. These are the so-called driver partitions.

During the investigation of a system, it is very important to test the research tools and validate the results with other research tools before evidence is brought to court.

Phones and PDAs

XRY, TuLP2G, Neutrino, PDA analyzer + PDA Express, and Paraben Device Seizure are examples of tools used for examining phones and PDA's.

PDA's and mobile phones nowadays have more and more the same functionality and have added the possibility to be used as Mp3 players, camera's, Internet (WiFi) connectivity and also as gaming platforms. Nokia is re-releasing the N-Gage phone as a gaming platform and Sony Ericsson has plans to release a gaming phone not to mention the recently added iPhone and new generation iPod touch by Apple.

Furthermore, we have to be aware that the same way laptop data is being protected by antivirus, firewall, and encryption software the same applies to PDA's and smart phones.

Dealing with encrypted PDAs and smart phone devices, it is best practice not to switch them off as this might result in the device not being accessible next time it is turned on. The accessible data of the device should be copied before it is switched off and therefore available to the forensic software.

The iPod, for instance, has been overlooked previously as a device where evidence could be extracted from. However, it has become clear that it can store vast amounts of data, has PDA like functions and it can be hacked and customized so it will be increasingly used in crimes.

One of the tools to analyze the iPod is EnCase by Guidance Software. Apart from being recognized by law enforcement and the courts, it does a good

job at recovering data (even deleted) from the iPod even if it has been (re-) initialized. Most of the iPod investigation can be done on a PC but when required a Mac can be used. Write blocking is in order when analyzing the iPod. It is best that for write blocking a hardware write blocker is used. When using the USB write blocking functionality in Windows XP SP2 investigators should be aware of the fact that it will allow write access to the device when it does not recognize the file system.

It is also recommended that during the investigation that the hash value[b] (a digital fingerprint of a file) is verified at various stages to ensure that the contents have not changed.

Also hard drives which contain read errors or which have been damaged are examined at forensic laboratories. When repairing hard drives one can chose to do that a specialized company or obtaining the knowledge of reparation yourself. One should have training, and it is advisable to have a collection of hard drives, in order to exchange, for example, a damaged controller.

Video

Due to the increase of digital closed circuit television (CCTV) systems in public areas, this area of work has added a complete new discipline to the forensic field workers. Very often the systems contain proprietary software, which makes it impossible to view the images with standard tools. In the worst case scenario, the hard drive with the images only work in the original machine. Also, the systems do not always have an input/output interface for making a right connection.

When dealing with CCTV systems, it is very likely that several cameras will be connected to it. Therefore, special forensic research computers with special software are needed. Examples of widely used software are IMPRESS from IMIX and and Video Investigator from Cognitech, which are forensic suites of tools that enables demultiplexing, frame averaging, duplication, video level adjustment, magnification, highlighting, and obscuring of multiple subjects and specific areas.

Challenges Ahead

- Addressing the fast growing complexity and diversity
  Different techniques and tools are needed to collect all the evidence from a device.
  An example of the wide range of knowledge demanded from forensic investigators is pointed out in the "iPod Forensic Update" [1]. In the default mode, the iPod can accept audio-, video-, and photo-based files. In the disk mode, however, it acts as a hard drive to which any type of file can be transferred through Windows Explorer, these files are not visible on the iPod while browsing it. Another example is the use of Encase on Apple computers, although a good thorough tool, there are some items it does not show.
- Storage of data online as a trend in society.
  Especially, in the child porn business both the use of encryption and central data storage (on the internet) is a growing problem. Different legislation in different countries severely restricts the investigation.
- Storage of the increasing amount of collected multimedia evidence in the law enforcement directorate itself. For this a solution has not yet been found.
- Proprietary software this is especially a problem in the camera industry, video is often stored in a proprietary software format on hardware sometime does not even have an output channel. To collect this kind of evidence means often making a copy with data loss and therefore not having an exact image. By well documenting the followed actions, the evidence is still acceptable in court;
- Vulnerability of tools (they are not well protected against viruses and Trojans) encryption is an increasing problem, as it is ever more widely used. If encryption is implemented not in the right way, it sometimes can be broken easily. However, if strong encryption is used, it will take much more effort and computing power to break the encryption. Depending on the passwords and the passphrases that have been used, one can do password guessing with dictionary attacks or try to receive the keys from the suspect.

The advantage in forensics examinations of an encrypted drive/device is that when the encrypted partition is in use special forensic tools give you the ability to obtain data and possible valuable information when the encrypted partition is unlocked. This is called a *live investigation*, which as a rule should be

done with the utmost care and procedures. Several best practice manuals exist, for example, from the Working group forensic IT from the European Network of Forensic Science Institutes, www.enfsi.eu. The live investigation can also be helpful to do a search for child pornography images at different systems without having to image the complete system, and realizing a faster response rate.

Conclusion

The challenge ahead is that the storage space grows very rapidly, and people distribute their information (encrypted) on the Internet. Furthermore, malicious software also influences the evidence, since in practice it is possible that this software is used in a distributed network without the person possessing the PC being aware of it. Also storage devices get smaller each year, and for having all digital evidence it is also important to look for other digital devices. Also, forensic laboratories will have to cope with an ever-growing storage space to several thousands of terabytes in a few years, with appropriate backup facilities.

End Notes

[a.] http://www.forensicfocus.com/

[b.] www.wikipedia.org/ A hash function takes a long string (or "message") of any length as input and produces a fixed length string as output, sometimes termed a *message digest* or a *digital fingerprint*. A hash value (also called a "digest" or a "checksum") is a kind of "signature" for a stream of data that represents the contents. One analogy that explains the role of the hash function would be the "tamper-evident" seals used on a software package.

Reference

[1] Kiley, M., Shinbara, T. & Rogers, M. (2007). iPod forensics update, *International Journal of Digital Evidence* **6**(1), Spring.

Related Articles

Image Processing and Analysis

K. van Waalwijk van Doorn, M. den Dunnen and Zeno Geradts

Confession *see* Interrogative Suggestibility

Confessions: Evidentiary Reliability of

Introduction

A "confession" is any written or oral statement in which a person admits to having committed some transgression, often indicating acknowledgement of guilt for a crime. In some settings, confessions are considered necessary for absolution, social acceptance, freedom, or physical and mental health, making it easy to understand why people often exhibit an "urge to confess". In other settings, however, confessions predictably result in personally damaging consequences to the confessor – such as a loss of money, liberty, or life itself – making it difficult to understand this aspect of human behavior.

In criminal law, confession evidence is considered a highly potent weapon of prosecution, suggesting that "the introduction of a confession makes the other aspects of a trial in court superfluous" [1]. On the one hand, confessions play a vital role in law enforcement and crime control. On the other hand, they are fallible and serve as a source of recurring controversy, with questions often arising about whether a statement is authentic, voluntarily given, the product of a competent waiver of rights, and in accord with the law. For all of these reasons, confessions have been described as "troubling" [2].

Contradicting the pervasive myth that people do not confess to crimes they did not commit are numerous instances throughout American history, beginning with the Salem witch trials of 1692, in which men and women were wrongfully prosecuted, convicted, imprisoned, and sometimes sentenced to death, because of false confessions. The prevalence rate is unknown, but analyses of wrongful convictions reveal that between 14 and 25% of prisoners exonerated by DNA or other means had confessed to police. Importantly, it is clear that these discovered instances represent the tip of a larger iceberg [3–5].

After reviewing a number of past cases, and drawing on psychological theories of social influence, Kassin and Wrightsman, in 1985, proposed a taxonomy that distinguished three types of false confessions: voluntary, compliant, and internalized [6]. Still used today, this classification scheme has provided a useful framework and has since been used, critiqued, extended, and refined in subtle ways [7].

"Voluntary false confessions" are those in which people claim responsibility for crimes they did not commit without prompting or pressure from police. Often this occurs in high profile cases. When Charles Lindbergh's baby was kidnapped in 1932, 200 people stepped forward to confess. When "Black Dahlia" actress Elizabeth Short was murdered and her body sawed in half in 1947, more than 50 men and women confessed. More recently, in 2006, John Mark Karr volunteered a confession, replete with details, to the unsolved murder of young JonBenet Ramsey. There are many reasons why innocent people might voluntarily confess – such as a pathological need for attention, notoriety, or self-punishment; residual feelings of guilt; delusions of their own involvement; the perception of tangible gain, real or imagined; or the desire to protect a parent, child, or someone else.

In contrast to voluntary false confessions are those in which suspects are induced to confess through the processes of police interrogation. In "compliant false confessions", the suspect acquiesces to the demand for a confession in order to escape from a stressful situation, avoid an expected punishment, or gain a promised or implied reward. Akin to the classic forms of influence observed in social psychological studies of conformity, compliance, and obedience to authority, this type of confession is a mere act of capitulation by a suspect who knows that he or she is innocent. Psychologically, these confessions are easy to understand, as they arise whenever a suspect comes to believe that the short-term benefits of confessing – such as being left alone, fed, or released – outweigh the long-term costs associated with prosecution, the loss of reputation, and incarceration.

In the extreme, this type of false confession was seen in *Brown v. Mississippi* (1936), a case in which three Black tenant farmers confessed to murder after they were whipped with a steel-studded leather belt [8]. More recently, it was multiply illustrated in the 1989 Central Park jogger case, where five New York City teenagers confessed after lengthy interrogations, in vivid detail, each claiming he expected to go home afterward. On the basis of these confessions, the boys were convicted and sent to prison, only to be exonerated 13 years later when the real rapist gave a confession that was confirmed by the original DNA evidence [9].

"Internalized false confessions" are those in which innocent but vulnerable suspects, exposed to highly suggestive interrogation tactics, come not only to capitulate in their decision to confess but also to believe that they were culpable, sometimes confabulating false memories in the process. Gudjonsson and MacKeith argued that this effect stems from "memory distrust syndrome", whereby people come to distrust their memory, which renders them vulnerable to misinformation from external sources [10]. This phenomenon is closely related to psychological studies involving the creation of false memories for words in a list, visual images, and past experiences.

The case of 14-year-old Michael Crowe, whose sister Stephanie was stabbed to death in her bedroom, illustrates the problem. Michael had adamantly asserted his innocence, but police targeted him for suspicion. After a series of interrogation sessions, during which time Michael was deceived into thinking there was substantial physical evidence of guilt (e.g., that his hair was found in Stephanie's grasp), he concluded that he was a killer, saying: "I'm not sure how I did it. All I know is I did it." Eventually, he was convinced that he had a split personality – that "bad Michael" acted out of a jealous rage while "good Michael" blocked the incident from consciousness. The charges against Crowe were later dropped when a drifter in the neighborhood that night was found with Stephanie's blood on his clothing [11].

Recent research on the causes and consequences of police-induced confessions has focused on three processes: (i) the preinterrogation interview during which police judge people to be truthful or deceptive, (ii) the interrogation processes that lead certain suspects under certain conditions to confess, and (iii) the influence of confession evidence on judges, juries, and other decision makers.

The Preinterrogation Interview

Typically, the confrontational process of 'interrogation' is preceded by an information-gathering 'interview' designed by police to determine if a suspect is guilty or innocent. Sometimes, that initial judgment is

reasonably based on witnesses, informants, or other extrinsic evidence. At other times, it is based on a clinical impression formed during this interview. In "Criminal Interrogations and Confessions", first published in 1962 and still the most influential manual on interrogation, Inbau, Reid, Buckley, and Jayne (2001) advise police in the use of "behavior provoking questions" and observation of verbal cues, nonverbal cues, and behavioral attitudes to detect deception – they claim, at high levels of accuracy [12].

For people who stand falsely accused, this first impression often determines whether they are interrogated or sent home. Yet in laboratories all over the world, research has shown that people are only about 54% accurate in judging truth and deception; that training produces little, if any, improvement compared to naive control groups; and that police, customs inspectors, judges, psychiatrists, and other so-called experts perform only slightly better, if at all [13, 14].

Assessing performance in a forensic context, researchers randomly trained some lay participants but not others in a popular law enforcement method of truth and lie detection. These students then watched videotaped interviews of mock suspects, some guilty and others innocent, all of whom denied their involvement in various mock crimes. As in nonforensic studies, observers were generally unable to differentiate between guilty and innocent suspects. Moreover, those who underwent training were less accurate, more confident, and more biased toward seeing deception [15]. Using the same tapes, other researchers found that experienced samples of police detectives exhibited these same tendencies – a finding consistent with studies suggesting that current law enforcement training may lead police to make prejudgments of guilt, with confidence, that are frequently in error [16].

Recent studies suggest that human lie detection performance can be improved. In one research program, Hartwig and colleagues found that interviewers can increase their accuracy by withholding crime details while questioning suspects, which traps those who are guilty but not innocent in inconsistencies when these facts are disclosed [17]. In a second research program, Vrij and colleagues note that because lying is more effortful than telling the truth, interviewers who tax a suspect's cognitive load (for example, by asking challenging questions or reversing their order) can make more accurate true–false judgments by attending to effort cues – such as response hesitations or the use of simple language [18].

The Processes of Interrogation

To protect citizens against violations of their constitutional rights and to guard against wrongful convictions, American courts have set guidelines by which confessions are admitted into evidence at trial. For judges to consider on a case-by-case basis, in the context of a "totality of the circumstances", confessions are admitted if deemed voluntary but excluded if coerced.

Many years ago, interrogators used bright lights, brute force, the rubber hose, and physical intimidation to get confessions. Today, however, the police are required to warn suspects of their "Miranda" rights to silence and to an attorney and are legally prohibited from drawing confessions through violence, physical discomfort, threats of harm or punishment, or promises of leniency or immunity. Hence, the use of physical force has given way to more psychologically oriented techniques.

Observational studies have confirmed that modern police interrogation is psychologically oriented, often involving feigned sympathy and friendship, appeals to conscience and religion, confrontations with evidence, the alternation of good cop and bad cop, and other forms of trickery and deception [19]. In a recent survey, 631 police investigators reported that the tactics most frequently used were to physically isolate the suspect from family and friends, typically in a small private room; identifying contradictions in the suspect's account; trying to establish rapport in order to gain the suspect's trust; confronting the suspect with evidence of his or her guilt; and appealing to his or her self-interests [20].

Today, a number of manuals are available to train police in how to elicit confessions through interrogation. The text "Criminal Interrogation and Confessions", cited earlier, teaches the popular Reid technique (named for its founder, John Reid), a nine-step interrogation program that is essentially reducible to an interplay of three processes [21]: "*isolation*", which involves placing the suspect into a small, barely furnished, soundproofed room to increase stress, discomfort, and the suspect's desire to escape; "*confrontation*", in which the interrogator

strongly accuses the suspect of the crime, sometimes citing real or fictitious evidence to bolster the claim; and "*minimization*", in which a sympathetic interrogator morally justifies or excuses the crime, leading the suspect to infer that he or she will be treated with leniency upon confession. Once these processes elicit a suspect's admission of guilt, investigators are trained to convert that simple admission into a full and detailed narrative confession.

False Confessions

As prescribed, police interrogations should lead guilty suspects to incriminate themselves by reducing the perceived negative consequences of confessing while, at the same time, increasing the anxiety associated with continued deception. However, actual cases as well as psychological research indicate that this approach may induce innocent people to confess. As for risk factors, some suspects are particularly vulnerable to manipulation; some tactics are particularly coercive.

Dispositional Vulnerabilities

Some suspects are dispositionally more vulnerable to influence than others – and at greater risk for false confessions. Focusing on personality traits, Gudjonsson [22] has found that individuals who are prone to "compliance" in social situations are especially vulnerable because of their eagerness to please others and a desire to avoid confrontation, particularly with those in authority. Individuals who are prone to "suggestibility" – whose memories can be altered by misleading questions and negative feedback – are also more likely to confess. Most importantly, Gudjonsson notes that people who are highly anxious, fearful, depressed, delusional, or otherwise psychologically disordered are often at a heightened risk to confess under pressure.

Youth is a substantial risk factor. Statistics show that more than 90% of juveniles whom police seek to question waive their "Miranda" rights to silence and a lawyer [23]. In fact, the presence of a parent, guardian, or other "interested adult" – which many states require to protect young suspects – does not lower this waiver rate, as adults often urge their youths to cooperate with police [24]. This problem can be seen in the disproportionate number of juveniles appearing in the population of false confessors. Looking at 125 proven false confessions, for example, Drizin and Leo found that 33% involved juveniles [3].

That juveniles are vulnerable in this regard makes sense. Developmental psychology research shows clearly that adolescents are not only more compliant and suggestible than adults but that their decision-making is characterized by an "immaturity of judgment" – a pattern of behavior that is impulsive, focused on the present, and diminished in the capacity to perceive risk [25]. To the adolescent not fully focused on long-term consequences, confession may thus serve as an expedient way out of a stressful situation.

People who are intellectually impaired are also vulnerable in the interrogation room. Indeed, at least 22% of exonerees in a recent sample of false confessions were mentally retarded, as measured by conventional IQtests [3]. This result is not terribly surprising. Most people who are mentally retarded cannot fully comprehend their rights, leading some researchers to describe the "Miranda" warnings they receive as "words without meaning" [26]. People who are mentally retarded also exhibit an acquiescence response bias that leads them to say "yes" to a whole range of questions – even when an affirmative response is incorrect, inappropriate, or absurd [27]. They are also suggestible, as measured by the degree to which they are influenced by leading and misleading questions. Gudjonsson and others have found that people who are mentally retarded score higher than average on the Gudjonsson Suggestibility Scale, a commonly used test of interrogative suggestibility [22].

Finally it is important to note the paradoxical possibility that "innocence" itself may put "innocents" at risk. According to Kassin [21], anecdotal and research evidence indicate that people who stand falsely accused tend to believe that truth and justice will prevail and that their innocence will become transparent to investigators, juries, and others. As a result, they cooperate fully with police, often not realizing that they are suspects, by waiving their "Miranda" rights and speaking freely to defend themselves.

The problem is poignantly illustrated by the case of Jeffrey Deskovic, exonerated in 2006 after fifteen years in prison for a murder he confessed to while

in high school but did not commit. "Believing in the criminal justice system and being fearful for myself, I told them what they wanted to hear," Deskovic said. Certain that DNA testing on the semen would establish his innocence, he added that he confessed because "I thought it was all going to be okay in the end" [28].

Interrogation Tactics

It appears that certain aspects of interrogation, even if legal, put lead innocent people to confess. To begin with, people in general become vulnerable to influence if isolated for extended periods of time, becoming fatigued or sleep-deprived – states that induce a focus on short-term consequences in decision-making [29, 30]. Thus it comes as little surprise that whereas most police interrogations last for less than two hours, the average length of interrogation in false confession cases is 16.3 h [3].

One potentially problematic tactic, present in the vast majority of false confession cases, is the false evidence ploy. Following a US Supreme Court ruling in 1969, American police are permitted to bolster their accusations by telling suspects that there is incontrovertible evidence of their guilt (e.g., a fingerprint, blood or hair sample, eyewitness identification, or failed lie-detector test) even if such evidence does not exist [31].

Basic psychology research, in nonforensic contexts, has shown that presentations of false information – via confederates, witnesses, counterfeit test results, bogus norms, false physiological feedback, and the like – can alter people's visual judgments, beliefs, behaviors, emotions, physical attraction, self-assessments, perceptions of other people, memories for observed and experienced events, and even certain medical outcomes, as seen in studies of the placebo effect. Human malleability to influence in the face of misinformation is broad and pervasive.

In the laboratory experiment, Kassin and Kiechel [32] tested the hypothesis that false evidence would lead innocent people to confess and possibly to internalize blame for outcomes they did not produce. Subjects typed letters on a keyboard in what was supposed to be a reaction time study. At one point, they were accused of causing the experimenter's computer to crash by pressing a key they were instructed to avoid. They were asked to sign a confession. All subjects were truly innocent and all initially denied the charge. In some sessions but not others, a confederate reported that she witnessed the subject hit the forbidden key. This false evidence nearly doubled the number of students who signed a written confession. As measured moments later, it also increased the number of subjects who believed they were responsible for this outcome. Follow-up studies using this computer-crash paradigm have replicated this effect even when highly negative consequences followed from the confession [33, 34].

Another potentially problematic tactic is minimization, the process by which interrogators offer moral justification or face-saving excuses, minimizing the crime and making confession seem like a cost-effective solution. Interrogators are thus trained to suggest to suspects that their actions were spontaneous, accidental, provoked, peer pressured, or otherwise justifiable by external factors. Minimization tactics lead readers of an interrogation transcript to infer that leniency will follow from confession, even in the absence of an explicit promise [35]. In a laboratory study in which college students were accused of cheating, minimizing remarks increased the rate of true confessions from actual cheaters and false confessions from noncheaters – just as effectively as did an explicit offer of leniency [36].

Confession Evidence in the Courtroom

When a suspect retracts a confession, pleads not guilty, and goes to trial, a sequence of two decisions is set into motion. First, a judge determines whether the confession was voluntary and hence admissible as evidence. Then a jury, hearing the admissible confession, determines whether the defendant is guilty beyond a reasonable doubt. But can people distinguish between true and false confessions, and what effect does this evidence have within the context of a trial?

Addressing the first question, researchers interviewed male prison inmates on videotape providing true confessions to the crimes for which they were incarcerated and concocting false confessions to crimes selected by the experimenter that they did not commit. When college students and police investigators later judged these statements from videotapes or audiotapes, the results showed that neither group

was particularly adept, exhibiting accuracy rates that ranged from 42 to 64% [37].

Mock jury studies have further shown that confession is a highly potent form of evidence and that people do not fully discount confessions even when coerced. To illustrate, researchers presented mock jurors with one of three versions of a murder trial – a low-pressure confession version, a high-pressure confession version, and a no-confession control version. Confronted with the high-pressure confession, subjects judged the statement to be involuntary and said it did not influence their decisions. Yet when it came to verdicts, this same confession significantly boosted the conviction rate – even when subjects were specifically asked to disregard confessions they thought were coerced [38].

Criminal justice statistics reinforce the point that confessions unleash a chain of adverse legal consequences – from arrest through prosecution, conviction, and prison. Archival analyses reveal that roughly 80% of innocent confessors who pled not guilty and went to trial were convicted by juries [3]. This figure led Drizin and Leo to describe confessions as "inherently prejudicial and highly damaging to a defendant, even if it is the product of coercive interrogation, even if it is supported by no other evidence, and even if it is ultimately proven false beyond any reasonable doubt" (p. 959).

A jury's reaction to confession evidence also depends on how that evidence is presented. Today, it is mandatory in Great Britain and common in many American police departments to videotape confessions for presentation in court. This policy brings many possible advantages – such as deterring police from using overly aggressive tactics, preventing voluntary confessors from later claiming they were coerced, and providing an objective and accurate record of the whole process, a common source of dispute at trial.

At present, researchers are weighing in on the "how to videotape" question, focusing on how the interrogation is staged for the camera. Lassiter and colleagues reviewed a series of studies in which they taped mock confessions from three different camera angles so that either the suspect or the interrogator or both were visible. All subjects – including mock jurors – heard the same exchanges of words, but those who focused on the suspect saw the situation as less coercive – and the suspect as more guilty – than did those focused on the interrogator [39]. Recently, these investigators found that juries and judges also appreciate the situational factors that produce confessions more when the interrogator is visible on camera than when the focus is solely on the suspect [40, 41]. Hence, it is important not only that entire sessions be recorded but that the camera adopt a neutral "equal focus" perspective that shows both the accused and interrogators.

Psychologists as Confession Experts

The problems raised by confession evidence are fundamental to criminal justice, substantial, and important. Thus, there is a great deal about this aspect of the legal system that warrants the intervention of psychologists for educational and research purposes. Intervention in this regard may take several forms, including but not limited to in-court testimony from "confession experts". In recent years, a growing number of psychologists have testified, often on behalf of criminal defendants who confessed but then retracted their statements and went to trial.

Typically, these experts testify in general terms about the social influence processes of compliance, persuasion, and obedience to authority; suggestibility effects in memory; personality and psychopathology; and other relevant phenomena; in other cases, they offer opinions concerning a specific defendant based on interviews and test results. As both the law and empirical foundations for such expert testimony evolve, it remains to be seen what impact psychological research on confessions will have on police, judges, juries, defendants, and others in criminal justice.

References

[1] McCormick, C.T. (1972). *Handbook of the Law of Evidence*, 2nd Edition, West, St. Paul, p. 316.

[2] Brooks, P. (2000). *Troubling Confessions*, University of Chicago Press, Chicago.

[3] Drizin, S.A. & Leo, R.A. (2004). The problem of false confessions in the post-DNA world, *North Carolina Law Review* **82**, 891–1007.

[4] Gross, S.R., Jacoby, K., Matheson, D.J., Montgomery, N. & Patel, S. (2005). Exonerations in the United States, 1989 through 2003, *The Journal of Criminal Law and Criminology* **95**, 523–553.

[5] Scheck, B., Neufeld, P. & Dwyer, J. (2000). *Actual Innocence*, Doubleday, Garden City.

[6] Kassin, S.M. & Wrightsman, L.S. (1985). Confession evidence, in *The Psychology of Evidence and Trial Procedure*, S. Kassin & L. Wrightsman, eds, Sage, Beverly Hills, pp. 67–94.
[7] Kassin, S.M. & Gudjonsson, G.H. (2004). The psychology of confession evidence: A review of the literature and issues, *Psychological Science in the Public Interest* **5**, 35–69.
[8] Brown v. Mississippi, 297 U.S. 278 (1936).
[9] Saulny, S. (2002). Convictions and charges voided in '89 central park jogger attack, *The New York Times*, Section A, 1.
[10] Gudjonsson, G.H. & MacKeith, J.A.C. (1982). False confessions: psychological effects of interrogation, in *Reconstructing the Past: The Role of Psychologists in Criminal Trials*, A. Trankell, ed, Kluwer, Deventer, pp. 253–269.
[11] Drizin, S.A. & Colgan, B.A. (2004). Tales from the juvenile confessions front, in *Interrogations, Confessions, and Entrapment*, G.D. Lassiter, ed, Kluwer Academic, New York, pp. 127–162.
[12] Inbau, F.E., Reid, J.E., Buckley, J.P. & Jayne, B.C. (2001). *Criminal Interrogation and Confessions*, 4th Edition, Aspen, Gaithersberg.
[13] Bond, C.F. & DePaulo, B.M. (2006). Accuracy of deception judgments, *Personality and Social Psychology Review* **10**, 214–234.
[14] Vrij, A. (2008). *Detecting Lies and Deceit: Pitfalls and Opportunities*, Wiley, Chichester.
[15] Kassin, S.M. & Fong, C.T. (1999). "I'm innocent!" Effects of training on judgments of truth and deception in the interrogation room, *Law and Human Behavior* **23**, 499–516.
[16] Meissner, C.A. & Kassin, S.M. (2002). "He's guilty!": investigator bias in judgments of truth and deception, *Law and Human Behavior* **26**, 469–480.
[17] Hartwig, M., Granhag, P.A., Strömwall, L. & Vrij, A. (2005). Detecting deception via strategic closure of evidence, *Law and Human Behavior* **29**, 469–484.
[18] Vrij, A., Fisher, R., Mann, S. & Leal, S. (2006). Detecting deception by manipulating cognitive load, *Trends in Cognitive Sciences* **10**, 141–142.
[19] Leo, R.A. (1996). Inside the interrogation room, *The Journal of Criminal Law and Criminology* **86**, 266–303.
[20] Kassin, S.M., Leo, R.A., Meissner, C.A., Richman, K.D., Colwell, L.H., Leach, A.-M. & La Fon, D. (2007). Police interviewing and interrogation: a self-report survey of police practices and beliefs, *Law and Human Behavior* **31**, 381–400.
[21] Kassin, S.M. (2005). On the psychology of confessions: does innocence put innocents at risk? *The American Psychologist* **60**, 215–228.
[22] Gudjonsson, G.H. (2003). *The Science of Interrogations and Confessions: A Handbook*, Wiley, Chichester.
[23] Grisso, T. (1981). *Juveniles' Waiver of Rights: Legal and Psychological Competence*, Plenum Press, New York.
[24] Oberlander, L.B. & Goldstein, N.E. (2001). A review and update on the practice of evaluating Miranda comprehension, *Behavioral Sciences and the Law* **19**, 453–471.
[25] Owen-Kostelnik, J., Reppucci, N.D. & Meyer, J.R. (2006). Testimony and interrogation of minors: assumptions about maturity and morality, *The American Psychologist* **61**, 286–304.
[26] Cloud, M., Shepherd, G.B., Barkoff, A.N. & Shur, J.V. (2002). Words without meaning: the constitution, confessions, and mentally retarded suspects, *The University of Chicago Law Review* **69**, 495–624.
[27] Finlay, W. & Lyons, E. (2002). Acquiescence in interviews with people who have mental retardation, *Mental Retardation* **40**, 14–29.
[28] Santos, F. (2006). DNA evidence frees a man imprisoned for half his life, *The New York Times*, September 21, p. A1.
[29] Blagrove, M. (1996). Effects of length of sleep deprivation on interrogative suggestibility, *Journal of Experimental Psychology: Applied* **2**, 48–59.
[30] Harrison, Y. & Horne, J.A. (2000). The impact of sleep deprivation on decision making: a review, *Journal of Experimental Psychology: Applied* **6**, 236–249.
[31] Frazier v. Cupp, 394 U.S. 731 (1969).
[32] Kassin, S.M. & Kiechel, K.L. (1996). The social psychology of false confessions: Compliance, internalization, and confabulation, *Psychological Science* **7**, 125–128.
[33] Horselenberg, R., Merckelbach, H. & Josephs, S. (2003). Individual differences and false confessions: a conceptual replication of Kassin and Kiechel (1996), *Psychology, Crime and Law* **9**, 1–18.
[34] Redlich, A.D. & Goodman, G.S. (2003). Taking responsibility for an act not committed: the influence of age and suggestibility, *Law and Human Behavior* **27**, 141–156.
[35] Kassin, S.M. & McNall, K. (1991). Police interrogations and confessions: communicating promises and threats by pragmatic implication, *Law and Human Behavior* **15**, 233–251.
[36] Russano, M.B., Meissner, C.A., Narchet, F.M. & Kassin, S.M. (2005). Investigating true and false confessions within a novel experimental paradigm, *Psychological Science* **16**, 481–486.
[37] Kassin, S.M., Meissner, C.A. & Norwick, R.J. (2005). "I'd know a false confession if I saw one": a comparative study of college students and police investigators, *Law and Human Behavior* **29**, 211–227.
[38] Kassin, S.M. & Sukel, H. (1997). Coerced confessions and the jury: an experimental test of the "harmless error" rule, *Law and Human Behavior* **21**, 27–46.
[39] Lassiter, G.D., Geers, A.L., Munhall, P.J., Handley, I.M. & Beers, M.J. (2001). Videotaped confessions: is guilt in the eye of the camera? *Advances in Experimental Social Psychology* **33**, 189–254.
[40] Lassiter, G.D., Geers, A.L., Handley, I.M., Weiland, P.E. & Munhall, P.J. (2002). Videotaped confessions and interrogations: a change in camera perspective alters verdicts in simulated trials, *Journal of Applied Psychology* **87**, 867–874.

[41] Lassiter, G.D., Diamond, S.S., Schmidt, H.C. & Elek, J.K. (2007). Evaluating videotaped confessions: Expertise provides no defense against the camera-perspective effect, *Psychological Science* **18**, 224–226.

SAUL KASSIN

Confidentiality: Patient *see* Duty to Warn, Violence Risk Assessment for Mental Health Professionals

Confirmation Testing: Toxicology

Introduction

An axiom of the highest importance in forensic toxicology is the need to ensure that a reported result is reliable and cannot be successfully challenged in legal proceedings due to some weakness in the identification process.

The concept of confirmation testing is to use an alternative method based upon a different chemical principle than originally used to identify a substance, or group of substances. Confirmation is distinct to and should be more sensitive and specific than initial or screening tests covered elsewhere in this Encyclopedia.

What Techniques are Suitable?

Essentially all methods using some form of chromatographic separation are suitable, provided they are sufficiently different from the initial method and show very good to excellent discriminatory power.

Mass spectrometry (MS) is preferred, wherever possible, to chromatographic methods using a form of detection not capable of providing structural information. Most commonly these include "hyphenated" (one after the other) instrumental techniques in which MS is linked with gas chromatography (GC) or high performance liquid chromatography (HPLC, or often abbreviated as liquid chromatography LC) (Table 1).

Mass Spectrometry

This is generally regarded as the gold standard because of its ability to provide structural information in addition to retention time matching. Retention time matching means the coincidence, or almost coincidence, of chromatographic peaks obtained with an authentic standard and the compound being identified. The ability of a mass selective detector (MSD) to be discriminatory depends very much on the type i.e., single stage vs. tandem, the type of ionization (see below) and the compound(s) being detected.

MS is a technique in which a substance is subject to bombardment with high energy electrons in a high vacuum environment to create electrically charged particles (ions). Ionization of the substance may also be achieved by collision with other ionized small molecules. Initially the molecular ion is formed; however, the substance will often break down to smaller fragments, which provide clues as to the structural features of the original substance. The ions are separated by a mass analyzer according to their molecular weight-to-charge (m/z) ratios and each type of ion is counted (the response) by an ion detector, usually an electron multiplier. The response for all of the ions detected by the mass analyzer is compared with the response obtained from the most frequently encountered ion (the base ion), which is given a nominal response of 100%. A bar graph representation of the m/z of the detected ions against their relative abundances can then be drawn. This is the mass spectra of the substance and can be compared for identification purposes with the mass spectra of an authentic sample of the substance or an electronic mass spectral library.

When electrons are used to ionize substances and positive ions are counted, this is called *electron impact positive ion mass spectrometry* (EI-MS). When ionized small molecules such as methane (a reagent gas) are used to ionize the substance under question, it is called *chemical ionization mass spectrometry* (CI-MS). Negative ions can also be monitored and when this is used the technique is called *negative ion MS*.

Table 1 Confirmatory methods and relative discriminatory powers

Instrumental method	Relative discriminatory power[(a)]
GC-MS (full-scan EI)	Very high
GC-MS (SIM)	High
GC-MS/MS	Excellent
GC-MS (high resolution)	Very high–excellent
LC-MS (quadrupole)	Moderate–high
LC-MS/MS (and ion trap)	Excellent
GC (FID, ECD, NPD)	Moderate
HPLC (UV)	Low
HPLC (DAD)	Moderate

DAD, photodiode array detection; ECD, electron capture detection; EI, electron impact; FID, flame ionization detection; GC, gas chromatography; LC, liquid chromatography; MS, mass spectrometry;, MS/MS, tandem mass spectrometry; NPD, nitrogen phosphorus detection; SIM, selected ion monitoring; UV, ultraviolet detection

[(a)] These are relative and will also depend on compound being confirmed and application method

Some substances do not produce a mass spectra with much discriminating power e.g., the electron impact EI mass spectra of amphetamine and indeed a range of other substances which have poor mass spectra with "common" low-molecular-weight ions dominating the spectra. Under these conditions, MS is little better than a conventional detector, such as a flame ionization detector (FID) or a nitrogen phosphorus detector (NPD).

It is not possible to provide straightforward guidelines as to what constitutes a sufficiently adequate mass spectrum. Chemists and analytical toxicologists are trained to assess identification requirements. However, relatively common ions such as m/z 44, 58, 77, 91, 105 should not be used as the main identifying ions since these are found in many compounds and represent "simple" fragment ions. Rather, ions of higher molecular weight or preferably a number of ions of higher molecular weight are preferred. Drugs that have relatively simple MS include amphetamine, amitriptyline, and doxepin. These substances need to be chemically altered (derivatized) to improve MS characteristics or need to be detected by other techniques.

MS can be run under full scan conditions in which all ions from a low cutoff (often m/z 40 or 50) to the highest measurable mass (m/z 500 or more) are detected by the mass analyzer of the MS. When this occurs, any compound that is required to be identified has a "full" mass spectrum. This is preferred over "selected ion monitoring" or SIM, in which only selected ions known to be present in the suspected compound are measured. However, SIM is commonly used as a confirmatory method, and provided this is conducted properly and uses well-defined ions it can be suitably discriminatory and become a confirmation method.

Mass Spectral Matching Criteria

When performing full mass spectral matches, it is essential that the retention time (the chromatographic detection time) of the presumed substance matches that of an authentic standard. This means that an authentic standard must be run under the same conditions as the unknown sample in the same batch of analyses. All ions in the mass spectra need to be consistent with that of the authentic compound, i.e., there should be no missing ions or additional ions that do not form part of the background signals. The relative abundances of the ions should also match that of the authentic standard. A usual guide is that they should match within a relative abundance of $\pm 30\%$. Some laboratories will apply even tighter matching guidelines.

When running SIM an authentic standard must also be run in the same batch as the unknown(s). Again the retention time as demonstrated by the coincidence of the ions being monitored should match within a small allowance for measurement uncertainty and the ion ratios of the monitored ions should agree within $\pm 20\%$.

Methods to Enhance Spectral Information

When compounds provide poor mass spectral definition, they need to be run under different ionization conditions or be analyzed as a chemical derivative.

In GC other ionization conditions include chemical ionization (CI). This can be operated under positive and negative ion modes. This mode employs a reagent gas (R) in the ionization chamber (source), such as methane, isobutane, or ammonia. The reagent gas is ionized by the electrons, but since these molecules have a far higher mass than electrons, the collision energies with the compounds eluting from the GC column are far lower. This results in much less fragmentation of the target molecules and can often provide strong signals for the molecular ion, usually as $M + H^+$ or $M + R^+$ ions. As a consequence, CI is often called a *soft ionization technique*, and is required when using LC to prevent substantial ionization of the volatile buffers and solvents used. However, less structural information can be obtained than EI ionization and depending on the other techniques used it may not be sufficiently discriminatory to confirm the presence of a compound.

Alternatively, negative ions produced in the ionization process can be monitored. These will be from different structural parts of the molecule than the positive ions. Compounds with functional groups or moieties that capture electrons are most sensitive using this form of ionization, e.g., compounds containing halogen atoms or nitro groups.

The use of tandem mass spectrometry (MS/MS or MS^n) has revolutionized the discipline of MS and is now a major technique to confirm the presence of foreign substances such as drugs. In this technique, one or more ions derived from the initial ionization can be induced to fragment in a collision chamber adjacent to the mass analyzer and these further fragments characterized by a second mass analyzer. Consequently, derivative spectra are obtained. This type of MS is more discriminatory than first-generation mass spectra and is often also more sensitive and will allow ultralow detection limits to be obtained. The cost of the instrument is higher than a simple mass spectrometer, but the costs are affordable by major laboratories.

Tandem MS is particularly important when LC is used. LC has some inherent advantages over GC in that derivatization may not be required and thermal decomposition should not be a problem; however, liquid chromatography mass spectrometry (LC-MS) has been hampered by the necessity to use a soft ionization technique producing fewer ions with consequent less discriminatory power. With the advent of tandem MS, LC-MS/MS has become a very powerful technique with high discrimination. Ion trap mass spectrometers utilize a different process to generate ions compared to conventional quadrupole or magnetic sector mass spectrometers, but can also provide MS^n spectra. These are often at a lower price than tandem mass spectrometers.

Time-of-flight (TOF) MS, like other forms of high resolution MS, is another instrumental technique used in laboratories. This is more commonly used to screen samples for the presence of drugs and other compounds. It does, however, provide accurate molecular weights of ions that in association with another suitable technique can be used in confirmatory analysis.

Chromatography Tests

Traditionally, before the routine use of mass spectrometers, laboratories would use a chromatographic technique to confirm the presence of drugs based on an immunoassay. A colorimetric method or even an older chromatography technique such as thin layer chromatography (TLC) has also been used. Indeed in the developing world this practice is still being applied. Although the discriminatory process is less used than MS, a well-developed and validated chromatographic method can be suitable for those laboratories that do not have mass spectrometers.

Chromatography has various forms but is generally based on separating substances by virtue of a partitioning between two phases i.e., a stationary and a mobile or moving phase. In TLC separation of substances is caused by their relative solubility in solvents (moving phase by capillary action) and affinity for the support i.e., the stationary phase (often silica based). In HPLC this is the relative solubility of the substance between the moving liquid phase and a molecular layer (pseudo liquid phase) bound to a silica support. In GC this is a relative solubility between the moving gas phase (usually helium or nitrogen) and a micro-liquid layer coated (or even bonded) onto a silica support.

HPLC with ultraviolet (UV) or, better, photodiode array (DAD) or multiwavelength detection (MVD)

can be used to provide the second method of identification. DAD can produce some clue as to the identity of the eluting molecule provided the molecule has sufficiently distinctive UV spectra. The potential information obtained from DAD would be more useful than the simple retention time matching that would occur with a fixed wavelength UV detector. Published spectra are available for initial matching purposes. However, as for MS, direct comparison with a standard is preferred to confirm retention time matching and UV spectra under the conditions used. UV spectra of eluting substances are compared with library entries, usually electronically in a similar manner to that used for comparing MS, except that here it is the closeness of a UV spectral fit – the UV maxima and minima wavelengths and general similarity of spectra.

GC can be also used as the second method using detection systems such as FID, NPD, or electron capture detector (ECD). However, this technique is more commonly used as a screening method since it does not provide any structural information and relies exclusively on retention time matching. It is however possible to increase the discriminatory power of GC by splitting the injection into two separate columns with different stationary phases with their own detectors, usually both NPD. This arrangement produces two distinct sets of retention time data greatly facilitating identification. Laboratories have also been known to use this arrangement using an NPD and an MSD. The use of capillary columns allows better separation characteristics than HPLC since the resolving power of capillary columns is higher than that of columns used in HPLC.

Further details of how chromatography systems work and their relative advantages and disadvantages can be found in some of the reading materials.

Cutoffs

A term often used in forensic testing is cutoff. This refers to a concentration that is being applied as an effective reporting limit and is usually higher than the detection and quantitation limits of an analytical assay (see next section).

For example, it is often used in urine analysis for drugs of abuse to assist in the interpretation of the possible significance of a drug result. When testing for opiates such as morphine and codeine, a cutoff of $2000\,ng\,ml^{-1}$ will mean that most results of morphine over $2000\,ng\,ml^{-1}$ in the absence of substantial levels of codeine probably mean exposure to morphine or heroin (morphine is major metabolite of heroin). A cutoff of $15\,ng\ ml^{-1}$ of cannabis metabolite in urine will exclude most cases of passive exposure of cannabis smoke.

It should be stressed that applying cutoffs does not mean that an interpretation does not need to be performed since there may still be alternative explanations for a positive drug result than personal drug use. However, it does guide laboratories in using more relevant concentrations when performing analyses.

Extraction and Isolation Techniques

Analyses of tablets, powders, or residues usually do not require sophisticated sample isolation or purification step. Rather they involve the dilution of an accurate mass of material with a solvent such as an alcohol or simple ester (i.e., ethyl acetate) and direct injection into the analytical equipment such as an MS.

However, when dealing with biological samples, e.g., blood/plasma, urine, hair, and saliva, an isolation process needs to be employed to extract the analyte(s) from the matrix. This not only isolates the analyte(s) to facilitate some form of chromatography but also removes potential interferences from the matrix.

Isolation procedures can involve the use of solvents such as the extraction of codeine from blood using butyl chloride, or the extraction of the active substance in cannabis, Δ^9-tetrahydrocannabinol (THC) from blood using hexane. The pH of the blood (or other fluid) is adjusted prior to extraction to ensure the nonionized form of the substance is present. This assures a better recovery of the analyte(s) in the solvent. The solvent is then usually evaporated and the residue containing the drug reconstituted into a small volume of an injecting solvent. An aliquot of this extract is then injected into the chromatograph.

Other common isolation procedures include the use of solid-phase extraction (SPE) columns or solid-phase microextraction (SPME). SPE involves the use of commercial minicolumns or cartridges packed with a coated silica-based phase similar to that used

in HPLC. The sample is diluted into a buffer of correct pH and applied to the column/cartridge and the analyte(s) of interest are selectively absorbed onto the column. A wash-step removes the excess of unwanted material in the sample and a final step elutes the analyte(s) with a suitable solvent into a small tube. The solvent is either injected directly into the chromatograph or concentrated by evaporation prior to chromatography. This method is most commonly used for urine analyses for drugs of abuse.

SPME involves the use of fibers that selectively retain the analyte(s) after dipping the fiber into the sample and applying this directly to a GC. A heated injection port volatilizes the substance and introduces it to the GC column. Fibers used include dimethylpolysiloxane for low-molecular-weight compounds and volatiles, polyacrylate for polar compounds, dimethylpolysiloxane–divinyl benzene for volatile alcohols and amines, and a carbowax–templated resin for surfactants.

Another technique used occasionally is direct precipitation using methanol, acetonitrile, or other water-soluble solvents. This precipitates proteins and allows direct injection of the supernatant into a chromatograph (usually HPLC) or it becomes the first step of a further purification stage, e.g., solvent extraction.

Whatever the technique used, the overall recovery of the analyte(s) must be reproducible and efficient (>50%) (see sections on internal standards) with no chromatographic interferences caused by extraneous or co-extracted substances.

Chiral Analysis

Many drugs used in medicine and illicit drugs exist as optically active compounds i.e., existing in mirror image forms (stereoisomer) with one or more optically active (chiral) centers. Often, only one of these stereoisomers is pharmacologically active, or is much more active than the other isomer(s). For example, methamphetamine has one chiral center, with the S-configuration having the most activity as a central nervous system stimulant. Illicit methamphetamine has the S-configuration while the R-form is legally available in some countries (e.g., United States) as a weak stimulant for use as a nasal decongestant.

Since conventional chromatography does not separate stereoisomers, it is necessary to use chiral methods to distinguish the isomeric form. This can be accomplished by the use of either derivatization techniques to produce a diastereomer that can be separated on a conventional column or use of a column coated with an optically active phase that preferentially retards one isomer over another.

Quantitative Analysis

Many laboratories combine the confirmation step with a series of calibrators that enable the amount of identified substance to be quantified. Quantitative results are most useful for blood/plasma measurements, but are also useful in some applications of oral fluid, hair, and urine testing. See the respective chapters on "alternative specimens . . . " (**Hair: Toxicology**; **Oral Fluid Toxicology**; **Sweat: Toxicology**) for further details. Quantitative results provide a basis for interpretation of the possible significance of a drug presence, in terms of how much was taken, when the substance may have been taken, and what the likely effects of the substance may have been on the individual.

To achieve quantitative results, a series of calibrators are prepared at varying concentrations covering the anticipated concentration range expected for the samples. Typically this range extends from the lower limit of quantitation (LLOQ) to the highest concentration that still enables linear performance of the detector. Samples that contain concentrations of analyte(s) that exceed the highest calibrator will usually need to be repeated as a diluted sample. The response of the calibrators is calculated as a ratio of the peak area of the analyte(s) to the peak area of the internal standard IS (see below).

Performance measures of quantitative assays include the LLOQ, the linearity of the response over a concentration range, the precision of the analysis at particular concentrations over the calibration range, and the accuracy of a calibration. Typically precision should be less than ±20%, and the accuracy should be within 20% of the true concentration. Oftentimes these parameters are far tighter. For example, in alcohol analyses precisions of better than 2% can be obtained using GC techniques. These parameters all need to be properly established before an assay

is conducted on real samples and form part of the validation process.

Internal Standards

Analytical assays used for quantitative analyses must contain a marker to ensure the chromatography is adequate and any variations in extraction efficiency can be corrected. Indeed this marker, known as an *internal standard* (*IS*), should also be used in qualitative assays including confirmation assays. The IS is a substance with very similar physiochemical characteristics to the analyte(s) and is added to all calibrators and unknown samples before the start of the extraction or isolation procedure. The IS can be separated from the analyte(s) either because it has slightly different chromatographic performance (to the analyte) or it has different ions that can be selected by the mass spectrometer.

For example, if an IS in a sample chromatogram is half its usual area (or peak height), then by using the ratio of responses of the analyte(s) to the IS the concentration of the analyte(s) can still be calculated. In reality, this is no different from injecting half of the usual volume into the chromatograph since the ratio of the analyte(s) to the IS will still be the same although the individual peak areas have halved. The use of an IS therefore makes the response of any sample independent of the extraction efficiency (see the section Quantitative analysis). This improves the overall precision of quantitative analyses. It is good practice to calculate the retention time of a substance relative to the retention time of the IS as this value is independent of variations in column performance. Unfortunately this is not possible when isotopically labeled ISs are used (see below).

In the past ISs were often structurally related compounds, such as another barbiturate for a barbiturate assay, or another benzodiazepine for a benzodiazepine assay. Modern chromatographers prefer the use of stable isotopes of the analyte(s), invariably deuterated analogs. These can only be used when some form of MS is being used, since they co-chromatograph with the analyte and can only be distinguished from the analyte by the selection of the appropriate ion(s). Deuterated ISs are available for most common drugs and provide the best match with the analyte in terms of closeness of physiochemical characteristics.

Method Intelligence and Validation

All methods used in laboratories need to be properly validated prior to use such that they are fit for the intended purpose. The validation process ensures that the method is reproducible and rugged, i.e., it always works even when different analysts are using the method. The detection limits for confirmatory testing need to be well established, as well as the recovery and precision of the method at different drug concentrations. This is particularly important if the confirmatory method is also used to quantify the amount of compound in the sample. Moreover, the method should be checked for specificity by potentially interfering compounds, either for their known elution at or near selected compounds or because they are frequently coassociated with the compound(s) being confirmed.

This validation process together with its use in case work enables intelligence to be obtained of what can and cannot be detected using this method. Thus, a robust and well-developed confirmatory method will be far more useful in a laboratory than *ad hoc* or ill-defined methods.

Uncertainty of Measurement

Uncertainty of measurement is a concept used in laboratories where some form of quantitative result is being produced, or a result is compared to a decision limit, such as cutoff. It simply means that laboratories recognize that all measurements are subject to some error of measurement and that laboratories need to estimate or calculate this error and where necessary apply this to their work.

Typically laboratories will assess all steps of an assay for error of measurement, i.e., purity of substance, weighing errors, pipetting or other dilution errors, recovery of drug from a matrix, errors in determining size of a chromatographic peak, etc. If the combined error of a quantitative measurement is 10% and to show the distribution about a mean of about 95%, a coverage factor of 2 is often applied. This means that for a cutoff of $100\,\mathrm{ng\,ml^{-1}}$ the uncertainty of measurement is $\pm 20\%$ (2 times 10%) with a confidence of 95% or a coverage factor of 2.

References

[1] Moffatt, A.C., Osselton, W.D. & Widdop, B. (eds) (2004). *Clarke's Analysis of Drugs and Poisons*, 3rd Edition, Pharmaceutical Press, London.

[2] Peters, F.T., Drummer, O.H. & Musshoff, F. (2007). Validation of new methods. *Forensic Science International* **165**, 216–224.

[3] Peters, F.T., Jung, J., Kraemer, T. & Maurer, H.H. (2005). Fast, simple, and validated gas chromatographic–mass spectrometric assay for quantification of drugs relevant to diagnosis of brain death in human blood plasma samples. *Therapeutic Drug Monitoring* **27**(3), 334–344.

[4] Walker, V. & Mills, G.A. (2002). Solid-phase extraction in clinical biochemistry. *Annals of Clinical Biochemistry* **39**(Pt 5), 464–477.

[5] Ulrich, S. (2000). Solid-phase microextraction in biomedical analysis. *Journal of Chromatography. A* **902**(1), 167–194.

[6] Franke, J.P. & de Zeeuw, R.A. (1998). Solid-phase extraction procedures in systematic toxicological analysis. *Journal of Chromatography. B, Biomedical Sciences and Applications* **713**(1), 51–59.

[7] Lu, N.T. & Taylor, B.G. (2006). Drug screening and confirmation by GC-MS: comparison of EMIT II and Online KIMS against 10 drugs between US and England laboratories. *Forensic Science International* **157**(2–3), 106–116.

[8] Srinivas, N.R. (2004). Simultaneous chiral analyses of multiple analytes: case studies, implications and method development considerations. *Biomedical Chromatography* **18**(10), 759–784.

[9] Herraez-Hernandez, R., Campins-Falco, P. & Verdu-Andres, J. (2002). Strategies for the enantiomeric determination of amphetamine and related compounds by liquid chromatography. *Journal of Biochemical and Biophysical Methods* **54**(1–3), 147–167.

Related Articles

Toxicology: Initial Testing

OLAF H. DRUMMER

Congressman Sickles ***see*** Temporary Insanity

Context Effects ***see*** Interpretation: Observer Effects

Convulsions ***see*** Seizures: Behavioral

Copycat Crimes ***see*** Homicide: Multiple (Behavior)

Counterfeit Documents ***see*** Forged and Counterfeit Documents

Courts: Mental Health ***see*** Mental Health Courts

Courts: Specialized Mental Health ***see*** Therapeutic Jurisprudence

Crime Scene Documentation

Introduction

A crime scene investigator (CSI) can be unrivaled in his ability to identify, recover, and interpret evidence, but unless he can correctly document the evidence and its passage through the various processes required, it will be useless in court.

A CSI generates several documents during an investigation. These are necessary to demonstrate the integrity of the evidence and the continuity of the chain of custody (*see* **Chain of Possession of Tangible Evidence**). In many jurisdictions, all evidence recovered by a CSI is passed on to another unit or agency. Each of these diverse entities has its own requirements for information and therefore its own documents. The design and layout of documents vary greatly between agencies and legal systems but they fall into basic categories described.

To understand the complexity of the documentation needed, it is necessary to explain the various paths that evidence takes.

Physical and Trace Evidence

When recovered and packaged, it is recorded on the following documents:

- chain of custody form (for one such example, see Figure 1);
- crime scene report (Figure 2); and
- exhibits list (Figure 3).

It may also be recorded on the following:

- photo log (Figure 4) and
- sketch of the scene.

Most departments also design a multitude of forms to document each of the following activities. Upon return to the CSI office it is also recorded in

- exhibit store register.

At a later stage, it may be transferred to a forensic science laboratory for examination and analysis, which entails the exhibit being recorded on a

- laboratory submission form.

On removal from the exhibit store and its subsequent return to the CSI unit following analysis, the items' movements in and out are recorded in the exhibit store register.

DNA Evidence

When recovered and packaged, it is recorded on the following documents:

- chain of custody form;
- crime scene report; and
- exhibits list.

It may also be recorded on the following:

- photo log and
- sketch of the scene.

Upon return to the CSI office it is also recorded in

- CSI freezer register or exhibit store register.

At a later stage, it may be transferred to a forensic science laboratory for examination and analysis, which entails the exhibit being recorded on a

- laboratory submission form.

On removal from the CSI freezer or exhibit store and its subsequent return to the CSI unit following analysis, the item's movements in and out are recorded in the exhibit store register.

Footwear, Tire, and Tool Mark Evidence

When recovered and packaged, it is recorded on the following documents:

- chain of custody form;
- crime scene report; and
- exhibits list.

It may also be recorded on the following:

- photo log and
- sketch of the scene.

Upon return to the CSI office it is also recorded in

- exhibit store register.

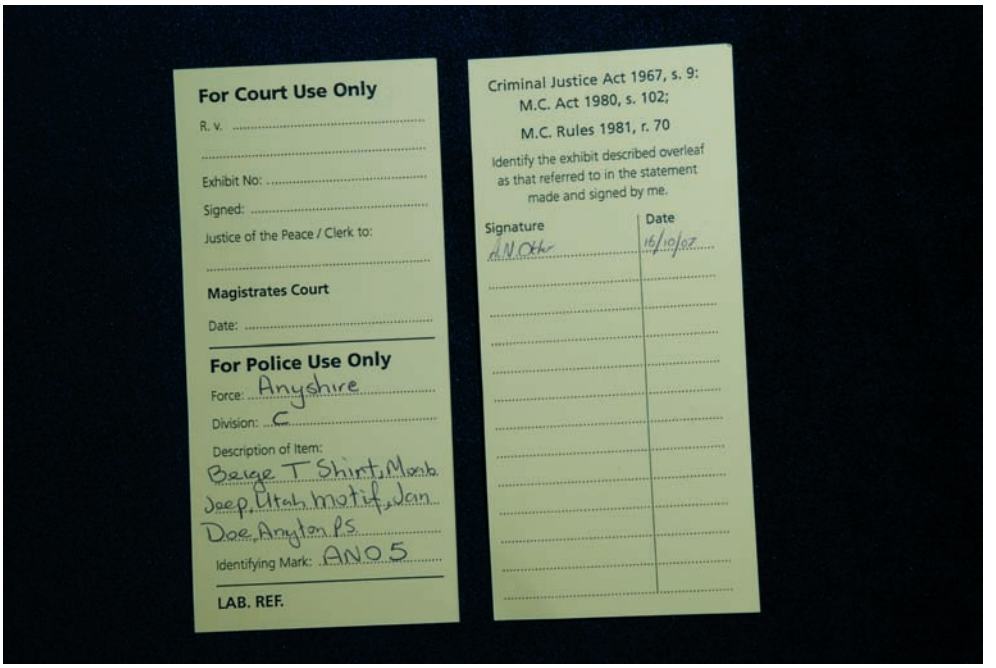


Figure 1 An example of a chain of custody document

At scene, prior to recovery by casting or other means, footwear, tire, and tool marks are photographed (*see* **Evidence Collection and Preservation: Casting**), so in addition their details are entered on an

- imaging submission form.

In addition, many police forces in the United Kingdom employ footwear analysts as part of the forensic intelligence section of their scientific support units. In those cases, a copy of the footwear mark would be forwarded to the analyst, along with a copy of the crime scene report and the copy of the exhibit would be recorded on a

- footwear submission form.

At a later stage, it may be transferred to a forensic science laboratory for examination and analysis, which entails the exhibit being recorded on a

- laboratory submission form.

On removal from the exhibit store and its subsequent return to the CSI unit following analysis, it is returned to the CSI unit, the item's movements in and out are recorded in the exhibit store register.

Photographs and Videos

When completed, films and/or disks are packaged and the following documents are completed:

- chain of custody form;
- crime scene report; and
- exhibits list.

They may also be recorded on the following:

- photo log.

They may also possibly be entered on the

- sketch of the scene.

Upon return to the CSI office, they will also be recorded on a

- imaging submission form.

And forwarded, either by hand or secure dispatch, to the relevant photographic or imaging unit.

Greater Mornington Police

Crime Scene Report

Offence: ..... Date: .....

Location/ Name/ Reg. No : .....

CSI: .....Date attended.....From:..... To: .....

Modus Operandi

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

ID Ref	Article	Where Found

Figure 2 Example of a crime scene report

Latent and Patent Fingerprints

When recovered, they are packaged and the following documents are completed:

- chain of custody form;
- crime scene report; and
- exhibits list.

Identification Ref.No:	Full Description:
Date: ..... Time:.....	Where Found:
Movement of Item:	Date:..... Time:.....

Identification Ref.No:	Full Description:
Date: ..... Time:.....	Where Found:
Movement of Item:	Date:..... Time:.....

Identification Ref.No:	Full Description:
Date: ..... Time:.....	Where Found:
Movement of Item:	Date:..... Time:.....

Identification Ref.No:	Full Description:
Date: ..... Time:.....	Where Found:
Movement of item:	Date:..... Time:.....

Identification Ref.No:	Full Description:
Date: ..... Time:.....	Where Found:
Movement of Item:	Date:..... Time:.....

Identification Ref.No:	Full Description:
Date: ..... Time:.....	Where Found:
Movement of Item:	Date:..... Time:.....

Identification Ref.No:	Full Description:
Date: ..... Time:.....	Where Found:
Movement of Item:	Date:..... Time:.....

Identification Ref.No:	Full Description:
Date: ..... Time:.....	Where Found:
Movement of Item:	Date:..... Time:.....

Figure 3 A basic exhibit list

Crime Scene Photo Log

Case #: 08-01-0002 Date: January 2, 2008

Camera Used: Nikon Time: 1130 – 1340 Hrs

Film Type: Media Scene: 905 Windmill CT

Photographer: SA Gardner Film Roll #: 1

Time	Photo #	Depicting	Remarks
1130	1	Overall of hillside looking from the SW-NE	Looking toward Spivey Road
1140	2	Overall of hillside looking from the NW-SE	
1147	3	Overall up hillside looking from the NE-SW	905 residence in background
1158	4	Overall of hillside looking from SE-NW	
1210	5	Evidence establishing looking from top of hillside back toward residence	
1222	6	Evidence establishing of body and shoes. Looking south.	Placards # A, B and C
1234	7	Evidence close-up of key ring.	Item E
1238	8	Evidence close-up of key ring with scale.	Item E

Figure 4 A typical photo log sheet [Reproduced with permission from R M Gardner.]

If patent fingerprints, they are also recorded on the following:

- photo log.

They may possibly also be entered on the

- sketch of the scene.

Upon return to the CSI office, dependent on the job description and qualifications of the CSI unit, they may be recorded on a

- fingerprint submission form.

Where the fingerprints have been photographed, they are also recorded on an

- imaging submission form.

Once fingerprints have been recovered from a scene, it is normal practice that elimination fingerprints are obtained from those persons with legitimate access.

- At this stage, agency protocols can vary considerably, so suffice to say they ensure that lifted fingerprints, photographs of fingerprints, and elimination fingerprints are collated and forwarded to the responsible unit or agency for comparison and identification, along with the fingerprint submission form.

Required Contents of Crime Scene Documents

Chain of Custody Document

When an item is recovered it must be packaged immediately and the packaging sealed to prevent any contamination from another source (*see* **Packaging and Transport**). Some packaging have the chain of custody document printed on it, or a label with the required details attached to the packaging.

In some jurisdictions this is called an *exhibit label*, while in others it may be referred to as an

evidence voucher. It should be noted that in Scotland, exhibits are referred to as *productions* as they are "produced" in court. All references to exhibits should be read as productions for the purposes of Scots law and documents include variations to include the requirement for corroboration of evidence.

Whatever its name, a chain of custody document is used to record the following details:

- Law enforcement agency.
- Basic command unit, where applicable (division, precinct, etc.).
- Description of item (This should be detailed enough to clearly distinguish similar items).
- Location recovered from (The location within a scene where the item was recovered).
- Address or scene where it was recovered.
- Identifying mark or reference (This varies between agencies. In England and Wales an identifying mark is the initials of the person recovering the item plus a chronological number starting at 1 e.g., ABC/1).
- Signature and date of person recovering and space for the signature of every other person taking possession of the item up to its presentation in court.

Whenever an item of evidence is passed on to another party, it is good practice to ensure that person signs for possession in front of the person relinquishing custody of the item. That way a CSI can be sure that if the chain of custody is broken it is not broken with them.

Crime Scene Report

While at a crime scene, it is imperative that a CSI maintain a detailed record of actions, observations, information, and evidence recovered.

A CSI not only recovers evidence but has an important part to play in the investigative and intelligence gathering processes. A crime scene report is an important part of both the investigation and intelligence gathering.

The style and layout of a crime scene report varies greatly between agencies. Even in England and Wales, with only 43 law enforcement agencies, there is little commonality of style and layout.

No matter where an agency is, a crime scene report includes the following details:

- case number and agency
- details of crime or incident
- address or scene
- details of CSI
- details of victim(s)
- details of first officer(s) attending
- date and time of incident, or time frame
- date and times CSI arrived and departed
- modus operandi and other pertinent details.

This should be as detailed as possible and include the following:

- description of the scene;
- actions and apparent actions taken by offender(s);
- point and method of entry to the scene;
- items moved by offender(s);
- items stolen by offender(s);
- items from scene not taken by offender(s);and
- items abandoned by offenders(s)

Some jurisdictions may also require that a detailed record is kept of the examination equipment and methods used.

Most jurisdictions prefer a narrative style of crime scene report, but some may have a checklist requirement for standardized information gathering for intelligence and statistical purposes. Care should be taken that a crime scene report remains an investigative tool and does not become solely a means of gathering statistical information that does not assist the investigative process.

- **Elimination actions taken or required**

This includes details of, where required, elimination fingerprints, footwear impressions, and DNA samples taken from those with legitimate access to the scene. Where elimination databases of police personnel are maintained, details of all officers attending should be included.

- **Sketch of crime scene**

Where necessary a sketch of the scene should be made and if available to an agency, professional plan drawers should be utilized to provide an exact scale plan of the scene.

List of exhibits

This should be maintained as exhibits are recovered and the details should be an exact copy of those

documented on the chain of custody document. This ensures that challenges as to verification of exhibits are minimized. Dependent on legal requirements and agency policies this may be part of the main crime scene report or a separate document (Figure 3). Legal requirements may vary but the main information recorded is the following:

- item number or identifying mark (see the section "Chain of Custody Document")
- description of item
- location recovered from
- person recovering item
- time item recovered
- signature.

If the exhibits list is a separate document, additional information such as scene, offense, and case number are required along with any other information such as search warrant number.

Where a serious incident is under investigation in England and Wales an exhibits register is maintained by an exhibits officer. This can be a CSI or experienced detective and all exhibits recovered are passed to them for documentation and secure storage. This is used to document all exhibit movements and also serves as a source of information that can guide the forensic recovery and submissions strategies (*see* **Crime Scene Management**).

Photographic Log

In many jurisdictions, it is a requirement to provide a log detailing the serial number of the camera equipment used and if a traditional film camera, the film batch number(s).

For each photograph taken, all or some of the following details are required:

- shutter speed
- film speed
- lens number and size, e.g., 55 mm
- aperture
- distance from camera to object in question and distance from reference points
- direction of photograph, i.e., north
- photographic placards and what they indicate.

Some jurisdictions may also require the height at which the camera was.

Post Scene Documentation

Once work at the scene has been completed, or in the case of a major incident where work stopped for the day, the onward movement of all evidence is required. Unlike their television counterparts, CSIs cannot simply hand an exhibit to someone, nor are they experts in all fields of forensic science. Almost everything a CSI recovers is either placed in a secure store or passed onto another unit or agency for further examination or analysis (Figure 5). This documentation is necessary to ensure a fully recorded and transparent movement of exhibits between the scene, storage, the receiving unit(s), and onward until the case is heard in court.

With several thousand agencies worldwide, it is impossible to establish a common style of documentation. As agency protocols and CSI job descriptions vary enormously some of the described documents may not be used.

Exhibits Store Register and CSI Freezer Registers

In major investigations, exhibits may be immediately dispatched to another unit or agency. However, in routine volume crime cases exhibits will normally be stored in secure locations pending decision on their submission for examination or analysis. Some agencies may have electronic registers while the majority of medium and smaller agencies use a paper-based system.

The sections requiring completion in those registers include all the details of the exhibit, which must match all the details of the exhibit marked on the evidence voucher (see the section "Chain of Custody Document" above). The last person to sign the evidence voucher must be the person who signs the item into the store or freezer. The register should include the following details for each item:

- date and time items deposited
- person depositing
- location in store/freezer
- movement from store
- time and date removed and reason (to laboratory, etc.).

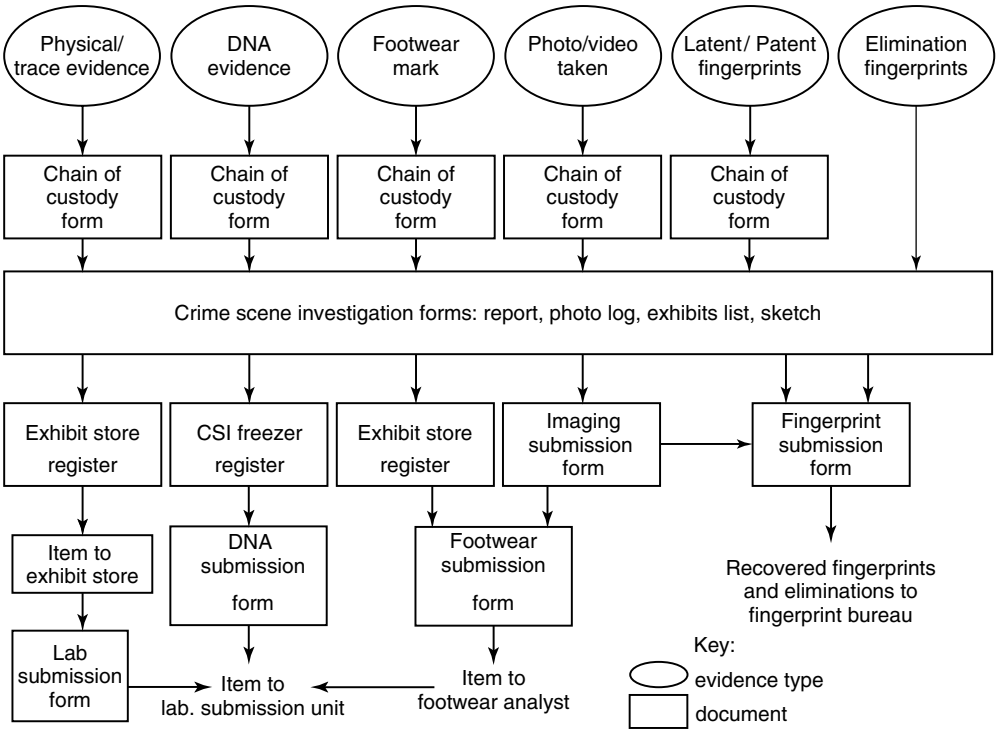


Figure 5 A typical flowchart of CSI actions and document generation

Imaging Submission Form

Once images of a scene have been made, the resultant film(s), tape(s), and disk(s) are either handed over to the imaging unit or dispatched to them via secure internal mail. This form not only documents the submission of the film(s), tape(s), and disks but the required details include such things as the following:

- CSI case number and agency
- imaging unit case number (if different)
- date and time items received
- person depositing
- person receiving
- number of working copies made and who they were provided to
- number of evidential copies made for, prosecutors, defense attorneys, jurors, and court officers.

Fingerprint Submission Form

In some agencies, a CSI may also be a latent print analyst, while in others, the latent print comparison and identification is the responsibility of a court recognized fingerprint expert. In the United Kingdom for example, police forces maintain fingerprint bureaus. Where the latent fingerprints are passed on to another unit or agency, a fingerprint submission form is required. The detail on this form varies from agency to agency but includes the following:

- agency and CSI case number
- fingerprint bureau case number (if different)
- list of lifted and photographed fingerprints
- date and time items received
- person depositing
- person receiving
- identity of persons whose elimination prints are enclosed.
- details of personnel attending scene
- results of comparison (eliminated, identified to person on record, and outstanding).

Dependent on local protocols or legal requirements, disposal of elimination fingerprints may be required.

Footwear Submission Form

Where an agency uses a footwear analyst to provide forensic intelligence on footwear impressions found at scenes a submission form is needed. The details on this form includes

- agency and CSI case number
- imaging unit case number (if different)
- footwear analysis case number
- date and time items received
- person depositing
- person receiving
- details of when entered on intelligence systems
- details of onward submission to a forensic science laboratory.

Laboratory Submission Form

When submitting items to a forensic science laboratory, it should be remembered that all forensic science interpretations are context based. Given one set of circumstances, the results of examination and analysis may refute the prosecution version of events. In another set of circumstances, the results may support the suspect's version of events. To this end the detail required on any laboratory submission form is substantial. Some details may vary but in general they include

- agency and investigating officer details
- agency case number
- CSI case number
- incident details
  - date, time, and offense type
  - full details of the offense
- details of all subjects (victims, deceased, suspects, and persons for elimination)
- occupation of subject (This could have great bearing on the interpretation, e.g., a glazing contractor could have glass on their hair clothing, some of which has the same refractive index as glass from a scene. This does not necessarily place him at the scene)
- name and signature of person delivering or dispatching the items and method used
- name and signature of person receiving
- list of exhibits, including date time and location of seizure
- previous submission reference numbers.

In some jurisdictions, where an invoice may be levied against the submitting agency, a submission form also includes the following:

- person authorizing submission.

Following receipt at the laboratory, the case is given a laboratory case number.

DNA Submission Form

In jurisdictions where DNA viable material is submitted for inclusion on a DNA database the requisite form normally contains the following details:

- agency and investigating officer details
- agency case number
- CSI case number
- incident details
  - date, time, and offense type
  - details of the offense
- name and signature of person delivering or dispatching the items and method used
- name and signature of person receiving
- list of exhibits, including date, time, and location of seizure
- previous submission reference numbers.

As can be seen from the previous lists, a considerable amount of administration effort is required of a CSI. Fortunately, in many larger jurisdictions, much of the very necessary but tedious work has been computerized, allowing several forms to be populated with the required information from a single entry or "one key" operation. This allows the CSI to simply read and sign the submission forms.

Major Incident Forms

Routine or volume crime generates substantial administration but major incidents create even more. This is necessary to demonstrate the thoroughness of the investigation.

Scene Entry Log

Before entering a scene, any personnel encounters an officer assigned to document all persons entering the

scene. In many jurisdictions, this is an informal list of personnel and times entering and departing the scene. In others, this is a formal document and includes the following details:

- *Cover*
  - offence
  - scene
  - crime scene manager
  - detailed instructions to the officer(s) on completion of the log and handover to a relieving officer.
- *Inner pages*
  - time and date
  - details of person requesting entry (instructions from Crime Scene manager should be specific about refusal of entry to all except those necessary for the examination function)
  - reason/authority for entry and reason for refusal
  - signature of person entering
  - time and date of entry to the scene
  - details of protective clothing worn
  - whether person is on a DNA/fingerprint elimination database
  - time and date of leaving scene
  - name and badge number of logging officer.

All of the above details are necessary to show that there was no contamination of or interference with the scene. Any nonpolice personnel attending the entry and exit point should have their details taken. Where no statutory power exists to compel them to provide their name etc., their vehicle license plate should be noted for future reference. This may prevent any future defense of the defendant having attended or entered the scene.

Crime Scene Manager and Crime Scene Coordinator Log

In the United Kingdom, the roles of crime scene manager and crime scene coordinator have been established for well over a decade (*see* **Crime Scene Management**). In the United States and other jurisdictions, the roles may be called a *lead CSI*. The United Kingdom and other European Union states have adopted the European Crime Scene Management Manual. Amongst the many recommendations of the manual is the need to maintain detailed logs for every scene involved in an incident and where a coordinator is appointed to oversee the various scenes, that person should also maintain a log. "Discovery" or "disclosure" legislation varies considerably and while some legislation may allow a degree of flexibility, other jurisdictions may make it compulsory that every written word in the course of an investigation has to be copied to the defense, including in some areas, a photograph of the notes on white dry wipe board. In reality, this may include a note of a family telephone number on the same piece of paper as the contact details of an official involved in the investigation, or the local pizza shop. The implications of either piece of information being divulged to the defense could present a personal security risk or the potential for an unscrupulous attorney to attempt to show that staff welfare was more important than their client's liberty.

The logs may vary in content and size but the general information is very similar.

The following descriptions of the contents outline the details required and the rationale for their inclusion.

Reference Number and Contact Details. It is a good practice and an efficient way of collating information to record the following information:

- Case reference numbers (dependent on agency guidelines).
- Details of personnel and their office phone number, cell phone/pager number. For example,
  - senior investigating officer and deputy
  - exhibits officer
  - command post senior staff
  - crime scene manager
  - CSI and lab staff
  - FME/coroners staff.

Under "discovery" or "disclosure" rules, it may be possible to withhold copies of this page from the defense and reference should be made to the applicable laws.

History of the Scene. Recording the details of all first responders ensures that if required all unidentified fingerprints, footwear marks, tire tracks, and possibly DNA can be checked against those person

for the purpose of elimination. Documenting the position of cordons and any adjustments prevent any false claims by defendants and others.

Initial Observations. Several pages are devoted to recording the details of the scene. This may seem a time consuming exercise but they are vital in transparently demonstrating that an open minded and thorough investigation has been conducted. Several pages may not be used in an investigation and are simply scored across diagonally and marked "N/A" (not applicable).

A full description of the scene would be documented, along with any approach path and the use of forensic stepping plates to facilitate activity in the scene without destroying any evidence on floors.

Any disturbance, forced entry, and items obviously foreign to the scene would be recorded.

If the scene of homicide is a residence, the details of all rooms are noted. The following should be recorded:

- Furnishing and contents.
- The status of doors and windows – open/ closed? Locked/ unlocked? Marks indicating forced entry? Keys in locks?
- Are any drapes/curtains/blinds open or closed?
- Lights – are they on? Position of switch? Does the light bulb work?
- Are sinks, washing machines and dishwashers, full, blocked or inoperative?
- Are the heating system and its timing mechanism working? What time does it switch on and off and what temperature is it set at?
- Are all utilities working in the premises?

The above may seem overly detailed but in all democracies the prosecution must prove the defendant guilty "beyond all reasonable doubt" and an inability to answer those questions later in court could prove problematic at least for the prosecution case.

Vehicle Details. When examining a vehicle, most crime scene investigators record the basic details of make, model, color, and license plate. However, all eventualities should be considered and the following details should be recorded:

- make, model, and color
- any damage
- true license plate details and any false details displayed
- chassis, engine, and vehicle identification number
- any required taxation or insurance details displayed
- odometer and trip meter readings
- fuel level and note any oil leaks
- where and when the vehicle was stolen, used in the commission of an offense and recovered
- make, model, size, and position of all tires on or in the vehicle.

The depth of detail recorded may seem excessive but later in an investigation several questions may be asked or new possible lines of inquiry opened. By recording more than the basic information most unexpected twists in a case can be dealt with. Consider the possibility of organized criminals having a decoy vehicle with an identity duplicating that of another. If all the additional details are recorded, it should be possible to demonstrate the existence of different vehicles.

Deceased Check List. In most areas of the United States, the deceased is the province of the medical examiner or coroner and not the responsibility of the CSI. In other countries, notably in the United Kingdom, it is the norm for a CSI team to assist the forensic medical examiner at the scene, attend the autopsy to record information, and recover evidence for onward movement to the necessary agencies. The autopsy would be considered a secondary crime scene.

The crime scene manager (CSM) would record details of the deceased.

- If known, their name, address, date of birth, and occupation.
- On the scene
  - the position of the body, visible marks of violence, tattoos, piercings, state of their clothing, and if areas of exposed skin have been tape lifted. Any body and air temperatures taken would be recorded
  - details of surrounding areas, including vegetation, that may affect the body
  - weapons
  - water.

At autopsy, detailed records of the clothing would be taken, as some garments may be missing or on the

body in an incorrect manner if the deceased had been stripped and redressed postmortem by the offender.

Sketch Plan. A sketch plan should be made by the CSM for easy reference.

Information from Fire Officers. Where fire is involved, many US jurisdictions call on the services of Fire Marshals or the Bureau of Alcohol, Tobacco, Firearms, and Explosives, (ATFE, formerly Bureau of Alcohol, Tobacco and Firearms (ATF)) to investigate. In the United Kingdom, a joint investigation among Fire Service, Police, and a forensic scientist would be undertaken.

The CSM would record such information from the first fire unit as

- Security of premises on arrival and any forcible entry made.
- Location and extent of burning on arrival.
- Effects of firefighting on spread of the fire and final burning pattern.
- Any relevant observations, including persons on scene.
- Weather.
- If any video surveillance equipment was fitted to fire trucks.

Scene Attendance Cross Reference. Where there are multiple scenes involved in a major investigation, the crime scene coordinator should maintain a scene attendance cross reference or matrix.

These pages have a grid or matrix with the scenes noted in separate boxes along the top and the details of the personnel attending down the left side. Where a person enters a scene, an "X" and the date of attendance are placed in the box across from their name and beneath the scene. This provides an easy and rapid check that no person has attended two scenes that could be linked by forensic evidence. This helps eliminate the possibility of cross-contamination of scenes.

Scene Records. In the United Kingdom, although the scene is managed by the CSM, it is still under the control of the senior investigating officer (SIO).

Following discussion, the SIO and CSM agree the search parameters and if necessary adjust cordons accordingly. Under the Health and Safety at Work Act 1974, they carry out a risk assessment and ensure all persons in attendance are aware of any potential risks. This could be the possibility of infected blood, sharp instruments trip hazards. Control measure to remove or reduce those risks must always be implemented. They also decide on the forensic recovery strategy. All of the above decisions are documented.

While at scene, all decisions made by the CSM, as circumstances change, are documented along with the rationale. This may seem excessive but it does ensure the transparency of the investigation. Several months or years later, a court may ask why a decision was made and with the benefit of hindsight may attract criticism. However, if, based on information available at the time, the decision was correct, any reasonable criticism can be answered.

The CSM also notes all relevant information from briefings, instructions received, instructions, and briefings given to personnel and any informal contacts and meetings. This not only ensures transparency but provides reminders about information that needs to be disseminated to staff and higher authority.

Aide Memoir

The last pages contain an A*ide Memoir* to remind the CSM of items that can easily be overlooked when multitasking.

An example of the *Aide memoir* is

Scene manager/scene coordinator to be appointed.

Scene manager to receive comprehensive briefing by officer/SIO at the scene. Ensure logs have been commenced.

Review scene preservation methods/cordons. Establish rendezvous point and car parking area.

Ensure full protective clothing, suits/overshoes worn by all personnel entering the scene and that they have a valid reason for doing so. Brief scene entry loggist on the requirements/restrictions on entry to scene.

Establish common approach path and consider stepping plates. Consider alternative routes. Consider tents, screens, sheeting and lighting, etc.

Consider closing the scene until daylight.

Commence initial scene video/still photography. Consider aerial photography.

Scene manager to decide on and arrange attendance of pathologist, forensic scientist and any other specialists.

Forensic medical examiner/pathologist to certify death.

Air/room temperatures, Meteorological Office data, heating settings, etc.

Check all electrical switches/sockets and appliances are working.

Check all gas appliances/pipes in working order.

Consider 1/2 h rectal temperatures (Not in sexual offences) and air temperatures if the body is to remain in situ for a prolonged period.

Scene manager to liaise with plan drawer(s) regarding scene requirements.

ALL firearms found to be made safe by authorized officers. Consider the attendance of a ballistics specialist.

Consider preservation of insect life on body – especially in cases of advanced decomposition or if the body may have been moved.

Consider the use of an entomologist, biologist, and anthropologist.

Complete relevant examination of the body prior to removal. Tape lift as appropriate. In cases of burning, tape lift all exposed areas of skin as it is prone to break away from body.

Bag head, hands and feet before removal. If possibility cocaine use involved paper bag on head.

Obtain samples from underneath the body. (For example, soil, carpet, etc.)

Ascertain identities of all persons entering the scene prior to introduction of control measures and obtain elimination fingerprints/footwear. Consider seizing their clothing.

Arrange removal of deceased.

Recover "fragile" evidence from entry point (for example, fibers, etc.) Tape lift contact areas on common approach paths.

Latent footwear impressions in dust. ESLA as soon as possible and always before further examinations undertaken.

Consider chemical enhancement of footwear marks in blood.

Tool/tire/footwear impressions – photograph and cast.

Blood distribution – mark and photograph. Consider attendance of scientist. Agree plan of action E.G. swab or removal of wallpaper, flooring etc.

Fingerprint examination – agree parameters with SIO. Consider use of Quasar/chemical treatment.

Arrange for extended search by operations support unit.

Consider seizure of CCTV video tapes.

Even if not apparently involved, vehicles of victims and suspects.

Further Reading

Gardner, R.M. (2005). *Practical Crime Scene Processing and Investigation*, CRC Press, Boca Raton, pp. 201–218.

ALLAN MATHIESON SCOTT

Crime Scene Investigation

Introduction

The basis of crime scene investigation, as with every other aspect of forensic science, is Locard's theory, sometimes called *Locard's Law*. Simply stated it is: "Every contact leaves a trace." Considering its application to a crime scene, it was described very eloquently in a US Federal Court decision (*Harris vs. United States,* 331 U.S. 145 (1947).

> Wherever he steps, whatever he touches, whatever he leaves, even unconsciously, will serve as silent witness against him.
>
> Not only his fingerprints or his footprints, but his hair, the fibers from his clothes, the glass he breaks, the tool mark he leaves, the paint he scratches, the blood or semen he deposits or collects – all of these and more bear mute witness against him.
>
> This is evidence that does not forget.
>
> It is not confused by the excitement of the moment.
>
> It is not absent because human witnesses are, it is factual evidence, physical evidence cannot be wrong; it cannot perjure itself; it cannot be wholly absent, only its interpretation can err.
>
> Only human failure to find it, study and understand it can diminish its value.

In the United States, larger agencies' workloads and crime rates can justify crime scene investigation units or even full crime laboratories. Others must rely on state and adjacent law enforcement agencies, joint task forces, officers with minimal training and experience, and even volunteers.

Around the world, the police are organized in different ways. Some countries have single national police forces, while others have local and federal agencies. Those countries following variations of Napoleonic Law (*see* **Civil Law Systems of Justice**) can have policing systems, where the uniformed patrol force, the detective force, and the magistrates who interview witnesses and suspects are responsible to different national government departments. Other countries, such as the United Kingdom, have large police forces serving one or a combination of local government areas.

In the British Isles, the examination of a crime scene is undertaken by trained police personnel, formerly police officers but now more likely to be civilian support staff. They are known variously as *scenes of crime officer* (SOCO) or now more commonly crime scene investigator (CSI). In other countries, there are various titles for the role: crime scene examiner (CSE), crime scene analyst (CSA), crime scene technician, evidence technician, forensic specialist, etc. For simplicity, the CSI is used for those who examine a "scene".

A scene is described as any premises, location, vehicle, person, or indeed an item involved in an offense or incident.

The First Officer(s) Attending (FOA)

Any officer performing this role has a responsibility to preserve the scene (*see* **Crime Scene Management**).

Actions of a CSI at a Volume Crime Scene

As Gardner [1] and White [2] correctly state, a CSI must carry out an independent assessment of a crime scene. Witnesses can be confused and provide misleading information and only by careful examination of the scene, one can arrive at a true understanding of the sequence of events. The CSI continuously reassesses a scene, based on his observations, information received, and evidence revealed during the activities of scene processing.

According to Smith [3], when a CSI attends a volume crime scene, he should be aware of the contamination risks. While no specific guidance is given, it would be prudent to follow the simple anticontamination procedures that many police forces use:

- **Always wear a face mask.**

This not only prevents accidental contamination of potential DNA sources but also protects you from any hazards due to long-term exposure to fingerprint powders.

- **Change your disposable gloves after handling individual items.**

Touching your face, adjusting spectacles, and touching everyday items such as handles and switches could transfer DNA.

Smith [3] suggests an eight point strategy for the investigation of a scene such as a home burglary. Gardner [1] and White [2] recommend similar strategies.

Attend the Scene and Make Contact with the Victim

When the CSI attends a scene, he should make contact with the victim, explain why he is there, and what is his role in the investigation. To assist in establishing a time frame, the CSI needs to ascertain if victims were on scene at the time of the offense and, if they were present, whether they saw or heard anything related to the offense.

If they were not present, it is important to establish when they departed and returned. This is important information that can impact on the intelligence that can be gained from a scene investigation, even if no forensic evidence is recovered.

Assess the Scene

Reassessment of the initial preservation is the immediate priority. Various factors must be considered, such as prevailing weather conditions and the possibility of further unintended interference with evidence, due to location.

By visual examination and questioning the victim, the CSI should review the extent of the scene and establish the scene boundaries. Careful observation allows the establishment of the points of entry and egresses and assesses the potential evidence, including objects left at the scene by the offenders.

Health and Safety Risks

Consideration should be given to any potential health risks present and the appropriate safety measures followed.

Carry out the forensic examination (process the scene).

It is important to adopt search and examination techniques appropriate to the scene and circumstances of the offense.

At burglary scenes, fencing, gates, and vegetation should be checked for fibers (*see* **Crime Scene Photography: US Perspective**; **Packaging and Transport**). The ground inside and outside the property line should be checked for footwear and tire marks (*see* **Evidence Collection and Preservation: Casting**; **Packaging and Transport**).

There may be window glass or items moved or abandoned by the perpetrators placed outside the premises. If so, as soon as the necessary photography and documentation is complete, those items should be examined for trace, biological, and fingerprint evidence, recovered and removed to a place of safety, to ensure the evidence is not damaged, degraded, contaminated, or lost and to prevent any injury to members of the public.

As every crime scene is individual, it is difficult to prescribe which items and indeed which examination precedes another. While Association of Chief Police Officers (ACPO) [4] recommend that DNA recovery should take precedence, this is not always possible. Where other evidence is located on the path to the possible DNA source, that evidence must first be recovered to prevent its loss.

Provide advice and guidance when examination is complete.

When the examination is complete, the victim should be advised as to what has been found and removed from the scene and given advice on what the best method is to clean up after concluding the examination.

The CSI also tries to ascertain which officers or other persons have entered the scene after discovery of the offense, and where they have walked. This permits determining whose fingerprints, DNA, or shoe impressions need to be acquired for elimination purposes.

Where elimination fingerprints, footwear marks, and DNA are required, an explanation of the need to obtain them and what will happen to those elimination samples, will be appreciated by the victim.

Complete the relevant documentation (see **Crime Scene Documentation***).*

Maintain continuity and integrity of samples (see **Crime Scene Documentation***;* **Packaging and Transport***).*

Disseminate the results of the examination (see **Crime Scene Documentation***).*

Burglary Scene – Potential Evidence to be Considered

By the careful questioning of the victim, it is possible to establish what has been touched, moved, or stolen. All of which allows the CSI to complete a detailed crime scene investigation report (*see* **Crime Scene Documentation** and Phys Ev as Intel).

The CSI has to consider the potential of all evidence types. Starting at the point where the offender(s) entered the property, the possibility of the following evidence types should be considered:-

Hairs and fibers snagged on fences, gates, bushes, doors, windows, and walls can help place a perpetrator, or at least their clothing, at a scene. Fibers can help to link scenes in a series of crimes, where an offender consistently wears the same clothing or gloves. They also provide an indication of where the property was entered (*see* **Crime Scene Documentation**).

Where necessary, botanical evidence can be utilized. In some cases, this can be in the form of a wooden weapon or safe ballast [2].

When a person walks across rough surfaces or debris such as broken glass, unique damage occurs to the tread of sole of the footwear. The human gait and various other factors produce different wear patterns on similar shoes worn by different individuals. Similarly, tires also receive unique damage. These impressions can be recorded and casts can be made (*see* **Evidence Collection and Preservation: Casting**).

Pioneered by Cornelius Van der Lugt from the Netherlands and by others, ear impression comparisons are accepted as evidences in some jurisdictions. Similarly, lip impressions are currently being assessed as a means of identification.

When an implement such as a screwdriver, prybar, or case opener is used to force entry, there are three evidential possibilities to consider. First, if door or window is painted or varnished, some of the paint is likely to be transferred to the implement. If the tool is painted, some of its paint will transfer to the surface of the door or window. The blade of a screwdriver can leave impressions on both the frame and the closing edge of a door. The grinding process used to shape the blade can produce unique striation marks which are visible under magnification. These transferred paints and impressions can be used both as evidence and intelligence (*see* Phys Ev as Intel and **Evidence Collection and Preservation: Casting**).

When breaking glass or forcing entry, an offender may have cut himself and leave vital DNA evidence, which assists in establishing his identity. In addition to blood, the possibility of discarded items, such as cigarettes, chewing gum, drink containers, and partly consumed food, having DNA on them may be recovered for analysis.

One should bear in mind not only the impressed footwear impressions in soil but also the possibility of their presence on floors, work surfaces, and paper strewn on floor. Even where impressions are indiscernible, the possibility of electrostatic recovery should be considered.

Where footwear indentations are visible on paper, it is likely to subject it to an ESDA examination as such impressions can sometimes be visualized or developed. When muddy footwear impressions are left on floors and work surfaces, developing them by the application of fingerprint powder can provide unique evidence or intelligence. While the more visible impressions may have been preserved by the FOA and the victim, it should be noted that they are often the least useful for evidential purposes. Further, the impressions in the crime scene having excess mud and debris removed by the simple act of walking are more likely to have detail useful to a footwear analyst.

The mainstay of crime scene investigation, finger and palm prints, can be left by even the most careful criminal (*see* **Evidence Collection and Preservation: Casting**).

Although a glove mark is not always useful evidence, it should never be discounted as an evidence type or intelligence tool. Damage to rubber, vinyl, or leather gloves may, just like footwear, turn out to be unique. Where a suspect is detained and his gloves are seized, they could provide evidence of his glove being used at a scene. Where a perpetrator has abandoned gloves, the chemical development of latent finger and palm impressions from the inside of suitable materials may be attempted wherever possible.

In the heat of the moment, many items such as implements, weapons, clothing, and possessions are lost or abandoned at the scene by perpetrators. Not only those items can bear forensic evidence, but also they can have other, possibly documentary, evidence that can lead to the identification of offenders.

A forenic scientist must have something against which the comparison can be made. Thus, where forced entry has been made, it is imperative that exemplars be taken. In the United Kingdom and other countries, these are referred to as control samples, e.g., glass, wood and paint, fibers from long pile carpets. If, for example, a suspect is arrested close to a crime scene, in possession of a screwdriver, the cast tool mark can place the screwdriver at the scene. This, however, does not positively establish that he was at the scene. If minute fragments of paint from the scene are found on his clothing, this can place him, at least close to, at the forced door. In this case, both pieces of forensic evidence, presented together, could be compelling.

On completion of the examination, the CSI must establish the persons with legitimate access that need to be eliminated by checking against fingerprints, biological, and physical evidence found. Wherever possible, the requisite elimination fingerprints and samples should be taken at this time by the CSI. Details of those who are no longer present at the scene and whose elimination evidence is needed should be obtained as soon as is practicable.

Vehicles – Potential Evidence to be Considered

Developed nations are wedded to motor vehicles to different degrees and in many criminal investigations vehicles are subject to crime scene investigation. At the lower end of the scale, the vehicle may have been forced open and items stolen, or it may have been used in the commission of a homicide or other crime. The extent of examination depends on agency policies and priorities.

As with all other scenes, one must remain aware of the contamination risks. It would be wise to follow the simple anticontamination procedures used at a burglary.

Where a vehicle is being examined as part of a serious crime investigation, full protective clothing should be worn.

It is important that where there is the possibility of a victim having been placed in the trunk, it should be established whether that area has been checked by officers attending the report of the vehicle being found. If this has not been done, examining and operating the release mechanism for opening the trunk should be undertaken.

Before processing the vehicle, the appropriate photographs should be taken and wherever visible potential evidence is found, additional close-up photographs should be taken.

The vehicle license plates and any corresponding state DMV indicators on the windshield should be checked and documented. When appropriate, the manufacturer's vehicle identification numbers should always be checked. They are located normally under the hood and on the floor pan next to the driver's seat, or, in some cases, are impressed on the dash of the car so as to be visible from the outside. It should always be remembered that if a vehicle bears a false identity, the engine serial number should also be checked.

If a window has been broken to gain entry, the frame should be examined for any fibers or hairs. If any are found, they should be recovered. Exemplars should be taken from all broken windows (*see* **Evidence Collection and Preservation: Casting**).

Throughout the examination, one should always be careful to recover blood evidence at a time and using a method that will not compromise the examination of other forensic evidences. Likewise, it is important to ensure that other evidence recovery does not compromise DNA analysis.

If the vehicle is believed to collide with a person, one should conduct a careful examination of the exterior, including windows, roof, and rear. Fibers hair and dried blood can be attached to the trim and damaged areas. Even the washing of the vehicle may not have destroyed this evidence.

Where the offenders sat in a vehicle, tape lifts should always be taken from the seat belts and seats (*see* **Packaging and Transport**). The material used in the manufacture of seat belts is particularly good at retaining extraneous hairs and fibers. The use of separate tape lifts for the seat back, seat, head restraint, and the rear of the front seats is recommended. It may be possible to determine the location within the vehicle occupied by the offenders.

Once this has been done, the investigation can focus on to the foot wells or floor pan. Any papers, food wrappings, partially consumed food, and used drinks containers found, should be recovered for possible fingerprint, footwear, and DNA analysis. Where loose mats are used on the floor, they should be removed carefully as there may be vital evidence, such as hairs, fibers, glass, paint etc, that could link the vehicle to a location, other vehicle, victim, or suspect (*see* **Packaging and Transport**). Where none are used, the fitted carpet area should be brushed carefully and, if necessary, the carpet is tape lifted (*see* **Packaging and Transport**). Only when the above evidence recovery processes have been undertaken, the search and examination can carry on.

At this point, if the vehicle is involved in a serious offense, the mileage and trip meters readings should be recorded, as well as the fuel level.

As with any scene, one must constantly assess both the potential evidence and hazards. A careful search of the glove box, door pockets, seatback pockets, and underside of seats should be made and any possible evidence should be recovered. It is important to be mindful of the possibility of encountering sharp objects that present a safety hazard to the investigators.

The mechanical fit of objects to items recovered elsewhere should also be considered. This is particularly important where the vehicle has possibly been in collision with another vehicle, object, or person. There may be broken parts, light lens plastic fragments, damage to bulbs, trim, and even panels with missing paint. Where a vehicle has been repaired using filler compound, this can also furnish valuable evidence.

Plastic bags from a roll of bags, cable or broken tools, or weapons are all possible sources of mechanical fit evidence. Items such as ropes and cords and pieces of tape such as packaging or duck tape are not just possible sources of a mechanical fit or fracture match. Rope may have other fibers, hair and epithelial cells caught in the plait, as well as possible microscopic pieces of dried blood. The mechanical fit of one section of tape to a roll of tape can often be matched. Fibers and hairs can attach not only to the

adhesive side but also to the edge of the roll. Fingerprints, DNA, and even bite marks are other evidential possibilities to consider.

Once you have recovered physical, trace, and visible DNA evidence, you should consider the use of alternate light sources as well as luminol.

Areas inside the vehicle covered by "orange peel" plastic surfaces, such as steering wheels, gear shift and parking brake, may not be suitable for latent print development but may be suitable for LCN DNA.

Where it is important to determine whether a vehicle has been at another location, one should consider foreign and exemplar paint samples, tire impressions, and recovering the debris from inside wheel arches and underside. When examining tires, their make, model, size, condition and position on the vehicle should be noted. This evidence could provide important information during the inquiry.

One should always consider the possibility of oil and grease leaking or being rubbed off. This, too, may provide important evidence that helps to place a vehicle at a scene. For example, if a vehicle has an oil leak and is stationary for a period of time, oil could be deposited on the ground. The oil sample should be taken from the sump area.

Outside Scenes – Potential Evidence

With the exception of possible livestock thefts in rural areas, most scenes in fields, woods, parks etc, will probably be limited to the investigation of serious offenses or related to vehicle involvement.

Where the scene is in or near water, samples of the water should be obtained.

The search and recovery process at a major outside scene would be decided by the Senior Investigator in consultation with the Lead CSI or crime scene manager (*see* **Crime Scene Management**). Crime scene investigation is the same for all offenses but serious offences prompt a greater response in terms of resources.

References

[1] Gardner, R.M. (2005). *Practical Crime Scene Processing and Investigation*, CRC Press, Boca Raton, pp. 76, 80, 87–128.

[2] White, P. (ed) (2004). *Crime Scene to Court, the Essentials of Forensic Science*, 2nd Edition, The Royal Society of Chemistry, Cambridge, pp. 28–29, 74–75.

[3] Smith, K. (2002). *ACPO Investigation of Volume Crime Manual*, National Crime Faculty, Bramshill, pp. 19–21.

[4] ACPO (2005). *DNA Good Practice Manual*, Association of Chief Police Officers, London, p. 16.

ALLAN MATHIESON SCOTT

Crime Scene Investigation: Evidence Collection *see* Evidence Collection and Preservation: Casting

Crime Scene Investigation: Mass Grave Investigation *see* Mass Grave Investigation

Crime Scene Management

Volume Crime Scene Management – First Report to Police

Crime scene management of these offenses begins with the first report to police. As it may be some time before a law enforcement officer attends the scene, the advice given by the personnel receiving the report can have important consequences on the final resolution of the case.

This advice should be to touch as little as possible and move about as little as possible. Once officers have been dispatched to the scene, subsequent questioning will allow dispatchers to evaluate the prevailing weather conditions and advise on retrieval

of items outside the premises to protect them from inclement weather.

First Officer Attending

The first officer attending should be aware of the potential that forensic science has to detect the crime. When approaching the scene, they should consider the following possibilities:

- Stolen property being hidden nearby with a view to later retrieval by the offenders.
- Fibers from offenders' clothing attached to any rough or broken surfaces with which an offender could have had contact.
- Footwear impressions:
  - In soil or mud where the offenders have breached the perimeter fencing of the scene.
  - In mud on smooth surfaces inside the premises. The best impressions for evidential purposes will be those further from the initial impression inside the point of entry. Any areas where illegal entry from outside has been gained should be entered with care.
  - In dust on smooth surfaces, as they could be recovered using an electrostatic lifting apparatus (ESLA) (see section on footwear recovery).
  - On papers scattered about during the commission of the offense. Papers should be left *in situ*, pending the attendance of a crime scene investigator (CSI). Footwear marks on papers can be developed using ESLA. Where footwear has left an indented impression on a piece of paper, its image may be enhanced at the crime laboratory by using an electrostatic development apparatus (ESDA), normally used for detecting indented handwriting or typescript on documents.
- Instrument marks, on doors and windows, that have been forced open by items such as a pry bar or screwdriver.
- Where doors and windows have been forced or the glass broken, officers should remember that, not only may offenders have snagged their clothing, but they may have cut themselves. Great care should be taken to ensure that this evidence is neither disturbed nor contaminated.

Major Crime Scenes – First Officer(s) Attending

When the first officer or officers arrive at a major scene, they have several, often conflicting, tasks to perform.

Their first priority is to preserve life but they must always consider the possibility of forensic evidence and, where offense details are apparent, take the route least likely to be used by the offender. This route is known as the common approach path (CAP) and should be used by all subsequent personnel entering the scene.

They must check for witnesses and ensure they are retained at scene and if possible kept apart from one another. It should also be considered that any witness might in fact be an offender.

They must also be aware of persons entering and leaving the immediate area of the scene. In addition to all the above, they need to ensure that the scene is secured and protected. This can be done by using police cordon tape or by improvization using garbage can lids, boxes, or even a car hubcap to protect fragile evidence. Always remember to "think big". Consider where the scene could end and keep people out. For example, if an incident occurs in a front yard or street, the whole length of the block should be cordoned off. This can always be modified and reduced later. Consider the impact on the case if CSI needs to extend a cordon and evidence is found in an area that anyone had access to.

Major Crime Scene Management in England and Wales

There is no legal definition in the United Kingdom of a major crime scene. It normally applies to scenes of murder or serious assaults that may prove fatal. A major scene ranges from a relatively simple domestic homicide where one partner has killed the other to serial killings, terrorist incidents, and mass disaster. However, large or small the scene, it is imperative it is managed in a professional structured way.

Prior to the 1990s there was an almost unprofessional approach to the management and examination of major crime scenes. The senior investigating officer (SIO), normally of superintendent rank, would decide on the sequence of examinations. Those officers often had little or no understanding of forensic science. Although by this time, police forces outside

London had from about 500 to 7000 officers, there was a lack of suitable expertise. In some forces the Scenes of Crime Unit, Photographic Unit, and Fingerprint Bureau were part of different divisions, reporting to different top-level management. Although the three departments interacted on a daily basis, there was no coordination of their activities. At a major scene, there was a reliance on the expertise of forensic science specialists but with no coordination at scenes some examinations were carried out at the expense of others and there was a real possibility that crucial evidence could be lost.

There was little incentive to change as confessions were much more common than they are today. Expert evidence was very rarely challenged and even the integrity of evidence, which was open to question in some cases, was almost always accepted.

The training available to an SIO and a CSI varied greatly and for those involved in what we now call "Scientific Support" there were very few career opportunities for personal development. Although there had been civilian staff in the departments for over 25 years in some cases, the management of the department was still the responsibility of a police officer. Police officers performing management roles in the Scenes of Crime, Photographic, and Fingerprint Units were sometimes in those posts for possibly only two to three years, as part of their career development. In some cases, this led to a lack of continuity and a very low status and esteem for those involved.

By the 1980s, several high profile appeals in criminal and especially terrorist cases demonstrated there was the possibility of unsafe convictions. This led to a lack of confidence in the police. This confidence is vital to the democratic principle of policing by consent.

From 1986 onwards, there were several reviews of the structure and use of forensic science by the police in England and Wales. However, those reviews only made recommendations and none of them actually placed any requirements on police forces. Successive recommendations were wide ranging and included training, organizational structures, roles, and equipment.

Forensic science laboratory provision in England and Wales has gone through radical changes over the last 15 years. The Home Office Forensic Science Service became an agency of the Home Office and later a Government owned company. As a result, forensic science moved from being a central service on behalf of the police to an effectively free market, where several competing organizations provide forensic services to police forces.

Recommendations for change with police forces were also made. They included the roles of scientific support manager (SSM), crime scene manager (CSM), and crime scene coordinator (CSC).

The Crime Scene Unit, Fingerprint Bureau, and Photographic Unit were all integrated into one Scientific Support Department (SSD) under the SSM with overall responsibility for all aspects of forensic science within the force. In what was then viewed as a radical policy, the SSM could be a police officer, a forensic scientist brought in from a laboratory or a civilian with the right skills and background. This move paved the way for a progressive career path for SSD staff and also to a more professional approach to forensic science within police forces.

With this professional approach, it quickly became evident that a far more scientific approach to managing forensic science at major scenes was required. The roles of CSM and CSC quickly evolved.

A CSM is normally a supervisory CSI. They have operational control of a major scene. They are appointed for individual major scenes and are part of the incident management team and forensic management team. The first CSM may well take a dual role as CSM and CSC in a relatively simple case.

In complex cases, a CSC is normally appointed to oversee and coordinate actions at all the scenes involved. In many cases as well as the primary scene there could be several secondary scenes. Any of the following could be a primary or secondary scene:

- scene of assault or killing;
- deposition site;
- vehicle(s) of victim and suspect(s); and
- residence(s) of victim or suspect(s).

It is imperative that there can be no suggestion of cross-contamination of those scenes by personnel going directly from one scene to another. It will not always be clear which scenes need to be connected to one another or to any of the subjects involved in the offense. Good practice is that as a CSC, you keep an anticontamination matrix (*see* **Crime Scene Documentation**).

Crime Scene Manager – Responsibilities

As a CSM, you would be a member of incident management team. This recognizes the vital nature of the forensic evidence available at a scene. As such you would have regular meetings with the senior investigator, exhibits officer, and others and would be involved in decisions as well as advising on forensic matters.

CSI Activities at Major Scenes

As either the first CSI or as a CSM arriving at a scene, you should always review and, if necessary, revise scene protection and cordons. It is worth repeating the advice to "think big". Consider where the scene could end and keep people out. For example, if an incident occurs in a front yard or street, the whole length of the block should be closed off. This can always be modified and reduced later and it allows you carry out a "flash search". Information gained about the exit route is always useful. No matter how careful the planning of a criminal enterprise, once the crime has been committed, caution can be thrown to the wind and perpetrators start to try and distance themselves from the offense. A point to point or "flash search" along this route may reveal garments, weapons, tools, or even documentary evidence. Consider the impact on the case if you need to extend a cordon and evidence is found in an area that anyone has access to.

You should also consider the need for inner and outer cordons at extensive scenes. All nonemergency service personnel are prevented from entering the outer cordon. This allows you to designate a rendezvous point and park essential vehicles between the two cordons. Only those personnel essential to forensic operations should be allowed inside the inner cordon.

Although it should already be in use, you should ensure that a scene entry log (*see* **Crime Scene Documentation**) is maintained. Always reinforce the need for exclusion and record keeping, especially where inexperienced officers are involved. In addition to any persons allowed access to a scene, all perimeter guards should be instructed to record details of all persons seeking entry. Where statutes permit, the identity of those persons enquiring what is happening should be obtained and what, if anything, those persons observed. This can possibly obtain useful information and where they later decide to give evidence for the defense, negate any perjury they may commit. Any vehicles that continuously drive past the scene should always have their details noted. The occupants may be purely inquisitive but they may have ulterior motives for their interest.

Many agencies operate a system that requires those entering a scene to sign in and out. This can eliminate any future doubts as to who entered a scene. A simple way to deter inquisitive officers of any rank is a large notice (*see* **Crime Scene Documentation**)

All CSIs should maintain a record of their actions. As a CSM you should also initiate a record of scene examination or CSM log. (*see* **Crime Scene Documentation**) Aide memoirs are useful but should not be treated as exclusive checklists. Always consider all options, possibilities, and the "I wonder if?" factor.

As the first CSI attending, you should always establish a CAP. This may require clearing a path from the perimeter to the deceased or center of the scene. This should always be the unlikeliest route the offender will have used. At external scenes, this may mean cutting a path through thick woodland to avoid any footpaths or tracks. You may also consider using planks laid on metal ladders to cross watercourses or any other means of access to a scene.

In an internal scene, best practice is the use of stepping plates. These allow investigators access to the scene without destroying any possible footwear evidence on the floor.

The CSM should undertake an initial assessment of scene. You should take time to get the full picture and consult with those who have information. Always remember to keep an open mind as your assessment must be an independent assessment of the scene and all hypotheses must be tested and no assumptions made. All observations, information received, and advice given must be recorded in the CSM log (*see* **Crime Scene Documentation**).

It is advisable to evaluate the extent and boundaries of your scene and plan accordingly.

When arriving at a strategy, a structured approach is critical. You must ensure that you have the correct resources and effective safety and welfare measures. Again, those decisions must be documented (*see* **Crime Scene Documentation**). At night, you should always consider securing the scene, pending daylight.

Before scene-processing activities commence, you must record the scene, using photographs and video.

A sketch plan or plans should be commenced and maintained as the processing progresses (*see* **Crime Scene Documentation**).

Secure fragile evidence to allow body removal, tape lifting, and fingerprint recovery from door frames, walls, etc. In many US jurisdictions, the deceased is the province of the medical examiner's or coroner's office and the body and its examination and recovery is their responsibility.

Where the police agency has sole responsibility for scene processing this can involve swabbing and recovery of fragile evidence from the body. You must always ensure that life has been pronounced extinct by a medical practitioner prior to any movement of the deceased. You should always ensure body recovery by CSIs to safeguard forensic evidence.

When all of the above has been completed, you should then begin the core process of crime scene investigation, the search for, identification, preservation, and recovery all types of contact trace evidence (*see* **Crime Scene Investigation**).

Throughout the scene processing, as a CSM must always ensure that techniques and decisions are within the areas of competence of personnel present and seek expert assistance where needed; serology, ballistics, blood spatter analysts, anthropology and cadaver, accelerant, and explosives dogs.

As a general rule, fingerprinting is the final process, as powder and chemical development could contaminate or destroy other evidence.

Search Strategies

Search strategies are dependent on various factors, including the operating environment. As part of your risk assessment (*see* **Crime Scene Investigation**; **Crime Scene Documentation**) you should consider any potential dangers, such as perpetrators returning to scene. The item may also be hazardous and you must consider how to preserve and recover it (see section Risk Assessment).

Methodologies. Most CSIs agree that there are five basic search patterns used for forensic purposes:

- **Circle or Spiral Search.** As described by many authors, it is a useful method where resources are limited or a specific item is being sought.

- **Strip and Line Search.** Strip search of a scene is a very effective method of searching a large area, such as fields or parking lots. The width of strip will be determined by the physical environment of the scene. Where necessary, string or tape such as those used to cordon a scene can be used to delineate the area a single person has to search. A strip search can be a useful method where human resources are limited.

  A line search is an ideal technique where there are sufficient personnel or the area concerned is on a slope. You must always remember to search downhill as any disturbed item will then still be within the area still to be searched.

- **Grid Search.** A variation of the strip search, it is a useful method to ensure small items are not overlooked. The scene is searched using the strip or line search technique, then re-searched at 90°. If on a slight slope, you should search across it first and then downhill. Start across at top of hill, then always down hill.

- **Zone Search.** This is a useful technique for searching areas such as cars, industrial areas, and multi-zone areas such as gardens. It is also an effective top tier division of extremely large scenes such as terrorist incidents.

- **Point-to-point Search.** This method is used to clear a path to the deceased, to enable the necessary processes to be undertaken, and to allow the body to be moved for autopsy. It is rarely used, except as a "flash" search, following the known or presumed route of suspects away from the scene. When fleeing a scene, offenders have been known to abandon, weapons, clothing, and even personal effects, in an effort to "distance" themselves from a crime.

Deciding on a Search Strategy. When deciding the type of search to undertake you have to consider the possible size of the object. If the size of the item is small or unknown, you need to consider a fingertip search. This is exactly what it says, hands and knees on the ground in a line. You may be searching for a spent cartridge case or spots of blood on leaves on a forest floor. This can be a very slow painstaking method, involving the careful movement or removal

of objects. No matter how time consuming, it can be essential to the evidence recovery process.

If the item is visible from head height, you have to consider if a line search would suffice, and if it could be above head height, the search of roofs gutters and down pipes needs consideration.

Where an item could be buried by soil movement, human, or animal activity, you should consider the use of ground penetrating radar.

A useful member of the team in this scenario would be a forensic archaeologist. If searching for human remains, the assistance of an anthropologist can be beneficial in determining if remains are human.

The physical environment of a scene inside premises such as stairwells, yards, elevators, and even fire and water damage can affect the processing of a scene.

The topography affects how a scene is processed and searched, as it may be difficult to traverse the terrain. If it is hilly, flat, marsh, pasture, thickly wooded, or brush you must be able to adapt your methods. If you are searching a wooded area, despite the obvious uncomfortable position, you should always search in a downhill direction as any item disturbed will move downhill and still be in the area to be searched. You may have to consider a double line approach in a wooded area. You would have two lines of personnel, one behind the other, the first line would search the ground and the second would scour the trees for obvious evidence.

Inclement weather can have a devastating effect on a crime scene. Rain can wash away evidence and light items can be blown away by high winds. You need to consider what measure if any you can take to prevent or lessen this. If necessary, you can the block drains to prevent some evidence loss but ultimately you will need to deviate from original plans and possibly recognized protocols. However, if you can justify these decisions and actions and document them in your CSM's log (see **Crime Scene Documentation**), any criticism will be removed or minimized.

Risk Assessment. In most developed countries, all work tasks must be carried out in a manner that is safe for any personnel and members of the public. In the United States, this is controlled by the Occupational Safety and Health Administration (OSHA).

In the United Kingdom, the Health and Safety Executive (HSE) has responsibility for ensuring compliance with the Health and Safety at Work Act 1972. Routine crime scenes can have generic risk assessments that cover low level risks that are normally encountered, such as broken glass, potential biohazards, and trip hazards (*see* **Crime Scene Documentation**). By their nature, any police operation is a fluid and major operations require dynamic risk assessments, continuously reviewed in the light of changing circumstances. Any CSM's log should include documentation to demonstrate the identified risks and the measures taken to eliminate or reduce them.

Threat Assessment. In the United States, it is normally recommended that at least one armed member of the agency remains at a major scene during the processing activity. Even in the United Kingdom, in rural areas, you should always consider that the offender may be living rough and that a search may reveal a camp or hide. Whatever the potential threat, a contingency plan, including an extraction plan should be decided upon and communicated to all personnel involved.

Resources. All of the above are affected by resource availability. The ideal equipment may not be available. Human resources are finite and the number required may not be immediately available. Both of these affect your strategy and mean that your strategies need to be flexible.

A very important consideration at any major scene is the welfare of personnel. Provision should always be made for an area where staff can relax for a short while and have some refreshment. Where a double cordon is in place the area between the inner and outer cordon should be used.

Exhibit Handling. To ensure integrity and security of recovered evidence, proper packaging is vital (*see* **Packaging and Transport**). As soon as an exhibit is recovered, it must be packaged, sealed, and the chain of custody documentation (*see* **Crime Scene Documentation**) completed. This should be double checked prior to removal from scene, handover to other agencies or exhibits officer. These simple steps ensure there is no possibility of contamination.

Quality Assurance Measures. To ensure that no potential evidence is missed, periodic reviews of

progress should be undertaken in consultation with the team(s) at scene and personnel involved in managing the investigation. Once the forensic examination of a scene is complete, where another trained CSM is available, a peer review walkthrough should be undertaken. This will minimize any possible oversights. This reflects good practice in homicide investigations as identified in the United Kingdom. Once this has been done, a walkthrough with the senior investigator highlighting activity is essential to confirm that all expected actions have been completed.

Once the examination has been confirmed as complete, a thorough search of the scene should be undertaken. Where possible, this should be carried out by a trained search unit, with a CSI present to photograph, recover, and document any additional evidence found. Where necessary an inventory of the scene should be completed. It should be noted that the examination and search are integral parts of a two-part operation and discovery of evidence during a search that includes dismantling or total removal of items does not mean it was missed by the CSI personnel.

Only when all activity completed can the scene be decommissioned and released. Where statutes permit, it is best to retain secure if possible, pending the completion of the investigation.

End Notes

ACPO, MIRSAP, Dovaston *et al.*, Derbyshire Constabulary

Related Articles

Footwear and Foot Impressions: Foot Impressions and Linking Foot to Shoe

ALLAN MATHIESON SCOTT

Crime Scene Photography: US Perspective

Introduction

The primary role of the crime scene technician is to document the context of the scene before it is altered or destroyed and to recover any physical evidence that may be present. These actions, if accomplished properly, ultimately assist the trier of fact in determining what occurred at the scene in question. In pursuing this goal, the crime scene technician accomplishes six basic activities. These include assessing the scene, making observations about the scene, documenting the scene, searching the scene for additional data or evidence, collecting all of the evidence located, and analyzing certain aspects of the scene (e.g., on-scene bloodstain pattern analysis or external ballistics).

Since every action taken by the technician has the potential to alter or destroy scene context, these six activities are accomplished in an order that places the least intrusive first, followed by the more intrusive actions. Assessing and observing the scene, in and of themselves, are not intrusive. They demand only that the crime scene technician move carefully in the scene, without disturbing physical evidence as they assess and make their initial observations. The subsequent steps of documenting, searching, collecting, and analyzing are quite intrusive; each demanding the crime scene technician interact directly with the scene. For this reason, scene documentation is always accomplished before any invasive search or collection activity.

The express purpose of crime scene documentation is to capture the scene context *in situ*, or as found. On-scene documentation involves several different elements including

- creating detailed notes on observations;
- capturing the scene context in photographs and video; and
- creating crime scene sketches.

Each of these elements of documentation is important to the overall investigation. No single element

stands alone, as each supports the other. Together they allow individuals who were not physically present at the scene to "see" and understand what the crime scene technician observed. From this documentation, theories of what happened are developed. When proper documentation exists, questions posed by the court about where an item of evidence was or in what condition it was found can be succinctly answered without hesitation. Without proper documentation anything goes. Counsel can make claims about items of evidence, with no functional means for the crime scene technician to refute or corroborate the claim. As important as crime scene notes and sketches are to this purpose, photographs of the scene are the most effective means of sharing what the crime scene technician observed there.

Basic Considerations for Crime Scene Photography

Photography of the scene is a critical element of crime scene documentation. The intent of the crime scene photographs is to lead the viewer from an overall perspective of the scene, to the details of specific items of evidence. Each photograph builds upon the next so the viewer is never lost or left wondering what they are viewing. The importance of capturing the scene context through the crime scene photographs cannot be overemphasized. A crime scene technician could spend an hour on the stand trying to explain some aspect of the scene; but using a single photograph, in a literal instant the court and jury are able to take in details such as relationships between items, color, and texture. Through photographs the crime scene technician allows the jury to see and experience the scene themselves.

To achieve this result, a standard photographic methodology is necessary. Such a methodology eliminates errors of omission and if applied appropriately produces functional and complete crime scene photographs. What is entailed in such a methodology? First and foremost it involves understanding and using three basic evidential photographs. These are the overall, evidence establishing and evidence close-up photo. Each photograph is discussed in detail later in the article, but the use of all three types of photographs allows the crime scene technician to capture scene context completely. Additionally, any practical methodology for crime scene photography should involve the following:

- The entire scene is photographed *in situ*, as soon as practical after arriving on-scene.
- Items of evidence considered to be fragile (e.g., footwear marks in dust, bloodstain patterns on the body) are documented completely before they can be damaged or altered.
- All observable items of evidence are photographed using the three basic scene photograph types before disturbing them in any manner.
- After completing a first series of photographs, photo/evidence placards are added and a second series of photographs are created. The placards allow the viewer to differentiate the various items of evidence from one another.
- If additional evidence is located in latter stages of the processing methodology (e.g., while collecting an item of evidence or when moving large items in the scene) the crime scene technician stops all activity and takes appropriate photographs of the new evidence. It is understood that these photographs may show the scene in an altered condition, but they are needed to functionally place the new evidence within the scene.
- For any on-scene analysis efforts, (e.g., a trajectory analysis of a bullet, defining the area of origin for a bloodstain pattern) the technician creates photographs that clearly demonstrate the result of that analysis. Once again, it is recognized that these photographs will show the scene in an altered state. Alteration of the surrounding scene is unimportant; the purpose of these photographs is to demonstrate the analysis result.
- A film roll reference card and photo log should be included in the documentation efforts. The film reference card is nothing more than a placard on which the technician writes where, when, and who produced the photos. The first photograph exposed on the media is of this card. The purpose of the photographic log is to keep track of the intended content of each photograph. **Crime Scene Documentation**

Composition of Crime Scene Photography, Understanding the Basic Types of Crime Scene Photographs

Composition is important in artistic photography, where the photograph is judged on its aesthetic value. Photographic composition is generally accomplished by organization of the space and items depicted

in the photo as well as consideration of how the various visual elements relate to one another. Artistic photographic composition involves various aspects including the following:

- **Contrast**

Contrast requires the photographer properly frame and light a subject in such a way that the background does not compete for attention.

- **Framing**

Framing requires simply that the subject of the photograph be captured completely in the field of view.

- **Simplicity**

A photograph should have one primary subject (e.g., a group of people, a landscape, or a butterfly). The value of a photograph with numerous visual centerpieces is often diminished by the resulting clutter and confusion it creates.

- **Rule of thirds**

The rule of thirds defines subject placement in the photograph. Applying the rule, the subject is always placed slightly off center rather than in the center of the viewfinder.

- **Viewpoint**

Viewpoint is a function of altering the camera position to create different perspectives. Lowering the camera height or aligning the camera's vanishing point with a skyscraper's lines are examples of altering viewpoint.

In artistic efforts, all of these composition elements are necessary to create aesthetically pleasing photographs. Crime scene photography is not the same. The purpose of the crime scene photograph is to accurately depict the scene without introducing distortion. For crime scene photography the concepts of simplicity, framing, and contrast are all applicable, but altering the viewpoint and the rule of thirds are not generally applied in crime scene work. For example, in terms of contrast, proper lighting of the subject of a photograph is always a challenge. The crime scene technician is limited in that they cannot alter the scene to change a background. Contrast and proper lighting are achieved through the use of fill-flash, off-camera flash, and bounce-flash. Simplicity is achieved by concentrating on the purpose of each photograph, asking "Why am I taking this photo and what do I want to show?" Simplicity requires a thorough understanding of the three basic types of crime scene photographs. Framing a subject is always important in the crime scene photograph, but as the technician is not concerned with aesthetic value, off center placement is rarely used.

To achieve composition simplicity and meet the underlying purpose of crime scene photography, each photograph must assist in leading a viewer through the crime scene and capturing the scene context. This is accomplished using three types of photographs. The three photographs are known as the following:

- *overall photographs;*
- *evidence establishing photographs; and*
- *evidence close-up photographs*

Overall Photographs

The purpose of the overall photograph is to capture the general condition and layout of an area. They are also effective in showing interrelationships between different areas of the scene. The crime scene technician seeks to show how the scene is oriented and capture major landmarks or focal points (e.g., doors, furniture items, bodies). Overall photographs are always the first photographs taken on-scene. Of all of the photographs created, overall photographs are the single best source to demonstrate that the scene was not intentionally altered by processing efforts. If someone questions some aspect of the scene (e.g., claims that the specific location of some item was altered by the investigators), overall photographs often show the location of the item when the initial processing began.

Overall photographs are generally exposed with a wide-to-normal angle lens and are taken from head height in a standing position (a normal human viewpoint). The use of a wide-to-normal angle lens (e.g., 28–50 mm) allows the photographer to capture a broad area in a single photograph, which is certainly helpful when one considers its purpose. The use of a 50–55 mm lens produces a photograph consistent with a human perspective and this lens is generally considered the standard for crime scene work. Nevertheless, wide angle lenses like the 28 mm lens can be used effectively in creating overall photographs without altering their value. The use of an extreme wide angled lens, such as a "fish-eye" lens is not recommended as it produces significant distortion. Although

a fish-eye lens allows a larger area to be captured in a single photograph, the level of distortion is so distinct that the photograph may not be allowed in court.

An effective method for creating overall photographs is to simply shoot the scene from various overlapping angles. In an interior scene the technician goes to a corner of the room or in an exterior scene to the perimeter and then takes a photograph across the area to an opposite corner or side (Figure 1). This process is repeated for each corner of the scene. The end result of the process is four or more photographs that provide overlapping coverage of the entire area. The photographs are shot at a distance, so small details may be lost (e.g., a shell casing may not be visible), but large items such as furniture, bodies, weapons, and other evidence items that are evident. Remember the purpose of the overall photograph is to capture the general condition of the scene as it was found. The four-corner method captures quite a bit of detail, but does not capture everything. If areas exist where specific evidence or initial conditions are not visible in the four-corner photographs, additional overall photographs of that area are required as well. As needed, the additional photographs are taken from whatever position is required for these odd areas; the sole purpose of which is to establish the initial condition and orientation of that area alone.

As the methodology described, overall photographs are typically taken in two series. During the first series, the crime scene technician does nothing to the scene. No scales or photo placards are introduced. This first series is taken as soon as possible after arriving on-scene and certainly before any significant action by the crime scene team. These overall photographs are used to demonstrate scene integrity and eliminate any claim that the scene, as shown in later photographs, was altered from its original condition. The initial series of overall photographs always depict the scene prior to any invasive processing activity.

A second series of overall photographs is exposed after an assessment of the scene. At that point various items of evidence are known to the technician. Photo placards are introduced into the scene adjacent these known items of evidence and the second series are shot in exactly the same way (e.g., corner to corner) (Figures 2–5). The introduction of the photo placards increases the value of the photograph. The placards allow the viewer to recognize that items that are not visible (e.g., the example of the shell casing) are, in fact, present in the same location.

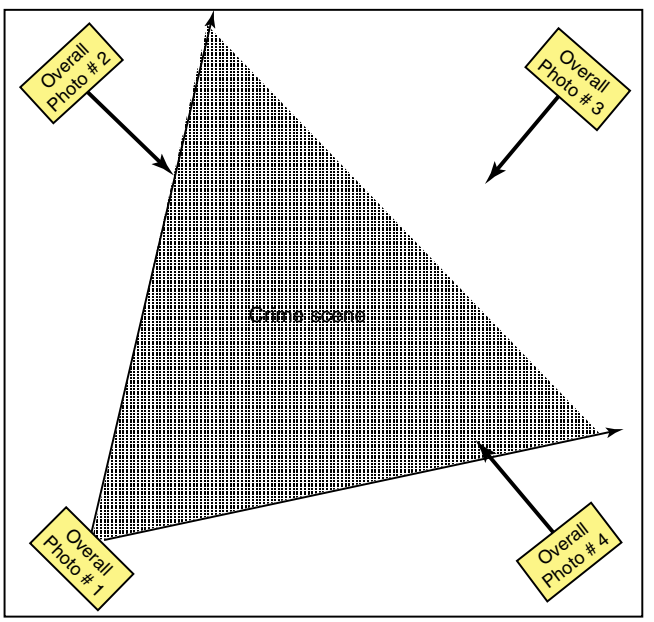


Figure 1 Overall photographs are most effective when they capture overlapping aspects of the scene. The easiest method to accomplish this is to take a photograph from each corner of the room or scene. The four resulting photographs will effectively capture the full scene context



Figure 2 Figures 2–5 demonstrate the concept of the overall photograph in an exterior scene. In this first photograph, the camera is oriented looking down the hill (SW-NE) into the scene. Note the trailer hitch in the foreground. Its inclusion is purposeful and will serve to orient other photographs back to this one



Figure 3 This overall photograph is taken looking NW to SE in the scene. The evidence flags are already in place to help illustrate where items of evidence are located

Another use of the overall photograph is to demonstrate the relationship between various areas in a scene. For scenes involving multiple rooms or dispersed areas outside, the use of a properly framed overall photograph allows the viewer to understand how one area relates to another (Figure 6).

Evidence Establishing Photographs

The purpose of the evidence establishing photograph is to frame an item of evidence in relationship with a known landmark in the scene. This is particularly important when we consider that many items of



Figure 4 This overall photograph is shot from the bottom of the hill looking NE–SW, back toward the position Figure 2 was taken from. Including the trailer hitch in Figure 3, allows the viewer to clearly orient this photo back to that photograph



Figure 5 This overall photograph is taken looking SE–NW across the scene. Beyond the presence of the evidence items, the pink fluorescent tape on the right side of the photograph was also visible in Figure 3. This is another landmark that allows us to orient the two photographs to some extent

evidence (e.g., bullet holes, bloodstains, and shell casings) as well as the surfaces on which they occur often look quite similar in close-up photographs. Without the evidence establishing photograph, the viewer is often confused by the close-up photograph, unable to understand what surface they are viewing (Figure 7). In combination with the evidence close-up photograph, the evidence establishing photograph allows the viewer to clearly orient themselves as to what item they are viewing and exactly where it is in the scene (Figure 8).

This particular photograph is often described as a mid-range photograph in which the crime scene technician is advised to shoot an overall or distant photo, followed by a midrange photograph and then finally a close-up photograph. Range or distance



Figure 6 Overall photographs can be used to orient separate areas of a scene as well. The wooded site offers little that allows us to know where it is. This photograph is taken at the top of the hill looking back at the adjacent residence. By framing the trailer in the photograph, when considered with Figure 5, the viewer has a clearer understanding of where the wooded site is in relation to the residence



Figure 7 Close-up photographs rarely assist the viewer in understanding where an item of evidence is. Items such as bloodstains, bullet casings, or bullet holes are often found on similar looking backgrounds; which only adds to the confusion. The function of an evidence establishing photograph is to remove this confusion. Consider Figure 8 in combination with this evidence close-up photograph

is not the issue and thus the term mid-range is misleading. The purpose of this photograph is to frame an item of evidence in relationship to a landmark in the scene. How that is accomplished is based on the context of the scene. Depending upon the context, an evidence establishing photo might be taken from 2 or 10 ft. That decision is based solely on the landmarks available. If the photograph



Figure 8 An evidence establishing photograph for Figure 7. This photograph captures the position of the various bloodstains including Stains #5 and #6. Together the two figures allow the viewer to see both the detail of the stains and understand where in the scene the stains are. Combined with an overall photograph of the scene that shows the door, one can quickly and easily understand the full context of where these stains were located

is approached arbitrarily from a distance aspect, then each establishing photograph would be shot from an arbitrary distance (e.g., halfway between the overall and close-up). No matter what term is used to describe this photograph, its purpose must be understood and met; it is taken singularly to identify where the item is in the scene. This photograph bridges the gap between the overall and evidence close-up photographs, allowing the close-up photographs to be understood in relation to the scene.

A 50–55 mm lens is usually adequate for shooting evidence establishing photographs. The viewpoint (camera position) used in the evidence establishing photographs can vary widely; however, avoid interjecting unnecessary angles that may distort the photograph or confuse the viewer. The importance is to frame the landmark and evidence item together in such a fashion that the viewer can identify each.

In small or uncluttered scenes, the second series of overall photographs with the placards in place often serve effectively as the evidence establishing photographs (Figure 9). If this is the case, additional evidence establishing photographs are unnecessary. In complex and larger scenes, the overall photos are not enough. Whatever the situation, the addition of the photo placards in the photograph enhances the viewers understanding. Photo placards are available from numerous sources and although commercial placards are relatively cheap, there is no requirement that the technician use a commercial product. Disposable placards are easily created from index cards and a bold marking pen. An additional technique for exterior scenes involves the use of grading flags and plastic lettering kits. Large adhesive plastic letters or numbers are attached to the grading flag and the grading flag is placed into the soil adjacent to the item of evidence (Figure 10).

Once the photo placards are in place, items should not be arbitrarily renumbered at some later point in the report, evidence voucher, or on the sketch.



Figure 9 In uncluttered scenes, the introduction of photo placards in the overall photograph may eliminate the need for a specific "evidence establishing" photograph. This overall photo effectively serves both purposes for items 1–5

During the initial placement, choose a numbering and lettering scheme that work throughout the entire crime scene processing and allow for the addition of as-yet discovered items. A simple method is to use a combination of alpha-numeric associations. Numbers are assigned to major items of evidence (e.g., item 1 and item 2) and if subassignment is necessary letters are used in conjunction with the number (e.g., 1a, b). Changing or using duplicate numbers or letters in subsequent documents or photographs create confusion for those viewing the documentation.

Evidence Close-up Photographs

The purpose of the evidence close-up photographs is to show detailed aspects of a particular item of evidence. As the name implies the photograph is taken from a distance that is as close as possible to the evidence. Far too often these photographs are taken from a distance of 3 or 4 ft away, when the subject of the photograph is the size of a coin (Figure 11). Such effort fails to capture any detail, yet they are still represented as "close-up" photos. The viewpoint for the evidence close-up photograph is very simple. Proper close-up photographs are achieved by placing the camera at a distance that allows the evidence to fill the frame of the viewfinder. Whenever possible position the camera, so the film plane is horizontal to the primary surface being photographed. After items are collected from the scene, additional close-up photos may be necessary as well (e.g., to show a fingerprint or small bloodstain on a murder weapon) (Figure 12). When creating these after-collection photographs, the technician should consider that recent advances in digital cameras now provide the means of photographing evidence that in the past was simply impossible to capture. Cameras like the Fuji S3 Pro UVIR allow the technician to document items like bloodstains on dark backgrounds, gunshot residues, bruises, and biological stains (Figure 13).

The best lens for creating a close-up photograph is either a 55-mm lens or a macrolens manufactured for that specific purpose. Macrolenses typically range from 50 to 105 mm, the latter often offering 1 : 1 aspect ratios. Proper lighting of the close-up photograph often requires effort. External strobes controlled in automatic mode by the camera function fine when taking overall and evidence establishing photographs. In close-up work, the camera body almost always ends up in close proximity to the surface being photographed. In this situation the automatic flash, more often than not, bleaches out surface details. Off-camera strobes held at an angle to the surface being photographed or adjustable beam strobes that allow the flash to be bounced off ceilings or adjacent surfaces are the primary way to properly light the



Figure 10 In an exterior scene grading flags can be used as an evidence placard. The grading flags are available from any hardware store and adhesive plastic lettering is found in any office supply store. The lettering is added as needed and the flag, which is on a metal rod, is inserted into the soil adjacent the evidence

close-up photograph. An additional method for lighting the close-up photograph is the use of a ring light.

Evidence close-up photographs without a scale of reference are not absolutely necessary in terms of investigative value; however, in the past, lawyers have argued that placing a placard in the photograph materially alters the essence of the scene. To prevent such attacks, close-up photographs can be shot along with the overall photographs without placards (Figure 14). Keep in mind that nothing is altered or manipulated in the process. For instance, it would be inappropriate to take close-up photographs of both sides of a revolver in the scene and then replace it and then take a second series with the photo placards. No matter how effectively the technician replaced the item, the second series would not be a true representation of the original condition of the revolver.

During the second series of photographs, placards are placed in close proximity to each item of evidence (Figure 15). To get both the detail of the item and include the placard number (the identity of the evidence), two approaches may be necessary when shooting photographs of small items of evidence. In some instances the placard itself may be repositioned to functionally place it as close as possible to the evidence. In other instances, particularly when shooting photographs of very small items, a smaller scale may have to be introduced to get as much detail as possible. In this instance, the small scale is annotated with the appropriate placard number. The

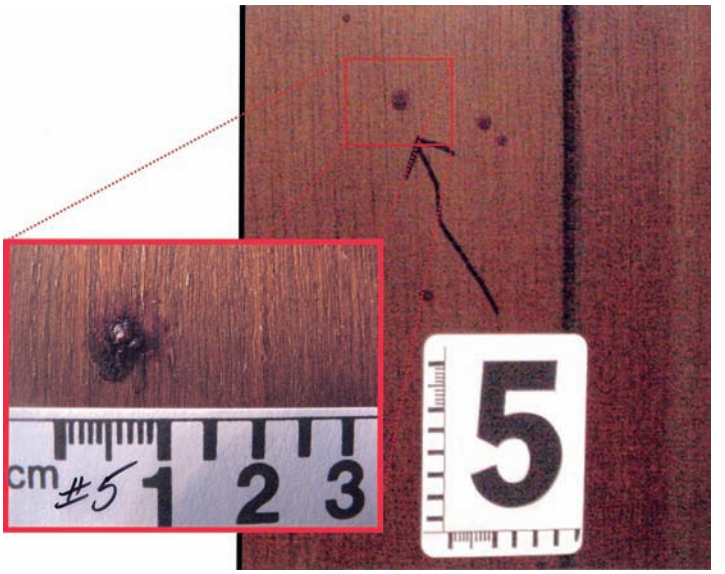


Figure 11 The purpose of an evidence close-up photograph is to show detail. The photograph on the right is typical of those represented as "close-up", but note that the detail of the stain in question cannot be seen. The inset photograph is framed much closer and now the small details (e.g., the bubbles in the stain) are evident. Close-up photographs should be framed as close as possible

scale of reference used for any given close-up photo may change, but a close-up photograph that fails to include some scale of reference is of limited value to the viewer. Placards, ABFO scales, or small rulers can all be used for this purpose.

Once the full context of the evidence is documented (e.g., observed in situ, detailed notes created, overall, evidence establishing, and close-up photos taken of the item as it lay in the scene, and the item's position is measured and sketched) the item can be collected and or manipulated without losing information. Additional close-up photographs are created when the item is collected. The item is carefully handled and photographed to capture all aspects and surfaces. Care is always in order when accomplishing this to prevent the loss or alteration of additional evidence (e.g., fingerprints and DNA). These additional photographs are often created after returning to the office, where it is easier to control lighting and the background environment. For this purpose, the technician should have an area at their office with a white nonglare background and off-camera lighting.

Using the three basic types of photographs, each item of evidence and all general aspects of the scene are captured in their unaltered condition. Once this is accomplished, the crime scene technician can move on to the remaining processing steps (e.g., sketching the scene, searching for additional evidence, and collecting all of the evidence). At any point in these efforts, if additional evidence is located or some aspect of the scene is discovered that has not been documented, the crime scene technician stops the ongoing activity and takes the necessary photographs. This is particularly true of on-scene fingerprinting efforts. Developed fingerprints should be documented *in situ* and then collected. This requires an evidence establishing and close-up photograph of each. If on-scene analysis is pursued in some fashion (e.g., bullet trajectory analysis, area of origin for bloodstains) the results of these efforts have to be photographed as well. These photographs are approached in much the same manner as any evidence photograph; however, these situations may require a little more composition effort to properly frame and demonstrate the analysis result.

Camera Control

There are two primary reasons why crime scene photography fails to achieve its intended purpose. The first reason is a failure to understand and apply the

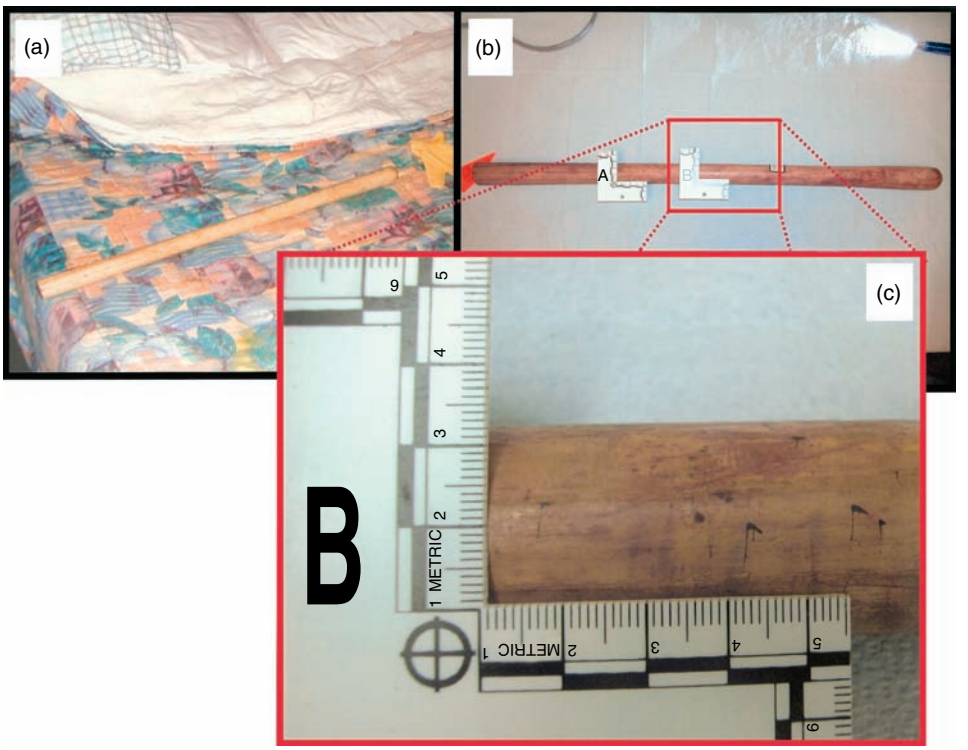


Figure 12 After collection at the scene, additional close-up photographs are often required. Note that in photo(a), the crime scene photograph, we can see the location of the wooden club, but its details are not evident. Photograph (b) was taken after collection, with the additional placards (a and b) added to show items of interest are on the club. Photograph (c) shows a close-up of the stains found at location (b)

composition aspects. The crime scene technician must understand the nature and purpose of each of the three basic crime scene photograph types (overall, evidence establishing, and evidence close-up). A second reason for poor crime scene photography is the inability to control the camera properly.

To create quality crime scene photographs that accurately depict the scene and have value to subsequent viewers, the technician must do the following:

- control camera stability;
- control light coming into the camera; and
- control camera focus and use depth of field effectively.

Camera Stability

Proper photography always begins with a stable camera. If the camera is in motion while the shutter is open, the result is blurring and loss of focus. Camera stability is accomplished by using a two hand grip or through the use of a tripod. A two-handed grip is more than adequate for a conventional camera using a 28–55-mm lens at any shutter speed of 1/60th or shorter. A shutter speed of 1/250th of a second is possible when operating with a lens of 100 mm or more in a handheld mode [1]. If the situation demands the technician to operate the camera with a longer shutter speed than those described, a tripod is necessary. With the advent of the digital single lens reflex (SLR) camera, the range of shutter speed for handheld operation is broader. Depending upon the camera, the technician may find they can functionally operate the digital camera in a handheld mode at much longer shutter speeds. In close-up work, no matter what the conditions or equipment, the use of a tripod enhances the quality of the resulting photos. The only downside to using a tripod for every close-up photograph is the increased time required to take



(a) (b)

Figure 13 Advances in digital photography allow the crime scene technician to capture details that in the past could not be captured. The red colored sweat pants in Photo (a) have bloodstains present on them. On the red background they can barely be visualized. Photograph (b) was taken using Fuji's S3 Pro UVIR. Using the cameras IR mode the stains are evident, but the conflicting background color is removed. This capability is particularly helpful when dealing with bloodstains on dark clothing. Photographs courtesy of Jason Guffey, Program Developer-General Forensics, Central Piedmont Community College, Charlotte NC

the photos; but the results are worth the trade-off of time spent on-scene.

Light Control

Proper composition in both artistic and evidential photographs requires proper lighting. Owing to the emphasis on close-up work in evidence photographs, improper lighting often leads to poor quality photographs. Use too little light and the picture is dark and details of the evidence can't be seen. Use too much light and the subject is bleached out with loss of detail as well. There are three techniques of lighting that aid the technician in creating properly exposed photographs.

The technique known as *fill-flash* is one approach. Without consideration for on-scene lighting conditions, all photographs are shot with the flash. What fill-flash achieves is consistent lighting of all surfaces in the photograph. In the past, manual flash systems required adjusting the flash settings for each exposure when using the fill-flash technique, which also increased the amount of time spent on-scene. Now most digital cameras have the ability to read the level of lighting and adjust the level of the flash accordingly. When shooting close-up photographs fill-flash is important as well, but the distances involved



Figure 14 Evidence close-up photographs are often taken both with and without a scale of reference or evidence placard. The only purpose of the close-up photo without a placard is to prove that the context of the scene was not altered by introducing the placard



Figure 15 An evidence close-up photograph for Figure 15, including a scale of reference. Many evidence placards come with a scale on them. This allows the viewer to identify it (e.g., item E) and understand its size. A close-up photograph without some form of scale of reference has limited investigative value

require the use of additional techniques to prevent overexposure.

One technique for preventing overexposure is the use of a flash diffuser. These devices generally snap over the flash facing and are made of an opaque white plastic. They diffuse the flash strobe creating a gentler more effective flash when the subject is close to the camera. An alternative technique is to use an

off-camera flash. The flash is connected to the camera by a sync cord. The camera lens is positioned as close as required for the subject matter, while the flash is held obliquely to the subject and at a greater distance (e.g., very often at arms length) (Figure 16). A final method for eliminating overexposure in photographs is to bounce the flash. Many flash systems have a flash facing that rotates from a normal 90° position up to a vertical position. By altering the angle of the flash facing, light is bounced off adjoining surfaces. At intermediate distances, bounce-flash effectively illuminates the subject and eliminates overexposure; however, for extreme close-up photographs, where the distance is less than 12 in., bounce lighting tends not to be effective.

Paint-with-Light

At night or in low light conditions, when shooting overall photographs a single strobe no matter how strong it is, may not be enough to achieve sufficient flash (Figure 17). In these instances, the technician should apply the technique known as "*paint with light*". Paint with light uses multiple flashes along with an extended exposure time to adequately light a scene and capture detail. The most typical circumstance requiring paint with light is when taking an overall photograph of an exterior scene at night, particularly when evidence is widely scattered across the scene. A single flash does not sufficiently light the scene from the camera position, resulting in poorly illuminated photographs. To paint with light, the camera is mounted on the tripod and once focus is set, the camera is placed on a bulb (B) or extended time setting. On this setting, as long as the shutter release is pushed, the aperture of the lens remains open. On a conventional film camera, a shutter release cable is used to keep the shutter open. On a digital SLR camera, there is an electronic shutter release cable or a menu option that allows the shutter to remain open. Whatever the particular mechanism, once the shutter is open the technician moves to positions on either side of the camera. There they manually initiate the flash toward the subject matter. The combination of the multiple flashes provides for a detailed photograph (Figure 18). If the area in question is in total darkness, the technician can walk into areas contained in the viewfinder, allowing them to initiate the flash from several angles but at a position closer to the evidence than the camera. To prevent any ghost image from appearing in this situation, the technician must keep the flash in front of them and always point it away from the camera. Paint with light works effectively during lowlight conditions as well, making efforts at dawn or dusk more effective. However, low light conditions demand the technician to remain outside of the area in the camera viewfinder when positioning the flash. If they fail to do this, the limited light present is usually enough to create a ghost image of the technician.



Figure 16 Particularly when taking evidence close-up photographs, it may be necessary to use an off-camera flash. The technician places the camera lens as close as required, while holding the flash at a greater distance

Focus and Depth of Field

Most digital SLR cameras in use today are automatic. The camera's computer determines the optimum focus, depth of field, shutter speed, and even adjusts the flash as needed. But understanding the effect of the depth of the field is still important. On occasion, the crime scene technician may have to use the camera in a manual mode. The subject matter of



Figure 17 Depending on the flash used and the size of the scene, when taking overall photographs at night the flash unit may not adequately illuminate the scene. In these instances, the technique known as *paint with light* is very effective. This photograph was taken on a tripod using an on-camera flash, compare it to Figure 18



Figure 18 This photograph was taken at the same time as Figure 17, but uses the technique known as *paint with light*. The camera was set on a tripod and the lens opened to the bulb (B) setting. The flash was manually initiated in three positions, first at the camera then forward of the camera to the right, and then once again to the left. The result is a well exposed photograph in which scene detail is evident

the photo is always the primary focal point of the picture, but there may be objects in front of or behind it, that are of interest as well. Depth of field is a description of how in focus objects in front of and beyond the subject are. This is a critical concern when creating the evidence establishing shot, where both the landmark and evidence item need to be in focus. Depth of field is controlled by changing the

f-stop setting of the camera. The f-stop defines how wide the lens aperture opens, which, in turn, controls how much light is used in the exposure. As the f-stop number increases the aperture becomes smaller, resulting in the introduction of less light. By limiting the amount of light for any given exposure (e.g., increasing the f-stop) there is a corresponding increase in the depth of field for the resulting photograph (Figures 19 and 20). For close-up photographs, a short depth of field is usually not an issue, so using a lower f-stop will not hurt the resulting photo. When shooting both overall and evidence establishing photographs the technician wants as much depth of field as possible. The more items and areas in focus in an overall or evidence establishing shot, the greater the value of the photograph to the viewer.

Focus is an obvious concern. Achieving focus for the most part is determined by the technician. Although many digital SLR cameras come with automatic focus systems, in many crime scene instances this function must be turned off. Too often the camera computer chooses the wrong item to focus on, or the focus may continually adjust when taking close-up photographs. Under these circumstances the only solution is to use a manual focus. Manual focus begins by understanding how the camera's focus system works. There are various manual focus systems available for SLR cameras. The most effective system for crime scene work is the split image. When the focus ring on the lens barrel is turned, the viewfinder "splits" the image into three parts. The top and bottom thirds in the focus circle remain aligned, but movement of the lens barrel causes the middle third to move in and out of alignment. By slowing and turning the lens barrel focus ring, the middle portion of the image is brought into alignment with the two outer areas. When the three sections are aligned the subject is in focus. Split image focus can be used effectively with photo placards and scales. The technician positions the camera at the distance they require, but instead of focusing on the evidence itself, they use the straight thin lines of the ruler to achieve an optimum focus. Once optimum focus is achieved, the evidence is centered back in the viewfinder and the photograph taken. This technique is effective when focusing on objects that lie close to or in the same plane as the ruler, but the technician must take depth of field into consideration. If the evidence itself extends any distance beyond the plane where the ruler is located and a low f-stop is in use, the entire item may not be in clear focus in the photograph.



Figure 19 A photograph taken to demonstrate the effect of f-stop. This photograph was taken at f3.5, with the focus set at the D placard. Note that the tree in the foreground and placard are in focus; however, the background items appear fuzzy



Figure 20 A photograph taken from the same position and with the same focus point as Figure 20. In this instance the f-stop was set at $f18$. Note that the background details are in focus as well. As we increase the f-stop (decreasing the amount of light entering the lens) we increase the area in the fore and background that appears in clear focus. Turning off the automatic mode of the camera and manually controlling f-stop can be helpful when taking overall or evidence establishing photographs

Video Photography

The use of the video camera at the crime scene can certainly supplement the crime scene documentation, but video footage does not replace standard crime scene photography. Video is an excellent tool for establishing scene integrity. After creating the initial series of crime scene photographs, the technician completes a nonintrusive walkthrough of the scene using the video camera. If at some later stage of the investigation, someone claims alteration of the scene by the crime scene team, the video footage along with the initial crime scene photographs should allow the team to establish nearly all of the initial conditions. The video camera can also be used in later stages of processing, documenting actions, and analysis efforts such as trajectory or bloodstain pattern analysis.

When using the video camera, a few added precautions are in order. The first concern is that the video camera does not function like the human eye. Humans can quickly avert their eyes from one area to the next, producing a seamless stream of images that remain in focus. If the video camera lens is moved in the same fashion from one focal point to the next, the resulting footage is jerky and difficult to watch. When panning with the video camera, eliminate any abrupt movements and keep the camera focused on areas of interest for more than just a few seconds. Any movement of the camera must be deliberate, panning slowly from one point to the next. Plan the panning effort in advance; don't pan back and forth from one side of a room to the other. Move in a methodical and logical fashion across the room or area in question. Once the camera is on an area of interest, obtain a good focus and keep the camera in place. An effective technique is to quietly count to 10, before moving the camera to a new point of interest. The use of the zoom lens should be limited and controlled as well. Treat the zoom function the same as panning with the camera; use a slow and deliberate effort when transitioning between one zoom angle to the next. Once the zoom position is achieved, maintain the camera in that position for at least 10 s. Just as lighting was important in still photography, lighting for the video camera is important as well. The human eye may be able to see effectively in low light, but video footage shot

under the same conditions without a strobe is often useless. Always use a strobe with the video camera unless outside and in daylight. A flashlight used as additional lighting for the video camera does not suffice.

A final precaution when using the video camera is to turn off the audio. There is rarely a reason to capture audio when shooting the crime scene. Far too often, comments that were unintended or inappropriate are caught on the audio track. The easiest method of turning the audio off is to insert a microphone jack (without the microphone attached) into the external microphone plug of the video camera. Beyond these considerations, video footage of the crime scene is approached in much the same fashion as still photography. Know what you're shooting and why you're shooting it. Then control the camera, the lighting and be sure to get a good focus.

Summary

Crime scene documentation consists of detailed notes, sketches, and photographs. Each of these three elements has a purpose but crime scene photographs are critical. Using the three basic photographs, the overall, evidence establishing and evidence close-up photograph, the crime scene technician leads the viewer through the scene from an overall perspective to the small details. Overall photographs provide overlapping coverage of the entire scene, showing major landmarks and items. Details of specific items of evidence are captured with the evidence close-up photograph, where the viewfinder is filled with the item of concern. Bridging the gap between the overall and close-up photograph is the evidence establishing photo. This photo is taken in such a manner that the evidence is framed with a recognizable landmark in the scene. The evidence establishing photograph prevents the viewer from becoming lost in the scene. Together these three basic photographs allow the technician to functionally prove to a jury or judge the actual conditions and context of the scene, as the technician viewed it. They are the heart of crime scene documentation and critical to establishing the integrity of both the scene and the evidence removed from it.

Reference

[1] Walt, S. (1979). *Surveillance Photography Guides*, U.S. Dept. of Justice, Federal Bureau of Investigation, Washington, DC, p. 7.

Ross M. Gardner

Crime Victims' Decision to Report Crime

The police rarely discover a crime on their own. In the vast majority of cases they learn of a crime via notification by the victim [1]. This has led researchers to conclude that crime victims are the "gatekeepers of the criminal justice system" [2]. Indeed, without notification by crime victims, few crimes would reach the attention of the police. The discretionary power accorded to victims is reflected by the consistent finding that fewer than half the nonlethal crimes (41%) that occur in a given year are reported to the police [3]. Of the estimated 18 million property crimes (e.g., burglary and theft) committed in 2005, just 40% were reported to the police [3]. In comparison, 47% of the estimated 5.2 million violent crimes (e.g., robbery, rape, and assault) committed in 2005 were reported to the police.

These figures derive from an ongoing telephone survey of 77 000 households involving 134 000 individuals conducted jointly by the Department of Justice and the Census Bureau since 1972. Known as the *National Crime Victimization Survey* (NCVS), the survey is considered to be the most reliable source of information about reported as well as unreported crime in the United States.

Findings from the NCVS reveal that, with few exceptions, features of the crime and its circumstances are better predictors of reporting than characteristics of the victim. Thus, with regard to violent crime, reporting of aggravated assaults (62%) is greater than reporting of robbery (52%) and rape/sexual assault (38%). With regard to property crime, motor vehicle theft (83%) is more likely to be reported than burglary (56%) or theft outside the home (32%). Further, the more severe the harm, the

higher the rate of reporting. Thus, higher reporting is associated with completed versus attempted crimes as well as crimes involving greater financial loss and injury [4]. Rapes and sexual assaults committed by strangers are more likely to be reported than similar crimes committed by nonstrangers.

However, the victim–offender relationship does not appear to affect the reporting of other types of crimes. With few exceptions, reporting does not appear to be a function of victim characteristics: the exceptions being that women and blacks are more likely to report violent victimizations to the police than men and whites [4].

Making Sense of the Data: Theoretical Approaches

The rational choice perspective is the dominant theoretical approach to explaining crime victim decision making [5, 6]. This approach assumes that when deciding whether or not to call the police, victims rationally calculate the potential benefits and costs of each option and then choose the one that appears most "profitable". Thus, victims who incur injury or large financial losses may conclude that they have more to be gain by notifying the police than by not doing so. Moreover, they may reason that the police will be more successful in apprehending the suspect because they will invest more effort in solving such crimes.

This perspective can also account for why rape and sexual assault victims who are acquainted with the perpetrator are less likely to report the crime than those who are attacked by a stranger. Victims who are acquainted with their attacker may feel that the costs of reporting outweigh the benefits. They might view it as a personal matter or not report out of fear reprisal from the perpetrator. In addition, they might reason that because of their prior acquaintanceship with the perpetrator, the police might question their credibility and therefore not vigorously pursue the case.

While the rational choice perspective has intuitive appeal and has proven useful in explaining some aspects of victim decision making, recent research suggests that modifications are necessary. First, researchers have come to recognize the importance of emotional factors in decision making [7, 8]. This may be particularly true for crime victims who, in the aftermath of their victimization, experience intense negative emotions of fear, anger, and sadness [9–11]. Such intense emotional experiences can impair the rationality of their decision making. Research suggests that emotional arousal can influence crime victims' attention, perceptions, interpretations, judgments, their ability to process information, and their processing strategies [12–14] (*see* **Posttraumatic Stress Disorder**; **Rape Trauma Syndrome**).

Second, interviews with crime victims reveal that they often do not make their decision in a social vacuum. A substantial number of crime victims consult with family, friends, and bystanders when deciding whether or not they should call the police and that such contacts are a key factor in their decision making [15–17].

Third, the decision process of victims is not a single "yes" "no" decision, but rather represents the culmination of a series of decisions or stages [16]. In the following paragraphs, a broad integrative rational choice model that incorporates these three qualifications is presented.

The Greenberg–Ruback Model of Crime Victim Decision Making

Greenberg and Ruback propose a three-stage model of victim decision making [16]. In order for victims to report a crime, they must first label a suspicious event a crime; second, they must decide that it is sufficiently serious; and third, they must decide that notifying the police is the best course of action. The model proposes that victims may not call the police because they fail to define the event as a crime, because they do not view it as sufficiently serious, or because they conclude that calling the police is not the best option. The model further proposes that the decisions at each stage are subject to influence by emotional and social factors. Let us examine the model more closely.

Stage 1: Defining the Event as a Crime

Whether or not a suspicious event is defined as a crime depends on how closely the event matches the victim's personal definition of a crime. For example, victims of attempted crimes may not view themselves as having been victimized despite the fact that such acts are legally defined as criminal. Similarly, victims

of date rape may not view the event as a rape since it does not match the classic definition of an attack by a stranger [18]. That an individual has a personal definition of a crime does not guarantee that the definition is psychologically accessible. Victims of crimes that occur in unexpected circumstances may be slow to label the event as a crime because noncriminal labels are more readily accessible [19]. Such unexpected circumstances could include crimes committed in broad daylight or by trusted others.

Alternatively, circumstances and emotions can enhance the accessibility of crime labels such as when one is alone at night on a dimly lit street and/or when one feels particularly fearful. Social influence can play a role in how an event is labeled. This is particularly true in discovery crimes, such as burglary or theft, where victims are not typically present during the commission of the crime. In these instances, victims engage in information-gathering activities, such as consulting with neighbors, friends, and coworkers, in order to help them define event. Such consultations are a major source of delay in police notification [20, 21].

Stage 2: Determining the Seriousness of the Crime

Research shows that one of the best predictors of victim reporting is the perceived seriousness of the crime [6, 16]. Across a wide variety of crimes, the more serious victims perceive their victimization to be, the more likely they are to call the police. According to the NCVS, one of the reasons most often cited by victims for *not* reporting the crime is that "the crime was not important enough" [22]. The model proposes that perceived seriousness is directly related to the magnitude of distress – the more serious the victimization, the greater the distress. Perceived seriousness depends on the combined effects of two sources: (i) victims' sense of being "wronged" by the attack and (ii) their feelings of vulnerability to subsequent victimization. The sense of being wronged involves beliefs of injustice and inequity with accompanying feelings of anger, resentment, and, sometimes, revenge.

Feelings of vulnerability are accompanied by the emotion of anxiety and fear. Consistent with this reasoning, research shows that anger and fear are the emotions most frequently reported by crime victims [16]. Studies show that beliefs about being wronged and vulnerable depend on three features of the crime: its perceived unexpectedness, the magnitude of subjective harm suffered by victims, and their perception of potential harm in the situation. The sense of being wronged is greatest when the crime is unexpected, subjective harm is severe, and the potential harm is low. Feelings of vulnerability are enhanced when the crime is unexpected, subjective harm is severe, and potential harm is great [16]. To the extent that victims judge the crime not to be serious, they will be less inclined to call the police. Victims' perception of being wronged and being vulnerable to future harm are subject to social influence. Thus, others in their interactions with victims can either amplify or decrease such feelings.

Stage 3: Deciding What to Do

Having labeled the incident a crime and evaluated its seriousness, victims next must decide what action to take. The model proposes that they can choose from among four broad options: (i) dealing with the matter privately, (ii) cognitively reevaluating the situation, (iii) notifying the police, or (iv) simply deciding to do nothing. From a rational choice perspective, victims will choose the option that from a costs/benefits analysis appears to yield the best outcome. Presumably, this involves selecting the option that they believe is most likely to reduce the distress deriving from their feelings of being wronged and vulnerable.

To the extent that a particular option successfully ameliorates their distress, victims will have less incentive to avail themselves of other options. This has clear implications for why many victims fail to call the police. If victims can reduce their distress via private solutions or by restructuring their perceptions of the situation, they will have less incentive to notify the police.

The emotional distress typically experienced by crime victims impacts their decision making in two ways: first, the stress is likely to make their decision fall short of a perfectly rational decision; and second, it increases the victims' susceptibility to social influence. With regard to the first impact, research shows that severe stress interferes with the decision makers' ability and patience to conduct a rationally exhaustive costs/benefits analysis of available options [12–14]. High stress has been found to narrow the focus of

attention, to impair the assimilation of information, and to restrict the range of considered options. Under such conditions, decision makers tend to opt for a less cognitively demanding strategy that meets some standard of *sufficiency* rather than employing a more cognitively demanding *maximizing strategy* [23]. The maximizing strategy involves a detailed examination of the costs and benefits of multiple options. One might say that under highly stressful conditions, victims' decision making may best be described as "semirational".

Second, research shows that under stress, victims are highly amenable to social influence. As Bard and Sangrey [24] noted, "A crime victim's entire structure of defenses becomes weakened under the stress of violation, leaving him or her unusually accessible to the influence of others" (pp. 37–38). Studies have demonstrated that immediately after the crime, a substantial percentage of victims turns to others for support, information, and advice, which has a decisive impact on their decision making [15–17]. This is true for victims of violent crime as well as property crime. Moreover, the data clearly indicate that victims tend to follow the advice offered them. It is noteworthy that social influence is not confined to the third stage of the model, deciding what to do, but, as previously stated, can apply to how victims label the event and how they evaluate its seriousness.

Option 1: Dealing with the Matter Privately. Rather than calling the police, victims can reduce their sense of being wronged by a variety of private means. In some cases where the perpetrator's identity is known, victims have sought private vengeance by individually retaliating, or by retaliating in concert with others (vigilantism) [25]. Property crime victims can threaten to harm the perpetrator in order to obtain compensation. They have also sought compensation from third parties (e.g., insurance companies) or have filed a civil suit against responsible third parties (e.g., a hotel for failing to maintain adequate security) [26]. Alternatively, some have obtained compensation by stealing from others [27].

Victims can also employ private solutions to deal with their sense of vulnerability. They can take actions to reduce potential perpetrators' *ability* to harm them. For example, burglary victims often "harden the target" by purchasing an alarm system, dead bolt locks, and window bars [28]. Other private options that have proven effective in reducing fear of crime include purchasing a weapon, participating in a self-defense course or a neighborhood watch program, and changing one's residence from a high- to a low-crime area [29–31]. Victims can also engage in actions designed to reduce harm-doers' *motivation* to harm them. Thus, victims of domestic abuse have attempted to avoid triggering subsequent attacks by pleasing the potential abuser and complying with his wishes [32]. Similarly, business owners pay "protection" money to avoid damage to their business.

These tactics share in common the belief that it is beneficial to provide potential perpetrators with benefits in order to avoid more serious harm. Finally, victims have retaliated against the perpetrator. By so doing, victims can reduce perpetrators' ability and motivation to victimize them in the future. Physical harm directed at perpetrators can serve to incapacitate them as well as cause them to think twice about repeating the attack. An example of an extreme case would be a battered woman who kills her abuser when he is asleep (*see* **Battered Spouse Syndrome**). In recent years, some courts have accepted a broader definition of imminent danger and therefore have shown more sympathy toward victims who kill their abusers [33].

Option 2: Cognitively Reevaluating the Situation. Victims have also been shown to attempt to reduce their distress by reassessing their earlier decision that a crime had been committed and/or by reevaluating the seriousness of the harm. Research has documented numerous such mechanisms employed by crime victims. Two of the most commonly documented mechanisms are self-blame and making downward comparisons. With regard to the first, victims commonly ask themselves "Why me?"

One consequence of such questioning is that victims blame themselves in part for the crime [34]. Such efforts at fault finding can reduce their sense of being wronged. And, if they can blame their behavior for contributing to the victimization, they can reassure themselves that they will not repeat the mistake in the future, which can reduce their vulnerability and fear [34]. Victims have been found to reduce the seriousness of their victimization by minimizing the magnitude of the harm. This can be achieved by making "downward comparisons", such as thinking "it could have been worse", or by comparing their

outcomes to those of others who have suffered worse fates [35].

Option 3: Notifying the Police. To the extent that the above options fail to reduce victims' stress, they will have more incentive to report the crime to the police. This point is supported by the finding that reporting is significantly correlated with victims' levels of anger and fear [16, 36]. It is not only the magnitude of victims' distress that motivates reporting, but in addition, the belief in the efficacy of the police (although some studies do not support this conclusion) [37]. However, results from the NCVS tend to support the importance of perceptions of police efficacy. Some of the most frequently cited reasons for not calling the police include "Nothing could be done" and "The police would not want to be bothered" [38].

What benefits do victims hope to obtain by calling the police? One motive for notifying the police is to reduce the sense of being wronged. In the case of property crime, if the police can apprehend the criminal, the stolen property might be returned or the criminal might be ordered to make restitution to the victim. Even if the police fail to apprehend the perpetrator, reporting the crime is necessary for receiving compensation from a third party, such one's insurance company. Victims' sense of injustice could also be reduced if the perpetrator is convicted and punished. Reporting to the police can reduce victims' sense of vulnerability, particularly when the offender is convicted and sent to prison. Victims might believe that the punishment will deter future crimes by the offender, as well as deterring other would-be offenders. Even when victims see little chance of the perpetrator being apprehended, they may feel less fearful because of the belief that reporting the crime will lead to more intensive police surveillance in their neighborhood.

Given these possible benefits of calling the police, why do victims report less than half of the crimes committed? The answer to this question lies in part in the anticipated costs of reporting. Studies document numerous costs for getting involved with the criminal justice system [16]. They include costs of transportation, parking, lost time from work or school, and child care. Questioning by the police and defense attorneys, particularly in cases of sexual assault, represent additional costs. Fear of retaliation represents yet another potential cost of reporting. When victims weigh the balance of benefits and costs, they may conclude that the costs outweigh the anticipated benefits of notifying the police. This weighing of benefits is reflected in the findings from the NCVS showing that one of the most frequently offered reasons for not reporting the crime is that "The crime was not important enough" [22].

Option 4: Doing Nothing about the Crime. This is the default option. Some victims may reason that any attempt to right the wrong or reduce their vulnerability will meet with failure. As Greenberg, Ruback, and Westcott [17] have stated, "Perceiving no visible solution to their distress, these victims have no recourse but to live with the injustice and wait in fear for the inevitable occurrence of the next victimization" (p. 98).

References

[1] Bureau of Justice Statistics (1985). *Reporting Crimes to The Police*, U.S. Department of Justice, Washington, D.C.

[2] Hindelang, M.J. & Gottfredson, M. (1976). The victim's decision to invoke the criminal justice process, in *Criminal Justice and The Victim*, W.F. Mc.Donald, ed, Sage, Beverly Hills pp. 57–78.

[3] Bureau of Justice Statistics (2006). *Criminal Victimization, 2005*, U.S. Department of Justice, Washington, D.C.

[4] Bureau of Justice Statistics (2003). *Reporting Crimes to The Police, 1992–2000*, Department of Justice, Washington, D.C.

[5] Gottfredson, M.R. & Gottfredson, D.M. (1980). *Decision Making in Criminal Justice*, Ballinger, Cambridge.

[6] Skogan, W.G. (1984). Reporting crimes to the police: the status of world research, *Journal of Research in Crime and Delinquency* **21**, 113–137.

[7] Forgas, J.P. (2001). *Handbook of Affect and Social Cognition*, Erlbaum, Mahwah.

[8] Greenberg, M.S. & Beach, S.R. (2004). Property crime victims' decision to notify the police: social, cognitive, and affective determinants, *Law and Human Behavior* **28**, 177–186.

[9] Frieze, I.H., Hymer, S. & Greenberg, M.S. (1987). Describing the crime victim: psychological reactions to victimization, *Professional Psychology* **18**, 299–315.

[10] Denkers, A.J. & Winkel, F.W. (1998). Crime victims' well being and fear in a prospective and longitudinal study, *International Review of Victimology* **5**, 141–162.

[11] Elklit, A. (2002). Acute stress disorder in victims of robbery and victims of assault, *Journal of Interpersonal Violence* **17**, 872–887.
[12] Keinan, G. (1987). Decision making under stress: scanning of alternatives under controllable and uncontrollable threats, *Journal of Personality and Social Psychology* **52**, 639–644.
[13] Niedenthal, P.M., Setterlund, M.B. & Jones, D.E. (1994). *Emotional Organization of Perceptual Memory*, in P.M. Niedenthal & S. Kitayama, eds, Academic Press, New York, pp. 87–113.
[14] Petty, R.E., Cacioppo, J.T. & Kasmer, J.A. (1988). The role of affect in the elaboration likelihood model of persuasion, in *Communication, Social Cognition, and Affect*, L. Donohew, H.E. Sypher, E.T. Higgins, eds, Erlbaum, Hillsdale, pp. 117–146.
[15] Ruback, R.B., Greenberg, M.S. & Westcott, D.R. (1984). Social influence and crime-victim decision making, *Journal of Social Issues* **40**(1), 51–76.
[16] Greenberg, M.S. & Ruback, R.B. (1992). *After The Crime: Victim Decision Making*, Plenum, New York.
[17] Greenberg, M.S., Ruback, R.B. & Westcott, D.R. (1983). Seeking help from the police: the victim's perspective, in *New Directions in Helping (Vol. 3)*, A. Nadler, J.D. Fisher & B.M. De.Paulo, eds, Academic Press, New York, pp. 71–103.
[18] Williams, L.S. (1984). The classic rape: when do women report? *Social Problems* **31**, 459–467.
[19] Lejeune, R. & Alex, N. (1973). On being mugged: the event and its aftermath, *Urban Life and Culture* **2**, 259–287.
[20] Spelman, W. & Brown, D.K. (1981). *Calling The Police: Citizen Reporting of Serious Crime*, Police Executive Research Forum, Washington, D.C.
[21] VanKirk, M. (1978). *Response Time Analysis: Executive Summary*, Law Enforcement Assistance Administration, Washington, D.C.
[22] Bureau of Justice Statistics (1990). *Criminal Victimization in the United States, 1988*, Department of Justice, Washington, D.C.
[23] Simon, H.A. (1976). *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization*, 3rd Edition, Free Press, New York.
[24] Bard, M. & Sangrey, D. (1979). *The Crime Victim's Book*, Basic Books, New York.
[25] Shotland, R.L. (1976). Spontaneous vigilantism: a bystander response to criminal behavior, in *Vigilante Politics*, H.J. Rosenbaum & P.C. Sederberg, eds, University of Pennsylvania Press, Philadelphia, pp. 30–44.
[26] Barkas, J.L. (1978). *Victims*, Scribner's, New York.
[27] Van Dijk, J.J.M. & Steinmetz, C.H.D. (1979). *The RDC Victim Surveys 1974–1979*, Research and Documentation Centre, Ministry of Justice, The Hague.
[28] Conklin, J.E. (1975). *The Impact of Crime*, Macmillan, New York.
[29] (1980). *The Figgie Report on Fear of Crime: America Afraid (Part 1)*, A-T-O, Inc., Willoughby.
[30] Cohn, E.S., Kidder, L. & Harvey, J. (1978). Crime prevention vs. victimization prevention: the psychology of two different reactions, *Victimology: An International Journal* **3**, 285–296.
[31] Skogan, W.G. & Maxfield, M.G. (1981). *Coping With Crime: Individual and Neighborhood Reactions*, Sage, Beverly Hills.
[32] Ferraro, K.J. (1983). Rationalizing violence: how battered women stay, *Victimology: An International Journal* **8**, 203–212.
[33] Walker, L.E. (1989). When the battered woman becomes the defendant, in *Crime and Its Victims: International Research and Public Policy Issues*, C.E. Viano, ed, Hemisphere, New York, pp. 57–69.
[34] Janoff-Bulman, R. (1979). Characterological versus behavioral self-blame: Inquiries into depression and rape, *Journal of Personality and Social Psychology* **37**, 1798–1809.
[35] Wills, T.A. (1981). Downward comparison principles in social psychology, *Psychological Bulletin* **90**, 245–271.
[36] Maguire, M. & Corbett, C. (1987). *The Effects of Crime and the Work of Victim Support Schemes*, Gower, Aldershot.
[37] Schneider, A.L., Burcart, J.M. & Wilson II, L.A. (1976). The role of attitudes in the decision to report crimes to the police, in W.F. Mc.Donald, ed, *Criminal Justice and The Victim*, Sage, Beverly Hills, pp. 89–113.
[38] Law Enforcement Assistance Administration (1982). *Criminal Victimization in The United States, 1980*, U.S. Government Printing Office, Washington, D.C.

MARTIN S. GREENBERG

Criminal Defense: Insanity *see* Insanity: Defense

Criminal Discovery and Disclosure Rules in the United States *see* Discovery in the United States: Criminal Cases

Criminalization of the Mentally Ill

Abramson coined the term *criminalization of the mentally ill* in 1972 to describe his observation that persons with mental illness were increasingly being arrested and prosecuted for minor offenses [1]. Since 1972, a number of researchers have addressed the criminalization hypothesis. This article reviews the definition of criminalization of persons with severe mental illness (SMI), discusses possible causes of criminalization of this population, and analyzes findings from empirical studies that examine the criminalization hypothesis.

The criminalization hypothesis has not been consistently defined in the literature [2]. The common underpinning of the various definitions used for the criminalization hypothesis is the process in which the care of a subgroup of persons with mental illness shifts from the mental health system to the criminal justice system. Most studies examine criminalization by focusing on one decision point in the criminal justice system, such as police–citizen encounters, arrest, or detention [3–23].

For this article, we define criminalization as persons with SMI being treated more punitively than persons without SMI for minor or nonviolent offenses at the point of arrest, prosecution, or detention, [2, 24]. The term *severe mental illness* refers to the following diagnostic categories: schizophrenia/psychotic disorders, major depressive disorder, or bipolar disorder [25].

Hypotheses on why Persons with SMI may be Criminalized

According to the hydraulic theory of social control, the criminal justice and mental health systems share responsibility for controlling deviant behaviors in a society [26]. As a result, changes in the capacity of one system may cause a shift in the other system [26, 27]. Three changes in the mental health system could result in the criminalization of persons with SMI: (i) deinstitutionalization, (ii) legal initiatives, and (iii) length of psychiatric hospitalization. Each has brought persons with SMI increasingly into the community.

Deinstitutionalization

State mental hospitals in the United States were systematically emptied of patients in the late 1960s and 1970s. Nearly 600 000 persons were in state mental hospitals in 1955 [28]. By 2000, approximately 55 000 persons were in state mental hospitals [29].

Deinstitutionalization and community care has benefited persons with mental illness. Mental health treatment is increasingly focused on psychosocial rehabilitation and addressing individualized needs [30]. However, deinstitutionalization has not benefited all persons with SMI. Many persons with SMI were released into communities that were inadequately prepared to provide needed services [28, 31, 32]. In addition, treating persons with SMI in community settings is itself a challenge. Many persons with SMI living in the community abuse substances [33, 34], do not comply with treatment protocols [33, 35, 36], and do not have adequate community support systems [28]. By living in the community without receiving adequate care, individuals with SMI may be more likely to come to the attention of law enforcement officials. Therefore, one unintended consequence of deinstitutionalization may be that persons with SMI have been shifted from the mental health system to the correctional system.

Legal Initiatives

Civil commitment (*see* **Civil Commitment**) laws are those which stipulate the criteria that must be met to commit persons with mental illness to mental health hospitals or treatment against their will. Increasingly strict civil commitment laws, were passed in the late 1960s and early 1970s, resulting in fewer and shorter involuntary mental health facility admissions of persons with SMI [28]. Persons with SMI who are both disruptive and refusing psychiatric treatment in the community no longer meet criteria for involuntary hospitalization [15]. In contrast, jails have a "no decline" policy [37]. Police officers may manage disruptive persons with SMI who refuse psychiatric treatment through arrest and/or detention in the absence of a better alternative in the community.

Length of Psychiatric Hospitalizations

In 1965, Medicaid and Medicare programs were introduced. These programs shifted the cost of most psychiatric care except residential psychiatric care from state governments to the federal government [38, 39]. The effect of this economic incentive is evident in the accelerated pace of deinstitutionalization after the introduction of Medicaid and Medicare. State psychiatric populations decreased 4% between 1955 and 1960 after the introduction of the antipsychotic drug chlorpromazine (also known as Largactil® and Thorazine®). In contrast, state psychiatric populations decreased 29% between 1965 and 1970 after the introduction of Medicaid and Medicare [39]. Moreover, the length of all (voluntary and involuntary) psychiatric inpatient hospital stays has decreased. In 1970, the median length of stay in psychiatric hospitals was 41 days [40]. In 1997, the median length of stay in psychiatric hospitals was less than 1 week (5.4 days) [41]. Because there are economic incentives to decrease hospital stays, persons with SMI may be released into the community prematurely [42]. Consequently, with few options for adequate treatment, housing, and social support, persons with SMI may be managed through arrest or detention by police officers in the community rather than being managed through the mental health system [36, 43].

Reviewing the Literature

Arrest of Persons with SMI and Police–Citizen Encounters

Many self-report and archival studies examine the criminalization hypothesis at the point of arrest in the criminal justice system [7–9, 11, 44–51]. Several of these studies compare arrest rates among persons in mental health treatment with general population arrest rates [9, 11, 46, 47, 50, 51]. Most studies found that arrest rates among persons in psychiatric treatment were greater than arrest rates among persons in the general population [9, 11, 46, 47, 50, 51].

Although studies on arrest provide important information on the elevated arrest rates of persons with SMI compared with the general population, using findings from these studies to evaluate the criminalization hypothesis is problematic. These studies do not measure how many persons with and without SMI come into contact with police officers, making it impossible to assess if having a SMI increases the probability of arrest.

The criminalization hypothesis is more effectively investigated by studying encounters between police and persons with SMI. Such studies can evaluate if having an SMI increases the likelihood of being arrested among all criminal suspects. Police officers use their discretion to determine the appropriate response to persons with mental illness [52]. Criminalization occurs if police officers respond differently to persons with versus without SMI, choosing to arrest those with SMI because of lack of a better alternative in the community. Criminalization may also occur because of an arresting officer's implicit beliefs that persons with SMI may be more violent or likely to persistently cause trouble.

Two common law principles guide police when responding to persons with mental illness in the community: a mandate to (i) "protect the safety and welfare of the community" and (ii) "protect disabled citizens" unable to care for themselves [52]. Within these broad guidelines, police officers respond to persons with mental illness informally or formally. Examples of informal dispositions between police officers and persons with mental illness that do not result in criminalization of persons with SMI include police helping a person with SMI find a homeless shelter and police attempting to calm a person with SMI through "counsel and release" [4]. More formal dispositions that do not result in criminalization of persons with SMI involve emergency apprehension; that is, bringing a citizen to the hospital for psychiatric evaluation. In contrast, criminalization occurs if police officers arrest persons with SMI because more suitable alternatives in the community are not available. Criminalization also occurs if police officers arrest persons with SMI who have been released prematurely from psychiatric facilities and who continue to be disruptive in the community [36, 43].

To evaluate the criminalization hypothesis via police encounters, two large-scale field studies compared arrest rates of criminal suspects with mental illness to those without mental illness [3, 53].[a] Teplin's study was designed to examine the criminalization hypothesis; Engel and Silver conducted secondary analyses of two general studies of police decision-making: (i) the Project on Policing Neighborhoods Study (data collected in 1996–1997) and

(ii) the Police Service Study (data collected in 1977). Both studies found that fewer than 6% of suspects were identified as having a mental illness, limiting ability to examine additional predictive variables.

The studies differ substantially in the measure used to assess mental disorder. Teplin used a measure of mental disorder based on the Diagnostic Statistical Manual of Mental Disorders-Version III (DSM-III); her measure was tested for reliability and validity. Trained observers used a symptom checklist to assess the presence of symptoms associated with SMI, e.g., "confusion/disorientation, withdrawal/unresponsiveness, paranoia, inappropriate or bizarre speech and/or behavior, self-destructive behavior", and a global rating scale to account for the environmental context of the behavior [53]. Significantly more persons with (46.7%) versus without (27.9%) SMI were arrested [53]. Qualitative analyses revealed that police arrested persons with SMI in three types of circumstances: (i) when hospitalization was the police officer's preference but the police officer did not think the suspect met civil commitment criteria; (ii) when persons with SMI publicly exceeded the community's tolerance of deviant behavior; and (iii) when the police "felt there was a high probability that the person would continue to 'cause a problem' (and thus result in a 'callback')" [54].

In contrast, Engel and Silver's definition of mental disorder relied on the definition developed for the original studies, the Project on Policing Neighborhoods Study and the Police Service Study. They relied solely on untrained officers' and observers' perception of mental disorder; it was not on the basis of *DSM* criteria or tested for validity and reliability. In contrast to the study by Teplin [53], Engel and Silver did not find higher arrest rates among persons with mental illness after controlling for a variety of situational, legal, and suspect characteristics [3]. However, the lack of replicable measures based on *DSM* criteria introduces a serious bias. Although police officers are adept street corner psychiatrists [54], police chronically underestimate the presence of SMI [55]. As many as one-half of persons with SMI are not detected by police [53]. Thus, the study by Engel and Silver is valuable because it shows that police do not knowingly arrest persons with mental illness. But, these findings are less relevant for the criminalization hypothesis.

In short, there is a paucity of recent methodologically sound studies of police decision-making. Future studies must investigate the extent to which prior findings pertain to today's milieu.

Court Processing of Persons with SMI

Many arrests do not result in criminal prosecution [56, 57]. Prosecutors exercise a great deal of discretion in how a criminal case is processed [2]. Persons with SMI may be criminalized by being treated more punitively at the point of criminal prosecution than defendants without SMI.

Only a few studies have examined the criminalization hypothesis at the point of court processing [2, 58, 59]. Hochstedler[58] found that during court processing, 85% of defendants identified as having a mental illness were charged with misdemeanor charges [58]. The court was as likely to sanction mental health treatment (61%) as it was to sanction punishment (59%) in these cases. Persons with mental illness were granted leniency in the criminal court system only if the offense did not reach "an unspoken threshold or seriousness" [2, 58].

Hochstedler [59] later compared the individual-level data from these defendants identified as having a mental illness to aggregate data on criminal sanctions. She concluded that defendants with mental illness may be "twice cursed". When traffic offenses were excluded, defendants with mental illness were less likely to be charged with felony crimes than individuals in the comparison group but were more likely to receive stricter criminal sanctions, with many sentenced to both mental health treatment and criminal punishment. However, because she compared individual-level data with aggregate data, the conclusions from this study are not decisive [59].

One study on court processing directly evaluated if misdemeanant defendants with mental illness were treated more punitively than misdemeanant defendants without mental illness [2]. After controlling for the seriousness of the charge, defendants with mental illness were given harsher sentences for all charges except for battery compared with those without mental illness. Misdemeanants with records of mental illness and who were recognized by the court as being mentally ill (with the exception of those found incompetent to stand trial) were treated the most punitively – more were "taken into custody, held in custody, and convicted" [2].

Findings from these studies suggest that misdemeanant defendants with mental illness are treated more punitively in the criminal court system than offenders without mental illness [2]. However, more research is needed to determine if these findings are robust.

SMI in US Jail Populations

US jail populations increased 97% per 100 000 US civilians between 1986 and 2000 [60, 61]. At the same time, 24 hour psychiatric residential treatment beds have decreased 30% per 100 000 US civilians, and the number of persons in state and county mental hospitals have decreased 57% per 100 000 US civilians [29, 61]. Consequently, the hydraulic theory of social control [27], which postulates shifting responsibility between the mental health and the criminal justice systems depending on the capacity of each, could account for some of the increase in US jail populations between 1986 and 2000.

Several empirical studies have been conducted to evaluate whether changes in mental health policy – including deinstitutionalization and changes in Medicaid and Medicare managed-care programs – have caused increases in the number of persons with mental illness processed in the correctional system. Studies have found that more persons with mental illness are processed through the correctional system after changes in mental health policy that restrict or reduce inpatient mental health care [9, 49, 62–65].

To the extent that persons with SMI are criminalized, we would expect prevalence rates of SMI to be higher in jails than in the general population. In the past 20 years, many epidemiological studies of jail populations have examined prevalence rates of SMI in US jails [15, 17–23]. However, only three studies compared the prevalence of SMI in jails with that in the general population. All three studies measured psychiatric disorder using the Diagnostic Interview Schedule (version III-R), a valid and reliable measure. Two of these studies found that in the Cook County Department of Corrections, a typical urban jail, the prevalence of SMI in male and female jail detainees was significantly higher than in the general population [18, 20]. In a rural state, Powell examined a random sample of 118 prison inmates from three small state prisons, and 95 jail inmates from three regional jails. The prevalence of SMI in the rural jails was only slightly higher than that in rural community settings [17]. These findings suggest that the process of criminalization may differ in rural and urban settings.

Conclusions

Studies have shown that persons with SMI have extensive contact with the criminal justice system at each point in the system – the police–citizen encounter, arrest, court processing, and jail detention. Studies of police decision-making underscore that many persons with SMI come into contact with police officers because they need assistance, not because they are criminal suspects [4, 66]. During court processing, 85% of defendants identified as having a mental illness were charged with misdemeanor charges, and misdemeanant defendants with mental illness are treated more punitively in the criminal court system than offenders without mental illness [2]. Several epidemiological studies found higher prevalence rates of SMI in jails than in the general population [18, 20]. Finally, findings from studies analyzing the overlap of jail and clinical populations demonstrate that persons with SMI have extensive contact with the criminal justice system. A substantial number of persons with SMI in outpatient (41%) [67] and inpatient (76%) [68] settings have been incarcerated in a jail at least once in their lifetime.

The net result of deinstitutionalization, legal initiatives making involuntary treatment more difficult, and shorter length of psychiatric hospitalizations has been an increase in the number of individuals with SMI, often untreated, who live within the community [6, 24]. Persons with SMI may become involved with the criminal justice system because they fall through the cracks of the mental health system [69]. Unfortunately, society's tolerance for these individuals and their behavior is limited, especially given the stereotype of the "dangerous" patient and the sometimes bizarre behavior that can accompany SMI [6]. As a result, many residents in the community frequently summon the police to deal with individuals with mental illness [52, 54, 55]. Furthermore, many persons with SMI also have cooccurring substance use disorders [69, 70], making placement and treatment in appropriate care facilities even more problematic. One study notes a finding from its qualitative data:

"police officers would often make the rounds of the various service agencies – from halfway house to hospital to 'detox' – before resorting to the disposition of arrest" [53]. Because of problems in securing a place in the treatment system, police may view arrest as the only alternative [54].

However, many studies of the criminalization of persons with SMI were conducted before special programs were developed to divert persons with SMI from the correctional system into the mental health system, e.g., special training programs for police, mental health crisis teams integrated into police departments, jail diversion programs, and mental health courts [52, 71–73]. Future studies should address the extent to which the recent programs designed to reduce criminalization of persons with SMI have been effective.

End Notes

[a.] One study [5] reanalyzed data from Teplin (1983). However, their analysis was flawed because they included nonsuspects who, by definition, cannot be arrested [3].

References

[1] Abramson, M.F. (1972). The criminalization of mentally disordered behavior: possible side-effect of a new mental health law, *Hospital & Community Psychiatry* **23**, 101–105.

[2] Steury, E. (1991). Criminology: specifying "criminalization" of the mentally disordered misdemeanant, *The Journal of Criminal Law & Criminology* **82**, 334–359.

[3] Engel, R.S. & Silver, E. (2001). Policing mentally disordered suspects: a reexamination of the criminalization hypothesis, *Criminology* **39**, 225–281.

[4] Green, T.M. (1997). Police as frontline mental health workers: the decision to arrest or refer to mental health agencies, *International Journal of Law and Psychiatry* **20**, 469–486.

[5] Kalinich, D.B. & Senese, J.D. (1987). Police discretion and the mentally disordered in Chicago: a reconsideration, *Police Studies* **10**, 185–191.

[6] Teplin, L.A. (1983). The criminalization of the mentally ill: speculation in search of data, *Psychological Bullettin* **94**, 54–67.

[7] Belcher, J.R. (1988). Are jails replacing the mental health system for the homeless mentally ill? *Community Mental Health Journal* **24**, 185–195.

[8] Bloom, J.D., Shore, J.H. & Arvidson, B. (1981). Local variations in arrests of psychiatric patients, *The Bulletin of the American Academy of Psychiatry and the Law* **9**, 203–209.

[9] Harry, B. & Steadman, H.J. (1988). Arrest rates of patients treated at a community mental health center, *Hospital & Community Psychiatry* **39**, 862–866.

[10] Holcomb, W.R. & Ahr, P.R. (1988). Arrest rates among young adult psychiatric patients treated in inpatient and outpatient settings, *Hospital & Community Psychiatry* **39**, 52–57.

[11] McFarland, B.H., Faulkner, L.R., Bloom, J.D., Hallaux, R. & Bray, J.D. (1989). Chronic mental illness and the criminal justice system, *Hospital & Community Psychiatry* **40**, 718–723.

[12] Munetz, M.R., Grande, T.P. & Chambers, M.R. (2001). The incarceration of individuals with severe mental disorders, *Community Mental Health Journal* **37**, 361–372.

[13] Schellenberg, E.G., Wasylenki, D., Webster, C.D. & Goering, P. (1992). A review of arrests among psychiatric patients, *International Journal of Law and Psychiatry* **15**, 251–264.

[14] Lafayette, J.M., Frankle, W.G., Pollock, A., Dyer, K. & Goff, D.C. (2003). Clinical characteristics, cognitive functioning, and criminal histories of outpatients with schizophrenia, *Psychiatric Services (Washington, D.C.)* **54**, 1635–1640.

[15] Fisher, W.H., Packer, I.K., Banks, S.M., Smith, L.J. & Roy-Bujnowski, K. (2002). Self reported lifetime psychiatric hospitalization histories of jail detainees with mental disorders: comparison with a non-incarcerated national sample, *The Journal of Behavioral Health Services Research* **29**, 458–465.

[16] McNiel, D.E., Binder, R.L. & Robinson, J.C. (2005). Incarceration associated with homelessness, mental disorder, and co-occurring substance abuse, *Psychiatric Services (Washington, D.C.)* **56**, 840–846.

[17] Powell, T.A., Holt, J.C. & Fondacaro, K.M. (1997). The prevalence of mental illness among inmates in a rural state, *Law and Human Behavior* **21**, 427–438.

[18] Teplin, L.A. (1990). The prevalence of severe mental disorder among male urban jail detainees: comparison with the Epidemiologic Catchment Area Program, *American Journal of Public Health* **80**, 663–669.

[19] Teplin, L.A. (1994). Psychiatric and substance abuse disorders among male urban jail detainees, *American Journal of Public Health* **84**, 290–293.

[20] Teplin, L.A., Abram, K.M. & McClelland, G.M. (1996). Prevalence of psychiatric disorders among incarcerated women: pretrial jail detainees, *Archives of General Psychiatry* **53**, 505–512.

[21] Guy, E., Platt, J.J., Zwerling, I. & Bullock, S. (1985). Mental health status of prisoners in an urban jail, *Criminal Justice and Behavior* **12**, 29–53.

[22] Fisher, W.H., Packer, I.K., Simon, L.J. & Smith, D. (2000). Community mental health services and the prevalence of severe mental illness in local jails: are they

related, *Administration and Policy in Mental Health* **27**, 371–382.

[23] Washington, P. & Diamond, R.J. (1985). Prevalence of mental illness among women incarcerated in five California county jails, *Research in Community and Mental Health* **5**, 33–41.

[24] Lamb, H.R. & Weinberger, L.E. (1998). Persons with severe mental illness in jails and prisons: a review, *Psychiatric Services (Washington, D.C.)* **49**, 483–492.

[25] National Institute of Mental Health (2006). *The Numbers Count: Mental Disorders in America*, Available at http://www.nimh.nih.gov/publicat/numbers.cfm#Intro (accessed April 2, 2007).

[26] Fisher, W.H., Dickey, B., Normand, S.L., Packer, I.K., Grudzinskas, A.J. & Azeni, H. (2002). Use of a state inpatient forensic system under managed mental health care, *Psychiatric Services (Washington, D.C.)* **53**, 447–451.

[27] Penrose, L.S. (1939). Mental disease and crime: outline of a comparative study of European statistics, *The British Journal of Medical Psychology* **18**, 1–15.

[28] Lamb, H.R., Weinberger, L.E. & Gross, B.H. (2004). Mentally ill persons in the criminal justice system: some perspectives, *The Psychiatric Quarterly* **75**, 107–126.

[29] Manderschein, R.W., Atay, J.E., Male, A., Blacklow, B., Forest, C., Ingram, L., Maedke, J., Sussman, J. & Ndikumwami, A. (2004). Highlights of organized mental health services in 2000 and major national state trends, in *Center for Mental Health Services: Mental Health, United States, 2002*, R.W., Manderscheid & M.J., Henderson, eds, DHHS Pub No. (SMA) 3938. Substance Abuse and Mental Health Services Administration, Available at: http://www.mentalhealth.samhsa.gov/publications/allpubs/SMA04-3938/Chapter18.asp.

[30] Lamb, H.R. & Bachrach, L.L. (2001). Some perspectives on deinstitutionalization, *Psychiatric Services (Washington, D.C.)* **52**, 1039–1045.

[31] French, L. (1987). Victimization of the mentally ill: an unintended consequence of deinstitutionalization, *Social Work* **32**, 502–505.

[32] Pepper, B., Kirshner, M.C. & Ryglewicz, H. (2000). The young adult chronic patient: overview of a population, *Psychiatric Services (Washington, D.C.)* **51**, 989–995.

[33] Cournos, F. & Le Melle, S. (2000). The young adult chronic patient: a look back, *Psychiatric Services (Washington, D.C.)* **51**, 996–1000.

[34] Lamb, H.R. & Weinberger, L.E. (2001). Persons with severe mental illness in jails and prisons: a review, *New Directions for Mental Health Services* **90**, 29–49.

[35] Lamb, H.R., Weinberger, L.E. & Gross, B.H. (1999). Community treatment of severely mentally ill offenders under the jurisdiction of the criminal justice system: a review, *Psychiatric Services (Washington, D.C.)* **50**, 907–913.

[36] Laberge, D. & Morin, D. (1995). The overuse of criminal justice dispositions: failure of diversionary policies in the management of mental health problems, *International Journal of Law and Psychiatry* **18**, 389–414.

[37] Borzecki, M. & Wormith, J.S. (1985). The criminalization of psychiatrically ill people: a review with a Canadian perspective, *Psychiatric Journal of the University of Ottawa* **10**, 241–247.

[38] Pogrebin, M.R. & Poole, E.D. (1987). Deinstitutionalization and increased arrest rates among the mentally disordered, *The Journal of Psychiatry & Law* **15**, 117–127.

[39] Torrey, E.F. (1997). New initiatives in funding, in *Out of the Shadows: Confronting America's Mental Illness Crisis*, John Wiley & Sons, Inc., New York, pp. 91–140.

[40] Grobb, G.N. (2001). Mental health policy in 20th-century America, in *Center for Mental Health Services: Mental Health, United States, 2000*, R.W. Manderscheid & M.J. Henderson, eds, DHHS Pub No. (SMA) 01-3537. Substance Abuse and Mental Health Services Administration; Available at: http://www.mentalhealth.samhsa.gov/publications/allpubs/SMA01-3537/chapter2.asp (accessed April 2, 2007).

[41] Milazzo-Sayre, L.J., Henderson, M.J., Manderscheid, R.W., Blacklow, B.G., Evans, C. & Male, A.A. (2004). Selected characteristics of adults treated in specialty mental health care programs, United States, 1997, in *Center for Mental Health Services: Mental Health, United States, 2002*, R.W. Manderscheid & M.J. Henderson, eds, DHHS Pub No. (SMA) 3938, Substance Abuse and Mental Health Services Administration; Available at http://www.mentalhealth.samhsa.gov/media/KEN/pdf/SMA01-3938/MHUS02_Chapter_19.pdf (accessed April 2, 2007).

[42] Lamb, H.R. & Weinberger, L.E. (2005). The shift of psychiatric inpatient care from hospitals to jails and prisons, *The Journal of the American Academy of Psychiatry and the Law* **33**, 529–534.

[43] Rogers, A. (1990). Policing mental disorder: controversies, myths, and realities, *Social Policy and Administration* **24**, 226–236.

[44] Borum, R., Swanson, J., Swartz, M. & Hiday, V. (1997). Substance abuse, violent behavior, and police encounters among persons with severe mental disorder, *Journal of Contemporary Criminal Justice* **13**, 236–250.

[45] Hiday, V.A. (1992). Civil commitment and arrests: an investigation of the criminalization thesis, *Journal of Nervous & Mental Disease* **180**, 184–191.

[46] Brekke, J.S., Prindle, C., Bae, S.W. & Long, J.D. (2001). Risks for individuals with schizophrenia who are living in the community, *Psychiatric Services (Washington, D.C.)* **52**, 1358–1366.

[47] Schuerman, L.A. & Kobrin, S. (1984). Exposure of community mental health clients to the criminal justice system, in *Mental Health and Criminal Justice*, L.A. Teplin, ed, SAGE Publications, Inc., Beverly Hills, CA, pp. 87–118.

[48] Wolff, N., Diamond, R.J. & Helminiak, T.W. (1997). A new look at an old issue: people with mental illness and the law enforcement system, *Journal of Mental Health Administration* **24**, 152–165.

[49] Fisher, W.H., Normand, S.L., Dickey, B., Packer, I.K., Grudzinskas, A.J. & Azeni, H. (2004). Managed mental

health care's effects on arrest and forensic commitment, *International Journal of Law and Psychiatry* **27**, 65–77.
[50] Sosowsky, L. (1980). Explaining the increased arrest rate among mental patients: a cautionary note, *The American Journal of Psychiatry* **137**, 1602–1605.
[51] Hiday, V.A. (1991). Arrest and incarceration of civil commitment candidates, *Hospital & Community Psychiatry* **42**, 729–734.
[52] Lamb, H.R., Weinberger, L.E. & DeCuir Jr, W.J. (2002). The police and mental health, *Psychiatric Services (Washington, D.C.)* **53**, 1266–1271.
[53] Teplin, L.A. (1984). Criminalizing mental disorder: the comparative arrest rate of the mentally ill, *The American Psychologist* **39**, 794–803.
[54] Teplin, L.A. & Pruett, N.S. (1992). Police as streetcorner psychiatrist: managing the mentally ill, *International Journal of Law and Psychiatry* **15**, 139–156.
[55] Husted, J.R., Charter, R.A. & Perrou, B. (1995). California law enforcement agencies and the mentally ill offender, *The Bulletin of the American Academy of Psychiatry and the Law* **23**, 315–329.
[56] Federal Bureau of Investigations (2005). *Crime in the United States, 2004: Uniform Crime Reports*. Table 29: Estimated number of arrests, Available at http://www.fbi.gov/ucr/cius_04/documents/CIUS2004.pdf (accessed April 2, 2007).
[57] Bureau of Justice Statistics (2005). *Criminal Case Processing Statistics*, Available at http://www.ojp.usdoj.gov/bjs/cases.htm (accessed May 9, 2006).
[58] Hochstedler, E. (1986). Criminal prosecution of the mentally disordered, *Law & Society Review* **20**, 279–292.
[59] Hochstedler, E. (1987). Twice-cursed? The mentally disordered criminal defendant, *Criminal Justice and Behavior* **14**, 251–267.
[60] Glaze, L. & Palla, S. (2005). *Bureau of Justice Statistics report correctional surveys. Key Facts at a Glance: Correctional Populations*, Available at http://www.ojp.usdoj.gov/bjs/glance/tables/corr2tab.htm (accessed April 2, 2007).
[61] US Census Bureau (2001). *Monthly Estimates of the United States Population: April 1, 1980 to July 1, 1999, with Short-term Projections to November 2000*, Available at http://www.census.gov/population/estimates/nation/intfile1-1.txt (accessed April 2, 2007).
[62] Bonovitz, J.C. & Bonovitz, J.S. (1981). Diversion of the mentally ill into the criminal justice system: the police intervention perspective, *The American Journal of Psychiatry* **138**, 973–976.
[63] Arvanites, T.M. (1990). A comparison of civil patients and incompetent defendants: pre and post deinstitutionalization, *The Bulletin of the American Academy of Psychiatry and the Law* **18**, 393–403.
[64] Dickey, W. (1980). Incompetency and the nondangerous mentally ill client, *Criminal Law Bulletin* **16**, 22–40.
[65] Steadman, H.J. & Ribner, S. (1980). Changing perceptions of the mental health needs of inmates in local jails, *The American Journal of Psychiatry* **137**, 1115–1116.
[66] Teplin, L.A. (1985). The criminality of the mentally ill: a dangerous misconception, *The American Journal of Psychiatry* **142**, 593–599.
[67] Wright, R.G., Heiman, J.R., Shupe, J. & Olvera, G. (1989). Defining and measuring stabilization of patients during 4 years of intensive community support, *The American Journal of Psychiatry* **146**, 1293–1298.
[68] Sullivan, G. & Spritzer, K. (1997). The criminalization of persons with serious mental illness living in rural areas, *The Journal of Rural Health* **13**, 6–13.
[69] Abram, K.M. & Teplin, L.A. (1991). Co-occurring disorders among mentally ill jail detainees: implications for public policy, *The American Psychologist* **46**, 1036–1045.
[70] Abram, K.M., Teplin, L.A. & McClelland, G.M. (2003). Comorbidity of severe psychiatric disorders and substance use disorders among women in jail, *The American Journal of Psychiatry* **160**, 1007–1010.
[71] Meyers, L. (2006). $5 million will fund mentally ill offender treatment, *Monitor on Psychology* **37**, 14.
[72] Steadman, H.J., Davidson, S. & Brown, C. (2001). Law & psychiatry: mental health courts: their promise and unanswered questions, *Psychiatric Services (Washington, D.C.)* **52**, 457–458.
[73] Reuland, M. (2004). *A Guide to Implementing Police-Based Diversion Programs for Persons with Mental Illness*. TAPA Center for Jail Diversion, Substance Abuse and Mental Health Services Administration, Available at http://gainscenter.samhsa.gov/pdfs/jail_diversion/PERF.pdf (accessed April 2, 2007).

Related Articles

Crime Victims' Decision to Report Crime
Mental Health Courts
Sentencing: Demographic Factors in
Treatment, Mandated: Mental Health
Treatment, Right to: Mental Health
Treatment, Right to Refuse: Mental Health

ERIN G. ROMERO, KATE S. ELKINGTON AND LINDA A. TEPLIN

Critical Incident Teams and Policing *see* Policing and Critical Incident Teams

Cross-Examination: Impact on Testimony

Introduction

Cross-examination is the process by which evidence presented in court is scrutinized by the opposing counsel. It is one of the defining characteristics of the adversarial legal system. In theory, the primary purpose of cross-examination is to uncover errors and inconsistencies in a witness's testimony, thereby increasing the accuracy of the evidence that is presented in court. In practice, however, cross-examination is commonly used to discredit a witness's testimony, *regardless* of whether that testimony is correct or not [1].

Most legal textbooks describe a number of techniques that can be used to discredit a witness's testimony [2, 3]. For example, lawyers are taught how to lead witnesses to contradict themselves or to enlarge a story until it is improbable or unbelievable. Lawyers are also instructed to fire damaging facts at the witness during cross-examination, to attack the witness's credibility or credentials, and to ask questions in an illogical sequence in order to prevent the witness from becoming aware of the purpose of questioning [2]. In discrediting a witness's testimony, cross-examining lawyers frequently put forward an alternative version of events, the plausibility of which the witness is forced to acknowledge [4].

In light of these techniques, it is not difficult to see why cross-examination is traditionally viewed as a negative and aggressive procedure relative to the other components of the evidential process. Most witnesses, including police, find being cross-examined both intimidating and confusing. In fact, even experienced expert witnesses are typically given training designed to help them respond to cross-examination questions coherently and without undue anxiety [5, 6].

Cross-Examination of Children

Over the past three decades, researchers have firmly established the conditions under which forensic interviewers can obtain the most complete and accurate accounts from child witnesses and, conversely, the conditions that typically lead to errors (for reviews, see [7, 8]). In line with these findings, many countries have made changes to the way in which children's testimony is elicited [9] and presented in court (e.g., prerecorded videotape or closed-circuit television; [10, 11]).

Despite these advances, almost all of the recent recommendations and reforms have focused on the interviewers who solicit children's primary evidence. The effect of cross-examination on the accuracy of children's testimony has received very little attention.

In fact, the typical cross-examination of a child witness has been described as "a virtual *how not to* guide to investigative interviewing" [1, p. 279]. That is, the questioning style used during cross-examination directly contravenes almost every principle established for obtaining reliable and accurate reports from children. On the basis of prior empirical research, at least three aspects of cross-examination are likely to cause problems for children.

Leading and Closed Questioning. Although witnesses of all ages sometimes succumb to the effects of suggestive or leading questions, children are particularly susceptible (for reviews, see [7, 12]). For this reason, the use of suggestive or leading questions is strongly discouraged during direct examination whether it takes place in court or via prerecorded videotape. During cross-examination, however, lawyers can ask questions in which the desired answer is suggested or disputed facts are assumed. In fact, the right to use leading questions has been described as "one of the great advantages of cross-examination" [2, p. 105], and legal textbooks often encourage lawyers to suggest the desired answer wherever possible during cross-examination [2, 13]. Inspection of court transcripts of trials involving child witnesses has shown that leading and suggestive questions make up the bulk of the questions that are asked during cross-examination [14, 15].

Linguistic Complexity. Research with children has repeatedly shown that they will attempt to provide an answer, even if they do not understand the question [16, 17], or if the question does not make any sense (e.g., "Is milk bigger than water?") [18, 19].

Given this, most current standards of best practice for interviewing children recommend that the questions that are posed to children are developmentally appropriate for the child's cognitive skill and linguistic competence.

Despite these recommendations regarding direct examination, lawyers typically use a large number of linguistically complex questions during cross-examination [14, 15, 20–23]. In practice, complex questions are much more common during cross-examination than during other aspects of evidential questioning [14, 15, 23].

There are a number of linguistically complex features inherent in cross-examination. First, cross-examination questions often include a vocabulary of words that children do not understand, including a large number of legal terms, and children often overestimate their understanding of these words. For example, both Flin *et al.* [24] and Saywitz *et al.* [25] have shown that although children may claim to know the meaning of a legal term, when they are required to give a definition, they often provide an erroneous answer (e.g., confusing "jury" with "jewelry").

Second, cross-examination questions often require children to use complex concepts such as height, weight, age, time, and distance. Again, research has shown that children's understanding of these concepts has a very long developmental trajectory [26].

Finally, cross-examination questions often include complex syntactic features. For example, the questions posed to children during cross-examination are frequently multifaceted (e.g., "So he picked you up and then the two of you went to the movies and then he dropped you at the bus stop – is that correct?"), ambiguous (e.g., "When you and your father picked up your brother from school, did he seem happy?"), or may contain embedded clauses (e.g., "Did you talk to the lady that your father told you about?"), or inappropriate negatives (e.g., "Did you not leave the house at three o'clock?"). During cross-examination, lawyers also change topics abruptly, a tactic that they are advised to avoid when questioning their own witnesses because it is extremely confusing.

Credibility Challenges. By definition, cross-examination is highly confrontational because it involves the lawyer openly disagreeing with the witness's testimony. While lawyers sometimes allege that children are mistaken or confused about the event they are recalling, most cross-examinations of sexual abuse complainants involve the child being directly or indirectly accused of fabricating the allegation [4, 27]. For example, during cross-examination, children may be accused of lying because their disclosure was delayed, because the seriousness of their allegations have increased over time, or because, from the outside, it appears as if the complainant and the alleged perpetrator have a harmonious relationship [4].

Children's vulnerability to leading questions is higher under intimidating conditions [28] and when the child believes the interviewer to be authoritative [29], but it is important to note that challenges to a child's credibility during cross-examination are not always outwardly aggressive. Over 70% of defense lawyers surveyed by Davies *et al.* [4] indicated that the best manner in which to cross-examine a child was to be nonthreatening and gentle, at least initially. In fact, some experts have argued that children's evidence can be discredited even *more* readily by interviewing them in a friendly, charming manner and engaging trust before confronting the child [5].

Delay. Many countries have acknowledged the potential negative effects of delay on children's testimony by allowing child witnesses to present their direct evidence via prerecorded videotape [11]. Cross-examination, however, does not take place until the time of the trial. Although trial delays vary widely across countries, jurisdictions, and individual cases, children often wait several months between making an allegation and being cross-examined in court [23, 30–32], and delays of several years are not uncommon [30]. Given that the accuracy of children's recall declines over time [33], and declines more quickly than that of adults [34], delays leading up to cross-examination may be especially problematic for children.

Cross-Examination and Accuracy. Although the rules for direct and cross-examination differ markedly, and many of the questions that are asked of children during cross-examination appear to be developmentally inappropriate, the real question is whether there is any evidence that cross-examination has a negative impact on the accuracy of children's testimony. In one of the first studies designed to address this question, Brennan and Brennan [20] tested 6- to 14-year olds' understanding of cross-examination-style questions obtained from court

transcripts. They reviewed court transcripts for commonly asked questions and then asked same-aged children in the laboratory to repeat these questions back to the experimenters. Brennan and Brennan then measured the extent to which the child's response accurately captured the question's meaning. They found that many common courtroom questions were misunderstood by children.

In another study, Zajac *et al.* [15] compared child witnesses' responses to lawyers' questions during direct and cross-examination. Despite the fact that cross-examining lawyers asked a high proportion of complex and grammatically unsound questions, children rarely requested clarification, and the nature of their answers indicated that they often misunderstood lawyers' questions. Children also exhibited a high degree of compliance with the leading questions that they were asked during cross-examination. Most notably, Zajac *et al.* found that 75% of the children made at least one change to their earlier testimony under cross-examination. Almost all of these changes (95%) were preceded by a leading question that challenged the child's credibility (e.g., "But that's not really what you told her, is it?"). The changes that children made during cross-examination were not restricted to peripheral details of the alleged events. In fact, some children retracted their abuse allegations altogether.

Are the changes that children make to their testimony during cross-examination directed toward or away from the truth? For obvious reasons, the use of court transcript data does not allow researchers to evaluate the effect of cross-examination on children's accuracy. In principle, the changes that children made during cross-examination in Zajac *et al.*'s study could have increased their overall accuracy. Given the complex and challenging nature of the cross-examination process, however, it is equally likely that these changes may have made children *less* accurate.

In order to explore and empirically evaluate the effect of cross-examination on the accuracy of children's testimony, Zajac and Hayne [35] developed a laboratory analog of cross-examination. Zajac and Hayne's methodology was unique in that they examined the effect of an ecologically valid, yet standardized, cross-examination interview on children's reports of a personally experienced event. Because the researchers were knowledgeable about the target event, they could evaluate the accuracy of the changes children made to their reports during cross-examination.

In Zajac and Hayne's [35] study, 5- and 6-year-old children visited the local police station, where they participated in four unique activities (e.g., getting their "mugshots" taken). Six weeks after the visit, children participated in a simulated direct-examination interview. During this interview, they were asked general (e.g., "Tell me everything that you can remember") and yes/no (e.g., "Did you see the police car?") questions about the visit to the police station. Eight months later, children were shown their direct-examination interview on videotape and were then interviewed with a standardized laboratory analog of cross-examination. These procedures are analogous to those currently used in both New Zealand and the United Kingdom.

During the cross-examination phase of the Zajac and Hayne [35] study, the specific questions that children were asked were modeled after the kinds of questions that had been asked by defense lawyers in Zajac *et al.* [15] (e.g., "Maybe you *did* get to try on handcuffs, but you just can't remember. That's what *really* happened, isn't it?"). In response to cross-examination, 85% of the children changed at least one of the answers that they had provided during their direct-examination interview. In fact, one-third of the children changed *all* of their previous responses. Overall, cross-examination-style questioning significantly decreased the accuracy of children's reports to a point where following cross-examination their accuracy levels were not significantly different from chance. Similar findings have since been obtained with older children [36], and using forensically relevant events (e.g., bodily touch by a male authority figure) [37].

More recent empirical research on cross-examination in children has set out to examine some of the other factors that might impact the degree to which their accuracy is compromised.

Delay. In an attempt to model the conditions in actual forensic settings, Zajac and Hayne's [35] cross-examination interview occurred after an 8-month delay. In light of research showing that suggestibility increases with delay [38], Righarts *et al.* [37] hypothesized that children might perform better if the delay between direct examination and cross-examination was eliminated. To test this hypothesis,

Righarts *et al.* cross-examined 5- to 6-year-old children either 1–3 days or 8 months after their initial, direct-examination interview. Despite highly accurate initial reports, children's performance during cross-examination was very poor, even when they were cross-examined shortly after the target event. In fact, children's cross-examination accuracy scores did not differ as a function of delay. Thus, it appears that the cross-examination questions *per se* pose a difficulty for children, and reducing the delay between the allegation and the trial or conducting pretrial cross-examination may do little to facilitate children's performance.

Individual differences. It is widely acknowledged that the accuracy of children's eyewitness reports hinges largely on the way in which they are questioned, but even when external factors are held constant, children do not respond uniformly to forensic questioning. For example, under direct examination, individual differences in children's social, emotional, biological, and cognitive functioning can influence the content and accuracy of their reports [39]. What about the role of individual differences in children's response to cross-examination?

Preliminary research on this question has been conducted by Zajac *et al.* [40]. They examined the role of several psychosocial variables on children's performance under cross-examination questioning.

Although all children in that study made mistakes during cross-examination, children's level of self-esteem, self-confidence, and assertiveness exerted a small but significant contribution to cross-examination performance. That is, the same factors that may make children targets for abuse, or may be the consequences of abuse, could also make children particularly susceptible to the cross-examination process.

Preparation for cross-examination. Many jurisdictions have implemented formal preparation programs for children who are required to testify in court. These programs generally involve reducing the stress associated with testifying by familiarizing children with their role as a witness and with courtroom personnel and procedures. Although these programs are not directly designed to facilitate accuracy, are there ways of preparing children for cross-examination that might help them to provide accurate testimony?

A growing body of research has examined the use of preparation techniques aimed specifically at helping to reduce children's suggestibility. Many of these interventions involve simple verbal warnings, for example, telling children that the interviewer does not know what happened, that the questions might be tricky, and that it is okay to correct the interviewer. Many of these interventions have met with at least some success [41]. These types of warnings, however, appear insufficient to buffer children from the negative effects of cross-examination on accuracy [42].

Recently, Righarts and Zajac [42] developed a comprehensive method of preparing children for cross-examination questioning. As in earlier research, children participated in a unique staged event and were then interviewed with an analog of direct examination. Prior to the cross-examination interview, the researchers showed half of the children a short film about a girl who gets lost, and then gave children both practice and feedback on the types of questions that they were to be asked in the cross-examination interview. The entire preparation session took approximately 20 min, and was conducted 1–2 days before the cross-examination interview.

Relative to control children, those children who received preparation made fewer changes to their earlier responses, and changed a smaller proportion of their correct responses during the subsequent cross-examination interview. Furthermore, overall accuracy levels during the cross-examination interview were significantly higher in the intervention group than in the control group. In short, the intervention was successful.

The preparation intervention developed by Righarts and Zajac [42] has several distinct advantages. First, during the preparation session, children were asked questions that were entirely unrelated to their "testimony", making allegations of coaching less feasible. Second, the intervention increased children's accuracy even though it was delivered by an unfamiliar interviewer, which would be the case in real-life situations, where court preparation would be conducted by an independent third party. Third, the intervention did not reduce the number of prior mistakes that children *corrected* during cross-examination. Finally, for the 9- and 10-year-olds, the success of the intervention was unrelated to children's performance during the preparation session. That is, for these older children, mere

exposure to the preparation session was sufficient to increase accuracy during cross-examination.

Cross-Examination of Adults

The negative effects of cross-examination on children's accuracy are of concern, but they also raise questions about how adults respond to this particular questioning style. Although many aspects of cross-examination are more likely to affect children than adults (e.g., complex language, intimidation, and long delays), this does not mean that older witnesses will be buffered from the negative effects. Like children, adults are susceptible to leading questions and suggestion [43], linguistic complexity [44], social pressure or intimidation [45], and interviewers of high status [46]. Furthermore, there are several categories of witness who may be particularly susceptible to being discredited during cross-examination (e.g., witnesses who recount traumatic events, elderly witnesses, or witnesses with communication or learning difficulties).

The cross-examination of adult witnesses is likely to be qualitatively and quantitatively different from that of child witnesses. For example, we know that adult witnesses, on average, are cross-examined for a longer period of time than are children [47]. We also know that cross-examining lawyers ask adults a higher proportion of complex and credibility-challenging questions than they ask children [47]. Furthermore, when cross-examining adults, lawyers may use different reasons for challenging a witness's story. In cases of sexual assault, for example, adult witnesses are likely to be challenged on issues of consent, whereas this issue is not valid when cross-examining children.

Preliminary research conducted by Zajac and Cannan [47] on the cross-examination of adult sexual abuse complainants has shown that adults are not immune to the negative effects of cross-examination on their testimony. While the adult complainants in the Zajac and Cannan study were less likely than child complainants to comply with leading questions and more likely to give clarification in response to closed and leading questions, their responses did provide reason for concern. For example, like children, adults showed very low levels of clarification-seeking and expression of uncertainty, even when questions were ambiguous, complex, or nonsensical. Most concerning, adult complainants were just as likely as child complainants to change their earlier testimony under cross-examination. Many of these changes were made in response to credibility-challenging and leading questions. It is now imperative that laboratory research examines the effect of cross-examination-style questioning on adults' accuracy, in an attempt to determine how this process might affect witnesses' ability to provide accurate details of their past experiences.

Conclusion

Although the ultimate goal of any legal investigation should be to ascertain the truth, a number of recent studies have cast doubt on whether cross-examination is an effective means of obtaining accurate eyewitness reports. Findings from a number of studies have shown that the cross-examination process may pose particular problems for child witnesses. It is particularly concerning that the types of questions typically employed during cross-examination have been shown to exert a negative effect on the accuracy of children's reports. While recognizing that cross-examination is a mainstay of adversarial trial procedure, further research is required to identify the factors that may reduce or exacerbate the negative effects of cross-examination on the testimony provided by both children and adults.

References

[1] Henderson, E. (2002). Persuading and controlling: The theory of cross-examination in relation to children, in *Children's Testimony: A Handbook of Psychological Research and Forensic Practice*, H.L. Westcott, G.M. Davies & R. Bull, eds, John Wiley & Sons Ltd, Chichester, England, pp. 279–294.

[2] Glissan, J.L. (1991). *Cross-Examination Practice and Procedure: An Australian Perspective*, 2nd Edition, Butterworths, Sydney.

[3] Stone, M. (1988). *Cross-Examination in Criminal Trials*, Butterworths, London.

[4] Davies, E., Henderson, E. & Seymour, F.W. (1997). In the interests of justice? The cross-examination of child complainants of sexual abuse in criminal proceedings, *Psychiatry, Psychology and Law* **4**, 217–229.

[5] Flin, R. (1993). Hearing and testing children's evidence, in *Child Victims, Child Witnesses: Understanding and Improving Testimony*, G.S. Goodman & B. Bottoms, eds, Guilford Press, New York, pp. 279–299.

[6] Brodsky, S.L. (2004). *Coping with Cross-Examination and Other Pathways to Effective Testimony*, American Psychological Association, Washington, DC.
[7] Bruck, M. & Ceci, S.J. (1999). The suggestibility of children's memory, *Annual Review of Psychology* **50**, 419–439.
[8] Pipe, M.E., Lamb, M.E., Orbach, Y. & Esplin, P.W. (2004). Recent research on children's testimony about experienced and witnessed events, *Developmental Review* **24**, 440–468.
[9] Home Office (2002). *Achieving Best Evidence in Criminal Proceedings: Guidance for Vulnerable or Intimidated Witnesses, Including Children*, Home Office Communication Directorate, London.
[10] Myers, J.E.B. (1996). A decade of international reform to accommodate child witnesses: steps toward a child witness code, in *International Perspectives on Child Abuse and Children's Testimony: Psychological Research and Law*, B.L. Bottoms & G.S. Goodman, eds, Sage Publications, Thousand Oaks, pp. 221–244.
[11] Pipe, M.E. & Henaghan, M. (1996). Accommodating children's testimony: legal reforms in New Zealand, in *International Perspectives on Child Abuse and Children's Testimony: Psychological Research and Law*, B.L. Bottoms & G.S. Goodman, eds, Sage Publications, Thousand Oaks, pp. 145–167.
[12] Ceci, S.J. & Bruck, M. (1993). Suggestibility of the child witness: a historical review and synthesis, *Psychological Bulletin* **113**, 403–439.
[13] Eichelbaum, R (ed) (1993). *Mauet's Fundamentals of Trial Techniques: New Zealand Edition*, Oxford University Press, Oxford.
[14] Davies, E. & Seymour, F.W. (1998). Questioning child complainants of sexual abuse: analysis of criminal court transcripts in New Zealand, *Psychiatry, Psychology and Law* **5**, 47–61.
[15] Zajac, R. & Hayne, H. (2003). I don't think that's what *really* happened: the effect of cross-examination on the accuracy of children's reports, *Journal of Experimental Psychology: Applied* **9**, 187–195.
[16] Flavell, J.H., Speer, J.R., Green, F.L. & August, D.L. (1981). The development of comprehension monitoring and knowledge about communication, *Monographs of the Society for Research in Child Development* **46**(5), Serial No. 192.
[17] Markman, E.M. (1979). Realizing that you don't understand: Elementary school children's awareness of inconsistencies, *Child Development* **50**, 643–655.
[18] Hughes, M. & Grieve, R. (1980). On asking children bizarre questions, *First Language* **1**, 149–160.
[19] Waterman, A.H., Blades, M. & Spencer, C. (2001). Interviewing children and adults: the effect of question format on the tendency to speculate, *Applied Cognitive Psychology* **15**, 521–531.
[20] Brennan, M. & Brennan, R.E. (1988). *Strange Language: Child Victims Under Cross-Examination*, 3rd Edition, CSU Literary Studies Network, Wagga Wagga, NSW.
[21] Flin, R., Bull, R., Boon, J. & Knox, A. (1992). Children in the witness-box, in *Children as Witnesses*, H. Dent & R. Flin, eds, John Wiley & Sons, Chichester, pp. 167–181.
[22] Walker, A.G. (1993). Questioning young children in court: a linguistic case study, *Law and Human Behavior* **17**, 59–81.
[23] Goodman, G.S., Taub, E.P., Jones, D.H.P., England, P., Port, L.K., Rudy, L. & Prado, L. (1992). Testifying in criminal court: emotional effects on child sexual assault victims, *Monographs of the Society for Research in Child Development*, **57**(5), Serial No. 229).
[24] Flin, R.H., Stevenson, Y., & Davies, G.M. (1989). Children's knowledge of court proceedings, *British Journal of Psychology* **80**, 285–297.
[25] Saywitz, K., Jaenicke, C. & Camparo, L. (1990). Children's knowledge of legal terminology, *Law and Human Behavior* **14**, 523–535.
[26] Friedman, W.J. & Lyon, T.D. (2005). Development of temporal-reconstructive abilities, *Child Development* **76**, 1202–1216.
[27] Westcott, H.L. & Page, M. (2002). Cross-examination, sexual abuse and child witness identity, *Child Abuse Review* **11**, 137–152.
[28] Goodman, G.S., Bottoms, B.L., Schwartz-Kenney, B.M. & Rudy, L. (1991). Children's testimony about a stressful event: improving children's reports, *Journal of Narrative and Life History* **1**, 69–99.
[29] Ceci, S.J., Ross, D. & Toglia, M. (1987). Suggestibility of children's memory: psycholegal implications, *Journal of Experimental Psychology: General* **116**, 38–49.
[30] Eastwood, C. & Patton, W. (2002). *The Experiences of Child Complainants of Sexual Abuse in the Criminal Justice System*. Report to the Criminology Research Council, Australia.
[31] Lash, B. (1995). *Time Taken to Process Sexual Offence Cases Through the Courts*, Department of Justice, Wellington, Unpublished Manuscript.
[32] Plotnikoff, J. & Woolfson, R. (1995). *Prosecuting Child Abuse: An Evaluation of the Government's Speedy Progress Policy*, Blackstone Press, London.
[33] Brainerd, C.J., Reyna, V.F., Howe, M.L. & Kingma, J. (1990). The development of forgetting and reminiscence, *Monographs of the Society for Research in Child Development* **55**(3–4), Serial No. 222.
[34] Lee, K. & Bussey, K. (2001). Children's susceptibility to retroactive interference: the effects of age and degree of learning, *Journal of Experimental Child Psychology* **80**, 372–391.
[35] Zajac, R., Gross, J. & Hayne, H. (2003). Asked and answered: questioning children in the courtroom, *Psychiatry, Psychology and Law* **10**, 199–209.
[36] Zajac, R. & Hayne, H. (2006). The negative effect of cross-examination style questioning on children's accuracy: older children are not immune, *Applied Cognitive Psychology* **20**, 3–16.

[37] Righarts, S., Zajac, R. & Hayne, H. (2008). *Children's Responses to Cross-Examination Style Questioning: Suggestibility or Compliance?* Manuscript submitted for publication.
[38] Zaragoza, M.S. & Lane, S. (1994). Source misattributions and the suggestibility of eyewitness memory, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **20**, 934–945.
[39] Bruck, M., Ceci, S.J. & Melnyk, L. (1997). External and internal variation in the creation of false reports in children, *Learning and Individual Differences* **9**, 289–316.
[40] Zajac, R., Jury, E. & O'Neill, S. The role of psychosocial factors in young children's responses to cross-examination style questioning, *Applied Cognitive Psychology* (in press) .
[41] Warren, A., Hulse-Trotter, K. & Tubbs, E.C. (1991). Inducing resistance to suggestibility in children, *Law and Human Behavior* **15**, 273–285.
[42] Righarts, S. & Zajac, R. (2008). *The Negative Effect of Cross-Examination Questioning on the Accuracy of Children's Reports: Can we Intervene?* Manuscript in preparation.
[43] Kebbell, M.R., Hatton, C., Johnson, S.D. & O'Kelly, C.M.E. (2001). People with learning disabilities as witnesses in court: what questions should lawyers ask?, *British Journal of Learning Disabilities* **29**, 98–102.
[44] Perry, N.W., McAuliff, B.D., Tam, P., Claycomb, L., Dostal, C. & Flanagan, C. (1995). When lawyers question children: is justice served? *Law and Human Behavior* **19**, 609–629.
[45] Kassin, S.M. & Kiechel, K.L. (1996). The social psychology of false confessions: compliance, internalization, and confabulation, *Psychological Science* **7**, 125–128.
[46] Roper, R. & Shewan, D. (2002). Compliance and eyewitness testimony: do eyewitnesses comply with misleading 'expert pressure' during investigative interviewing? *Legal and Criminological Psychology* **7**, 155–163.
[47] Zajac, R. & Cannan, P.N. (in press). *Cross-Examination of Sexual Assault Complainants: A Developmental Comparison*, *Psychiatry*, *Psychology and Law*.

Related Articles

Child Sexual Abuse

Children: Suggestibility of

Children: as Witnesses

Eyewitness: Suggestibility of

Eyewitness Testimony

RACHEL ZAJAC AND HARLENE HAYNE

Cross-Examination of Experts

Nature, Purpose, and Scope

When a witness, whether lay or expert, has been called by a party to a law suit to give sworn testimony in a court or court-related procedure, and that witness has been subjected to questions by the party who called him on direct examination (*see* **Direct Examination of Experts**), the opposing party has thereafter an opportunity to ask questions of that same witness. This stage of the trial process is called the *cross-examination*. During the direct examination, the cross-examiner would have listened very carefully to the answers given by the expert, who must be prepared to be challenged on all the answers that he or she had given. The properly prepared witness would also have discussed, with the direct examiner, any areas in his testimony that are open for possible challenge.

Cross-examination is permitted so that the veracity and accuracy of the witness' testimony can be questioned and challenged where appropriate. In the context of testimony by experts, the litigant may seek to explore potential weaknesses in the expert's qualifications, question the appropriateness and accuracy of the methods of examination used, explore potential insufficiency and inaccuracy of the data obtained, and throw doubt on the credibility of the conclusions drawn from the data. In this article, only issues relating to the above are discussed.

Additionally, witnesses may be cross-examined on issues unrelated to job performance but relating only to credibility or honesty of the experts. These issues are typically referred to as *matters impeaching the credibility of the expert* and are discussed separately.

Not all possible avenues are explored on cross-examination in each case. Prior to commencing the cross-examination, a litigant will typically evaluate which approach is most likely to result in information useful to him, and forgo other possible avenues of exploration. Some cross-examiners are extremely skilled and conduct searching cross-examinations that, in important cases, may take many hours, even days, to complete. Others may, for whatever reason, be more perfunctory. Effective cross-examination

skills are believed to be an art and are highly individual to each trial attorney.

The form, duration, and manner of cross-examination of experts may differ from jurisdiction to jurisdiction. Local procedure dictates, to a large extent, how cross-examination is to be conducted.

In the adversary system, the person against whom a witness has testified has an absolute right to subject the witness to cross-examination. In criminal cases, if the right of cross-examination is abridged, the direct testimony of the witness cannot stand and must be stricken from the record.

The right of cross-examination is not without limits. Local statutes, court rules, custom, or common law may dictate the extent, duration, and scope of such questioning.

It must be noted that the right of cross-examination extends to court-related procedures only. This includes not only actual trials but also such pretrial proceedings of a judicial nature as are depositions, pretrial hearings, and even some posttrial hearings.

The style of advocates conducting a cross-examination differs also according to the personal attributes of the advocates, the credentials of the witness, and even the locale where the trial occurs. In the United Kingdom, Canada, and other adversary system jurisdictions, the tone tends to be less adversarial and strident as may be seen in many American courtrooms. When called to give testimony as a court-appointed expert – a process that while widely permitted by law is invoked seldom – both sides to the dispute are permitted to cross-examine the witness.

In civil law systems, the right of cross-examination serves a more limited function. When the examining magistrate questions witnesses while compiling the file that would make up the *dossier* upon which the decision to prosecute would be made, such questioning of witnesses by the magistrate may not allow for cross-examining the witnesses by an attorney for an accused, assuming that a potential accused has already been identified. Local practices may vary in this regard.

Cross-Examination – Some Additional Concepts

Preliminarily, a litigant who has the right to cross-examine an expert witness must decide whether it is strategically wise to do so. Secondarily, recognizing that answers given by an opposing party's witness cannot always be anticipated with confidence, the structuring of an effective cross-examination requires skillful preparation on the part of the interrogator. If a cross-examination is not carefully prepared, the questioning may well have the undesired result of strengthening, rather than weakening, an opponent's case. Along with books about brilliant cross-examinations resulting in true reversals of fortune, there are also many examples to be found, where inept cross-examinations simply permitted the expert to repeat answers that were already given on direct examination.

Thus, as a matter of tactics and in recognition of the imponderables of asking questions to which the attorney may not know the answer, lawyers are ordinarily taught that if an opposing witness has not damaged the client's case on direct, little is to be gained by cross-examining the witness at all, unless an opportunity exists to elicit evidence from the witness that has not yet been brought out and may be helpful in clarifying or supporting the opponent's position.

Generally, all testimony must be relevant to the disputed issues before the court. During cross-examination, that means the questions posed are limited to matters discussed during direct examination. Most courts, however, do not impose strict limitations on the scope of cross-examination and will permit questioning on any issues that are raised, either expressly or implicitly, by the direct examiner or by the issues before the court. In addition, impeachment regarding matters that pertain to the witness' honesty and credibility are always appropriate within the limits of the law.

Cross-Examining the Qualifications of the Witness

In adversary system jurisdictions, after a witness has been found qualified to give opinion evidence as an expert following direct examination, it is still permissible to cross-examine an opponent's expert as to his qualifications, because the extent of the expert's education and experience as a witness affects the weight and credibility of his opinion.

Litigants have many opportunities to explore the extent of an expert's knowledge-base and experience. There are professional societies who maintain

the curriculum vitae of their members or have, at the least, registers that show whether a person is a member, has obtained certain certifications and distinctions, or has been subject to ethics complaints and/or sanctions. In addition, much information about the extent of an expert's knowledge is currently available on the Internet from a great variety of sources. This includes not only professional sources but also contact information about attorneys who have opposed the same expert in other cases. These lawyers may be willing to furnish transcripts of earlier testimony that appear to contradict what an expert is stating in the current case.

It becomes important for the meticulous expert to constantly review the standard texts, protocols, and tenets of his profession, so as to be properly prepared to answer challenges on every aspect of his professional experience. It is often said among forensic scientists that they can never stop learning and must not only "keep up" with the latest advances in their field but they must also constantly review what the accepted wisdom of the past was. Nothing in the expert's professional field is immune from challenge on cross-examination.

Attempts seeking to exclude an expert witness's testimony can be made prior to trial by the filing of a motion *in limine* (*see* ***In Limine*** **Motions and Hearings**). Such a motion may be based on the lack of appropriate qualifications of the proffered witness, weaknesses in the training of the witness, or other factors that may disqualify the expert from giving opinion evidence about a specific issue – such as professed unreliability of a technique espoused. Even if the pretrial motion to exclude the expert's testimony is unsuccessful and the witness is allowed to testify, the same questions posed to the expert prior to trial in support of exclusion may again be made on cross-examination during the trial. These questions remain relevant to the credibility of the expert – an issue the jury must evaluate.

Any expert witness who (i) possesses appropriate qualifications for a profession, (ii) has the needed experience, (iii) has performed a professional examination following approved protocols, (iv) has accurately obtained certain results, (v) has drawn conclusions from these observations that are readily recognized as valid in the profession, and (vi) has properly reviewed all data before taking the witness stand has little to fear from cross-examination. The important admonition to the expert is to always state the truth, not deceive expressly or by omission, and to follow standards of the profession.

Cross-Examining with Contrary Opinions Expressed in the Expert's Professional Literature

In most jurisdictions, challenging an expert's testimony on cross-examination on the basis of professional literature or textbooks in the expert's discipline is subject to evidentiary limitations. This is true particularly in adversary system jurisdictions where stringent rules against the use of hearsay evidence may remain. Views expressed by authors of published materials who are not present in court are often believed to be hearsay because they represent assertions of persons who are not subject for cross-examination.

However, when such views by acknowledged authorities in the field are brought in only for the purpose of contradicting the testifying expert's views, the questions may be appropriate because they are offered, not for the truth of the published views, but to reflect on whether the witness on the stand is believable. The cross-examiner simply attempts to show, by this method, that there are distinguished authorities in the profession who hold views differing from those of the testifying expert.

Despite such evidence not being "hearsay", the use of treatises and professional literature is subject to some limitations. They can be used to impeach an expert only if the text or treatise is first recognized as "authoritative" through the testimony of an expert. In limited situations, such treatises or publications may indeed themselves become positive proof of the truth of their content (*see* **Learned Treatises as Evidence**).

There are two different rules concerning the use of written authorities on cross-examination to impeach or discredit an expert. The first rule allows cross-examining an expert on a published statement only if the expert has relied on the written authority. A less restrictive rule also permits cross-examination on the basis of published professional literature if the expert admits that the particular publication is a recognized authority, even though the expert did not rely on it in reaching an opinion. If an expert being cross-examined refuses to recognize a

particular publication as authoritative, such fact may be established by another expert, or even by judicial notice (*see* **Judicial Notice of Scientific Principles and Facts**).

In most jurisdictions that permit the admission in evidence of a published treatise or periodical, the publication itself is not admissible as substantive evidence if its excerpts are used solely for impeachment purposes. Excerpts that call in doubt the expert's conclusions may be read to the jury, but the publication itself is not received in evidence. The rationale for that limitation on evidentiary use is that while the reading of the author's text may be justified as an exception to the rule against hearsay, it should not be permitted to assume additional dignity by being taken to the jury room during deliberations because there was no cross-examination of its author.

Redirect and Re-Cross-Examination

In most jurisdictions, after an expert has been cross-examined, the litigant who called the witness has an opportunity to conduct a redirect examination. This process is often referred to as *rehabilitating the witness*. It consists of asking explanations about answers given during cross-examination that appear to have damaged the witness's credibility. Redirect is limited only to matters raised during cross-examination.

When the process of redirect examination has been concluded, the challenging litigant may have a further opportunity, subject only to limitations imposed by court rules or the judge's discretion, to further cross-examine the expert on answers given during redirect examination. In some jurisdictions, i.e. Canada, re-cross-examination is either prohibited entirely, or subject to stringent limitations.

Related Articles

Adversary Systems of Justice
Civil Law Systems of Justice
Direct Examination of Experts
Discovery: Depositions
Hearsay Evidence
***In Limine* Motions and Hearings**
Learned Treatises as Evidence
Ultimate Issue Evidence by Experts

ANDRE MOENSSENS

CSAAS *see* Child Sexual Abuse Accommodation

Custody Disputes *see* Parental Alienation



Dangerous Patient Exception ***see*** Duty to Warn, Violence Risk Assessment for Mental Health Professionals

Dangerous Severe Personality Disorder ***see*** Dangerousness: Risk of

Dangerousness: Risk of

An Introduction to Dangerousness

The question of an individual's dangerousness is raised often within both civil and criminal law systems (e.g., parole decisions, civil commitment, child custody cases; [1, 2]). Frequently when this issue arises, mental health professionals are asked to provide an assessment of the individual's "dangerousness", a broad concept that indicates an individual's propensity to commit a dangerous act [3]. Historically, it was assumed that dangerousness resided primarily within the individual, and external factors were minimized. Psychologists and psychiatrists, whether in practice or research, therefore tended to sort people into two groups – dangerous and not dangerous. From such assessments, dichotomous clinical predictions were then tendered as to whether individuals were likely to commit dangerous acts in the future. However, research during the late 1960s through the 1980s suggested that mental health professionals were quite inaccurate prognosticators, and they faced severe criticism for participating in these assessments. In response to the strong criticism, spurred on by legal decisions within the United States, substantial research effort was focused on methods for improving these predictions.

In addition, the very task was reconceptualized throughout the 1980s to the present day. Rather than construing individuals to be "dangerous", locating the vast majority of cause within the individual, and tending to make one-time dichotomous predictions of violence, the field has moved toward the concept of "risk assessment", an ongoing process of evaluating a person's risk factors for specific behavior(s), with the acknowledgment that many risk factors may be external to the individual [4]. Despite the movement away from the term *dangerousness* by mental health professionals, this term continues to have important meaning within a legal context, and in fact it is still used within some mental health settings as well (primarily psychiatric).

Jurisdictions across the United States, Canada, and the United Kingdom include the criterion of dangerousness in various civil and criminal laws. This article provides a brief review of the concept of

dangerousness from a legal stance. In addition, the methods for assessing dangerousness and how risk assessment has evolved over the years into its current form are discussed.

Dangerousness

It is difficult to provide a single definition of dangerousness. The definition of the term varies between jurisdictions and has changed over time. For instance, in the early part of the twentieth century, property offenses were considered "harmful acts" and property offenders could be considered dangerous and incarcerated indeterminately [5]. However, over the years, the definition of dangerousness tightened considerably and by the 1960s, the term reflected a risk of harm, particularly violence, to others, typically borne of individual or constitutional factors. That is, dangerousness was assumed to reside within the individual. It was at this time that "dangerousness to others" became a widely accepted criterion for civil commitment in most jurisdictions throughout the United States [6], Canada [1], and the United Kingdom [7].

Dangerousness in Civil Law

Risk assessments of dangerousness occur in many contexts within civil law, such as child protection and custody disputes, immigration cases, and labor regulations [1]. Civil psychiatric commitment is another area where assessments of dangerousness are often relied upon. As mentioned, most jurisdictions now have "dangerousness to others", in some form or another, as a criterion for civil commitment. Up until the 1970s, the main criterion for involuntary hospitalization in the United States was the "need for treatment" [8]. During the 1970s, most states revised their civil commitment statutes to require an individual to be both mentally ill and dangerous to self or others for involuntary hospitalization to occur [9]. Although many states have since revised their commitment laws to allow the commitment of persons assessed as "gravely disabled", the criterion of dangerous to self or others remains prominent in all state civil commitment statutes [10]. A similar legal standard has been established across Canadian jurisdictions. Following the civil commitment legislation in the United States, Canada moved to include the dangerous criterion to its commitment laws. Currently, every provincial Mental Health Act includes dangerousness to self or others criteria, although the wording and the specificity of the definition of dangerousness may vary across provinces [1].

More recently and controversially, there is also specific legislation for the *civil* commitment of *criminal* offenders, particularly sex offenders, who are deemed to pose an ongoing threat to society. Historically, in the United States, the first "sexually dangerous person" law was enacted in 1937 in Michigan [8], and other states soon adopted similar provisions. At this time, the purpose of the sexually dangerous person laws was to help society deal in a nonpenal manner with sexual offenders who were too ill to deserve punishment [11]. During the 1970s, many of the statutes were repealed owing to the increasing criticism over the lack of scientific validity of the categories used by these laws (e.g., sexual psychopathy), ineffective treatment, and inaccurate prediction methods [11].

However, since 1990 a second generation of commitment laws, generally referred to as *sexually violent predator* laws (SVP laws), have been passed in 20 states as of the time of writing this article [12]. As opposed to the purpose of the first set of sexual predator laws, the new SVP laws were established as a public safety measure and allow the precautionary, indeterminate detention of sex offenders who have completed their criminal sentences [13]. "Patients" are able to be released back into the community once treatment has reduced their risk to manageable levels. In practice, however, given the general ineffectiveness of sex offender treatment, very few people have been released back into the community during the tenure of these laws [11]. This state of affairs has drawn criticism based on civil rights grounds.

The first of the new SVP legislation was established in the State of Washington [14] in response to public outcry for greater society protection following the abduction, rape, and sexual mutilation of a young boy by a sex offender who had admitted his intent to torture children upon release from prison [15]. The Washington SVP statute defines an SVP as "a person who has been convicted of or charged with a crime of sexual violence and who suffers from a mental abnormality or personality disorder which makes the person likely to engage in predatory acts of sexual violence" [13]. The Washington statute served as a model for other state SVP legislation, and although the exact wording may vary from one state to another, all legislation requires similar criteria:

1. The individual must have a history of criminal sexual behavior.
2. The individual must have a mental abnormality or personality disorder predisposing the individual to sexual violence.
3. There must be a likelihood of future sexually violent behavior.
4. There must be a connection between the mental abnormality and the potential sexual harm [13].

Although the final decision as to whether the person will be designated an SVP is left to the court, mental health professionals play an integral role in the process. In order to determine whether the individual meets the above criteria, each potential SVP must undergo an evaluation by a mental health professional. Mental health professionals provide their opinion during the trial phase of the SVP proceedings, which occurs only if there is probable cause that an individual is an SVP [16]. If the offender is determined to be an SVP, the individual is committed indefinitely, until the person is judged to no longer pose a threat to society.

The constitutionality of SVP laws has been challenged before the courts, with *Kansas v. Hendricks* [17] being one of the most well-known cases. In this case, Mr Hendricks, a diagnosed pedophile who admitted being unable to control his urge to molest children, challenged his commitment on the basis of double jeopardy, *ex post facto* law making, and substantive due process grounds [18].

In response to the arguments presented by Mr Hendricks, the court ruled that the proceedings in the Kansas statute were civil in nature, not criminal, and therefore deemed it unnecessary to address his claims of double jeopardy and *ex post facto* law making. With respect to Hendrick's third claim, the court ruled that the Act's definition of mental abnormality satisfied substantive due process requirements. As a result, the constitutionality of the act and Mr Hendrick's commitment were upheld [18]. Since this time, the constitutionality has been challenged again (e.g., *Kansas v. Crane* [19]); however, the US Supreme Court has continued to uphold the SVP laws.

The United Kingdom has civil commitment laws that are similar to those found in Canada and the United States. According to the *Mental Health Act* [20], an individual can be involuntarily committed if "it is necessary for the health or safety of the patient or for the protection of other persons" [20]. In addition, the person being committed must be treatable [21]. However, recently an amendment to the *Mental Health Act* has been made that removes the treatability criterion, at least under certain conditions, for the purpose of managing the problematic behaviors presented by people with personality disorders [22, 23]. The policy is referred to as the *dangerous severe personality disorder* (*DSPD*) program. Under the DSPD policy, there are three admission criteria:

1. more likely than not to commit an offense that might be expected to lead to serious physical or psychological harm from which the victim would find it difficult or impossible to recover;
2. the presence of a severe personality disorder; and
3. an established link between the personality disorder and the risk of offending [24].

The reform to the Mental Health Act was prompted by two cases in the 1990s in which persons with personality disorders committed horrible crimes upon their release from prison [25]. This new policy allows for the detention of people with a severe personality disorder, regardless of whether they have previously committed a crime. This latter point is incredibly contentious, and, on its face, is inconsistent with the principles of fundamental justice that persons can only have their rights abrogated (i.e., sentenced to imprisonment) if they have engaged in officially proscribed behaviors, such as criminal behavior. It indicates that the concept of "dangerousness" is still well-entrenched in UK law.

The United Kingdom has established a number of beds for patients who are admitted under this legislation, half of which are in high security prisons and half in high security hospitals. Similar to the SVP statutes in the United States, an individual who is assessed to have a dangerous and severe personality disorder is detained indefinitely, until a comprehensive risk assessment is completed, which demonstrates that risk has been reduced [23]. Additionally, patients may be subject to a period of compulsory care and treatment in the community following discharge. The community care and treatment order would include specific requirements for the patient, and specify the actions that the clinical supervisor is empowered to take if the patient fails to comply with the order, including a provision to recall the patient to hospital. Similarly, a recall to hospital may occur

where the care and treatment could no longer be given effectively or safely in the community, or where the patient is posing a risk of serious harm to others [23].

One final area in civil law in which the issue of dangerousness, and the ability to predict violence, is important is with respect to a mental health professionals' duty to protect third parties (*see* **Duty to Warn**). In both the United States and Canada (depending on jurisdiction), judicial decisions have clearly demonstrated that mental health professionals can be held liable for failing to take reasonable steps to protect third parties from the violence of their patients or clients [1]. This was most famously illustrated in the case of Tarasoff v. The Regents of the University of California [26].

In Canada, this issue was raised in the case of *Wenden v. Trikha* [26]. In this case, Mr Trikha was admitted to hospital and was under the care of a psychiatrist. Soon after his admission, Mr Trikha left the hospital without permission and was involved in a car accident in which Ms Wenden was seriously injured. In the lawsuit brought against the hospital, the court held that when a psychiatrist becomes aware that a patient presents a serious danger to the physical well-being of a third party or parties, there arises a duty to take reasonable steps to protect such a person or persons provided they can be identified. In this case, the question of identifiability of the victim was difficult; therefore, the court ruled that the psychiatrist had not acted negligently as no specific victim could be identified.

The role of mental health professionals' duty to protect third parties also arose in the Supreme Court of Canada case of *Smith v. Jones* [27]. In this case, the defense counsel hired a psychiatrist to assess the accused and told the accused that the interview was legally privileged. During the assessment interview, the accused revealed in substantial detail his plan to kidnap, rape, and kill prostitutes. The psychiatrist reported to defense counsel that the accused was dangerous and likely to commit future offenses unless he received adequate treatment. The accused later pled guilty to the charge (aggravated assault against a prostitute). The psychiatrist contacted the defense counsel to inquire about the proceedings and learned that his concerns regarding the defendant would not be presented to the court. Upon learning this, the psychiatrist commenced action to disclose the information he had in the interest of public safety. The psychiatrist filed an affidavit describing his interview with the accused and his opinion based upon the interview. The trial judge ruled that the public safety exception to the doctor–patient confidentiality and solicitor–client privilege released the psychiatrist from his duties of confidentiality and concluded that the psychiatrist was under a duty to disclose the information to the police and Crown counsel. The accused appealed the trial judge's decision; however the ruling was upheld by the Supreme Court of Canada and the psychiatrist was permitted to reveal the statements made by the accused and his opinion based upon the interview. *Smith v. Jones* specified the conditions under which legal privilege can be violated for public safety reasons. The court identified three factors that should be taken into consideration in determining whether public safety outweighs solicitor–client privilege: (i) Is there a clear risk to an identifiable person or group of persons? (ii) Is there a risk of serious bodily harm or death? and (iii) Is the danger imminent? These three factors must be defined within each separate context and different weights will be allocated to each factor, as well as the various aspects of each, in any particular case. The court also indicated that, since legal privilege is the strictest of all types of privilege or confidentiality, the exception would apply to other forms of confidentiality, such as that between mental health professional and patient or client. These cases suggest that the judiciary believes that mental health professionals have the ability to recognize the risk individuals pose and to intervene effectively [28].

Dangerousness in Criminal Law

Canada has also established legislation to incarcerate offenders indeterminately if they are deemed to pose undue risk to society. However, as opposed to the civil SVP laws found in the United States, the Canadian Dangerous Offender (DO) and Long-Term Offender (LTO) legislation falls within the criminal law system. Additionally, whereas the SVP laws in the Unites States vary from one state to another, the Canadian Criminal Code, and therefore DO/LTO legislation, applies across the country. In the 1940s, Canada paralleled the movement in the United States for handling sexual offenders and enacted the first DO provisions with the *Habitual Offender Act* (1947) and the *Criminal Sexual Psychopath Act* (1948) [29]. Both of these acts survived until 1977, when the

DO provisions were passed, which remain in place today [30].

To designate an offender as a DO, the Crown must present an application to the court after a conviction but prior to the sentencing. In order for an offender to be designated a DO, the index offense must be a serious personal offense other than homicide (punishable by imprisonment of at least 10 years) and the offender must have been assessed as a "threat to the life, safety and physical or mental well-being of other persons" [31].[a] If an offender is found to be a DO, the court imposes an indefinite sentence and the offender can only be released upon a judgment that the person no longer poses a threat to society.

In addition to the DO statute, Canada more recently enacted LTO provisions. The purpose of the LTO provision, enacted in 1997, is to ensure that sexual offenders are provided with long-term supervision once they are released into the community [30]. Similar to the DO application process, an LTO application is made by Crown Counsel following a conviction but prior to sentencing. To be designated an LTO, the offender must have committed an offense that is punishable by a sentence of two years or more, there must be a substantial risk that the offender will reoffend and there must be a reasonable possibility of eventual control of risk in the community [30]. Unlike a DO or an SVP, an LTO is not given an indefinite sentence. In fact, no additional custody time is added to an offender who is designated an LTO. Instead, the LTO designation allows for community supervision of the offender for up to 10 years after release.

The question of dangerousness can be raised in multiple other criminal law contexts. The dangerousness of an offender may be considered in decisions to release an individual on bail, transfers of juveniles to adult court, parole decisions, and institutional classification decisions, to name a few [1, 2]. In the United States, dangerousness can also be raised when deciding whether to impose the death penalty. Many states in the United States continue to use capital punishment as a sentencing option for offenders who commit serious crimes. Within this process, mental health professionals can be asked to assess whether an individual is likely to commit a dangerous act in the future. If the individual is deemed likely to pose a threat to the community in the future, which in some jurisdictions includes fellow prisoners, the death penalty could be imposed.

Similar to the SVP civil commitment cases, the constitutionality of psychiatrists providing assessments of future violent behavior was challenged in the United States death penalty case of *Barefoot v. Estelle* [32]. In this case, the offender was convicted of capital murder in Texas and the question of whether the death penalty should be imposed was deliberated at sentencing. As required by Texas statute, the question of whether there was a probability that the offender would commit further criminal acts of violence in the future and would constitute a continuing threat to society was submitted to the jury. One element of the evidence presented by the State was the testimony of two psychiatrists who testified that the defendant would probably commit future acts of violence and represent a continuing threat to society.

On the basis of the evidence presented, the jury answered affirmatively to the above questions and imposed the death penalty. The offender appealed his sentence, arguing that the psychiatrists' testimony was unconstitutional because psychiatrists are not competent to predict future dangerousness and their predictions are so likely to produce erroneous sentences that their use violated the Eighth and Fourteenth Amendments of the United States Constitution. However, the Texas Court of Criminal Appeals rejected this argument and the death sentence was upheld.

Risk Assessment

Clearly, mental health professionals are relied upon often and within many different legal contexts to assess dangerousness, or predict violence. Unfortunately, mental health professionals have not always demonstrated a clear ability to perform this task. During the 1960s, 1970s, and early 1980s, mental health professionals were seriously challenged regarding their ability to predict violence. At this time, the prevailing view was that dangerousness was a stable disposition and that only a one-time dichotomous yes/no prediction of future dangerousness was required [4]. Following important legal decisions in the United States (e.g., *Baxstrom v. Herold* [33]; *Dixon v. Attorney General of the Commonwealth of Pennsylvania* [34]), researchers were provided with a unique opportunity to assess the predictive ability of mental health professionals. As a result of these court cases, a number of presumed "dangerous" patients

were released from involuntary hospitalization under court orders and returned to the community or less secure institutions [35]. After following up with these patients, it was found that very few of the patients were involved in a violent incident after release [35, 36]. In a four-year follow-up period, only 7 of 98 *Baxstrom* patients released to the community committed new violent offenses, despite the fact that all had been predicted to be violent [35]. The incidence of violence was slightly higher, though still low given that all patients had been predicted to be violent, for the *Dixon* patients, with 14.5% of patients released involved in a violent incident in a three-year follow-up [36]. The results from the *Baxstrom* and *Dixon* studies illustrated two important points: (i) violence was overpredicted and (ii) there was a low base rate of violence during the follow-up periods [28]. Monahan [37] reviewed the accuracy of clinical prediction in these and several other early studies, which he referred to as *first-generation studies*, and concluded that

> *psychiatrists and psychologists are accurate in no more than one out of three predictions of violent behavior over a several-year period among institutionalized populations that had both committed violence in the past (and thus had a high base rate for it) and who were diagnosed as mentally ill* (p. 77, emphasis in original).

Monahan was not the only person who criticized the ability of mental health professionals. Ennis and Litwack [38] concluded from the first-generation research studies that mental health professionals had no ability to assess risk and suggested that mental health professionals should be barred from offering testimony because their judgments were not valid. Professional organizations also voiced their opinions regarding the poor ability of mental health professionals. The American Psychiatric Association [39] concluded that psychiatrists had not demonstrated any ability to predict future violence or dangerousness and that no special psychiatric expertise in the area of violence prediction had been established.

Despite the poor outlook on mental health professionals' ability to predict future behavior, the legal system continued to demand violence predictions by mental health professionals [40, 41]. Attempts to challenge the use of violence predictions as evidence in both the United States (e.g., *Barefoot v. Estelle* [32]) and Canada (*Re Moore and the Queen* [42]) were rejected by the courts [1]. It appeared that the role of mental health professionals as predictors of violent behavior within the legal context was solidified.

Considering the important role that mental health professionals were continuing to play in the legal world, it was imperative that mental health professionals and researchers worked toward improving predictions of violence. At the conclusion of his review of clinical prediction, Monahan [37] urged researchers to move toward a "second generation" of studies and provided numerous suggestions for improving research. As a result of the criticism and the suggestions provided, researchers moved into the second generation of research studies and began to use more sophisticated methodology. Following the suggestions provided by Monahan [37] and others [3], researchers began to conduct studies that used shorter follow-up times [43–45], focused on risk factors in greater detail [44, 45], measured violence from multiple sources [43, 46] and began to examine the role that situational/environmental factors might play [46].

Monahan [37] also argued for a more actuarial, or statistical, approach to violence prediction. In line with this suggestion, numerous actuarial measures, such as the Violence Risk Appraisal Guide (VRAG) [47] and the Static-99 [48], were developed. The actuarial method is based on empirically established correlations between a set of risk factors, on the one hand, and violent outcome on the other [49]. On the basis of foundational work of Paul Meehl [50], the actuarial method uses set rules (i.e., algorithms, equations) to combine risk factors in order to derive estimates of the risk of future behavior for any given individual [51].

To develop the actuarial measures, researchers attempted to identify the characteristics that could be used to distinguish those individuals who are likely to behave violently in the future from those who were not likely to behave violently [1]. Through this research, a number of potential risk factors were explored and certain risk factors associated with violent behavior were identified. Actuarial measures focused mostly on static, or historical, risk factors and there was a growing consensus that this type of factor was the most strongly predictive of violence [28]. Static factors that demonstrated an association with violence included previous violent behavior [52, 53], psychopathy [54], a history of substance use [55], and young age [56] (for review of risk factors, see [1, 28]).

The results of the second-generation research demonstrated that mental health professionals have some ability to assess risk and make predictions [9]. Contrary to the results from the first-generation research that suggested that mental health professionals successfully predicted future violence in only one of three cases, the results from the second-generation studies suggested that the ability to predict future violence improved to one of every two cases [9]. Further, the new research often demonstrated large statistical effect sizes for the prediction of violence (for reviews, see [28] [57]).

Perhaps, one of the most important developments that occurred with the increased focus on improving dangerousness predictions was the shift in the conceptualization of the task itself [28]. The change in the conceptualization saw the terms *prediction* and *dangerousness* replaced with "risk" and "assessment". Steadman and colleagues [4] explained that the reconceptualization was a result of the emerging public health perspective that viewed violence as a threat to the health of society, rather than merely as a discrete event. The public health perspective, it was argued, would be better able to balance the rights of mentally ill persons not to be unfairly detained, while at the same time protecting the rights of society to not be victims of violence [4].

A number of conceptual differences between predicting dangerousness and risk assessment have been described. First, the term *dangerousness* itself was criticized as being vague and not able to delineate what the danger is or for whom the danger exists [58]. In contrast, the terms *risk* and *risk assessment* are viewed as terms that acknowledge the multifaceted nature of violence by giving consideration to the type of violence, the population at risk, possible contextual inhibitors or disinhibitors, and the type of harm [1]. Further, risk is not viewed as a stable disposition as dangerousness was. Instead, risk should be seen as a continuous probability statement that can fluctuate over time [4]. It is a state of potential, and an assessment of risk is an estimate of how likely that state of potential is to materialize.

A prediction of dangerousness or violence, on the other hand, was a definitive statement about the future. As such, estimates of risk should take the form of ongoing assessments of a potentially changing state of potential, as opposed to the one-time predictions of dangerousness of the past. Finally, the purpose of a risk assessment is not violence *prediction*. Mental health professionals are asked to conduct risk assessments for the purpose of managing risk and violence *prevention* [58], even if it means that prevention takes the form of incarceration as opposed to treatment. With this goal in mind, risk assessment research has now moved into the third generation.

Although there is general agreement that an actuarial approach improves consistency and validity of group-based predictions beyond that of unstructured clinical prediction, commentators have noted several shortcomings of a purely actuarial approach [58–60]. First, actuarial prediction approaches tend to overoptimize predictive accuracy. Many, though not all, actuarial devices were developed in single samples, and not subjected to cross-validation. Upon cross-validation, there is often attenuation in observed accuracy. Second, and relatedly, actuarial devices that were developed within specific settings cannot be assumed to generalize to novel settings, in that their predictive properties depend to some extent on the nature of the development sample. Third, many actuarial instruments focus upon static risk factors that are less sensitive to change than dynamic risk factors [61]. As such, they have limited utility to the fundamental task of risk reduction and management.

Finally, it is not clear that an actuarial prediction, which is derived from the behavior of groups of individuals, necessarily applies at the individual level. That is, if 60% of a group of offenders in Category *X* of Risk Measure *Y* are violent, this does not mean that any subsequent individual who scores in Category *X* has a 60% chance of future violence. As Hart *et al.* [60] have argued and illustrated, such attempts at precise numerical prediction at the individual, as opposed to group, level are highly prone to error.

In response to these problems, another structured risk assessment model was developed, termed *Structured Professional Judgment* (*SPJ*) [58, 59, 62]. Like actuarial approaches, it includes risk factors with empirical support, but SPJ instruments select risk factors based on a thorough review of the literature, rather than performance in single studies, with the goal of enhancing generalizability across samples and applications. Similarly, SPJ instruments have a set of operationally defined risk factors in order to enhance reliability. Unlike actuarial approaches, SPJ instruments strive for comprehensive domain coverage in terms of risk factors, rather than relying solely upon

the set of risk factors that emerge from a single development study.

A major difference between SPJ and actuarial approaches is that SPJ approaches do not involve strict algorithmic decision making. That is, there are no numeric cutoffs, or risk categories based on a person's score on an instrument, that determine decisions about risk. Although, as its name suggests, SPJ measures provide structure to the decision-making process, they do not adopt algorithms for several reasons: (i) as mentioned, algorithms derived in a given sample often degrade in predictive accuracy when applied to new samples; (ii) though algorithms promote consistency, they may diminish individual relevance, in that the assumption is that the algorithm applies in the same manner to all persons (i.e., assumes that all risk factors are of equal relevance for all people); (iii) algorithmic procedures presume a fixed future – they cannot take ebbs and flows in future circumstances into account; and (iv) cases in which only a few risk factors are present, but those risk factors are highly salient for a given individual, tend to result in "low risk" classifications, even if individual risk is high.

SPJ approaches rely both on nomothetic data and individual relevance. Concerning the former, they include operationally defined risk factors with support in the scientific literature. Concerning the latter, their clinical assessment protocols encourage evaluators to give due attention to how these nomothetically supported risk factors apply in the given case. That is, rather than simply noting that a person has substance abuse problems, SPJ approaches then require evaluators to consider (i) the way in which substance use has manifested for the given individual and (ii) the extent to which substance use appears critically tied to the violent behavior of an individual.

Critics of the SPJ approach [63] have argued that it lowers reliability and validity through the allowance of discretion at the variable integration phase of decision making. Though this is a controversial aspect of SPJ, research to date suggests that the reliability and predictive validity of the SPJ approach are at least comparable to that of the actuarial approach, and in some studies, exceed it [64–66]. There are two main types of validity analysis that are important within the SPJ context: (i) do the risk factors, as defined within the given SPJ instrument, relate to violence? and (ii) do the judgments of low, moderate, and high risk – based on a consideration of the presence and relevance of risk factors present in a given case, and the anticipated intensity of risk management strategies necessary to reduce risk – relate to violence?

In an illustrative example of whether SPJ-defined risk factors relate to violence, Douglas, Ogloff, Nicholls, and Grant [67] followed 193 civil psychiatric patients after release from psychiatric hospitalization. In an average two-year follow-up, risk factor scores on the SPJ instrument under investigation (the HCR-20, [62]) were strongly related to violence, with an area under the curve (AUC) of 0.80 for violence leading to criminal conviction. Similarly, those who scored above the median of the HCR-20 were 13 times more likely to record a criminal conviction for a violent offense than those who scored below the median score of the HCR-20. Concerning the issue of whether judgments of low, moderate, or high risk, predict violence, de Vogel and de Ruiter [68] reported that even with the numeric (actuarial) HCR-20 scores in a predictive model, the ratings of low, moderate, and high risk improved upon their predictive accuracy. Although there are exceptions, these studies are typical of findings on SPJ risk assessment instrument (i.e., that their risk factors predict violence, and that judgments of low, moderate, and high risk predict violence as well or better than numeric prediction (see Douglas, Guy, & Weir [69], for an annotated bibliography).

In conclusion, the concept of "dangerousness" has evolved substantially over the past 40 years, when it first occupied a central stage in forensic mental health. Credible studies no longer lump all persons with mental disorder into a single gross category. Meaningful predictions are not offered in simple "yes" or "no" terms. Contemporary research on "dangerousness" – now known as risk assessment and management – attempts to fuse the vast literature on what is known about violence risk factors with principled and structured decision making. Most recently, risk assessment approaches attempt to integrate the concept of violence risk management, with the goal of reducing violence. Future research should continue in this tradition and incorporate even more attention to the concept of risk reduction, since there is little point in predicting violence if we do not attempt to prevent it (see Hart [58]).

End Notes

[a.] It should be noted that the designation "dangerous offender" is not strictly for sexual offenders. Violent offenders can be designated as DOs as well [30]. However, most people who are designated DOs are in fact sexual offenders.

References

[1] Lyons, D.R., Hart, S.D. & Webster C.D. (2001). *Violence and Risk Assessment*, in R.A. Schuller & J.R.P. Ogloff, eds, University of Toronto Press, Toronto, pp. 314–350.

[2] Shah, S. (1978). Dangerousness: a paradigm for exploring some issues in law and psychology, *American Psychologist* **33**, 224–238.

[3] Mulvey, E.P. & Lidz, C.W. (1984). Clinical considerations in the prediction of dangerousness in mental patients, *Clinical Psychology Review* **4**, 379–401.

[4] Steadman, H.J., Mulvey, E., Monahan, J., Robbins, P.C., Appelbaum, P.S., Grisso, T., Roth, L.H. & Silver, E. (1994). Designing a new generation of risk assessment research, in *Violence and the Mental Disorder: Developments in Risk Assessment*, J. Monahan & H.J. Steadman, eds, University of Chicago Press, Chicago, pp. 297–318.

[5] Pratt, J. (1997). *Governing the Dangerousness*, Federation Press, Sydney.

[6] Monahan, J. (1988). Risk assessment of violence among the mentally disordered: generating useful knowledge, *International Journal of Law and Psychiatry* **11**, 249–257.

[7] John Howard Society of Alberta (1999). *Dangerous Offender Legislation around the World*, http://www.johnhoward.ab.ca/PUB/C20.htm.

[8] Monahan, J. & Shah, S. (1989). Dangerousness and commitment of the mentally disordered in the United States, *Schizophrenia Bulletin* **15**, 541–553.

[9] Otto, R. (1994). On the ability of mental health professionals to "Predict Dangerousness": a commentary on interpretations of the "Dangerousness" literature, *Law and Psychology Review* **18**, 43–68.

[10] Miller, R.D. (1992). Need-for-treatment criteria for involuntary civil commitment: impact in practice, *American Journal of Psychiatry* **149**, 1380–1384.

[11] Janus, E.S. (2000). Sexual predator commitment laws: lessons for law and the behavioral sciences, *Behavioral Sciences and the Law* **18**, 5–21.

[12] Hafemeister, T.L. & Woolard, A. (2007). Book review: failure to protect: America's sexual predator laws and the rise of the preventive state, *Law and Politics Book Review* **17**, 326–333.

[13] Jackson, R.L., Rogers, R. & Schuman, D.W. (2004). The adequacy and accuracy of sexually violence predator evaluations: contextualized risk assessment in clinical practice, *International Journal of Forensic Mental Health* **3**, 115–129.

[14] Community Protection Act (1990). *Chapter 71.09 Sexually Violent Predators*, http://apps.leg.wa.gov/RCW/default.aspx?cite=71.09.

[15] Mercado, C. & Ogloff, J.R.P. (2007). Risk and the preventive detention of sex offenders in Australia and the United States, *International Journal of Law and Psychiatry* **30**, 49–59.

[16] Sreenivasan, S., Weinberger, L.E. & Garrick, T. (2003). Expert testimony in sexually violent predator commitments: conceptualizing legal standards of "mental disorder" and "likely to reoffend", *Journal of the American Academy of Psychiatry and Law* **31**, 471–485.

[17] Kansas v. Hendricks, 521 U.S. 346 (1997).

[18] Miller, H.A., Amenta, A.E. & Conroy, M.A. (2005). Sexually violent predator evaluations: empirical evidence, strategies for professionals, and research directions, *Law and Human Behavior* **29**, 29–54.

[19] Kansas v. Crane, 534 U.S. 407 (2002).

[20] Department of Health and the Home Office (1983). *Mental Health Act*.

[21] Maden, A., Williams, J., Wong, S.C.P. & Leis, T.A. (2004). Treating dangerous and severe personality disorder in high security: lessons from the regional psychiatric centre, Saskatoon, Canada, *The Journal of Forensic Psychiatry and Psychology* **15**, 375–390.

[22] Appelbaum, P.S. (2005). Dangerous severe personality disorders: England's experiment in using psychiatry for public protection, *Psychiatric Services* **56**, 397–399.

[23] Department of Health and Home Office (2000). *Reforming the Mental Health Act*, Stationary Office, London.

[24] National Probation Service (2005). Dangerous and severe personality disorder (DSPD) programme, *Probation Circular* p. 40.

[25] Howard, R. (2006). What is the link between personality disorder and dangerousness? a critique of 'dangerous and severe personality disorder', *The British Journal of Forensic Practice* **8**, 19–23.

[26] Wenden v. Trikha, 116 A.R. 81 (Q.B. 1991).

[27] Smith v. Jones, 1 S.C.R. 455 (1999).

[28] Douglas, K.S. & Webster, C.D. (1999). Predicting violence in mentally disordered individuals, in *Psychology and Law: The State of the Discipline*, R. Roesch, S.D. Hart, eds, Kluwer Academic/Plenum Publishers, New York, pp. 175–239.

[29] Grant, I. (1985). Dangerous offenders, *Dalhousie Law Journal* **9**, 347–382.

[30] Heilbrun, K., Ogloff, J.R.P. & Picarello, K. (1999). Dangerous offender statutes in the United States and Canada: implications for risk assessment, *International Journal of Law and Psychiatry* **22**, 393–415.

[31] Criminal Code of Canada S.752 (1985).

[32] Barefoot v. Estelle, 463 U.S. 880 (1983).

[33] Baxstrom v. Herold, 383 U.S. 107 (1966).

[34] Dixon v. Attorney General of the Commonwealth of Pennsylvania, 325 F.Supp. 966 (1971).

[35] Steadman, H.J. & Cocozza, J. (1974). *Careers of the Criminally Insane*, Lexington Books, Lexington.

[36] Thornberry, T. & Jacoby, J. (1979). *The Criminally Insane: A Community Follow-up of Mentally Ill Offenders*, University of Chicago Press, Chicago.
[37] Monahan, J. (1981). *Predicting Violent Behavior: An Assessment of Clinical Techniques*, Sage, Beverly Hills.
[38] Ennis, B.J. & Litwack, T.R. (1974). Psychiatry and the presumption of expertise: flipping coins in the courtroom, *California Law Review* **62**, 693–752.
[39] American Psychiatric Association (1974). *Report of the American Psychiatric Association Task Force on Clinical Aspects of the Violent Individual*, American Psychiatric Association, Washington, DC.
[40] Monahan, J. (1996). Violence prediction: the last 20 and the next 20 years, *Criminal Justice and Behavior* **23**, 107–120.
[41] Mulvey, E.P. & Lidz, C.W. (1995). Conditional prediction: a model for research on dangerousness to others in a new era, *International Journal of Law and Psychiatry* **18**, 129–143.
[42] Re Moore and the Queen, 10 C.C.C. (3d) 306 (1984).
[43] Lidz, C.W., Mulvey, E.P. & Gardner, W. (1993). The accuracy of predictions of violence to others, *Journal of the American Medical Association* **269**, 1007–1111.
[44] McNiel, D.E. & Binder, R.L. (1994). The relationship between acute psychiatric symptoms, diagnosis, an short-term risk of violence, *Hospital and Community Psychiatry* **45**, 133–137.
[45] McNiel, D.E. & Binder, R.L. (1995). Correlates of accuracy in the assessment of psychiatric inpatients' risk of violence, *The American Journal of Psychiatry* **152**, 901–906.
[46] Klassen, D. & O'Connor, W.A. (1988). A prospective study of predictors of violence in adult male mental health admissions, *Law and Human Behavior* **12**, 143–158.
[47] Quinsey, V.L., Harris, G.T., Rice, M.E. & Cormier, C.A. (1998). *Violent Offenders: Appraising and Managing Risk*, American Psychological Association, Washington, DC.
[48] Hanson, R.K. & Thornton, D. (1999). *Static-99: Improving Actuarial Risk Assessment for Sex Offenders*, User report 1999-02, Department of the Solicitor General of Canada, Ottawa.
[49] Gendreau, P. & Goggin, C. (1996). Principles of effective assessment for community corrections, *Federal Probation* **60**, 64–71.
[50] Meehl, P. (1954). *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*, University of Minnesota Press, Minneapolis.
[51] Hanson, R.K. (1998). What do we know about sex offender risk assessment?, *Psychology, Public Policy, and Law* **4**, 50–72.
[52] Bonta, J., Law, M. & Hanson, R.K. (1998). The prediction of criminal and violent recidivism among mentally disordered offenders: a meta-analysis, *Psychological Bulletin* **123**, 123–142.
[53] Gardner, W., Lidz, C.W., Mulvey, E.P. & Shaw, E.C. (1996). A comparison of actuarial methods for identifying repetitively violent patients with mental illness, *Law and Human Behavior* **20**, 35–38.
[54] Harris, G.T., Rice, M.E. & Cormier, C.A. (1991). Psychopathy and violent recidivism, *Law and Human Behavior* **15**, 625–637.
[55] Swanson, J.W. (1994). Mental disorder, substance abuse, and community violence: an epidemiological approach, in *Violence and Mental Disorder: Developments in Risk Assessment*, J. Monahan & H.J. Steadman, eds, University of Chicago Press, Chicago, pp. 101–136.
[56] Harris, G.T., Rice, M.E. & Quinsey, V.L. (1993). Violent recidivism of mentally disordered offenders: the development of a statistical prediction instrument, *Criminal Justice and Behavior* **20**, 315–335.
[57] Monahan, J. & Steadman, J. (1994). *Violence and Mental Disorder: Developments in Risk Assessment*, University of Chicago Press, Chicago.
[58] Hart, S.D. (1998). The role of psychopathy in assessing risk for violence: conceptual and methodological issues, *Legal and Criminological Psychology* **3**, 121–137.
[59] Douglas, K.S. & Kropp, P.R. (2002). A prevention-based paradigm for violence risk assessment: clinical and research applications, *Criminal Justice and Behavior* **29**, 617–658.
[60] Hart, S.D., Michie, C. & Cooke, D.J. (2007). Precision of actuarial risk assessment instruments: evaluating the 'margins of error' of group v. individual predictions of violence, *The British Journal of Psychiatry* **190**, s60–s65.
[61] Douglas, K.S. & Skeem, J.L. (2005). Violence risk assessment: getting specific about being dynamic, *Psychology, Public Policy, and Law* **11**, 347–383.
[62] Webster, C.D., Douglas, K.S., Eaves, D. & Hart, S.D. (1997). *HCR-20: Assessing Risk for Violence (Version 2)*, Mental Health, Law, and Policy Institute, Burnaby, pp. 297–318.
[63] Quinsey, V.L., Harris, G.T., Rice, M.E. & Cormier, C.A. (2006). *Violent Offenders: Appraising and Managing Risk*, 2nd Edition, American Psychological Association, Washington, DC.
[64] Douglas, K.S., Ogloff, J.R.P. & Hart, S.D. (2003). Evaluation of a model of violence risk assessment among forensic psychiatric patients, *Psychiatric Services* **54**, 1372–1379.
[65] Douglas, K.S., Yeomans, M. & Boer, D.P. (2005). Comparative validity analysis of multiple measures of violence risk in a sample of criminal offenders, *Criminal Justice and Behavior* **32**, 479–510.
[66] Doyle, M. & Dolan, M. (2006). Predicting community violence from patients discharged from mental health services, *The British Journal of Psychiatry* **189**, 520–526.
[67] Douglas, K.S., Ogloff, J.R.P., Nicholls, T.L. & Grant, I. (1999). Assessing risk for violence among psychiatric patients: the HCR-20 violence risk assessment

scheme and the Psychopathy Checklist: Screening Version, *Journal of Consulting and Clinical Psychology* **67**, 917–930.

[68] de Vogel, V. & de Ruiter, C. (2006). Structured professional judgment of violence risk in forensic clinical practice: a prospective study into the predictive validity of the Dutch HCR-20, *Psychology, Crime and Law* **12**, 321–336.

[69] Douglas, K.S., Guy, L.S. & Weir, J. (2006). *HCR-20 Violence Risk Assessment Scheme: Overview and Annotated Bibliography*. Available at http://kdouglas.files.wordpress.com/2007/01/annotate9-current-through-10jan2007.pdf.

Further Reading

Tarasoff v. Regents of the University of California, 17 Cal. 3d 425, 131 Cal. Rptr. 14, 551 P.2d 334 (1976).

Webster, C.D., Dickens, B.M. & Addario, S.M. (1985). *Constructing Dangerousness: Scientific, Legal and Policy Implications*, University of Toronto Centre of Criminology, Toronto.

Related Articles

Civil Commitment

Catherine M. Wilson and Kevin S. Douglas

Database: Footwear and Foot Impression *see* Footwear and Foot Impressions: Databases

Databases

Introduction

Typically forensic DNA databases consist of two separate collections of profiles: a database of the profiles of individuals who have either volunteered or been compelled to submit samples, and a database of profiles obtained from samples from crime scenes or from exhibits associated with an alleged offense (Figure 1). The administrator of a database typically has capacity to compare profiles from

1. individuals to individuals;
2. crime samples to individuals; and
3. crime samples to other crime samples.

The overall database comprises (at least) two separate indices. One contains the DNA profiles from individuals, whereas the other stores DNA profiles from crimes. These separate databases can be matched internally (as depicted by arrows (a) and (c) in Figure 1). These matches seek to locate duplicate entries on the "offender database" or crimes with a common DNA profile, respectively. The databases are also matched against each other (as depicted by arrow (b) in Figure 1). This is often regarded as the most informative match process, as it links individuals on the "offender database" with profiles associated with crimes.

In a relatively short period, the growth of DNA databases internationally has been rapid with tens of millions of short tandem repeat (STR) profiles now held from convicted offenders, suspects, and unsolved crimes. Links provided through DNA database searches have contributed valuable intelligence to literally millions of criminal investigations. Often links are provided for crimes which are notoriously difficult to resolve, such as property crime (such as burglary and vehicle theft) and historic unsolved crimes (commonly referred to as "cold cases"). The global scale of DNA database use, its relative infancy, the complexity of addressing requisite socio-legal concerns, and the expanding capability of forensic DNA profiling combine to create a challenging law enforcement tool that demands careful assessment and management. This section introduces the emergence of forensic DNA databases in more detail and isolates aspects of their development for more detailed consideration.

A Brief Summary of National DNA Database Programs

Over the past decade, the establishment of a forensic DNA database has been a focal point of development for police and forensic agencies. There has been wide acceptance of the concept, and large-scale DNA

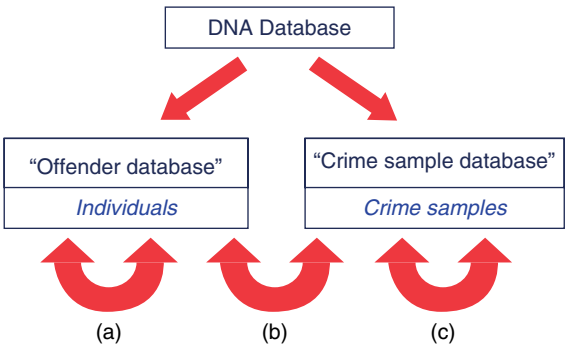


Figure 1 Summary of functionality of a standard DNA database system

database operations now exist in most developed countries.

As the initial National DNA Database, the United Kingdom has benefited from a broad legislative regime and consistent funding and has grown remarkably since 1995. The UK National DNA Database (NDNAD) now contains over 4 400 000 person profiles (as at October 2008) and over 350 000 crime profiles, and has contributed investigative links in over 750 000 cases. National DNA databases also exist in 27 European Union countries: Austria, Belgium, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, The Netherlands, Norway, Portugal, Slovenia, Spain, Sweden, Switzerland, and Ukraine (data provided by European Network of Forensic Science Institutes (ENFSI), see http://www.enfsi.eu/). By late 2007, there were over 1 500 000 million person profiles and 350 000 crime profiles on European databases. These had led to over 200 000 links involving previously unsolved crimes.

In 1989 the Federal Bureau of Investigation (FBI) launched the Combined Offender DNA Index System (CODIS) as a pilot program. This was fortified in 1994 with the creation of the *DNA Identification Act*. The technology was standardized to a panel of 13 STR loci (known colloquially as the CODIS loci). All 50 states have enacted legislation to establish a State Index, and once uploaded at the State level, the data are combined at the National level through CODIS. In August 2008, there were over 6 200 000 individual profiles and 233 000 crime profiles on CODIS. There have been more than 74 000 investigations aided through CODIS over its 10-year history. The information management system that operates CODIS is made freely available by the FBI and has been adopted in at least 27 countries.

The Canadian Government committed to DNA databasing by introducing the *DNA Identification Act in 1998*, and implementing the national database in June 2000. The National DNA Database of Canada has shown consistent growth over its eight-year history. Currently (August 2008) the National DNA Database holds over 143 000 person profiles and 43 000 crime profiles and has contributed links to over 11 500 previously unsolved crimes.

China recently established a National DNA Database program that grew from 28 000 in 2005 to over 1 500 000 by 2008. Profiles are contributed by over 200 laboratories including Hong Kong, which itself has over 20 000 profiles on a database that was established in 2001. Japan began DNA database operations in December 2004 and each province is equipped for standardized DNA analysis. Singapore has over 10 000 profiles on a successful National system. Smaller, more restrictive national database programs exists in Korea and Taiwan but other major Asian countries such as Malaysia, Thailand, and Indonesia currently have no National DNA database system. There is no National DNA database in India, although mechanisms to develop one are underway. A draft Bill was released in February 2006 advocating the sharing of State data at the national level.

Australia is a federation of six States and two Territories. Each has implemented a DNA database, with the first (Victoria) beginning in 1997. Because of difficulties, harmonizing laws data was only combined onto a National system in 2007. The National Criminal identification DNA Database (NCIDD) is managed by the Federal agency CrimTrac and now

holds over 315 000 person profiles and 110 000 crime profiles. Since linking the databases, there have been over 9000 interstate links. The New Zealand National DNA database began in 1995 as the second national database in the world. In global terms, it remains small with over 75 000 person samples and 18 000 crime profiles but it contains profiles from 2.1% of the New Zealand population which is a higher proportion than countries such as the United States (1.7%) and Canada (0.5%) but lower than the United Kingdom (7.0%). The New Zealand DNA database has a high crime-to-person hit rate (almost 60% of crimes loaded) and has contributed over 10 500 links to unsolved investigations.

There are major initiatives to develop the capacity for DNA profiling and database use in Africa. Botswana established a DNA database in 2002. A DNA Project underway in the Republic of South Africa has seen over 80 000 profiles loaded on the Republic of South Africa database. A Bill was proposed in 2004 to allow the construction of a DNA database in Israel, although operations are yet to commence. Moves are also underway to unify database operations across the United Arab Emirates.

An appealing aspect of forensic DNA technology has always been the potential for the standardized STR outputs to be shared widely by police and/or forensic agencies. While this widespread international exchange has yet to be reached, the recent heightened awareness around the threat of transnational crime and terrorism has precipitated significant early steps toward the formation of international DNA database capabilities. In Europe, the Council of the European Union released the first resolution covering the exchange of DNA analysis results in 1997. The resolution called on Member States to consider establishing their own national DNA databases and to agree common standards for DNA profiling to facilitate the exchange of data. This approach was acknowledged as an important tool for investigating and combating cross-border crime. Work on agreeing common standards for DNA profiling was being advanced under the umbrella of the ENFSI. The Prüm Treaty, signed by Belgium, Germany, France, Luxembourg, the Netherlands, Austria, and Spain in March 2006, provides for enhanced cross-border cooperation of the police and judicial authorities. The signatories agreed to give one another access to their DNA and fingerprint files using a *hit/no hit* system. This approach creates a necessary distinction between databasing and speculative searching of DNA profiles between states and the preferred process of exchanging specific profiles during serious international investigations.

An example of bona fide international databasing is the Interpol DNA Database, set up by the Interpol General Secretariat in 2003. The database (or Gateway) provides a resource through which Interpol's member countries can exchange and compare DNA profile data. Access to the Interpol database (by what are termed "beneficiaries" or "users") is allowable only following a written undertaking. Existing users can also object to any new beneficiary being granted access. The submitting countries retain ownership of the profiles and have direct control of submission, access, and deletion, in accordance with (their own) national legislation. Once a match occurs and the submitting country has been notified, then that country can communicate or request additional material to or from another country, subject to restrictions imposed by that country. This framework is aimed at meeting understandable concerns about privacy of, and control of, profiles once they leave national borders.

To date, wholesale contributions to the Interpol Database have been limited to a few countries. Presumably, this is due to the realization that for most crime types there is limited intelligence value in exporting large quantities of DNA profiles for international searching. By September 2006 Interpol's DNA database held just over 60 000 profiles from 39 countries and had produced 95 hits. Several of the successes have involved multiple crimes, some of them most serious, over multiple countries. Given that in all cases the act of searching involved only adding a profile already collected and analyzed, each of these instances of success must be seen as adding quite significant value toward achieving an otherwise improbable identification.

From the selected countries mentioned in the above summary alone, there are now over 15 000 000 profiles held on National DNA database systems around the world. Collectively they have contributed links to over 1.1 million unsolved crimes. This represents the product of considerable investment from national governments and the police and forensic agencies responsible for law enforcement in these territories. Although, while this progress is impressive, DNA database operations are still relatively recent

additions to the criminal justice system and the principal focus to date has been on their establishment and growth. It is important that we also examine the operational and socio-legal impact the databases have had on the criminal justice system and how we can learn from experiences to date and ensure that continued positive outcomes will flow from their future use.

Operational Impact of Forensic DNA Databases

As the technology that forms the basis for DNA intelligence databases is specialized, the operational components have remained the responsibility of forensic biology laboratories. In general, the database and its products are the property of law enforcement agencies with the analytical and matching processes administered on their behalf by forensic institutions. All aspects of the process, whether handled by police or scientists, are subject to governing legislation. Often this legislation contains clauses that facilitate external review of operations by delegated parliamentary authorities. From a forensic scientist's perspective, the legal basis for the administration of DNA databases represents an additional level of governance over their work. DNA laws typically contain sections that prescribe the appropriate conditions under which a DNA sample can be collected, analyzed, and stored, and the criminal sanctions that are enforceable for individuals in breach of these requirements. Although not possible to itemize all the various offense categories here, they generally include intentionally or recklessly supplying forensic material for analysis, improperly accessing or disseminating information stored on the DNA database, and matching profiles on the database unlawfully. Penalties can include fines and/or prison sentences.

In a practical context, the impact of the operational management of forensic DNA databases has had a much more profound effect on forensic organizations than the need to adjust processes to adhere to the governing legislation. The snapshot of global database models provided above illustrates that a unifying trend in major jurisdictions has been a steady increase in the scale of database operations. While this has occurred in concert with increasingly broad legislative regimes, it is more likely to have been a critical impetus for iterative legislative expansion.

The workload generated through DNA database operations has compelled forensic laboratories to find efficiency gains in the analytical process. The workload generated through DNA database operations has been a major burden on forensic DNA providers. Many have simply not coped with the volume of submissions and have experienced considerable backlogs and case processing delays. In the United States, in 2003, there were over 540 000 unsolved homicide, rape, and property crime cases in the hands of police or forensic agencies awaiting DNA testing [1]. Considerable resources have been directed toward this issue by the US Government (USDoJ, 2003) [2], the National Institute of Justice, and the American Society of Crime Laboratory Directors (ASCLD) [3], but the issue remains an intractable one and one that commonly effects major forensic organizations.

This circumstance has compelled forensic laboratories to find efficiency gains in the analytical process and has resulted in a vastly different model of case management. Historically, police and forensic scientists have had a focus on clearing one crime at a time. Cases were submitted to the forensic laboratory after considerable investigation had occurred and, in most cases, a suspect had been identified. The role of the scientist was to process these cases and determine whether there was evidence that could assist either the prosecution or defense of that specific crime. Occasionally, there were cases where the forensic analysis contributed vital investigative information that assisted the location and arrest of a suspect. The broad utilization of DNA technology and, in particular, the advent of DNA databasing has seen this paradigm change. Essentially, the analysis of evidence items occurs earlier in the investigative process with the hope that the scientist can produce intelligence information (such as a database link) in crimes for which no suspect has been identified through other means. The increased volume of submissions has also seen a trend away from individual case management and toward batch processing. It has also led to a greater prevalence of automated laboratory techniques that seek to achieve high-throughput analysis without extensive human involvement at all.

The combination of high analytical demand and the considerable funding associated with forensic DNA analysis has created an opportunity for commercial genetic testing providers to enter the forensic *market*. This is a common feature of the US and UK landscapes. "Forensic" samples (exhibits relating to the alleged offense) typically remain the responsibility of the *crime lab*, but private companies are

contracted to analyze *offender samples* on behalf of the State.

Another trend associated with the emergence of DNA database operations has been the broadening range of crimes submitted to forensic biology laboratories for analysis. In the 1980s, DNA profiling was primarily used to solve serious crimes. It now contributes to the investigation of a broad spectrum of crimes, including property offenses such as burglary. The variation in jurisdictional, legislative, and operational frameworks can impact the case submission profile; for example, certain countries in Europe initially focused on the investigation of serious violent and sexual offenses [4]. In general, there has been a clear pattern of decrease in the proportion of cases from serious crime categories and an increase in the proportion of cases submitted from volume crime categories [5]. Data from the New South Wales state laboratory in Australia demonstrates a significant decrease in proportional levels of sexual assault (53.0–8.0%), murder and manslaughter (8.7–1.0%), malicious wounding (8.9–2.9%), and attempt to murder (3.6–0.3%) submissions between 1998 and 2005 (database operations began in 2001). The case categories where there was a significant increase in proportional submissions include break, enter, and steal (2.7–37.8%); stolen vehicle (0.0–15.6%); robbery (6.3–12.8%); stolen from vehicle (0.0–4.9%); and malicious damage (0.0–3.3%) [5]. In all categories the most significant change in the case submission profile occurred after 2001, indicating that the profile of casework received by the laboratory has changed since the beginning of DNA database operations.

The changing nature of case submissions, represented above, has also been accompanied by an associated change in the type of exhibit located, and hence the type of sample(s) presented for DNA analysis. For example, serious, violent crimes would be expected to result in injuries and bleeding from those associated and therefore a higher likelihood of receiving blood as the principal evidence type. Likewise, crimes of a sexual nature predominantly involve a male offender and hence will typically be associated with semen evidence. Property crimes and drug crimes are less simple to classify in this way and are not thought to be strongly associated with a characteristic evidence (or sample) type. A general trend, however, is that these types of incidents often result in the submission of more discrete evidence types, such as cigarette butts, drinking containers, food remnants, tools or swabs from surfaces, or objects that the offenders are believed to have touched or handled (commonly referred to as "trace DNA") [6].

Investigative and Intelligence Use of Forensic DNA Databases

As mentioned earlier, the establishment of forensic DNA databases has been a contributor to a change in thinking within forensic science. Historically, each case was processed as a unit or perhaps as a small series. The analyst adopted a *cradle-to-grave* approach and usually sought to refute or corroborate the version(s) of events suggested during either the investigation or trial. Typically, a suspect had been identified through other means and the forensic analysis was comparative rather than suggestive.

This model has changed significantly. Forensic DNA databases have catalyzed profound changes in the volume and profile of forensic case submissions, meaning that with the exception of certain serious crimes, the focus has shifted to a batch-processing model within which the DNA database is an embedded component and where all cases and individual samples are viewed as potentially linked. At an organizational level, the critical objective is efficiency: maximizing the processing capacity in the minimal time and with the minimal associated expense. These objectives have typically been set and driven by police users as they align best with their investigative priorities.

The ability to link cases and identify a crime series has been an effective strategy in law enforcement for some time [7]. Incorporating DNA outcomes into the existing array of investigative data provides an additional mechanism to link cases committed by the same individual or organization, adding also a highly discriminating mechanism for identification. Through the ability to provide technical information capable of directing police investigations (such as by identifying a suspect or a crime-to-crime link), DNA databases are able to operate as an intelligence tool. Another important investigative feature of DNA databases is their ability to transcend jurisdictional boundaries, which may have hampered abilities of law enforcement agencies to link crimes and offenders which transcend such boundaries. This is particularly true in large countries, for example, the United States,

Canada, and Australia, where states or provinces have the responsibility for law enforcement within particular physical and legislative boundaries. This more proactive use of forensic DNA outcomes also presents opportunities for more integrated and intuitive use of the technology. Achieving this relies upon a complete understanding of the nature and scope of the investigative contribution that forensic DNA profiling can make, presently and in the future.

The amalgamation of DNA profile data from numerous crimes also provides a means to examine forensic outcomes holistically for tactical or operational intelligence. The cases and crime information residing on forensic DNA databases (and indeed other forensic holdings) contain vast records of links made through scientific analyses, covering an amalgamation of crime types and localities that often remain separate in police investigative and intelligence structures. Analyzed collectively, database outcomes also offer significant value to fields such as criminology. A future step in the use of forensic DNA databases could be to extend the operational success of DNA databases to allow greater analysis of trends in criminal behavior. This could lead to proactive policing strategies that are based on a contemporary understanding of criminality.

Familial Searching

Recently, DNA databases have been utilized to assess profiles that are returned after a search with a high level of allele sharing. This is due to the increased likelihood that such a result would be returned from a person related to the donor of the comparison profile, than from an unrelated, random individual. This application of the use of forensic DNA databases is known as *familial searching* and was pioneered by the UK Forensic Science Service (FSS) as a legitimate investigative use of the UK NDNAD. These searches produce leads that are refined through combination with other investigative information such as geography, age of the donor, Y-chromosome testing, and so on.

Familial searching was first applied by the FSS in 2002 in the case of serial murder of three women in South Wales, which had occurred in 1973. A DNA profile taken from crime scene evidence was used for comparison and provided investigators with a familial match to Paul Kappen. Police investigation then led them to his father, Joseph Kappen, who was identified as the rapist and murderer of the three women.

While there have been celebrated examples of success from cases of familial searching, a degree of challenge and adverse commentary has begun to surface. For example, concerns have been raised that minority groups will be unfairly targeted when from a sweep of partial matches, as these groups are overrepresented in convicted offender databanks. Other opponents to this technique claim that law enforcement will waste time and resources following up on a variety of leads, many of which may be false and have no bearing on the case. However, when investigating a crime, a variety of leads must be considered and pursued; some will have merit and others will not. Through this technique forensic science has the potential to generate leads for investigative consideration, some of which return results of great significance.

Cold Case Review and Postconviction Testing

Forensic DNA testing and state and national databases have been used retrospectively to reexamine evidence from historic, unsolved offenses originating as far back as the 1970s. In some circumstances, this analysis has not only identified a suspect for the crime in question but has proven the innocence of an already incarcerated person. In America, through the New York Innocence Project (http://www.innocenceproject.org/) based at the Cardozo School of Law, New York, 217 persons have been exonerated (as on June 1, 2008) following DNA testing of exhibits related to their original convictions. Collectively, these men have served over 2500 years in prison for crimes that they never committed, with the average jail term being in excess of 12 years. In 16 cases, the death penalty had been imposed before an individual's innocence was proved. About 70% of those exonerated were members of minority groups. Additionally, in over 35% of cases involving postconviction DNA exoneration, the true perpetrator was also correctly identified through the use of DNA testing.

As of July 2007, 42 states of the United States had some form of law permitting inmates' access to DNA testing. The remaining eight states had no law granting such access (National Conference of State Legislatures). The legislative models vary in

the extent of provisions they afford to inmates seeking DNA retesting. In some states, all incarcerated felons are granted access to postconviction DNA testing with the associated costs borne by state authorities. In others, there are restrictions on the eligibility of certain inmates: for example, those who pleaded guilty or whose lawyers failed to request DNA testing at trial. In some models there are time limits on when an application can be made, and the petitioning inmate must meet the cost of reanalysis. As with all laws, there is a need to strike a balance, in this case between the rights of incarcerated felons to have their convictions reviewed, and the potential for misuse of state resources and the protracted continuation of criminal matters.

The revelations of programs such as the innocence project provide an example of the potential for DNA testing to uncover the truth – whether it does so on the "side" of the victim or their wrongly convicted attacker. However, it is important that we also remember that DNA is simply another example of forensic evidence. While it allows a high level of certainty regarding its conclusions, it is not infallible and does not have the same effect when used in a stand-alone capacity, in the absence of support from investigative outcomes or other forensic evidence. Also, obtaining meaningful DNA evidence relies on the accurate localization of biological material at the crime scene and the demonstration of appropriate chain of custody and procedural handling of the subsequent analysis. Operating in this way is commonplace in contemporary investigations; however; it is less assured in historic investigations, and in some cases the lower level of rigor applied through outdated practices can jeopardize the usefulness or admissibility of forensic evidence recovered later using contemporary techniques.

For postconviction testing to be of future use, it is incumbent that all relevant evidence be collected and stored appropriately so as to allow for subsequent reanalysis. Again, it is necessary to strike a balance between collecting and storing everything indefinitely and destroying vital items of evidence too hurriedly. This is another area to be considered in postconviction legislative models and in best-practice policies and procedures for police and forensic agencies.

The realization of the ability of a "new" technology such as DNA analysis to provide greater clarity and certainty in historic investigations tells us something else about our field of science. That is, that we should continue to strive to adapt and develop our technical capability. The impressive outcomes of postconviction use of DNA testing provide a stark reminder of how the technologies of tomorrow may well provide the ability to reflect with more clarity and certainty on cases occurring today.

Socio-Legal Issues Associated with Forensic DNA Databases

It is important to remember that although DNA profiling is arguably the most powerful and rapidly developing forensic investigative tool, its use is not without controversy. This is due largely to its enormous potential to implicate individuals in a crime and the risk of its deliberate or accidental misuse. The present and future uses of forensic DNA profiling are a worrying prospect for some members of the international criminal justice community and for members of the public as well.

Almost ubiquitously, the establishment of DNA databases on a state or national level has required substantial changes in legislation. There is considerable variation in the legislative models applied in different jurisdictions. Initially, offenders convicted of serious violent or sexual crimes were seen as the most suitable candidates for DNA database inclusion. Over time, the gambit of offenses for which a compulsory DNA sample can be obtained has increased to include property crimes and, in some places, summary offenses. In addition, the need for an individual to be convicted is no longer mandatory in many jurisdictions, and individuals can be sampled upon suspicion or arrest. In all cases, the legislation seeks to strike a balance between the desire for the state to develop and utilize a significant forensic resource for the purpose of more effective crime resolution and reduction, and the rights of individuals exposed to the criminal justice system.

Legislative amendments of the type that lead to the establishment of forensic DNA databases nature are destined to be controversial. In some cases, observers feel that these new laws have been enacted too hastily, have lacked suitable public and legal scrutiny, and have been justified under misleading, populist "law and order" politicking. Others worry that extending police powers to allow the collection of DNA samples represents an encroachment into the previously sacred territory of criminal law and

a diminution of basic individual rights, in particular the right to silence and the right against self-incrimination. Other observers are concerned by the storage of human genetic information and its potential for future misuse.

Forensic scientists utilizing forensic DNA databases have an increased need for awareness of relevant socio-legal issues. In a practical context, DNA-based legislation represents an additional level of governance for forensic professionals and one of the first pieces of law that places direct requirements on the manner in which they undertake their professional work. In addition, it prescribes sanctions for individuals or institutions who contravene the administrative processes detailed in these laws. The debate that has continued on many of the issues associated with the use of forensic DNA profiling in the criminal justice system has expanded to encompass applications of the scientific process that are primarily the responsibility of the forensic community. Forensic professionals (and particularly the administrators of forensic institutions) must acquaint themselves with these issues and enter the existing debate. Failure to do so could mean that the direction for the application of our scientific tools will become the responsibility of people from outside the forensic community itself.

Summary

Despite the rather spectacular results that have emerged from the use of forensic DNA databases, the technology remains a relatively recent development and one that will continue to develop over time. It is important to remember that as databases grow and age, they will become more challenging to manage, as much of the information stored on them will become increasingly redundant. It is important that forensic practitioners and administrators continue to research and refine DNA database applications to ensure they continue to have an effective, positive impact.

References

[1] Lovrich, N.P., Gaffney, M.J., Pratt, T.C., Johnson, C.L., Asplen, C.H., Hurst, L.H. & Shellberg, T.M. (2003). *National Forensic DNA Study Report*, 12 December 2003.

[2] Office of the Inspector General (2004). *Report on the No Suspect Casework DNA Backlog Reduction Program*, Audit Report No. 05-02, November 2004.

[3] 180 Day Study Report: Status and Needs of United States Crime Laboratories, Report of the American Society of Crime Laboratory Directors, 28 May 2004.

[4] Schneider, P.M. & Martin, P.D. (2001). Criminal DNA databases: the European situation, *Forensic Science International* **119**, 232–238.

[5] Walsh, S.J. (2007). Current and future trends in forensic molecular biology in *Molecular Forensics*, R. Rapley & D. Whitehouse, eds, John Wiley & Sons, London, pp. 1–20.

[6] Raymond, J.J., Walsh, S.J., van Oorschot, R.A., Gunn, P.R. & Roux, C. (2004). Trace DNA: an underutilised resource or Pandora's Box? *Journal of Forensic Identification* **56**(4), 668–686.

[7] Gotleib, S. (1998). *Crime Analysis – From First Report to Final Arrest*, Alpha Publishing, Montclair, CA.

Related Articles

DNA

DNA Databases and Evidentiary Issues

DNA: an Overview

SIMON J. WALSH

Dating: Bomb Pulse *see* Bomb-Pulse Dating

Dating: Document

Introduction

In questioned document examinations, the age of a document is often an important criterion to determine authenticity or highlight fraud. The following questions are therefore frequently asked to the expert: Was a receipt actually produced at the mentioned date? Was a contract signed at the same time by the involved stakeholders? Was a paragraph added several years after the establishment of a testament? To answer these questions, it is necessary to determine the date at which a document was written or printed.

The time frame may be a few months or several years. As a result, many forensic scientists investigated the issue of dating documents. Every book about document analysis has a chapter or paragraph about dating methods [1–8]. The dating can focus on different aspects of questioned documents. Whereas dating through paper, toner, or handwriting examinations is less considered in the literature, the potential of inks has been more particularly studied and many methods have been proposed so far [9–35]. Several reviews have been published on the topic over the last 30 years [36–45]. Although a number of laboratories apply ink dating in caseworks, the method raises a large amount of controversy and debates among the scientific community [39, 45–61].

The objective of this contribution is to give a comprehensive overview on the dating of questioned documents, and more particularly their ink. In the first part, three fundamental approaches to dating are presented with examples of dating methods. This article also introduces the problematic of method validation and interpretation of dating evidence.

Dating Principles

Three fundamental approaches can be defined and formalized [62] (Figure 1). The first, often called the "*static approach*" [41], focuses on the production date of documents raw constituents and their introduction on the market. The second, addressed as the "dynamic approach" [42], is based on aging processes of documents. The third approach aims at reconstructing the chronology among documents or ink entries by ordering them in sequence. Differentiation between "absolute" and "relative" age of documents is usually made in the literature [15, 16, 18, 46, 50,

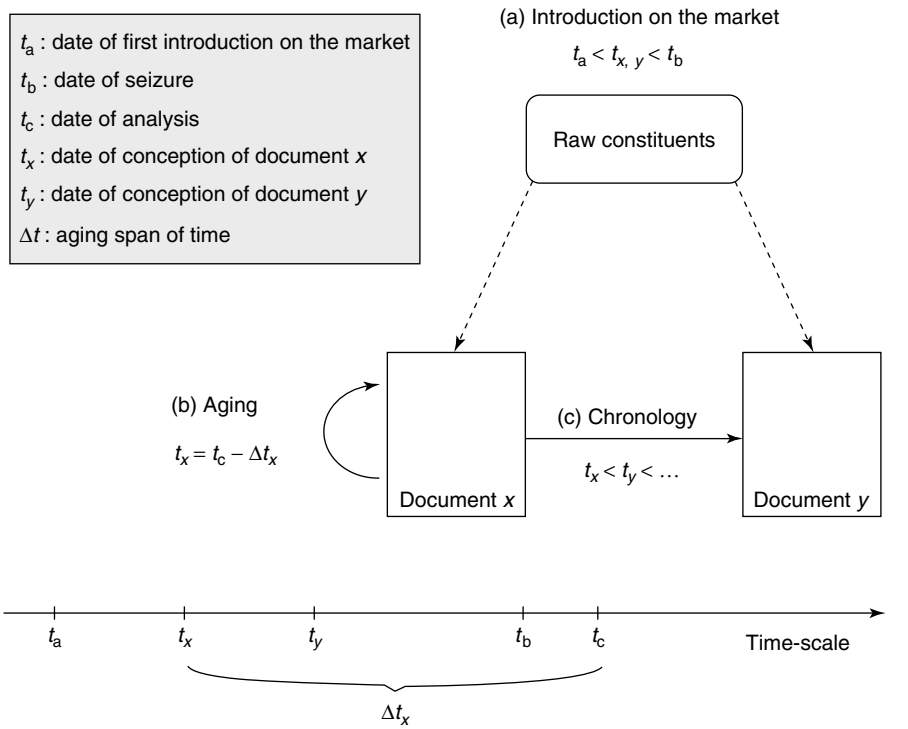


Figure 1 Age determination of questioned documents relies on (a) the date of introduction of raw constituents on the market, (b) aging of ink since document conception, and (c) reconstruction of the chronology of document conceptions. The date of first introduction of the market (t_a), the date of seizure (t_b), the date or analysis (t_c) and the aging time span (Δt) are known and/or yielded by the investigation. These pieces of information then help answer the question about the conception time of the documents (t_x and t_y)

52]. Determining the relative age of a document compared to others is equivalent to reconstructing their chronology.

Introduction on the Market

An estimation of the age of documents can be based on the analysis of stable compounds, found in the paper, ink, or toner, that are specific to a certain period in time. Fabrication properties change and evolve with time as a function of new industrial developments and progress. For example, ballpoint pens appeared on the market in 1944 and were characterized by oil-based inks up to 1951, when glycols solvents were introduced in their ink composition [63, 64]. Fluorescent brighteners have been introduced in paper composition in the 1950s [65]. Copper phthalocyanine dyes and pigments were introduced in 1954 in blue ink for their good light stability. Gel pens were first found in Japan in 1986 [66]. This approach is generally called "*static*" in the literature because the measured parameters are invariable in time. It allows the determination of the first possible date of existence for a given composition of ink, paper, or toner found on a questioned document. Incoherencies and anachronisms can be highlighted, as was the case for the paper analysis of Hitler's diaries [65].

The developments mentioned above are quite important and relatively easy to measure, but they occur quite infrequently. Most evolutions in the composition of paper, ink, and toner are less easily detected and are closely guarded industrial secrets. To highlight the date of introduction on the market of minor characteristics, two conditions must be respected:

1. First the determination of the composition must be based on validated and reproducible analytical methods **Ink Analysis** in order to insure that the selected set of parameters does not alter as a function of time (i.e., their aging is negligible).
2. Secondly, it is essential to have access to a comprehensive database containing the selected parameters of all inks, toners, and paper introduced on the market. Ideally this should cover a broad spatial and temporal range.

An alternative method is to introduce chosen markers in the composition that point at the time of introduction on the market. For example, the Bureau of Alcohol, Tobacco, and Firearms (ATF) started a yearly ink tagging program in collaboration with ink manufacturers in the mid-1970s [8, 41, 67]. The objective was to introduce a different tag each year that was easily analyzable. This required close collaboration with ink manufacturers and a constant monitoring of the introduced tags (which should never repeat) and can generate a nonnegligible increase of the ink costs. Some special papers also enclose their date of production in the form of microimpressions or indirectly through specific watermarks [68].

Aging

The second approach is based on measurements of the questioned documents parameters that change as a function of time. Some older methods such as chloride and sulfate ions migration from the ink into the paper [9–12] can no longer be used, as most inks are now free of these ions. Earlier gallotannic inks were also acidic, and caused paper deterioration. Moreover, they contained iron that oxidized, provoking a change of color [1, 2, 13]. The introduction of new composition along the years also meant changes in the aging processes and dating methods.

Paper was studied to a lesser extent than ink [69, **Paper Analysis**], because it starts aging directly after manufacture (or earlier when the tree was felled), and that can be well before the document conception. On the contrary, it is expected that ink does not age in the cartridge [70], but only after it is apposed on paper: dyes fade [13, 15, 16, 19, 22, 71–75], solvents diffuse and evaporate [24–28, 67, 76], and resins polymerize [76–79]. Aging processes of ink follow complex pathways that are considerably influenced by a number of factors other than time, inducing acceleration or quenching of the aging. The influencing factors can be ordered in three main classes [45, 74, 76]: (i) initial composition of the ink (in the cartridge), (ii) physical and chemical properties of the substrate (paper porosity and coatings, etc.), and (iii) storage conditions (temperature, light, air flux, humidity, etc.). In practice, no information on these factors is generally available. This is why the determination of the absolute age of an ink entry remains truly difficult and in most cases impossible. The more important objective is therefore the determination of a time range rather than a precise date. The

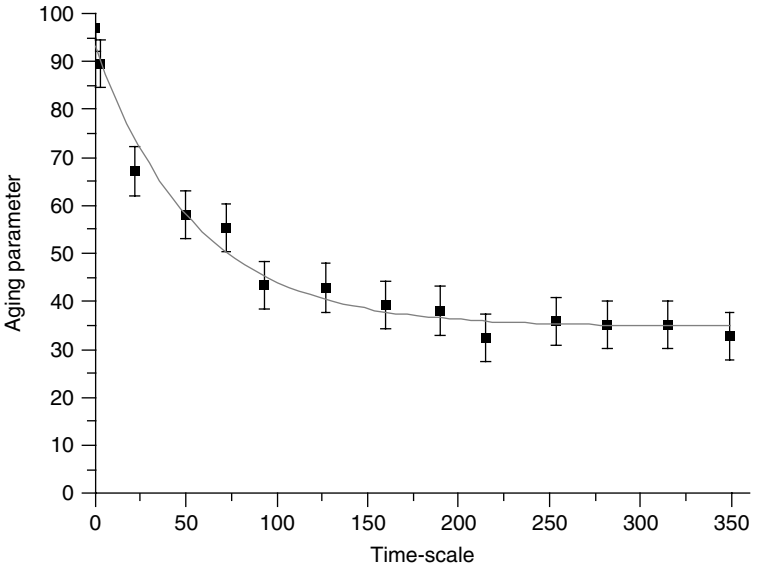


Figure 2 Typical aging curve (exponential decrease of first order): an aging parameter of the ink is plotted as a function of time (for example in seconds, hours, days, or years) [Reproduced with permission from Polymedia Meichtry SA.]

considered scale can significantly vary depending on the measured parameters. Thus, while solvents disappear very quickly from the ink, dyes degradation occurs more slowly over several years. Measured changes are reported as a function of time in order to establish an aging curve (Figure 2).

As fading of dyes is visible to the naked eye, many dating methods based on the aging approach were developed to measure color degradation with time [13, 20, 22, 71, 73]. Unfortunately for forensic scientists, dyes that are unstable in the presence of light do not degrade in the dark, or do so only very slowly [72, 75]. Therefore, dating relying on dye degradation usually is carried out only by comparing ink entries from the same pen that were stored in the same conditions. Unlike reported in [35], the absence of difference in the degradation states of two ink entries does not necessarily mean that they are contemporaneous. In fact, the inks could also have been stored for a significantly different number of years in the absence of light [45].

Lately, interest was revived into a method first proposed in 1985 by Stewart [24], in which evaporation of the volatile components of ink is measured using gas chromatography mass spectrometry (GC/MS) [26–28]. However, after a few months their quantity is generally too low to be reproducible and pertinent, as evaporation and diffusion of solvents happen promptly after ink apposition on the paper [27, 57, 76].

When ink dries, resins polymerize inducing fastening of the ink on paper. Resins are high-molecular-weight molecules, which are present in low quantity in the ink entries. They are therefore very difficult to analyze [77–79]. An indirect way to evaluate their hardening derives from the measurements of ink extractability over time. Observations have shown that generally as ink ages it becomes more difficult to extract. Earlier, changes in the extractability of ink have been investigated by measuring the dissolution rates in acids [4]. Later, sequential extraction of the ink was proposed, in order to eliminate the dependence of the extracted mass (i.e., difference in ink thickness provoked difference in the mass obtained). Thus, two extractions were carried out consecutively, first in the "weak" solvent, then in a "strong" solvent. Then extractions are analyzed to quantify either dyes [15, 16, 21, 23] or solvents [25, 51, 52] in order to calculate an extraction percentage (P). The extracted quantity in the "weak" solvent (M_1) is divided by the total amount extracted in the two solvents ($M_1 + M_2$) in order to obtain the following ratio:

$$P(\%) = \frac{100 \cdot (M_1)}{(M_1 + M_2)} \quad (1)$$

These methods are based on the following hypothesis: a "weak" solvent extracts only the fraction of the ink that is still fresh, while a "strong" solvent extracts the totality of dyes of solvents in the ink entry. Thus, a large P means that the ink is still fresh, while a low P indicates an old ink.

Since then, many authors showed that the measurements based on dyes extraction were not reproducible [39, 47, 48, 51–53, 58]. This is due to the fact that the quantity of dye extracted is dependent not on the capacity of the solvents to extract them, but also on their initial quantity in the ink and their subsequent degradation. It was observed that every ink did require different extraction solvents and their aging curves were significantly divergent. Some inks even showed an inversion of the tendency as a function of their age: their P value increased with time instead of decreasing as was expected [23, 55].

New methods based on sequential extraction of solvents were recently developed [27]. Earlier, Aginsky [25] proposed a method that made use of artificial aging in order to calculate a portion of the aging curve. However, the relevance of artificial aging was also questioned [45, 46, 50].

Chronology

This approach aims to determine the relative age of a document in comparison to others (i.e., to order them in chronological sequence). Different methodologies can be applied, such as the study of latent writings by oblique lighting or electrostatic detection apparatus (ESDA). If latent writings from a document are found on a second document, then the latter was physically placed under the first one during writing. This information may be useful in sequencing the order of writings.

The determination of the sequence of crossing lines proved to be very useful in certain cases [29–33, 45, **Intersecting Lines: Documents**]. It is, however, not always easy to determine optically which line is above the other. Therefore, chemical [45] and spatial techniques [30–33, 79, 80] were developed and proved useful in some cases.

Finally, the comparison of ink aging states may help to reconstruct the sequence of apposition of ink entries on documents [15, 16, 18, 46, 50, 52]. This can only be applied for inks stored in the same conditions on the same type of paper (i.e., diaries). The general evolution of the aging curve must be known (Figure 3). If a diminution is expected, it is imperative to insure the aging parameters never increase whatever the conditions.

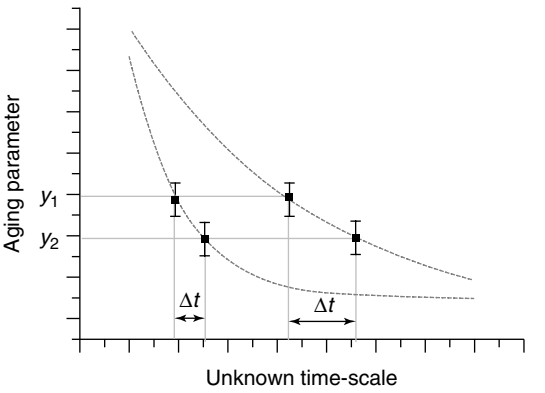


Figure 3 If the aging curve is unknown (dots), the values obtained from two ink entries can be ordered in chronological order at the imperative condition that the general evolution of the curve is known (increase or decrease). Storage conditions, paper type, and ink composition must be the same. Measurement error must be taken into account [Reproduced with permission from Polymedia Meichtry SA.]

Validation

The analytical development of dating methods (Table 1) requires a considerable amount of time and resources. It is therefore important not to underestimate the task of ensuring their scientific validity before implementing them in practice. Some important aspects of analytical reliability were enounced by Horwitz [82]: reproducibility (between laboratories precision), repeatability (within laboratory precision), systematic error or bias (accuracy), selectivity, and limit of reliability. The robustness of the method is also an important factor.

It is of concern that errors are very rarely mentioned in literature and are generally not represented in the figures. It is, however, essential to make certain that predicted differences provoked by aging are in fact higher than measurement errors. Further, the available ink in practical cases is generally not sufficient to repeat analysis several times in order to obtain a mean and a standard deviation [61]. Developments of dating methods are carried out with known

Table 1 Summary of document dating possibilities: relevant application time frame and main limitations of dating methods. All methods must be validated before their implementation

Approach	Time variable	Time frame	Main limitations
(1) Introduction on the market	Tags	1970s – end of project	Limited number of inks with tags
	New developments	Not defined	Secret of fabrication
(2) Aging	Dyes	Months	Unreliable
	Solvents	Weeks	Possible only for a limited number of inks
	Resins	Years	No data
	Paper	Years	Limited to historical document
(3) Chronology			
	Dyes	Months	Same ink and support (e.g., diary)
	Solvents	Weeks	Same ink and support (e.g., diary)
	Line crossings	Years	Aging of the crossings
	Latent writings	Years	Their absence gives no information

ink samples prepared and stored in controlled conditions, and blind testing on realistic samples is therefore imperative. When low quantities are analyzed, such as solvents in ink entries, the detection and quantification limits play an important role to determine a threshold at which the method is not applicable anymore. The most demanding aspect is the interlaboratory validation. In fact, till date most dating methods are used only by one single laboratory. In order to apply dating in casework, the forensic experts should deploy their efforts most particularly in adequate validation of their method [49, 51, 54, 61].

Interpretation

Interpretation of evidence, like in other forensic fields, plays an essential role in the dating of ink and should already be included in method developments. In fact, the question about the age of an ink entry lies perhaps more on the inference of sources rather than on the technological or analytical aspects. Thus, it is important to consider all the possible sources (alternative hypotheses) to allow for a balanced interpretation of the evidence [61].

A Bayesian statistical framework can be applied to assist in reaching an opinion regarding the date of a document. In this case, when one has to look at the probability of the evidence (E) given the ink entry has been made at a time t_1 (hypothesis of the plaintiff, H_p) compared to the probability of this same evidence given the ink entry has been written at a different time t_2 (hypothesis of defense, H_d). Then the prior odds of the hypotheses of the plaintiff H_p and the defense H_d existing before observation of the evidence E are multiplied by a factor called *likelihood ratio* (LR) to obtain the posterior odds that account for the new evidence E. The LR is an indication on the strength of the evidence in supporting one of the hypotheses in Bayesian logic [61, 83]. It is defined by the probability of E given H_p is true divided by the probability of E given H_d is true:

$$LR = \frac{P(E/H_p)}{P(E/H_d)} \tag{2}$$

For example, when comparing two ink entries to determine their relative age, one has to evaluate the probability of obtaining the observed results, if the entries have the same age, and the probability of obtaining these results, if the entries do not have the same age. If the dye compositions of the two ink entries were determined and found to be nondifferentiable, at least three explanations should be taken into account. In fact, the analysis of two ink entries having the same age, the same initial composition, paper, and storage conditions would logically lead to nondifferentiable results. However two inks having different ages, but the same initial composition, on the same paper and stored in the dark, would also lead to nondifferentiable results, because dye fading occurs very slowly in the absence of light [75]. Finally, two inks having different compositions and/or stored in different conditions did sometimes also yield nondifferentiable results [45].

Conclusion

From its beginning, the field of questioned documents has been concerned with dating. Proposed methods usually lean upon complex processes and controversy among the scientific community is still high. Every document dating method, whose objective is to be applied in forensic caseworks, must fulfill validation requirements. Moreover, source inference must also be taken into account in the interpretation of the dating evidence. Till date, most methods still fail to be adequately validated and should be applied with extreme caution. The limitations of the methods used must be adequately disclosed and documented.

Acknowledgments

The author would like to thank all the people who helped to achieve this work, and especially Prof. Bernahrd Spengler of the University of Giessen, Dr. Dieter Kirsch of the Bundeskriminalamt in Wiesbaden and Prof. Pierre Margot of the University of Lausanne.

References

[1] Osborn, A.S. (1910). *Questioned Documents*, The Lawyers' Co-operative Publishing Company, Rochester.

[2] Brewster, F. (1932). *Contested Documents and Forgeries*, The Book Company, Calcutta.

[3] Grant, J. (1937). *Books and Documents – Dating, Permanence and Preservation*, Grafton & Company, London.

[4] Locard, E. (1959). *Les faux en écriture et leur expertise*, Bibliothèque Scientifique, Paris.

[5] Harrison, W.R. (1966). *Suspect Documents – Their Scientific Examination*, Sweet & Maxwell, London.

[6] Brunelle, R.L. & Reed, R.W. (1984). *Forensic Examination of Ink and Paper*, Charles C Thomas Publisher, Springfield.

[7] Ellen, D. (1997). *The Scientific Examination of Documents – Methods and Techniques*, *Forensic Science Series*, Taylor & Francis, Salisbury.

[8] Levinson, J. (2001). *Questioned Documents – A Lawyer's Handbook*, Academic Press, London, California.

[9] Metzger, D., Rall, D. & Heess, W. (1933). Ein neues Verfahren, Identität und Alter von Tintenschriften festzustellen, *Archiv für Kriminologie* **92**, 105–113.

[10] Türkel, S. (1933). Eine Schnellmethode zur ersten Orientierung welche von zwei Tintenspuren auf einem Papiere jüngeren oder älteren Datums ist, *Archiv für Kriminologie* **92**, 224–226.

[11] Heess, W. (1935). Ein neues Verfahren, Identität und Alter von Tintenschriften festzustellen-2. Mitteilung, *Archiv für Kriminologie* **96**, 13–17.

[12] Heess, W. (1937). Ein neues Verfahren, Identität und Alter von Tintenschriften festzustellen-3. Mitteilung, *Archiv für Kriminologie* **101**, 7–37.

[13] Sen, N.K. & Ghosh, P.C. (1971). Dating iron-base ink writings on documents, *Journal of Forensic Sciences* **16**(4), 511–520.

[14] Frei-Sulzer, M. (1971). Messungen zeitlicher Veränderungen der Schreibrillentiefe, *Kriminalistik – Ein Hilfsmittel bei der Alterfeststellung von Urkunden* **7**, 348–349.

[15] Brunelle, R.L., Breedlove, C.H. & Midkiff, C.R. (1987). Determining the relative age of ballpoint inks using a single-solvent extraction technique, *Journal of Forensic Sciences* **32**(6), 1511–1521.

[16] Cantu, A.A. & Prough, R.S. (1987). On the relative aging of ink – the solvent extraction technique, *Journal of Forensic Sciences* **32**(5), 1151–1174.

[17] Cantu, A.A. (1988). Comments on the accelerated aging of ink, *Journal of Forensic Sciences* **33**(3), 744–750.

[18] Brunelle, R.L. & Lee, H. (1989). Determining the relative age of ballpoint ink using a single-solvent extraction, mass-independent approach, *Journal of Forensic Sciences* **34**(5), 1166–1182.

[19] Aginsky, V.N. (1993). Some new ideas for dating ballpoint inks – a feasibility study, *Journal of Forensic Sciences* **38**, 1134–1150.

[20] Aginsky, V.N. (1994). A new version of the internal standard method in quantitative thin layer chromatography, *Journal of Planar Chromatography* **7**, 309–314.

[21] Aginsky, V.N. (1994). Determination of the age of ballpoint pen ink by gas and densitometric thin-layer chromatography, *Journal of Chromatography A* **678**, 119–125.

[22] Aginsky, V.N. (1995). A microspectrophotometric method for dating ballpoint inks – a feasibility study, *Journal of Forensic Sciences* **40**(3), 475–478.

[23] Brunelle, R.L. & Speckin, E.J. (1998). Technical report with case studies on the accelerated aging of ballpoint inks, *International Journal of Forensic Document Examiners* **4**(3), 240–254.

[24] Stewart, L.F. (1985). Ballpoint ink age determination by volatile component comparison – a preliminary study, *Journal of Forensic Sciences* **30**(2), 405–411.

[25] Aginsky, V.N. (1996). Dating and characterizing writing, stamp pad and jet printer inks by gas chromatography/mass spectrometry, *International Journal of Forensic Document Examiners* **2**(2), 103–116.

[26] Andrasko, J. (2003). A simple method for distinguishing between fresh and old ballpoint pen ink entries (Abstract), *Forensic Science International* **136**(1), 80–81.

[27] Bügler, J.H., Buchner, H. & Dallmayer, A. (2008). Age Determination of Ballpoint Pen Ink by Thermal Desorption and Gas Chromatography-Mass Spectrometry, *Journal of Forensic Sciences* **53**(4), 982–988.

[28] Brazeau, L. & Gaudreau, M. (2007). Ballpoint pen inks: the quantitative analysis of ink solvents on paper by solid-phase microextraction, *Journal of Forensic Sciences* **52**(1), 209–215.

[29] Poulin, G. (1996). Establishing the sequence of strokes: the state of the art, *International Journal of the Forensic Document Examiners* **2**, 16–32.

[30] Kasas, S., Khanmy-Vital, A. & Dietler, G. (2001). Examination of line crossings by atomic force microscopy, *Forensic Science International* **119**, 290–298.

[31] Berx, V. & De Kinder, J. (2002). A 3D view on the 'crossing lines' problem in document investigation, in *Investigative Image Processing II*, Crossing lines problem in document investigation, in Investitive Image Processing II, *Proceedings of SPIE*, Z.J. Geradts & L.I. Rudin, eds, Vol. 4709, pp. 102–110.

[32] Berx, V. & De Kinder, J. (2003). Line crossings and laser profilometry (Abstract), *Forensic Science International* **136**(1), 81.

[33] Schirripa Spagnolo, G. (2006). Potentiality of 3D laser profilometry to determine the sequence of homogenous crossing lines on questioned documents, *Forensic Science International* **164**(2–3), 102–109.

[34] Grim, D.M., Siegel, J. & Allison, J. (2002). Evaluation of laser desorption mass spectrometric and uv accelerated aging dyes on paper as tools for the evaluation of a questioned document, *Journal of Forensic Sciences* **47**(6), 1265–1273.

[35] Hofer, R. (2004). Dating of ballpoint pen ink – case report, *Journal of Forensic Sciences* **49**(6), 1–5.

[36] Brunelle, R.L. & Cantu, A.A. (1987). A critical evaluation of current ink dating techniques, *Journal of Forensic Sciences* **32**(6), 1522–1536.

[37] Stewart, L.F. & Guertin, S.L. (1991). Current status of ink age determination, *International Criminal Police Review* (March–April), **429**, 9–12.

[38] Brunelle, R.L. (1992). Ink dating – the state of the art, *Journal of Forensic Sciences* **37**(1), 113–114.

[39] Hicks, T. (1993). Détermination de l'âge des encres – possibilités de standardisation, *Séminaire de 4ème Année*, Institut de Police Scientifique, Université de Lausanne et de Criminologie.

[40] Stewart, L.F., Becker, J.L. & Fitzgerald, M. (1995). Multi-agency response to the advancement of ink age determination. Brief communication. *47th AAFS Meeting*, Seattle.

[41] Cantu, A.A. (1995). A sketch of analytical methods for document dating. Part I. The static approach: determining age independent analytical profiles, *International Journal of Forensic Documents Examiners* **1**(1), 40–51.

[42] Cantu, A.A. (1996). A sketch of analytical methods for document dating. Part II. The dynamic approach: determining age dependent analytical profiles, *International Journal of Forensic Documents Examiners* **2**(3), 192–208.

[43] Dormann, P. (2000). Les différentes techniques et la controverse sur la datation des encres, *Séminaire de 2ème Année*, Institut de Police Scientifique et de Criminologie, Université de Lausanne.

[44] Brunelle, R.L. & Crawford, K.R. (2003). *Advances in the Forensic Analysis and Dating of Writing Ink*, Charles C Thomas Publisher, Springfield.

[45] Weyermann, C. (2005). *Mass Spectometric Investigation of the Aging Processes of Ballpoint Ink for the Examination of Questioned Documents*. PhD Thesis, Justus-Liebig University, Giessen, http://geb.uni-giessen.de/geb/volltexte/2006/3044/, dernier accès 15.10 .2007.

[46] Lyter, A.H. (1994). Relative aging of ball pen ink: natural aging vs. artificial aging. *American Academy of Forensic Sciences Meeting*, San Antonio.

[47] Starrs, J.E. (1994). Forensic science in law enforcement – ink analysis, *The Scientific Sleuthing Newsletter* **18**(4), 11.

[48] Hicks Champod, T., Khanmy, A. & Margot, P. (1995). Ink aging: perspectives on standardization, in *Proceedings of the 13th Meeting of the International Association of Forensic Sciences, Düsseldorf, 1993*, *Advances in Forensic Sciences 3, Forensic Criminalistics 1*, B. Jacob, W. Bonte, W. Huckenbeck & P. Pieper, eds, Verlag Dr. Köster, Berlin, pp. 304–309.

[49] Margot, P., Hicks Champod, T. & Khanmy, A. (1995). Ink dating up-dated and up-ended – correspondence of Prof. Starrs, *International Journal of Forensic Document Examiners* **1**(1), 2–5.

[50] Stewart, L.F. & Fortunato, S.L. (1996). Distinguishing between relative ink age determinations and the accelerated aging techniques, *International Journal of Forensic Document Examiners* **2**(1), 10–15.

[51] Aginsky, V.N. (1996). Accelerated aging – its use in methods for dating ink, *International Journal of Forensic Document Examiners* **2**(3), 179–181.

[52] Aginsky, V.N. (1998). Measuring ink extractability as a function of age – why the relative aging approach is unreliable and why it is more correct to measure ink volatile components than dyes, *International Journal of Forensic Document Examiners* **4**(3), 214–230.

[53] Andermann, T. & Neri, R. (1998). Solvent extraction techniques – possibilities for dating ball point pen inks, *Journal of Forensic Sciences* **4**(3), 231–239.

[54] Starrs, J.E. (2000). Review: the many use of the forensic sciences – document examination, *The Scientific Sleuthing Newsletter* **24**(1), 12, 17.

[55] Fortini, A. (2000). Tester la reproductibilité des mesures de l'une des méthodes de datation des encres d'Aginsky (GC/MS), *Séminaire de 4ème Année*, Institut de Police Scientifique et de Criminologie, Université de Lausanne.

[56] Weyermann, C., Kirsch, D., Costa-Vera, C. & Spengler, B. (2003). Dating of ink entries by MALDI/LDI-MS and GC/MS analysis: reality or utopia? (Abstract), *Forensic Science International* **136**(1), 71–72.

[57] Lociciro, S., Mazzella, W., Dujourdy, L., Lock, E. & Margot, P. (2004). Dynamic of the aging of ballpoint pen ink, *Science and Justice* **44**(3), 165–171.

[58] Jahns, K. (2004). Altersbestimmung von Schreibmitteln durch chemische Analysverfahren, in *Mannheimer Hefte Für Schriftvergleichung*, P.E. Baier, ed, Universität Mannheim, Schmidt/Römhild, p. 29, Jahrgang, 3/2004.
[59] White, P. (2004). The age of an ink: an age old problem, *Oral Presentation at the 17th International Symposium on Forensic Science*, ANZFSS, Wellington.
[60] Schirripa Spagnolo, G., Zavattaro, D., De Santis, M. & Gennaro Ferrillo, L. (2007). Max Frei theory revisitation: does really strokes depth change along time? *Forensic Science International* **167**, 163–166.
[61] Weyermann, C., Schiffer, B. & Margot, P. (2007). A logical approach to ballpoint ink dating, *Science and Justice* available online. DOI:10.1016/j.scijus.2007.10.009.
[62] Weyermann, C. (2008). Revue des méthodes de datation de l'encre: Est-il possible de déterminer l'âge d'un document en analysant l'encre? *Revue Internationale de Criminologie et de Police Technique et Scientifique* **3**(LXI), 359–376.
[63] Hilton, O. (1957). Characteristics of the ball point pen and its influence on handwriting identification, *Journal of Criminal Law, Criminology and Police Science* **47**, 606–613.
[64] Hung, P.S. & Leung, S.C. (1995). Some observations on the morphology of a ballpoint pen stroke, *International Journal of Forensic Document Examiners* **1**, 18–31.
[65] Grant, J. (1985). The diaries of Adolf Hitler, *Journal of the Forensic Science Society* **25**, 189.
[66] Mazzella, W.D. & Buzzini, P. (2005). Raman spectroscopy of blue gel pen inks, *Forensic Science International* **152**(2–3), 241–247.
[67] LaPorte, G.M., Wilson, J.D., Cantu, A.A., Mancke, S.A. & Fortunato, S.L. (2004). The identification of 2-phenoxyethanol in ballpoint inks using gas chromatography/mass spectrometry – relevance to ink dating, *Journal of Forensic Sciences* **49**, 155–159.
[68] Caywood, D.A. (1995). Watermarks and the questioned document examiner – an historical retrospective on watermarks and document examination, *International Journal of Forensic Document Examiners* **4**(1), 399–304.
[69] Stewart, L.F. (1982). Artificial aging of documents, *Journal of Forensic Sciences* **27**(2), 450–453.
[70] Grim, D.M., Siegel, J. & Allison, J. (2002). Does ink age inside of a pen cartridge, *Journal of Forensic Sciences* **47**, 1294–1273.
[71] Andrasko, J. (2001). HPLC analysis of ballpoint pen inks stored at different light conditions, *Journal of Forensic Sciences* **46**(1), 21–30.
[72] Andrasko, J. (2001). Changes in composition of ballpoint pen inks on aging in darkness, *Journal of Forensic Sciences* **47**(2), 324–327.
[73] Grim, D.M., Siegel, J. & Allison, J. (2001). Evaluation of desorption/ionization mass spectrometric methods in the forensic applications of the analysis of inks on paper, *Journal of Forensic Sciences* **46**(6), 1411–1420.
[74] Weyermann, C., Kirsch, D., Costa Vera, C. & Spengler, B. (2006). Forensic investigation of ageing and degradation of ballpoint dyes on paper by LDI- and MALDI-MS analysis, *Journal of the American Society for Mass Spectrometry* **17**, 297–306.
[75] Weyermann, C., Kirsch, D., Costa Vera, C. & Spengler, B. (2007). A study of the drying of ballpoint pen ink on paper by GC/MS, *Forensic Science International* **168**(2–3), 119–127.
[76] Bügler, J., Buchner, H. & Dallmayer, A. (2005). Differenzierung von Kugelschreiberpasten durch Thermodesorption und Gaschromatographie – Massenspektrometrie, *Archiv für Kriminologie* **214**, 141–148.
[77] Bügler, J., Buchner, H. & Dallmayer, A. (2005). Characterization of ballpoint pen inks by thermal desorption and gas chromatography-mass spectrometry, *Journal of Forensic Sciences* **50**(5), 1–6.
[78] Kirsch, D., Weyermann, C., Koehler, F. & Spengler, B. (2005). Mass spectrometric investigation of the aging of ballpoint inks, *53rd ASMS Conference*, American Society for Mass Spectrometry, San Antonio.
[79] Bojko, K., Roux, C. & Reedy, B.J. An examination of the sequence of intersecting lines using attenuated total reflectance – Fourier transform infrared spectral imaging. *Journal of Forensic Sciences* (corrected proof).
[80] Gal, T., Sandor, J. & Karoly, A. (2007). Determining the sequence of crossed lines by FT-IR-ATR-microscopy, *Global Forensic Science Today* **1**, 8–11.
[81] Horwitz, W. (1982). Evaluation of analytical methods for regulation of foods and drugs, *Analytical Chemistry* **52**, 67A–76A.
[82] Evett, I.W. (1987). Bayesian inference and forensic science: problems and perspectives, *The Statistician* **36**, 99–105.

CÉLINE WEYERMANN

Dating: Radiocarbon *see* Radiocarbon Dating

Daubert v. Merrell Dow Pharmaceuticals

Facts and Issues Summarized

Two infants, one with a surname Daubert, were born with limb-reduction birth defects. Their parents and the infants sued Dow Pharmaceutical Company, the manufacturer of Bendectin, the drug that the plaintiff

mothers took during pregnancy to alleviate morning sickness. The suit claimed that the drug caused the limb reduction. The drug company provided the trial court with 30 studies involving more than 130 000 patients who had used Bendectin. None of the studies had found a statistically significant correlation between use of the drug and birth defects in offspring. The plaintiffs' experts sought to counter with a reanalysis of these studies, in which reanalysis showed a possible link, but the trial court refused to admit this evidence and granted a summary judgment for the pharmaceutical company.

An intermediate appellate court – the Ninth Circuit Court of Appeals – affirmed the trial court's decision to exclude plaintiffs' expert evidence, reasoning that their testimony did not satisfy the *Frye* standard. The *Frye* standard, which had become widely accepted by American state and federal courts, mandated that for novel scientific evidence to be admissible in court, the specialized subject matter had to be generally accepted as reliable in the field to which it belongs (*see* ***Frye v. United States***).[b]

In the *Daubert* case, the relevant science was epidemiology and the appeal's court found that the plaintiffs' expert testimony was not based on published, accepted, and peer-reviewed scientific knowledge in that the plaintiffs' experts' recalculations had indeed been created for the sole purpose of the litigation, and had not been published in peer-reviewed journals (*see* **Federal Rule of Evidence 702**).

On review of this decision, the US Supreme Court had to decide whether the 1923 *Frye* rule of general acceptance had survived the passage of the Federal Rules of Evidence (FRE) (*see* **Federal Rule of Evidence 702**). The Court held that it did not, deciding that proof of reliability of the expert's methodology was to be the sole criterion for its admissibility. The case was remanded for reconsideration by the lower court in light of the Supreme Court's mandate.

Upon remand, the Ninth Circuit Court applied the "reliability" criterion and again granted summary judgment for the defendant, finding that the plaintiffs' expert testimony that had failed to satisfy *Frye* had also failed to meet the *Daubert* standard.[c]

Discussion of the *Daubert* Opinion

The Court's primary holding in *Daubert* was that the 1975-promulgated FRE superseded *Frye*'s requirements of general acceptance in federal civil and criminal trials.

The decision also explained how Rule 702 was to be interpreted. *Daubert* held that the trial court's gatekeeping function requires proof of reliability of the methodology used by experts in arriving at their opinions. *Frye* had cast the function of determining whether an expert's testimony was "demonstrable" on the scientific community by requiring proof of its general acceptance by the field in which it belonged.

Daubert, by contrast, assigned the role of determining reliability of the testimony, as well as its relevance to the issues, to be decided to the trial judge. Although Rule 702 is applicable to a broad range of expert testimony other than that presented by scientists (*see* **Federal Rule of Evidence 702**), the *Daubert* decision dealt only with scientific expert testimony because that was the nature of the case at hand[d] (*see* ***Kumho Tire v. Carmichael***).

The Court stated that the trial judge must determine at the outset whether a qualified expert is proposing to testify to (i) scientific knowledge that (ii) will assist the trier of fact to understand or determine a fact in issue.

In determining reliability, by which the Court must have meant "reproducibility", *Daubert* cautioned that the trial courts must evaluate whether the expert's methodology was likely to lead to replicable (reproducible) results. The inquiry was to focus on the methodology used, and not on whether the conclusion reached by the expert in the particular case was accurate.

To guide the courts in this effort, the Supreme Court suggested several factors for courts to use in assessing the reliability of proffered scientific evidence, though the Court was at pains to stress that these factors were not to be considered litmus tests. Indeed, trial judges were recognized to have a wide range of discretion in the admissibility decision and were to use a "flexible" approach for determining whether a results of a particular technique were reproducible and trustworthy.[e]

The Daubert Factors Explained

The four suggested *Daubert* factors are as follows:

1. Whether the relevant science can be and has been tested by a scientific method

This was said to mean that the theory on which the science relies has been empirically tested. A

hypothesis is conjectured upon scientific principles, investigated with observed data, and determined to be false or true. The Court cited Karl Popper's criterion for the scientific status of a theory or methodology to be its *falsifiability, refutability*, or *testability* (*see* **Falsifiability Theory**) [4].

2. Whether the underlying theory, technique, or methodology has been subjected to peer review or has been published in the professional literature

The term *peer review literature* generally refers to professional journals or periodicals, wherein inclusion of an article has preliminarily been screened by other scientists who deem it worthy to be submitted for comment and scrutiny to the relevant scientific community. Published works, which the Court considered to be one aspect of "peer review", receive even more analysis and critical comment, but the Court recognized that some propositions may be too new or too limited in scope to have been widely published.

3. Whether the results of the chosen methodology have been subjected to an error-rate determination

To what extent the theory or conjecture is likely to be wrong is a relevant issue in determining its reliability or reproducibility and, hence, its trustworthiness. Error rates, when they exist, are meaningful in that they help confirm the degree of confidence with which scientific conclusions reached in a particular area are regarded by professionals.

Finally, in an almost complete reversal of its rejection of the *Frye* rule as a standard for admissibility, the Court added that

4. Whether a theory had received general acceptance in the relevant scientific community may also be a meaningful consideration

Using the *Frye* terminology at (4), the Court admitted that general acceptance could remain as a factor in determining the reliability of a technique, but its significance was no longer primary or determinative of the outcome. The Court found that *Frye* aspired to an austere standard that was incompatible with the more lenient approach to admissibility of the FRE. Use of these four criteria would allow newer and likely more sophisticated scientific methodology to be admitted as testimony if other criteria safeguarding its reliability were met.

Other Considerations Expressed by the Court

FRE 402 further guided the Supreme Court's reasoning. Rule 402 mandates that all evidence be relevant to the case and defines "relevant" evidence in Rule 401 as that which has "any tendency to make the existence of any fact that is on consequence to the determination of the action more probable or less probable than it would be without the evidence."

The Court also dealt with two other concerns noted by the litigants in *Daubert*. First, that abandoning *Frye* would lead to a "free-for-all" causing potential juror confusion by "absurd and irrational pseudoscientific assertions". The adversary system properly conducted, the Court countered, would sort shaky from firm evidence in the crucible of vigorous cross-examination and the presentation of contrary evidence by the opposing parties.[f]

Another concern that had been expressed suggested that the judge's screening role may preclude an adequate consideration of insights and innovations that may well eventually prove to be valid.[f] The Court stressed that there are "important differences between the quest for truth in the courtroom and the quest for truth in the laboratory". The need to arrive at a final and binding result in legal controversies requires that such decisions be made on admissibility of evidence. The Court suggested that judges are competent to apply its rules and strike a balance between these competing interests.

Impact of the Decision

Daubert was sweeping in its influence, if not always a champion of clarity, logic, or realistic in the assumption that the trial court judges would be able to display sufficient scientific acumen in determining the reliability of sophisticated methodologies. Indeed, it is said that to cast judges in the role of amateur scientists was ill conceived. Although it is true that the misuse of scientific evidence has always been seen as a fundamental assault on justice, the Court obviously believed that its *Daubert* discussion and standard would provide adequate safeguards against the use of unreliable or untested scientific evidence in court.

Because the decision was directed at federal courts, the individual states were free to adopt or

reject it at their discretion. Many states have since decided to follow the federal model, but a sizable number of American states' judicial systems still use the *Frye* standard in its original form or as modified locally. It is also recognized, however, that even in states that have not adopted the *Daubert* formulation of admissibility of expert opinion evidence, court decisions since 1993 have, nevertheless, been greatly influenced by the Supreme Court's *Daubert* opinion and holding.

A curious consequence of *Daubert* has been that while the decision clearly stated the *Frye* test to be too rigid a requirement for admissibility, trial courts interpreting it have found the *Daubert* mandate to be, instead, more demanding in its need for proof of reliability than the *Frye* test of general acceptance in a professional field had been. In addition, surveys of judges have shown that most of them continue to attach the greatest weight to the general acceptance of a technique [5].

Several studies have been conducted to measure the effect of *Daubert* on trial court decisions subsequently. Most show that the courts scrutinize more closely the reliability of expert opinion evidence today than they did before the *Daubert* decision date [13].

Daubert standards have also been useful for analyzing evaluations in disciplines somewhat removed from *Daubert's* model of scientific methodology. For example, the Millon Clinical Multiaxial Inventory (MCMI) and its later versions have been used by mental health professionals to assist with psychiatric screening and clinical diagnoses. In an article by Richard Rogers, Randall Salekin and Kenneth Sewell, the MCMI-III was found to have fundamental problems with its scientific validity and error rates precluding its admissibility under *Daubert*. Reconstructing validity through a "meta-analysis" did render portions of MCMI-II acceptable in *Daubert* measures [6].

The *Daubert* decision and its US Supreme Court progeny (*see* ***General Electric v. Joiner***; ***Kumho Tire v. Carmichael***) [4, 5] continue to have a major impact on the development of proof by forensic scientists presenting opinion testimony in courts, even in courts beyond the United States.[a] Where it is applied as a criterion for admissibility, the public as well as judicial acceptance of forensic evidence as valid and credible continues to grow.

End Notes

[a.] The decision is also followed as the law of a number of American states. But even beyond USA borders, it is quoted by courts in other adversary system jurisdictions as well as in civil law countries. E.g., making reference to *Frye* and *Daubert:* In R. V. Dallagher, [2003] 1 Cr. App. R. 12; 2002 WL 1446192 (CA. Crim. Div. 2002), the U.K. Court of Appeal (Criminal Division) cited the United States *Frye* decision as being in accord with the English approach when it quashed a conviction obtained on a crime scene ear mark identified as having been made by defendant as "unsafe". The Court also cited *Daubert* as being a more lenient standard of admissibility than *Frye*. It should also be noted that civil law countries quote from reports by local researchers on reliability of forensic techniques in which the researchers' opinions are based on *Frye* and/or *Daubert* criteria, which the courts then adopt in deciding specific cases. See, e.g., http://www.forensic-evidence.com/site/ID/IDearNews.html mentioning reports of the Netherlands Forensic Institute at the Netherlands Ministry of Justice and by university professors from Holland and Belgium, which in turn were relied on by the Netherlands' "Gerechtshof te Amsterdam" in the Netherlands (Dutch court of appeal at Amsterdam) in reversing a conviction based on methods that find "insufficient support in accepted evidentiary principles"

[b.] See *Frye v. United States* and General Acceptance Test for Novel Expert Testimony. See also, *Daubert v. Merrell Dow Pharmaceuticals* about the FRE incompatibility with *Frye*: "Nothing in the text of this Rule (702) establishes "general acceptance" as an absolute prerequisite to admissibility. Nor does respondent present any clear indication that Rule 702 or the Rules as a whole were intended to incorporate a 'general acceptance' standard. The drafting history makes no mention of Frye, and a rigid general acceptance requirement would be at odds with the 'liberal thrust' of the Federal Rules and their general approach of relaxing the traditional barriers to opinion testimony". (1923).

[c.] See *Daubert v. Merrell Dow Pharmaceuticals*, 43 F.3d 1311 (Ninth Circuit, 1995), on remand from the Supreme Court. This subsequent decision is often referred to as *Daubert II*. The author of the *Daubert II* opinion, Judge Alex Kozinski, wrote about *Daubert*'s "Brave New World" in which judges who must apply the *Daubert* mandates will henceforth live.

[d] Whether *Daubert* also applied to nonscientifically based expert opinions and skill-based testimony was left an open question. The Court addressed this issue in 1999 in the *Kumho Tire* case. See, *Kumho Tire v Carmichael*.

[e] The Court stated: "... [I]t would be unreasonable to conclude that the subject of scientific testimony must be 'known' to a certainty; arguably, there are no certainties in science... ." Daubert, 509 U.S. at 590, 113 S.Ct. at 1295 (1993).

[f] See, *Daubert* opinion, 509 U.S. at 596, 113 S.Ct at 2798 (1993).

References

[1] Cheng, E.K. & Yoon, A.H. (2005). See, A study of scientific admissibility standards, *Virginia Law Review* **91**, 471–513.

[2] Criticizing the Supreme Court's definition of scientific knowledge and the scientific method, see, Schwartz, A. (1997). A 'Dogma of Empiricism' revisited: *Daubert v Merrell Dow Pharmaceuticals, Inc* and the need to resurrect the philosophical insight of *Frye v United States*, *The Harvard Journal of Law and Technology* **10**(2), 149.

[3] Kozinski, A. (1997). Brave new world, *The University of California, Davis Law Review* (30), 997–1011.

[4] The Court referred to: Popper, K.R. (1968). *Conjectures and Refutations: The Growth of Scientific Knowledge*, 1962, Basic Books.

[5] Cecil, J.S. (2005). Ten years of judicial gatekeeping under *Daubert*, *American Journal of Public Health* **51**(95), (Suppl 1) 574–580. The survey conducted by the Federal Judiciary Center of federal judges and attorneys found that judges were more likely to scrutinize expert testimony before trial and to limit or exclude proffered testimony than they did before *Daubert*.

[6] See Rogers, R., Salekin, R.T. & Sewell, K. (1999). Validation of the millon clinical multiaxial inventory for axis II disorders: Does it meet the Daubert standard? *Law and Human Behavior* **23**, 425–443.

Related Articles

Expert Opinion: United States

General Electric v. Joiner

Kumho Tire v. Carmichael

Weisgram v. Marley

ANN C. SMITH

Deadly Force ***see*** Police Use of Force, Policing and Critical Incident Teams

Death: Anoxia ***see*** Asphyxia

Death: Asphyxia ***see*** Asphyxia

Death: Blunt Force Trauma ***see*** Blunt Force Trauma

Death: Cardiac ***see*** Cardiac and Natural Causes of Sudden Death

Death: Multiple ***see*** Homicide: Multiple (Behavior)

Death: Natural ***see*** Natural Causes of Sudden Death: Noncardiac

Death: Natural Causes ***see*** Cardiac and Natural Causes of Sudden Death

Death: Time of

Introduction

Death and dying are processes that are characterized by loss of the function of the great organ systems (cardiovascular system, respiratory system, and nervous system) and their coordination. The loss of the coordination of the great organ systems reveals a dissociation of the function of the different organs [1]. The agonal period may be initiated by either disease or trauma; the final crisis leads over a vita minima in which no vital signs may be apparent to the death of the individual characterized by irreversible cessation of circulation or respiratory arrest (Figure 1). Under special clinical conditions, brain death may replace the classical signs of death (irreversible circulatory or respiratory arrest and their consequences postmortem lividity and rigor mortis). The duration of the agonal period may differ largely:

- Ultrarapid agonal periods may be found in explosions with a total dismemberment of the body in milliseconds.
- Short agonal periods of some minutes' duration are found in many violent deaths and acute natural deaths.
- Long terminal episodes lasting hours to days are seen as terminal episodes of longlasting diseases, e.g., cancer.

In the final crisis of an agonal period, vital signs may not be apparent and persons still alive may falsely be pronounced dead (Figure 2). In most cases where death was certified for persons still alive an intoxication with central depressants in combination with hypothermia was the reason for the lifeless condition. Therefore, especially in cases of hypothermia, the old proverb of emergency physicians that no one is dead until he is warm and dead should be kept in mind [2].

Supravitality

The irreversible circulatory – or respiratory – arrest is, in most cases, the main criterion for death [1]. Metabolism of tissue, however, does not cease immediately after death but runs on for some hours. The main energy-producing metabolic processes are in the early postmortem period: the creatine-kinase

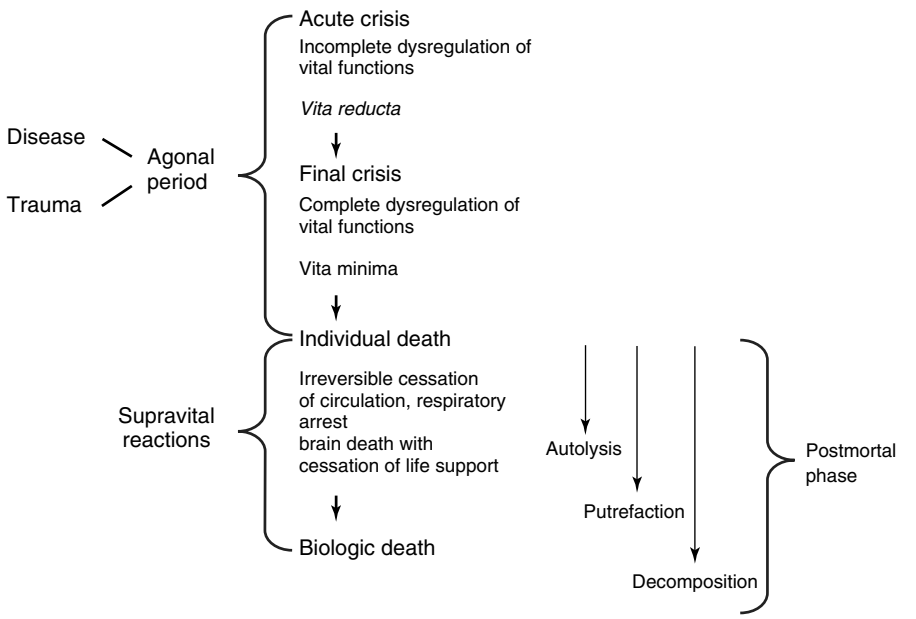


Figure 1 Scheme and terms of the agonal period. The agonal period is initiated by either disease or trauma



Figure 2 An old lady lying dressed on the bank of a river. Vital signs were missing. Although postmortem changes (lividity, rigor mortis) were missing, she was pronounced dead by an emergency physician. Probably cold stiffening was mistaken for rigor mortis. Cause of death was intoxication with central depressants

reaction and then anaerobic glycolysis [3–6]. During this period of intermediary life, supravital reactions can be examined, which are defined as reactions of tissues on excitation postmortem [6]. The duration of the supravital period is much longer than that of the resuscitation period, known from physiology and experimental surgery (Figure 3). Resuscitation period is the time of global ischemia after which the ability to recover expires; the supravital period covers also the period after, which is characterized by an increasing irreversible damage of structure and function. The resuscitation period of skeletal muscle under normothermia lasts, e.g., 2–3 h; supravital electrical excitability of skeletal muscle may be preserved up to 20-h postmortem. Some supravital reactions are of great practical importance in forensic medicine since they can easily be examined at the scene of crime and give an immediate result on the time elapsed since death: these are the mechanical and electrical excitability of skeletal muscle and the pharmacological excitability of the iris.

Mechanical Excitability of the Muscle

The mechanical excitability of skeletal muscle is examined by rigorously hitting a muscle with the back of a knife or a chisel, e.g., the biceps brachii muscle at right angles to the arm axis. Of course other muscles can be examined as well, but reference values for estimating the time since death are available only for the biceps brachii muscle [5, 7]. There are three reaction patterns of the muscle depending on the postmortem interval (Figure 4):

- In the first phase, mechanical excitation of the muscle reveals a contraction of the whole muscle (propagated excitation). This first phase of idiomuscular contraction is identical to Zsako's muscle phenomenon. The first degree of idiomuscular contraction, a propagated excitation, can be observed up to 1.5–2.5 hpm.
- In the second phase, a strong and typically reversible idiomuscular pad develops. This phase may be seen as long as 4–5 hpm (Figure 5).
- In the last phase, a weak idiomuscular pad develops, which may persist over a longer period (up to 24 h). This weak idiomuscular pad can be seen in the time interval up to 8–12 hpm. If the idiomuscular pad is not visible it should be palpated for; otherwise, skin incisions may be necessary to demonstrate the presence of an idiomuscular pad.

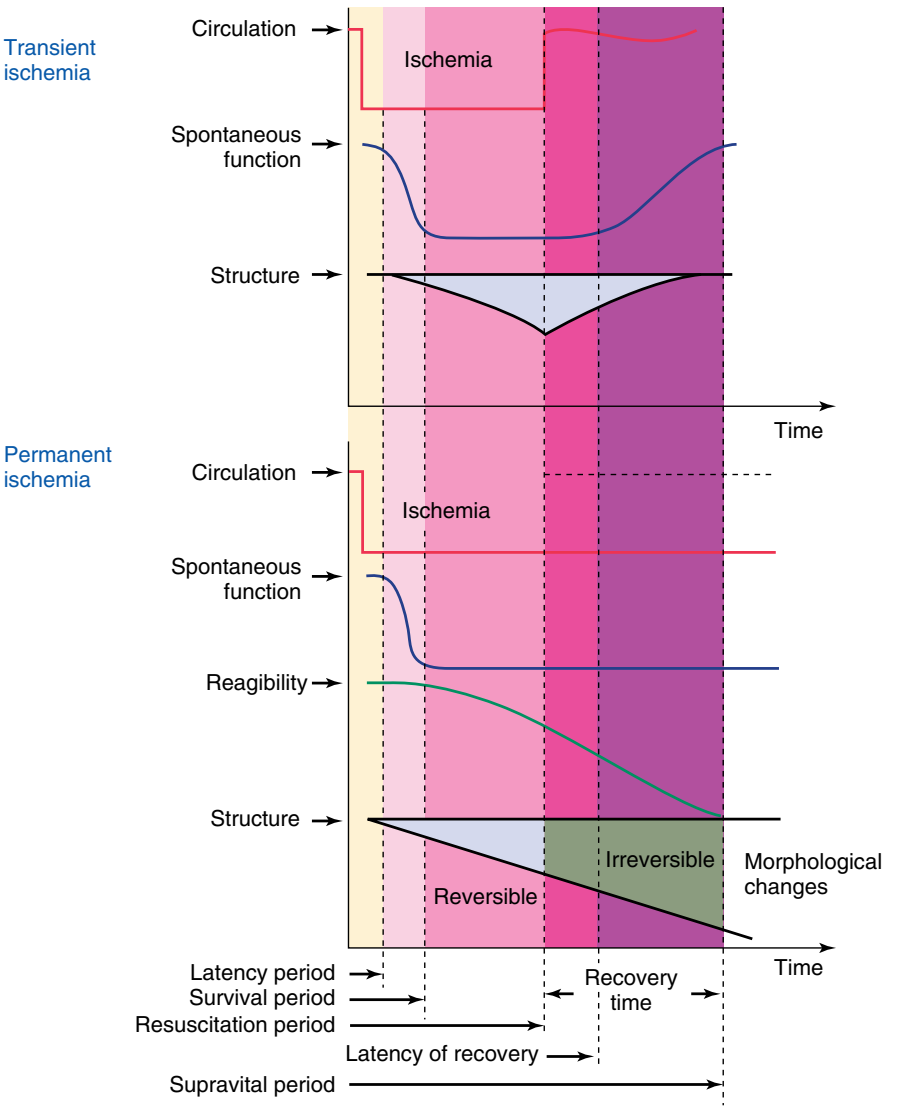


Figure 3 Diagram of the duration of the supravital period (bottom) compared to the resuscitation period (top). The supravital period after irreversible circulatory arrest exceeds the duration of the resuscitation period after transient ischemia by far

Electrical Excitability of Skeletal Muscle. The earliest investigations on electrical excitability of skeletal muscle were carried out at the end of the eighteenth century and beginning of the nineteenth century and electrical excitability was recommended as a method for estimating the time since death [8]. Subsequently, extensive investigations have been carried out using techniques to objectify muscular contraction [9–12]. However, for practical purposes, most investigations on postmortem electrical excitability of skeletal muscle are based on a verbal description and subjective grading of the muscular response to excitation – the muscular contraction – according to

- the strength of contraction and
- the spread of movement to areas distant from the electrodes.

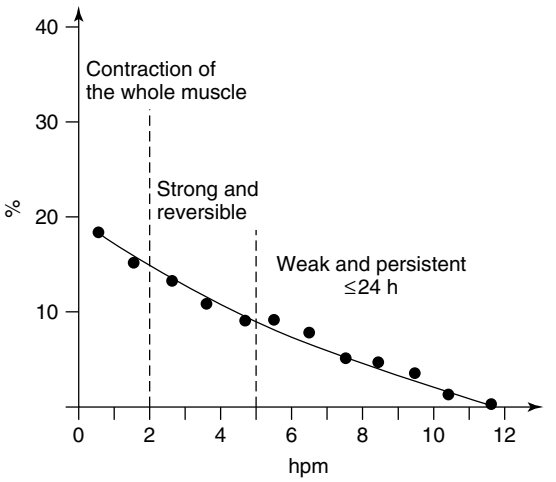


Figure 4 Three phases of idiomuscular contraction after mechanical excitation of the muscle; frequency of a positive reaction (y axis) over the postmortem interval (x axis)

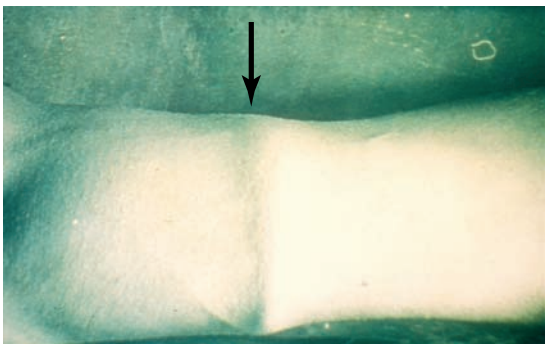


Figure 5 Typical idiomuscular pad of the biceps brachii muscle

In the early postmortem interval, excitation leads to a strong contraction of muscles and the excitation spreads to muscles distant from the electrodes, while with increasing postmortem interval, the contraction becomes weaker and the muscular response is confined to the place of excitation. This reaction pattern can be seen more or less in all muscles. Lastly, only a fascicular twitching or movement of the electrodes is visible. The most extensive investigations have been carried out for the orbicularis oculi muscle since movements of this muscle are easily visible [5, 13–16]. For the orbicularis oculi muscle puncture, electrodes are inserted in a distance

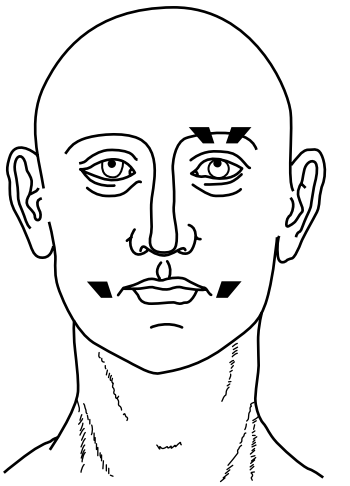


Figure 6 Position of electrodes for examining electrical excitability of mimic muscles

of 15–20 mm in the nasal part of the upper eyelid 5–7-mm deep (Figure 6). The muscular response is graded into six stages. In the very early postmortem interval (degree VI), the whole ipsilateral muscle contracts, in degree V only the upper and lower eyelids and forehead contract, and with increasing postmortem interval the reaction is confined to the place of excitation (Figure 7, Table 1).

For stimulation, a small square wave generator producing constant current rectangular impulses of 300 mA, 10 ms duration in a repetition rate of 50 ms is used (Figure 8). The grading mean values and 95% limits of confidence for the stages are presented in Figure 7.

The 95% limits of confidence for the six stages of electrical excitability of the orbicularis oculi muscle have been proved to be reliable on independent case material [15] and in field studies [17–19]. However, in cases with hematoma or emphysema of the eyelid, electrical excitability may last much longer than corresponding to the upper 95% limits of confidence for the special stage (Table 1, last column). The same is true for cases of hypothermia [15]. On the other hand, in cases of longlasting diseases with a long terminal episode, the duration of electrical excitability is shorter than in cases of sudden death due to the fact that glycogen, which is responsible for the resynthesis of ATP, may be already depleted during life.

Other muscles can be examined as well and reliable reference data are available, e.g., for the

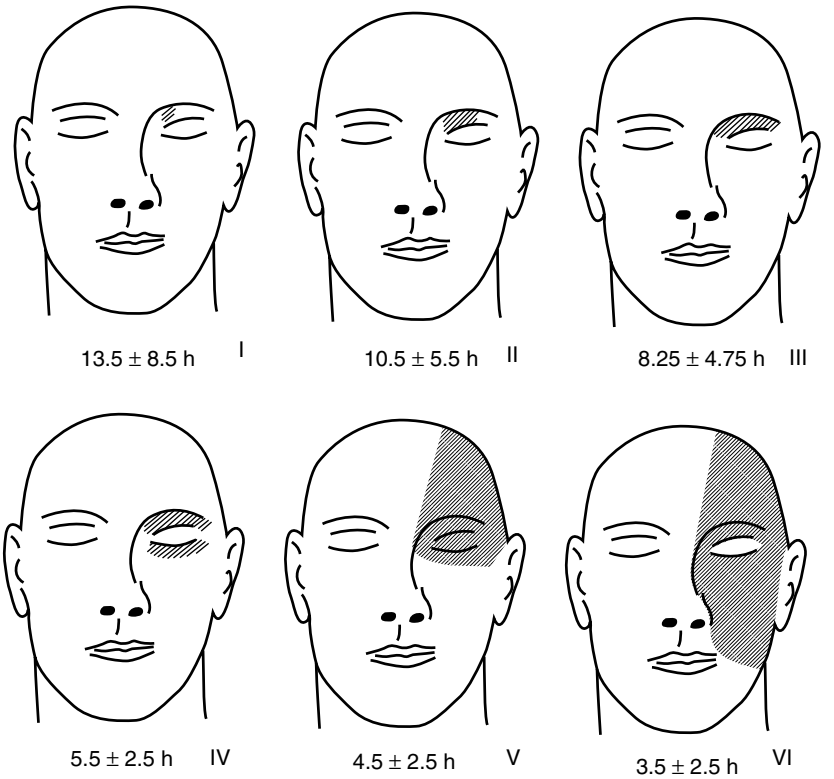


Figure 7 Six stages of positive reaction after stimulation of the orbicularis oculi muscle (see also Table 1)

Table 1 Upper and lower confidence limits for the six levels of electrical excitability in hours in different random samples: forensic pathology with a short terminal episode; clinical pathology with a longer terminal episode; and cases with emphysema and hematoma of the eyelid. In cases of clinical pathology, the duration of electrical excitability is shorter than in cases of forensic pathology

Stage		Forensic pathology	Clinical pathology	Hematoma/emphysema of the eyelid
I	Local upper eyelid	5–22	3–16	–29 traumatic emphysema 27.3–52 postmortem emphysema –32 traumatic hematoma
II	1/3–2/3 upper eyelid	5–16	0–16	
III	Whole upper eyelid	3.5–13	1.5–9	
IV	Upper and lower eyelid	3–8	1–7	
V	Upper and lower eyelid + forehead	2–7	1–7	
VI	Upper and lower eyelid + forehead + cheek	1–6	1–6	

orbicularis oris muscle (Figure 9). The muscles of the thenar or hypothenar may react on stimulation up to about 11-h postmortem [9–11].

Pharmacological Excitability of the Iris. Compared to skeletal muscle, the smooth iris muscle is excitable by electrical and pharmacological

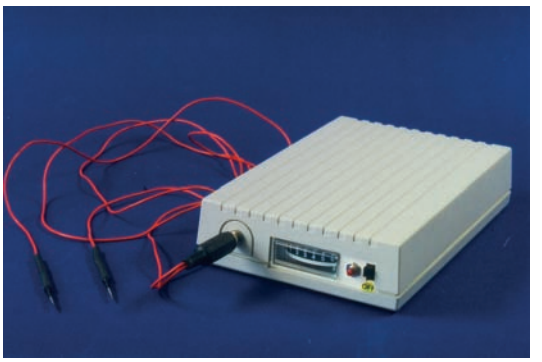


Figure 8 Square wave generator for defined stimulation of, e.g., facial muscles enabling the procurement of the data in Table 1 and Figure 7. Producer and supplier: Peschke J, http://home.t-online.de/home/j-peschke/rztg1.htm

stimulation for a much longer period [13, 14, 20]. In some cases, excitability on subconjunctival injection of noradrenalin or acetylcholine may be preserved up to 46 hpm (Table 2). For practical purposes, the pharmacological excitability of the iris after subconjunctival injection of drugs can be recommended, whereas the drugs should not be injected into the anterior chamber. Of course, the starting diameter of the pupil should be stated before injection using a transparent, multidiameter pattern. About 0.5 ml solution of noradrenalin, tropicamide, atropine, or acetylcholine should be injected. A positive reaction can be seen within 5–30 min, with the diameter becoming greater (atropine, tropicamide, and noradrenaline) or smaller (acetylcholine). The duration of reaction lasts at least 1 h. When no change is visible after this time, the reaction is negative (Table 2). It is sufficient to examine the pharmacological excitability of the iris just with one drug since examining a double reaction after injection of a mydriatricum first, then a mioticum, gives no further information on the time since death.

Table 2 Postmortem excitability of the iris after injection of different drugs

Drug		Postmortem excitability (h)
Mydriatics		
Norepinephrine/epinephrine	1%	14–46
Tropicamide	0.25%	5–30
Atropine/cyclopentolate	1%/0.5%	3–10
Miotics		
Acetylcholine	5%	14–46

Further supravital reactions are the reagibility of sweat glands on the injection of drugs, sperm motility, vitality of leukocytes, DNA incorporation, and postmortem blood clotting. However, these

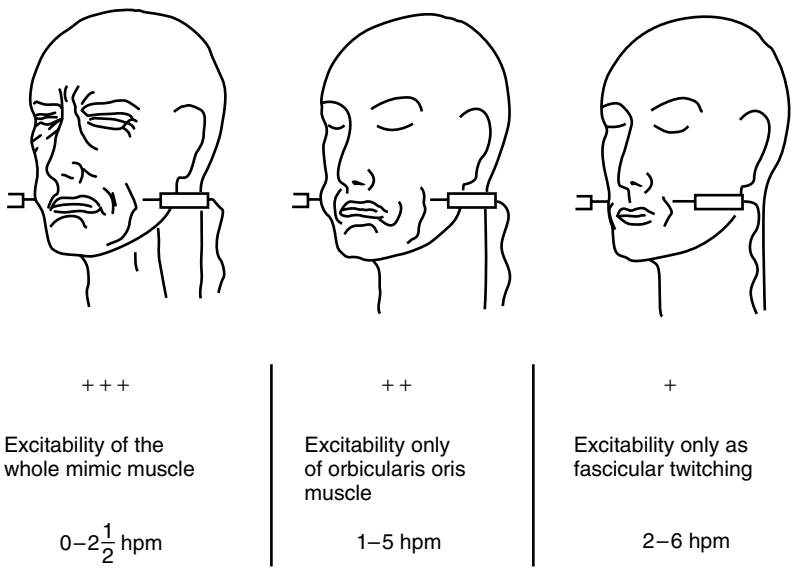


Figure 9 Three stages of positive reaction after stimulation of the orbicularis oris muscle

Table 3 Lividity: causes, consequences, and phenomena checked on the body

Causes		Consequences		Phenomenon
Decrease of force of myocardial contraction	⟶	Stasis	⟶	"Kirchhofrosen", local stasis with patchy discoloration during agonal period due to centralization of circulation
Cardiac arrest, hydrostatic pressure	⟶	Hypostasis	⟶	Livores with a "shifting" quality and "disappearance on pressure"
Vascular permeability	⟶	Hemoconcentration	⟶ ↘	Decrease of shifting and disappearance on pressure
Autolysis, putrefaction	⟶	Diffusion of hemoglobin	⟶	No shifting, no disappearance on pressure

supravital reactions have not gained any practical relevance in forensic medicine.

Postmortem Lividity

After irreversible circulatory arrest, postmortem lividity develops as the earliest postmortem change [1, 16, 21]. After the circulatory arrest, the hydrostatic pressure becomes the leading force within the parallelogram of forces consisting of blood pressure, structural barriers, tissue turgor, and pressure of underlying surfaces [13]. Hypostasis means the movement of body fluids under the influence of gravity. All fluid compartments are involved in hypostasis; that is, not only the intravascular but the transcellular fluids as well. Influenced by gravity, blood is moved into the lowest parts within the vascular system of the body; in a supine position the blood moves into the back, the buttocks, thighs, calves, and back of the neck. Irregular pink patches on the face, especially the cheeks, in the agonal period are caused by local stasis and are called *Kirchhofrosen* (Table 3). Postmortem lividity visible in the skin is a consequence of the movement of blood into the capillaries of the corium. Postmortem lividity may be visible after 20–30 min postmortem, in the early stages as still pink patches, which gradually coalesce with increasing postmortem interval. Owing to the consumption of oxygen, the pink color changes to dark pink or blue (Figure 10).

In the area of deep hypostasis, cutaneous petechial hemorrhages due to capillary rupture may develop and are called *vibices* (Figure 10b). Of diagnostic and criminalistic relevance is not only the development of lividity but also the color, the distribution on the body, the phenomena of fixation (disappearance after turning the body), and disappearance (blanching) on thumb pressure.

In carbon monoxide intoxication and cyanide intoxication, the color of hypostasis is cherry pink, and in methemoglobin intoxication it is brownish (Table 4, Figure 10d). Owing to the lack of dissociation of oxygen from hemoglobin, a bright pink color may be seen in hypothermia as well. In a body brought from cold environment into normal room temperature, a typical zonal segmentation of hypostasis may be seen with a dark blue color in the rewarmed areas (Figure 10c).

Of predominant criminalistic significance are the phenomena of *disappearance on pressure and disappearance after turning the body*. In the early stages, lividity will completely disappear on soft thumb pressure, with increasing postmortem interval the pressure will increase as well, later the lividity will disappear only incompletely on pressure, and lastly it will not disappear at all.

If the body is turned in the early postmortem interval, some or all of the hypostasis may move down to new areas. In a comparatively later postmortem interval, only some of the hypostasis will slip down to the new area and only a slight blanching will be noted in the original area (Figure 11, Table 5).

With increasing postmortem interval, the disappearance of lividity on thumb pressure and relocation after shifting decreases, and then completely ceases. This is caused by an increasing hemoconcentration of intravascular erythrocytes due to transcapillary plasma extravasation (i.e., the fluid moves from the blood leaving the red cells behind, which cannot move without the liquid). The intravascular hemoconcentration is the main reason for the gradually decreasing disappearance on thumb pressure and after

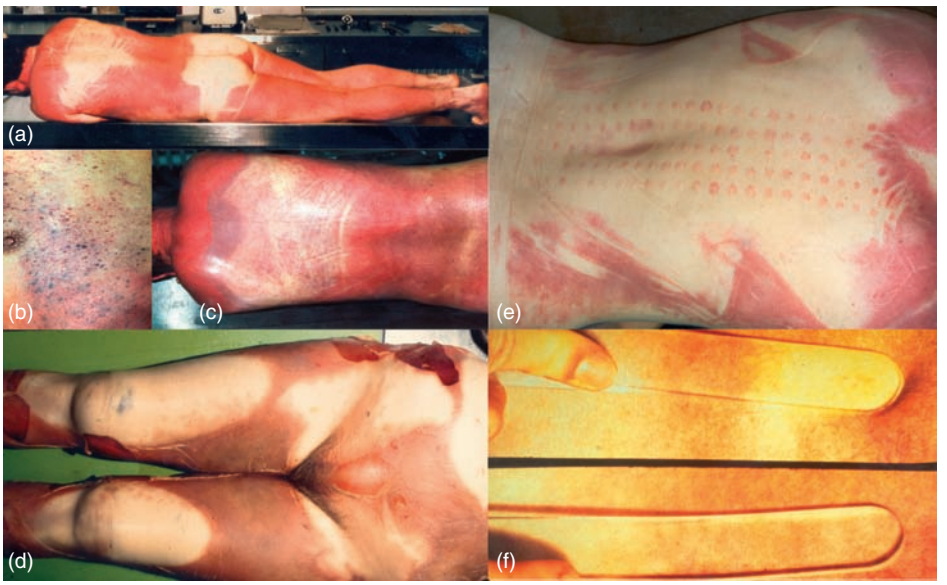


Figure 10 Postmortem lividity. (a) Postmortem lividity in a supine position, the pink color being due to storage in a refrigerator. (b) Petechial-like hemorrhages due to capillary rupture in areas of deep hypostasis (vibices). (c) Zonal segmentation of hypostasis: partly dark blue, partly red. (d) Brownish color of hypostasis in methemoglobin intoxication. (e) Patterned hypostasis of the back due to the underground. (f) Complete disappearance of lividity on light pressure

Table 4 Postmortem lividity discoloration[(a)]

Etiology	Color	Mechanism
Normal	Blue–purplish	Venous blood
Carbon monoxide	Pink, cherry red	Carboxy hemoglobin
Cyanide	Pink, cherry red	Excessively oxygenated blood due to inhibition of cytochrome oxidase
Fluoroacetate	Pink, cherry red	Same as above
Refrigeration/ hypothermia	Pink, cherry red	Oxygen retention in cutaneous blood by cold air Left shifting of Hb O_2 dissociation curve
Sodium chloride/nitrite, nitrate	Brown	Methemoglobin
Hydrogen sulfide	Green	Sulfhemoglobin

[(a)] Modified according to Spitz and Fischer [22]

shifting. In a comparatively later postmortem interval, hemolysis and hemoglobin diffusion into the perivasal tissue starts. However, this is only a secondary mechanism contributing to the fixation of hypostasis [6, 13].

Phases of hypostasis are, therefore, beginning, confluence, maximum, disappearance on thumb pressure, complete or incomplete disappearance after shifting, and are changing with time.

Longitudinal studies on these criteria on large random samples are missing. Other studies are of limited value. The best statistical data available were summarized by Mallach [16, 25], who calculated mean values, standard deviations, and 95% limits of confidence based on textbook reports (Table 6). Since better data are missing, these data are still undisputed. However, it should be kept in mind that these data do not represent absolute limit values. Investigations with a quantitative measurement of livor mortis [26] have not yet gained practical importance.

Beginning is evidenced if mottled patches at lower parts of the body (for instance, neck in supine position) are observed.

Confluence is identified if separate areas of discoloration of moderate intensity are observed.

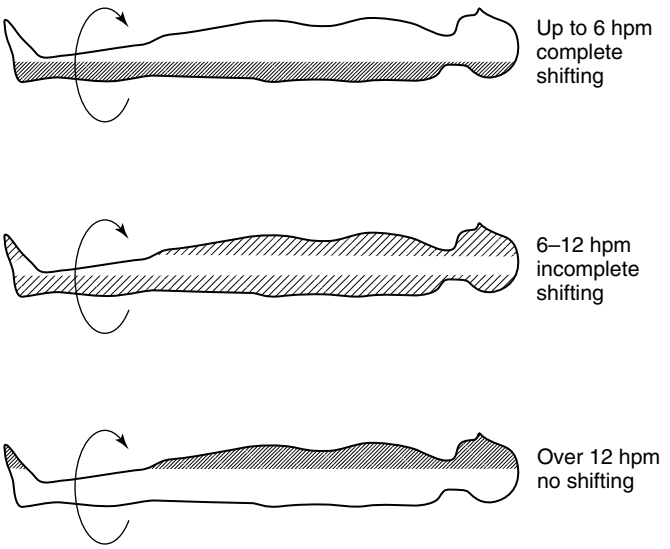


Figure 11 Shifting of lividity after turning the body (modified according to [23]

Table 5 Time course of different criteria of lividity according to W. Naeve [24]

Lividity	Time postmortem
Beginning	15–20 min
Confluence	0.5–2 h
Maximum confluence	4–10 h
Complete disappearance on pressure	10–20 h
Incomplete disappearance	10–30 h
Complete shifting	2–6 h
Incomplete shifting	4–24 h
Only slight pallor after shifting	20–30 h

Maximum can be identified if, during crime scene investigation and autopsy, hypostasis did not increase.

The criterion *thumb pressure* is positive if lividity disappears completely on light thumb pressure.

Complete relocation is when all hypostasis shifts down to the new dependent areas. This may be observed at the scene when a body is turned from a facedown to a supine position.

In *incomplete shifting*, hypostasis not only remains in the former dependent areas but shifts as well, to a more or lesser extent, to the new dependent areas.

Rigor Mortis

The second postmortem change and sign of death, developing in normal ambient temperature about 3–4 hpm (hours postmortem) after primary flaccidity, is rigor mortis. Rigor mortis was misjudged as a sign of death up to the nineteenth century, although Shakespeare (Romeo and Juliet, Act. IV, Scene 1) described all elements of rigor mortis very well:

> Each part, deprived of supple government, shall, stiff and stark and cold, appear like death.

With irreversible circulatory arrest, all muscles of the body will become completely flaccid due to the loss of tone. In the early postmortem interval, adenosine triphosphate (ATP, the chemical source of energy) can be resynthesized due to the creatin-kinase reaction and anaerobic glycolysis. Once the ATP level has fallen under 85% of the initial value, actin and myosin filaments will contract and the subjective impression of stiffening of the muscle can be noticed (Table 7) [1, 3, 13, 27, 28]. In practice, development and state of rigor mortis are proved subjectively by flexing a joint: either the muscles are flaccid or during development of rigor mortis resistance may be felt (Figure 12). If rigor develops completely, even a strong investigator cannot flex or stretch a joint. Biochemical, physiological, and

Table 6 Time course of different criteria of lividity[a]

			2 *s*		Range of scatter		
	$\bar{x}$	*s*	Lower limit	Upper limit	Lower limit	Upper limit	Number of quotations
Development	3/4	1/2	–	2	1/4	3	17
Confluence	2 1/2	1	3/4	4 1/4	1	4	5
Greatest distension and intensity	9 1/2	4 1/2	1/2	18 1/4	3	16	7
Displacement							
1. Complete on thumb pressure	5 1/2	6	–	17 1/2	1	20	5
2. Incomplete on sharp pressure (forceps)	17	10 1/2	–	37 1/2	10	36	4
Displacement after turning the body							
1. Complete	3 3/4	1	2	5 1/2	2	6	11
2. Incomplete	11	4 1/2	2 1/4	20	4	24	11
3. Only little pallor	18 1/2	8	2 1/2	34 1/2	10	30	7

[a] Statistical calculations by Mallach on textbook reports. The statistical calculations are not based on cross-sectional or longitudinal studies but on empiric knowledge quoted in textbooks. $\bar{x}$, mean value and s, standard deviation

Table 7 Overview of biochemical, mechanical, morphological, and physiological basis of rigor mortis

	Delay period	Establishment of rigor	Rigor phase	Postrigor phase, secondary flaccidity
Biochemistry	ATP level: $0.435 \pm 0.555\,mg\,g^{-1}$ muscle	ATP level decreased to less than 85% of the original value	–	NH_3 ↑
Mechanics	–	Stiffness ↑, plasticity first increased, then decreased. Contraction of loaded muscle	–	Spontaneous elongation of loaded muscle; plasticity ↑
Morphology	–	Appearance of fine transverse striations (bridging between A and I filaments) with a periodicity of 400 Å	Swelling and destruction of mitochondria and the sarcoplasmic reticulum	Irreversible elongation of muscle; decoupling of myofilaments, disintegration of structure
Physiology	Exponential decrease of membrane potential: above −55 mV propagated excitation possible; below this level up to −30 mV only local concentration on excitation	–	–	–



Figure 12 Owing to rigor mortis, the lower leg is fixed against gravity (a). Objects in the hands must not be mixed up with instantaneous rigor mortis or cadaveric spasm (b)

mechanical properties of the rigor (pre- and postrigor) phases are summarized in Table 7.

Rigor must not be mixed up with cold stiffening. When rigor is present, hypostasis must be present as well; in cases of cold stiffening (body core temperature 30–33 °C), hypostasis is absent.

The development and state of rigor should be examined not only in one but also in several joints (mandibular joint, finger, knee, and elbow joint) to get an impression if it is still in progression or has already reached its maximum.

Rigor mortis does not start in all muscles simultaneously. Nysten's rule (1881) that rigor starts in the mandibular joint, muscles of the trunk, then in the lower extremities, and lastly in the upper extremities is true for most cases who die in a supine position [8]. However, in cases with glycogen depletion during agony in the lower extremities, rigor will start here.

Resolution of rigor mortis is due to protein degradation (increase of ammonia – NH_3)

In different fibers of a muscle, rigor does not start simultaneously but successively. This phenomenon can be used for a rough estimation of the time since death as well. If some fibers have already become stiff, this stiffness can be broken by flexing a joint; rigor may now develop in other, not yet stiffened fibers. Depending on the time when stiffness has been broken, rigor may develop again on a higher or lower level unless it was already fully developed (Figure 13). This phenomenon of reestablishment of rigor mortis may be seen up to 6–8 hpm, in very low ambient temperatures up to 12 hpm.

Secondary flaccidity may become apparent in normal ambient temperature after two days. In deep ambient temperatures (winter <10 °C), fully

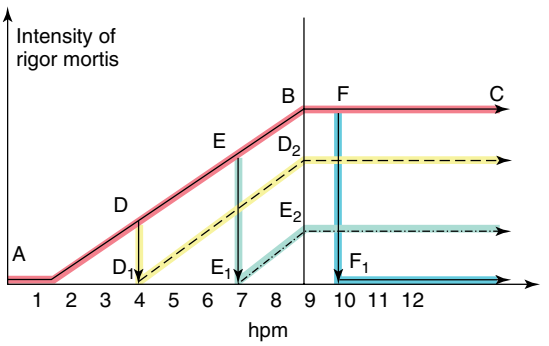


Figure 13 Re-establishment of rigor mortis after breaking. The later during development of rigor mortis rigor is broken, the lower will be the level after re-establishment. If rigor is broken after it has already been fully developed (F) it will not be re-established at all

developed rigor mortis may be preserved two weeks or longer.

Cadaveric rigidity, cataleptic spasm, or instantaneous rigor is a phenomenon always mentioned in textbooks but nonexistent in practice. No case reported in the literature stands up to criticism.

Rigor mortis is established not only in skeletal muscle but also in smooth muscle, e.g., of the skin, as well. Rigidity of the smooth musculi arrectores pilorum can be seen as gooseflesh (cutis anserina) (Figure 14).

Development, duration, and resolution of rigor mortis are dependent on the amount of glycogen in the muscle at the moment of death, ambient temperatures, and so on.

Therefore, rigor may develop very fast in persons who die during or soon after physical exertion



Figure 14 Gooseflesh of the right forearm due to rigidity of the smooth musculi arrectores pilorum

Table 8 Time course of different criteria of rigor mortis according to calculations by Mallach on textbook reports

Rigor state	Average in hpm and standard deviation	Range of scatter in hours (2 s)		Number of quotations
		Lower limit	Upper limit	
Beginning	3 ± 2	–	7	26
Maximum	8 ± 1	6	10	28
Duration	57 ± 14	29	85	27
Complete resolution	76 ± 32	12	140	27

or exhaustion or from electrocution. All above-mentioned criteria of rigor mortis (development, reestablishment, fully developed, duration, and resolution) are time dependent. This becomes already evident from one of the rare studies of the nineteenth century on rigor mortis. Niderkorn, who determined the time necessary for the completion of rigor mortis in 113 bodies, found it fully established after 4–7 h in 76 corpses (67%), in 2 cases rigidity was complete within 2 hpm, and in 2 others it was complete within 13 hpm.

However, the interindividual variability due to endogenous and exogenous factors is great. Longitudinal studies on a great random sample are missing; however, animal experiments, taking into account various factors influencing the time course of rigor mortis, have been published [29–35]. Devices for an objective measurement of rigor mortis have been developed [36] but have not yet gained practical importance.

The best available data in spite of all justified criticism are again from Mallach [16, 25], based on a literature compilation (1811–1969) with statistical analysis (Table 8). These values again cannot claim to be absolute limit values. As lividity, rigor mortis can give only a rough estimation and no accuracy can be expected. It never should be examined alone.

Rigor is examined by flexing or stretching joints. Beginning can be stated to have occurred if a slight rigor can be observed in some joints.

Complete development or maximum means that it has strongly developed in all joints.

Reestablishment can be stated to have occurred if rigor is found again in a joint (mostly elbow) some time (hours) after breaking it. Mostly rigor is broken during crime scene investigation and examined again at autopsy.

Algor Mortis

The fall of body temperature at postmortem is due to four factors: convection, conduction, radiation, and evaporation, the first two being the leading causes.

The rate of cooling depends on various conditions and varies with several factors:

- ambient conditions (temperature, wind, rain, and humidity);
- weight of the body, mass/surface area ratio;
- posture of the body (extended or thighs flexed on the abdomen); and
- clothing/covering.

Different temperature probe sites were used (surface skin temperature, axilla, liver, rectum, and brain temperature); for practical purposes, only central core temperatures (rectum, brain) are of value today [13, 37, 38]. Body cooling does not follow Newton's law of cooling. The early authors in the nineteenth century using rectal temperature described a lag time, the postmortem temperature plateau, before an exponential body cooling according to Newton's law starts (Figure 15) [8, 13].

The temperature plateau is due to the fact that central axial temperatures cannot begin to decrease until a heat gradient is set up between the core of the body and the surface. This delay is variable and may last for some hours.

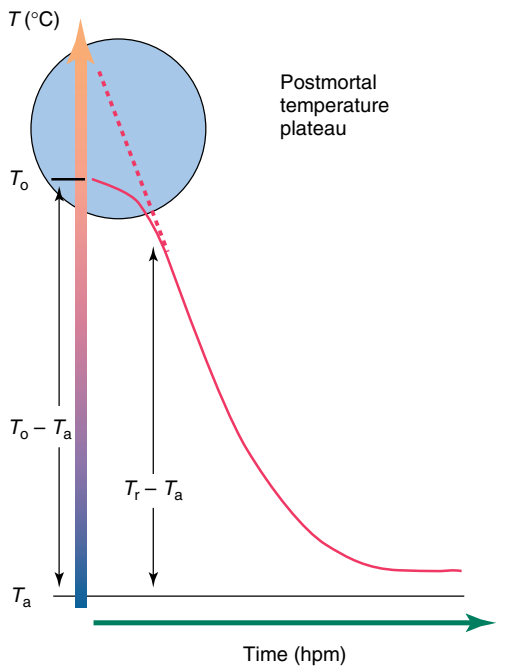


Figure 15 Sigmoidal shape of the cooling curve, which is best described by Marshall and Hoare's two-exponential term. The quotient $\frac{T_r - T_a}{T_o - T_a}$ is a good measure of the progress of cooling. If this quotient Q is < 0.3, only a minimum interval of the time since death should be given. T_o, rectal temperature at death ($T = 0$); T_r, rectal temperature at any time; and T_a, ambient temperature

A mathematical expression of the rectal cooling after death was published by Marshall and Hoare [39], taking into account the whole sigmoidal shape of the cooling curve. Marshall and Hoare performed body cooling experiments under "standard conditions of cooling", which are defined as "naked body with dry surfaces, lying extended on a thermically indifferent base, in still air." Their mathematical model of body cooling was expressed in a two-exponential term:

$$Q = \frac{T_r - T_a}{T_o - T_a} \quad A \times \exp[(B \times t) + (1 - A)] \times \exp\left(\frac{(A \times B)}{A - 1}\right) \times t \qquad (1)$$

where Q, standardized temperature; T_r, rectal temperature at any time t; T_o, rectal temperature at death ($t = 0$); T_a, ambient temperature; A, constant; B, constant; t, and time of death.

The second exponential term is subtracted from Newton's exponential term taking into account that temperature plateau is a from null increasing temperature decrease.

This mathematical expression is valid for all central axial temperatures and is the ultimate success in modeling body cooling for purposes of estimating the time since death [39, 40].

The exponential form with the constant B expresses the exponential drop of temperature after the plateau, according to Newton's law of cooling; the term with the constant A as part of the exponent describes the postmortem temperature plateau.

The experimental work that led to an identification of these constants is outlined in several original papers [41–45] and two monographs [13, 37], which are referred to here.

Empirically, it was found that the individual value of B can be computed under chosen standard conditions of cooling by

$$B = -1.2815(bw^{-0.625}) + 0.0284 \qquad (2)$$

where bw is body weight (kilograms).

This reflects that the body cooling after the temperature plateau is mainly influenced by the body weight.

Values for the constant A could be empirically identified on the material studied under standard conditions and the material of De Saram and Webster [46], reevaluated by Henssge:

A is 1.25 for ambient temperatures up to 23 °C and 1.11 for ambient temperatures above 23 °C.

The temperature at the time of death, T_o, is the third constant of Marshall and Hoare's formula. For this constant, the value of 37.2 °C is used.

With this empirical solution of Marshall and Hoare's formula, the time since death can either be computed according to

$$Q = \frac{T_r - T_a}{37.2 - T_a} = 1.25 \times \exp(B \times t) - 0.25 \times \exp(5 \times B \times t) \qquad (3)$$

(for ambient temperatures up to 23 °C) or

$$Q = \frac{T_r - T_a}{37.2 - T_a} = 1.11 \times \exp(B \times t) - 0.11 \times \exp(10 \times B \times t) \qquad (4)$$

(for ambient temperatures over 23 °C)

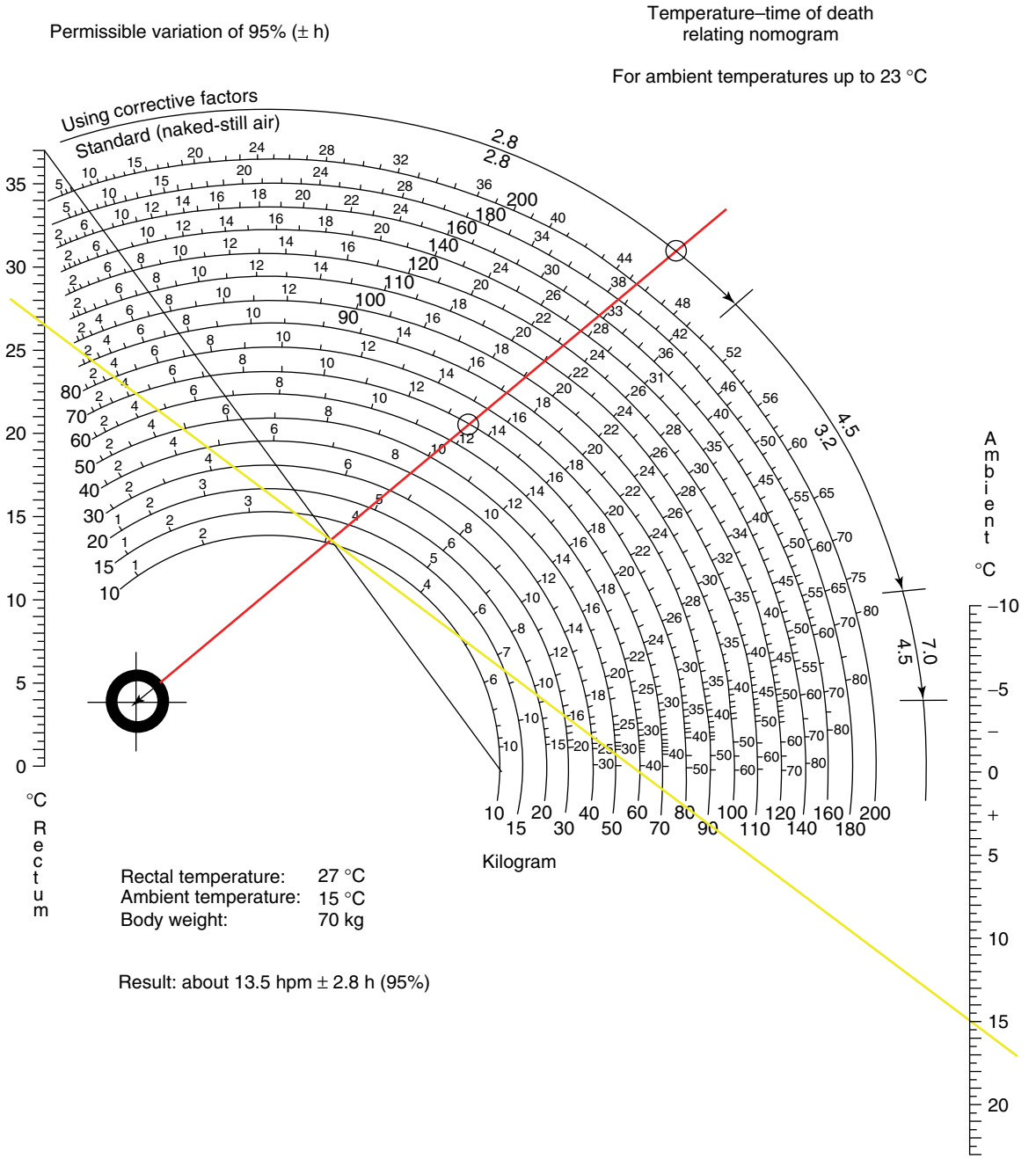


Figure 16 Temperature–time-of-death nomogram for ambient temperatures up to 23 °C. At the scene of crime, for instance, a rectal temperature of 27 °C at an ambient temperature of 15 °C was measured. At first, the point of the scale of the measured rectal temperature and ambient temperature has to be joined by a straight line (yellow), which crosses the diagonal of the nomogram at a specific point. For the second step, a second straight line has to be drawn passing through the center of the circle (below left of the nomogram) and the intersection of the first line and the diagonal (red line). The second line crosses the semicircles for different body weights. The time since death (in this case, for 70 kg body weight) can be read off at the intersection of the semicircle of the given body weight. The intersection gives the mean time since death, and the intersection with the outer circle the 95% limits of confidence, which may be higher if corrective factors have to be used

using computer programs developed by Henssge or using a nomogram (Figure 16).

The nomogram is valid for the chosen standard conditions of cooling (naked body with dry surfaces, lying extended on a thermally indifferent base, in still air). Conditions that improve or delay body cooling compared to standard conditions (lying in water and wind on one hand, clothing and covering on the other hand) reduce or increase the real body weight. Extensive cooling experiments under varying cooling conditions led to empirically found corrective factors for the body weight (Table 9). With these corrective factors, the nomogram can be used for nonstandard cases. These corrective factors themselves are dependent on the body weight (Table 10) in cases of a low and high body weight under higher thermic insulation conditions. Higher body weights need lower factors and, vice versa, lower weights need higher factors [37]. The nomogram can also be used in cases of sudden change of ambient temperature [47].

Application of the Nomogram Method in Casework

1. Inspection of the scene of crime, body posture, clothing, covering, sunshine on the body, windows (closed, opened: when?), radiators, and floor (thermally indifferent?).
2. Measurement of ambient temperature (air) close to the body and at the same level (10–20 cm above the base); measuring the temperature of

Table 9 Empiric corrective factors (CF) for body weight[(a)]

Dry clothing/ covering	Air	Corrective factor	Wet through clothing/ covering wet body surface	In air	In water
		0.35	Naked		Flowing
		0.5	Naked		Still
		0.7	Naked	Moving	
		0.7	1–2 thin layers	Moving	
Naked	Moving	0.75			
1–2 thin layers	Moving	0.9	2 or more thicker layers	Moving	
Naked	Still	1.0			
1–2 thin layers	Still	1.1	2 thicker layers	Still	
2–3 thin layers		1.2	More than 2 thicker layers	Still	
1–2 thicker layers	Moving or still	1.2			
3–4 thin layers		1.3			
More thin/thicker layers	Without influence	1.4			
Thick blanket		1.8			
and		2.4			
Clothing combined		2.8			

[(a)] – The listed values of CF apply to bodies of average weight (reference: 70 kg) (see Table 10) in an extended position on a thermally indifferent supporting base.
– "Thermally indifferent" supporting bases are, e.g., usual floor of rooms, dry soil, lawn, or asphalt. In comparison, bases that appear more thermically insulating or heat conducting should additionally be taken into account.
– Excessively thick upholstered bases require CF of 1.3 for naked bodies. In cases of clothed bodies, CF should be increased by 0.1 units (thickly clothed) to 0.3 units (very thickly clothed).
– Insulating but not excessively thick upholstered bases such as a mattress (bed) or thick carpet require CF of 1.1–1.2 for naked bodies.
– Bases that accelerate cooling, e.g., concrete, stony or tiled bases on ground require CF up to 0.75 for naked bodies.
– In cases of clothed bodies lying on bases, CF should be reduced by 0.1 unit (thicker clothes) or by 0.2 units (very thin clothes)

Table 10 Chart of the dependence of corrective factors on the body weight

	Real body weight (kg)																	
Cooling conditions	4	6	8	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150
	Average range																	
										1.3								
Clothing, more layers	1.6	1.6	1.6	1.6	1.5	–	–	–	–	**1.4**	–	–	–	–	1.3	1.2	1.2	1.2
	2.1	2.1	2.0	2.0	1.9	1.8	–	–	–	**1.6**	–	–	–	1.4	1.4	1.4	1.3	1.3
Bedspread	2.7	2.7	2.6	2.5	2.3	2.2	2.1	2.0	–	**1.8**	–	–	1.6	1.6	1.6	1.5	1.4	1.4
	3.5	3.4	3.3	3.2	2.8	2.6	2.4	2.3	–	**2.0**	–	1.8	1.8	1.7	1.6	1.6	1.5	1.5
	4.5	4.3	4.1	3.9	3.4	3.0	2.8	2.6	2.4	**2.2**	2.1	2.0	1.9	1.8	1.7	1.7	1.6	1.6
Clothing + bedspread	5.7	5.3	5.0	4.8	4.0	3.5	3.2	2.9	2.7	**2.4**	2.3	2.2	2.1	1.9	1.9	1.8	1.7	1.6
	7.1	6.6	6.2	5.8	4.7	4.0	3.6	3.2	2.9	**2.6**	2.5	2.3	2.2	2.1	2.0	1.9	1.8	1.7
Feather mattress	8.8	8.1	7.5	7.0	5.5	4.6	3.9	3.5	3.2	**2.8**	2.7	2.5	2.3	2.2	2.0	1.9	1.8	1.7
	10.9	9.8	8.9	8.3	6.2	5.1	4.3	3.8	3.4	**3.0**	2.8	2.6	2.4	2.3	2.1	2.0	1.9	1.8

the underlying surface as well. Have any changes of thermic conditions been made since the body was found?

3. Single measurement of deep rectal temperature at the scene using an officially calibrated electronic thermometer with probes for measuring air, surface, and rectal temperature. The deep rectal temperature must be at least measured 8 cm within the anal sphincter.
4. At autopsy control, whether the estimation of the body weight was correct by weighing the body.
5. Evaluation of the corrective factor. Are there any conditions that accelerate or delay cooling compared to standard conditions? For rectal temperature, only the thermical conditions of the lower trunk are relevant:
   (a) clothing/covering;
   (b) resting or moving air;
   (c) kind of supporting base.
   In cases of strong insulation conditions and very high or low body weight, the corrective factor must be adapted to the body weight.
6. Using the nomogram.

Connect the points on the scales for rectal and ambient temperature by a straight line. This line crosses the diagonal at a particular point. Draw a second straight line going through the center of the circle, below the left of the nomogram and the intersection of the first line and the diagonal. The second line crosses the semicircles that represent the body weights. At the intersection of the semicircle with the body weight, the time of death can be read. The second line touches a segment of the outermost semicircle. Here the permissible variation of 95% can be seen for standard cases or cases using corrective factors. If ambient temperature or corrective factors are in question, repeat the procedure with other appropriate values. If the chosen ranges of ambient temperature and corrective factors are rather wide, it is recommended to give two values of time: the shortest time resulting from the combined lower limits of the evaluated ranges of body weight, ambient temperature, and corrective factor and the longest time resulting from the upper limits of body weight, ambient temperature, and corrective factor.

The range of time of death can be seen in this way.

If the rectal temperature has nearly reached the ambient temperature ($Q < 0.3$), only a minimum interval of the time since death should be given (mean value minus appropriate 95% limit of confidence).

For temperatures above 23 °C, another nomogram developed on the data of De Saram must be used (Figure17). The difference between this nomogram and that for temperatures up to 23 °C is only due to the relative lengths of the postmortem temperature plateau; it is shorter in higher ambient temperatures.

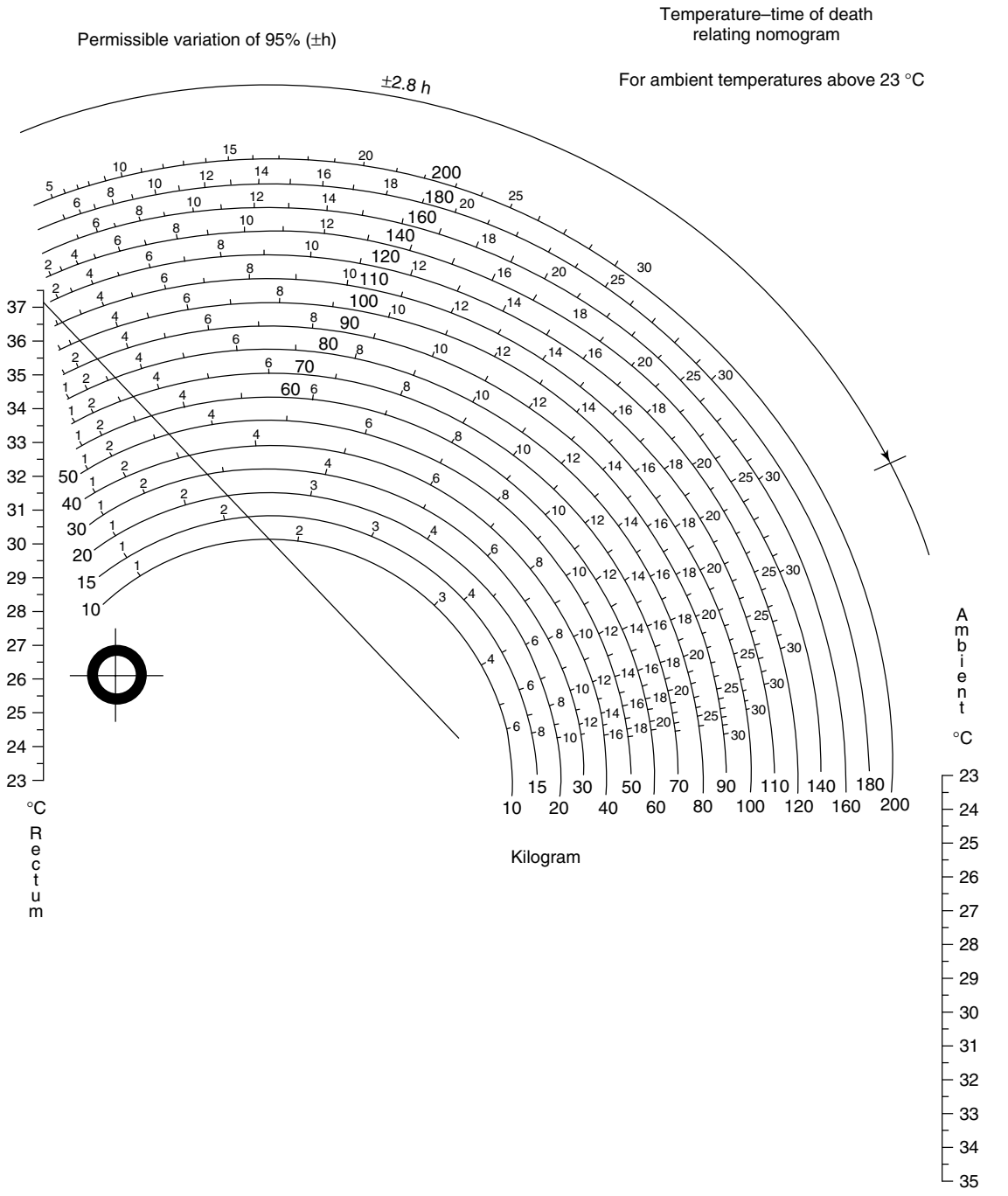


Figure 17 Temperature–time-of-death nomogram for ambient temperatures above 23 °C

Requirements for use:

- no strong radiation (for example, sun, heater, cooling system);
- no strong fever or general hypothermia;
- place of death must be the scene where the body was found;
- no uncertain severe changes of the cooling conditions in the period between death and examination;
- no high thermal conductivity of the surface beneath the body; and
- no longer terminal period after fatal injury (time since death may be inconsistent with time of assault).

Meanwhile, a multicenter [48] and a single center [18] study on the accuracy of the nomogram method compared to the investigations of the police comprising 76 and 72 cases, respectively, were carried out. These are two of the very rare field studies on the accuracy of proposed methods. Both studies covered a wide range of ambient temperatures, body weights, and corrective factors. The estimated period did not contradict the investigated period in any of the cases. Especially in the earlier stages, investigations of the police can be supported effectively by the method.

Cooling conditions can be simulated using cooling dummies. For further information, see References [37, 49, 50].

From all methods developed to estimate the time since death, the nomogram method based on the two-exponential term by Marshall and Hoare is by far the most successful and reliable one since

- actual cooling conditions can be taken into consideration quantitatively;
- data on the precision of the method are available; and
- field studies have confirmed the accuracy and reliability of the method.

Further Postmortem Changes

Autolytic changes, especially in vitreous humor [51–57] and putrefactive changes [58, 59], to the extent they are of importance for determining the time since death [60, 61], are addressed in the chapter (*see* **Time of Death Determinations**; **Chain of Possession of Tangible Evidence**), time of death determinations. There are also suggestions on how the postmortem changes discussed here can be used for estimating the time since death.

References

[1] Madea, B. (2006). *Praxis Rechtsmedizin. Befunderhebung, Rekonstruktion, Begutachtung*, Springer Berlin, Heidelberg, New York, p. 2.

[2] Madea, B., Preuß, J. & Musshoff, F. (2005). Fälschliche Todesfeststellung bei einer Lebenden mit Paroxetinintoxikation, *Notfall + Rettungsmedizin* **8**, 548–551.

[3] Bate-Smith, E.C. & Bendall, J.R. (1949). Factors determining the time course of rigor mortis, *The Journal of Physiology* **110**, 47–65.

[4] Madea, B. (1994). Importance of supravitality in forensic medicine, *Forensic Science International* **69**, 221–241.

[5] Madea, B., Krompecher, T. & Knight, B. (2002). Muscle and tissue changes after death, in *The Estimation of the Time Since Death in the Early Postmortem Period*, 2nd Edition, C. Henssge, B. Knight, T. Krompecher & B. Madea, L. Nokes, eds, Edward Arnold, London.

[6] Madea, B. & Henssge, C. (2003). Timing of death, in *Forensic Medicine: Clinical and Pathological Aspects*, J. Payne-James, A. Busuttil & W. Smocx, eds, Greenwich Medical Media Limited, London, pp. 91–114.

[7] Dotzauer, G. (1958). Idiomuskulärer Wulst und postmortale Blutung, *Deutsche Zeitschrift fur Die Gesamte Gerichtliche Medizin* **46**, 761–771.

[8] Madea, B. & Henssge, C. (1985). Historisches zur Todeszeitbestimmung, *Zeitschrift für Rechtsmedizin-Journal of Legal Medicine* **95**, 19–25.

[9] Madea, B. (1990). Längsschnittuntersuchungen zur supravitalen elektrischen Erregbarkeit der Skelettmuskulatur. I. Objektivierung der supravitalen Muskelkontraktion, *Zeitschrift fur Rechtsmedizin-Journal of Legal Medicine* **2**, 107–121.

[10] Madea, B. (1990). Längsschnittuntersuchungen zur supravitalen elektrischen Erregbarkeit der Skelettmuskulatur. II. Quantifizierung der supravitalen Muskelkontraktion, *Zeitschrift fur Rechtsmedizin-Journal of Legal Medicine* **3**, 44–50.

[11] Madea, B. (1992). Estimating time of death from measurement of electrical excitability of skeletal muscle, *Journal of the Forensic Science Society* **32**, 117–129.

[12] Madea, B. & Rödig, A. (2006). Precision of estimating the time since death using different criteria of excitability, *Forensic Science, Medicine and Pathology* **2**(2), 127–133.

[13] Henssge, C. & Madea, B. (1988). *Methoden zur Bestimmung der Todeszeit an Leichen*. Schmidt-Römhild-Verlag, Lübeck.

[14] Klein, A.& Klein, S. (1978). *Todeszeitbestimmung am Menschlichen Auge*. MD Thesis, Dresden University.

[15] Madea, B. & Henssge, C. (1990). Electrical excitability of skeletal muscle postmortem in casework, *Forensic Science International* **47**, 207–227.

[16] Mallach, H.J. & Mittmeyer, H.J. (1971). Totenstarre und Totenflecke, *Zeitschrift für Rechtsmedizin-Journal of Legal Medicine* **69**, 70–78.

[17] Henssge, C., Knight, B. & Krompecher, T., Madea, B, Nokes, L (2002). *The Estimation of the Time Since Death in the Early Postmortem Period*, 2nd Edition, Edward Arnold, London.

[18] Henssge, C., Althaus, L. & Bolt, J., Freislederer, A, Haffner, H.T, Henssge, C.A, Hoppe, B, Schneider, V (2000). Experiences with a compound method for estimating the time since death. I. Rectal temperature nomogram for time since death. II. Integration of non-temperature-based methods, *International Journal of Legal Medicine* **6**, 303–319, 320–331.

[19] Henssge, C.& Madea, B. (2004). Leichenerscheinungen und Todeszeitbestimmung, in *Handbuch Gerichtliche Medizin Bd. I*, B. Brinkmann, & B. Madea, eds. Springer Berlin, Heidelberg, New York, pp. 79–150.

[20] Henssge, C., Madea, B. & Gallenkemper, E. (1988). Death time estimation in case work II. Integration of different methods, *Forensic Science International* **39**, 77–87.

[21] Camps, F.E. (1959). Establishment of the time of death – a critical assessment, *Journal of Forensic Science* **4**, 73–82.

[22] Spitz, W.U. & Fisher, R.S. (1993). *Medicolegal Investigation of death: Guidelines for the Application of Pathology to Crime Investigation* 3rd Edition, C. Charles Thomas Publisher, Springfield.

[23] Patscheider, H. & Hartmann, H. (1993). Leitfaden der Gerichtsmedizin, Vol. 3. Auflage, Hans Huber, Bern, Stuttgart, Toronto.

[24] Naeve, W. (1978). *Gerichtliche Medizin für Polizeibeamte*, Kriminalistik, Heidelberg.

[25] Mallach, H.J. (1964). Zur Frage der Todeszeitbestimmung, *Berlin Medicine* **18**, 577–582.

[26] Hunnius, P.V., Mallach, H.J. & Mittmeyer, H.J. (1973). Quantitative pressure measurements of livores mortis relative to the determination of the time of death, *Zeitschrift fur Rechtsmedizin-Journal of Legal Medicine* **73**, 235–244.

[27] Bate-Smith, E.C. & Bendall, J.R. (1947). Rigor mortis and adenosinetriphosphate, *The Journal of Physiology* **106**, 177–185.

[28] Bendall, J.R. (1973). Postmortem changes in muscles. in *The Structure and Function of Muscle*, G.H. Bourne, eds, Vol. 2. Structure part 2, London Academic press, pp. 244–309.

[29] Krompecher, T. (2002). Rigor mortis: estimation of the time since death by the evaluation of the cadaveric rigidity, in *The Estimation of the Time Since Death in the Early Postmortem Period*, 2nd Edition, C. Henssge, B. Knight, T., Krompecher, B. Madea & L. Nokes, eds, Edward Arnold, London, pp. 144–160.

[30] Krompecher, T. & Bergerioux, C. (1988). Experimental evaluation of rigor mortis. VII. Effect of ante- and postmortem electrocution on the evolution of rigor mortis, *Forensic Science International* **38**, 27–35.

[31] Krompecher, T. & Fryc, O. (1978). Experimentelle Untersuchungen an der Leichenstarre unter Einfluss von körperlicher Anstrengung, *Beiträge zur Gerichtlichen Medizin* **36**, 345–349.

[32] Krompecher, T. & Fryc, O. (1978). Experimental evaluation of rigor mortis. III. Comparative study of the evolution of rigor mortis in different sized muscle groups in rats, *Forensic Science International* **12**, 97–102.

[33] Krompecher, T. & Fryc, O. (1978). Experimental evaluation of rigor mortis. IV. Change in strength and evolution of rigor mortis in the case of physical exercise preceding death, *Forensic Science International* **12**, 103–107.

[34] Krompecher, T. & Fryc, O. (1979). Zur Frage der Todeszeitbestimmung auf Grund der Leichenstarre, *Beitrage zur Gerichtlichen Medizin* **37**, 285–289.

[35] Krompecher, T. & Fryc, O. (1981). Experimental evaluation of rigor mortis. V. Effects of temperature on the evolution of rigor mortis, *Forensic Science International* **17**, 19–26.

[36] Beier, G., Liebhardt, E. Schuck, M., & Spann, M. (1977). Measurement of rigor mortis on human skeletal muscles in situ, *Zeitschrift fur Rechtsmedizin-Journal of Legal Medicine* **79**, 277–283.

[37] Henssge, C. (2002). Temperature based methods II, in *The Estimation of the Time Since Death in the Early Postmortem Period*, 2nd Edition, C. Henssge, B. Knight, T. Krompecher, B. Madea & L. Nokes, eds, Edward Arnold, London.

[38] Knight, B.& Nokes, L. (2002). Temperature based methods. I, in *The Estimation of the Time Since Death in the Early Postmortem Period*, 2nd Edition, Henssge, C., Knight, B., Krompecher, T., Madea, B. & Nokes, L., eds, Edward Arnold, London, pp. 3–42.

[39] Marshall, T.K. & Hoare, F.E. (1962). I estimating the time of death. The rectal cooling after death and its mathematical expression. II The use of the cooling formula in the study of postmortem body cooling. III The use of the body temperature in estimating the time of death, *Journal of Forensic Science* **7**, 56–81, 189–210, 211–221.

[40] Brown, A. & Marshall, T.K. (1974). Body temperature as a means of estimating the time of death, *Forensic Science* **4**, 125–133.

[41] Henssge, C. (1992). Rectal temperature time of death nomogram: dependence of corrective factors on the body weight under stronger thermic insulation conditions, *Forensic Science International* **54**, 51–56.

[42] Henssge, C. (1979). Precision of estimating the time of death by mathematical expression of rectal body cooling, *Zeitschrift fur Rechtsmedizin-Journal of Legal Medicine* **83**, 49–67.

[43] Henssge, C. (1988). Death time estimation in case work I. The rectal temperature time of death nomogram, *Forensic Science International* **38**, 209–236.

[44] Henssge, C., Beckmann, E.R., Wischhusen, F. & Brinkmann, B. (1984). A determination of time of death by measuring central brain temperature, *Zeitschrift fur Rechtsmedizin-Journal of Legal Medicine* **93**, 1–22.

[45] Henssge, C., Frekers, R. Reinhardt, S. & Beckmann, E.R. (1984). Determination of time of death on the basis of simultaneous measurement of brain and rectal temperature, *Zeitschrift fur Rechtsmedizin-Journal of Legal Medicine* **93**, 123–133.
[46] De Saram, G.S.W., Webster, G. & Kathirgamatamby, N. (1955). Postmortem temperature and the time of death, *The Journal of Criminal Law and Criminology* **46**, 562–577.
[47] Althaus, L. & Henssge, C. (1999). Rectal temperature time of death nomogram: Sudden change of ambient temperature, *Forensic Science International* **99**, 171–178.
[48] Albrecht, A., Gerling, I. & Henssge, C. (1990). Zur Anwendung des Rektaltemperatur-Todeszeit-Nomogramms am Leichenfundort, *Zeitschrift fur Rechtsmedizin-Journal of Legal Medicine* **103**, 257–278.
[49] Henssge, C., Hahn, S. & Madea, B. (1986). Praktische Erfahrungen mit einem Abkühlungsdummy, *Beitrage zur Gerichtlichen Medizin* **44**, 123–126.
[50] Henssge, C., Madea, B. Schaar, U. & Pitzken C (1987). Die Abkühlung eines Dummy unter verschiedenen Bedingungen im Vergleich zur Leichenabkühlung, *Beitrage zur Gerichtlichen Medizin* **45**, 145–149.
[51] Lange, N., Swearer, S. & Sturner, W.Q. (1994). Human postmortem interval estimation from vitreous potassium: An analysis of original data from six different studies, *Forensic Science International* **66**, 159–174.
[52] Madea, B. (2005). Is there recent progress in the estimation of the postmortem interval by means of thanatochemistry? *Forensic Science International* **151**, 139–149.
[53] Madea, B., Herrmann, N. & Henssge, C. (1990). Precision of estimating the time since death by vitreous potassium – comparison of two different equations, *Forensic Science International* **46**, 277–284.
[54] Madea, B., Käferstein, H. Herrmann, N. & Sticht, G. (1994). Hypoxanthine in vitreous humour and cerebrospinal fluid – a marker of postmortem interval and prolonged (vital) hypoxia? Remarks also on hypoxanthine in SIDS, *Forensic Science International* **65**, 19–31.
[55] Madea, B., Kreuser, C. & Banaschak, S. (2001). Postmortem biochemical examination of synovial fluid – a preliminary study, *Forensic Science International* **118**(1), 29–35.
[56] Madea, B. & Henssge, C. (2002). Eye Changes After Death, in *The Estimation of the Time since Death in the Early Postmortem Period*, 2nd Edition, C. Henssge, B. Knight, T. Krompecher, B. Madea & L. Nokes, eds, Edward Arnold, London, pp. 103–133.
[57] Madea, B. & Rödig, A. (2006). Time of death dependent criteria in vitreous humor – precision of estimating the time since death, *Forensic Science International* **164**, 87–92.
[58] Madea, B. (2001). Estimation of duration of immersion, *Scandinavian Journal of Forensic Medicine* **8**(1), 4–9.
[59] Reh, H. (1969). *Diagnostik des Ertrinkungstodes und Bestimmung der Wasserzeit*, Triltsch, Düsseldorf.
[60] Henssge, C. & Madea, B. (2004). Estimation of the time since death in the early postmortem period, *Forensic Science International* **144**(2–3), 167–175.
[61] Henssge, C. & Madea, B. (2007). Estimation of time since death, *Forensic Science International* **165**, 182–184.

Further Reading

Madea, B., Oehmichen, M. & Henssge, C. (1986). Postmortaler Transport von Mageninhalt, *Zeitschrift für Rechtsmedizin-Journal of Legal Medicine* **97**, 201–206.
Madea, B., Henssge, C., Hönig, W. & Gerbracht, A (1989). References for determining the time of death by potassium in the vitreous humour, *Forensic Science International* **8**, 231–243.

Related Articles

Time of Death Determinations

BURKHARD MADEA

Death: Timing ***see*** Death: Time of, Time of Death Determinations

Death: Vehicle ***see*** Traffic Fatalities

Death, Cause of: Retrospective Psychological Analysis ***see*** Psychological Autopsy

Death Penalty: Mitigating Testimony *see* Mitigation Testimony

Death Penalty and Age

At the end of 2005, 36 states within the United States of America and the federal prison system held 3254 prisoners under sentence of death [1]. The youngest inmate under sentence of death was 20 and the oldest was 90. Those executed in 2005 had been under sentence of death an average of 12 years and three months.

Lifespan developmental perspectives emphasize continuity and change from birth to death, with notable changes observed in both cognitive capacity and physical status. It has been noted that the period between infancy and adulthood is a time of tremendous growth and change in which an individual's cognitive abilities broaden and intensify, and their physical capabilities are fully developed. In contrast, the period between early adulthood and late adulthood is one of relative quiescence in which both cognitive and physical abilities remain relatively stable while gains in emotion regulation, motivation, and concepts of self predominate. This period is then followed by yet another period of tremendous change that is typified by cognitive and physical decline that ultimately results in death.

In two recent landmark decisions, the United States Supreme Court (USSC) recognized that cognitive capacity was relevant in deciding what class of individuals should be subject to the death penalty. In the 2002 decision in *Atkins v. Virginia* [2], and the 2005 decision in *Roper v. Simmons* [3] the USSC made it unconstitutional to execute individuals with mental retardation and individuals who were under the age of 18 years at the time they committed a capital offense (*see also* **Juvenile Justice: Adolescent Development**). To date, the USSC has been silent to the issue of cognitive decline and its relevance to the stated goals of the death penalty.

The End of an Era: The United States Supreme Court Abolishes the Death Penalty for Juveniles

As previously noted, over the years the legal system has recognized the importance of age. The importance was recognized as far back as 1899 when the American juvenile justice system decided that adolescents who violate certain laws should be provided with treatment, rather than be punished for their actions [4]. In cases such as *Belotti v. Baird* [5] and *Eddings v. Oklahoma* [6] the USSC recognized that the period of adolescence was distinct from adulthood and that during this period, individuals were vulnerable to the influence of others and did not possess the level of experience, perspective taking, or maturity required to make critical decisions in an informed manner.

Despite the recognition that the period of adolescence was different from that of adulthood, the USSC did not intervene on behalf of juveniles until 1988 in the case of *Thompson v. Oklahoma* [7]. In *Thompson*, the USSC stated that the execution of a person who was 15 years of age or younger at the time of the crime violated the Eighth Amendment against cruel and unusual punishment. The Court was guided by "evolving standards of decency" [8] that rendered such punishment unconstitutional. By 1988, 18 states had restricted the death penalty to defendants aged 16 years or above at the time of the crime, and there had not been an execution of someone under the age of 16 years for 39 years. In addition to noting the evolving standards of decency, the Court in *Thompson* directed attention to the importance of inexperience, limited education, and limited intellectual functioning in making sentencing decisions. Quite simply, the Court believed that the goals of retribution and deterrence could not be met with the execution of youthful defendants. Lastly, the Court commented that their decision was supported by the fact that professional organizations and leading members of the "Western European Community" believed that executing a defendant who was under the age of 16 years at the time of the crime offended civilized standards of decency. Of note, one year later, the application of the death penalty for youth who were between the ages of 15 and 18 years at the time of the crime was upheld in *Stanford v. Kentucky* [9].

It took 17 years for the USSC to take a major step in changing the laws regarding capital punishment in the United States. In the case of *Roper v. Simmons*

[2], the Court decided that the death penalty was not appropriate for defendants who were under the age of 18 years at the time of the crime. The decision came on the heels of the Court's decision to abolish the death penalty for individuals who have mental retardation [2]. In their 6-3 opinion, the Court stated that the execution of a mentally retarded individual violates the Eighth Amendment right against cruel and unusual punishment. Writing for the majority, Justice Stevens stated that due to disabilities in areas of reasoning, judgment, and control of their impulses, individuals with mental retardation do not act with the level of moral culpability that characterizes the most serious adult criminal conduct. It was further noted that the presence of these impairments can put into question the reliability and fairness of capital proceedings. Citing *Atkins* as a backdrop, Simmons filed an appeal of his death sentence with the Missouri Supreme Court stating that the same reasoning used in *Atkins* should apply to juvenile offenders. The Missouri Supreme Court agreed and resentenced Simmons to life without the possibility of parole. In 2004 the USSC granted *certiorari* to determine if imposition of the death penalty for an individual under the age of 18 years violated the Eighth Amendment against cruel and unusual punishment. In a 5-4 decision, the USSC banned the execution of individuals that were under the age of 18 years at the time of the offense [3].

As would be expected, the decision in *Roper* was not met with universal acceptance. In their 5-4 decision, the dissenting justices were unified in their opinion that the Missouri court should be admonished for not adhering to precedent as set forth in *Stanford* and that the difference in maturity between adults and juveniles was not substantial, nor was it universal. Justices Scalia, Rehnquist, and Thomas further stated that by including nondeath penalty states in the purported national consensus, the Court artificially elevated their position, and that the use of international law has no place in the opinions that arise from the USSC.

The Emergence of an Era: Aging Criminals and the Death Penalty

Older inmates are the fastest-growing segment of the population in federal, and state prisons in the United States [10]. Aday [11] reported that the number of prisoners 50 years and older in state and federal correctional facilities in 2001 was 113 358, three times as many as the number in 1990. This growth was thought to have resulted from three primary forces: (i) the overall aging of the population, (ii) tougher sentencing laws (e.g., mandatory minimums, and "three strikes" laws), and (iii) the elimination of parole [12]. Although age 50 is not considered old in community settings, research has found that incarcerated individuals of age 50 and older physiologically resemble community-dwelling individuals approximately 10 years older than their chronological age [13]. In other words, a 60-year-old inmate under sentence of death is likely to resemble a community-dwelling person who is 70 years of age. Though the exact cause of this rapid aging process is not known, it may be due to an inmate's history of excessive drug and alcohol use, poor nutrition, stressful life experiences, personal neglect, and lower socioeconomic status in comparison with nonoffenders. As of December 2005, 672 death row inmates were over the age of 50 years, which accounted for approximately 21% of the United States death row population [1].

Although little attention has been given to the enactment of the death penalty among older offenders, questions about the constitutional justification for executing these individuals can be raised due to potential disabilities in reasoning, judgment, and control of their impulses. Such deficits are associated with cognitive incapacity secondary to dementias such as Alzheimer's disease. Alzheimer's disease is the most common form of dementia in the United States, 5.1 million Americans had Alzheimer's disease in 2007, and it is estimated that 11–16 million older Americans will suffer from the disease by the year 2050 [14]. In brief, dementia is a clinical term that describes a variety of illnesses with hallmark symptoms of cognitive decline from previous levels of functioning. These cognitive declines primarily involve memory or executive functioning deficits (e.g., the ability to sustain attention, to plan, to make informed decisions, and to solve problem). Additionally, cognitive decline associated with dementia is accompanied by deficits in adaptive functioning, particularly in the personal care and community functioning domains [15]. With the gradual progression of the disease comes an increasing level of impairment in adaptive functioning. Ultimately, an adult with dementia will forget to carry out day-to-day activities (e.g., they may forget to

take their medication or to pay bills) and over time will forget personal historical events. Older inmates with dementia may not remember the circumstances surrounding the crime for which they are incarcerated or the rationale behind their death sentence.

Numerous case studies have suggested the important role of dementia as a factor in late life criminal activity [11]. Rosner and colleagues [16] examined the cognitive abilities of 52 defendants between the ages of 62 and 88 who were referred for assessment of competence to stand trial. Eighty-three percent of the defendants were facing charges related to violent crimes. Rosner and colleagues found that a substantial number of these older defendants suffered from dementia.

To date, the USSC has not addressed the issue of the execution of elderly inmates under sentence of death or the need to assess their competence for execution. The issue, however, has been cited as a serious concern by death penalty experts in the United States and by experts internationally [17]. According to USA Today February 10, 2005 as cited in the death penalty information site [17], two unidentified elderly inmates have asked federal judges to rule on the constitutionality of executing inmates with Alzheimer's disease, dementia, or other age-related illnesses. As stated by Jonathan Turley of George Washington School of Law, "Dead man walking is one thing ... dead man being pushed along to the execution chamber in a wheelchair is another thing."

By way of case example, in 2004, J.B. Hubbard, a 74-year-old man, was put to death in Alabama for the murder of a woman in 1977. At the time of his execution he had been under sentence of death for 27 years and was the oldest inmate put to death since the reinstatement of the death penalty in 1976. His attorneys argued that executing Mr. Hubbard would be a violation of the Eighth Amendment due to his age and mental incompetence. Specifically, at the time of his execution J.B. Hubbard exhibited multiple symptoms of dementia including the inability to remember names of significant individuals in his life, inability to remember significant events in his life, and deficits in self care [17]. Additionally, he suffered from colon and prostate cancer and was so weak that other inmates helped him walk to the shower and to comb his hair.

Setting the Example: The USSC Prohibits the Execution of Inmates Who Have a Mental Illness and do not Understand or Rationally Appreciate the Reason behind Their Execution

In 2007 the USSC took the position that, to be executed, an individual must understand and appreciate the application of the death penalty in their individual case [18]. Specifically, the Court determined that conducting an execution in the absence of rational thought and understanding violated the Eighth Amendment of the United States Constitution. Thus, in order to execute a mentally ill offender, the offender would need to evince: (i) the ability to understand the purpose of punishment and (ii) the ability to appreciate the personal consequences of previous actions necessitating punishment. As noted in the majority opinion, an offender is incompetent for execution if he does not "have a rational understanding of the penalty about to be imposed". Such understanding and appreciation are necessary to fulfill the purpose of retribution and establish the offender's competence to be executed. It can readily be seen that the cognitive deficits associated with dementia, mimic almost exactly the concerns put forth in *Panetti*.

It is important to note that the decision in *Panetti* was consistent with the constitutional protections established in *Ford v. Wainwright* [19]. In *Ford,* the Court established that "the Eighth Amendment forbids the execution only of those who are unaware of the punishment they are about to suffer and when they are to suffer it". Like *Ford,* the decision in *Panetti* supports the notion that to be executed a person must be aware of the reason for this punishment; however, *Panetti* extends the protections by including the need for a rational, as well as factual, understanding of the basis for the decision. The Court noted that mental health experts can assist the Court in determining competence, and specifically stated that "there is precedent to guide a court in conducting Eighth Amendment analysis". Not surprisingly, the cases cited in *Panetti* were *Atkins v. Virginia* and *Roper v. Simmons.*

References

[1] http://www.ojp.usdoj.gov/bjs/cp.htm, 2006.
[2] Atkins v. Virginia, 536 U.S. 304 (2002).

[3] Roper v. Simmons, 543 U.S. 551 (2005).
[4] Feld, B.C. (1987). The juvenile court meets the principle of the offense: Legislative changes in juvenile waiver statutes, *Journal of Criminal Law and Criminology* **78**, 471–533.
[5] Belotti v. Baird, 443 U.S. 622, 99S.Ct. 03035, 61 L.Ed.2d 797 (1979).
[6] Eddings v. Oklahoma, 455, U.S. 104, 102S.Ct. 869, 71 L.Ed.2d1 (1982).
[7] Thompson v. Oklahoma, 487 U.S. 815 (1988).
[8] Trop v. Dullas, 356 U.S. 86, 101 (1958).
[9] Stanford v. Kentucky, 492 U.S. 361 (1989).
[10] Williams, J.L. (2006). The aging inmate population: southern states outlook, *The Southern Legislative Conference*, Atlanta.
[11] Aday, R.H. (2003). *Aging Prisoners: Crisis in American Corrections*, Praeger, Westport.
[12] Yates, J. & Gillespie, W. (2003). The elderly and prison policy, *Journal of Aging and Social Policy* **11**, 167–175.
[13] Dawes, J. (2003). Dying with dignity: Prisoners and terminal illness, *Illness, Crisis and Loss* **10**, 188–203.
[14] Alzheimer's Association (2007). *Alzheimer's Disease Facts and Figures*, Alzheimer's Association.
[15] American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders*, 4th Edition, text revision. American Psychiatric Association, Washington, DC.
[16] Rosner, R., Wiederlight, M., Harmon, R.B. & Cahn, D.J. (1991). Geriatric offenders examined at a forensic psychiatry clinic, *Journal of Forensic Sciences* **36**, 1722–1731.
[17] http://www.deathpenaltyinfo.org/time-death-row#aging, 2008.
[18] Panetti v. Quarterman, 551 U.S. http://www.oyez.org/cases/2000–2009/2006/2006_06_6407/ (2007).
[19] Ford v. Wainwright, 477, U.S. 399 (1986).

Related Articles

Juvenile Justice: Adolescent Development

KAREN L. SALEKIN AND REBECCA S. ALLEN

Death Penalty and Mental Retardation *see* Mental Retardation: Death Penalty

Debris: Explosion, Analysis *see* Explosion Debris: Laboratory Analysis of

Debris: Fire, Analysis *see* Fire Debris: Laboratory Analysis of

Deception *see* Deception: Truth Serum

Deception: Detection of

Introduction

Deception detection is a useful and critical ability for humans in variety of settings. Growing concerns about terrorism have highlighted the need for accurate detection of deception. All the 19 of the September 11, 2001 terrorists lied to the officials on at least three occasions (when applying for an entry visa, when crossing the border, and before boarding their airplanes) and none were detected in their deception. This is perhaps not surprising given the substantial body of research showing that human beings are not innately skilled in detecting deception [1, 2].

Moreover, even those whose professions appear to require deception detection are not good at deception detection. Recognizing this weakness in human ability; police, national security experts, and scientist have explored other strategies to detect deception. One alternative approach to deception detection is to use technology to examine various physiological responses that may provide cues for deception detection. To date, only one of those approaches, the polygraph, has received widespread application

around the world [3], although other approaches have been attempted (*see* **Deception: Detection of and Brain Imaging**).

Psychophysiological Detection of Deception (Polygraph)

Psychophysiological detection of deception (PDD), commonly referred to as *polygraph testing*, involves the use of autonomic physiological measures to make inferences about a subject's credibility. Typically, modern PDD instruments take measures of respiratory activity, relative blood pressure, skin conductance (palmar sweating), and vasomotor activity (blood flow at the surface of the skin, usually at the thumb) [4]. Movement can also be monitored in an attempt to detect artifacts and deliberate distortion. There are two major approaches to PDD, information-based tests and deception-based tests. Information-based tests attempt to determine whether the suspect has specific information that only a guilty person should have. Deception-based tests take a more direct approach and ask direct accusatory questions. These two approaches and their variations are described below.

Information-Based Psychophysiological Detection of Deception. In taking the information-based approach [5], often called the *guilty knowledge test*, or more correctly the *concealed information test* (CIT), if a murder occurred, and the weapon was a 38-caliber revolver, and if that information was not made public, then a suspect could be asked: If you committed the murder you would know the murder weapon. Was the murder weapon: a 45-caliber automatic, a 38-caliber revolver, a 9-mm automatic, a 44-caliber revolver, a 25-caliber automatic, a 22-caliber revolver? The critical item is never placed in first position as the first in such series almost always evokes a physiological response. Several critical items developed from the investigation would be used in a complete information-based examination. The expectation would be that a naïve individual would respond to items at random, whereas the guilty would show a nonrandom pattern with consistent physiological arousal associated with the items identified by the investigation. One advantage of this approach is that the false-positive rate (that is, the rate at which the innocent fail the test) can be predicted based on statistical models of chance.

Laboratory research using the CIT in a variety of mock crime scenarios consistently produced results for CIT that support the statistical models for false-positive rates and suggest a true positive (correct identification of the guilty) rate of around 80% [6]. Unfortunately, the field studies of the CIT have failed to produce equivalent true positive rates. Data from the field suggest that true positive rates for the CIT in the field are about 50% [7, 8]. The difference between the laboratory and the field is perhaps not surprising. The general tact in laboratory studies of the CIT is to choose items for the CIT that have been pilot tested to assure that they are remembered by persons committing the mock crime. Such pilot testing is not possible in field situations and there are no predictive theories of eyewitness memory to use to confidently determine what may be remembered from a crime scene by a highly aroused perpetrator. This inability to assess the memorability of potential CIT items would seem to be a serious problem for the application of the CIT in real cases [8]. Moreover, at least in American forensic cases, the opportunity to apply the CIT seems to be severely limited. Podlesny and his colleagues studied Federal Bureau of Investigation case files and found that only about 10% of their cases were amenable to CIT assessment, even under ideal conditions [9, 10]. To date, Japan is the only country to report widespread application of the CIT [11]. Unfortunately, the supporting evidence for the use of the CIT in Japan is still primarily anecdotal.

Deception-Based Physiological Detection of Deception. The deception-based approaches to PDD use direct accusatory questions, known as *relevant questions*, that address the issue under investigation. For example, in a murder investigation where the victim died from a gunshot, a relevant questions would be, "Did you shoot John Doe?" The earliest deception-based approach was known as the *relevant–irrelevant test*. The relevant–irrelevant test contained only relevant questions and simple irrelevant questions to which the subject was assumed to be answering truthfully, "Are the lights on in this room?" The working theory of the relevant–irrelevant test was that only guilty individuals would show dramatic and consistent responses to the relevant questions. Virtually all contemporary scientists who study PDD state a belief that the working theory of the relevant–irrelevant test is naïve and incorrect. The current belief is that

innocent subjects will easily recognize that the relevant questions are much more important than the irrelevant questions. This recognition is likely to result in a substantial number of false-positive outcomes [4]. Research has supported that belief with the relevant–irrelevant test consistently producing large number of false-positive outcomes [4]. As a result the relevant–irrelevant test currently has little application in forensic testing, although it continues to have application in employment screening [4].

Reid [12] recognized the problems with the relevant–irrelevant test and developed an alternative that replaced the earlier test and is now the test of choice worldwide for forensic PDD tests [3]. Reid's innovation was the introduction of a question that was presented to the subjects in such a manner that it would probably elicit a response that was a lie, "Before the year 2007, did you ever lie to a person in a position of authority?" This question was referred to as a *comparison question*, and the new technique was referred to as a *comparison question test* (CQT). The comparison question is left deliberately vague and covers a long time period. Subjects are told that deception to the comparison questions will make the subject look like the kind of person who could have committed the crime under investigation.

The rationale and the various variations of the CQT have been described in detail elsewhere [4, 8]. The working rationale of the CQT is that guilty individuals will produce strong physiological responses to the relevant questions because they know they are lying to those questions, they have memory of committing the crime, and they are afraid to be caught in their deception. Guilty subjects are not expected to respond strongly to the comparison questions because they address relatively unimportant matters as compared to the relevant questions concerning the crime under investigation. Innocent subjects are expected to produce strong physiological responses to the comparison questions to which they are deceptive and because they have been told that deception to the comparison questions is indicative of criminal involvement. Innocent subjects are not expected to respond strongly to the relevant questions because they are answering them truthfully. The intentional focus of the relevant questions is expected to be on the comparison questions. Overall, the expectation is for an interaction of responding between the question types and guilt. That is, guilty subjects are expected to respond with relevant question responses greater than comparison question responses. Innocent subjects are expected to respond with the opposite pattern, that is, comparison question responses greater than relevant question responses. Currently, there are a number of versions of the CQT in use, in practice, but all of them rely on the same basic idea of comparing the responses to relevant and comparison questions.

During the last 35 years, a large number of studies were conducted on the validity of CQTs. That research was conducted in both the laboratory and field settings. Although those studies have produced a range of validity estimates, it now seems that those studies converge on an estimate of validity around 90%, with some tendency for more false-positive errors, when only the polygraph charts are used for decisions [3, 4, 8]. The tendency toward false-positive errors disappears if only the original examiner's final decisions are evaluated [8].

The CQT has found widespread application in the United States and elsewhere. In the State of New Mexico, the results of polygraph tests are generally admitted as evidence at trial [13]. Across the other states polygraph are sometimes admitted, sometimes used in special proceedings (e.g., sentencing), are frequently used in charging decisions, and they have nearly universal use as an investigative tool. The CQT also has international use currently being used by police in many countries worldwide [8].

There is, however, controversy around the CQT. Some scientists think the working rationale of the CQT is naïve and cannot work. Some have persisted in that belief even in the face of overwhelming data indicating that their position is wrong [14, 15]. Moreover, survey research clearly shows that the majority of the informed scientific community does not support their position [8]. Other scientists criticize the CQT because a complete scientific theory to explain all aspects of the phenomenon is lacking [14]. This position ignores the mass of published research showing high validity for the CQT in application. This is an odd position for scientists to take and it is roughly the equivalent of saying because we lack a complete understanding of gravity, we must therefore not rely on things accelerating toward large masses like a planet. By such reasoning, the pharmaceutical industry and the medical profession should abandon the use of most prescription medicines! In reality, polygraph decision making has been shown to be as accurate as or more accurate than many medical and psychological diagnostic procedures in common use [16].

There are, however, some valid concerns about the application of the polygraph in practice, particularly in the United States. The polygraph profession is poorly regulated and at this time there are no binding standards of practice beyond those provided by those states that license polygraph examiners. The American Society of Testing and Materials International has promulgated a set of nonbinding standards of practice [17] that would seem to be a step toward greater standardization and professionalism in the practice.

There is also a concern about the use of countermeasures to defeat the polygraph. A contemporary review of the scientific research on countermeasures [18] has reached the following conclusions. General state countermeasures that affect the subject throughout the examination (e.g., drugs) are ineffective against the CQT. Specific point countermeasures that can be applied during the test (e.g., muscle tension or mental arithmetic) can be effective for some subjects, if and only if they follow specific training. Spontaneously attempted countermeasures are ineffective. Reading accurate information about the structure of the CQT and possible countermeasures has no effect on the validity of the CQT [19]. Statistical decision models may be more resistant to countermeasure attacks than human-based analysis methods [18].

The use of the polygraph in other than forensic settings is considerably more controversial. A recent review by the National Research Council of The National Academies [20] concluded that there was insufficient data to support the use of polygraph as a screening tool for national security.

Voice Stress Analysis

Over the last 35 years a variety of machines have been marketed claiming to detect deception through the detection of a microtremor in the voice. If such a device worked, it would offer many advantages to those with a need to detect deception. Proponents of these devices, known as a *voice stress analyzers* (VSA) note that their use requires much less training than for PDD, their use is much quicker, and they can be clandestinely. Following their initial introduction. Horvath, in a 1982 review concluded, "Without exception, however, the scientific evidence reported to date shows that voice stress analyzers are not effective in detecting deception; none of these devices has yet been shown to yield detection rates above chance levels in controlled situations." [[21], p. 340]. Although a number of additional studies have been reported since 1982, many of them examining the new computerized version of the VSA, there is no reason to update Horvath's 1982 conclusion. To our knowledge, no controlled study of the VSA has ever produced better than chance results in detecting deception. Given the unanimity of the scientific research, the continued use of the VSA by the police is simply quixotic.

References

[1] Vrij, A. (2008). *Detecting Lies and Deceit: Pitfalls and Opportunities*, Wiley, New York.

[2] Bond, C. & DePaulo Jr, B. (2006). Accuracy of deception judgments, *Personality and Social Psychology Review* **10**, 214–234.

[3] Honts, C. (2004). The psychophysiological detection of deception, in *Detection of Deception in Forensic Contexts*, P. Granhag & L. Strömwall, eds, Cambridge University Press, London, pp. 103–123.

[4] Raskin, D. & Honts, C. (2002). The comparison question test, in *Handbook of Polygraph Testing*, M. Kleiner, ed, Academic, London, pp. 1–49.

[5] Ben-Shakhar, G. & Elaad, E. (2002). The Guilty Knowledge Test (GKT) as an application of psychophysiology: future prospects and obstacles, in *Handbook of Polygraph Testing*, M. Kleiner, ed, Academic, London, pp. 87–102.

[6] MacLaren, V. (2001). A quantitative review of the Guilty Knowledge Test, *Journal of Applied Psychology* **86**, 674–683.

[7] Elaad, E., Ginton, A. & Jungman, N. (1992). Detection measures in real-life criminal guilty knowledge tests, *Journal of Applied Psychology* **77**, 757–767.

[8] Honts, C., Raskin, D. & Kircher, J. (2005). Scientific status: the case for polygraph tests, in *Modern Scientific Evidence: The Law And Science of Expert Testimony*, Forensics 2005–2006 Edition, D. Faigman, D. Kaye, M. Saks & J. Sanders, eds, Vol. 4 Thompson West, Eagan, pp. 571–605.

[9] Podlesny, J. (1993). Is the guilty knowledge polygraph technique applicable in criminal investigations? A review of FBI case records, *Crime Laboratory Digest* **20**, 59–63.

[10] Podlesny, J. (2003). *A Paucity of Operable Case Facts Restricts Applicability of The Guilty Knowledge Technique in FBI Criminal Polygraph Examinations, Forensic Science Communications* **5**,. http://www.fbi.gov/hq/lab/fsc/backissu/july2003/podlesny.htm (accessed Jul 2003).

[11] Nakayama, M. (2002). Practical use of the concealed information test for criminal investigation in Japan, in *Handbook of Polygraph Testing*, M. Kleiner, ed, Academic, London, pp. 49–86.

[12] Reid, J. (1947). A revised questioning technique in lie detection tests, *Journal of Criminal Law, Criminology and Police Science* **37**, 542–547.
[13] Daniels, C. (2002). Legal aspects of polygraph admissibility in the United States, in *Handbook of Polygraph Testing*, M. Kleiner, ed, Academic, London, pp. 327–338.
[14] Iacono, W. & Lykken, D. (2005). Scientific status: the case against polygraph tests, in *Modern Scientific Evidence: The Law And Science Of Expert Testimony*, Forensics 2005–2006 EditionD. Faigman, D. Kaye, M. Saks & J. Sanders, eds, Thompson West, Eagan, Vol. 4, pp. 605–655.
[15] Ben-Shakhar, G. (2002). A critical review of the control questions test (CQT), in *Handbook of Polygraph Testing*, M. Kleiner, ed, Academic Press, London, pp. 103–126.
[16] Crewson, P. (2001). *A Comparative Analysis of Polygraph with Other Screening and Diagnostic Tools*, Report on Contract No. DABT60-01-P-3017, Department of Defense Polygraph Institute.
[17] ASTM E 2062-00. (2002). *Standard Guide of PDD Examination Standards of Practice. Annual Book of ASTM Standards*, ASTM Committee E-52 on Forensic Psychophysiology, Vol. 14.02.
[18] Honts, C. & Amato, S. (2002). Countermeasures, in *Handbook of Polygraph Testing*, M. Kleiner, ed, Academic Press, London, pp. 251–264.
[19] Honts, C. & Alloway, W. (2007). Information does not affect the validity of a comparison questions test, *Legal and Criminological Psychology* **12**, 311–320.
[20] Committee to Review the Scientific Evidence on the Polygraph (National Research Council (U.S.)) (2003). *The Polygraph and Lie Detection*, The National Academies Press, Washington, DC.
[21] Horvath, F. (1982). Detecting deception: the promise and the reality of voice stress analysis, *Journal of Forensic Science* **27**, 340–351.

CHARLES R. HONTS

Deception: Detection of and Brain Imaging

Introduction

In this article we summarize the field of deception detection by means of brain activity measures. In the following, we use the terms *deception* and *lying* as synonyms, although some scholars have distinguished between the two (but cf. [1]).

Over the last century, the field of deception detection has gone through three phases. In the first phase, deception detection methods were based on behavioral cues, such as blinking rate and voice pitch, brought about by increases in arousal generated by the act of deceiving. However, no behavioral cues have proven reliably to signal deception [2, 3].

In a second phase, aided by technological advances, the field focused on the detection of subtler changes associated with shifts in the level of arousal that are not visible to the naked eye; that is, researchers focused on changes in peripheral psychophysiological variables, such as heart rate and skin conductance. This phase led to the development of the polygraph that is still the instrument most widely used for deception detection (*see* **Deception: Detection of**). However, despite their widespread application, these measures also proved to be unreliable [4].

During the third phase, in order to find variables that were more diagnostic of deception, researchers began to look at measures of processes that index more directly activity in the brain, the system that produces the lies. That is, rather than rely on the downstream results of brain processing, be they overt behaviors or psychophysiological changes, researchers have turned to the source – the brain – and tried to discover whether deception is signaled by distinctive patterns of neural activation.

Brain Measures and the Detection of Deception

Although there are emerging technologies that may be useful in the field (e.g., optical imaging), the primary technologies used to detect deception with brain-based measures have been event-related potentials (ERPs) and neuroimaging, primarily functional magnetic resonance imaging (fMRI).

ERPs and Detection of Deception

ERPs are derived from the electroencephalogram (EEG), a noninvasive measure of electrical brain activity obtained from electrodes placed on the scalp of an individual. Given the portability and low cost of the technology, it is not surprising that there has been considerable interest in using ERPs in efforts to detect deception – with the hope of eventual

field applications. The EEG is the summation of a myriad of neural processes occurring in the brain at the same time. Thus, to extract measures of brain responses specific to an event of interest, such as the presentation of a word or a picture, one needs to average the EEG data over tens or hundreds of repetitions of the same type of event. ERPs are obtained by time-locking the averages to the onset of the event of interest, which decreases the influence of processes that are not related to the event of interest while at the same time increases the influence of processes that are in fact related to it. The resulting average shows a characteristic sequence of positive and negative deflections that unfold over time, referred to as *ERP components*.

One of the most reliable ERP components is the P300, a positive peak that occurs about 300 ms after a rare and meaningful stimulus is perceived. The P300 is the ERP component that has been most commonly used in deception research, almost exclusively in *concealed information tests* (CITs), also known as *guilty knowledge tests* (GKTs). Note that, technically, these tests do not detect deception *per se*, because the neural signatures of stimulus recognition are present regardless of whether or not the person is being deceptive. However, researchers hope that these tests will enable one to infer the presence of deceptive behavior.

These tests involve presenting three types of stimuli: (i) *probes*, that is, stimuli related to the information that is concealed and that are known to the guilty individual but not to innocent individuals; (ii) *irrelevants*, that is, stimuli that are unrelated to the concealed information and unknown to both the perpetrator of a crime and innocent individuals; and (iii) *targets*, that is, stimuli included to ensure that participants pay attention to the task. The proportion of probe stimuli is very low, compared to that of irrelevant stimuli and the probe stimuli, ideally, are recognized only by the guilty individuals. Thus, probe stimuli generate larger P300s than irrelevant stimuli in guilty but not in innocent individuals. This is the core logic of the ERP approach for detection of deception. Variants of these paradigms differ primarily in the details of the analyses.

Rosenfeld and his collaborators reported the first peer-reviewed ERP studies that used the P300 as an index of concealed knowledge in an effort to detect deception [5, 6]. In these studies, the researchers found that the P300 could be used as an index of knowledge that the participant possesses, even if the participant denied possessing it. Thus, the P300 could be employed to determine what the participants know and also to determine whether they may be lying about what they know.

Numerous methods based on the P300 have been proposed by different laboratories. However, the efficacy of these methods is still controversial, and different laboratories have reported different accuracy results. For example, Farwell and Donchin [7] reported accuracy of 87.5% (accuracy is usually defined as the average of the guilty individuals judged to be guilty and innocent individuals judged to be innocent by the method at hand). Farwell is also associated with the most controversial P300-based method, referred to as *Brain Fingerprinting*. This term relies on a false analogy with actual fingerprinting, implying that the brain stores an immutable replica of an external event. The assumption is that such replica can be accessed via brain fingerprinting. Farwell has claimed 100% accuracy with an improved method, the "Memory and Encoding Related Multifaceted Electroencephalographic Response" (MERMER) but, to date, no peer-reviewed evidence of the accuracy of the method has been published in psychophysiology or neuroscience journals [8].

In numerous publications on the topic, Peter Rosenfeld and collaborators have reported accuracy rates between 80 and 95%, depending on the paradigm and analyses employed. Similarly, John Allen and collaborators reported detection rates of over 90% with overlearned information [9], but less than 50% with mock crime scenarios [10]. The only independent field study using these methods reported accuracy rates of only 48% [11]. Finally, a recent comparison of methods for ERP assessment in the P300-based CIT paradigms reported that accuracy rates varied within the 74–80% range [12].

fMRI and Detection of Deception

fMRI, unlike ERPs, has very good spatial resolution, and so it can localize brain processes with great precision. However, the current temporal resolution of the technique is rather poor, compared to the speed of processing in the human brain; fMRI is based on detecting relatively slow changes in regional cerebral blood flow produced by neural activity. Furthermore, the lack of portability and cost of

the technology (fMRI scanners weigh several tons and require complex infrastructure to be maintained) raises critical issues regarding its application in the field.

Since 2001, there have been numerous fMRI studies of deception using various paradigms [13–27]. Most studies have employed versions of the CIT, although some have used more ecologically valid paradigms. These studies have generally shown statistically significant differences in group analyses in the brain activation that underlies deceptive versus truthful responses. Although there is variability from study to study, the brain regions that are most diagnostic of deception are usually found in lateral and medial portions of the frontal cortex.

Most studies have only examined group data, leaving open the critical issue of how accurate this technology is in single individuals. However, a few of these fMRI studies have estimated the accuracy of the methods in single individuals, using CIT paradigms. For example, Davatzikos and collaborators [15] used high-dimensional nonlinear pattern classification methods (Support Vector Machines) in a group of 22 individuals to discriminate patterns of brain activation associated with producing deceptive and truthful responses in a simple CIT paradigm. The results indicated that predictive accuracy, as measured by determining the proportion of individuals correctly classified in a group not used to train the classifier, was about 88%. Kozel and collaborators [17] used a mock crime scenario involving stealing a ring or a watch. Participants were then scanned with fMRI while they were telling the truth or lying. A group of participants was used to define brain regions that distinguished truthful from deceptive responses and these brain regions were then used to determine whether participants in new group were lying or telling the truth. Results indicated accuracy rates of about 90% in this second group. However, these findings have not yet been replicated by other laboratories, and their generalization to situations relevant to national security and forensic science is unknown.

Countermeasures

Countermeasures are methods used to confound deception detection procedures. For instance, a classic countermeasure is to increase arousal intentionally, and thereby confuse the polygraph – such as by biting one's tongue right after comparison questions (*see* **Deception: Detection of**). Although these kinds of physical countermeasures are not likely to work with brain-based techniques, cognitive countermeasures that rely on inducing specific changes in brain activation are highly problematic. It has been shown that simple countermeasures in which participants generate covert responses to some irrelevant stimuli in CIT paradigms (effectively transforming these stimuli into targets) result in a large P300 being produced by these stimuli [28]. This finding invalidates the logic on which P300 methods are based, and drastically reduces their accuracy [28].

There have been no studies on the effects of countermeasures on fMRI deception detection accuracy rates, but it is very likely that fMRI methods would be affected in similar ways as ERPs by these types of cognitive countermeasures: the brain regions usually found active during deception are also engaged during many other cognitive tasks.

Other Technologies

Some emerging neuroimaging technologies may be used for detecting deception in future applications. One such technology is noninvasive optical imaging [29]. The potential advantages of this technology are its low cost, its portability, and its sensitivity to some of the same signals measured by the most common form of fMRI. Although noninvasive optical imaging has produced reliable results in numerous cognitive tasks [30], there have been no peer-reviewed publications using this technique in deception tasks.

Conclusions

Although brain-based methods for detecting deception are relatively young and show considerable promise, the published accuracy rates so far are not clearly better than those obtained with conventional techniques based on peripheral psychophysiological measures, such as the polygraph. Many researchers are working intensely on different approaches to using brain-based methods; this research focuses on improving paradigms, developing more sensitive brain measures, and devising more sophisticated data analyses. Practical issues also pose important, although not insurmountable, questions about the applicability of brain-based techniques such as fMRI in the field.

Acknowledgments

This material is based partially upon work supported by the National Science Foundation under Grant BCS-0322611. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors wish to thank Dr Peter Rosenfeld for helpful discussion.

References

[1] Bok, S. (1978). *Lying: Moral Choice in Public and Private Life*, Pantheon, New York.

[2] Bond Jr, C.F. & DePaulo, B.M. (2006). Accuracy of deception judgments, *Personality and Social Psychology Review* **10**(3), 214–234.

[3] DePaulo, B.M., Lindsay, J.J., Malone, B.E., Muhlenbruck, L., Charlton, K. & Cooper, H. (2003). Cues to deception, *Psychological Bulletin* **129**(1), 74–118.

[4] National_Research_Council (2003). *The Polygraph and Lie Detection*.

[5] Rosenfeld, J.P., Cantwell, B., Nasman, V.T., Wojdac, V., Ivanov, S. & Mazzeri, L. (1988). A modified, event-related potential-based guilty knowledge test, *The International Journal of Neuroscience* **42**(1–2), 157–161.

[6] Rosenfeld, J.P., Nasman, V.T., Whalen, I., Cantwell, B. & Mazzeri, L. (1987). Late vertex positivity in event-related potentials as a guilty knowledge indicator, *International Journal of Neuroscience* **34**, 125–129.

[7] Farwell, L.A. & Donchin, E. (1991). The truth will out: interrogative polygraphy ("lie detection") with event-related brain potentials, *Psychophysiology* **28**(5), 531–547.

[8] Rosenfeld, J.P. (2005). 'Brain fingerprinting': a critical analysis, *The Scientific Review of Mental Health Practice* **4**(1), 20–37.

[9] Allen, J.J., Iacono, W.G. & Danielson, K.D. (1992). The identification of concealed memories using the event-related potential and implicit behavioral measures: a methodology for prediction in the face of individual differences, *Psychophysiology* **29**(5), 504–522.

[10] Mertens, R., Allen, J., Culp, N. & Crawford, L. (2003). The detection of deception using event-related potentials in a highly realistic mock crime scenario, *Psychophysiology* **4**, S60.

[11] Miyake, Y., Mizutanti, M. & Yamahura, T. (1993). Event related potentials as an indicator of detecting information in field polygraph examinations, *Polygraph* **22**, 131–149.

[12] Abootalebi, V., Moradi, M.H. & Khalilzadeh, M.A. (2006). A comparison of methods for ERP assessment in a P300-based GKT, *International Journal of Psychophysiology* **62**(2), 309–320.

[13] Abe, N., Suzuki, M., Mori, E., Itoh, M. & Fujii, T. (2007). Deceiving others: distinct neural responses of the prefrontal cortex and amygdala in simple fabrication and deception with social interactions, *Journal of Cognitive Neuroscience* **19**(2), 287–295.

[14] Abe, N., Suzuki, M., Tsukiura, T., Mori, E., Yamaguchi, K., Itoh, M. & Fujii, T. (2006). Dissociable roles of prefrontal and anterior cingulate cortices in deception, *Cerebral Cortex* **16**(2), 192–199.

[15] Davatzikos, C., Ruparel, K., Fan, Y., Shen, D.G., Acharyya, M., Loughead, J.W., Gur, R.C. & Langleben, D.D. (2005). Classifying spatial patterns of brain activity with machine learning methods: application to lie detection, *Neuroimage* **28**(3), 663–668.

[16] Ganis, G., Kosslyn, S.M., Stose, S., Thompson, W.L. & Yurgelun-Todd, D.A. (2003). Neural correlates of different types of deception: an fMRI investigation, *Cerebral Cortex* **13**(8), 830–836.

[17] Kozel, F.A., Johnson, K.A., Mu, Q., Grenesko, E.L., Laken, S.J. & George, M.S. (2005). Detecting deception using functional magnetic resonance imaging, *Biological Psychiatry* **58**(8), 605–613.

[18] Kozel, F.A., Padgett, T.M. & George, M.S. (2004). A replication study of the neural correlates of deception, *Behavioral Neuroscience* **118**(4), 852–856.

[19] Kozel, F.A., Revell, L.J., Lorberbaum, J.P., Shastri, A., Elhai, J.D., Horner, M.D., Smith, A., Nahas, Z., Bohning, D.E. & George, M.S. (2004). A pilot study of functional magnetic resonance imaging brain correlates of deception in healthy young men, *The Journal of Neuropsychiatry and Clinical Neurosciences* **16**(3), 295–305.

[20] Langleben, D.D., Loughead, J.W., Bilker, W.B., Ruparel, K., Childress, A.R., Busch, S.I. & Gur, R.C. (2005). Telling truth from lie in individual subjects with fast event-related fMRI, *Human Brain Mapping* **26**(4), 262–272.

[21] Langleben, D.D., Schroeder, L., Maldjian, J.A., Gur, R.C., McDonald, S., Ragland, J.D., O'Brien, C.P. & Childress, A.R. (2002). Brain activity during simulated deception: an event-related functional magnetic resonance study, *Neuroimage* **15**, 727–732.

[22] Lee, T.M.C., Liu, H.-L., Tan, L.-H., Chan, C.C.H., Mahankali, S., Feng, C.-M., Hou, J., Fox, P.T. & Gao, J.-H. (2002). Lie detection by functional magnetic resonance imaging, *Human Brain Mapping* **15**, 157–164.

[23] Mohamed, F.B., Faro, S.H., Gordon, N.J., Platek, S.M., Ahmad, H. & Williams, J.M. (2006). Brain mapping of deception and truth telling about an ecologically valid situation: functional MR imaging and polygraph investigation–initial experience, *Radiology* **238**(2), 679–688.

[24] Nunez, J.M., Casey, B.J., Egner, T., Hare, T. & Hirsch, J. (2005). Intentional false responding shares neural substrates with response conflict and cognitive control, *Neuroimage* **25**(1), 267–277.

[25] Phan, K.L., Magalhaes, A., Ziemlewicz, T.J., Fitzgerald, D.A., Green, C. & Smith, W. (2005). Neural correlates

of telling lies: a functional magnetic resonance imaging study at 4 Tesla, *Academic Radiology* **12**(2), 164–172.
[26] Spence, S.A., Farrow, T.F., Herford, A.E., Wilkinson, I.D., Zheng, Y. & Woodruff, P.W. (2001). Behavioural and functional anatomical correlates of deception in humans, *Neuroreport* **12**(13), 2849–2853.
[27] Spence, S.A., Hunter, M.D., Farrow, T.F., Green, R.D., Leung, D.H., Hughes, C.J. & Ganesan, V. (2004). A cognitive neurobiological account of deception: evidence from functional neuroimaging, *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **359**(1451), 1755–1762.
[28] Rosenfeld, J.P., Soskins, M., Bosh, G. & Ryan, A. (2004). Simple, effective countermeasures to P300-based tests of detection of concealed information, *Psychophysiology* **41**(2), 205–219.
[29] Villringer, A. & Chance, B. (1997). Non-invasive optical spectroscopy and imaging of human brain function, *Trends in Neurosciences* **20**(10), 435–442.
[30] Obrig, H., Wenzel, R., Kohl, M., Horst, S., Wobst, P., Steinbrink, J., Thomas, F. & Villringer, A. (2000). Near-infrared spectroscopy: does it function in functional activation studies of the adult brain? *International Journal of Psychophysiology* **35**(2–3), 125–142.

GIORGIO GANIS AND STEPHEN M. KOSSLYN

Deception: Truth Serum

"Truth serum" is the popular generic term for drugs and other substances used for the purposes of obtaining information from an unwilling subject; typically by a police, intelligence, or military organization seeking information from a prisoner or potential informant. Sometimes also known as *truth drugs*, these substances have also been used in conjunction with psychotherapy, and promoted as tools for determining the "truth" from persons not simply unwilling to cooperate but also ostensibly unable to reveal information on topics for which they have no conscious knowledge.

Origins of Belief in "Truth Serums"

Throughout recorded history, intentionally or by accident, human beings have been using many substances, in addition to alcohol, that affect mental functioning. For example, ergot, a fungus that can grow on rye and other grasses, was used in rituals in ancient Greece. In modern times, psychoactive substances related to lysergic acid diethylamide (LSD) have been isolated from ergot [1]. The long-term evolution of mind-altering experiences has been accompanied by many beliefs about the properties of psychoactive substances including attributions of their effectiveness as aphrodisiacs, potions for altering the behavior of others, and even beliefs about the ability of various substances to compel truthful responses.

Over time, science has given us a better understanding of the actual properties of these substances, provided drugs derived from basic psychoactive materials, and more recently synthesized new psychotropic drugs in the laboratory. Nevertheless, although significant advances in biochemistry have led to understandings of biochemical dynamics, full understanding of the relationship between biochemistry, behavior, and memory has presented formidable challenges [2] – in no small part because insights into behavior are always influenced by human dreams, fears, and vested interests.

Alcohol use has been traced back to the Stone Age [3], and the time of the early Roman Empire was recognized as having truth-inducing properties (*in vino veritas* or "in wine there is truth"). It took until the 1920s, however, for a Texas obstetrician named *Robert House* to give truth induction an *imprimatur* of science. House had been experimenting with the use of scopolamine in childbirth and concluded that when properly administered, it was "impossible to lie" when under its influence. Dr House promoted his discovery as a "humane third degree" that would protect the innocent from being "convicted upon circumstantial evidence" while eliminating "any excuse for . . . brutal third-degree methods" [4].

Law enforcement officials had long been aware of the effects of alcohol, and that withholding drugs from an addict could facilitate confessions, but the idea of an effective scientific truth drug had great appeal not only for the police but the general public. Soon the Chicago police, in particular, were using scopolamine with reported success [5, 6]. Although scientific research on the nature of memory had been making steady progress into the first decades of the twentieth century, the concept of a truth drug required that emerging concepts of memory be supplanted by a belief that memory was actually akin to a vast video recording, storing every detail of a person's life, permanently and without distortion, even if the

individual had no conscious knowledge of the thus-recorded facts [7].

Misconceptions about the nature of memory would surround the "truth" technologies until finally challenged at the end of the twentieth century. Erroneous beliefs persisted, however, and fostered the discovery of many successor drugs, including barbiturates, that were advanced as "truth" drugs by Horsley in the 1940s [8].

Drugs Promoted as "Truth Serums"

Alcohol

Not only is there a long-standing belief that alcohol can act as a truth drug but also its use for this purpose has persisted to the present day. The Russians, in particular, have an extensive history of using grain alcohol (ethanol) in large doses during interrogations, including interrogations of their own agents to determine loyalty [9] (*see* **Alcohol: Behavioral and Medical Effects**).

Scopolamine

Scopolamine is a botanical alkaloid of the Solanaceae, or nightshade family [10]. The use of scopolamine as a mind-altering substance (producing feeling of flying, hallucinations, and other behaviors that occasionally caused users to be identified as witches) has been traced back to 3000 BC [11]. It was the refinement of the drug in Germany in the 1890s that led the German doctor Von Steinbuchtel of Graz to develop a procedure for the use of intravenous scopolamine to induce "twilight sleep" (*Daemmerschlaf*) in women going into labor. While it was believed that the procedure would prevent pain, it was later discovered that the drug simply induced an amnesia that eliminated memory of the pain and the birth itself.

Moreover, scopolamine stimulated graphic erotic outcries and statements that physicians often assumed to be truthful, contributing to the belief that scopolamine was a "truth" drug (nurses often objected to the use of the drug for what they considered to be a source of amusement for male physicians). In addition, the procedure often led to subsequent psychotic-like symptoms and posttraumatic stresslike memories in the mothers. However, it was the discovery that the drug could cause depressed respiration, could be introduced to the neonate through breast milk; and, in rare instances, could cause death of the mother, the child, or both, that led to the eventual discontinuation of the procedure.

Scopolamine is an antimuscarinic agent, which is a competitive inhibitor of acetylcholine at some synapses, like those in the ciliary body, the iris, and secretory (salivary, bronchial, and sweat) glands. The mechanism of action for its extensive use in the prevention and treatment of motion sickness is unclear, but presumably involves direct anticholinergic action on the vomiting center. The Nazi doctor Josef Mengele reportedly experimented with scopolamine as an interrogation drug. Scopolamine's effectiveness as "truth serum" was investigated by various intelligence agencies in the 1950s, including the CIA as part of project MK ULTRA [12]. It continued to be used in this way into the 1970s.

Barbiturates

Barbiturates are sedatives and hypnotics derived from barbituric acid; they are controlled substances due to their high potential for abuse. Barbituric acid was first discovered by the German chemist Adolf von Baeyer in 1864, who combined urea (an animal waste product) with malonic acid (derived from the acid of apples). Barbiturates probably obtained their name from the fact that they were discovered on Saint Barbara Day. The generic names of barbiturates generally end in *al*. It is the class of barbiturate derivatives known as *thiobarbiturates*, which contain a sulfur molecule in place of one of the oxygen, that were found to have properties that led to their use as truth drugs.

The principal mechanism of action of barbiturates is believed to be their affinity for the gamma-aminobutyric acid $(GABA)_A$ receptor, the principal inhibitory neurotransmitter in the human central nervous system (CNS). In addition to this, GABA-ergic effect, barbiturates also block a subtype of glutamate: the principal excitatory neurotransmitter.

Sodium pentothal (thiopental sodium for injection) [13] is an ultrashort acting barbiturate, meaning that sedation lasts only for a few minutes. Sodium pentothal slows the heart rate, lowers blood pressure, and reduces CNS activity. Sedation occurs in < 1 min after injection. Sodium pentothal is also used as a general anesthetic for procedures of short duration, for induction of anesthesia given before other anesthetic drugs, as a supplement to regional anesthesia (such as a spinal block), and as an anticonvulsive.

Sodium amytal (amobarbital) [14] is an intermediate-acting barbiturate. Sedation occurs in 1 h or longer and lasts for 10–12 h. Sodium amytal depresses the central nervous system. It has been used as a sedative, hypnotic, an anticonvulsive, and has been used as a "truth" drug in what were termed *amytal interviews*.

Benzodiazepines

Benzodiazepines, often abbreviated to "benzos", are a class of psychoactive drugs with varying hypnotic, sedative, anxiolytic, anticonvulsant, muscle relaxant and amnesic properties, which are mediated by reducing CNS activity. Benzodiazepines are useful in treating anxiety, insomnia, agitation, seizures, muscle spasms, and alcohol withdrawal. The generic names of drugs in this class typically end with *pam*.

The use of barbiturates has largely been supplanted by benzodiazepines because of the increased safety of the latter. Barbiturates, benzodiazepines, and alcohol all bind to the $GABA_A$ receptor, but the barbiturates bind with the highest affinity with longer receptor binding half-lives. Barbiturates can activate GABA receptors in the absence of the GABA molecule, whereas benzodiazepines require the presence of GABA to have an effect. This explains why an overdose of barbiturates may be lethal, but overdoses of benzodiazepines alone are typically not lethal.

Temazepam (Restoril) [15] is very fast acting, and the benzodiazepine most frequently used as a "new generation truth serum".

Barbiturates, benzodiazepines, and scopolamine must be administered intravenously and with very careful titration, in order to maintain the dose at a sufficient level to induce the trancelike state required for interview, yet avoid death or other serious side effects. Ethanol must also be administered in very large quantities, and excessive amounts can cause death. Contrary to accounts in popular fiction, it is virtually impossible to follow a truth drug protocol surreptitiously or for a person to be under the influence of a "truth" drug without being aware that such a drug is being employed. Moreover, each of the above substances interferes significantly with judgment and higher cognitive function, with a corresponding impact on the potential for recovering verifiably accurate information.

Other Substances

A number of other substances have been touted as "truth" drugs but no reliable documentation exists to support claims of their effectiveness. The Soviets have boasted of various secret substances, including one codenamed SP-17, but there is no proof that this is anything other than vodka.

Although there is no documentation of oxytocin (sold as an injectable drug under the brand names *Pitocin* and *Syntocinon*) [16] ever being used in an attempt to elicit truthful answers to questions, this nonapeptide found in pituitary extracts from mammals has generated discussion in intelligence circles. From the Greek "quick birth", oxytocin acts as a neurotransmitter in the brain, and in women, it is released in large quantities after distension of the cervix and vagina during labor, and after stimulation of the nipples, facilitating birth and breast-feeding. In humans, oxytocin is thought to be released during hugging, touching, and orgasm in both sexes. In the brain, oxytocin is involved in social recognition and bonding, and may be involved in the formation of trust between humans [17] and in generosity [18]. Moreover, oxytocin increases social behaviors in autistic adults [19]. Although an interrogation setting would likely offset any psychodynamic of this drug, the biochemistry suggests possibilities for interview effectiveness and a rethinking of the use of drugs in intelligence operations.

The Nature of Truth and Memory

Even as the notion of "truth serum" was capturing the fancy of police and the public in the 1920s and 1930s, scientists were publishing research findings on the nature of memory that contradicted the assumptions in the foundation of truth technologies (*see* **Memory: Reconstructive**). Memories are actually edited and consolidated compilations of what may have even initially been flawed or only partial fragments of experience. Memory is not a vast video recording that stores every detail of a person's life. Perceptions are often very short term and fleeting, such as one's images of the terrain as one glances down when he or she is running over uneven ground – these serve a purpose and are then discarded. Typically, memories of the day may be detailed but are eventually sorted out and summarized. The images of what is seen on a commute to work in the morning are processed,

probably during sleep the same night, into simply a recollection that one went to work the previous day. Over time, "memories" of commuting to a previous job will be summarized into recognition of only the fact that one took a particular route to work on most days.

Subsequent additions to these compilations may be selected because they fit well with earlier impressions that were, in turn, shaded by subtle distortions and defenses. The brain's management of memory can be compared with management of a filing system – things that seem redundant are eventually discarded, recurrent information is summarized, and occasionally things are simply misfiled. Digging back into memory to reconstruct a reliable account of historical truth resembles the task faced by a reporter when preparing a story of a train accident from multiple fragments of conflicting eyewitness accounts.

Courtroom witnesses can be compared to people who have been involved in such a train accident. Even if they are fortunate enough to avoid the influences of personal bias and motive, they can find themselves attempting to reconcile perceptions of events they did not anticipate, having viewed the most critical aspects of the event from unique and often distorted vantage points, observed with less than fully focused attention. Like the task of the reporter, memory reconstruction requires weighing sources and credibility, and learning to live with irreconcilable inconsistencies.

Forensic experts therefore find it helpful to distinguish human memories, even when highly credible and consistent, as narrative truth rather than historical truth [20]. Such memories may be very vivid and real to the people experiencing them, and are often the closest available approximation to what the parties obviously accept as "truth", but they can never be the same as historical truth, any more than a movie about a train accident can be the same as the actual original accident.

Eyewitness observations are obviously of great interest to the criminal justice system when a potential witness is thought to have information that may relate to an open case (*see* **Eyewitness Testimony**). Witnesses who have seen perpetrators, getaway cars, or details relating to a crime are often interviewed yet remain unable to recall key details. Occasionally, such witnesses were asked to submit to hypnosis (*see* **Hypnosis and Memory**) or take a "truth serum" to better facilitate recall. Popular topics of movies and crime novels, these processes are often glamorized and easily misunderstood.

Both hypnosis and drug interviews lower inhibitions and minimize conscious editing of beliefs, but neither can facilitate recovery of information that the subject never possessed nor can they conclusively distinguish a conviction that the subject believes to be true from actual historical truth. There are, however, instances in which these techniques can be useful, and one way to think about how they do work is through a computer analogy. When you delete a file from the disk on your personal computer, the file is seldom immediately erased, but is instead deleted from the indexing capability that allows the computer to access and retrieve the data. The computer may eventually run out of storage space and write over some or all of the actual deleted information, but in the meantime, the original data simply sits, inaccessibly, on the disk itself.

This process is similar to what the brain does as it reviews and processes daily perceptions; it selects what it considers to be important and releases the rest to simplify processing and maximize the use of storage capacity. If something is "misfiled" it may be retained, but is "forgotten" unless or until there is a reason to recall the information with which it was actually stored.

However, while this compilation process is taking place, unprocessed details remain in the brain, just as "deleted" files can remain on a computer disk. Most people who have inadvertently erased an important file have had occasion to learn about computer utilities that allow you to "recover" data. These utilities do not work like application programs; they instead allow direct access to files no longer indexed for logical access, "forgotten" but as-yet-unerased data files. Hypnosis and drugs can serve a similar function with the human mind, allowing the bypass of normal mental processing logic, and possibly the recovery of images that have been misfiled or are awaiting deletion.

Suppose an investigator is seeking, for example, the license plate of a getaway car. This information could have been noticed in passing by someone on the street, but ignored and subsequently set aside by the mind as unimportant to that person. No longer accessible, the number sequence of the license plate may, nonetheless, sit in memory awaiting final processing and deletion. Hypnosis or medication may

provide a way to search through fragments of images and find what could be the elusive plate number.

Using hypnosis or medication does *not*, however, provide any indication of the historical truth of the information. The human mind is filled with vague perceptions, questionable beliefs, fantasies, images from favorite old movies, snatches of tales told around childhood campfires, and countless other vivid information of dubious authenticity. It is the job of the mind's central indexing system to keep track of all of this information and remember what is real and what is not. When this critical *source monitoring* is suspended, as it is in hypnosis or by "truth" drugs, it becomes impossible when attempting recollection to distinguish truth from fiction. This may not be a problem for an investigator who can check plate numbers against possible suspects and gather additional sources of physical evidence to verify and build a case, but to think of such interview results as "truth" is without scientific basis.

An even greater danger, however, lies in the potential for altering human memory once the mental safeguards designed to distinguish information by source and reliability – source monitoring – have been lowered. The application programs on a computer, for example a word processor, keep track of when and why files were created. Recovery utilities bypass this electronic form of source monitoring. Just as a computer user can then accidentally misuse a file recovery utility to misinterpret or corrupt files on a hard drive, an investigator can use hypnosis or drugs to interpret or suggest images and beliefs that never originated in actual human experience.

This "editing" of perceived reality does not even require special procedures or prescription drugs. Many over-the-counter and illicit drugs, including alcohol, have, among their properties, an ability to weaken that portion of the brain that monitors the origin and credibility of images and beliefs. Moreover, this reduction in source-monitoring efficacy, which increases susceptibility to influence, need not require drugs or external intervention of any type.

Alterations in belief and memory are influenced by both the nature and context of the messages to which people are subjected, as well as by individual characteristics and motivations. The former, which includes what may be also termed *propaganda and coercion*, can be understood collectively as forms of persuasion.

Persuasion

The experiences of American prisoners during the Korean War generated international attention to what had become popularly known as *brainwashing*. Following that War, extensive studies were conducted with former prisoners. The results, published in *Psychiatry* in 1956 [21], helped to shape the direction of research on social influence, and lay the foundation for what has since been confirmed in what we now know about the operation of the human brain.

The approach used by the Communists was, according to the report, "to gain complete control over those parts of the physical and social environment which sustain attitudes, beliefs, and values, breaking down interactions and emotional bonds which support the old beliefs and values and building up new interactions which will increase the probability of the adoption of new beliefs and values. If the only contacts a person is permitted are with persons who *unanimously* have beliefs different from his own, it is very likely that he will find at least some of them with whom, because of growing emotional bonds, he will identify and whose beliefs he will subsequently adopt". (p. 333–334).

"Taken singly," the study concluded, "there is nothing new or terrifying about these techniques... [some even]...have their counterparts in education and in psychiatry...[or are widely used]...by the *police* [emphasis added], by newspaper reporters, and by others interested in aggressively eliciting information... Forced confessions and self-criticism have been widely used techniques in religious movements as a basis for conversion...the only novelty... [here]... was the attempt *to use a combination of all these techniques and to apply them simultaneously* in order to gain complete control over significant portions of the physical and social environment..." (p. 334).

The role of persuasion or coercion in any given human decision is in part a matter of perspective, and we must always carefully weigh questions of free choice and duress [22]. Nevertheless, the impacts of persuasion and duress are real [23], and the concerns that they raise are reflected in the structure of our legal systems and the holdings of the courts [24] (*see* **Confessions: Evidentiary Reliability of**; **Interrogation**; **Interrogative Suggestibility**; **Capacity to Waive Miranda Rights**).

Suggestibility

Suggestibility is a function of both individual characteristics and personal hopes and fears (*see* **Eyewitness: Suggestibility of**; **Children: Suggestibility of**; **Children: as Witnesses**).

In June 1962, employees of a Southern textile mill were stricken with a mysterious illness. Investigators attributed the illness to the bite of insects that were brought into the factory in a shipment of cloth that had arrived from England. The insect attacked the skin, the bite leaving a wound similar to the bite of a gnat. In about 20 min, the victim was stricken with severe nausea and a rash that covered much of the body. Within a matter of days the resulting epidemic forced the closing of the factory.

Yet despite the consistent and specific symptomatology, so impressive that a vast array of professional talent was mobilized to cope with the problem, and the strength of the victim's belief in the reality of the "June bug", no acceptable medical explanation could be found. Given the demands of a peak production season, the personal needs of employees, and the fact that no other complaints could be found to lodge against a relatively new and progressive factory, the existence of an "epidemic" at that time and place was the ideal answer to the wishes of its victims [25]. If the wish can be the father of the dream, that dream is but a step away from belief. And belief, even in a factitious illness as seen in Munchausen syndrome and induced in children by their parents in Munchausen by proxy [26–30], can be powerful enough to produce the required physical symptoms. Powerful enough in the case of Voodoo, for example, to result even in death [31, 32].

The phenomenon is hardly new, nor is it confined to our culture. George Rosen, in his classic *Madness in Society* [33], traces similar examples back to ancient times. Madness is, moreover, the correct term. While such behavior is hardly productive – wiping out people's savings, threatening their health and even their lives, and often leading those caught up in its frenzy to neglect responsibilities and violate the law – it is neither the product of mental disease nor within any definition of legal insanity. It is simply people acting as though their fantasies were reality. And nothing fans the flames of madness like having the people around you, those to whom you usually turn to test reality, share your belief in the reality of your own dreams and delusions.

The ability to keep track of the source of perceptions and beliefs – source monitoring – like most individual traits, varies widely throughout the human population. In the early 1970s, Herbert Spiegel confirmed the variability of response to hypnotic suggestion, and identified a high level of susceptibility that occurs naturally in 1 out of every 10–20 people [34, 35].

Individuals in this grade 5 group frequently demonstrate spontaneous trance logic, the capacity to act as if unaware of the most extreme logical inconsistencies, and to suspend normal levels of critical judgment. Such individuals are particularly sensitive to the views and preferences of those around them, and outstanding in their ability to comply with the wishes and beliefs common to their environment. People with this characteristic can act as if hypnotized even when every effort is made to avoid hypnotizing them, and when in their own spontaneous trance states they are highly susceptible to enduring distortions of memory and the confusion of fantasy with reality.

Implanted Memories

It was not until the 1990s that events forced a reexamination of the assumptions implicit in the belief that "truth" drugs and procedures could yield scientifically reliable results. For most women, the 1980s had been a time of fruition in a decades-long campaign for expanding opportunities. At its root, however, many activists saw the women's movement as similar to the civil rights struggle. A faction of its constituency even came to envy the privilege, attention, and status accorded the dependent and the oppressed with whom they identified.

Throughout any culture, there are unhappy people searching for an explanation for their unhappiness [36]. There is no seemingly safer refuge than fantasy, and early in his career Freud was struck by the number of grown women who recounted childhood sexual encounters with parents and other adults [37, 38]. Freud eventually concluded that these stories had to be factually untrue, if only on the weight of the evidence and their internal inconsistency [39], but the accounts were understandable by their ability to interweave themes of sexual desire, seduction, dependency, and exploitation into a warm blend of desirability and justification for special dispensation.

As the women's movement progressed, the desire to satisfy this combination of fantasies found expression in the vast popularity of fictional works that were increasingly presented as true accounts; for example, *The Exorcist* [40], *Sybil* [41] *Michelle Remembers* [42], and *Communion* [43]. All of this may have come to nothing more than harmless diversion had it not been for the convergence of three larger social forces. The first was the women's movement that made it increasingly difficult to dismiss women's concerns as delusional nonsense. The second was the shocking discovery that all forms of sexual abuse *really were* much more common and destructive than previously believed. Finally, in the 1960s, psychotherapy itself had undergone a dramatic transformation that had brought in significant numbers of "alternative" clinicians – many of whom dedicated their practices to meeting the felt needs of special client populations.

Much of this new professional faction worked very hard to appear independent of previous "male" biases. Cloaking their activities with the respectability of sympathetic, well-credentialed psychiatrists [44], and evoking images of the Nazi war crime cover-ups, popular writers with few if any clinical credentials began to pour out mass-marketed books like *The Courage to Heal* [45] and *Breaking the Circle of Satanic Ritual Abuse* [46].

These books, and the individual and group "therapy" sessions, which they spawned, swept the country. Feminist and New Age bookstores could not keep the wave of materials on their shelves, and the "abuse therapy" business boomed – funded in large part by insurance dollars [47]. But once the movement ignited, it took on a life of its own. Before long, psychologist/lawyer teams sprang up to avenge the "victims", and the most popular adjunct treatment for the abused patient became litigation. Kits catering to those thirsty to sue and thus have the forum, which they were told would offer their essential therapeutic opportunity for confrontation and justice, joined the pop psychology books that packed bookstore shelves [48].

Starting with a trickle in the mid-1980s, the courts were soon inundated by a torrent of suits that by the 1990s would grow in number to over 800, with at least 17 000 threatened or in various stages of preparation [49]. From the beginning, it was the courts that stemmed the tide, using the statute of limitations under the discovery rule to prevent cases alleging old harms that no one can adequately defend against [50]. Rarely if ever did any plaintiff assert any corroborating evidence, and all of the cases seemed to boil down to the word of one family member against another. Nevertheless, the mass media quickly took up the quintessential feminist cause – how could the law presume that a woman was wrong when she accuses a man of the most unspeakable of exploitive crimes. In rapid succession, more than half of the states in the nation revised their statutes of limitation to facilitate suits alleging recovered memories of sexual abuse [51, 52].

Many of the allegations were outrageous, involving complex cults and accomplices from outer space [53, 54]. Although the accusers ranged across all genders and backgrounds, the vast majority were not stereotypical victims but white women in their 20s to mid-40s from American Gothic families who had been well educated and had found moderate success in middle-level professions. Most had few signs of serious disturbance until they entered therapy. In fact, this influx of persons seeking help required dusting off the previously discredited diagnosis of "multiple personality disorder" [55], transforming it into a condition characterized as "extreme and dramatic, involving significant amnesia ... [with] ... sudden loss of memory, change from a sad, dependent, and helpless personality state to an angry, demanding hostile one in seconds ... involving alterations in identity, memory, consciousness, and somatic function ... [that] ... occurs as a defense during and after trauma, a means of maintaining mental control just as physical control is lost" (for source of quotations *see* **Dissociative Disorders**).

"Dissociation may especially occur when role conflict with important family figures on whom one depends is involved, a condition that has been described as 'betrayal trauma' ... Keeping conflicting views of the same person in mind creates considerable tension and confusion, and complicates memory storage and retrieval." The mental health professions were divided by the controversy that resulted, clinicians on the one hand contended childhood sexual abuse opened up a vast new population in need of care, and researchers on the other hand pointing to a century of science that refuted the basic assumptions underlying the clinicians' assertions [56, 57, 58] (*see* **Memory: Repressed**).

Childhood sexual abuse therapists were convinced that the symptoms seen in their new patients were

proof of early sexual trauma, the memories of which had been repressed. If they could simply "recover" those memories, they would have the smoking gun needed to initiate the litigation that would "cure" the psychiatric condition [59, 60]. Amytal and hypnosis were employed, and, not surprisingly, the desired repressed memories were "recovered" [61]. These therapists solution reached back to Dr House and the "truth serum" that had captured the fancy of the public early in the century.

Concerned behavioral and forensic scientists came together to document the truth [62]. Vilified by the opposition as a refuge for sex perverts [63], their organization went to work taking on what had become a worldwide problem. Working with the accused, they lobbied to make their voice heard in the legislatures and turn back the popular tide of statute changes that had effectively targeted accused offenders in a way normally not used against any other class of defendants. Secondly, they learned to use the litigation process and in doing so, revolutionized third party standing by opening the door to suits by falsely accused family members against their accuser's psychotherapists – this despite a history of legal tradition that held that only patients or those paying for a patient's care could sue for malpractice [64–66]. As the opposition began to crumble, some 300 accusers recanted with 60–70% of these also lining up to sue their own former therapists [67–72]. Finally, the scientists attacked the credibility of the expert testimony used to support "recovered memory" litigation against family members in a systematic effort by experts to protect the integrity of behavioral science expert testimony [73–78].

Faced by objective data from respected scientists, a new perspective came to dominate the media. In 1993, Lawrence Wright published his account of the Paul Ingram case in the *New Yorker* [79], and in April 1995 *Frontline* ran a two-part special that brought to an end any momentum that may have remained in the recovered memory movement [80]. The tide of accuser suits crested in 1994 [81, 82], and turned to recede in the wake of the landmark case of *Ramona v. Ramona* [84].

What Kind of "Truth" Can "Truth Serums" Elicit

"Truth serums", of course, are neither serums nor scientific tools capable of reliably eliciting historical truth. Hypnosis and the various drugs touted as truth serums do generally lower inhibitions and inhibit the cognitive functioning needed to distinguish stored recollections from implanted beliefs, but this can also be said of the persuasive techniques that do not rely upon drugs or traditional hypnosis.

Interrogation employing any of these techniques, or simply providing an alcoholic drink, is likely to make the subject more talkative. If such actions result in any information, the accuracy of that information is likely to be greater if it involves recent events still in short-term memory. However, even under the best of circumstances, recall is always limited by the accuracy of the information stored, the quality of the storage, and the precision of the retrieval.

Lowering inhibition and reducing cognitive control encourage a person to be more talkative, but loss of inhibitions is not the same as loss of self-control. Subjects can still lie, fantasize, and even provide deliberately misleading information, although they can also be more easily manipulated by an interviewer's suggestions and cues into telling falsehoods. Although it may be possible to troll for miscellaneous facts such as a license plate number, such information would be of dubious value unless it could be independently verified. In short, the procedures used for most of the twentieth century to obtain secrets and information ostensibly beyond the conscious awareness of human subjects are inconsistent with established scientific findings and unsuitable for the purposes for which they have been used [84].

Legal Issues and Admissibility

Four major considerations govern the admissibility of all of the "truth technologies", and, in turn, any decision to employ the procedures in the first place.

Finder of Fact

Just who is or is not telling the truth in a legal situation in which the facts at issue are material to the outcome of the case is a question of fact usually left to the trier of fact – the jury or in some instances a judge. The judiciary is very reluctant to take such decisions away from the trier of fact, and therefore does not permit expert testimony to substitute for a party taking the stand for direct assessment by a jury. Similarly, although science has clearly demonstrated that eyewitnesses may provide the most unreliable

forms of evidence, we leave their evaluation to the jury and even make it difficult (depending upon jurisdiction) to provide expert testimony on the reliability of eyewitness testimony.

For the same reason, the courts generally do not allow admission of the results of lie-detector tests or other new high-technology deception technologies [85]. Not only are they of questionable scientific accuracy but they are also limited by the same human memory constraints as hypnosis and truth serums (*see* **Deception: Detection of**; **Deception: Detection of and Brain Imaging**).

"Truth serum" and hypnosis evidence also impinge upon the prerogatives of the finder of fact, and for this reason alone have been held inadmissible as legal evidence.

Self-Incrimination

Although the courts have never addressed the issue directly, "truth serum", as well as hypnotic evidence, is presumed to be coercive and therefore its results, as they relate to a criminal defendant, are inadmissible against that defendant [86].

Torture

Recently, the scope and intent of Miranda have been strengthened through the legal argument that the use of "truth serum" is torture, and therefore even if its results were not precluded under Miranda itself, it would be precluded as torture [87].

Court Holdings on Admissibility

Although the U S Supreme Court has never addressed the admissibility of "truth serum" evidence directly, the 1963 Chicago *Townsend* case [88] is frequently cited as binding precedent. In that case, the State did not dispute petitioner's assertion that the use of a "truth serum" to obtain his confession was unconstitutional. State courts have also consistently held that "truth serum" evidence is inadmissible, typically because it fails to meet scientific evidence standards [89]. Moreover, even if a suspect did not confess under the influence of a drug but simply provided a single scrap of information that ultimately led to evidence sufficient to convict, that entire chain of evidence could be excluded as "fruit of a poisonous tree".

Tradecraft and Clinical Practice

Obviously, substances that have been used as "truth serums" can have other legitimate medical applications, and nothing in the science of memory or truth precludes their use for those additional purposes. However, some gray areas that touch upon truth warrant mention. For example, one of the authors of this article was treating a patient who had suspected hysterical paralysis. Because this condition involves higher cognitive processes that may nevertheless be unconsciously preventing a physical act otherwise unimpeded by a physical anomaly, an intravenous barbiturate was administered to bypass higher cognition; whereupon the patient no longer suffered paralysis. In a sense, this was a test of truth, but here the truth was a function of altered behavior rather than reported memory [90].

The use of "truth" technologies, although attempted in the past by virtually all major intelligence agencies, is currently disavowed worldwide. The United Nations considers the use of "truth" drugs to be physical abuse and, therefore, a form of torture [91, 92]. Moreover, intelligence and interrogative tradecraft has generally come to a consensus that traditional interview techniques, without drug or high-technology supplementation, are the most effective way to gain information. As one operative observed, when one person acts as the "good cop" and creates a bond with the prisoner, the process probably makes use of the brain's own oxytocin. As the interrogator observed, recent science has just taught us "the neurophysiology . . ." [93].

The issue was revisited in 2002, when some authorities, including former Central Intelligence Agency and FBI chief William Webster, became frustrated by the lack of forthcoming information from suspected al-Qaeda and Taliban members held at the U.S. prison in Guantanamo Bay. Some, including Webster, advocated administering "narcoanalysis" drugs to uncooperative captives [94]. Because terrorists are known for their preoccupation with secrecy and loyalty, this suggestion alone may have served a propaganda objective. Nevertheless, considerable thought has been devoted to the "ticking bomb" scenario [95].

This is the situation in which a country's authorities know that terrorists have set in motion a weapon of mass destruction. One of the terrorists in custody will not talk. There is no real concern about

the conviction of this one terrorist or about the admissibility of any evidence that he may provide – only finding and defusing the bomb. Though the authorities know that statements made under the influence of any "truth" substance may not be true, they consider any information to be better than nothing and are prepared to follow any lead. Should they use the drug?

Conclusions

"Truth" substances are one member of a class of techniques and technologies that maximize suggestibility and facilitate persuasion. Although these all may make a subject or suspect more talkative or tractable, there is no scientific basis for the belief that statements made under their influence are necessarily true. Truth and memory are complex topics, and involve fragile mental processes. There is also no basis for believing that memories of distant events reflect historical truth or can be "recovered" from the subconscious. Although "truth" drugs can lower inhibition and reduce cognitive control, these effects are not the same as loss of self-control. Subjects can still lie, fantasize, and even provide deliberately misleading information, although they can also be more easily manipulated into telling falsehoods by an interviewer's suggestions and cues. The use of these substances is considered torture under international law, and evidence obtained through or as a result of their use is legally inadmissible in virtually all courts of law.

References

[1] Webster, P., Perinne, D.M. & Ruck, C.A.P. (2000). Mixing the *Kykeon*, *Eleusis: Journal of Psychoactive Plants and Compounds* New Series (4).

[2] Meyer, J.S. & Quanzer, L.F. (2004). *Psychopharmacology: Drugs, the Brain and Behavior*, Sinauer Associates, Sunderland.

[3] Hanson, D.J. (1995). *Preventing Alcohol Abuse: Alcohol, Culture and Control*, Praeger, Wesport.

[4] Rejali, D.M. (2007). *Torture and Democracy*, Princeton University Press, Princeton, pp. 386–387.

[5] Sheedy, C.E. (1959). Narcointerrogation of a criminal suspect, *The Journal of Criminal Law, Criminology, and Police Science* **50**(2), 118–123.

[6] Geis, G. (1959). In scopolamine veritas. The early history of drug-induced statements, *The Journal of Criminal Law, Criminology, and Police Science* **50**(4), 347–357.

[7] Historian Alison Winter (2005) has studied this topic extensively. Her Her "The making of 'truth serum", *Bulletin of the History of Medicine* **79**, 500–503; reports events from the 1920s and 1930sc and her work covering later time periods was summarized by Carrie, G., in "The facts about truth serum", *University of Chicago Magazine* **95**(6), (2003).

[8] Horsley, J.S. (1943). *Narco-Analysis. A New Technique in Short-cut Psychotherapy: A Comparison with Other Methods and Notes on the Barbiturates*, Oxford University Press, New York.

[9] Shvets, Y.B. (1994). *Washington Station: My Life as a KGB Spy in America*, Simon & Schuster, New York.

[10] See http://www.botgard.ucla.edu/html/botanytextbooks/economicbotany/Atropa/index.html. Because the Solanaceae contain a number of infamous poisons, members of the family including tomatoes, white potatoes and eggplant were viewed with great suspicion when proposed as foods for human consumption.

[11] For an overview see http://www.ucalgary.ca/~bali/chem353/project/00final/00final.htm#conclusions. (accessed Oct 2008).

[12] U.S. Congress (1976). The Select Committee to Study Governmental Operations with Respect to Intelligence Activities, Foreign and Military Intelligence (Church Committee report), report no. 94–755, 94th Congress, 2nd Session, GPO, Washington, DC, p. 392. According to the CIA, the project [MKULTRA] was decreased significantly each budget year until its complete termination in the late 1960s.

[13] See http://www.rxlist.com/cgi/generic/thiopental_ids.htm. (accessed Oct 2008).

[14] See http://www.netdoctor.co.uk/medicines/100002406.html. Official prescribing information on sodium amytal is now difficult to obtain in the United States.

[15] See http://www.rxlist.com/cgi/generic/temaz.htm and http://www.rxmed.com/b.main/b2.pharmaceutical/b2.1.monographs/CPS-%20Monographs/CPS-%20(General%20Monographs-%20R)/RESTORIL.html.

[16] See http://www.rxlist.com/cgi/generic/oxytocin.htm (for Pitocin), and http://www.drugs.com/pro/syntocinon.html (for syntocinon) as well as http://www.vivo.colostate.edu/hbooks/pathphys/endocrine/hypopit/oxytocin.html. (accessed Oct 2008).

[17] Kosfeld, M., Kosfeld1, M., Heinrichs, M., Zak, P.J., Fischbacher, U. & Fehr, E. (2005). Oxytocin increases trust in humans, *Nature* **435**, 673–676.

[18] Stanton, A.A. (2007). Neural substrates of decision-making in economic games, *Scientific Journals International* **1**(1), 1–64.

[19] Hollander, E., Bartz, J. & Chaplin, W. (2007). Oxytocin increases retention of social cognition in autism, *Biological Psychiatry* **61**(4), 498–503.

[20] Spence, D.P. (1984). *Narrative Truth and Historical Truth*, W.W. Norton, New York.

[21] Reprinted with minor revisions as Schein, E.H. (1958). The Chinese indoctrination program for prisoners of war: a study of attempted "brainwashing", in *Readings in*

Social Psychology, E.E. Maccoby, T.M. Newcomb & E.L. Hartley, eds, Holt, Rinehart & Winston, New York, pp. 311–334.

[22] Hindery, R. (2001). *Indoctrination and Self-deception or Free and Critical Thought?* Edwin Mellen Press, Lewiston.

[23] For comments on this topic by a former president of the American Psychological Association, see http://www.apa.org/monitor/nov02/pc.html.

[24] Escobedo v. Illinois, 378 U.S. 478 (1964).

[25] Kerckhoff, A.C. & Back, K.W. (1968). *The June Bug: A Study in Hysterical Contagion*, Appleton-Century-Crofts, New York.

[26] Feldman, M.D. & Ford, C.V. (1994). *Patient or Pretender: Inside the Strange World of Factitious Disorders*, Wiley, New York.

[27] Janofsky, J.S. (1994). The Munchausen syndrome in civil forensic psychiatry, *Bulletin of the American Academy of Psychiatry and the Law* **22**(4), 489–497.

[28] Barker, L.H. & Howell, R.J. (1994). Munchausen syndrome by proxy in false allegations of child sexual abuse: legal implications, *Bulletin of the American Academy of Psychiatry and the Law* **22**(4), 499–510.

[29] Goldman, L.H. & Yorker, B.C. (1999). Mommie dearest? Prosecuting cases of Munchausen syndrome *Criminal Justice* **13**(4), 26–33.

[30] Feldman, M.D. (2000). Factitious disorders *Audio-Digest Psychiatry* **29**(10).

[31] Cannon, W.B. (1942). Voodoo death, *American Anthropology* **44**, 2.

[32] Cannon, W.B. (1963). *Bodily Changes in Pain, Hunger, Fear, and Rage*, Harper and Row, New York.

[33] Rosen, G. (1968). *Madness in Society: Chapters in the Historical Sociology of Mental Illness*, University of Chicago Press, Chicago.

[34] Spiegel, H. (1974). The grade five syndrome: the highly hypnotizable person, *International Journal of Clinical and Experimental Hypnosis* **22**, 303–319.

[35] Lotto, D. (1994). On witches and witch hunts: ritual and satanic cult abuse, *Journal of Psychohistory* **21**(4), 373–396.

[36] One theory which has been advanced to explain this observation is presented by Hudson, J.I. & Pope, Jr, H.G. (1990). Affective spectrum disorder: does antidepressant response identify a family of disorders with a common pathophysiology? *American Journal of Psychiatry* **147**(5), 552–564.

[37] Freud presented these findings in 1896 in his famous paper (1953–1974). The etiology of hysteria which appears in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Hogarth Press, London.

[38] For an excellent discussion of recovered memories in the context of evolving psychoanalytic theory, see Michels, R. (1997). Psychodynamic psychotherapy in American Psychiatry and Health Care, *Audio-Digest Psychiatry* **26**(15).

[39] An account of the unfolding of this revelation can be found in Masson, J.M. (1984). *The Assault on Truth* HarperCollins, New York, p. 131.

[40] Blatty, W.P. (1971). *The Exorcist*, HarperCollins, New York.

[41] Schreiber, F.R. (1973, 1987). *Sybil*, Warner Books, New York.

[42] Smith, M. & Pazder, L. (1980). *Michelle Remembers*, Pocket Books, New York.

[43] Strieber, W. (1987). *Communion: A True Story*, Avon Books, New York.

[44] Harvard psychiatrist Herman, Judith, a colleague of respected researcher Bessel van de Kolk, became the spokesperson for this group. See generally, Herman, J.L. (1995). Crime and memory, *Bulletin of the American Academy of Psychiatry and the Law* **23**, 5–17.

[45] Bass, E. & Davis, L. (1988). *The Courage to Heal: A Guide to Women Survivors of Child Sexual Abuse*, HarperCollins, New York.

[46] Ryder, D. (1992). *Breaking the Circle of Satanic Ritual Abuse: Recognizing and Recovering from the Hidden Trauma*, CompCare Publishers, Minneapolis.

[47] Barden, C.R., a Ph.D. psychologist and clinic administrator, was one of many who became disgusted with the mounting trend in his profession. He took a J.D. at Harvard Law School, clerked for a future Supreme Court justice, and mounted a campaign to terminate insurance payments for questionable abuse therapy. Joining him were other respected clinicians, such as Campbell, Terence W. (1994). who would write *Beware the Talking Cure: Psychotherapy May Be Hazardous to Your Mental Health*, Upton Books, Baca Raton, who would form a new cadre of the psychology malpractice expert witnesses.

[48] Crnich, J. & Crnich, K. (1992). *Shifting Burden of Proof: Suing Child Sexual Abusers – A Legal Guide for Survivors and Their Supporters*, Portland.

[49] Based upon reports received by the False Memory Syndrome Foundation from persons reporting that they were falsely accused and possible targets of litigation.

[50] Most of the early cases originated in the Northwest, and the first case to reach the appellate courts was *Tyson v. Tyson*, 727 P.2d 226 (Wash. 1986) which barred pursuit of the litigation under the discovery rule; For a review of this legal issue, see (1996). Judicial response to repressed memory claims, *FMS Foundation Newsletter* **5**(10), 9–12.

[51] Hagen, A.M. (1991). Tolling the statute of limitations for adult survivors of childhood sexual abuse, *Iowa Law Review* **76**, 355–382.

[52] Bannon, C. (1994). Recovered memories of childhood sexual abuse: should the courts get involved when mental health professionals disagree? *Arizona State Law Journal* **26**, 835–856; Washington's *Tyson* decision was superseded by statute, and in all 24 states tolled the statute of limitations while Michigan and North Carolina bypassed their time limitations on the grounds of disability.

[53] The FBI was brought into the dispute but found no evidence to support the allegations. See Lanning, Kenneth Satanic, occult, ritualistic crime: a law enforcement perspective, *The Police Chief*, October, 1989, 1–11. While only a minority of clinicians accepted these accounts, that minority tended to accept their believability despite evidence of limited credibility.

[54] Bottoms, B.L., Shaver, P.R. & Goodman, G.S. (1996). An analysis of ritualistic and religion-related child abuse allegation, *Law and Human Behavior* **20**(1), 1–34 (reporting acceptance rates of such accounts by American Psychological Association member clinicians).

[55] Piper, Jr, A. (1994). Multiple personality disorder, *British Journal of Psychiatry* **164**, 600–612.

[56] Schacter, D.L. (1996). *Searching for Memory: The Brain, the Mind, and the Past* HarperCollins, New York.

[57] Edwards, C.N. (2001). *Responsibilities and Dispensations: Behavior, Science and American Justice*, Four Oaks Press, Dover.

[58] Piper, A., Lillevik, L., & Kritzer, R. (2008). What's Wrong with Believing in Repression: A review for Legal Professional, *Psychology, Public Policy, and Law* **14**(3), 223–242.

[59] Assertions of repressed memory and "dissociative amnesia" have been persistent but refuted by scientific evidence. See, e.g., Piper, Jr, A., Pope, Jr, H.G. & Borowiecki, III, J.J. (2000). Custer's last stand: Brown, Scheflin, and Whitfield's latest attempt to salvage 'dissociative amnesia', *Journal of Psychiatry & Law* **28**, 149–213.

[60] Piper, A. & Merskey, H. (2004). The persistence of folly: a critical examination of dissociative identity disorder (two parts), *Canadian Journal of Psychiatry* **49**, 592–600, 678–683.

[61] The International Society for the Study of Trauma and Dissociation (ISSTD) still includes a section on "Pharmacologically Facilitated Interviews" in its *Treatment Guidelines*, but advises that "Due to the current academic and forensic controversy surrounding dissociation disorders and traumatic memory, it is prudent to reserve these interventions for emergency situations . . ." (at p. 57). See http://www.isst-d.org/education/treatmentguidelines-index.htm#adults (accessed September 24, 2008).

[62] False Memory Syndrome Foundation, 3401 Market Street, Suite 130, Philadelphia, PA 19104-3318 (215-387-1865). The Foundation's board included Terence Campbell, and its Professional Advisory Board included Elizabeth Loftus from the University of Washington, Paul McHugh from Johns Hopkins, Richard Ofshe from the University of California, Martin Orne from the University of Pennsylvania, and Harrison Pope from Harvard Medical School.

[63] For an interesting assembly of contrasting views, see the special Spring 1994 issue of the *Journal of Psychohistory* (published in New York City).

[64] Rock, S.F. (1995). A claim for third party standing in malpractice cases involving repressed memory syndrome, *William & Mary Law Review* **37**, 337–79.

[65] Yamini, R.J. (1996). Repressed and recovered memories of child sexual abuse: the accused as direct victim, *Hastings Law Journal* **47**, 551–580.

[66] For a discussion of a related legal issue, see Loftus, E.F., Paddock, J.R. & Guernsey, T.F. (1995). Patient-psychotherapist privilege: access to clinical records in the tangled web of repressed memory litigation, *University of Richmond Law Review* **30**, 109–154. The limited precedents for third-party standing had been supported by psychotherapists providing social framework and expert testimony; typically in auto accident cases.

[67] Taub, S. (1996). The legal treatment of recovered memories of child sexual abuse, *Journal of Legal Medicine* **17**, 183–214.

[68] Sinnott, C.J. (1993). When defendant becomes the victim: a child's recantation as newly discovered evidence, *Cleveland State Law Review* **41**, 569–598.

[69] Clifford, R.A. (1996). Families target therapists in memory cases, *Chicago Lawyer* **19**(3), 8.

[70] More recently, the "survivors" of childhood sexual abuse have begun to sue therapists for failure to properly substantiate memories of abuse. Regehr, C. & Glancy, G. (1997). Survivors of Sexual Abuse Allege Therapist Negligence, *Journal of the American Academy of Psychiatry and the Law* **25**(1), 49–58.

[71] In late 1997, insurers for therapists and Rush-Presbyterian-St.Luke's Medical Center reached at $10.6 million settlement with former patient Patricia Burgus who had alleged that therapists at the Center had implanted false memories, and a federal grand jury in Texas handed down criminal false memory implantation charges for the use of techniques commonly associated with mind control and brain washing, *Journal of the American Academy of Psychiatry and the law* **26**(1), 146–147.

[72] For an empirical study of childhood memory, see Porter, S., Yuille, J.C. & Lehman, D.R. (1999). The nature of real, implanted, and fabricated memories for emotional childhood events: implications for the recovered memory debate, *Law and Human Behavior* **23**(5), 517–537.

[73] Lazo, J. (1995). True or false: expert testimony on repressed memory, *Loyola of Los Angeles Law Review* **28**, 1345–1414.

[74] Richmond, D.R. (1996). Bad science: repressed and recovered memories of childhood sexual abuse, *University of Kansas Law Review* **44**, 517–566.

[75] McAlister, C.V. (1996). The repressed memory phenomenon: are recovered memories scientifically valid evidence under Daubert? *North Carolina Central Law Journal* **22**, 56–82.

[76] Hough, J. (1996). Recovered, memories of childhood sexual abuse: applying the Daubert Standard in State Courts, *Southern California Law Review* **69**, 855–884.

[77] Choiniere, Monique The False Memory Syndrome Debate – will the victim please stand up? *Journal of Contemporary Health Law and Policy* **12**, (1996). 675–700 (proposing alternative evidentiary standards); This evidentiary assault has met with some success in the state courts, notably *State v. Hungerford*, 1995 WL 378571 (N.H. Super., May 23, 1995).
[78] Weidlich, T. (1995). Repressed memories: unreliable? Judges in four cases reject them before trial starts, *The National Law Journal* **17**(41), A7.
[79] Wright, L. (1993). Remembering Satan, *New Yorker*, May 17 and 24, 1993; The Ingram case is one of many recounted in the Pulitzer Prize winning Ofshe, R. & Watters, E. (1994). *Making Monsters*, Scribner's, New York.Ingram's daughter was first swept up in the movement at a church camp-out for girls, and Ingram himself, a deputy sheriff, in a dramatic demonstration of suggestibility, confessed to crimes that he could never have committed.
[80] *Frontline*, April 4 and 11, 1995.
[81] The false memory incident prompted extensive soul-searching within the mental health professions, as well as a number of books realigning the relationship between law and the behavioral professions. See, e.g., Appelbaum, P.S., Uyehara, L.A. & Elin, M.R., eds, *Trauma and Memory: Clinical and Legal Controversies*, Oxford University Press, New York.
[82] Brown, D. Scheflin, A.W.& Hammond, D.C. (1998). *Memory, Trauma Treatment, and the Law* W.W. Norton, New York.
[83] No. 61898 (Cal. Super.Ct., Napa County, 1994); The *Ramona* case is reported in Reisner, R. & Slobogin, C. (1995). *Law and the Mental Health System*, 2nd Edition Suppl., pp. 39–49; The human story surrounding the case is the subject of Johnston, M. (1997). *Spectral Evidence*, Houghton Mifflin, St. Paul, MN: West Group, Boston.
[84] Piper, Jr, A. (1993). 'Truth serum' and 'recovered memories' of sexual abuse: a review of the evidence, *Journal of Psychiatry and Law*, **Winter**, 447–471.
[85] Thompson, S.K. (2007). A brave new world of interrogation jurisprudence? *American Journal of Law and Medicine* **33**, 341–357.
[86] *Miranda v. Arizona*, 384 U.S. 436 (1966).
[87] Keller, L.M. (2007). Alternatives to Miranda: preventing coerced confessions via the convention against torture, *Chapman Law Review* **10**, 745.
[88] *Townsend v. Sain*, 372 U.S. 293 (1963).
[89] Some forty-seven cases appeared under Westlaw Key Cite 110k388.10 in September of 2008; the earliest from 1950 and the most recent being *Cogburn v. State*, 732 S.W.2d 807 (1987).
[90] For other examples, see "Treatment with Truth Serum" at http://echo.forensicpanel.com/1998/1/31/treatmentwith.html. (accessed Oct 2008).
[91] This view is reflected in the legal literature. See, e.g., Brugger, Winfried (2000). May government ever use torture? Two responses from German law, *The American Journal of Comparative Law* **48**, 661–678.
[92] Keller, L.M. (2005). Is truth serum torture? *American University International Law Review* **20**(3), 521–612.
[93] Brown, D. (2006). Some believe "truth serum" will come back, *Washington Post*, November 20, 2006, A08.
[94] Johnson, K. & Willing, R. (2002). Ex-CIA chief revitalizes 'truth serum' debate, *USA Today*, April 26, 2002, 12a.
[95] Lasson, K. (2008). Torture, truth serum, and ticking bombs: toward a pragmatic perspective on coercive interrogation, *Loyola University Chicago Law Journal* **39**, 329.

CARL N. EDWARDS, SAMUEL I. MILES AND AUGUST PIPER

Decision Making by Crime Victims *see* Crime Victims' Decision to Report Crime

Decriminalization of Mental Illness *see* Mental Health Courts

Defendant Characteristics in Sentencing *see* Sentencing: Demographic Factors in

Degraded Samples *see* DNA: Degraded Samples

Deinstitutionalization *see* Criminalization of the Mentally Ill

Deliberation: Jury *see* Jury Dynamics

Delusions

In the current diagnostic nomenclature of the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR), delusions are defined as false beliefs based on incorrect inference about external reality that is firmly sustained despite what constitutes incontrovertible and obvious proof or evidence to the contrary. Delusions are beliefs that are not ordinarily accepted by the affected person's social or cultural milieu [1, pp. 821–822]. A delusion needs to be distinguished from an overvalued idea in which an individual harbors an unreasonable belief or idea, but does not hold it as firmly as is the case with a delusion [1, pp. 821–822].

The aforementioned description of a delusion entails two important defining characteristics. First, it tends to view the nature of delusion as a seemingly unitary process. Second, delusional thinking tends to be viewed as a mutually exclusive binary process, i.e., delusions are fixed cognitions but not vice versa. More recently, delusions have been increasingly viewed as a multidimensional process in which some delusional characteristics are robustly representative of delusions, while others present with a more variable nature [2]. In this entry, we follow the latter view, because it provides a clearer and more systematic explanation of delusional experience. For example, a high degree of conviction continues to be an important characteristic of delusional thinking that frequently results in fixed delusions that constantly preoccupy the affected individual. Nonetheless the intensity of delusional conviction can fluctuate considerably, depending on factors such as social isolation and exposure to environmental stressors.

Delusions are amongst the most significant of psychiatric symptoms as they figure prominently in serious mental disorders such as schizophrenia, schizoaffective disorder, psychotic disorder due to general medical condition, substance-induced psychotic disorder with delusions, and of course, delusional disorders. A delusion has the distinction of being the only psychotic symptom for which separate category of psychotic disorders derives from the symptom itself.

Delusions may be divided into primary and secondary types. Primary delusions are explained by a psychological process that is essentially irreducible such as a religious delusion. Secondary delusions can be explained by other psychological processes such as hallucinatory experiences or affective factors. Differentiating primary versus secondary components may be of forensic relevance. For example, command hallucinations urging a person to kill others may increase a paranoid delusional conviction and the risk for violence toward others. This example also highlights the need to differentiate delusions from other psychotic symptoms as potential contributors to violent behavior.

Delusions may have varying degrees of organization. Some delusions are relatively simple, with a tendency to fragment, whereas others may present with very complex and highly stable plots involving many people, organizations, and conflicting concerns. Some delusions display a stable time course and may last for many years, while other delusions are relatively short-lived, though recent research suggests that certain factors such as never having married, older age, diagnosis of schizophrenia, delusions of body/mind control and thought broadcasting, having acted on a delusion, and higher levels of psychopathology and functional impairment are associated with persistence of delusional thinking [3].

Delusions are often classified according to content [4]. In "persecutory" delusions, the affected person believes he or she has become the object of mockery, harassment, persecution, or conspiracy. These delusions form the core feature of paranoid schizophrenia and delusional disorders. The central theme of "grandiose" delusions involves a grossly

exaggerated belief in the worth, importance, or power of the affected person. They are especially common in the manic states of bipolar disorder and mood disorder due to a general medical condition as well as occurring in schizoaffective disorder, schizophrenia or delusional disorder due to a general medical condition or induced by an exogenous substance. "Somatic" delusions encompass the belief that physical events in a person's body are occurring despite much evidence to the contrary. An example of a somatic delusion is a man who believes that he is pregnant and is about ready to deliver his own baby. A "nihilistic" delusion is characterized by the nonexistence of the self, part of the self, others, or the world. An example of this is a person who believes that he or she lost all his organs and is empty inside. Nihilistic delusions may be synonymous with somatic delusions if the emphasis of the delusional content involves nonexistence of the body or body part. "Jealousy" delusions involve the false belief that one's marital or sexual partner is being unfaithful. "Erotomanic" delusions involve the theme that the patient is the object of amorous pursuit by another. In "misidentification" delusions the affected person believes that the self or another presents with new physical or psychological attributes resulting in a new identity. A man who believes that his mother has been replaced by a physical double of his mother would be an example. Delusions have also been divided into a binary system of bizarre versus nonbizarre delusions. Bizarre delusions involve ideas that are not plausible such as, for example, if someone delusionally believes that most human beings are mechanical robots. An example of a nonbizarre delusion has an element of plausibility, as exemplified by the delusion that a spouse is sexually involved with a close friend of the man who harbors delusional jealousy.

Most cases of delusional thinking involve only one individual. However, on rare occasions, a delusional system may involve multiple individuals. For example, in the condition of folie-á-deux (currently subsumed under the diagnostic category of shared psychotic disorder), a second previously unaffected person develops the same delusion(s) as the primary or inducing person who already harbored an established delusion(s). The central feature of the induced or shared delusion(s) is its development in the context of a close relationship between the primary and secondary persons [4].

The Origin of Delusions

Although psychiatrists and other mental health professionals have in many cases constructed a psychodynamic explanation of the psychological and ecological factors that led to the development of an individual's delusional thinking, in recent times research has focused on the biological origin of delusions. Delusions are thought to be caused by abnormalities in cerebral physiology [5]. Although discrete physical causes are often not found, many delusions have been associated with an extensive array of biological factors. A large list of brain diseases can lead to the development of delusions. Delusional thinking can be found among many common neurological or medical conditions, including Parkinson's disease, Huntington's disease, syphilis, Alzheimer's disease, epilepsy, brain tumors, and traumatic brain injury. Medical conditions outside of the brain and central nervous system can also give rise to delusional thinking, including various endocrine disorders, vitamin deficiencies, and autoimmune disorders that would be recognizable to everyone [6]. Using various neuroimaging techniques [such as single proton emission computed tomography (SPECT), magnetic resonance spectroscopy (MRS), and positron emission tomography (PET)], researchers have been able to demonstrate brain changes on the cellular and neurotransmitter level in the development of delusions and other psychotic symptoms in individual who have used various psychoactive substances, including alcohol, amphetamines (primarily methamphetamine), cannabis (marijuana), cocaine, methylenedioxymeethamphetamine (MDMA, also known as Ecstasy), phencyclidine (PCP), ketamine, and inhaled hydrocarbons (such as gasoline, glue, paint, and related compounds) [7–12].

Treatment of Delusions

When there is an identified biologic causation to the delusions, such as a vitamin deficiency or endrocrine disorder, the treatment of the medical condition with the appropriate vitamin or hormone replacement/blocking agent would be the first step. Similarly, when treating a substance-induced psychotic disorder with delusions, abstinence from the offending drug should, but not always, ablate or attenuate the delusion(s) and other associated symptoms. Treatment of delusions accompanying a degenerative brain

disease such as Alzheimer's disease or Parkinson's disease has a lower success rate in view of the limitations of the therapeutic interventions for these medical illnesses.

In treating delusions in cases of schizophrenia, mood disorders with psychotic features, or other psychoses, use of antipsychotic medications has generally resulted in reduction of auditory hallucinations, delusions, and/or other manifestations of psychosis in most individuals. Interestingly, the clinical response in individuals with delusional disorders appears to be less successful [13]. Nonetheless, a significant proportion of individuals harbor delusions that appear to be refractory to clinical intervention.

Malingering of Delusions

The central feature of malingering involves the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives [1, pp. 739–740]. The frequency of feigned psychosis is unknown [14]. Research has not yet arrived at a foolproof method to determine the presence of genuine verses manufactured psychosis.

From reviewing the literature and his clinical experience, forensic psychiatric Philip Resnick has found that recognition of a malingered delusion requires clinical knowledge of the phenomenology of genuine delusions. Rensick has identified four particular signs that raise the suspicion of feigned delusions. These four signs include: abrupt onset or termination of the delusion; an individual's eagerness to call attention to the delusions; conduct markedly inconsistent with delusions; and bizarre content without disordered thinking [14].

However, on closer inspection of Resnick's inquiry, Resnick's method involves comparing the individual's description of the delusions with characteristics of delusions described by actual patients. Although the survey of the literature on the presentation of delusions shows likely characteristics as to what constitutes a genuine delusion, there remains some degree of uncertainty in some cases. Of course, during the course of ordinary and routine clinical practice of inpatient psychiatry, individuals may misrepresent the presence or extent of delusions to resist or gain admission, remain in the hospital, or seek release from the hospital.

At this point, keeping abreast of the clinical and research psychiatric literature may be best strategy for psychiatrists and other clinicians to know the typical characteristics of various delusions. The determination of the genuineness versus the manufacture of a reported delusion often does not present as a simple dichotomous choice. Individuals who have had previously experienced delusions, whether from a mental disorder, medical illness, and/or recreational or prescribed drug use, have first hand knowledge of the characteristics of a delusion, thereby complicating the task.

Probably the most frequent occurrence of manufacture or exaggeration of delusions involves requests for admission to psychiatric facilities, whether by individuals with no history of mental illness or by individuals who have previously harbored delusions. Also the most frequent manifestation of denial of actual delusions takes place when individuals seek to avoid psychiatric hospitalization, or if already hospitalized, to obtain release.

Delusions in Forensic Settings

Outside of the civil commitment setting, delusions arise in a variety of legal contexts. In a nutshell, forensic evaluators attempt to determine the presence of delusions in regard to the legal issue involved and if present, the effect(s) of delusion(s) on the legal matter at issue. Outside of the treatment-related civil commitment realm, common forensic evaluations in which delusions may be especially germane involve disability, competency to stand trial, mental state at the time of the offense, and dangerousness.

The so-called content-specific delusions, particularly those involving specific objects or targets of the delusion, may have increased forensic relevance. In particular, misidentification delusions, erotomanic delusions, and jealous delusions have been associated with cases of violence directed at the object of the delusion [15–17] though the relative risk of these delusions compared to delusions in general remains unclear.

Functional Capability Versus Disability

Delusions can be a significant factor in determining a person's work ability or degree of compensable disability. Individuals harboring paranoid or persecutory delusions may seek to avoid perceived persecutors, thereby precluding employment in positions that require interpersonal interaction. Certain

jobs may draw delusional individuals such as security guard work [18]. Delusions not involving specific paranoid or persecutory fears may be of such intensity that an individual becomes preoccupied to the extent that the person would not be able to focus on the tasks of the job. Delusions thus may form the central reason in evaluations for disability income, fitness for duty, or disability retirement.

Certain delusions, for example, nihilistic delusions may led to individuals refusing to eat because they harbor the delusion that they are already dead. When individuals reach the point that they lack the ability to provide for their basic survival needs, appointment of a guardian or conservator may be necessary.

Although not a disability issue, in the occupational context, workers who raise false allegations of harassment or discrimination based on sex, race, or gender that arise from delusional thinking may be especially problematic.

Competence to Stand Trial

Pretrial defendants who harbor delusions may have significant impairment in the capacities to meaningfully communicate with defense counsel to prepare a defense or to participate in the courtroom proceedings. Such impairment could jeopardize the defendant's competence to stand trial if the jurisdiction's criteria for incompetence are satisfied. The presence of delusions does not automatically confer incompetence as some individuals can proceed with their case if the delusions do not dominate their thinking or whose specific content is peripheral to the trial process. Defendants who harbor specific delusions revolving around defense counsel, for example, may preclude these individuals from even meeting with defense counsel.

Mental State at the Time of the Offense

Delusions often form part or all of the basis for a clinical-legal opinion supporting the insanity defense or diminished capacity. In regard to the insanity defense, delusions play a crucial role. An adjudication of not guilty by reason of insanity can be made if the trier of fact finds that the delusions negated the defendant's cognitive or volitional capacities as outlined by the jurisdictional insanity statute.

For example, a worker who develops the delusion that her boss is a nonhuman extraterrestrial creature who intends to kill her and then kills her boss in front of coworkers may qualify for the insanity defense, based on an incapacity to know right from wrong (or appreciate the criminality of her act) at the time of the killing.

Dangerousness

When assessing individuals for risk of harm to others, whether for initiation of civil commitment, or prior to releasing an insanity acquittee into the community, the presence of psychotic symptoms, particularly delusions, often merits close consideration. Delusions almost always are not a solitary factor in risk assessment, but need to be considered along with a host of other symptoms and factors [19]. Because of different study designs and different samples, there are varying findings as to effect of delusions in contributing to violent behavior [20–23].

References

[1] American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition, Text Revision (DSM-IV-TR), American Psychiatric Association, Washington, DC.

[2] Garety, P.A. & Hemsley, D.R. (1994). *Delusions: Investigatins into the Psychology of Delusional Reasoning*, Oxford University Press, New York.

[3] Appelbaum, P.S., Robbins, P.C. & Vesselinov, R. (2004). Persistence and stability of delusions over time, *Comprehensive Psychiatry* **45**, 317–324.

[4] Enoch, M.D. & Ball, H.N. (2001). *Uncommon Psychiatric Syndromes*, 4th Edition, Arnold, London.

[5] Richardson, E.D. & Malloy, P.F. (2001). The frontal lobes and content-specific delusions, in *The Frontal Lobes and Neuropsychiatric Illness*, S.P. Salloway, P.F. Malloy & J.D. Duffy, eds, American Psychiatric Publishing, Washington, DC, pp. 215–232.

[6] Cummings, J.L. & Mega, M.S. (2003). *Psychosis, Delusions, and Schizophrenia, in Neuropsychiatry and Behavioral Neuroscience*, Oxford University Press, New York, pp. 172–186.

[7] Iyo, M., Sekine, Y. & Mori, N. (2004). Neuromechanism of developing methamphetamine psychosis: a neuroimaging study, *Annals of the New York Academy of Sciences* **1025**, 288–298.

[8] Leweke, F.M., Gerth, C.W. & Klosterkotter, J. (2004). Cannabis-associated psyhchosis: current status of research, *CNS Drugs* **18**, 895–910.

[9] Arendt, M., Rosenberg, R., Foldager, L., Perto, G. & Munk-Jorgensen, P. (2005). Cannabis-induced psychosis and subsequent schizophrenia-spectrum disorders: follow-up study of 535 incident cases, *The British Journal of Psychiatry* **187**, 510–515.

[10] Chang, L. & Haning, W. (2006). Insights from recent positron emission tomographic studies of drug abuse and dependence, *Current Opinion in Psychiatry* **19**, 246–252.
[11] Quickfall, J. & Crockford, D. (2006). Brain neuroimaging in cannabis use: a review, *The Journal of Neuropsychiatry and Clinical Neurosciences* **18**, 318–322.
[12] Thirthalli, J. & Benegal, V. (2006). Psychosis among substance users, *Current Opinion in Psychiatry* **19**, 239–245.
[13] Manschreck, T.C. & Khan, N.I. (2006). Recent advances in the treatment of delusional disorder, *Canadian Journal of Psychiatry* **51**, 114–119.
[14] Resnick, P.J. (1997). Malingered psychosis, in *Clinical Assessment of Malingering and Deception*, 2nd Edition, R. Rogers, ed, Guilford Press, New York, pp. 47–67.
[15] Silva, J.A., Leong, G.B., Weinstock, R., Sharma, K.K. & Klein, R.L. (1994). Delusional misidentification syndromes and dangerousness, *Psychopathology* **27**, 215–219.
[16] Leong, G.B. (1994). De Clérambault's syndrome (erotomania) in the criminal justice system: another look at this recurring problem, *Journal of Forensic Sciences* **39**, 378–385.
[17] Leong, G.B., Silva, J.A., Garza-Treviño, E.S., Oliva Jr, D., Ferrari, M.M., Komanduri, R.V. & Caldwell, J.C.B. (1994). The dangerousness of persons with the Othello syndrome, *Journal of Forensic Sciences* **39**, 1445–1454.
[18] Silva, J.A., Leong, G.B. & Weinstock, R. (1993). The psychotic patient as security guard, *Journal of Forensic Sciences* **38**, 1436–1440.
[19] Monahan, J., Steadman, H.J., Appelbaum, P.S., Robbins, P.C., Mulvey, E.P., Silver, E., Roth, L.H. & Grisso, T. (2000). Developing a clinically useful actuarial tool for assessing violence risk, *The British Journal of Psychiatry* **176**, 312–319.
[20] Wessely, S., Buchanan, A., Reed, A., Cutting, J., Everitt, B., Garety, P. & Taylor, P.J. (1993). Acting on delusions: I. Prevalence, *The British Journal of Psychiatry* **163**, 69–76.
[21] Link, B.G. & Stueve, C.A. (1994). Psychotic symptoms and the violent/illegal behavior of mental patients compared to community controls, in *Violence and Mental Disorder*, J. Monahan & H. Steadman, eds, University of Chicago Press, Chicago, pp. 137–159.
[22] Swanson, J., Borum, R., Swartz, M. & Monahan, J. (1996). Psychotic symptoms and disorders and risk of violent behavior in the community, *Criminal Behavior and Mental Health* **6**, 317–338.
[23] Appelbaum, P.S., Robbins, P.C. & Monahan, J. (2000). Violence and delusions: data from the MacArthur violence risk assessment study, *The American Journal of Psychiatry* **157**, 566–572.

J. Arturo Silva, Gregory B. Leong and Robert Weinstock

Dementia as Elder Abuse Risk Factor *see* Elder Abuse: Risk

Demographic Factors in Sentencing *see* Sentencing: Demographic Factors in

Demonstrative Evidence

Demonstrative Evidence Is a Broad Label

Demonstrative evidence is a term used to describe a particular form of physical or documentary proof under the technical admissibility rules used primarily in the United States, the United Kingdom, Canada, and other countries, whose system is based on the adversary system of justice (*see* **Expert Opinion in Court: a Comparison of Approaches**; **Expert Opinion: United States**; **Expert Opinion: United Kingdom, Canada, and Australia**). Demonstrative evidence is not used as substantive proof of a fact but, instead, to explain the oral factual or opinion testimony of a witness.

Demonstrative evidence can be in the form of casts, models, drawings, photographs, motion pictures, videotapes, computer animation (*see* **Computer Animation and Simulation Evidence**), physical objects, and other types of an inanimate nature such as skeletons, toys, dolls, etc., [1–7]. The scope of what can be offered as demonstrative evidence is only limited by the imagination of either the witness or the litigator.

In the forensic sciences, demonstrative evidence is widely used by a great variety of skilled professionals. A representative list includes the following:

- plaster casts of impressions made by trace evidence analysts or other crime-scene investigators;

- scale models of places, buildings, or objects that are related to the place where the incident that is the subject of litigation occurred;
- skeletons or anatomical models to illustrate wounds or injuries;
- anatomically correct dolls to aid a victim of a sexual abuse and rape case;
- maps, plans, and sketches that relate to a crime or an accident scene, whether made free-hand, to scale, by an investigator, witness, or licensed surveyor;
- artist drawings of suspects, Identi-Kit likenesses, computer-generated composite sketches, and other representations of unknown persons from visual descriptions furnished by eyewitnesses or victims;
- charts and drawings produced photographically that illustrate an identification expert's conclusions, such as those used by fingerprint experts to draw attention to matching characteristics in latent marks and known exemplars, or by firearm and toolmark examiners on photomicrographs matching marks impressed on crime scene and exemplar bullets or cartridge cases;
- other photographs produced by film and paper as well as digital cameras, whether in black and white or in color or even when produced by specialized photographic techniques such as X rays or by other medical instrumentation, representing buildings, locations, persons, objects, or findings that are relevant to issues in civil as well as criminal incidents;
- motion pictures and videotapes of places or persons involved in litigation-related incidents, including but not limited to driver-arrest circumstances, day-in-the-life-of-a-person moving images, and similar occurrences;
- computer animations and simulations of events supporting accident reconstruction findings and other incidents of relevance to civil and criminal cases; and
- charts and summaries, compiled under a great variety of circumstances.

Weighing the Probative Effect against Potential Prejudicial Impact

The admission of such evidence is ordinarily within the discretion of the court. The judge will typically weigh the probative value of the evidence against the prejudicial effect that the evidence might have on the jury. It is always required that the proposed evidence illustrates or explains a relevant issue in the case.

As the admission of demonstrative evidence requires that there be a testimonial sponsor of the proof who has personal knowledge of its connection to the litigation, the sponsor–witness must lay a foundation before the demonstrative proof can be admitted. Depending on the nature of the item, such proof may include testimony of a substantial connection with the event or occurrence litigated, or a substantial similarity to it.

Thus, a witness proposing to testify with the aid of a photograph may be asked whether the photograph of a locality involved in an incident depicts the location substantially as it existed at the time of the event in issue. As part of "laying the foundation" for admission, the witness may further be asked to explain how his involvement in the events began that ultimately let to its recovery or production. The witness may also be asked to give an explanation of the mechanism or technique by which the photograph was produced.

In the case of the admission of movies or videotapes, a court suggested that necessary foundation testimony includes, first, a showing by someone having personal knowledge of the filmed object that the film is an accurate portrayal of what it purports to show and, second, that the evidence is needed to an understanding of what happened [8]. This will permit a court to decide whether the probative value of the demonstrative evidence outweighs the danger of unfair prejudice to the opposing party [9].

"Mugshots": Earliest Form of Photographic Evidence

Identification pictures (mugshots) are perhaps the earliest form of photographic demonstrative evidence known to legal jurisprudence. Such evidence has been admitted in courts worldwide since the very origin of photography. Thus, perhaps the first known example of "mugshots" produced by the early photographic processes called *tin-types* and *daguerreotypes* of arrested suspects were used in Belgium in 1843 [10]. The French police official Alphonse Bertillon, the inventor and originator of the anthropometric system of identifying individuals by bodily measurements [11], is credited with adding a profile view to the collection of identification photographs

recorded in the Paris *Préfecture de Police*, beginning in 1873 [10].

In the admission of handwriting comparison evidence, courts were almost as early in recognizing the use of photographic evidence as they were in admitting mugshot evidence. Thus, in an 1859 case, an American court approved of the admission of photographs in aid of testimony challenging the authenticity of the signature of Governor Pico of California on a land-grant document [12].

Today, because of the widespread, nay universal, use in all media of visual stimuli, courts are likely to admit demonstrative evidence of the widest variety, as long as the proper foundation testimony is presented.

References

[1] Moenssens, A.A., Henderson, C.E. & Portwood, S.G. (2007). *Scientific Evidence in Civil and Criminal Cases*, 5th Edition, Foundation Press, 111–191.

[2] Duckworth, J.E. (1983). *Forensic Photography*.

[3] Feller, H.J. (1993). Photographic evidence: more than meets the eye, *Maine Bar Journal* **372**(8).

[4] Mnookin, J.L. (1998). The image of truth: photographic evidence and the power of analogy, *Yale Journal of Law and the Humanities* **1**(10).

[5] Varner, C.D. (1999). Worth a thousand words: the admissibility of day-in-the-life videos, *Tort and Insurance Law Journal* **175**(35).

[6] Marcus, R.P. (2001). Admissibility of effects of videotapes of medical procedures in litigation, *Journal of Legal Medicine* **401**(22).

[7] Albrecht, A.W. (2003). Laying a proper foundation for computer generated demonstrative evidence, *Illinois Bar Journal* **261**(90).

[8] Flannery, J.R. (2007). Using videos at trial: the big picture, *Illinois Bar Journal* **642**(95).

[9] McGoorty v. Benhart, 305 Ill. App. 458, 27 N.E.2d 289 (1940).

[10] Several of these earliest known judicial uses of identification photographs in the world were illustrated in Moenssens, A.A. (1962). The origin of legal photography, *Finger Print and Identification Magazine* 3. The article and some of the illustrations are also available on the Internet at: http://www.forensic-evidence.com/site/EVID/LegalPhotog.html.

[11] Bertillon, A. (1896). *Signaletic Instructions – Including the Theory and Practice of Anthropometrical Identification*, R.W. McClaughry, eds, The Werner Company. Bertillon called his system the *portrait parlé*.

[12] Luco *et al.* v. United States, 23 Howard 515 (1859). The court marveled at the novel evidence and stated: "We have ourselves been able to compare these signatures by means of photographic copies and fully concur, from evidence *oculis subjecta fidelibus*, that the seal and the signatures of Pico are forgeries".

ANDRE MOENSSENS

Denial by Hostages ***see*** Stockholm Syndrome

Dentistry: Forensic ***see*** Odontology

Dependence on Alcohol and Drugs ***see*** Alcohol: Use, Abuse, Tolerance, and Dependency

Dependency Law ***see*** Parental Rights and Prerogatives

Depositions of Experts for Discovery Purpose ***see*** Discovery: Depositions

Depression: Postpartum ***see*** Postpartum Psychosis

Descriptive Psychopathology ***see*** Aggression, Mental Status: Examination, Psychopathology: Terms and Trends

Destruction ***see*** Aggression

Desyndromizing: Battered Woman's Reality ***see*** Battered Woman's Reality

Detection of Deception ***see*** Deception: Detection of

Developmental Immunity ***see*** Children: as Defendants

DFSA ***see*** Drug-Facilitated Sexual Assault

Diagnostic and Statistical Manual (DSM) ***see*** Psychopathology: Terms and Trends

Diagnostic Interview ***see*** Mental Status: Examination

Diatoms

Introduction

What are Diatoms?

Diatoms are unicellular microscopic algae belonging to kingdom Protista and class Bacillariophyceae. More than 200 genera and approximately 100 000 species of diatoms have been reported in the literature so far. Some of the commonly occurring diatom genera are listed in Table 1. Diatoms are found in fresh water, salt water, and in moist soil. Diatoms usually have yellowish or brownish tint, which is due to the presence of photosynthetic plastids. There is large variation not only in the size (which varies from few micrometers to more than 100 μm) but also in their shape, from box-shaped to cylindrical and symmetrical to asymmetrical (photomicrographs of some of the common genera of diatoms have been shown in Table 2).

The growth of diatom takes place by mitotic cell division as well as by the formation of an auxospore by sexual reproduction. Living diatoms often have specific salinity, temperature, and other

Table 1 Showing some commonly occurring diatom genus

Achhnanthes	*Geissleria*
Achnanthidium	*Gomphocymbella*
Actinocyclus	*Gomphoneis*
Amphipleuara	*Gomphonema*
Amphora	*Gyrosigma*
Anomoeoneis	*Hannaea*
Catacombus	*Hantzschia*
Cocconeis	*Melosira*
Coscinodiscus	*Navicula*
Craticula	*Neidium*
Cyclotella	*Nitzschia*
Cymatopleura	*Pinnularia burkii*
Cymbella	*Placoneis*
Diadesmus	*Pleurosigma*
Diatoma	*Rhoicosphenia*
Diploneis	*Stauroneis*
Encyonema	*Staurosurirella*
Epithemia	*Stenopterobia*
Eucocconeis	*Stephanodicus*
Eunotia	*Surirella*
Fragillaria	*Synedra*
Frustulia	*Thalassiosira*

environmental tolerances; therefore, growth of diatoms is directly related to the temperature, light intensity, and physiochemical characteristics of the water, i.e., pH, salinity, concentration of silicon, the presence of organic matter, etc. Their population increases during spring and autumn (because of the favorable climatic conditions), while it decreases during summer and winter [1].

Classification of Diatoms

On the basis of their symmetry, diatoms can be divided into two orders as follows:

- Centrales (*Biddulphiales*) – which have valve striae arranged basically in relation to a point and appear radially symmetrical; and
- Pennales (*Bacillariales*) – which have valve striae arranged in relation to a line and appear bilaterally symmetrical.

The most acceptable classification system was developed by Round *et al.* [2]. Centric diatoms (plankton) commonly found in the marine water can be divided as follows:

- *Coscinodiscineae* – having a marginal ring of processes and nonpolarity symmetry;
- *Rhizosoleniineae* – with no marginal ring of processes and have unipolar symmetry; and
- *Biddulphiineae* – with no marginal ring of processes and have bipolar symmetry.

The pennate diatoms can be divided into two suborders as follows:

- *Fragilariineae* – which is araphid and
- *Bacillariineae* – which posses a raphe.

Structure of Diatom

All the diatoms are covered by a "frustule", which is made up of two valves of different size. The larger valve is called *epitheca*, which connects the smaller valve or "hypotheca" with the help of a "cingulum", also called *set of girdle bands* (Figure 1).

The valves are composed of silica (SiO_2), which is a characteristic feature of their silica shells and can be used for their identification. The face side of the valve is decorated with various identifying features like pores (areolae), processes, spines, hyaline areas, etc. Another important characteristic feature of the diatoms is "raphe", which is found mostly in pennate diatoms and it divides the valve into two (Figure 2).

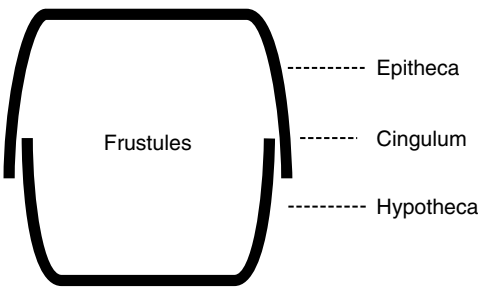


Figure 1 Diagrammatic representation girdle view of diatom

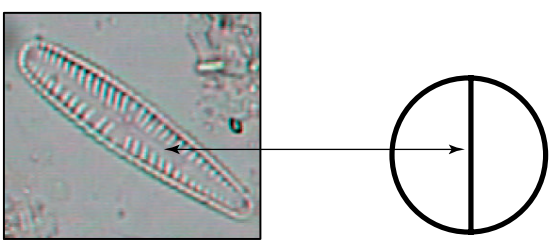
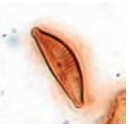
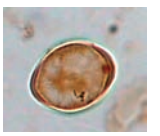
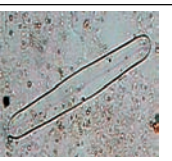
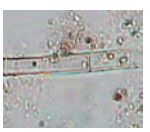
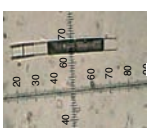
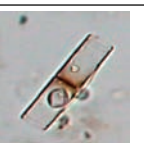
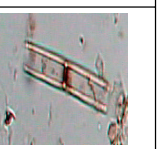


Figure 2 Valve view of diatom with raphe

Table 2 Showing photomicrographs of some commonly occurring diatoms

Cymbella species			
			
Cocconeis species			
			
Navicula species			
			
Pinnularia species			
			
Melosira species			
			

Forensic Diatomology

Historical Perspective

In some earlier attempts, ideas about the presence of aquatic debris in a drowned body were framed by Guy [3]. He stated that whenever drowning takes place, water, along with some mud and other debris, channels to the various organs. First, diatoms were discovered in the lungs by Hofmann in 1896. Revenstorf [4] used this evidence to help solve a drowning mystery in 1904. This test was improved by Kasparck [5], who digested lung tissue in acid medium to extract diatoms. Incze [6] successfully tested blood and parenchymatous organs for the presence of diatoms. Later Tamasaka [7] used bone marrow as a source material for the recovery of diatoms in drowned bodies. His studies helped to reach a conclusion that the presence of diatoms in bone marrow can indicate that the death is due to drowning. Timpermann [8] reported that the presence of even a single diatom in the bone marrow can establish that the cause of the death is due to drowning but later on in 1979, he suggested that it might not be a solid basis for such a diagnosis. Geissler and Gerloff [9] and Hendey [10] were the pioneer diatomologists, who first brought the specialized knowledge about this study in the forensic field. In 1977, Peabody further evaluated the results obtained from diatom tests and recommended this test for the diagnosis of the drowning cases. His work was a pioneering contribution in the field of "forensic diatomology". Pollanen [11–13] and his colleagues contributed significantly by conducting various experiments for utility and

validity of diatom test and proposed various factors that could interfere in the proper application of this test.

Ludes *et al.* [14–16] also conducted some experiments in order to locate the most probable site of drowning on the basis of diatom test and performed "Continuous River monitoring of diatoms". Recently, Rohn and Frade [17] published an article on diatom test and its significance. They mentioned the significance of diatoms in medicolegal investigations along with its history.

The origin of nondrowned diatoms has always remained a controversial issue because few workers have found diatoms even in the nondrowned bodies. Geissler and Gerloff [9] were of the opinion that this test should not be used on drowned infants. Mueller [18] also reported that diatoms can pass from the air to other organs. Spitz [19] conducted some experiments on rats and concluded that the presence of diatoms in the organs might be due to contamination by water and airborne diatoms. Apart from this, another possible cause was due to the contamination from laboratory glassware and reagents.

Why is Diatom Evidence So Good for the Diagnosis of Drowning Case?

Water contains a variety of other planktons but the reasons why only diatoms are used as criteria in these cases are as follows:

- These microorganisms have acid resistant frustules, which makes them easier to extract from postmortem tissues. Otherwise, any attempt made to extract any other fragile microorganisms from the tissue may end in failure.
- They are numerous and more importantly smaller in size to penetrate into the various distant organs.
- They can be easily identified and classified.
- In case of failure of chemical testing of blood for the diagnosis of drowning, this test can be administered.
- In case of skeletonized drowned body, the only alternative left is the diatom test.

Forensic Significance of Diatom Test

Diatoms can act as very sensitive and significant evidence in drowning cases. The test conducted to identify diatoms helps in distinguishing between antemortem and postmortem drowning, i.e., whether person was living or dead at the time of submersion. In other words, this test explains

- whether the cause of death was drowning or not and
- whether the person was alive at the time of immersion or not.

This test can also play an important role in searching the most likely drowning site [15, 16, 20], particularly in the cases

- when the body is found on land and no reference water body is available and
- when the body is found in the same water body away from the actual site of drowning, due to the body floating or dragging with the flow of water or other reason.

Fate of Diatoms Inside the Body

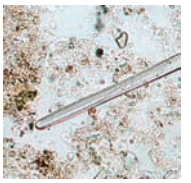
When a person breathes underwater (only in antemortem drowning), some water enters the lung cavity with force and exerts pressure on the lung walls. Owing to this pressure, lung alveoli get ruptured and water reaches the blood stream. Till the last heart beat, water keeps on mixing with the blood and during this course some diatoms (if at all present in the water) particularly the smaller ones are also pushed into various vital organs like liver, spleen, kidneys, brain, and even bone marrow. So it is only the respiratory pressure which helps diatoms to reach these distant organs.

On the other hand, the chance of diatoms entering in distant vital organs remains negligible if the person is already dead (postmortem immersion) and is thrown into the water. In certain instances, however, the diatoms can be detected in the lungs even in the case of postmortem drowning when, due to the passive absorption of water, diatoms can reach the lung cavity and their journey ends there.

Validity of Diatom Test

The validity of diatom test can be checked by applying the guidelines of "Criterion of Concordance" given by Pollanen in 1998. According to this criterion, the significant number of diatoms should be

Table 3 A few small sized diatoms

Some other diatom		
Gomphoneis	*Hantzschia*	*Diatoma*
		
Amphipleura	*Surirella*	*Synedra*
		

present inside the distant vital organs and they should match with the diatom species found in the water body from where the body has been recovered. This criterion can also be useful in providing solution to the controversial aspect of diatom test (diatoms are always present in human tissues) by stating that the diatoms species found in the blood must match with the species found in distant vital organs of the victim (such as the bone marrow). This refutes the doubt of many pathologists, who claim that diatoms are always present in human tissues. Therefore, application of the "Criterion of Concordance" is thought to be satisfactory enough to convince most critics in a court of law [13]. Similarly, Ludes *et al.* [14] studied the required number of diatoms in different vital organs of a drowned person. They concluded that diatom analysis should be considered positive only when a minimum established number of diatoms was present in the different vital organs, i.e., 20 diatoms per 100 µl of pellet (obtained from 10 g of lung samples) and 5 diatoms from other organs.

What are the Drowning-Associated Diatoms (DAD)?

For any forensic expert, the knowledge about drowning-associated diatoms (DAD) becomes necessary while dealing with drowning cases. Specific features of the diatoms like their type, size, and density can be crucial to get their entry into the different vital organs. No doubt smaller diatoms like *Achhnanthes, Hannae, Diatoma and Cyclotella*, etc. (Table 4) have more chances of entering into distant body organs, but fairly large diatoms like *Synedra* (Table 3) can also be found in these organs owing to their delicate nature. *Navicula, Nitzschia, Achnanthidium*, and *Cyclotella* are the most common types of diatoms found in the peripheral organs of drowned persons because they are widely distributed and have optimum size.

As the diameter of lung alveoli is very small, it is not possible for all the diatoms to penetrate into the organs through the lung cavity. Diatoms that can penetrate through this capillary network are called *drowning-associated diatoms*.

Large diatoms in the bone marrow can be indicative of laboratory contamination, as mostly small pinnate diatoms are found in the bone marrow [11]. After extracting some diatoms from the bone marrow, Pollanen concluded that strictly restricted

Table 4 Photomicrographs of some commonly occurring diatoms

Achhnanthes	*Cyclotella*	*Hannae*
		

sizes of diatoms are usually associated with drowning. This becomes a very useful guide in determining the type of diatoms that can be found in the bone marrow. Similarly, Hurlimann *et al.* [21] also stressed that penetration of a diatom in the lung capillaries depends on its size and density; therefore, it is required to emphasize on the morphological and morphometric studies of diatoms to diagnose death by drowning. In 1998, Lunetta *et al.* discovered the penetration of diatoms into the alveoli-capillary barrier by using transmission electron microscope (TEM) and scanning electron microscope (SEM). During this research, they observed various diatom species like *Diatoma moniliformis* ("penetrating" through the wall of a distal airway); *Navicula specula* (penetrating through Kohn's pore); *Tabularia fasciculat* (partially penetrating into a laceration of epithelial and endothelial lining of a distended alveolar septum); *Nitzschia paleacea* (partially penetrating through a laceration of alveolar wall); and *Mastogloia smithii* (penetrating the alveolar wall through a clearly visible laceration).

Forensic Analysis of Diatoms

Postmortem Samples

Usually postmortem samples sent for diatom examinations are of two types:

- hard bones like sternum and femur or
- soft tissues like lungs and liver.

Femur and sternum bones have been considered the most useful and least contaminated samples for this test [11]. These samples either can be cut open for the collection of marrow or they are directly dissolved by acid digestion method to extract diatoms. Even dissolution of these samples is not very difficult but complete extraction of diatoms frustules from these samples needs great care and skill.

Reference Water Samples

About 500 ml of water samples should be collected from the suspected site of drowning and stored in the refrigerator as a reference sample.

Extraction of Diatoms from Postmortem Samples

Chemical digestion method (by cooking the organs with acid) and enzymatic digestion method are most commonly for the extraction of diatoms from postmortem samples.

Various methods used for the extraction of diatoms are as follows:

- acid digestion method (nitric acid);
- ultrasonic radiation to solubilize tissue;
- enzymatic digestion with proteases such as proteinase-K;
- strong anionic detergents such as sodium dodecyl sulfate (SDS) and hydrogen peroxide; and
- membrane filtering method (blood samples)-not in use.

Acid Digestion Method

This is the most commonly used method for the dissolution of rigid postmortem samples like bones.

Method. About (50 g) of marrow is removed from the bone sample and put in a boiling flask. Approximately, 50 ml of concentrated nitric acid is added to it and marrow-acid suspension is simmered on a hot plate for approximately 48 h in a fume hood. After cooling the suspension at room temperature it is centrifuged (200–300 rpm/30 min). The distilled water is added to the obtained residue and is again centrifuged. The final supernatant is discarded and the pellet containing nitric acid-resistant material is aspirated using a Pasteur pipette and put on a clean microscopic slide for examination [12].

Enzymatic Method

The enzymatic method was considered more convenient in terms of rapidity, safety, and environmental protection than the chemical test.

Method. About 10 g each of the organs (peripheral lung tissue, kidney, liver, brain, and femur bone marrow) is minced with scissors. The sample is rinsed and mixed with 500 ml of 10 mg ml^{-1} proteinase-K and 100 ml of 0.01 M Tris–HCl buffer (pH 7.5) containing 2% SDS. The mixture is incubated at 50°C for overnight. In total, 500 ml proteinase-K is added and the solution is diluted with 100 ml of distilled water. It is then centrifuged at 3000 rpm for

15 min and the upper layer is discarded. The sediment (100 ml per slide) is transferred to a cover glass, mounted in Naphrax and examined under the light microscope Ludes *et al.* [15].

Extraction of Diatoms from Reference Water Samples

Modified Acid Digestion Method

This sample is mixed thoroughly and transferred to a sterilized beaker. Then 40–45 ml of concentrated nitric acid (HNO_3) is added to digest the unwanted material from the sample and then the sample is kept undisturbed for about 2 h. The whole sample is then centrifuged at 5000 rpm for 10 min. The supernatant is removed leaving behind only a pellet containing diatom frustules at the bottom of the tube. This pellet is suspended in distilled water and again centrifuged in the same way. This process is repeated twice to ensure all traces of the acid are removed Singh *et al.* [24].

Enzymatic Method

Method. Water samples are treated with hydrogen peroxide (130 vol%) at 80°C for 12 h. The solution is then cooled at room temperature and the second centrifugation is set at 2500 rpm for 15 min. After discarding the supernatant, residue is finally centrifuged with distilled water (3000 rpm for 10 min), which produces a pellet containing diatoms. After removing the supernatant, the sediment is air dried and mounted in Naphrax [15].

Acid Digestion Method for Sea Water

Water sample is incubated with fuming nitric acid for 30 min in a boiling water bath. Then the sample is washed with pure ethanol twice. This washing is made by simple centrifugation at 3200 rpm for 10 min. The residue is heated and dried on a glass slide and diatoms are examined with an optical microscope at higher magnification [23].

The "acid digestion" method is used in majority of the cases because it is easy and a better means to remove the organic materials. Besides having many merits, this method also has some demerits like

- substantial amounts of acids are required to digest 100 g of the tissue sample;
- a high-speed centrifugation is required to remove traces of acids that make this method lengthy;
- chemical digestion of fatty materials may not be completely removable;
- chances are there that diatoms may be lost in repeated washings and centrifugations;
- as Kjeldahl flasks and reflux apparatus are costly, therefore, they are reused after washing, so acid digestion is prone to contamination from the reagents as well as the glassware.

No doubt, the acid digestion method is effective for detecting diatoms, but necessarily not very safe. Despite having many limitations, acid digestion method is still better and is very frequently used for the extraction of diatoms from water and tissue samples.

The "enzymatic digestion" method is simple, safe, and effective for detection of both phytoplankton, including diatoms and zooplanktons, which is not possible with the acid digestion method. Therefore, this method is more advantageous than the acid digestion method in diagnosing drowning cases. Using subtilisin, this is an effective and rapid method of destroying tissue materials and Tris buffer can be easily cleaned by centrifugation. However, this method is bit costly compared to the traditional methods [24].

Preparation of Slides

Extracted diatoms need to be properly preserved for their noticeable visibility through microscope. For preservation, diatoms are permanently mounted on the microscopic slide.

Method

Residual material obtained after centrifugation is put on a microscopic slide and is allowed to fix. Then a drop of DPX mountant is poured on the center of a clean prelabeled slide, covered with a cover slip, and then examined under the microscope for possible diatoms.

Diatom Identification

Slides are examined using a binocular compound light microscope on 1000× (oil immersion) to maximize resolution.

Standards and Controls

- In diatom examination, the control water samples must be used for comparison purpose invariably.
- Standard diatom samples can be preserved on slides and can be used as standards for comparison purpose.

Role of Diatomological Maps in the Forensic Investigation of Drowning Site

Another important aspect of "forensic diatomology" is its use in search of most putative drowning site. Sometimes under suspicious circumstances

- the drowned body is recovered from a different location in the same water body, but away from the actual site of downing, i.e., the body has been dragged with the flow of water or has been done manually by somebody or
- the presumed drowning site may not be there at all, i.e., body is found on the land away from any water body.

The qualitative and quantitative analysis of diatom species can be performed to ascertain not only the cause of death but also the most probable site of drowning. Distribution of diatom species in any particular water body and their correlation with the diatoms recovered from the drowned body can be a method of choice to resolve the questions related to drowning site. After generating the profile of diatom flora of any water body, a diatomological map can be designed. These diatomological maps can be of great help in the characterization of the selected water bodies and are of immense help in diagnosing the suspected drowning cases occurring in any particular area [15, 20, 23, 24].

Authors have conducted studies dealing with the algological analysis from three different water bodies (Galta ji, Jal Mahal, and Mavath) of Jaipur city of Rajasthan [24]. Significant qualitative and quantitative differences in the distribution of some diatom species were observed. Certain site indicator diatom species were also found in the selected water bodies. On the basis of this database, diatomological maps of the selected sites were designed. Some characteristic diatom like *Rhoicosphenia* and *Cyclotella* were observed in Galta ji (a small waterfall). This water body was full of diatom blooms. Other large types of diatoms like *Synedra and Melosira* were also found here. However, in Jall Mahal (a pond) where water was stagnant and highly polluted fewer types and number of diatoms could be found. Except *Navicula*, which is the most commonly found diatom, *Cyclotella* was the only other dominating diatom species found. Mavath is a seasonal water pond, where elephants are brought to take bath. A great diversity of diatoms, i.e., *Synedra, Nitzschia, Diatoma, Navicula, Geissleria*, and *Achcanthidium*, was observed. *Geissleria* was found to be restricted to this particular water body, whereas *Navicula* and *Cyclotella* were found in all the three sites.

Conclusions

Diatoms are not found in all the drowning cases but if present in distant vital organs in sufficient numbers they provide positive evidence that death is due to antemortem drowning. A few authors still have some doubts about the authenticity of this test. According to them, the diatoms are not only inhaled through water but also through air. However, this position is difficult to sustain when the diatoms are found inside the peripheral organs.

Diatom analysis, like many forensic processes, requires proper care to avoid any sort of contamination while performing this test and necessary conditions like "Criterion of Concordance" should be followed. Application of this test would be a great assistance in the investigation of drowning cases.

Another aspect related to diatom test is the determination of the actual drowning site. Site-specific diatom species can also play a crucial role in the search for possible drowning site. But in order to minimize errors, results pertaining to such studies should be accurately interpreted. To achieve this, the studies related to monitoring of water body for diatoms need to be extended by conducting a comparative study to investigate the seasonal distribution of diatom species. These comparative studies can help to resolve many discrepancies or changes in diatomological database that can take place with the passage of time.

Scope of Biomolecular Studies in Forensic Diatomology

The methods related to molecular biology, especially DNA, occupy a secure and very significant place in various fields of the study. These methods may prove

to be beneficial for the future development of diatom tests for drowning [11]. This aspect has never been given much importance, but the use of polymerase chain reaction (PCR)-based methods for the detection of diatoms in human tissues would be an achievement in the field of "forensic diatomology". Sensitive methods seem to be an advantageous alternative over the average microscopic examination of diatoms, particularly in tissue. A high degree of uncertainty could be removed if diatom comparison is made on the bases of their genomes.

References

[1] Trent, G. (2004). Something in the water, *Law and Order* **52**(6), 92–93.

[2] Round, F.E., Crawford, R.M. & Mann, D.G. (1990). *The Diatom Biology and Morphology of Genera*, Cambridge University Press, Cambridge, pp. 7–20.

[3] Guy, W.A. (1861). *Principles of Forensic Medicine*, 2nd Edition, Renshaw, London.

[4] Revenstorf, V. (1904). Der Nachwies der aspirierten Ertrankungs flussigkeit als Kriterium des Todes, *Gerichtliche Medizin* **28**, 274–279.

[5] Kasparck, B. (1937). *Deutsche Zeitschrift fur die Gesamte Gerichtliche Medizin* **27**, 132.

[6] Incze, G. (1942). Fremdkörper in Blutkreislauf Ertrunkener, *Zentralblatt fur Allgemeine Pathologie und Pathologische Anatomie* **79**, 176.

[7] Tamasaka, L. (1949). Vizihullak Csontvelojneck Diatomattartalmarol (Diatom content of bone marrow in corpses in water), *Orvosi Hetilap* **16**, 509–511.

[8] Timpermann, J. (1972). The diagnosis of drowning- a review, *Journal of Forensic Sciences* **1**, 397–409.

[9] Geissler, U. & Gerloff, J. (1966). Das Vorkommen von diatomeen in menschlichen organen und in der luft, *Nova Hedwigia* **10**, 565–577 (cited from Peabody, 1980).

[10] Hendey, N.I. (1973). The diagnostic value of diatoms in case of drowning, *Medicine, Science, and the Law* **13**, 23–24.

[11] Pollanen, M.S. (1996). The diatom test for drowning in Ontario, *Journal of the Canadian Society of Forensic Science* **29**(4), 205–211.

[12] Pollanen, M.S., Cheug, C. & Chiasson, D.A. (1997). The diagnosis value of the diatom test for drowning. I. Utility: a retrospective analysis of 771 cases of drowning in Ontario, Canada, *Journal of Forensic Sciences* **42**(2), 281–285.

[13] Pollanen, M.S. (1998). *Forensic Diatomology and Drowning*, Elsevier, Amsterdam, pp. 125–147.

[14] Ludes, B., Quantin, S., Coste, M. & Mangin, P. (1994). Application of a simple enzymatic digestion method for diatom detection in the diagnosis of drowning in putrefied corpses by diatom analysis, *International Journal of Legal Medicine* **107**, 37–41.

[15] Ludes, B., Coste, M., Tracqui, A. & Mangin, P. (1996). Continuous river monitoring of the diatoms in the diagnosis of drowning, *Journal of Forensic Sciences* **41**(3), 425–428.

[16] Ludes, B., Coste, M., North, N., Doray, S., Tracqui, A. & Kintz, P. (1999). Diatom analysis in victims tissues as an indicator of the site of the drowning, *International Journal of Legal Medicine* **112**, 163–166.

[17] Rohn, E.J. & Frade, P.D. (2006). The role of diatoms in medicolegal investigations I: The history contemporary science and application of the diatom test for drowning, *The Forensic Examiner* December 22, https://www.highbeam.com/reg/reg1.aspx?full=yes&origurl=/doc/1G1-155870347.html 10–15.

[18] Mueller, B. (1963). [On the problem of the occurrence of diatoms in the organs of cadavers not having Lain in the water.], [Article in German – Original Title: Zyr Frage des Vorkommens von Diatomeen in Organen von Leichen die nicht in Wasser gelegen haben] *Deutsche Zeitschrift fur die Gesamte Gerichtliche Medizin* **54**, 267–272.

[19] Spitz, W.U. (1963). Diagnose des Ertrinkungstodes durch den diatomeen-Nachwies in organen, *Deutsche Zeitschrift fur die Gesamte Gerichtliche Medizin* **5**, 42–45 (cited from Peabody, 1980).

[20] Tyagi, G.D., Dogra, T.D. & Dikshit, P.C. (1985). Diatoms of Delhi, *Journal of Feline Medicine and Surgery* **2**(3), 18–23.

[21] Hurlimann, J., Feer, P., Elber, F., Niederberger, K., Dirnhofer, R. & Wyler, D. (2000). Diatom detection in the diagnosis of death by drowning, *International Journal of Legal Medicine* **114**, 6–14.

[22] Singh, R., Singh, R. & Thakar, M.K. (2006a). Extraction methods of diatoms-a review, *Indian Internet Journal of Forensic Medicine and Toxicology* **4**(2).

[23] Kazutoshi, A.G.O., Mihoko, A.G.O. & Mamoru, O.G.A.T.A. (2004). The distribution of diatoms in Yoronjima and application of the diatom test for the diagnosis of death by drowning in open sea islands, *Medical Journal of Kagoshima University* **56**(2), 25–29.

[24] Singh, R., Singh, R., Singh, R. & Thakar, M.K. (2006b). Diatomological studies from three water bodies of Jaipur, *Indian Internet Journal of Forensic Medicine and Toxicology* **4**, 3 Print ISSN: 0973–1970.

Further Reading

Nanikawa, R. & Kotoku, S. (1974). Medicolegal observations on a dead body drawn up from the sea bed, with special reference to ethanol and diatoms. Case report, *Journal of Forensic Sciences* **3**, 225–232.

Sidari, L., Di Nunno, N., Costantinides, F. & Melato, M. (1999). Diatom test with Soulene 350 to diagnose drowning in sea water, *Forensic Science International* **103**(1), 61–65.

Lunetta, P., Penttila, A. & Hallfors, G. (1998). Scanning and transmission election microscopical evidence of the capacity of diatoms to penetrate the alveoli-capillary barrier

in drowning, *International Journal of Legal Medicine* **111**, 229–237.
Pachar, J.V. & Cameron, J.M. (1992). Scanning electron microscopy: application in the identification of diatoms in cases of drowning, *Journal of Forensic Sciences* **37**(3), 860–866.
Peabody, A.J. (1980). Diatoms and drowning- a review, *Medicine, Science, and the Law* **20**(4), 254–261.
Mueller, B. & Gorgs, D. (1949). Studien uber das Elindringen von corpiscularen Wasserbetandtielen aus den Lungenaveolen, in des Kreislauf wahrend des Ertinkungsvorganges, *Deutsche Zeitschrift fur die Gesamte Gerichtliche Medizin* **39**, 25–71.
Timpermann, J. (1979). Personal Communication, (cited from Peabody, 1980).
Hofmann, E. (1878). *Lehrbuch der Gerichtlichen Medicin ("Textbook of Legal Medicine")*, p. 629.
Auer, A. (1991). Quantitative diatom analysis as a tool to diagnose drowning, *The American Journal of Forensic Medicine and Pathology* **12**(3), 213–218.
Peabody, A.J. (1977). Diatoms in forensic science, *Journal - Forensic Science Society* **17**, 81–87.
Yange, L., Hu, C., Chengxing, W. & Xu, W. (1999). Development of can for destruction of organic material in use for forensic diatom examination, *Forensic Science International* **101**(3), 163–166.

Related Articles

Autopsy

MUKESH KUMAR THAKAR AND RAJVINDER SINGH

Differential Extraction

As an initial step in the investigation of a sexual assault complaint, the complainant undergoes a medical examination to assess for medical trauma and to sample physical or biological evidence resulting from the attack. Commonly seminal staining is located on swabs taken from the complainants clothing, genital area, or body. In addition, the crime scene examination may yield other items of evidence such as bedsheets, condoms, tissues, and scene swabs. The common challenge with these samples is to separate epithelial cells from the spermatozoa. The epithelial cells typically arise from the complainant (such as from the lining of the vagina or mouth, or from the skin surface), whereas the spermatozoa typically arises from the seminal fluid of the perpetrator (assuming they ejaculated). Separation is achieved by exploiting the different biological constitution of the cells present. Spermatozoa have an extremely durable cell membrane containing thiol-rich proteins that are resistant to cell lysis in the absence of a reducing agent. By controlling the introduction of this reducing agent, it is possible to preferentially regulate the lysis of sperm cells and nonsperm (epithelial) cells. This is known as differential lysis, as it is a two-step cell lysis procedure.

Nonsperm cells are lysed first by a conventional lysis buffer (with the assistance of heat and proteinase K). The postlysis supernatant is removed for DNA analysis of the nonsperm fraction. The remaining fraction contains the sperm cells. These are lysed by the use of a reducing agent that breaks the disulphide bonds that form the extracellular membrane of spermatozoa. Dithiothreitol, or DTT, is a commonly used reducing agent [1]. For samples from sexual assaults that do not contain spermatozoa, it is not possible to separate the cellular content via differential lysis and alternative techniques such as Y-chromosome DNA profiling may be required. It is also important to note that the separation process is not 100% efficient and it is possible for epithelial cell DNA to appear in the sperm fraction, and vice versa. Nonetheless, the method has a satisfactory specificity for the typical samples encountered in forensic analysis.

Reference

[1] Budowle, B., Smith, J., Moretti, T. & DiZinno, J. (2000). *DNA Typing Protocols: Molecular Biology and Forensic Analysis*. Eaton Publishing, Natick.

Related Articles

Assault: Sexually Motivated

SIMON J. WALSH

Diminished Capacity *see* Behavioral Science Evidence

Direct Examination of Experts

Nature, Purpose, and Scope

When an expert witness is summoned to give testimony at a judicial proceeding, he is normally not permitted to get on the witness stand and simply tell the court what he knows. Instead, testimony is elicited in the form of questions posed by the litigant, to which inquiries the witness gives answers. This process is called *direct examination*. The prohibition against testifying in the narrative is to allow the opposing side to interpose objections to certain questions if they call for inappropriate or inadmissible answers under the rules of evidence. As with all trial proceedings, rules may vary from jurisdiction to jurisdiction. We discuss those that are most prevalently used in the adversary system.

Questions asked on direct examination must tend to assist the fact finder in understanding the fact and the issues of the case. Information that is irrelevant, incompetent, or immaterial to the issues in the case, cannot be elicited. There may also be countervailing policies that restrict potentially helpful information from being introduced. Thus, information about certain privileged matters may be restricted, as would information that tends to inflame the jury and has been determined to be more prejudicial than probative. The purpose of these restrictions on use of information is to protect the fairness of the judicial process and prevent decisions from being based on conjecture or on evidence of doubtful or limited probative worth.

In addition to restrictions on the content of information sought to be elicited, there are also evidentiary restrictions on the manner in which a question is phrased. These are called *objections* to the form of the question. Eliciting information may not be done on direct examination by asking leading questions. A leading question is one which suggests what the answer ought to be, for example, the question, "and then you identified the defendant as the perpetrator"?

There are, however, a number of exceptions to the prohibition against the use of leading questions. Leading questions can be used in the beginning of a direct examination to orient the witness to the subject matter about which questions are about to be posed. That information is only deemed to be preliminary and therefore not crucial. In many jurisdictions, questions about the qualifications of an expert may also be posed in a leading form. Lawyers are also permitted to ask leading questions when the witness has trouble understanding what is happening: witnesses who are old, feeble, young, or forgetful. Further, the rule against using leading questions is always relaxed or overlooked when questioning is addressed to a hostile witness, or to the adversary person in a civil dispute, or to someone who is closely associated with the adverse party. These witnesses may not freely supply the information that is sought to be elicited.

It also is always appropriate to use leading questions on cross-examination. Indeed, most cross-examiners will ask leading questions because it is the only way they retain some measure of control in seeking to avoid nonresponsive or volunteered statements.

There is one form of a leading question that is always improper, even when asked on cross-examination. It is the asking of questions that are misleading, argumentative, or compound – meaning two questions in one. A typical example of such a compound, argumentative, and improper question of a witness who has been confronted with a prior inconsistent statement would be: were you lying then or are you lying now?

In jurisdictions where rules of evidence still closely approximate the common law, there may be other rules that affect the form of questioning experts. Examples involve when a witness who is testifying orally may use a writing to assist in remembering information. Lawyers who call the expert, typically explain the manner in which present recollection may be refreshed by use of a writing that is not in evidence, or when and how past recollection recorded can be proved.

Expert Witnesses and Opinion Evidence

Some lay witnesses may offer opinion evidence about facts which the witness has observed. Thus a lay witness may be permitted to testify, for example, "I saw the defendant cross the street and, in my opinion, he was drunk." Lay witnesses are hardly ever permitted to offer an opinion that goes to the ultimate issue to be decided. This would be whether negligence existed in a tort case, or an opinion on the guilt of innocence of an accused in a criminal case.

Even if such an opinion were to be rationale based on perception, it is deemed not to be helpful to the fact finder.

What sets experts apart from ordinary fact witnesses is that an expert can not only testify to facts personally known or observed, or to data collected and examined, but will also be allowed a far greater latitude in expressing opinions based on the data collected, as long as the opinions are likely to be helpful to the fact finder. In jurisdictions, where a common law prohibition against "ultimate issue testimony" still exists, expert witnesses are most frequently exempted from that restriction. Thus, under Federal Rules of Evidence in the United States, and in those jurisdictions that have similar provisions, expert opinion evidence is not inadmissible, except for a few narrow exceptions, merely because the opinions embrace the ultimate issue to be decided.

Eliciting the Testimony of the Expert on Direct Examination

The law considers a witness to be an expert if he is shown to be qualified by knowledge, skill, experience, training, or education in a particular profession or occupation. While formal education is not required of experts, and a person can become an expert merely by having gained experience and skill in a particular task, there may be additional qualifications that courts exact for some professionals. Medical doctors, and other highly credentialed researchers and specialists, will be normally permitted to qualify only when they have obtained the required official certification or licensure that may be needed for practicing in their fields or the formal education required by the rules of their professions.

Thus, before a witness is permitted to offer opinion testimony as an expert, it must be shown that the witness is qualified in the discipline in which he proposes to offer testimony. It is the judge who initially makes the decision whether a witness has the necessary qualifications to be an expert. That process follows five distinct but fairly routine steps, consisting of (i) showing that the witness qualifies according to law to give opinion testimony; (ii) how the witness became involved in the case; (iii) how the witness conducted an examination or collected data; (iv) what conclusions were drawn from the findings; and (v) the probative meaning of the conclusions.

1. Qualifying process

The initial process of qualifying the witness is designed to enable the judge to rule that the proffered witness is permitted to give opinion evidence during the direct examination. This process elicits the following information:

(a) personal information about the witness, his employment history, and current position;
(b) formal education, training, and possession of current licensure or certification, if required;
(c) additional continuing education activities and attendance at professional symposia;
(d) membership in professional societies, and participation in society activities or leadership functions (officers, committees, etc.);
(e) specialized training beyond that which is obtained in colleges and universities;
(f) writings, publications, lecturing, and teaching;
(g) extent of the witness's experience in dealing with the specific tasks or examinations that are particularly applicable to the case before the court;
(h) how examinations of the type relevant to the case are conducted and what results may be expected; and
(i) prior testimony in similar cases.

After this information has been elicited, in some jurisdictions, the direct examiner moves that the witness be permitted to qualify as an expert. At this point, the opposing attorney may request permission to *voir dire* the witness. Such questioning, if permitted, will be limited to a cross-examination on the witness's qualifications. After the *voir dire*, the court will decide whether the witness is qualified to offer opinion testimony.

2. The witness's involvement in the case

In a fairly perfunctory manner, questions will next elicit when and how the expert became involved with the case at bar, and through what process evidence was obtained for examination. This stage is important in maintaining an unbroken chain of custody of any physical evidence that may be available.

3. The expert's examination

Questioning next focuses on the actual examination, wherein the witness will not only be asked to explain what was done but also to establish that each and every step of the recommended or required

examination protocol was followed. This is true also of a description of the analysis, comparison, evaluation, and verification process. In discussing the methods used in the analysis, the expert must also be mindful of other methods that were not performed, and the reasons why a certain process was preferred or selected. The witness may likely be cross-examined on any methods that were available but not utilized.

4. The results of the examination

An expert's appearance in court would be meaningless unless he had arrived at a conclusion. This is the part of the direct examination where the ultimate opinion of the expert will be elicited. Most courts require that the opinion be provided to the court with a reasonable degree of professional certainty. Opinions that are merely guesses, or cannot be substantiated with the data, are normally not admissible since they are deemed to be based on conjecture and speculation.

For that reason, it is important for the expert to be able to justify the basis for his opinion in a cogent, understandable manner.

5. The importance of the opinion in the context of the disputed issues

The direct examination testimony will conclude with an assessment of the value of the opinion in relation to the issues, which the fact finder must decide. This varies greatly depending not only on the discipline involved but also on the quantum of evidence that was available for examination. Some forensic disciplines engaged in impression comparisons, profess to express their findings as "matches" or "individualizations." Others merely state that a certain conclusion is likely, possible, improbable, or similar characterizations of the probative impact of the evidence.

Modern courts dealing with highly sophisticated crime laboratory techniques may want to know whether the methodology used is accepted as a standard of the profession. Of equal importance may be whether it has been widely publicized in the literature and subjected to peer review, and whether error rates have been established.

All of these facets of the examination will also be extremely important for the cross-examination of the evidence, which is ready to commence when the direct examiner hands over the witness for questioning by the opposing attorney.

Related Articles

Cross-Examination of Experts
***In Limine* Motions and Hearings**
Ultimate Issue Evidence by Experts

ANDRE MOENSSENS

Disaster Mental Health

Disaster Mental Health

People have always given aid and comfort to each other during times of disaster. However, attempts to structure and professionalize this assistance are fairly recent developments. Disaster mental health, as an evolving field of practice and study, is a collection of interventions and practices that are designed to address incident-specific stress reactions, rather than ongoing or developmental mental health needs. Traditional mental health practice is based on a medical model, with a clinician seeing a patient in an office setting. Disaster mental health introduces a paradigm shift, requiring that practitioners (clinicians and indigenous helpers) work with individuals and whole communities in the field rather than in an office.

This is similar to the clinical conceptualization of community psychology. Disaster mental health practitioners, like community psychologists, are likely to view emotional distress through a sociological lens that focuses on normal experiences rather than on pathological responses following disaster.

Different mental health disciplines (e.g., social work, psychology, and psychiatry) have varying levels of exposure to systemic interventions used in community psychology models and practices. The difference in how disaster mental health practice is viewed is critical to the development of organized systems of intervention that address individual and collective mental health needs after a disaster. If disaster mental health is viewed from the sociological standpoint, intervention systems are more likely to reflect the kinds of supports that people rely upon in day-to-day living. If disaster mental health is approached using a medical model, intervention systems are designed to

identify and treat maladaptive or pathological reactions to disaster. Disaster mental health interventions have evolved from both of these traditions.

Disaster Mental Health Interventions

Disaster mental health interventions typically include screening for symptoms of major disorders, outreach, and public education activities. The goal of these activities is to normalize stress reactions while both identifying those who may be at risk for developing more severe symptoms and avoiding any actions that may induce adverse outcomes [1]. This set of interventions is often called *psychological first aid* (PFA) and is gaining popularity as a skill set that can be taught to anyone and applied in a variety of situations.

The use of the phrase psychological first aid appeared infrequently in journal articles of the 1980s, and was typically described as a clinical intervention. In the 1990s, the American Red Cross began deploying licensed mental health personnel to carry out disaster mental health activities as part of their array of volunteer services offered after a disaster. More recently, the social science literature has begun to reflect discussion about PFA as both a tool for triage used by clinicians and as a set of skills that can be taught to other disaster responders to mitigate or normalize the psychological effects of disaster or a critical incident.

The international community (e.g., World Health Organization; United Nations) refers to the set of activities that make up PFA as psychological support, mental health, or psychosocial programming. The international view seems to predate, yet parallel the emerging United States movement toward dividing PFA into skills that can be carried out by indigenous helpers as well as trained clinicians [2].

Most disaster mental health interventions include the practice of encouraging survivors of disaster-related trauma or grief to talk about their experience. The effectiveness of this practice has been tested and debated in the literature, but is still considered central to all disaster mental health work. The PFA approach encourages people to talk with someone they trust, like a friend or family member. Medical models of intervention that rely on the special expertise of a clinician also encourage people to talk, though in a more guided format such as in the context of a cognitive behavioral approach. Disaster mental health practitioners in the field generally do not provide treatment for disorders; instead they provide triage or screening for problematic symptoms and refer to other clinicians, preferably based in the local community, who can then provide ongoing treatment services.

Research has not shown an undisputable link between most of the interventions presently used in disaster mental health and the prevention of major problems like posttraumatic stress disorder (PTSD) (*see also* **Posttraumatic Stress Disorder**). There have been efforts to gain expert consensus regarding potential best practices, but to date there are no universally accepted standards of care in disaster mental health. The field of disaster mental health has yet to standardize nomenclature and identify specific competencies that workers must have to function effectively as disaster mental health practitioners across jurisdictions. The lack of specific competencies has led to the development of a number of training curriculums, philosophies, and systems across voluntary and nongovernmental organizations that prepare workers to respond to the psychological, social, emotional, and spiritual needs of people after disasters or humanitarian emergencies. The American Red Cross and the US Department of Veterans Affairs National Center for Posttraumatic Stress Disorder both have PFA curricula and guides that are widely used in the United States [3].

The international disaster response community, unlike the United States, has focused less on competencies of clinical responders and more on widespread preparation of indigenous populations to provide psychological support to one another. Clinicians often serve as trainers or supervisors to indigenous helpers in mixed workforce models of service provision.

A broader evolving clinical role in disaster mental health is related to risk communication. This is assisting public officials to construct concise messages that can be relayed to the public via media about the disaster, its risks and potential consequences. Risk communication in the context of disaster response is a mechanism for communicating vital information that may increase compliance with directives, inform the public about common reactions to the event and help people gain some control over their lives after a disaster.

Who Provides Disaster Mental Health?

The division of labor between mental health clinicians and indigenous workers in a disaster mental health response varies according to the availability of clinicians, type and duration of disaster response, culture in which the disaster occurred, and level of involvement of outside entities (e.g., voluntary or nongovernmental organizations). The current lack of accepted standards for preparation of a disaster mental health workforce, both clinical and indigenous, is a glaring gap in the development of organized disaster mental health response. This is compounded by the lack of rigorous research on the effectiveness of interventions commonly used in disaster mental health.

Until there is hard science to support the field, a division of labor between clinical and indigenous personnel will be guided by history, culture, and context. The practice of disaster mental health within an ecological framework recognizes that one part of a system cannot be fully understood in isolation; and that each individual, family, and community's level of distress or resilience is influenced by a complex interplay of systems and events. The use of natural helpers within the local communities to augment the disaster mental health response creates community resilience that may ultimately mitigate negative psychological or social effects of some disasters. The role of indigenous helpers in an organized disaster mental health workforce is typically as culture brokers. They are often peers to those affected by the event (disaster or humanitarian emergency) and are therefore trusted sources of information. In organized systems of intervention, clinicians provide some supervision for indigenous workers.

Legal Issues and Disaster Mental Health Practice

Many disaster mental health practitioners are volunteers. Some are associated with specific aid or relief organizations and enjoy legal protections offered by the organization. Others volunteer their services more spontaneously after a disaster [4]. In the United States, state emergency response statutes typically immunize volunteers from civil suits arising from actions that may even be seen as negligent, so long as their conduct is provided gratuitously in the context of an emergency response. In the case of volunteers, it is likely that clinicians would be more vulnerable to any legal action than indigenous workers. Clinicians who provide supervision to indigenous workers have the additional concern that they may be held responsible for the actions of those they supervise. Clinical supervision under disaster response or relief conditions can be challenging, given the inherent chaos of the situation. Most clinicians prefer not to supervise more than 10 disaster workers at a time so that they can adequately track and manage the type of service being offered in the field.

Disaster mental health practitioners, like traditional purveyors of clinical service, are concerned about potential legal problems related to liability and malpractice. Historically, establishing legal liability for harms caused by mental health practitioners has been difficult. Demonstrating a causal link between a patient's psychological injury and a practitioner's act or omission can be a major obstacle. A central problem is distinguishing between the harm caused to a disaster survivor by virtue of experiencing the disaster *versus* that potentially caused by a practitioner. Additionally, without practices that are widely accepted to guide disaster mental health intervention choices, it is difficult to establish a legally recognized standard of care.

In the United States, Establishing negligence is the dominant legal theory employed to assert liability against mental health practitioners (*see also* **Medical Malpractice**). Negligence is behavior which falls below a legally recognized standard of care employed by a reasonable person in similar circumstances. Negligent conduct is not as culpable as gross negligence or intentional wrong doing, but it can serve as the basis for a successful malpractice lawsuit in many jurisdictions. Establishing negligence, and potentially malpractice, generally involves three factors: a treatment relationship must have existed between a practitioner and patient; the patient must have suffered an actual harm; and the cause of that harm was the practitioner's negligent behavior [5].

Establishing that a treatment relationship exists between disaster mental health practitioners and those they serve is difficult since practice is generally centered on the philosophy of normalizing symptoms rather than treating pathology. In disaster mental health there is also a noticeable absence of billing, medical records, or other formal indicia of such a relationship. The treatment relationship is probably most pertinent for clinicians who serve as the agent accepting referrals from disaster mental health practitioners in the field. Upon creation of such a treatment relationship, the mental health practitioner's

conduct toward the patient – whether that be an act or omission – must fall under the recognized reasonable standard of care for treatment of like conditions [6]. Generally speaking, the more experimental or unproven a treatment is, the less likely it will be considered an acceptable exercise of a professionally recognized standard of care [7].

In addition, in establishing that a treatment relationship existed, a case of malpractice could not proceed unless the plaintiff could prove that they experienced injury or harm [8]. This may take the form of a deteriorating mental or physical health condition, excessive alcohol or drug abuse, job loss or decline in job performance, and divorce or strained familial relationships. All these things could also be considered reactions to the stress of experiencing the disaster event [9]. Connecting such harm to negligent acts or omissions of a disaster mental health clinician instead of to the disaster experience would be difficult. This is known as the *proximate cause* – the cause that directly produced the harm and without which it would not have happened. The etiology of many mental illnesses is still unknown to the medical community. A plaintiff's poor mental health or emotional well-being could be the cause of a practitioner's substandard conduct, the natural development of his condition, or the influence of other factors.

Disaster mental health practitioners enjoy a number of protections from civil liability. These immunities are driven by policy concerns to encourage responders to provide help to people in need without fear of lawsuits. In the United States, Canada, Japan, and some European nations, there are "Good Samaritan" statutes which encourage medical professionals to come to the aid of injured persons. Generally, they shield practitioners from negligence liability if they render care that is free, in good faith, as part of a direct response to emergencies, and does not amount to reckless behavior [10].

In the United States, additional protections from liability exist in state emergency management statutes. Such laws are usually triggered by an official declaration by a state executive. These laws anticipate that some degree of disorder will characterize the immediate aftermath of a disaster and its response. They usually waive professional licensure and regulation requirements for a temporary amount of time in order to facilitate rapid response to an emergency situation. Importantly, they also typically bar civil suits against responders and their organizations that acted in good faith as part of a response effort, so long as their acts or omissions did not constitute reckless behavior or intentional wrong doings.

Determining whether behavior is protected under Good Samaritan laws and immunities in emergency management statutes is a matter of reasonableness. If a responder's actions were consistent with a good faith effort to provide assistance in an emergency context, courts will generally shield such behavior from liability. If, however, a person departs from a good faith and reasonable effort to assist in response activities, or engages in objectively reckless behavior, then liability protection ends.

Legal exposure for disaster mental health clinicians and indigenous workers has not been tested to date. Although the application of disaster mental health interventions is widespread, evaluation of the long-term effectiveness of their various forms of implementation will benefit from ongoing empirical documentation.

References

[1] Everly Jr., G.S. & Flynn, B.W. (2005). *Principles and Practice of Acute Psychological First Aid After Disasters*, Johns Hopkins Center for Public Health Preparedness, Baltimore, MD.

[2] Van Ommeren, M. & Saxena, S. (2004). *Mental Heath of Populations Exposed to Biological and Chemical Weapons*, World Health Organization.

[3] Brymer, M., Jacobs, A., Lyne, C., Pynoos, R., Ruzek, J., Steinberg, A., Vernberg, E. & Watson, P. (2006). *Psychological First Aid: Field Operations Guide*, National Child Traumatic Stress Network and National Center for PTSD.

[4] Hodge, J., Gable, L. & Calves, S. (2005). Volunteer health professionals and emergencies: assessing and transforming the legal environment, *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science* **3**(3), 1–8.

[5] Halleck, S. (1980). *Law in the Practice of Psychiatry: A Handbook for Clinicians*, Plenum Publishing Corporation, New York, pp. 13–54.

[6] Schultz, B. (1982). *Legal Liability in Psychotherapy*, Jossey-Bass Publishers, San Francisco, pp. 1–44.

[7] Scheflin, A. (2000). The evolving standard of care in the practice of trauma and dissociative disorder therapy, *Bulletin of the Menninger Clinic* **64**(2), 197–234.

[8] Simon, R. & Sadoff, R. (1992). *Psychiatric Malpractice: Cases and Comments for Clinicians*, American Psychiatric Press, Inc, Washington, DC, pp. 4–61.

[9] Davidson, R. & McFarlane, A. (2006). The extent and impact of mental health problems after disaster, *Journal of Clinical Psychiatry* **67**(suppl 2), 9–14.

[10] Hodge, J., Gable, L. & Calves, S. (2005). The legal framework for meeting surge capacity through the use of volunteer health professionals during public health emergencies and other disasters, *Journal of Contemporary Health Law and Policy* **22**, 7–71.

Further Reading

Hodge, J., Gable., L. & Calves, S. (2005). Volunteer health professionals and emergencies: assessing and transforming the legal environment, *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science* **3**(3), 1–8.

Inter-Agency Standing Committee (IASC) (2007). *IASC Guidelines on Mental Health and Psychosocial Support in Emergency Settings*, IASC, Geneva.

International Federation of Red Cross and Red Crescent Societies (2003). *Community-based Psychological Support Training Manual*.

National Institute of Mental Health (2002). *Mental Health and Mass Violence: Evidence-based Early Psychological Intervention for Victims/Survivors of Mass Violence. A Workshop to Reach Consensus on Best Practices*, Author, Washington DC.

Scheflin, A. (2000). The evolving standard of care in the practice of trauma and dissociative disorder therapy, *Bulletin of the Menninger Clinic* **64**(2), 197–234.

World Health Organization (2004). *Mental Health of Populations Exposed to Biological and Chemical Weapons*, pp. 1–13.

DENISE BULLING AND TARIK ABDEL-MONEM

Disaster Victim Identification

Introduction

The right to possess an identity is not automatically recognized by national or international law. The "Convention on the Rights of the Child" is the only international convention that addresses the subject of identity as a fundamental human right by proclaiming that childhood is entitled to special care and assistance [1].

Article 7 of the Convention establishes that, from birth, every child has the right to a name and to have that individuality recognized. Article 8 emphasizes the preservation of that identity and the fundamental importance of preserving that name, nationality, and family belonging.

The Geneva conventions [2], consisting of four treaties, set the standards for international law in relation to humanitarian issues. They chiefly concern the treatment of noncombatants and prisoners of war and, therefore, do not provide a legal framework for matters pertaining to the identification of the deceased if the mass fatality event is outwith a combat scenario.

The only formal resolution pertaining to identity in a nonwar situation was ratified by Interpol Standing Committee on DVI (Resolution AGN/65/Res/13 – 1996) [3]. This "recognizes that for legal, religious, cultural, and other reasons, humans have the right not to lose identities after death and that the identification of disaster victims may be of vital importance to police investigations".

Therefore, the identifications that were undertaken in the Former Republic of Yugoslavia as a result of the Balkan conflicts throughout the 1980s and 1990s are governed by the Geneva Conventions [2] and those who lost their lives in the Asian Tsunami have had their identities returned via observance of the Interpol resolution [3]. If the country involved is not an Interpol member, then there is no impediment to the deceased being interred in a mass grave without attempts to establish their identity [4]. Indeed, there is no legal requirement for any country to establish identity of its deceased unless it is written into their own national constitution as the Interpol resolution is not an international law, merely recognition of propriety i.e., a written motion adopted by a body without law making powers.

Legality notwithstanding, there are two important reasons why identification of the deceased is important and should be addressed wherever possible following a mass fatality event. In 1871, William Gladstone reminded us *''Show me the manner in which a nation cares for its dead and I will measure with mathematical exactness, the tender mercies of its people, their respect for the law of the land and their loyalty to high ideals."* Identifying the dead is a core aspect of the humanitarian response to a fatality. The grief of a family, a community, or indeed a nation can rarely be assuaged if the names of the deceased remain unknown. Suspicion, conspiracies, and abandonment of faith result and when these occur they can be of sufficient magnitude to unseat a government. Therefore, disaster victim identification (DVI) is extremely important in the aftermath of

a mass fatality event. Standards and commonality of approach are vital ingredients.

The second important reason for establishing identity is related to legal requirements. This is equally true whether the event is a natural disaster (e.g., hurricane), man made (e.g., nuclear reactor explosion), or terrorist (London bombings). Each will have a specific drive for the rationale of establishing identity as it may be related to an issue of probate, culpability and, therefore, prosecution or perhaps national and international security.

In the United Kingdom, the Civil Contingencies Act [5] recognizes the police service in a "first responder" role. More specifically, the police service is identified as being primarily responsible for the recovery and identification of the deceased. Assistance to undertake these tasks comes from outwith the police service and may include specialist search teams and specialists including pathologists, odontologists, anthropologists, archaeologists, anatomical pathology technologists, radiographers, biologists, etc. An ineffective recovery and identification procedure will most likely lead to a formal enquiry and while it may not be primarily driven by the need to blame, this often becomes a significant media focus. However, an enquiry can have beneficial effects for the discipline of DVI as it serves to assess objectively the operation and can make recommendations to ensure that any mistakes made are not repeated. The enquiry undertaken by Lord Justice Clark [6] into the events of the Marchioness disaster (the sinking of a pleasure boat on the River Thames, London, United Kingdom) have had long reaching beneficial effects on the way in which DVI is approached in the United Kingdom and led to the development of one of the most important roles in a mass fatality event – that of the senior identification manager (SIM). The SIM will always be a serving police officer who will work alongside the senior investigating officer (SIO), their prime role being to address all the issues pertaining to the identification of victims. This ensures that this important role is under the command and direction of one single point of contact.

A series of national and international events between 2004 and 2006 caused The Home Office and the Foreign and Commonwealth Offices of the UK Government to recognize that a national response capability was required to mass disasters. The Association of Chief Police Officers (ACPO) was tasked with developing a UK DVI task force. The vision statement for UK DVI stated "*The Police Service within England, Wales, Scotland, and Northern Ireland will provide an appropriate, effective and professional DVI contribution as part of the overall response to a mass fatality incident, either in the United Kingdom, or where requested by HM Government overseas*" [7].

The DVI Process in the United Kingdom

We assume DVI is primarily composed of those officers who are occupied within the mortuary but the process of DVI is a vast mechanism that must be delicately balanced to ensure that two sets of data are accurately recorded and brought together so that there is a strong likelihood that they will be recognized as a match and identity will be conferred by the appropriate legal authority [8, 9]. It is a simple concept but it is a complex mechanism which has the potential to go wrong at so many stages and the process begins with the Casualty Bureau.

The Casualty Bureau

Casualty Bureau is a complex telephone handling and investigative system. The Bureau is first activated within the police force (known as the *host force*), where the incident has occurred. It issues a single free phone number to the media for dissemination to the public as quickly as possible after the incident. Depending upon the size of the incident, the volume of calls received may exceed immediate host force capability in which case they then request assistance from neighboring forces to activate their Casualty Bureaus until the system is large enough to cope with the initial influx of calls.

When a mass fatality incident takes place, the public thirst for information is often insatiable and unfortunately many calls made are not about those who may have been caught up in the incident and therefore, a filtering mechanism is essential. This is achieved through a primary system whereby callers are initially asked a computerized set of questions to which they may respond by pressing an appropriate key on their phone. In theory, the only people who should ever talk directly to a Casualty Bureau operative are those who are reporting that friends or family may be potentially involved in the incident.

The primary aim of the Casualty Bureau is to enable the police to compile a list of people who

Table 1 Casualty Bureau grading system for missing persons

Assigned misper grade	Criteria
Grade 1	Highly likely that the missing person was involved in the incident
Grade 2	Likely that the missing person was involved in the incident
Grade 3	Unlikely that the missing person was involved in the incident
Grades 4–6	Often not allocated at the outset of the incident. Leaves room for expansion within the system as the incident develops
Grade 7	Missing person was an accounted casualty but not a fatality
Grade 8	Missing person was located safe and well
Grade 9	Missing person was not involved in the incident

are likely to have been involved in the incident. This allows them to target families for the collection of antemortem (AM) information. From the answers to the questions, it is then possible to assess the likelihood of the reported missing person being involved in the incident and the SIM will grade missing person (Misper) information as shown in Table 1.

When the Casualty Bureau receives a phone call reporting someone missing, the system generates a unique reference number (URN) for the reported Misper and a second URN relating to the person who made the call. It should be borne in mind though the public may not be the only source feeding information into the Casualty Bureau. A mass fatality incident often has many injured survivors who will be hospitalized and many may self evacuate if not injured. The rapid set up of a Humanitarian Assistance Centre or Survivor Reception Centre is a good method to ensure that maximum data concerning all those involved in the incident are obtained.

Casualty Bureaus are generally large operations when the mass fatality event is an "open" incident. This refers to a scenario when it is not clear who is most likely to be a victim of the incident e.g., the London tube bombings. It is a smaller affair if the incident is a "closed" event e.g., an airplane disaster where there is a passenger manifest indicating an initial identity for those who might be involved e.g., United Airlines Flight 93.

Eventually, as a result of the filter system provided by the Casualty Bureau, the SIM will be in possession of a list of Grade 1 Mispers that are believed to have been involved in the incident and have not been located as a casualty in a hospital, have not self evacuated or contacted the casualty bureau to confirm their safety. In theory, this should be a list of people who have been killed and whose bodies remain at the scene awaiting recovery by the DVI teams. In practice, however, this is almost never the case, as the list of Grade 1 Mispers will initially be far larger than the number of fatalities. What the list of Grade 1 Mispers provides to the SIM is a list of families to be interviewed by a family liaison officer (FLO) who can then commence the process of harvesting AM information.

Antemortem Information

When the national UK DVI team is invoked, AM data is collected by FLOs using only Interpol yellow AM forms [10]. These forms and instructions on how to complete them can be downloaded from the Interpol website.

The URN assigned to the Misper by the Casualty Bureau will be transferred to the yellow Interpol AM forms. The FLO must harvest as much information as possible about the Misper during the interview process with family members and friends and this will include, but is not limited to, those listed in Table 2. It is important that maximum information is harvested during a single interview to prevent the intrusion of subsequent interviews during this difficult time for family and friends. It is essential that the information is accurate as any mistake made at this stage may result in a missed-match with the appropriate deceased individual. Attention to detail is vitally important [9].

The yellow Interpol DVI Antemortem Form and all AM material is then submitted to the DVI Antemortem Coordination Centre as soon as possible after completion of the interview process and a missing person file is duly opened. In the context of forensic human identification, the following AM materials are the ones most commonly sought for establishing the identity of the deceased as they

Table 2 Information to be secured by the Family Liaison Officer during interview with family or friends of the missing person (this list is not exclusive)

	Information to be secured by the family liaison officer
1.	Any medical and/or dental/odontological X-rays, mouth guards, dentures, etc., in the possession of relatives or friends
2	Names and addresses of any medical practitioners consulted by the missing person/potential victim
3.	Names and addresses of dentists and/or odontologists consulted by the missing person/potential victim
4.	Descriptions of jewelry worn and property usually carried by the missing person/potential victim
5.	Recent photograph/s (showing full face and/or teeth, tattoos, etc.)
6.	Buccal smear or blood sample taken from the biological parents or children of the missing person/potential victim
7.	Any tattoos, scars, piercings, or other significant physical characteristics
8.	Any object that may contain the sole fingerprints and/or DNA of the missing person/potential victim

represent the primary sources of identity as accepted by Interpol [8]:

- AM dental records;
- AM latent prints (usually, but not exclusively, finger prints);
- AM reference or surrogate DNA samples; and
- AM medical records listing a unique medical condition (this equates to, for example, an implant with a unique serial number e.g., hip replacement or pacemaker).

Visual identification of the deceased is not accepted by Interpol as a primary means of identification. This is based on well founded experience which shows that relatives are unreliable at accurate identification of the deceased as on many occasions these have proved to be incorrect [11].

Once the SIM has devised a strategy regarding who should be targeted for the AM harvest, other aspects of this must be addressed. This includes selection of which AM materials are to be utilized, from whom they have to be sourced and how will this be achieved. The primary identifiers to be targeted will be unique to each disaster as they are governed by the potential data that will be retrieved from the remains of the deceased [8].

The FLO must be aware of the benefits of determining the daily habits and lifestyle of the missing person as opportunities to retrieve AM data are unlikely to be restricted to the home address. For example, physical evidence of identity may be retrieved from the work place, sports club, car, hospital histology department, dental surgery, etc.

Dental Records

Original dental records are an essential part of any AM harvest [12]. As well as charting teeth, the dentist, odontologist, or orthodontist may also be a source of additional information e.g., existing gum shields, radiographs bleaching trays, dentures, bridges, crowns, etc [13]. Gum shields and templates for dentures may also prove to be a rich ground for reference DNA sampling. For obvious reasons it is more likely that an adult will have usable dental records for the purposes of identification than will the child and so other means may require to be explored for the nonadult deceased [12]. Copying of dental information onto the yellow Interpol forms should be done under supervision of a forensic odontologist.

DNA

The collection and subsequent veracity of both DNA and latent prints is wholly reliant on the standard of collection. A qualified scene of crime officer should be deployed with the FLO to retrieve this information.

In any AM harvest, recourse should always be made to medical records for the existence of material capable of yielding DNA that can be reliably attributed to the individual. This is frequently referred to as reference DNA, because the source of the material is confirmed [14, 15].

When it is not possible to secure reference DNA for a missing person, attention turns to items believed to have been used exclusively by the missing person and from which useful DNA may be extracted. These may fall into several categories but the most common are as follows:

- toothbrush
- razors
- hairbrush (with hair roots visible)
- gum shields
- jewelry (particularly studs or rings used in piercings).

In these circumstances, the extraction of cellular material from the exhibit and the raising of a DNA profile is not a proof in itself that the profile relates to the person who was believed to have used the item. To verify the reliability of profiles obtained from such items, where possible, DNA buccal swabs should be obtained from parents/children who form a part of the missing person's genetic familial line. Occasions may occur when there are no surrogate items to compare against a familial line, and here the parent/sibling/offspring DNA can be run to assess the potential for matches [16].

Finger Prints

The scene of crime officer must attempt to identify good sources for the missing person's latent prints. Experience has shown that the retrieval of latent marks is a task best undertaken by a specialist fingerprint officer who is guided by intelligence provided by the investigator [8, 17]. One of the first sources to be checked for potential information regarding prints of the missing person should be the criminal records office. It should also be borne in mind that in our increasingly security conscious work environment there are many opportunities to secure biometric data including finger prints from the human resources element of any major company.

Should a putative identification have occurred in the postmortem (PM) environment, it is good practice for the fingerprint expert to gain access to the personal areas of the individual to see if direct matches on fingerprints can be located. In some instances, other latent prints may be targeted e.g., palm prints or foot prints. The latter have proved to be quite effective in the identification of children.

Nonprimary Indicators of Identity

Most judicial authorities will require at least one primary indicator of identity to confirm a match between the deceased and the missing person i.e., DNA, fingerprints, dental information, or a unique medical condition. They are more comfortable if more than one primary indicator is present. However, in certain instances, it is not possible for the identification to be made on primary indicators and so the SIM and the judiciary must agree on how many secondary or indeed tertiary indicators they will require to be present before they will confirm a match between AM and PM data. A list of nonprimary indicators of identity can be found in Table 3.

Table 3 Examples of nonprimary indicators of identity

Secondary	Tertiary
Jewelry	Visual
Personal effects	Photographic superimposition
Distinctive clothing	Generic clothing
Accidental scars	Surgical scars
Birth marks	Body location (e.g., number seat on plane)
Tattoos	
Blood group	
Deformity	
Developmental abnormality	

The aspect of DVI that concerns the Casualty Bureau, the AM harvesting process and the AM coordination center is frequently referred to as the prematching "dry" side of DVI. Therefore, there is a concomitant "wet" side to the process and this encompasses the fields of body recovery and PM harvesting of information pertinent to the identity of the deceased.

Body Recovery

In all disaster scenarios, securing the safety of survivors is paramount and the emergency services rightly give this aspect priority. Once it is clear that there are no more survivors, then the site is secured by the police service and health and safety personnel assess its safety for the recovery process to commence. Every scenario is different and no single generic approach suits every event and therefore, we can only touch on the very basic processes of efficient body recovery in this section.

It is important that this work is carried out by fully trained personnel who have full cognizance of the requirements of victim identification and full understanding of forensic procedures in relation to evidential recovery [8].

The disaster area may be searched by a variety of specialists and teams including search and rescue, health and safety, antiterrorist, air accident, rail accident, etc., prior to body recovery operations. Each body is photographed *in situ* and an ACPO body recovery booklet (with integral URN) is commenced

at this stage. In this way, the URN and recovery booklet remain with the deceased from the point of discovery to the final matching process. The position of the body is recorded and it will be removed to a body holding area where a preliminary visual search may occur that will fast track information in relation to possible identity. The holding area is usually located close to the incident site but in an area that is well screened from public or media observation. Once the resilience mortuary is ready – the bodies are transferred to the mortuary holding area prior to PM examination.

Postmortem Process

The PM process is designed to mirror the nature of the AM information that has been collected as the principal goal to achieve a match between two sets of information – AM and PM [8]. When the UK DVI team is deployed, all PM information is recorded on the pink PM forms which can be downloaded from the Interpol website [10]. The resilience mortuary is directed by the SIM, the mortuary manager, and the lead pathologist.

Triage

A body ready for PM examination is logged out of the holding area and transferred to the first station. This is usually where the ACPO body recovery booklet is checked and the commencement of PM pink documentation commences with transfer of the URN. If items had been identified in the body holding area as being of merit for fast tracking identity, then they will be readily apparent in the body bag and can be brought to the attention of the SIM who will task this information back to the casualty bureau. This may include items such as mobile phones, identification cards, etc.

In the event of fragmentation or disruption of bodies, the presence of an anatomically trained forensic anthropologist in triage can be valuable [18, 19]. They will quickly be able to identify body parts and therefore permit both fast tracking of the part or reconstruction of the individual.

Photography

A full record is created of the PM process from the commencement to the termination of the postmortem process. The photographer will remain with a single body as it passes through several stations in the resilience mortuary and will take photographs of the deceased at every stage, where a change has occurred and of every exhibit or production that is seized from the deceased as the process continues.

Fingerprints

Fingerprints will be recorded from the deceased as will palm prints and depending upon the directive from the SIM, foot prints may also be recorded. The means by which this is undertaken (powder, ink, etc.) will depend on the situation and requirements of the incident [17]. Fingerprints are usually taken prior to removal of clothing to ensure that there is minimal damage to the skin from which the print is being recorded. Loss of the outer layer of the skin (epidermis) does make fingerprinting more difficult, but the sloughed skin may be retrieved perhaps from the body bag and each finger slipped over that of the gloved finger of the officer and a print recovered in that way. In the absence of the epidermis, a print can still be recovered from the deeper dermal layer although it is less reliable as a means of achieving a positive match [20]. The hands of the deceased must be cleaned afterwards as families may wish simply to hold the hand of their loved one before they are buried or cremated.

Radiography

Radiographic imaging of the deceased may be performed in different ways. Fluoroscopy permits a real time image of the deceased and while this might be vital for the purposes of screening for devices that may affect the safety of the forensic teams e.g., explosives, it is not the best medium for visualization of internal structures. Either flat plate radiography or multiple slice CT images are preferred [21–23]. These will highlight the presence of implants and trauma to the pathologist and the anthropologist and may generate images that could at a later date be compared with AM radiographs.

Forensic Pathology and Anatomical Pathology Technology (APT)

The strip search process is supervised by the pathologist and physically undertaken by the APT. This involves the systematic search of the deceased for artifacts prior to removal of clothing. This is undertaken by a skilled APT as it can prove to be extremely

difficult to remove clothing or personal effects (e.g., rings) and they are trained to achieve this with minimal disruption to the deceased. Personal effects and clothing are seized by the exhibits (productions) officers and processed in parallel with the remains of the deceased.

Once the body has been stripped, the proper PM examination will commence. This will start with a detailed search of the surface of the remains in search of scars, wounds, birth marks, tattoos, etc., and will culminate in a full PM examination [8]. Although it may not be necessary to attribute the cause of death to the deceased (e.g., victim of a bomb explosion), it may be necessary to take tissue samples from the driver of a train or the pilot of a crashed plane to satisfy investigative requirements. In addition, the internal examination may well prove vital to confirm the reason behind surface scars, to confirm implantation of devices and their removal to secure a unique serial number. The pathologist will also supervise removal of tissue or fluids for DNA sampling. All information pertaining to the PM examination is relayed to the scribe who completes a full Interpol descriptor on the pink PM forms.

Exhibits/Productions of Personal Effects or Relevant Artifacts

All evidence retrieved from the deceased, whether in relation to personal effects, or evidence related to the incident e.g., bomb shrapnel must be processed through an approved process [8]. This includes photographing and labeling each exhibit/production before recording it in an exhibits book. Some of these images e.g., of jewelry may prove to be important to show to families to aid in the identification process. While one could argue that in an accidental or natural disaster, it is unnecessary for such a formal forensic procedure to be undertaken, it is prudent and rarely is there criticism of excess protocol but there certainly is justifiable criticism of incomplete documentation. All exhibits are also recorded in the pink Interpol PM form.

Forensic Odontology

The odontologists usually work in pairs – one clean to scribe for the dirty partner. The dental remains of the deceased are recorded on the pink forms as are any dental interventions e.g., crowns, bridges, etc. The odontologist will frequently take their own radiographs using specialized equipment that will allow them to directly compare with AM dental surgery radiographs [12]. The whole body radiography usually does not provide sufficient clarity of detail for their requirements. A significant number of photographs will be taken to facilitate comparison with AM photographs as teeth, in a smiling image, may be directly superimposed onto the PM photographs.

Forensic Anthropology

The anthropologist has an important role to play when the bodies are disrupted or severely decomposed. They have the ability to identify small body parts and it should not be confused that they are restricted to identification of bone alone, as anatomically trained anthropologists are equally comfortable at identification of soft tissue body parts. Bodies that have advanced decomposition may require an evaluation of biological identity (sex, age, stature, and ethnicity) performed from the skeletal remains. The anthropologist can also assist the pathologist with the reconstruction of body parts (usually skulls) when ballistic trauma has shattered the normal anatomy and trajectories of missiles need to be evaluated [18, 19].

Upon completion of the PM examination, the body is returned to the mortuary holding area until a match has been confirmed.

Matching Centers

Matching centers bring together all AM and PM information [8]. Depending upon the scale of the operation this may simply be a paper exercise where pink and yellow forms can be directly compared if there are few deceased in the incident. However, with larger events e.g., Asian Tsunami, a paper comparison is unworkable. In this situation, all AM and PM data is logged into a computerized system (the potential for mistakes is high and attention to detail remains at the forefront of operational procedures). The preferred computerized system is Plassdata, which was developed by a Danish company [9, 24]. This system will automatically search all the AM and PM data and highlight potential matches. The computer does not of course recognize the subtleties of identification and therefore any potential matches must be examined by an appropriate expert. If a match is established then a series of quality checks are performed prior to the completion of

a reconciliation report and an Interpol comparison form.

Identification Commission

Depending upon the country of location of the incident, the reconciliation report and the Interpol comparison form will be presented to an identification commission [8]. This is headed usually by a judicial authority of a particular country. The role of the commission is to undertake the legal responsibility for the confirmation of identity of the deceased, its match to a missing person, and ultimately the release of the remains to family and friends.

The Commission will likely include the lead judiciary, the SIM, the lead odontologist, the lead pathologist, and any other experts that it deems necessary to present information pertaining to a particular case. If the judicial authority agrees that a match has occurred, then the body can be released from the mortuary holding area, embalmed (if necessary) and released or repatriated to loved ones.

Any mistake made throughout this process will result in either a missed opportunity to achieve an identification or a misidentification. Should the latter occur then two families are involved, as a body may be released to the incorrect family leaving the second family bereft of their loved one. This has happened in many situations and it is regrettable and to be avoided at all costs. This can be difficult when the conditions of body recovery or in the mortuary may be less than optimal, consumes more time, and mortuary teams tire. These are, however, not acceptable excuses and teams are trained to the highest level of evidentiary standards to prevent mistakes.

The United Kingdom now has training in place for both the "dry" and the "wet" side of operations. Family liaison officers are trained both by the National Policing Improvements Agency (NPIA) and the Metropolitan Police Service (MPS). Body recovery training is provided by NPIA and advanced mortuary skills are taught by the University of Dundee in conjunction with the Centre for International Forensic Assistance (CIFA) and NPIA [25]. Training is not, however, sufficient on its own as there must be regular updates and exercises to ensure that skills are practised and updated.

Although the concept of DVI is relatively simple – comparison of two data sets to secure a match – it must be appreciated that protocols and procedures must be of the highest standard. This is rarely a speedy process, but it must be an accurate process and often it is being undertaken in less than optimal conditions.

Preparation for a disaster has to rely on the "when" and not an "if" a disaster occurs, as they usually arrive with absolutely no warning.

References

[1] UNICEF (1990). *Convention on the Rights of the Child*, [online] Available at: http://www.unicef.org/crc/ (accessed 9th Sept 2008).

[2] Attributed to International Commission of the Red Cross (1949). *Geneva Conventions*, [online] Available at: http://www.icrc.org/Web/eng/siteeng0.nsf/htmlall/genevaconventions (accessed 9th Sept 2008).

[3] Attributed to Interpol (1996). *Interpol Resolution AGN/65/RES/13*, [online] Available at: http://www.interpol.int/public/ICPO/GeneralAssembly/Agn65/Resolutions/AGN65RES13.asp (accessed 18th Jan 2008).

[4] Perera, C. (2005). After the tsunami: legal implications of mass burials of unidentified victims in Sri Lanka, *PLoS Medicine* **2**(6), e185.

[5] HMSO (2004). *Civil Contingencies Act* (c.36), London, [online] Available at: http://www.opsi.gov.uk/acts/acts2004/pdf/ukpga_20040036_en.pdf (accessed 27th Feb 2008).

[6] Clarke, L.J. (2001). *Public Inquiry into the Identification of Victims Following Major Transport Accidents*, [online] Available at: http://www.marchioness-nsi.org.uk/ (accessed: 18th Jan 2008).

[7] Sunderland, G. (2006). *Police Service Disaster Victim Identification (DVI) Team Strategy*, Association of Chief Police Officers of England, Wales and Northern Ireland, *London*.

[8] Interpol (2008). *Disaster Victim Identification Guide*, Interpol, Lyons, [online] Available at: http://www.interpol.com/public/disastervictim/default.asp (accessed 27th Feb 2008).

[9] De Valck, E. (2006). Major incident response: collecting ante-mortem data, *Forensic Science International* **159S**, S15–S19.

[10] Interpol (2005). *Disaster Victim Identification Forms*, Interpol, Lyon, [online] Available at: http://www.interpol.int/Public/DisasterVictim/Forms/Default.asp (accessed 28th Feb 2008).

[11] Lain, R., Griffiths, C. & Hilton, J.M.N. (2003). Forensic dental and medical response to the Bali bombing, *The Medical Journal of Australia* **179**(7), 362–365.

[12] Schuller-Gotzburg, P. & Suchanek, J. (2006). Forensic odontologists successfully identify tsunami victims in Phuket, Thailand, *Forensic Science International* **171**(2–3), 204–207.

[13] Keiser, J.A., Laing, W. & Herbison, P. (2005). Lessons learned from large-scale comparative dental analysis

following the South Asian Tsunami of 2004, *Journal of Forensic Sciences* **51**(1), 109–112.
[14] U.S. Department of Justice (2006). *Lessons Learned From 9/11: DNA Identification in Mass Fatality Incidents*, [online] Available at: http://www.ncjrs.gov/pdffiles1/nij/214781.pdf (accessed 12th Aug 2008).
[15] Graham, E.A.M. (2006). Disaster victim identification, *Forensic Science, Medicine and Pathology* **2**(3), 203.
[16] Prinz, M., Carracedo, A., Mayr, W.R., Morling, N., Parsons, T.J., Sajantila, A., Scheithauer, R., Schmitter, H. & Schneider, P.M. (2007). DNA Commission of the International Society for Forensic Genetics (ISFG): recommendations regarding the role of forensic genetics for disaster victim identification (DVI), *Forensic Science International: Genetics* **1**, 3–12.
[17] Interpol European Expert Group on Fingerprint Identification (IEEGFI) (2008). *Method For Fingerprint Identification*, [online] Available at: http://www.interpol.int/public/Forensic/fingerprints/WorkingParties/IEEGFI/ieegfi.asp?HM = 1 (accessed 9th Sept 2008).
[18] U.S Department of Justice (2005). *Mass Fatality Incidents: A Guide for Human Forensic Identification*, [online] Available at: http://www.ncjrs.gov/pdffiles1/nij/199758.pdf (accessed 9th Sept 2008).
[19] MacKinnon, G. & Mundorff, A.Z. (2007). The World Trade Centre – September 11, 2001, in *Forensic Human Identification. An Introduction*, T. Thompson & S. Black, eds, CRC Press, Baton Rouge.
[20] Hoover, J.E. (2006). The science of fingerprints, classification and uses, in *The Project Gutenberg EBook*, Online Distributed Proofreading Team, Federal Bureau of Investigation, [online] Available at: http://www.gutenberg.org/files/19022/19022-h/19022-h.htm (accessed 9th Sept 2008).
[21] Lichenstein, J.E., Fitzpatrick, J.J. & Madewell, J.E. (1988). The role of radiology in fatality investigations, *American Journal of Radiology* **150**, 751–755.
[22] Sidler, M., Jackowski, C., Dirnhofer, R., Vock, P. & Thali, M. (2007). Use of multislice computed tomography in disaster victim identification-advantages and limitations, *Forensic Science International* **169**, 118–128.
[23] Walsh, M., Reeves, P. & Scott, S. (2004). When disaster strikes; the role of the forensic radiographer, *Radiography* **10**, 33–43.
[24] Attributed to Plassdata, *Plassdata Software* (2008)[online]. Available at: http://www.plass.dk/ (accessed 9th Sept 2008).
[25] Black, S.M., Walker, G., Hackman, L. & Brooks, C. (2008). *Disaster Victim Identification: The Practitioner's Guide*, Dundee University Press.

Further Reading

Royal Thai Police – attributed to Metropolitan Police Service, (2006). *Royal Thai Police Take on Tsunami Victim Identification*, Metropolitan Police News Report, [online] Available at: http://cms.met.police.uk/news/major_operational_announcements/major_incident/royal_thai_police_take_on_tsunami_victim_identification (accessed 27th Feb 2008).

SUE BLACK AND LUCINA HACKMAN

Discovery: Depositions

Introduction

A deposition is a method of obtaining evidence in common law legal systems, prior to a trial or hearing, to assist in the preparation of a legal case. As part of the discovery process, depositions allow each side to obtain information about the evidence relied upon by the opposing side. This enables the parties to prepare their own case, and, if necessary, depose their own witnesses to meet the evidence of the opposing side. A deposition is but one tool of pretrial discovery; some jurisdictions (such as the United States of America) favor depositions, while others favor other forms of disclosure such as affidavits. In England and Wales, for example, depositions are most common in Magistrate Courts[a] and may also be used in cases where persons are dangerously ill and may be later unable to give evidence in person at a trial.[b] Depositions may also be used in cases where a witness is unwilling or unable to attend court to give evidence in a trial or hearing.[c] Depending on the jurisdiction, failure or refusal to attend a deposition may result in the person being fined[d] or ordered to pay any costs resulting from his failure or refusal.[e]

The person giving evidence at a deposition is called a "deponent". The evidence is given under oath, and for this reason, the deponent can be charged with perjury or obstruction of justice if the evidence given is untrue.

A deposition is conducted in the presence of the lawyers for all parties, never in the presence of a jury, and sometimes in the presence of a judge or examiner of the court.[f] The deponent is asked questions by the lawyer for the opposing party, and the deponent's lawyer has the right to object to questions that are irrelevant, inappropriate, vague, misleading, hectoring, and so on. Deponents may also be questioned by the lawyer who retained them, so

as to clarify or respond to the answers elicited by the opposing lawyer. If an issue arises as to whether a deponent should answer a question or not, a judge or court officer may decide the issue.

Depositions are recorded and the transcript of the deposition can be used by all parties to prepare for the legal proceedings. The evidence given in a deposition can be used to file pretrial motions (such as a motion for dismissal of the proceedings), to limit the issues in the subsequent trial (for example, by identifying the areas of agreement between opposing experts), to create an opportunity for settling the case, as evidence in a trial or hearing, and to provide material to be used during examination or cross examination in any subsequent trial or hearing.

Preparing for a Deposition

It is essential that a deponent properly prepare for a deposition. This includes reviewing all relevant notes, case records, calculations, files, drafts, and statements related to the evidence that the witness will give. In some jurisdictions, deponents will be asked to bring their notes, files, or other materials with them to the deposition. This enables them to refresh their memory while giving testimony, and makes it more difficult for them to justify any variations in their evidence, which may appear later in the trial proper.

The evidence given in a deposition can be referred to by lawyers during subsequent legal proceedings. This may be problematic for witnesses whose evidence in the trial is not consistent with the evidence they gave in the deposition. For example, during a cross examination in a trial, the deponent may be asked "Is it not true that in your deposition, you said [xyz]? Was your memory better then? Or were you lying then or now?" It is essential that deponents prepare not only for the deposition but also thoroughly review their deposition evidence before giving evidence in later legal proceedings.

Attending/Giving Evidence in a Deposition

Depositions may occur in courtroom facilities or in other locations such as lawyers' offices. Attending a deposition attracts the same responsibilities as giving evidence in ordinary legal proceedings. All deponents are obliged to honor their oath and expert witnesses must fulfill any duties which are attracted by their special status as experts. For example, the duty of an expert to provide impartial, objective, independent assistance to the court, rather than to the party that called them, prevails in a deposition as much as it does in a trial or hearing.

When attending a deposition, deponents should listen carefully to the questions asked, wait for the entire question, and provide only considered answers. Deponents should not provide unsolicited information or move beyond their area of expertise.

Just as in a trial or hearing, answers given in a deposition are given orally, and gestures or sounds to indicate affirmative/negative should be avoided as they are likely to be unclear on a transcript. Experts, who need to refer to body parts, exhibits, photographs, maps, diagrams, and so forth, should ensure that they use adequate verbal descriptions to convey their evidence.

If a question is unclear, the deponent is entitled to ask for clarification before providing an answer. This is important because the transcript of the deposition will not show that the deponent did not understand the question unless he makes it clear at the time. By answering a question he did not understand, a deponent runs the risk of having to alter or contradict his deposition in any subsequent trial or hearing.

Generally, all questions must be answered (even if the only answer is "I do not know" or "I cannot answer that because it is not within my area of expertise"). A deponent whose lawyer advises them not to answer a question should not answer the question. If the lawyer objects only to the form of the question (for example, because the question is unclear), the deponent should listen carefully to the nature of the objection, and if necessary, ask the opposing lawyer to rephrase the question.

Lawyers for all parties will be assessing the performance of the deponent; appearance, manner, clarity of explanations, confidence, and credibility are issues not only in the trial but also in the deposition.

After a Deposition

After a deposition, a transcript of the evidence is prepared and copies are provided to all parties. If there are reporter errors on the transcript, a deponent may be able to correct them. Errors in the content of the testimony, however, are more difficult to correct

and any such changes made by a deponent may be the subject of comment in the subsequent trial or hearing. A copy of the deposition testimony is filed with the court; if a deponent is later unable to attend the hearing or trial, their deposition evidence may be used instead.[g] Even if the deponent does attend the trial or hearing, and gives more evidence there, their testimony at the deposition is treated as an additional witness statement.

Attending a deposition is an opportunity for the deponent to assess how well he was able to give the evidence. This does not mean that the deponent should alter the substance of the evidence; however, it may be possible for him to work out how to better communicate his scientific, technical, or medical evidence in a legal forum. It is permissible and highly recommended for deponents to review their deposition testimony before giving evidence in the subsequent trial or hearing.

End Notes

[a.] Criminal Procedure Rules 2005 (UK) *s* 28.2.
[b.] Magistrates' Courts Act 1980 (UK) Deposition 105.
[c.] For example, the Crime and Disorder Act 1998 (UK) *Schedule 3* allows Justices of the Peace in England and Wales to compel unwilling witnesses to give evidence in the form of a deposition.
[d.] For example, the Magistrates Court Act 1980 (UK) *s* 97 A.
[e.] For example in the Civil Procedure Rules (UK) Rule 34.10.
[f.] In most jurisdictions within the United States of America it is rare for depositions to be conducted in the presence of a judge. In England and Wales, by contrast, depositions are usually conducted before a judge or examiner of the court.
[g.] For example, Criminal Procedure and Investigations Act 1996 (UK) Schedule 2 Part 2(2).

Related Articles

Discovery of Expert Findings
Discovery: Discovery Motions
Discovery in the United States: Civil Cases
Discovery in the United States: Criminal Cases

RHONDA M. WHEATE

Discovery: Discovery Motions

Introduction

Discovery motions are legal devices used by litigants in adversarial legal jurisdictions for compelling the opposing side to furnish information, which the law either permits or requires them to supply. The practice is highly regulated by statute or court rule in the state and federal court systems of the United States (*see* **Discovery in the United States: Civil Cases**; **Discovery in the United States: Criminal Cases**). This article explores the method used in the United Kingdom in both criminal and civil court cases.

Discovery motions as devices for eliciting information from an opponent are not used in civil law countries.

Discovery Motions in Criminal Proceedings

If the defense has reasonable cause to believe that the prosecution is not providing full disclosure (*see* **Discovery of Expert Findings**; **Discovery: Depositions**), the defense may apply for a court order compelling the discovery.[a] Since the advent of explicit disclosure requirements set out in the *Civil Procedure Rules* (which are expected to precipitate similar disclosure requirements in the *Criminal Procedure Rules*) courts in the United Kingdom have consistently approached disclosure as a significant element of a fair trial. It has been stated that "fairness ordinarily requires that any material held by the prosecution, which weakens its case or strengthens that of the defendant, if not relied on as part of its formal case against the defendant, should be disclosed to the defense. Bitter experience has shown that miscarriages of justice may occur where such material is withheld from disclosure. *The golden rule is that full disclosure of such material should be made*" [1].

Courts in the United Kingdom have also noted decisions of the European Court of Human Rights (ECHR) in which the ECHR has considered matters of disclosure. In any motion for discovery, considerations such as national security, the need to protect witnesses, or to keep secret police methods of investigation of crime, are likely to be weighed against

the needs of the party seeking discovery [2]. The ECHR has also specified additional reasons for which discovery may be denied, in that "the entitlement to disclosure of relevant evidence is not an absolute right. . . . In some cases, it may be necessary to withhold certain evidence from the defense so as to preserve the fundamental rights of another individual or to safeguard an important public interest" [3]. In any case, the ECHR has found that in order for a trial to be considered "fair" as per the requirements of *Article 6* of the *European Convention on Human Rights*,[b] any decisions to withhold material from the defense must be scrutinized by judicial authorities at a national level, on a case-by-case basis [4].

The current position in the United Kingdom in relation to motions for discovery in criminal proceedings appears to support three propositions [5]. First, the accused has a fundamental right to a fair trial, which includes an "equality of arms" between the prosecution and defense. This right entitles the defense, but not the prosecution, to seek discovery under both national law and under *Article 6* of the *European Convention on Human Rights*. The obligation on the prosecution arises even if it has not been specifically sought by the defense. Concomitant with this, the prosecution is obliged to disclose all material in its possession which would tend to undermine the prosecution's case or to assist the case for the defense. Thirdly, the right to disclosure is not absolute and may be curbed by competing interests, which include the public interest and the protection of witnesses. Any decision to withhold information is one which is not solely a matter for the prosecution. Material should not be withheld unless judicial scrutiny has deemed it strictly necessary on all the facts of the case.[c]

Civil Litigation in the United Kingdom

Under the *Civil Procedure Rules* for England and Wales, a party may put to an expert, instructed by another party or a court-appointed expert, written questions about his report. The questions may be put only once, must be put within 28 days of service of the expert's report, and must be for the purpose only of clarification of the report (unless the court gives permission or the other party agrees). Answers to such questions are then part of the expert's report.

End Notes

[a.] Criminal Procedure and Investigations Act 1996 (UK) §.8.
[b.] Adopted under the auspices of the Council of Europe, to which all European states (except Belarus, Kazakhstan and the Holy See) have acceded.
[c.] *Civil Procedure Rules* [1999 (currently up to 47 th update)] (UK) Rule 35.6.

References

[1] *R v H and C* [2004]. UKHL 3.
[2] Edwards and Another v United Kingdom, Application Nos 39647/98 and 40461/98 (2003). 15 BHRC 189.
[3] Edwards and Another v United Kingdom, Application Nos 39647/98 and 40461/98 (2003). 15 BHRC 189 at [53].
[4] *Jasper v UK* [2000]. ECHR 27052/95 at [53].
[5] *Sinclair v Her Majesty's Advocate (Devolution)* [2005]. UKPC D2 (11 May 2005) at [33].

Related Articles

Discovery of Expert Findings
Discovery: Depositions
Discovery in the United States: Civil Cases
Discovery in the United States: Criminal Cases

RHONDA M. WHEATE

Discovery in the United States: Civil Cases

Introduction

The general principles of discovery and disclosure of information in common law countries, excluding the United States, and in civil law countries, has been discussed elsewhere (*see* **Discovery of Expert Findings**; **Discovery: Depositions**; **Discovery: Discovery Motions**). In this article, the focus is on discovery and disclosure in civil litigation in US federal courts only.

"Discovery" is the legal term designating the duty of parties to civil litigation to disclose information about a pending action to the opposing side. While in years gone by, law suits in adversary system jurisdictions were seen as contests between parties during which each side sought to shield its evidence from disclosure so as to be able to surprise the opponent during trial, the modern approach has been to do away with what has sometimes been characterized as "trial by ambush". Current law favors free and full disclosure of all information relating to a case among the parties.

Recognizing that the broad availability of discovery methods may be misused by one side in litigation, so as to seek to overburden the opponent, special provisions have been included to make such abuses more difficult. To avoid misuse of the discovery process by serving upon the opponent excessive and oppressive demands for information of doubtful relevancy, provisions exist for the issuance of protective orders by courts. These provisions can also require the demanding attorneys to make payment of fees and expenses to the party retaining the expert.

Federal Rules of Civil Procedure

The Rules of Civil Procedure have seen significant changes since they were first enacted in 1937. Some major amendments were made to the discovery procedures in 1993, and again in 2006. The most significant rule for forensic experts relating to initially mandated pretrial discovery is contained in Rule 26(a)(2) – Disclosure of expert testimony – which provides the following:

> (A) . . . [A] party shall disclose to others the identity of any person who may be used at trial to present evidence under Rules 702, 703, or 705 of the Federal Rules of Evidence. (*See* **Federal Rule of Evidence 702** and **Expert Opinion: United States**).
>
> (B) Except as otherwise stipulated or directed by the court, this disclosure shall, with respect to a witness who is retained or specially employed to provide expert testimony in the case or whose duties as an employee of the party regularly involve giving expert testimony, be accompanied by a written report and signed by the witness. The report shall contain a complete statement of all opinions to be expressed and the basis and reasons therefor; the data or other information considered by the witness in forming the opinions; any exhibits to be used as a summary of or support for the opinions; the qualifications of the witness, including a list of all publications authored by the witness within the preceding ten years; the compensation to be paid for the study and testimony; and a listing of any other cases in which the witness has testified as an expert at trial or by deposition within the preceding four years.
>
> (C) . . . In the absence of other directions from the court or stipulation by the parties, the disclosure shall be made at least 90 days before the trial date or the date the case is to be ready for trial or, if the evidence is intended solely to contradict or rebut evidence on the same subject matter identified by another party under paragraph (2)(B), within 30 days after the disclosure made by the other party. The parties shall supplement these disclosures when required under subdivision (e)(1).[a]

As can be observed *supra*, subdivision (B) of Rule 26(a) sets out in great detail the type of information that must be contained in an expert's report of analysis, or in the appendices it must provide,[b] which also require that experts keep updating their curriculum vitae as they author new publications and are retained to testify as expert witnesses in other cases during a specified period of time – 10 years and 4 years, respectively.

In addition to the discovery provided by Rule 26(a)(2)(B), it is further provided, in subparagraph (5), that experts may also be compelled in some instances to participate in discovery requests made under other sections, such as

> (5) Methods to Discover Additional Matter. Parties may obtain discovery by one or more of the following methods: depositions upon oral examination or written questions; written interrogatories; production of documents or things . . . ; physical and mental examinations; and requests for admission.

The initial disclosures discussed so far are mandatory in all cases and require no court intervention or judicial order. The parties are required to provide voluntarily the information specified in the Rule. The scope of discovery is also specified in Rule 26. With respect to the counsel's preparation for trial, where experts have been involved either to testify at trial or to merely provide assistance in trial preparation, Rule 26(b)(4) – Trial Preparation: Experts – provides the following:

(A) A party may depose any person who has been identified as an expert whose opinions may be presented at trial. If a report from the expert is required under subdivision (a)(2)(B), the deposition shall not be conducted until after the report is provided.
(B) A party may, through interrogatories or by deposition, discover facts known or opinions held by an expert who has been retained or specially employed by another party in anticipation of litigation or preparation for trial and who is not expected to be called as a *witness at trial* only as provided in Rule 35(b) or upon a showing of exceptional circumstances under which it is impracticable for the party seeking discovery to obtain facts or opinions on the same subject by other means.
(C) Unless manifest injustice would result, (i) the court shall require that the party seeking discovery pay the expert a reasonable fee for time spent in responding to discovery under this subdivision and (ii) with respect to discovery obtained under subdivision (b)(4)(B) of this rule the court shall require the party seeking discovery to pay the other party a fair portion of the fees and expenses reasonably incurred by the latter party in obtaining facts and opinions from the expert.

Part (B), above, deals with the circumstance where a lawyer is informed that an opponent has received technical assistance from a consulting expert who is not expected to testify for the opponent. In that case, disclosure will be mandated only if the requesting party can show that there exists a special need to obtain information from the consulting expert, typically information that cannot be easily obtained in another manner. The provision may have been left intentionally vague to allow the courts to flesh out the Rule's parameters on a case-by-case basis. Elsewhere, the present author has explained under what circumstances this provision may apply:

> Among the exceptional circumstances which ought to warrant discovery are: (1) where the only known expert in the field has been retained by the opposing party; (2) where the expert retained by the opposing side has made an investigation which now, due to a change in the circumstances, can no longer be duplicated by a different expert, as where an analysis has consumed the substance to be analyzed, or materially altered it; and (3) where the information is available through effective discovery but only by the expenditure of excessive time and money and where a significant delay of the trial would be caused [1].

It is not always easy for a lawyer to discover whether an expert was retained in anticipation of providing opinion testimony at trial, or merely to consult on pretrial preparation. An attorney may have hoped to obtain a favorable opinion when he first contacted the expert, only to discover that the opinion received from the expert makes calling him as a witness undesirable. Perhaps the law ought to consider the information and advice given to the retaining attorney as privileged from disclosure in the latter case. To require disclosure would discourage lawyers from seeking as much information about the facts in the case as they possibly can learn, though there may be situations wherein the status of the expert might straddle both the role of witness and of assistant in trial preparation.

It is not uncommon for attorneys to seek off-the-cuff opinions on some aspect of a case by telephoning or consulting, informally, an expert without the conversation resulting in a retainer for either formal advice or an expert opinion related to disputed evidence. Rule 26 does not apply to such instances; it neither mandates nor regulates discovery of information from experts who were informally consulted but not retained.

The 2006 Electronic Discovery Amendments

Courts have dealt with electronic data storage for many years, while deciding disputed discovery issues on an *ad hoc* basis. In 2006, the Rules of Civil Procedure were amended to accommodate more formally the growing importance of the discovery of "electronically stored information" (ESI), a process that is also referred to as e-discovery. These amendments do not specifically mention expert witnesses, though they may have implication for laboratories and expert witnesses, generally, inasmuch as much of the data they consult and the examination results recorded and stored electronically.

Rule 26 includes, in its definition of matters subject to disclosure, "electronically stored information" and "data compilations" as a category of information, which must be disclosed at the outset of the litigation, and provides a procedure to be followed by attorneys when a party contends that the data is

inaccessible because of undue burden or excessive cost.

Other amendments to the rules relate to e-discovery as well. A new subsection added to Rule 16 addresses the scheduling of pretrial conferences, and provides that in its scheduling order a court may address ESI issues. Further, in Rule 33(d), it is made clear that ESI qualifies as a "business record" upon which answers to interrogatories may be based. Rule 34, on production of documents, now provides, as a default method, that ESI will be provided to the requesting party "in the form in which it is ordinarily maintained or in a form or forms that are reasonably usable." Whether this also includes metadata [2] is something that is not clear.

In order to know how to comply with today's e-discovery requirements, many attorneys will need to become far more familiar with computer operations and data storage systems than they were expected to be in years past. In the past, it was not unusual for attorneys trying civil law suits to bring into court boxes upon boxes of folders, files, and documents. In the future, discovery may well take a different form when seeking to obtain the knowledge contained in "gigabytes upon gigabytes of information that might be relevant to a case" [3].

Conclusion

The purpose of disclosure and discovery provisions applicable to civil litigation is to enable attorneys to learn all the information about the case that is possessed by the opposing side and which is not shielded from disclosure by evidentiary privilege rules, such as the attorney–client privilege, or some work-product rule applicable to certain trial preparation efforts of counsel. In civil litigation, disclosure obligations apply equally to plaintiffs and defendants.

End Notes

[a.] Federal Rule of Civil Procedure. Rule 26. General Provisions Governing Discovery; Duty of Disclosure, hereinafter Rule 26.

[b.] It will be observed that an expert's report in a civil case is required to contain far more detail than what is required of an expert in a criminal case (*see* **Discovery in the United States: Criminal Cases**).

References

[1] Moenssens, A.A., Henderson, C. & Portwood, S.G. (2007). *Scientific Evidence in Civil and Criminal Cases*, 5th Edition, Foundation Press, Chapter 1, §1.17, at 67.

[2] Moenssens, A.A., Henderson, C. & Portwood, S.G. (2007). *Scientific Evidence in Civil and Criminal Cases*, 5th Edition, Foundation Press, Chapter 1, § 1.18, at 71. When a document is retrieved in its "native" state, it contains information not visible in an ordinary printout of the text. Beyond the visible text and hidden in some system files are what is termed "metadata," such as the history of the manner in which documents were revised after first draft, and who the authors of the various drafts are. It also keeps a record of deletions believed to be sent to the "Recycle bin." All this additional information leads up to the final "saved" file that is visible when the document is retrieved. Metadata also may include the times when and by whom stored information was accessed. That metadata can be recovered by forensic computer analysts.

[3] Larlson, S.A. & Lipinski, R.L. (2007). e-Discovery: a new approach to discovery in federal and state courts, *Illinois Bar Journal* **96**, 184.

ANDRE MOENSSENS

Discovery in the United States: Criminal Cases

Introduction

Elsewhere, the general principles of discovery and disclosure of information that apply in common law as well as in civil law jurisdictions has been discussed (*see* **Discovery of Expert Findings**; **Discovery: Depositions**; **Discovery: Discovery Motions**). In this article, the focus is on discovery in criminal cases with special emphasis on discovery in federal criminal prosecutions in the United States.

As in civil litigation (*see* **Discovery in the United States: Civil Cases**), the term *discovery* embodies a legal concept delineating the duty of parties to a criminal prosecution to disclose information about a pending action to the opposing side. In the past half century, the trend has been toward free and open disclosure of most relevant information to all sides of the litigation. Discovery of information in criminal cases is not as wide or all-encompassing

as are the rules that pertain to civil litigation. The purposes of discovery are "preventing surprise at trial; narrowing the issues to be tried; and speeding the administration of justice by encouraging settlement (e.g., plea bargaining) of those cases where both sides know the strength or weakness of the evidence" [1].

Though the duty to disclose certain specific items of information about a case is mandated by either statute or court decision, today, much disclosure occurs informally. The informal discovery process starts with pretrial conferences between the lawyers, it continues thereafter at the preliminary hearing, and spans the entire duration of the trial. Today, many prosecutors and defense attorneys freely provide the information, which they know the law requires them to produce, obviating the need to file discovery motions. Resort to the courts needs to be followed only when the parties do not freely cooperate in the discovery process.

In earlier years, the duty to disclose information was a process, whereby information about a criminal prosecution flowed solely from the prosecution to the defense. Currently, however, many obligations of disclosure are mutual, requiring the defense to engage in reciprocal disclosure. While constitutional rights of an accused prevent the prosecution from compelling some specific types of disclosure from the defense [2], the prosecutor's right to obtain information from the defense may have particular applicability where forensic experts retained by the defense have examined evidence, or furnished opinions on technical issues relating to the evidence analyzed.

Constitutional Requirements

While the right to obtain disclosure of information from the prosecution is typically a matter of state or federal statute or court rule that depends for its specifics on the jurisdiction where the trial will be held, the US Supreme Court has overlaid the statutory duty of the prosecution by some universal constitutional mandates that apply regardless of jurisdictional boundaries and statutory provisions. While Congress and state legislatures may allow for additional discovery privileges, the constitutional safeguards afforded to criminal defendants cannot be lessened by local fiat.

While the US Constitution or the Bill of Rights (amendments to the Constitution) do not specifically mention a prosecutor's duty to disclose evidence to the defense, the Supreme Court has read this duty into several specific constitutional guarantees.

The Fifth Amendment's rights of a criminal defendant, as incorporated into the Fourteenth Amendment's due process clause, requires that evidence which exculpates the criminal defendant must be disclosed to the defense in state as well as in federal criminal trials [3]. In *United States v Agurs* [4], the Court clarified that evidence was "material" to guilt or innocence in one of three situations: (i) where the prosecution used false testimony it knew or should have known to be perjured [5]; (ii) where evidence both material and favorable to the defendant was withheld by the prosecution, after the defense had made a motion requesting the disclosure of such evidence [6]; and (iii) where the prosecution withholds other material evidence favorable to the defense.[a]

Federal (Statutory) Discovery Requirements

In addition to constitutional duties to disclose certain evidence to the defense, Federal Rule of Criminal Procedure 16,[b] governs pretrial discovery in federal criminal trials. There are two separate provisions contained in Rule 16 that impact on forensic experts: one deals with experts retained by defendant, and one applies to experts retained or employed by the prosecution.

The right of the defense to discover expert evidence in the possession of the prosecution is covered in Rule 16(a)(1). The text of this portion of the rule, as amended in 1993, provides the following:

> (a) Government's Disclosure.
> (1) Information Subject to Disclosure . . .
> (F) Reports of Examinations and Tests. Upon a defendant's request, the government must permit a defendant to inspect and copy or photograph the results of any physical or mental examination and of scientific test or experiment if:
>
> (i) the item is within the government's possession, custody, or control;
> (ii) the attorney for the government knows–or through due diligence could know–that the item exists; and

(iii) the item is material to preparing the defense or the government intends to use the item in its case-in-chief at trial.

(G) Expert Witnesses. At the defendant's request, the government must give to the defendant a written summary of any testimony that the government intends to use under Rules 702, 703, or 705 of the Federal Rules of Evidence1 during its case-in-chief at trial. If the government requests discovery under subdivision . . . [the reciprocal provisions compelling the defense to furnish its expert's reports to the prosecutor. Editor] and the defendant complies, the government must, at the defendant's request, give the defendant a written summary of the [expert witness] testimony that the government intends to use . . . as evidence at trial on the issue of the defendant's mental condition. The summary provided under this subparagraph must describe the witness's opinions, the basis and reasons for those opinions, and the witness's qualifications.

The right of the prosecution to discover evidence in the possession of the defense is contained in Rule 16(b)(1)(B) and (C). These provisions parallel those requirements imposed upon the defense using identical language, and thus make the discovery requirements reciprocal on both sides with one exception. The prosecution can discover expert information from the defense only if the defendant first requests and obtains disclosure of similar information from the government.

Rule 16 places some restrictions on the defendant's duty to disclose expert opinions by exempting material used by the defense in investigating the case or preparing a defense where the experts will not be asked to testify. These restrictions are somewhat technical. If forensic experts are in doubt as to their obligation under Federal Rule 16, they should seek guidance from the attorney who has retained them.

Forensic experts should also know that the duty to disclose evidence subject to discovery is a continuing one. Indeed, Rule 17 also provides that "A party who discovers additional evidence or material before or during trial must promptly disclose the evidence to the other party or the court. . .".[c]

While the Federal Rules of Criminal Procedure do provide, in Rule 17, for subpoenas to be issued by the court at the request of either side to produce books, papers, data, or other objects, the rules do not provide for the taking of discovery depositions in criminal cases.

State Statutes Regulating Discovery

In addition to the discovery obligations that arise from the US Constitution, all the States have their own disclosure and discovery provisions embodied in statutes or court rules. Many of the state provisions are patterned on Federal Rule 16, though considerable variation from the federal model can be noted. Most states also incorporate the constitutional requirements mandated by the US Supreme Court into the text of the state statutes or rules. Most states, but not all, make the right to obtain discovery by the defense reciprocal on the providing of similar information to the prosecution about defense experts. Some states provide for the right of litigants to depose prospective expert witnesses prior to trial–a right that is currently not available under Rule 16.

State discovery provisions may differ in at least two regards from the federal model:

1. whether expert information is available only upon motion and court order, as was the case under old federal rules, or self-executing upon mere request by the parties, as is the case in current Federal Rule 16; or
2. whether the state has a reciprocal right to discover expert information from the defense after the defendant has sought information from the state.

Because of the wide variety of provisions that exist, it is beyond the scope of this work to list or categorize the laws and rules of all 50 States on disclosure and discovery. As is true for experts retained or employed in federal prosecutions, forensic scientists must seek guidance from their individual state-licensed attorney to know the exact duties which the jurisdiction in which they practice imposes upon them, in regard to the duty to disclose information to the opposing side.

Conclusion

Since all jurisdictions have acquired a voluminous jurisprudence of decided court decisions on many aspects of the right of discovery and the duty to obtain disclosure of information, forensic experts should seek to become familiar with the local rules that apply specifically in their jurisdictions.

End Notes

[a.] This was the fact setting of United States v. Agurs, *supra* note 6, a murder prosecution wherein the defense pleaded self defense, and the state withheld evidence that the murder victim had a criminal record for violent crimes. Information becomes "material," the Court explained, "if the omitted evidence creates a reasonable doubt that did not otherwise exist."
[b.] Federal Rule of Criminal Procedure 16, hereinafter "Rule 16."
[c.] Federal Rules of Criminal Procedure, Rule 16(C) Continuing Duty to Disclose.

References

[1] Moenssens, A.A., Henderson, C. & Portwood, S.G. (2007). *Scientific Evidence in Civil and Criminal Cases*, 5th Edition, Foundation Press, Chapter 1, § 1.09, p. 42.
[2] The Fifth Amendment Privilege against compelled self-incrimination prohibits the prosecutor from requiring a defendant to submit to interrogation if the defendant has invoked his constitutional rights. See, e.g., Miranda v Arizona, 384 U.S. 436 (1966).
[3] Mooney v Holohan, 294 U.S. 103 (1935). The Court held that a criminal conviction procured by the knowing use of perjured testimony must be set aside as violative of the Fourteenth Amendment.
[4] United States v Agurs 427 U.S. 97 (1976).
[5] Miller v Pate, 386 U.S. 1 (1967). A rape-murder prosecution, the United States Supreme Court, when reviewing an Illinois conviction, concluded that a state chemist had falsely testified at trial that defendant's shorts contained bloodstains of the same blood type as that of the victim and that the shorts, instead, contained a red paint stain. The Supreme Court erred in some of these factual assumptions. See The vindication of a prosecutor, *The Journal of Criminal Law, Criminology, and Police Science* **59** 335 (1968), reporting on the subsequent investigation by Illinois State Bar authorities who, after the Supreme Court's opinion was handed down, sought to determine whether the prosecutor should be disbarred for having knowingly elicited perjured testimony. The investigation revealed that, at the trial, the prosecutor had not misrepresented the evidence because the underwear contained both blood and paint, though the blood stain was barely visible years after the original trial when the Bar's investigation was conducted.
[6] Brady v Maryland, 373 U.S. 83 (1963). In a murder prosecution, the prosecutor failed to disclose evidence that the defendant's companion had admitted doing the killing. However, the Supreme Court later qualified the *Brady* decision by deciding the case did not create a generalized constitutional right to discovery in a criminal case. See, Weatherford v Bursey, 429 U.S. 545 (2977).

Related Articles

Expert Opinion: United States
Federal Rule of Evidence 702

ANDRE MOENSSENS

Discovery of Expert Findings

Introduction

Discovery (also termed *disclosure* in some jurisdictions) is a pretrial process by which a party to a legal action seeks to gain access to documents or other evidence held by, or for, another party. Discovery allows the parties to find evidence, confirm evidence, identify issues, identify areas of agreement, and to avoid surprise evidence ("trial by ambush").

In forensic science, discovery is not limited to the expert report prepared for court. Discovery is intended to enable a party to ascertain the underlying method and results which created the opinion and conclusions reported in the expert's statement. In addition, discovery may be necessary to determine the significance of items not seized, examined, tested, or reported in the expert's statement. Discovery is also a means of investigating the validation and accreditation of methods, the training and competence of analysts, and contamination issues. The additional materials sought through the process of discovery may be contained in briefing papers, internal reports, staff training and performance assessments, laboratory records, case files, submission forms, labels, electronic data, apparatus printouts, calculations, notes, drafts, and correspondence.

Expert witnesses in most jurisdictions have a formal duty to keep accurate, up-to-date and thorough records of all dealings which relate to the matters set out in their reports for court. For example, the code of conduct of the Council for the Registration of Forensic Practitioners (CRFP) in the United Kingdom requires its members to "make and retain full, contemporaneous, clear and accurate records of the

examinations conducted, the methods, and the results, in sufficient detail for another forensic practitioner competent in the same area of work to review the work independently" [1].

Discovery in Common and Civil Law Systems

Formal processes of discovery are mainly a feature of *common law* jurisdictions (for example, the United States of America, the United Kingdom, Canada, and Australia), which use an adversarial legal process. Under common law principles, "the law" is sourced not only from legislation but also from precedents laid down by judges in equivalent or higher courts (*stare decisis*). In this system, the parties (rather than the judge) largely control the direction of the matter, the nature of the investigations, the list of witnesses, and the extent to which discovery of information is sought and pursued.

Conversely, legal systems based on Roman and/or Germanic legal traditions (for example, in France, Germany, and Italy) center upon codified laws. In these systems, legal decisions are based on legislative codes rather than on precedents from other courts (although judicial decisions may, in practice, be persuasive). Judges play a more active role in investigating incidents and determining how procedural matters are handled. Discovery in this context is usually actively controlled and investigated by the judge, rather than by the parties. The emphasis in this system is on the substance of the material sought for the investigation, rather than on the form of the discovery procedures.

In Germany, for example, discovery is governed by the *Criminal Procedure Code* (*StPO*) and the *Civil Procedure Code* (*ZPO*). Judges are duty-bound to *investigate* the charges against the accused, by finding and examining whatever material will assist in determining the truth. Formal procedures of discovery and considerations of rules of evidence are of minor importance in this model, compared with their importance in common law systems.

In 1989, Italy attempted a large-scale legislative reform to change the existing inquisitorial system to an "accusatorial" model (a combination of inquisitorial and adversarial features). This may have had some impact on the procedural criminal law relating to discovery; however, frequent subsequent legislative revisions largely seem to have returned to an inquisitorial, judge-led investigation of all relevant material held by expert witnesses for both the prosecution and defense.

Criminal and Civil Law Procedures

Even in a discussion of "discovery of expert findings", a distinction must also be drawn between *criminal* and *civil* matters. While the former is usually the prosecution of an individual by the state (for example, a murder trial), the latter is usually the action of an individual against another individual (for example, a negligence claim by a patient against a doctor).

Discovery in Criminal Matters

In criminal matters, in most jurisdictions, there is a constitution-level standard of protection for the accused. For example, *Article 6* of the *European Convention on Human Rights* [2] – *The Right to a Fair Trial* impliedly enables the accused in a criminal matter to have access to the evidence against him, including the expert evidence. Most of the laws of discovery in criminal proceedings focus on broad and automatic disclosure by the prosecution to the defense. This reflects the inherent imbalance of power between the state and the individual and seeks to bring an equality of arms to the criminal trial [3].

In England and Wales, the *Criminal Procedure and Investigations Act* (*CPIA*) 1996 and the *Criminal Justice Act* (*CJA*) 2003 govern the disclosure of "material of all kinds", which includes "information", "objects of all descriptions", writing, and tape-recorded information. The prosecution bears an *ongoing* duty [4] to disclose material to the defense, which might reasonably be considered capable of undermining the case for the prosecution or assisting the case of the accused [5]. The only legislative exceptions to this general rule of primary disclosure by the prosecution, are for material that a court has ordered it is not in the public interest to disclose, or which was obtained under a warrant issued under *s*2 of the *Communications Act* 1985 [6]. Although the *CPIA* and *CJA* confer responsibility for disclosure onto the prosecution and the police, rather than directly onto the expert witness, it is imperative that expert witnesses enable the prosecution and police to comply with their obligations under the Acts. Therefore, for most expert witnesses called by the prosecution, the range of material required for discovery is wide and the obligation is ongoing.

Discovery is also governed by case law in the United Kingdom, where full disclosure has been termed a *golden rule* of the criminal trial [7]. Cases such as *R v Ward* [8] have firmly established that a duty of disclosure exists irrespective of any request by the defense, it extends to *all* material matters that affect the scientific case relied upon by the prosecution, and that the duty is ongoing. *Ward* placed a clear obligation on expert witnesses who have carried out or know of experiments or tests, which tend to cast doubt upon the opinion being expressed ... to bring records of those tests to the attention of the instructing solicitor [9].

Failure by prosecution experts to disclose material (either to the prosecution or the defense) may result in a delayed prosecution, or cessation of the prosecution altogether, a successful appeal by the defense (that is, a finding that a conviction is unsafe), censure of the expert [10], or the inability to use expert evidence, which was not properly disclosed [11].

In the United Kingdom, primary disclosure by expert witnesses is usually initiated by the prosecution serving the expert's report on the defense. If the defense, or defense-retained experts, require further material from the prosecution, arrangements are made through the prosecution legal team [12].

In contrast, in France, the *Code of Criminal Procedure* enables investigating judges to control the disclosure of "available information helpful for the discovery of the truth," a list of expert witnesses, timely disclosure of expert reports, and so on [13]. Broad provisions go on to provide that a judicial police officer may order any person, public, or private establishment or organization, likely to possess "any documents relevant to the inquiry in progress", including those produced from a computer or data-processing system, to provide them with these documents. Failure to comply as quickly as possible can result in a fine or criminal sanction [14].

Discovery in Civil Matters

For discovery in civil actions, rules are usually set out in legislation and regulations. Arguments ordinarily arise as to the extent of the disclosure sought and the likely burden imposed on the party from whom disclosure is sought.

Disclosure in civil matters in England and Wales is governed by the *Civil Procedure Rules* (*CPR*) [15]. One of the objectives of the *CPR* is to ensure that parties are on an equal footing [16]. This harks back to the "equivalence of arms" theory, which posits that both parties in a legal action are entitled to know what weapons are held by the opposing side and to have the opportunity to arm themselves accordingly. The *CPR* defines "document" broadly, as meaning "anything in which information of any description is recorded" [17]. Disclosure is required of any documents relied upon by a party, and documents that support or adversely affect a party's case or another party's case [18]. The duty of disclosure is ongoing in civil matters; it continues until proceedings are concluded [19]. A party who fails to disclose or permit inspection of documents may not rely on that document in the proceedings [20].

Civil procedure in Germany is governed by the *Civil Procedure Act* (*Zivilprozessordnung, ZPO*), which specifies procedures from prehearing stages to the decision. Parties and witnesses can be sworn to tell the truth, with perjury or false statements being classified as criminal offences. Evidence is adduced in a number of forms, including site inspection, witness testimony (expert and lay), private or public documents, and evidence from the parties. As in criminal disclosure procedures, the evidence-taking procedures in civil matters are controlled by the judge.

References

[1] Council for the Registration of Forensic Practitioners (UK), Code of Conduct, Item 12.

[2] Adopted in 1950 under the auspices of the Council of Europe, to which all European states (except Belarus, Kazakhstan and the Holy See) have acceded.

[3] *Sinclair v Her Majesty's Advocate* (2005). SCCR 446 per Lord Hope.

[4] *Criminal Procedure and Investigations Act* (1996). (UK) *s* 9.

[5] *Criminal Procedure and Investigations Act* 1996 (UK) *s 3(1)(a)* as amended by the *Criminal Justice Act* (2003). (UK) *s* 32.

[6] *Criminal Procedure and Investigations Act* (1996). (UK) *ss 6,7*.

[7] *R v H and Others* (2004). 2 WLR 335 at 14–17.

[8] *R v Ward* (1993). 1 WLR 619.

[9] *R v Ward* (1993). 1 WLR 619.

[10] Crown Prosecution Service, Disclosure: Experts' Evidence and Unused Material – Guidance Book for Experts, http://www.cps.gov.uk/legal/section20/chapter a_a_annex_k.html [accessed 30 Dec 2007].

[11] For example, *Criminal Procedure Rules* (2005). (UK) Rule 24.3.

[12] Crown Prosecution Service, Disclosure Manual Chapter 23 at "FSP actions: access arrangements for the

defence" http://www.cps.gov.uk/legal/section20/chapter_a.html#126 (accessed 30 Dec 2007).
[13] *Code of Criminal Procedure* (France) *Article* 60–2.
[14] *Code of Criminal Procedure* (France) *Article* 60–1.
[15] *Civil Procedure Rules* (UK) *Part 31:* Disclosure and Inspection of Documents.
[16] *Civil Procedure Rules* (UK) *Part 1:* Overriding Objective.
[17] *Civil Procedure Rules* (UK) *Part 31:* 31.4.
[18] *Civil Procedure Rules* (UK) *Part 31:* 31.6.
[19] *Civil Procedure Rules* (UK) *Part 31:* 31.11.
[20] *Civil Procedure Rules* (UK) *Part 31:* 31.21.

Related Articles

Discovery: Depositions
Discovery: Discovery Motions
Discovery in the United States: Civil Cases
Discovery in the United States: Criminal Cases

RHONDA M. WHEATE

Dissocial Personality Disorder *see* Psychopathy

Dissociation *see* Deception: Truth Serum

Dissociative Disorders

Despite the fact that dissociative disorders have been observed and written about since the founding of the disciplines of psychiatry (e.g., Freud [1] (reprinted 1995)) and psychology (e.g. Morton Prince, the founder of the *Journal of Abnormal Psychology* [2] and Pierre Janet [3, 4], they continue to be the focus of surprisingly contentious debate in the forensic setting. In the proposed revision of the Diagnostic and Statistical Manual of Mental Disorders (DSM IV-R), Dissociation is defined as a failure of the usually integrative aspects of identity, memory, perception, and consciousness [5].

Controversy has swirled around the disorder, in part because it is extreme and dramatic, involving significant amnesia and frequent alteration in identity, memory, and consciousness [6–9]. Patients with dissociative identity disorder (DID) experience sudden loss of memory, change from a sad, dependent, and helpless personality state to an angry, demanding hostile one in seconds, and may therefore find themselves in situations they cannot understand, such as awakening in a strange hotel room. One "identity" may inflict physical damage on their body as "punishment" for another, e.g., carving "I hate Mary" onto her forearm with a knife. The amnesia is often asymmetrical, with certain components of their personality structure being aware of what is going on while others are out, but not vice versa. The problem is not, however, that there are actually "multiple personalities" existing in one body, as the old name of the disorder implied, but rather a failure of integration of various aspects of identity, memory, and consciousness.

We are normally "different people" at work and at a party (hopefully), but we have continuity of memory and identity across these differences. Patients with DID do not. Their problem is not that they have more than one identity, but rather that they have less than one. The dissociative disorders involve fragmentation of self rather than a proliferation of selves. One DID patient told me she needed "Krazy glue" to put herselves back together. Other dissociative disorders include dissociative amnesia, in which periods of forgetting too extensive to be explained by ordinary forgetfulness, usually of a traumatic or stressful nature, occur, but not associated with fragmentation of identity [5].

Dissociative fugue involves sudden and unexplained travel coupled with loss of customary identity, without necessary assumption of a new identity [5]. Depersonalization disorder involves feeling detached from one's body – it is primarily a dissociative disorder of perception [5]. Dissociative disorder not otherwise specified involves dissociative symptoms that do not meet full criteria for the disorders described above [5].

Controversies Regarding the Diagnosis of a Dissociative Disorder

Legally, skepticism about the diagnosis of DID or dissociative amnesia has led to contentions that such a diagnosis could fail to meet Frye [10] or Daubert [11] standards. While such objections have been raised, the fact that these diagnoses continue to evolve in debates related to the *DSM* and other authoritative resources speak to their level of endorsement and support within the psychiatric community.

Diagnostic Criteria

DID was formerly called *multiple personality disorder* (*MPD*). Diagnostic criteria in the DSM-III included "The existence of more than one identity or personality state [12]." This seemed to imply belief in the independent existence of multiple personalities within one person. However, the crucial problem is really the failure of integration of various elements of identity, perception, memory, and consciousness. The DSM-IV workgroup on dissociative disorders took care to bring the nosology into conformity with descriptions of other major psychiatric disorders, relying on observable symptoms rather than inferences about mechanisms. Most importantly, it changed the name to DID to emphasize the key problem, which is fragmentation of identity. Also, the word "presence" was substituted for "existence" of "two or more distinct identities or personality states" (p. 529) [5, 13, 14].

Cross-cultural Generalizability

Dissociative symptoms involving alterations in identity, memory, consciousness, and somatic function have been observed in cultures around the world. In many Hispanic cultures, *ataques de nervios* has been associated with alterations in consciousness and physical function [15], and follows a history of physical and sexual abuse and drug-abusing caretakers [16]. Dissociative symptoms comparable to those described in the West have been observed in China [17], Japan [18], where they are often associated with a history of childhood abuse [19], India [20–24], Turkey [25–33], and the Netherlands [34–38], with a population prevalence for all dissociative disorders of 2.3% among young adults [39]. One community study found a surprisingly high prevalence of DID of 1.5% in the United States [40]. Despite this, it is frequently underdiagnosed [41].

Etiology: Trauma and Dissociation

The majority of dissociative disorders, especially DID and dissociative amnesia, have their origins in trauma. The essence of traumatic stress is helplessness and loss of control over one's body. There is growing clinical and some empirical evidence that dissociative symptoms, including alterations in identity, memory, and consciousness, occur as a defense during and after trauma, a means of maintaining mental control just as physical control is lost [42–49]. Furthermore, dissociation of emotions such as numbing of response to trauma may impede cognitive and affective processing of traumatic experience, leading to increased vulnerability to the development of posttraumatic stress disorder [50, 51]. Many people during and in the immediate aftermath of acute trauma report being dazed, unaware of serious physical injury, or experiencing the trauma as if they were in a dream. Sexually or physically abused children often report seeking comfort from imaginary playmates or imagined protectors, or by absorbing themselves in the pattern of the wallpaper. Many rape victims report floating above their body, feeling sorry for the person being assaulted below them. Similar experiences accompanied by physical anesthesia have been reported by corrections officers being beaten during a prison riot [52] and by accident victims [53].

There is strong evidence of a connection between trauma and dissociation [44, 48, 54, 55]. Trauma, especially early in life, repeated, and from relatives or caretakers, predisposes to dissociative disorders [9, 17, 41, 56–59]. DID can be thought of as a chronic, severe form of posttraumatic stress disorder [44, 48, 60]. For example, in a sample of 55 delinquent children, pathological dissociation was found to be associated with a history of intrafamilial, but not extrafamilial, trauma [61]. This suggests that dissociation may especially occur when role conflict with important family figures on whom one depends is involved, a condition that has been described as "betrayal trauma." [62, 63]. Keeping conflicting views of the same person in mind creates considerable tension and confusion and complicates memory storage and retrieval [64, 65].

Does the traumatized child remember the abusing parent as "friend" or "enemy," associated with feelings of warmth or terror? One study demonstrated

that high dissociators have more impaired memory for trauma-related words (e.g., incest) but not for neutral words, than do low dissociators. The high dissociators also had a greater history of trauma and significantly more abuse by a caregiver than did low dissociators. The authors concluded: "These results are consistent with the proposal that dissociation may aid individuals with histories of betrayal traumas to keep threatening information out of awareness" [64] (p. 488).

Controversies Regarding Amnesia for Trauma

Dissociative Amnesia

The most frequent legal question involving dissociative amnesia is the tolling of the statute of limitations. This can be a crucial problem in both criminal and civil cases involving childhood sexual and physical abuse. Typically for adults, the statute of limitations for events that occurred in childhood will have long passed. However, typically courts will allow a claim to be brought if the victim could not have reported the crime(s) within the tolling period due to amnesia. Does functional amnesia for traumatic experiences happen? Considerable controversy about this has emerged, e.g., [66, 67]. While it is true that emotional arousal typically leads an increase, not a decrease, in recall, there is also scientific evidence that blocking emotional arousal will block the emotion-related increase in recall. Cahill *et al.* [68], for example, administered a β-blocker which reduces activity of the sympathetic nervous system (common effects of its activation are increased heart rate and blood pressure) after showing an emotionally arousing film to subjects. The β-blocker interfered with the arousal-related increase in recall. This means that controlling emotional arousal also reduces the effect of arousal on recall.

While voluntary suppression of emotionally laden memories is less likely to be successful than that for neutral stimuli [69, 70], when people are motivated to forget, they are more likely to do so for trauma related than neutral memories [71]. The pressure to forget is greater when children are abused by a trusted caregiver: "a trusted caregiver might represent an unavoidable cue for memory retrieval. The only way to prevent persistent recall of damaging memories would be to adapt internally and to deliberately avoid thinking of such memories – in Freud's terms, to push them away from consciousness. Anderson and Green have now shown that even in the innocuous setting of the laboratory, and with stimuli as trivial as randomly paired words, powerful inhibition can be evoked. How much stronger must this inhibition be for objects central to our thoughts and emotions" (p. 319).

We process such vast amounts of information that we would be immobilized by it if we were consciously aware of even minor portions of our total information stores at the same time. We can function only by being strategically selective and accessing the needed information which is then restored when appropriate use has been made of it. To do otherwise would be like having every stored file in a computer open at once, which would overwhelm the central processor, or having all the contents of one's office file cabinets on the desk at the same time. Memory encoding, storage, and retrieval must involve decision rules that limit access of memories to consciousness.

There is indeed strong evidence that even extremely traumatic events may be forgotten. Leading memory researchers have carefully documented cases of it [72]. Numerous other investigators have reached the same conclusion [73–77]. Williams [78, 79] showed that 38% of a sample of women with documented medical histories of physical and sexual abuse (emergency room records) could not recall the incident when debriefed approximately seven years later. Another 14% reported that for some period of time they had not recalled the traumatic incident.

A recent article by Geraerts *et al.* [80] confirms that people do have discontinuous memories – available at some times and not others (*see* **Recollective Accuracy of Traumatic Memories**). Furthermore, they show that such discontinuous memories ($N = 41$; 37% corroborated) are no less reliable than those that have been held continuously ($N = 57$; 45% corroborated). They reported that independent corroboration of such forgotten and then retrieved memories was comparable to that obtained for continuous memories, with the exception of those memories recovered by a subgroup ($N = 16$) in therapy. The authors concluded: "This finding indicates that discontinuous memories are not, as has sometimes been suggested, inherently unreliable. This idea is supported by recent research showing that people reporting spontaneously recovered memories show a striking tendency to forget prior incidences of remembering when those prior

retrievals have taken place in a different retrieval context. This finding suggests that this group, as a whole, may simply be failing to remember their prior thoughts about a genuine incidence of childhood sexual abuse (CSA) (p. 566)." This study proves that memories of traumatic events can be forgotten for some period of time and then reliably recovered. Excluding the 16 who recovered memories that could not be corroborated in therapy, 41 of 98, or 40% of this sample responding to a newspaper advertisement about an abuse history reported discontinuous memories of their abuse history. If one makes the conservative assumption that only corroborated abuse actually occurred, this means that 15 out of a total sample of 114, or 13%, clearly had amnesia for documented childhood abuse. While this is a minority, it provides further evidence that dissociative amnesia for trauma can and does occur.

Amnesia for trauma can be understood in part as a reflection of the immediate effects of trauma on consciousness. To the extent that attention is narrowly focused during trauma, information may be encoded in rather limited form, for example, the "weapon focus" of assault victims, who can give elegant descriptions of the gun pointed at them, but recall virtually nothing else about the assailant [81]. Dissociation can further isolate memories, by separating them from common associative networks that would make memory retrieval easier [82–84]. There is evidence that congruence in mood between the state in which memories were stored and that in which they are retrieved improves recall [85]. Similarly, another form of salient state dependency involves the dissociative state itself. To the extent that individuals do enter a spontaneous dissociated state during trauma, the memories may be stored in a manner that reflects this state (e.g., narrower range of associations to context). There may be fewer cross-connections to otherwise related memories [82, 86].

Trauma can be understood as a sudden discontinuity in experience. This may lead to a process of memory storage that is similarly discontinuous with the usual range of associated memories. This may explain the "off/on" quality of dissociative amnesia and its reversibility with techniques such as hypnosis [87, 88]. Dissociated information is out of sight, but not out of mind. The information kept out of consciousness nonetheless has effects on it. Thus trauma can elicit dissociation, complicating the necessary working through of traumatic memories. The nature of the acute response to trauma may also influence long-term adjustment to it.

Concern has been expressed that individuals with dissociative disorders, because they tend to be hypnotizable [60, 74, 89–93], are therefore suggestible and subject to deliberate or inadvertent influence that could affect memory [6, 94–98]. All memory retrieval is subject to potential bias, including influence by the nature of the questions asked and other subtle or overt coercive aspects of the inquiry. Highly hypnotizable individuals are more easily influenced by the nature of questioning, whether in or out of formal hypnosis [99]. At the same time, repeated retrieval efforts will in general produce more correct new information [100, 101] as well as more incorrect responses, so repeated efforts to induce recollection can in fact improve it. No simple memory report should be taken at face value, and external corroboration is always helpful in evaluating memory reports.

Testimony following dissociative amnesia has been challenged on the basis of the reliability of formerly "forgotten" or dissociated information [8] p. 4169. For example, George Franklin was prosecuted for the brutal murder of a little girl decades later on the basis of the sudden recollection of his daughter, Eileen Franklin-Lipsker, allegedly because of her glance at her daughter when she was roughly the same age as the victim. Eileen claimed to have suddenly visualized witnessing her father commit the murder. I testified for the defense that while it was possible that she could have dissociated conscious memory of the murder, it was not plausible that there would have been of evidence of her having witnessed it in the form of a changed relationship with her father or willingness to travel with him in the van that was used in the murder. Furthermore, the witness lied about the circumstances of her sudden memory retrieval, initially claiming that it occurred under the influence of hypnosis, then changing her story after a visit to her mother, who was an attorney.

A fourth crucial issue was corroboration. The defense attorney was not allowed to show the jury newspapers that portrayed clearly all the details she had reported to the police, i.e., she reported nothing that only an eyewitness could have know. While Franklin was convicted, the conviction was overturned on constitutional grounds involving his right to present such evidence in his defense ([102] p. 11848). Clearly, testimony based upon information

that has not been held continuously in consciousness will be the subject of intense legal scrutiny.

Controversies Regarding Dissociation and Criminal Responsibility

Does the diagnosis of a dissociative disorder qualify for a not guilty by reason of insanity defense? The diagnosis of a dissociative disorder has been used as a basis for a not guilty by reason of insanity plea [103–105]. There are obvious reasons for concern about the easy application of such an argument [103]. In theory, anyone with a dissociative disorder could claim that some portion of their personality was unaware of what another was doing when a crime was committed. The judge in the Hillside Strangler case, in which a dissociative disorder was claimed as grounds for an not guilty by reason of insanity (NGRI) defense, opined that he was going to take the personality who committed the murders, put him in jail, throw away the key, and whatever the other personalities wanted to do was up to them.

Other than the practical problem of potentially absolving a whole class of psychiatric patients from criminal responsibility, the deeper issue is whether situations can occur, consistent with, for example, the *M'Naughten* standard, in which a person committing a crime is either unaware of the meaning and nature of his act, or unable to know that the act was wrong. In order to hold a defendant legally culpable, the law makes certain assumptions regarding the *mens rea*, the presumption that the perpetrator had the capacity and did indeed plan to conduct a criminal act, such as a murder, with malice aforethought. Typically, some component of a DID patient's personality structure will know what they are doing, and some component may well know it was wrong, but it may not be the same component in control of the person at the moment of the crime. This potential separation is not automatically grounds for an NGRI claim, since potential awareness of the meaning and nature of the act and its implications can exist even in a dissociated state. Courts have ruled that as long as whatever component of the personality meets the prevailing standard of criminal responsibility, the diagnosis of a dissociative disorder is not a mitigating factor [106, 107]. Other rulings have included evidence of cooperation among personality components in committing or covering up the crime [108].

A patient with a relatively unusual form of DID, which had begun not in late childhood but in his twenties, was accused of murdering his girlfriend, shooting her in the face. He had wandered around aimlessly after killing her and turned himself in to the police later that day. When I examined him, I learned that he had one dissociative "alter" personality, that of a former friend who had committed suicide several years earlier. This alter was, unlike the "host" personality, hostile and threatening to the examiner. The story that emerged was that his girlfriend was having an affair with a friend of his, and placed hearts on the calendar in their apartment on days when she had had an assignation.

On the day of her murder, she informed the defendant that she was leaving to see her lover. He went to the closet, took a gun out of a box, and told her he was leaving to kill the boyfriend. She tried to prevent him from leaving, apparently to protect her lover from harm. He then apparently hallucinated the face of his own alter superimposed on his girlfriend's face. The alter reported that he "said" to the defendant, "Go ahead, shoot." He reported shooting what he thought was the alter, thereby killing his girlfriend.

My conclusion at the time was that he did indeed have DID. He was, on formal testing, highly hypnotizable [99], and had clear dissociative symptoms. However, I told the counsel that I could not help provide an NGRI defense, because he knew he was killing someone (the "alter," if not his girlfriend), and he knew it was wrong. Counsel concurred, and did not call me as a witness. However, during the trial the defendant insisted on an NGRI defense. Counsel called another psychiatrist, who testified on the basis of my videotaped interview with the defendant that the condition met the criteria for an NGRI defense, and the court indeed found the defendant not guilty by reason of insanity, much to the surprise of his counsel.

Conclusion

Dissociation is a common response to trauma [42, 46, 77, 89]. Dissociative disorders are rare, but occur around the world, and can profoundly interfere with the integration of identity, memory, consciousness, and perception. Dissociative amnesia for trauma occurs, and may legitimately affect such legal issues

as the tolling of the statute of limitations. Dissociation has been employed as a mitigating defense in criminal cases, with varying success. The difficulty individuals with dissociative disorders have in integrating various aspects of their experience places extra responsibility on professionals helping them to integrate information about the mental processes that surround responses to traumatic experience.

References

[1] Breuer, J. & Freud, S. 1893–95. (reprinted 1995). Studies in Hysteria, in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, J. Strachey, ed, Hogarth Press, London, Vol. 2, pp. 183–251.

[2] Prince, M. (1906). *The Dissociation of a Personality*, Longmans, Green, New York.

[3] Janet, P. (1889). *L'automatisme Psychologique*, Felix Alcan, Paris.

[4] Janet, P. (1920). *The Major Symptoms of Hysteria*, Macmillan, New York.

[5] APA (2000). *Diagnostic and Statistical Manual of Mental Disorders*, Text Revision (DRM IV-R), 4th Edition, American Psychiatric Press, Washington, DC.

[6] Dorahy, M.J. (2001). Dissociative identity disorder and memory dysfunction: the current state of experimental research and its future directions, *Clinical Psychology Review* **21**(5), 771–795.

[7] Kluft, R.P. (1999). Current issues in dissociative identity disorder, *Journal of Practical Psychiatry and Behavioral Health* **5**, 3–19.

[8] Spiegel, D. (1994). Dissociative identity disorder, in *DSM-IV Case Book: A Learning Companion to the Diagnostic and Statistical Manual of Mental Disorders*, R. Spitzer, M. Gibbon, A. Skodol, J. Williams & M. First, eds, American Psyhiatric Press, Washington, DC, London, pp. 56–58.

[9] Spiegel, D. (2006). Recognizing traumatic dissociation, *The American Journal of Psychiatry* **163**(4), 566–568.

[10] Frye v United States (1923).. 54 App. D. C. 46, 293 F. 1013.

[11] Daubert v. MDP (1993). In: Ct. S, editor.

[12] APA (1980). *Diagnostic and Statistical Manual of Mental Disorders*, (**DSM III)**, 3rd Edition, American Psychiatric Association, Washington, DC.

[13] Cardena, E., Lewis-Fernandez, R., Bear, D., Pakianathan, I. & Spiegel, D. (1996). Dissociative disorders, in *DSM-IV Sourcebook*, T. Widiger, A.J. Frances, H.A. Pincus., R. Ross, M.B. First & W.W. Davis, eds, American Psychiatric Association, Washington, DC, Vol. 2, pp. 973–1006.

[14] APA (1994). *American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders*, (DSM-IV), American Psychiatric Press, Washington, DC.

[15] Lewis-Fernandez, R., Garrido-Castillo, P., Bennasar, M.C., Parrilla, E.M., Laria, A.J., Ma, G. & Petkova, E. (2002). Dissociation, childhood trauma, and ataque de nervios among Puerto Rican psychiatric outpatients, *The American Journal of Psychiatry* **159**(9), 1603–1605.

[16] Schechter, D.S., Marshall, R., Salman, E., Goetz, D., Davies, S. & Liebowitz, M.R. (2000). Ataque de nervios and history of childhood trauma, *Journal of Traumatic Stress* **13**(3), 529–534.

[17] Xiao, Z., Yan, H., Wang, Z., Zou, Z., Xu, Y., Chen, J., Zhang, H., Ross, C.A. & Keyes, B.B. (2006). Trauma and dissociation in China, *The American Journal of Psychiatry* **163**(8), 1388–1391.

[18] Berger, D., Ono, Y., Nakajima, K. & Suematsu, H. (1994). Dissociative symptoms in Japan, *The American Journal of Psychiatry* **151**(1), 148–149.

[19] Nagata, T., Kiriike, N., Iketani, T., Kawarada, Y. & Tanaka, H. (1999). History of childhood sexual or physical abuse in Japanese patients with eating disorders: relationship with dissociation and impulsive behaviours, *Psychological Medicine* **29**(4), 935–942.

[20] Adityanjee A. (1992). Multiple personality disorder: a factual error, *The American Journal of Psychiatry* **149**(10), 1417.

[21] Adityanjee A., Raju, G.S. & Khandelwal, S.K. (1989). Current status of multiple personality disorder in India, *The American Journal of Psychiatry* **146**(12), 1607–1610.

[22] Downs, J., Dahmer, S.K. & Battle, A.O. (1990). Multiple personality disorder in India, *The American Journal Psychiatry* **147**(9), 1260–1261.

[23] Saxena, S. & Prasad, K.V. (1989). DSM-III subclassification of dissociative disorders applied to psychiatric outpatients in India, *The American Journal of Psychiatry* **146**(2), 261–262.

[24] Varma, V.K., Bouri, M. & Wig, N.N. (1981). Multiple personality in India: comparison with hysterical possession state, *American Journal of Psychotherapy* **35**(1), 113–120.

[25] Akyuz, G., Dogan, O., Sar, V., Yargic, L.I. & Tutkun, H. (1999). Frequency of dissociative identity disorder in the general population in Turkey, *Comprehensive Psychiatry* **40**(2), 151–159.

[26] Chodoff, P. (1997). Turkish dissociative identity disorder, *The American Journal of Psychiatry* **154**(8), 1179.

[27] Sar, V., Tutkun, H., Alyanak, B., Bakim, B. & Baral, I. (2000). Frequency of dissociative disorders among psychiatric outpatients in Turkey, *Comprehensive Psychiatry* **41**(3), 216–222.

[28] Sar, V., Yargic, L.I. & Tutkun, H. (1996). Structured interview data on 35 cases of dissociative identity disorder in Turkey, *The American Journal of Psychiatry* **153**(10), 1329–1333.

[29] Tezcan, E., Atmaca, M., Kuloglu, M., Gecici, O., Buyukbayram, A. & Tutkun, H. (2003). Dissociative disorders in Turkish inpatients with conversion disorder, *Comprehensive Psychiatry* **44**(4), 324–330.

[30] Tutkun, H., Sar, V., Yargic, L.I., Ozpulat, T., Yanik, M. & Kiziltan, E. (1998). Frequency of dissociative disorders among psychiatric inpatients in a Turkish University Clinic, *The American Journal of Psychiatry* **155**(6), 800–805.

[31] Yargic, L.I., Sar, V., Tutkun, H. & Alyanak, B. (1998). Comparison of dissociative identity disorder with other diagnostic groups using a structured interview in Turkey, *Comprehensive Psychiatry* **39**(6), 345–351.

[32] Yucel, B., Kiziltan, E. & Aktan, M. (2000). Dissociative identity disorder presenting with psychogenic purpura, *Psychosomatics* **41**(3), 279–281.

[33] Zoroglu, S.S., Sar, V., Tuzun, U., Tutkun, H. & Savas, H.A. (2002). Reliability and validity of the Turkish version of the adolescent dissociative experiences scale, *Psychiatry and Clinical Neurosciences* **56**(5), 551–556.

[34] Boon, S. & Draijer, N. (1993). Multiple personality disorder in The Netherlands: a clinical investigation of 71 patients, *The American Journal of Psychiatry* **150**(3), 489–494.

[35] Friedl, M.C. & Draijer, N. (2000). Dissociative disorders in Dutch psychiatric inpatients, *The American Journal of Psychiatry* **157**(6), 1012–1013.

[36] Merckelbach, H., Dekkers, T., Wessel, I. & Roefs, A. (2003). Dissociative symptoms and amnesia in Dutch concentration camp survivors, *Comprehensive Psychiatry* **44**(1), 65–69.

[37] Sno, H.N. & Schalken, H.F. (1999). Dissociative identity disorder: diagnosis and treatment in the Netherlands, *European Psychiatry* **14**(5), 270–277.

[38] Vanderlinden, J., Van Dyck, R., Vandereycken, W. & Vertommen, H. (1993). Dissociation and traumatic experiences in the general population of The Netherlands, *Hospital and Community Psychiatry* **44**(8), 786–788.

[39] Ferdinand, R.F., van der Reijden, M., Verhulst, F.C., Nienhuis, J. & Giel, R. (1995). Assessment of the prevalence of psychiatric disorder in young adults, *The British Journal of Psychiatry* **166**(4), 480–488.

[40] Johnson, J.G., Cohen, P., Kasen, S. & Brook, J.S. (2006). Dissociative disorders among adults in the community, impaired functioning, and axis I and II comorbidity, *Journal of Psychiatric Research* **40**(2), 131–140.

[41] Foote, B., Smolin, Y., Kaplan, M., Legatt, M.E. & Lipschitz, D. (2006). Prevalence of dissociative disorders in psychiatric outpatients, *The American Journal of Psychiatry* **163**(4), 623–629.

[42] Bremner, J.D., Southwick, S., Brett, E., Fontana, A., Rosenheck, R. & Charney, D.S. (1992). Dissociation and posttraumatic stress disorder in Vietnam combat veterans, *The American Journal of Psychiatry* **149**(3), 328–332.

[43] Cardena, E. & Spiegel, D. (1993). Dissociative reactions to the San Francisco Bay area earthquake of 1989, *The American Journal of Psychiatry* **150**(3), 474–478.

[44] Kluft, R.P. (1985). Dissociation as a response to extreme trauma, in *Childhood Antecedents of Multiple Personality*, R.P. Kluft, ed, American Psychiatric Press, Washington, DC, pp. 66–97.

[45] Koopman, C., Classen, C. & Spiegel, D. (1994). Predictors of posttraumatic stress symptoms among survivors of the Oakland/Berkeley, California, firestorm, *The American Journal of Psychiatry* **151**(6), 888–894.

[46] Marmar, C.R., Weiss, D.S., Schlenger, W.E., Fairbank, J.A., Jordan, B.K., Kulka, R.A. & Hough, R.L. (1994). Peritraumatic dissociation and posttraumatic stress in male Vietnam theater veterans, *The American Journal of Psychiatry* **151**(6), 902–907.

[47] Putnam, F.W. (1985). Dissociation as a response to extreme trauma, in *Childhood Antecedents of Multiple Personality*, R.P. Kluft, ed, American Psychiatric Press, Washington, DC, pp. 65–97.

[48] Spiegel, D. (1984). Multiple personality as a posttraumatic stress disorder, *The Psychiatric Clinics of North America* **7**(1), 101–110.

[49] Spiegel, D. (1988). Dissociation and hypnosis in posttraumatic stress disorder, *Journal of Traumatic Stress* **1**(1), 17–33.

[50] Foa, E.B. & Hearst-Ikeda, D. (1996). Emotional dissociation in response to trauma, in *Handbook of Dissociation*, L.K. Michelson & W.J Ray, eds, Plenum, New York, pp. 207–226.

[51] North, C.S. (1999). Psychiatric disorders among survivors of the Oklahoma City bombing, *The Journal of American Medical Association* **282**(8), 755–762.

[52] Hillman, R.G. (1981). The psychopathology of being held hostage, *The American Journal of Psychiatry* **138**(9), 1193–1197.

[53] Noyes, Jr, R. & Kletti, R. (1977). Depersonilization in response to life-threatening danger, *Comprehensive Psychiatry* **18**(4), 375–384.

[54] Vermetten, E., Dorahy, M. & Spiegel, D. (eds) (2007). *Traumatic Dissociation: Neurobiology and Treatment*, American Psychiatric Publishing, Washington, DC.

[55] Coons, P.M., Bowman, E.S. & Milstein, V. (1988). Multiple personality disorder. A clinical investigation of 50 cases, *The Journal of Nervous and Mental Disease* **176**(9), 519–527.

[56] Ehlers, A. (2006). More evidence for the role of persistent dissociation in PTSD, *The American Journal of Psychiatry* **163**(6), 1112.

[57] Nakdimen, K.A. (2006). Dissociative disorder underdiagnosed due to undescriptive criteria? *The American Journal of Psychiatry* **163**(9), 1645; author reply 1646.

[58] Simeon, D., Knutelska, M., Yehuda, R., Putnam, F., Schmeidler, J. & Smith, L.M. (2007). Hypothalamic-pituitary-adrenal axis function in dissociative disorders, post-traumatic stress disorder, and healthy volunteers, *Biological Psychiatry* **61**(8), 966–973.

[59] van der Hart, O. & Dorahy, M. (2006). Pierre Janet and the concept of dissociation, *The American Journal of Psychiatry* **163**(9), 1646; author reply 1646.

[60] Spiegel, D., Hunt, T. & Dondershine, H.E. (1988). Dissociation and hypnotizability in posttraumatic stress disorder, *The American Journal of Psychiatry* **145**(3), 301–305.

[61] Plattner, B., Silvermann, M.A., Redlich, A.D., Carrion, V.G., Feucht, M., Friedrich, M.H. & Steiner, H. (2003). Pathways to dissociation: intrafamilial versus extrafamilial trauma in juvenile delinquents, *The Journal of Nervous and Mental Disease* **191**(12), 781–788.

[62] Freyd, J.J., Deprince, A.P. & Gleaves, D.H. (2007). The state of betrayal trauma theory: reply to McNally – conceptual issues, and future directions, *Memory* **15**(3), 295–311.

[63] Freyd, J.J., Klest, B. & Allard, C.B. (2005). Betrayal trauma: relationship to physical health, psychological distress, and a written disclosure intervention, *Journal of Trauma and Dissociation* **6**(3), 83–104.

[64] DePrince, A.P. & Freyd, J.J. (2004). Forgetting trauma stimuli, *Psychological Science* **15**(7), 488–492.

[65] Freyd, J.J. (1999). Blind to betrayal: new perspectives on memory for trauma, *The Harvard Mental Health Letter* **15**(12), 4–6.

[66] McNally, R. (2003). *Remembering Trauma*, Belknap Press, Cambridge.

[67] Spiegel, D. (2003). The vagaries of memory, *Nature Medicine* **9**(10), 1251–1252.

[68] Cahill, L., Babinsky, R., Markowitsch, H. & McGaugh, J.L. (1994). Beta-adrenergic activation and memory for emotional events, *Nature* **371**, 702–704.

[69] Schacter, D.L. (2001a). *The Seven Sins of Memory: How the Mind Forgets and Remembers*, Houghton Mifflin, Boston.

[70] Schacter, D.L. (2001b). Suppression of unwanted memories: repression revisited? *Lancet* **357**(9270), 1724–1725.

[71] Conway, M.A. (2001). Cognitive neuroscience: repression revisited, *Nature* **410**(6826), 319–320.

[72] Kritchevsky, M., Chang, J. & Squire, L.R. (2004). Functional amnesia: clinical description and neuropsychological profile of 10 cases, *Learning and Memory* **11**(2), 213–226.

[73] Briere, J. & Conte, J. (1993). Self-reported amnesia for abuse in adults molested as children, *Journal of Traumatic Stress* **6**(1), 21–31.

[74] Bryant, R.A., Guthrie, R.M. & Moulds, M.L. (2001). Hypnotizability in acute stress disorder, *The American Journal of Psychiatry* **158**(4), 600–604.

[75] Classen, C., Koopman, C. & Spiegel, D. (1993). Trauma and dissociation, *Bulletin of the Menninger Clinic* **57**(2), 178–194.

[76] Schacter, D.L., Koutstaal, W. & Norman, K.A. (1996). Can cognitive neuroscience illuminate the nature of traumatic childhood memories? *Current Opinion in Neurobiology* **6**(2), 207–214.

[77] Spiegel, D. & Cardena, E. (1991). Disintegrated experience: the dissociative disorders revisited, *Journal of Abnormal Psychology* **100**(3), 366–378.

[78] Williams, L.M. (1994). Recall of childhood trauma: a prospective study of women's memories of child sexual abuse, *Journal of Consulting and Clinical Psychology* **62**, 1167–1176.

[79] Williams, L.M. (1995). Recovered memories of abuse in women with documented child victimization histories, *Journal of Traumatic Stress* **8**, 649–673.

[80] Geraerts, E., Schooler, J.W., Merckelback, H., Jelicic, M., Hauer, B.J.A. & Ambadar, Z. (2007). The reality of recovered memories: corroborating continuous and discontinuous memories of childhood sexual abuse psychological, *Science* **18**(7), 564–568.

[81] Loftus, E.F. & Burns, T.E. (1982). Mental shock can produce retrograde amnesia, *Memory and Cognition* **10**(4), 318–323.

[82] Hilgard, E. (1977). *Divided Consciousness: Multiple Controls in Human Thought and Action*, Wiley-Interscience, New York.

[83] Kihlstrom, J.F. (1987). The cognitive unconscious, *Science* **237**(4821), 1445–1452.

[84] Kihlstrom J. & Hoyt I. (1990). Repression, dissociation and hypnosis, in *Repression and Dissociation: Implications for Personality Theory, Psycholpathology and Health*, J. Singer, ed, Chicago University Press, Chicago.

[85] Bower, G.H. (1981). Mood and memory, *The American Psychologist* **36**(2), 129–148.

[86] Evans, F.J. & Kihlstrom, J.F. (1973). Posthypnotic amnesia as disrupted retrieval, *Journal of Abnormal Psychology* **82**(2), 317–323.

[87] Lowenstein, R.J. (1991). Psychogenic amnesia and psychogentic fugue: a comprehensive reveiw, in *American Psychiatric Press Review of Psychiatry*, A. Tasman & S.M. Goldfinger, eds, American Psychiatric Press Review of Psychiatry, Washington, DC, Vol. 10, pp. 189–222.

[88] Spiegel, H. & Spiegel, D. (1978). *Trance and Treatment: Clinical Uses of Hypnosis*, American Psychiatric Press, Washington, DC.

[89] Butler, L.D., Duran, E.F.D., Jasiukatis, P., Koopman, C. & Spiegel, D. (1996). Hypnotizability and traumatic experience: a diathesis-stress model of dissociative symptomatology, *The American Journal of Psychiatry* **153**(7), 42–63.

[90] Carlson, E.B. & Putnam, F.W. (1989). Integrating research on dissociation and hypnotizability: are there two pathways to hypnotizability? *Dissociation* **2**, 32–38.

[91] Frankel, F.H. (1990). Hypnotizability and dissociation [see comments, *The American Journal of Psychiatry* **147**(7), 823–829.

[92] Frischholz, E.J., Lipman, L.S., Braun, B.G. & Sachs, R.G. (1992). Psychopathology, hypnotizability, and dissociation [see comments, *The American Journal of Psychiatry* **149**(11), 1521–1525.

[93] Maldonado, J.R. & Spiegel, D. (1998). Trauma, dissociation and hypnotizability, in *Trauma, Memory and Dissociation*, R. Marmar & D. Bremmer, eds, American Psychiatric Press, Washington, DC.
[94] Barnier, A.J., Bryant, R.A. & Briscoe, S. (2001). Posthypnotic amnesia for material learned before or during hypnosis: explicit and implicit memory effects, *The International Journal of Clinical and Experimental Hypnosis* **49**(4), 286–304.
[95] Quas, J.A., Goodman, G.S., Bidrose, S., Pipe, M.E., Craw, S. & Ablin, D.S. (1999). Emotion and memory: children's long-term remembering, forgetting, and suggestibility, *Journal of Experimental Child Psychology* **72**(4), 235–270.
[96] Bryant, R.A. & Barnier, A.J. (1999). Eliciting autobiographical pseudomemories: the relevance of hypnosis, hypnotizability, and attributions, *The International Journal of Clinical and Experimental Hypnosis* **47**(4), 267–283.
[97] Pezdek, K. & Roe, C. (1997). The suggestibility of children's memory for being touched: planting, erasing, and changing memories, *Law and Human Behavior* **21**(1), 95–106.
[98] Spiegel, D. (1995). Hypnosis and suggestion, presented at Memory Distortion Conference, in *Memory Distortion: How Minds, Brains, and Societies Reconstruct the Past*, D.L. Schacter, J.T. Coyle, G.D. Fishback, M.-M Muesulam, L.E. Sullivan, eds, Harvard University Press, Cambridge, Chapter 4, pp. 129–149.
[99] Spiegel, H. & Spiegel, D. (2004). *Trance and Treatment: Clinical Uses of Hypnosis*, American Psychiatric Publishing, Washington, DC.
[100] Erdelyi, M.H. & Kleinbard, J. (1978). Has Ebbinghaus decayed with time? The growth of recall (hypermnesia) over days, *Journal of Experimental Psychology. Human Learning and Memory* **4**, 275–289.
[101] Klatzky, R.L. & Erdelyi, M.H. (1985). The response criterion problem in tests of hypnosis and memory, *The International Journal of Clinical and Experimental Hypnosis* **33**(3), 246–257.
[102] People v Franklin. (1993). San Mateo Superior Court, No C-24395.
[103] Aldridge-Morris, R. (1989). *MPD. An Exercise in Deception*, Lawrence Erlbaum, Hillside.
[104] Alexander, G.J. & Scheflin, A.W. (1998). *Law and Mental Disorder*, Carolina Academic Press, Durham.
[105] Kluft, R.P. (1987). First-rank symptoms as a diagnostic clue to multiple personality disorder, *The American Journal of Psychiatry* **144**(3), 293–298.
[106] Kirkland vs. S. 304 S.E. 2d 561 (Ga. Ct. App., 1983).
[107] Ohio v. Grimsely, Ohio App 3d 265 (1983); also known as 444 N.E. 2d 1071 (1983).
[108] State v Mack (1980). 292 N.W. 2d 764 (Minn).

DAVID SPIEGEL

Distance: Shooting *see* Shooting Distance: Estimation of

Diversion of Mentally Ill Accused *see* Mental Health Courts

Divorce *see* Parental Alienation

DNA

Introduction

Forensic DNA profiling combines scientific disciplines of anatomy, cell and molecular biology, genetics, mathematics, and statistics. This article aims to explain some of the fundamental science underpinning the use of DNA in a forensic context. In particular, it examines the natural attributes of DNA that contribute to its suitability for use in the forensic field and a summary of the practical tools routinely applied to extract the vital information encoded within human DNA.

Fundamental Human Anatomy

The human body is a universe of working parts and functional interactions. We observe the physical manifestations of these interactions all the time, as we walk, talk, breathe, think, or eat. These gross or macroscopic functions of the human body, however, are driven by extremely complex interactions, occurring at the cellular and subcellular level.

Our bodies are made up of trillions of cells. Each cell has a prescribed function relative to its position in the body that is essential to healthy human life. The cells themselves are extremely complex and advanced

pieces of biological machinery. A cell is comprised of a cytosol, which is bound by a permeable membrane, and contains a host of miniature organs (organelles) including the nucleus. The nucleus contains DNA – the material that prescribes the cell's principal functional characteristics.

It may be useful to think of the cell as being like a factory. Membranes enclose the structure and separate different organelles, which can be thought of as departments with specialized functions. The nucleus is the central administration, containing in its DNA a library of information that determines cellular structure and processes. From it instructions are issued for proper regulation of the business of the cell. The mitochondria are the power generators. The cytosol can be thought of as the general work area, where protein machinery (enzymes) carries out the formation of new molecules from imported raw materials. There are special molecular channels in the membranes between compartments and between the cell and its external surroundings. These monitor the flow of molecules in the appropriate directions, similar to personal assistants and receptionists. Like factories, cells tend to specialize in function. For example, many of the cells in higher organisms are largely devoted to the production and export of one or a few molecular products.

Despite the diverse functions of the different types of cells that constitute the human body, each nucleated cell contains an identical copy of a common DNA molecule from which genetic information is read in a linear fashion. Since the amount of information needed to specify the structure and function of a multicellular organism such as a human is immense, the DNA molecule is extremely long. In fact, if the DNA from a single human cell were stretched end to end it would extend approximately 2 m.

Human cells live by a well-defined life cycle, itself, separated into several distinct phases. Similarly, the process of eukaryotic cell division can be divided into principle phases. Cell division can occur by mitosis that produces two diploid cells, or meiosis, which produces four haploid cells. Mitosis results in the production of two daughter cells that are identical to each other and to the parent cell and is the process of somatic cell replication (Figure 1). Our genetic information is carried to the next generation by the haploid cells known as *gametes* (spermatozoa in the male and the oocyte in the female). These cells are produced by meiosis during gametogenesis (or specifically, spermatogenesis or oogenesis; Figure 2). In either mitosis or meiosis, the DNA (on the chromosomes) must be duplicated and distributed to each daughter cell. An error in DNA replication is called

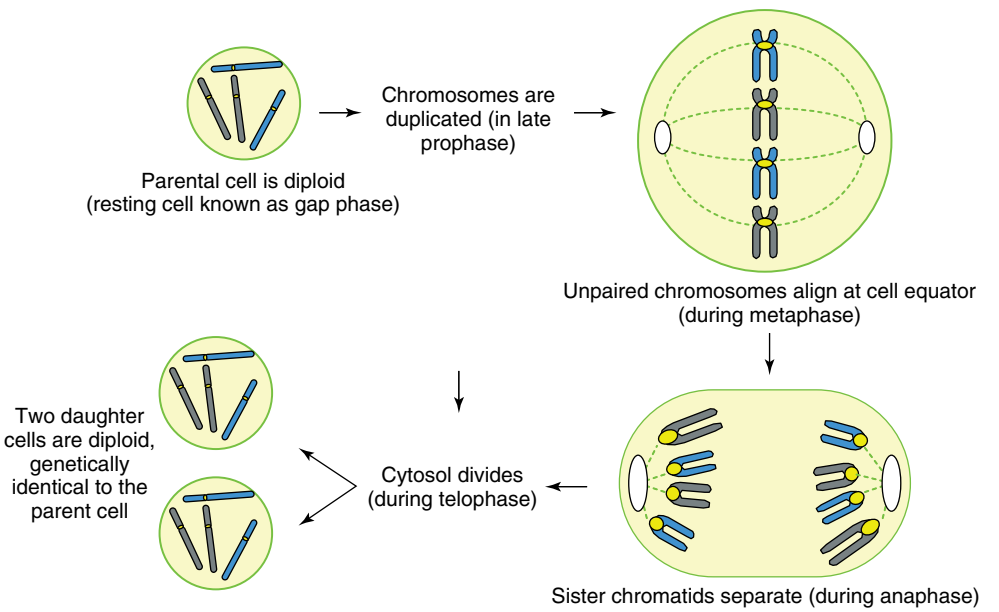


Figure 1 Diagrammatic representation of mitosis, the process of cell replication for adult or somatic cells

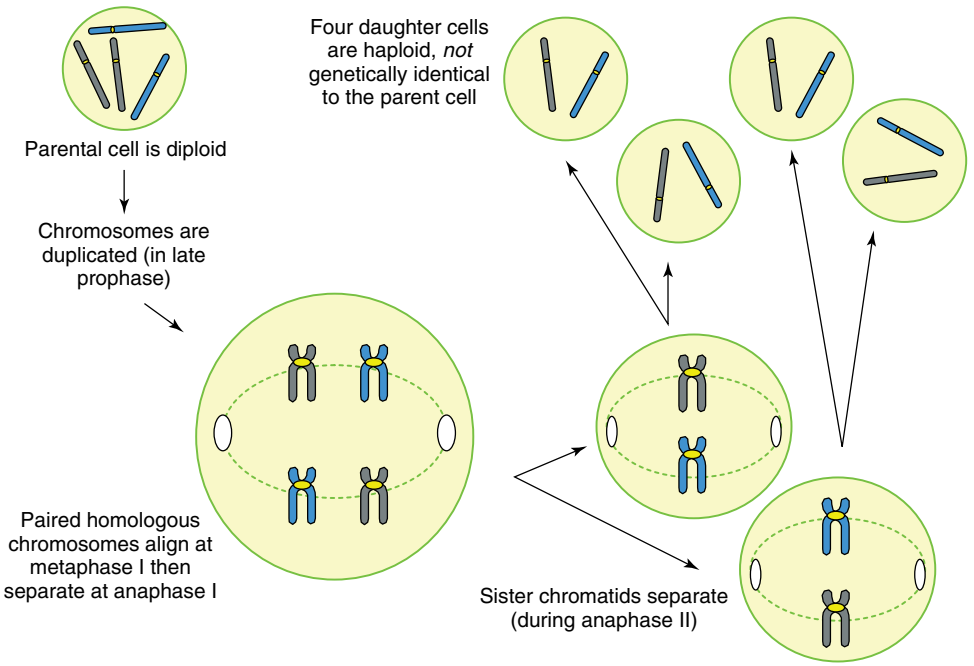


Figure 2 Diagrammatic representation of meiosis, the process of cell replication for germline cells or gametes

a *mutation*. If mutation occurs during replication of an adult cell only progeny cells within that person/organism are affected. This is known as *somatic mutation*. If mutation occurs in the gametes (or germinal cells) there is a possibility that it will be passed onto the offspring of the person/organism. This is referred to as a *germline mutation*.

This summary of basic cellular biology illustrates the first point as to the applicability of DNA analysis for forensic purposes. A replica DNA molecule exists in all nucleated cells in the human body. This means that cells existing in any individual's tissues, fluids, or organs carry a copy of the same DNA molecule. This means there are trillions of possible sources of DNA that can be targetted for forensic purposes. It also means that DNA can be cross-compared among various tissue types from the same donor.

The heritable characteristic of DNA is another advantage to its use in the forensic sciences. Familial, and particularly parental, relationships can provide information to the origin of a given DNA profile due to the known presence of inherited characteristics. This application of DNA profiling technology is often used in the identification of human remains or in the investigation of crimes involving disputed paternity.

Fundamental Genetics

A striking attribute of a living cell is its ability to transmit hereditary properties from one cell generation to the next. This power of self-replication is often attributed as the defining difference between the living and the nonliving. Since the beginning of human history, people have wondered how traits are inherited from one generation to the next. Although children tend to resemble one parent more closely than the other, most offspring tend to be a blend of the characteristics of both parents. The idea of a gene, a unit of hereditary information, arose in the mid-nineteenth century from the famous work of an Augustinian monk named Gregor Mendel. Predominantly working in agricultural science, Mendel found that by beginning with "parents" of known genetic background one had a baseline against which the pattern of inheritance could be measured.

The total genetic information carried by a cell, all the DNA in the nucleus, is referred to as the *genome*. Strictly speaking, this DNA is referred to as *nuclear DNA*. A small amount of the cells' total DNA content exists outside the nucleus, in the

mitochondria. Mitochondrial DNA (mtDNA) is also widely applied in a forensic context.

The genome of higher species is packaged into bundles known as *chromosomes*. Most nucleated human cells are diploid, meaning that the overall content of 46 chromosomes actually comprises one pair of each of the 23 chromosomes (two copies of each chromosome). Exceptions include the sex cells or gametes (spermatozoa or ova), which are haploid (a single copy of each chromosome). The human chromosomes are numbered from 1 to 22 according to their size (chromosome 1 being the largest, 2 the second largest, and so on). These 22 chromosomes are referred to as *autosomes*, as they do not play a role in sex-determination. The 23rd pair is the X and Y chromosomes that play a direct role in sex determination. This chromosomal pair is often referred to as *nonautosomes*.

During conception, parental haploid gametes fuse and form a progenitor diploid cell that is the origin for embryonic development. Owing to this diploidy, individuals who are DNA profiled show either one or two alleles at each DNA site (or locus). If an individual shows one allele, this implies that there are two copies of the same allele present on each of the chromosomes of that particular pair. This event is referred to as *homozygosity*, and we refer to the particular individuals' profile at that DNA site (or locus) as being homozygotic. In such cases, the individual has inherited the same allele from each parent. If an individual shows two alleles at a locus, it implies that there are two different alleles present on each of the chromosomes of that particular pair. This event is referred to as *heterozygosity*, and we refer to the particular individuals' profile at that locus as being heterozygotic.

The DNA molecule encodes for the functional activities of the cells in higher organisms with the packages of genetic information on the DNA referred to as *genes*. The DNA regions containing genes are called *coding regions* (or *exons*). Genes have a dedicated role in protein synthesis, with these products being in turn linked to one or a number of biological or biochemical functions. For the most part, these essential biological functions are the same from individual to individual. As such, exonic regions are highly conserved across individuals of the same species – that is they are almost identical genetically. Another reason for DNA sequence conformity in coding regions, or genes themselves, is that alterations to DNA sequences in these areas usually have some deleterious effect on the physical well being of the individual. The health effects reduce the likelihood of survival, mate selection, and reproduction and therefore are directly linked to the propagation of this individual's particular genotype. In population genetic terms, this link between genotype and ability to survive is referred to as *genetic fitness*. By this mechanism, a genotype that affects the fitness of an individual will become less prevalent, or even extinct, within a population. This is the notion of selection.

The entire DNA molecule is not made up of genes; in fact, only a little over 5% of the genome codes for the production of proteins. The remainder of the DNA is made up of noncoding regions, or introns. Despite their prevalence, the prescribed function of noncoding regions is poorly understood. Noncoding regions were initially presumed to be functionless and as a result were commonly, but inappropriately, referred to as *junk DNA*. Recent evidence proposes some functionality for certain noncoding DNA regions [1–3]. Interestingly, large areas of noncoding DNA, many of which are not implicated in regulation, are strongly conserved between species. This may be strong evidence that they too are functional. Although some ambiguity remains as to their exact role, it is foolhardy to assume that noncoding regions are redundant.

However, as noncoding regions appear to have less responsibility when it comes to the messenger characteristics of the DNA, they have not been exposed to the same evolutionary selection pressures as the coding regions. In essence, mutations can occur in the noncoding regions without affecting the genetic fitness of the individual. Over time, this has led to a high degree of polymorphism in the noncoding regions of the human DNA molecule as mutation has been allowed to continue relatively unchecked. Hypervariable loci, such as those used for forensic identification purposes, are common in the intronic regions.

There are three major varieties of polymorphisms that exist on the genome: minisatellites, microsatellites, and single nucleotide polymorphisms (SNPs). Minisatellites (or variable number of tandem repeats (VNTRs)) are large fragments of DNA which are comprised of sequentially aligned homologous units. The individual units are typically between 20 and 100 base pairs (bps) in length and are repeated consecutively up to 100 times. The overall molecular weight

(MW) of the locus is therefore determined by the number of times that the units are repeated. This is the polymorphic feature of these loci. Minisatellite loci are highly polymorphic and for that reason are powerful identification markers.

Microsatellites (or short tandem repeats: STRs), as the name suggests, are smaller versions of minisatellites. The same structural conformation exists in principle but, in comparison to minisatellites, the size of the individual units (2–8 bp) and the number of times they repeat (2–20) are reduced. Microsatellites were originally named as such as they were thought to solely be repeats of (cytosine-adenine) CA dinucleotide stretches. Microsatellites have been detected in every studied organism occurring at a higher rate than would be predicted purely on the basis of base composition. Microsatellites can be termed *simple* (an uninterrupted array of homologous repeat motifs), *compound* (repeat motif changes – two or more adjacent simple repeats), or *complex* (array is interrupted or contains several repeats of variable unit length). The overall MW of a microsatellite locus is still determined by the number of times that the units are repeated; however, the number of repeats is typically smaller for STR loci. Both VNTR and STR loci are length-based polymorphisms. This means that polymorphisms can be detected by standard electrophoretic techniques.

SNPs are loci where there is a variance in the individual base that exists at a particular position on the genome. SNPs are classified, and distinguished from single base changes, if the frequency of occurrence of the minor (less frequent) allele exceeds 1%. SNPs are the most common form of polymorphism on the genome, occurring approximately every 1000 bp (in unrelated individuals). SNPs can be bi-, tri-, or tetra-alleleic, meaning that the base that exists at a particular SNP position can vary between two, three, or four possible types. In practical terms, however, biallelic SNP loci are the only form of the polymorphism that are routinely detected and analyzed.

Aside from the polymorphic loci themselves, DNA has an additional level of variability that is introduced through processes such as assortment and recombination. These occur as part of the meiotic division that cells undergo during gametogenesis. Both assortment and recombination have the effect of shuffling genetic material and essentially randomizing the distribution of the diploid genotype into the haploid gametes. In assortment, the order in which chromosomal pairs align at the equator of the cell during metaphase is random. This means that the total number of gametic chromosomal combinations that can be formed from n chromosome pairs is 2^n (Figure 3). In humans ($n = 23$), meaning that over eight million combinations are possible among the haploid gametes of any individual.

Recombination, also known as *crossing-over*, involves the physical exchange of genetic material. In recombination, the arms of sister chromatids of a homologous chromosome pair can overlap during prophase I. This contact allows for the physical exchange of chromosomal segments and the genetic material that they carry (Figure 4). Without recombination the arrangement of alleles on a particular chromosome would remain coupled together. Recombination allows new (and possibly advantageous) combinations to be produced and adds another element of variability to the inheritance of DNA.

Aside from the shuffling effect of assortment and recombination, DNA may be subject to further mutational events during meiotic or mitotic replication. At the outset of cell division the DNA must replicate itself faithfully. Any errors during this replication process will introduce a difference between the parent and progeny cells. Such a mutation occurring during meiosis can lead to a Mendellian inconsistency between the biological parent and their offspring, as the parents' gametes have mutated to a haplotype that is inconsistent with the parental genotype. These mutations are more commonly observed on the paternal side, i.e., occurring during spermatogenesis, and are an important consideration in parentage or kinship investigations.

Microsatellite loci have a high mutation rate, in comparison with other polymorphic loci such as SNPs. This trend is observed for both autosomal and nonautosomal loci. In fact, microsatellite loci are thought to be located in mutational "hot spots" [4], that is, regions of the genome that are more susceptible to mutation due to their structure or location. A range of mutational models exist that describe the pattern of mutation of a particular locus. The model favored for microsatellite mutation is the model of slipped strand mispairing (SSM) during DNA replication.

Through the basics genetics, we understand that our genetic composition is comprised of equal proportions of DNA from each of our parents. We also

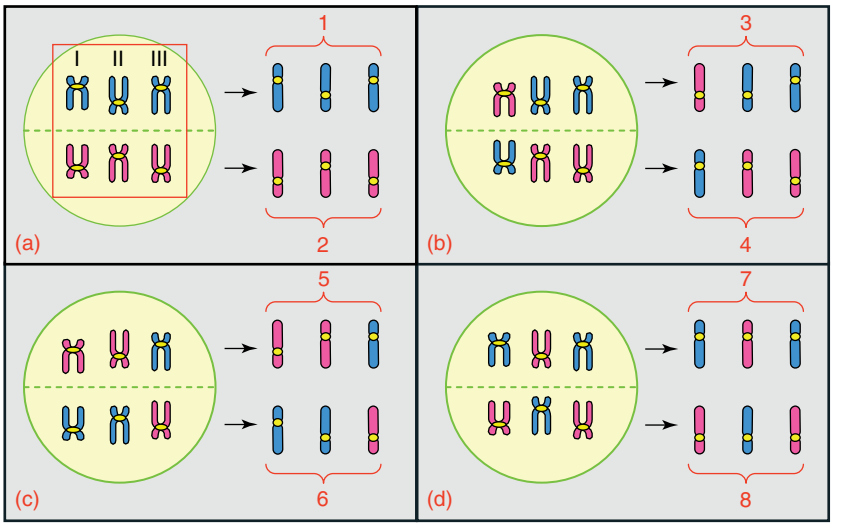


Figure 3 An example of the effect of assortment in generating haploid diversity. If there are three pairs of blue chromosomes (I, II, and III) which are duplicated during prophase to give three pink duplicate chromosome pairs. There are four different ways in which these three pairs of pairs can align at the equator of the cell during metaphase (a–d). After the two rounds of meiotic division this produces eight different possible haploid sets (1–8). The number of theoretical possibilities in a diploid organism is equal to 2^n, where n is the number of chromosomes in the haploid gamete

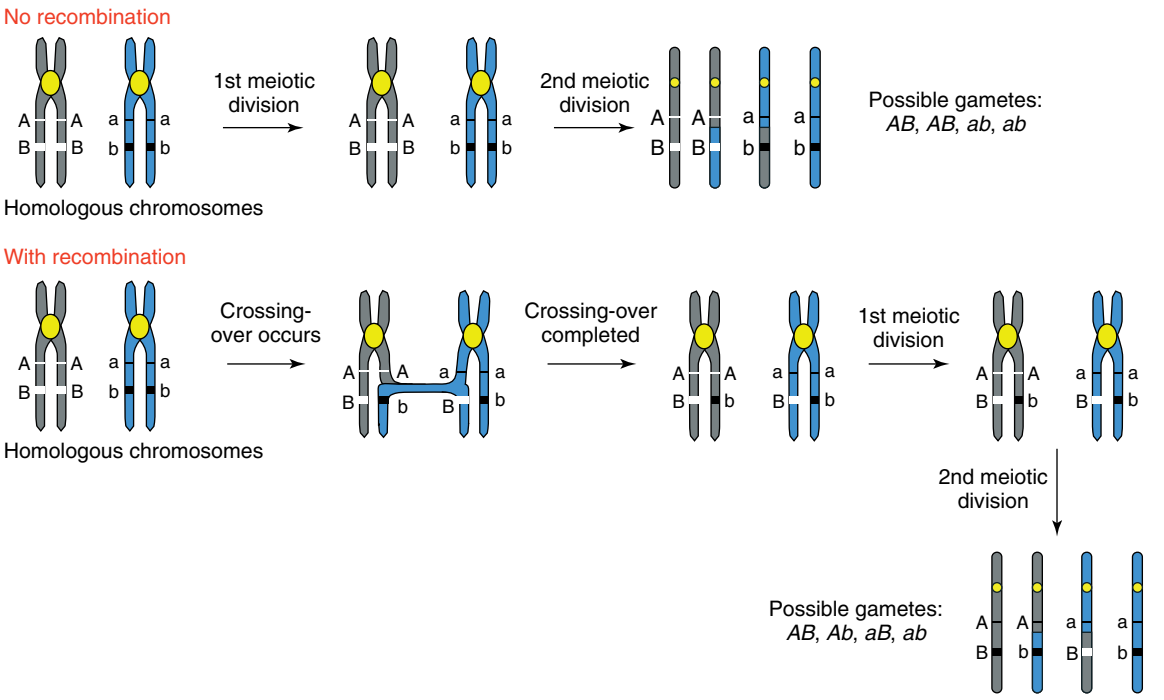


Figure 4 Representation of recombination showing its effect in shuffling the distribution of genetic material in the resultant gametes

understand that there are coding and noncoding segments of the molecule and types of loci that demonstrate considerable variability. We also acknowledge the role that mutation plays in shuffling and altering DNA characteristics, creating additional diversity within populations. The practical suitability of these characteristics is clear; however, it is the physical composition of the DNA molecule itself and its own inherent variation that provides scientists such a suitable template for analysis in a forensic context.

Fundamental Molecular Biology

The realization that DNA is the principle genetic molecule immediately focused attention on its structure. Likewise, the revelation that the structure of DNA was relatively simple [5] helped scientists to understand that genes had relatively similar three-dimensional structure and that differences between two genes resided in the order and number of the structural building blocks along the molecules and not in the molecules overall shape.

Structurally, DNA comprises two complementary chains of nucleotides twisted about each other in the form of a right-handed double helix. A nucleotide unit consists of a five-carbon sugar (2′-deoxyribose), a phosphate residue, and a nitrogenous base (either a purine or pyrimidine; Figure 5). The linking of the nucleotide units is formed via the phosphate residue through a phosphodiester bond between adjacent sugar groups. This series of bonds forms the sugar–phosphate backbone of the DNA strand and allows for the formation of extremely long polynucleotide chains, containing upward of 3 billion nucleotide units.

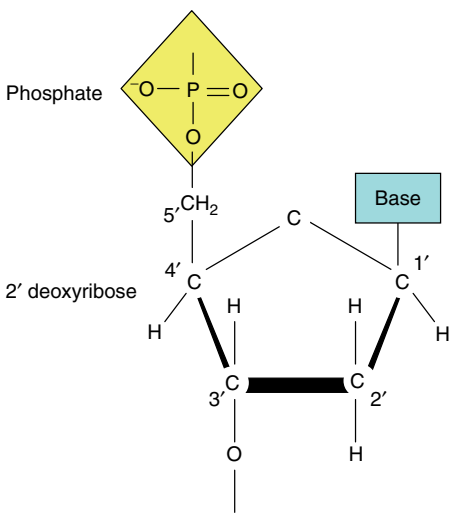


Figure 5 A nucleotide unit (or monomer). The nitrogenous base is either a purine or a pyrmidine. Nucleotide units are linked through a phosphodiester bond via successive phosphate residues and the free oxygen on the 3′ carbon

The backbone of the DNA is therefore a repetitious structure, which, due to its uniformity, is incapable of encoding information. The importance of the DNA structure is derived from the bases that are attached to each of the sugars of the polynucleotide chain. There are two types of bases which attach to the sugar group of the nucleotide, the purines (adenine (A) and guanine (G)) and the pyrimidines (cytosine (C) and thymine (T)). In contrast to the regular structure of the sugar phosphate backbone, the order of the purine and pyrimidine bases along the chain is highly irregular. The order of the bases along the polynucleotide chain is referred to as the *sequence of the DNA molecule* or *DNA sequence*.

Another vital discovery of the Watson and Crick model is that the two chains of the DNA molecule are made up of complementary pairs of bases, linked together by hydrogen bonds. The pairing of the bases follows simple and strict rules. Adenine is always paired with thymine and guanine is always paired with cytosine. No other pairings are possible due to the physical structure of the bases. The strictness of the pairing rules results in a complementary relationship between the DNA sequences on the two intertwined strands of the double helix. For example, a sequence of 5′-AAGCTG-3′ on one chain, the opposite chain must have the sequence 3′-TTCGAC-5′.

The revelation that the twisted strands of the double helix were always complementary meant that if the strands could be separated, and new DNA synthesized along each, the strict base-pairing relationship would see two identical double-stranded DNA molecules produced. In 1985, Saiki *et al.* [6] published an paper in which they described for the first time a method to simulate the DNA replication process in a laboratory environment. The technique was called the *polymerase chain reaction* (PCR). The PCR is essentially a sample preparation step. It allows for the continuous replication of subanalytical quantities of DNA to amplify it to a level such that routine analytical methods can be used for genotyping.

The discovery of the PCR has proven to be the catalyst for the modern biotechnology revolution. By 1988, the PCR had already ranked itself alongside cloning and DNA sequencing as an indispensable tool of molecular biology. The PCR has rapidly advanced the field of molecular biology. If the DNA sequence of a target region is known, the PCR allows for the replication and amplification of that region to the exception of the DNA, which is not required for analysis [7]. The *in vitro* replication of DNA, via the PCR, models the natural process of DNA replication. Each PCR cycle is one cycle of replication that theoretically doubles the amount of template DNA present in the sample. Each PCR cycle consists of three steps that are mediated by the temperature at which the reaction is proceeding and the relevant chemical and biological additives. The PCR is thermally controlled, readily automated, and is complete in 1–3 h.

The sequence of bases on the entire DNA molecule has now been determined [8]. It is accepted that with the exception of identical twins, no two humans contain identical DNA sequences. Therefore, an individual's DNA sequence is unique. Although extremely significant in principle, this fact does not have high practical relevance purposes as scientists are not yet able to designate the complete sequence efficiently and cost effectively. It is of more significance that differences exist on the DNA molecule and that this characteristic is another attribute of DNA that aids in its application in the forensic context.

Summary

In forensic DNA profiling, investigators collect biological evidence from crimes primarily with the purpose of assisting in the identification of the donor. The fundamental objective therefore is to differentiate individuals through analysis at the DNA level. In general, therefore, it is areas of difference on the DNA that are of more importance than those that are highly conserved. Both sequence variation and length variation have been utilized in forensic DNA profiling for many years. More recently, sophisticated techniques examining length variation at microsatellite loci are favored as the routine target of analysis.

To be useful in forensic context, the scientific procedure employed must be robust (able to produce a result from compromised samples), sensitive (able to produce a result from small amounts of original material), highly discriminating (able to provide a result which conveys a satisfactory degree of confidence), and accurate (involve minimal subjectivity or risk of error). The forensic community has many analytical tools already that satisfactorily address these requirements, but despite considerable progress the scientific community is still only beginning to understand the vast amounts of information resident on the DNA molecule and to develop technologies to allow its exploitation. Undoubtedly there is much more development to come, but already the implementation of DNA science into the forensic arena has had a tremendous and positive impact. This is primarily due to the inherent sophistication and informativeness of the DNA molecule itself.

References

[1] Johnston, M. & Stormo, G.D. (2003). Heirlooms in the attic, *Science* **302**, 997–999.
[2] Mattick, J.S. (1994). Introns: evolution and function, *Current Opinions in Genetics and Development* **4**, 823–831.
[3] Mattick, J.S. (2001). Non-coding RNA's: the architects of eukaryotic complexity, *European Molecular Biology Organisation Reports* **2**, 986–991.
[4] Chambers, G.K. & MacAvoy, E.S. (2000). Microsatellites: consensus and controversy, *Comparative Biochemistry and Physiology Part B* **126**, 455–476.
[5] Watson, J.D. & Crick, F.H.C. (1953). Molecular structure of nucleic acids: a structure for deoxyribonucleic acids, *Nature* **171**, 738–740.
[6] Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. & Arnheim, N. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction analysis for diagnosis of sickle cell anemia, *Science* **230**, 1350–1354.
[7] Saiki, R., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. & Erlich, H.A. (1988). Primer-directed enzymatic amplification of DNA with thermostable DNA polymerase, *Science* **239**, 487–491.
[8] Lander, E.S., *et al.* (2001). International Human Genome Sequencing Consortium, *Nature* **409**, 860–921.

Related Articles

Biological Stains

DNA: Sources of

SIMON J. WALSH

DNA: Allelic Designation *see* Allelic Designation

DNA: an Overview

History of DNA Profiling

DNA profiling, as we know it today, developed thanks to two independent breakthroughs in molecular biology that occurred at the same time on different sides of the Atlantic. In the United States, the polymerase chain reaction (PCR) was developed by Kary Mullis of Cetus Corporation [1–3]. Almost simultaneously, the individual-specific banding patterns observed after restriction fragment length polymorphism (RFLP) analysis of repeated DNA sequences were discovered by Professor Sir Alec Jeffreys at the University of Leicester [4–6]. In its earliest incarnation, this technique termed as *DNA fingerprinting* by its creators was performed by restriction of 0.5–10 μg of extracted DNA using the restriction enzyme *Hin*FI, followed by Southern blotting hybridization with probes termed as *33.5, 33.6*, and *33.15*, designed to bind to multiple "minisatellites" present in the restricted DNA [6]. This multilocus probing (MLP) technique would result in the binding of probes to multiple independent DNA fragments at the same time giving rise to the traditional "barcode" pattern, that is, often visualized when one discusses about DNA profiling, even today. Differences in the number of times the probe sequence is repeated in each DNA fragment form the basis of the individual patterns observed on the autoradiogram image.

The Mendelian inheritance of these markers was established by the pedigree analysis of 54 related persons [4], and the individual nature of the banding pattern was further established by the examination of 20 unrelated persons [6]. The probability of two unrelated individuals carrying the same fingerprint was calculated from these data and was estimated as 3×10^{-11} for probe 33.15 alone, provided 15 bands could be resolved in the 4–20 kb size range on the autoradiogram (*see* **Autoradiograph**). The potential application of this technique to maternity/paternity disputes and to forensic investigation was recognized immediately by Jeffery's *et al.*, and was demonstrated by DNA fingerprinting of forensic-type samples, such as bloodstains and semen, in the same year [5]. Difficulties in interpretation of MLP images quickly resulted in single locus probes (SLP) for variable number of tandem repeat (VNTR) loci becoming the markers of choice for DNA profiling (*see* **Variable Number Tandem Repeats**).

The first report concerning the use of DNA profiling in a criminal investigation was published in 1987 [7]. This investigation used two unpublished SLPs to link semen stain samples, collected from two rape and murder cases that had occurred three years apart in 1983 and 1986 in Leicestershire, United Kingdom. The probability of this match occurring by chance was calculated as 5.8×10^{-8}. This result not only linked the two crimes but also exonerated an innocent man implicated in the murders and led to the first mass screening project undertaken for DNA profiling in the world [8].

The potential of DNA analysis for forensic science had now been demonstrated; the technology now required statistical validation by analysis of population frequencies and application to casework samples before it could progress. Early evaluation studies on MLP 33.15 provided optimistic support for the use of DNA for the personal identification and the identification of male rapists from a mixed male/female sample [9]. It does, however, also begin to uncover the limitations of this method. A mean success rate of only 62% for the DNA fingerprinting of donated vaginal swabs was observed and no typing was possible for blood or semen stains that had been stored for 4 years at room temperature, and difficulty in directly comparing related samples run on different gels was also cited as a potential problem [9]. Similar studies and European collaborations were undertaken on SLPs such as YNH24 and MS43a [10, 11]. Difficulties were again observed when interpreting gel images, with only 77.9% of 70 samples distributed between nine laboratories producing matching results when a 2.8% "window" for size differences between gel runs and laboratories was used [10]. It was recognized that subtle differences between laboratory protocols were responsible for some of the observed discrepancies, leading to a requirement for the standardization of laboratory methodology [10] and DNA profile interpretation [12, 13].

Such standardizations could improve the reproducibility of DNA typing results for MLP and SLP

marker systems, but in order to be applicable to forensic investigation, DNA systems must be robust and must be applicable to samples of a less than pristine nature or that consists of only a few cells. PCR was first applied to forensic DNA profiling for the investigation of the HLA-DQ-α1 gene (*see* **DQα**), a polymorphic gene that encodes a human leukocyte antigen cell surface protein, located in the major histocompatibility complex (MHC) class II region on chromosome 6 [14].

Two big breakthroughs occurred between the late 1980s and early 1990s that would form the basis of DNA profiling techniques that are recognized today. An alternative class of DNA marker, the microsatellite or short tandem repeat (STR) marker (*see* **Short Tandem Repeats**), was described by Weber *et al.* [15] and an alternative method for DNA visualization, PCR amplification, and fluorescent labeling of VNTR markers was also introduced [16, 17].

STR Analysis

STR markers are similar to the VNTR markers that were originally identified and utilized in DNA fingerprinting and SLP profiling. The difference between the classes of DNA marker lies in the length of the tandemly repeated DNA sequence. VNTRs contain 10–33 bp hypervariable repeat motifs [4] that must be observed over a size range of 4–20 kb [9] and SLP markers that are observed over a size range of 1–14 kb [10]. An STR marker repeat is composed of 1–6 bp repeat motifs [18], making the region of DNA that must be scrutinized very short (<1 kb). This length reduction is immediately beneficial to one of the problems encountered in SLP profiling; difficulty in analyzing degraded DNA [5]. The use of the multiplex PCR to amplify target sequences before visualization significantly reduces the amount of DNA required for analysis from microgram to nanogram amounts [18, 19]. The detection and visualization method of polyacrylamide gel electrophoresis and fluorescent detection using an automated DNA sequencer (model 370, Applied Biosystems, Foster City, CA, USA), in combination with an internal size standard (GS2500, Applied Biosystems) and GENESCAN 672 software (Applied Biosystems), also allowed for precise band sizing, answering problems of intra- and interlaboratory allele designation discrepancies that had been observed in SLP analysis [12, 20, 21].

One of the earliest multiplexed STR systems developed for forensic DNA analysis was a quadruplex reaction that amplified the STR markers HUMVWA31/A, HUMTHO1, HUMF131A1, and HUMFES/FPS [22]. These particular STR markers were selected from the hundreds of STRs identified throughout the human genome [23] based on a number of important parameters. Each STR must have a high level of allelic variability to maximize the discriminating power of each marker. Markers should have short PCR product length (<500 bp) to aid the analysis of degraded DNA. The chromosomal location of any potential marker should be checked to avoid closely linked loci, and tetranucleotide repeat motifs are preferred due to low artifact production during PCR amplification [24]. The overall match probability using this system was calculated as 1.3×10^{-4} for white Caucasian populations [18]. Validation studies carried out on this quadruplex system determined that 1 ng template DNA was optimal for amplification and analysis, that stutter bands up to 11% were observed at some loci, especially HUMVWA21/A, and that DNA mixtures in a ratio of 1 : 2 or 2 : 1 could be distinguished using this system [22]. Further validation studies were carried out on extremely compromised casework samples, collected from the victims of the Waco siege, which resulted in the identification of several individuals that was not possible by any other means [25, 26].

The quadruplex system described above was next developed into an octoplex as the application of DNA profiling to forensic casework increased. The new octoplex system coamplified the tetranucleotide STR loci HUMVWFA31/A (vWA), HUMTHO1 (THO1), D8S1179, HUMFIBRA (FGA), D21S11, and D18S51 [27]. The nomenclature used for naming these STR markers had been standardized now, to allow for easy comparison between different working groups [28]. This system also included primers for the amplification of a region of the amelogenin gene, which could be used to deduce the sex of the DNA sample being analyzed [29]. Optimization and validation studies on this octoplex system reduced the amount of template required for the generation of full DNA profile to just 500 pg, with partial profiles being generated from as little as 50 pg, the equivalent of just 10 diploid cells [27]. This octoplex became known as the *second-generation multiplex* (*SGM*) system, [30] and was used to populate the first national criminal intelligence DNA database,

which became operational in the United Kingdom in April 1995 [27]. The SGM system was human-specific, highly discriminating, with a probability of chance association calculated as 1×10^{-8} [31]. It could also be applied to degraded DNA samples and was capable of detecting and resolving mixed DNA profiles at ratios between 1 : 10 and 10 : 1 [30]. One final evolution would take place, the inclusion of a further four STR markers (D3S1358, D19S433, D16S539, and D2S1338), to produce the STR profiling kit, that is, currently used in the United Kingdom for forensic DNA profiling, population of the national DNA database (NDNAD) and for paternity testing.

The AmpFlSTR® SGM Plus™ System

The AmpF*l*STR® SGM Plus™ system, commercially produced by Applied Biosystems (ABI), a division of Perkin Elmer, Foster City, California, USA, was introduced in June 1999, and was validated for use in forensic casework in 2000 [32]. It was designed to replace the SGM system for forensic casework in the United Kingdom, to decrease the probability of a chance match occurring from 1×10^{-8} to 1 in trillions for unrelated individuals [32]. This greatly increased the statistical power of DNA evidence to be taken for scrutiny in the courtroom, while being back-compatible with DNA profiles already stored in the NDNAD. The statistical power of this new system was deemed so great that instead of calculating the exact match probability for a full DNA profile, it was recommended that an arbitrary conservative estimate of 1 in a billion be reported for the match probability between unrelated individuals, 1 in 1 million for parent/child relations, and 1 in 10 000 for siblings [33]. The characteristics of each STR marker are detailed in Table 1.

The AmpF*l*STR® SGM Plus™ PCR amplification kit can be analyzed by two DNA separation methods, capillary electrophoresis, or polyacrylamide gel electrophoresis. In this article, the analysis was performed by polyacrylamide gel electrophoresis using an ABI Prism™ 377XL DNA Sequencer (ABI). The ABI Prism™ 377XL DNA Sequencer was introduced by Applied Biosystems in 1995 and stopped its use in 2001 [44]. Automated fragment sizing of fluorescently labeled DNA fragments was achieved by the use of a scanning argon ion laser, which tracks back and forth across a "read-region" at the lower end of a vertical polyacrylamide gel. As each labeled DNA fragment passes the laser, the fluorescent dye is excited, resulting in emission of light. This light is then collected and separated by wavelength onto a charged coupled device (CCD) camera. The camera used on the 377 model is capable of detecting four different wavelengths simultaneously, allowing for the detection of three similarly sized PCR products in a single gel run, with inclusion of a separately colored size standard in each lane. The 377XL model was validated for forensic STR analysis in 1996, using the original SGM octoplex system [45]. The set of validation experiments carried out determined that complete resolution of 1 bp differences between fragments could be achieved up to 350 bp, sizing precision was increased twofold and sensitivity was increased by one-third, compared to the predeceasing 373A DNA sequencer [45].

Alternative DNA Markers

Autosomal STR markers have become the most utilized ones in both forensic and paternity DNA profiling. However, there are numerous alternative markers that can be interrogated when required. Another class of autosomal marker, the single nucleotide polymorphism (SNP), has been investigated for application to forensic casework and identification projects. SNPs, as the name suggests, are alterations of a single base pair and occur on average every few hundred base pairs throughout the human genome [46, 47]. The major advantage of SNP markers over STRs is the small size of the DNA target, making them very useful for degraded DNA analysis and disaster victim identification projects [48]. Reduced size STR amplicons have, however, been developed to analyze degraded DNA whilst remaining compatible with current DNA databases to allow easy searching in identification projects [49].

There are a number of disadvantages associated with the use of SNP markers instead of STRs. The comparatively low discrimination power of each locus requires that approximately 50 SNPs must be investigated to give match probabilities equal to 10 STR loci [50]. Mixture analysis is also complicated for SNP analysis are usually bi-allelic, but have also been observed to be tri- and even tetra-allelic. The use of a predominantly biallelic marker makes the distinction between mixtures and homozygotes difficult, especially for minor contributors, which may

Table 1 Characteristics of SGM Plus STR loci[a]

Marker	Chromosome location	No. of alleles in allelic ladder	Size range (bp)	Repeat motif	Dye label	References
D3S1358	3p	8	114–142	TCA $(TCTG)_{1-3}$ $(TCTA)_n$	5-FAM	[34]
vWA	12p12-pter	14	157–209	TCA$(TCTG)_{3-4}(TCTA)_n$	5-FAM	[35]
D16S539	16q24-qter	9	234–274	$(AGAT)_n$	5-FAM	GenBank G07925
D2S1338	2q35–37.1	14	289–341	$(TGCC)_n(TTCC)_n$	5-FAM	[36]
Amelogenin	X p22.1–22.3	NA	107	NA	JOE	[29]
	Y p11.2		113			
D8S1179	8	12	128–172	$(TCTR)_n$	JOE	[37]
D21S11	21q11.2-q21	24	187–172	$(TCTA)_n(TCTG)_n$ [$(TCTA)_3$ TA$(TCTA)_3$ TCA $(TCTA)_2$ TCCA TA] $(TCTA)_n$	JOE	[38]
D18S51	18q21.3	23	262–345	$(AGAA)_n$	JOE	[39]
D19S433	19q12–13.1	15	106–140	(AAGG)(AAAG)(AAGG) (TAGG)$(AAGG)_n$	NED	[40]
THO1	11p15.5	10	165–204	$(AATG)_n$	NED	[41]
FGA	4q28	28	215–353	$(TTTC)_3$TTTT TTCT $(CTTT)_n$ CTCC $(TTCC)_2$	NED	[42]

[a] Adapted from AmpF*l*STR SGM Plus PCR Amplification kit manual and [43]

also display allelic dropout [50]. Owing to these problems and the financial implications of repopulating national DNA databases with SNP profiles, it is unlikely that SNP markers play a major role at the forefront of forensic DNA analysis in the foreseeable future. There are still some applications in related fields in which SNP markers are proven to be useful: determination of phenotypic characteristics, such as eye color [51, 52] and hair color, by variation in the melanocortin 1 receptor (*MC1R*) gene [53–55]. SNP markers can also be used to analyze the uniparentally inherited mitochondrial DNA (mtDNA) (*see* **Mitochondrial DNA: Profiling**) and Y chromosomes (**Y-Chromosome Short Tandem Repeats**), which are discussed in [56, 57].

DNA Extraction

In order for DNA from any given biological sample to be analyzed by a PCR-based method, it must first be purified from all organic and inorganic substances with which it is associated (*see also* **Extraction**). This process can become complicated during forensic investigation as DNA is often present on or in materials that are not usually encountered in the molecular biology laboratory. The extraction stage of DNA analysis is also the most susceptible stage to the occurrence of laboratory-induced contamination of sample material. For this reason, DNA extraction protocols, no matter which one is employed, must always be carried out in a dedicated laboratory, physically separated from downstream processes, especially PCR. Literature searching for DNA extraction of forensically interesting materials reveals tens, of various techniques, and variations on each of these. A brief listing of DNA extraction methods for commonly encountered forensic evidentiary samples is given in Table 2. Of the many techniques that have been proposed, the most commonly utilized ones are based on three techniques: organic cell lysis with phenol–chloroform purification, Chelex®100, and silica-based extractions.

The organic, phenol–chloroform method is the original DNA extraction technique to be applied to forensic (and archaeological) specimens [5, 58–60].

Table 2 List of materials commonly collected for DNA analysis during forensic investigations with reference to extraction methods

Substrate	Type of extraction	References
Blood	Silica	[63, 64]
	Chelex®	[62]
Bone	Silica	[65–70]
	Lysis/precipitation	[71]
	Phenol/chloroform	[66, 67, 72, 73]
Teeth	Chelex®	[74, 75]
	Phenol/chloroform	[76–79]
Hair	Phenol/chloroform	[58, 59]
	Chelex®	[62, 80]
	Alkaline digestion	[81]
	Silica	[80, 82]
Saliva	Chelex®	[62, 83–85]
Buccal cells	Phenol/chloroform	[86, 87]
	Silica	[88]
Maggots	Phenol/chloroform	[89–91]
Feces	Silica	[92–95]
	Phenol/chloroform	[94, 95]
Urine	Silica	[69, 96, 97]
Semen	Chelex®	[98]
Trace/cells	Silica	[99, 100]
	Organic	[99, 101–103]
	Chelex®	[104]

Although many subtle variations in buffer pH, chemical concentration, incubation time, and temperature have been introduced over the years, the basis of the technique remains the same. The detergent sodium dodecyl sulfate (SDS) is used to break down cell walls and lipids present in the sample, and proteinase K, a strong protease enzyme, is used to digest the proteins that protect the DNA molecule in its natural state. An additional component, dithiothreitol (DTT), can also be used to digest more robust materials such as keratinized hair and sperm cells and forms the basis of the differential extraction technique used for vaginal swabs with semen present. The differential extraction of the two cell types is possible as the sperm nuclei are impervious to SDS/proteinase K digestion because the nuclear membrane is reinforced with cross-linked thiol-rich proteins. This allows for digestion and removal of the female component from the mixture leaving intact sperm nuclei to be digested by the addition of a SDS/proteinase K/DTT mixture [5]. Once digested, phenol–chloroform is used to separate proteins into organic phenol layer from the nucleic acids in the aqueous chloroform layer. The nucleic acids must then be concentrated and purified from the hazardous chloroform solvent before further analysis. This is classically achieved by ethanol precipitation but can also be carried out using a commercially available centrifugal filter device, such as the Centricon®100 system [61].

Chelex®®100 was introduced as a medium for the simple extraction of DNA from forensic materials in 1991 [62]. This technique was designed specifically for use with forensic specimens and was intended to replace the organic, phenol–chloroform technique. The Chelex® method has three major advantages over the organic, phenol–chloroform methods: it is much faster, taking only 1 h compared to up to 24 h for the organic, it does not require multiple tube transfers, reducing the risk of laboratory-induced contamination, and it does not require the use of hazardous chemicals. Chelex® resin is composed of styrene divinylbenzene copolymers containing paired iminodiacetate ions, which act as chelating groups, binding polyvalent ions such as Mg^{2+} [62]. In this method, Chelex® resin is added directly to the sample from which DNA is to be extracted. Cell lysis and DNA liberation is achieved by a combination of alkalinity (pH 10–11) and heating at 100 °C. Initial testing performed on forensic-type samples such as blood, blood stains, semen stains, and hair demonstrated that the performance of Chelex® extraction was equal to that of phenol-chloroform methods and, in the case of blood, was less likely to allow the carryover of PCR inhibitors such as heme. Similar to organic methods, an additional DTT digestion step is required for DNA extraction from sperm [62].

Silica particles were first used for DNA extraction from human serum and urine samples. This method, similar to Chelex®, was designed to be more rapid and involve fewer tube transfer stages, to reduce the risk of sample contamination or DNA loss, than organic methods. The method uses a chaotropic agent, guanidinium thiocyanate (GuSCN) to lyse cells and inactivate nucleases, while simultaneously facilitating the binding of the freed nucleic acids to silica particles [69]. Once bound, the silica–DNA complex can be pelleted, allowing cellular debris and other non-DNA components to be removed and discarded. After washing, the purified DNA can then be eluted from the silica in sterile water or TE buffer. This method is extremely sensitive due to the strong binding affinity of silica particles for nucleic acids in

the presence of chaotropic agents and is subsequently adapted for use with ancient bone samples [70, 105]. The method does require some caution as GuSCN produces HCN gas on contact with acids and should therefore be kept in an alkaline solution at all times, and disposed of in 10 M NaOH.

The silica–DNA binding affinity has been utilized in the development of commercial kits such as the QIAamp® DNA kits produced by QIAGEN (Qiagen, West Sussex, United Kingdom) (described in **QiaAmp**) and the DNAIQ™ system produced by the Promega Corporation. Both systems have been tested and validated for use with forensic casework samples [63, 64] and both systems are compatible with automated robotic workstations [106–108]. The versatility and reproducibility of such commercial DNA extraction kits, especially QIAamp® products, has made them the choice of the forensic research community over recent years, as illustrated by their usage in the majority of research articles published over recent years.

DNA Quantification

Commercially produced DNA profiling kits, such as the AmpF*l*STR® SGM Plus™ PCR amplification kit described above, are optimized to produce DNA profiles from a narrow range of template DNA concentration. AmpF*l*STR® kits produced by Applied Biosystems are optimized to amplify 1–2.5 ng template DNA. The addition of insufficient template DNA can result in stochastic amplification. Stochastic amplification manifests as unbalanced amplification of heterozygote loci and can, if severe, result in allelic dropout, making heterozygotes to appear as homozygotes at the affected loci [109]. The addition of template DNA in excess can lead to the production of large stutter peaks that complicate the interpretation of DNA profile. The term *stutter* refers to the production of natural biological artifacts during the PCR. The currently accepted model for stutter generation is by a "polymerase slippage" model that results in the addition or, more often, deletion of a single repeat unit from the actual template size. The detection and classification of stutter play an important role in profile interpretation, especially if a mixture is anticipated or observed. If the stutter characteristics are unknown for a DNA profiling system, it is possible that such artifacts may falsely be reported as "actual alleles". It has been observed that stutter does not usually exceed 10% of the associated allele peak height, but with variation between both STR loci and alleles in the locus, stutter has been observed to approach 15% [110] in some cases. Percentage stutter is calculated using the peak height in relative fluorescent units (RFU) of the observed peaks. Figure 1 shows a typical stutter pattern at the STR locus FGA.

The most commonly observed nonbiological artifacts caused by the addition of template DNA in excess are termed as *pull-up* or 'bleed through' peaks. Pull-up peaks are produced when the GeneScan software program is unable to distinguish between the emission spectra of the fluorescent colors in the STR system. This phenomenon is visualized by the appearance of false bands in the overamplified fragment's size range in differently colored markers.

By quantifying all DNA samples before performing DNA profiling PCR, the production of such artifacts can be reduced or completely avoided. As with DNA extraction, there are numerous methods available to perform this function. The traditional method of DNA quantitation involves measuring the absorbance of the sample at 260 nm on a spectrophotometer. This method is simple to perform

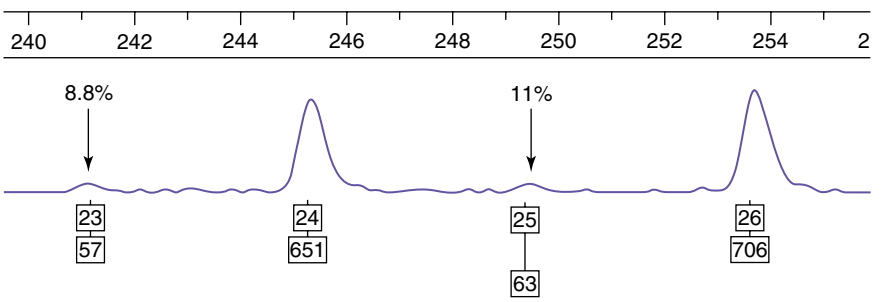


Figure 1 Example of typical stutter pattern of alleles 24 and 26 at STR locus FGA

and shows little sample-to-sample variation [111], making it a desirable technique. This technique does, however, suffers because of its simplicity. The technique is not species-specific, indicating that any bacterial or fungal DNA that has copurified with the DNA of interest cannot be distinguished from the human target DNA. Additionally, single-stranded DNA and nucleotides cannot be distinguished from double-stranded DNA by this method, which can also lead to falsely high quantitation readings [112]. A different approach to DNA quantitation is taken by the slot blot technique. This method is based on the specific hybridization of a 40-bp probe that binds to the human alpha satellite locus, D17Z1 [113]. Using this technique, DNA is quantified by visual comparison of sample band intensities with the band intensities of standards, produced from known amounts of DNA. This comparison can be performed either manually – but this allows for the subjectivity of individual interpretation, which can lead to differences between operators in quantitation – or electronically by computationally converting the band intensities into numerical values [114]. This technique shows a high degree of species specificity (human and primate) and can detect as little as 150 pg DNA in a standard assay [113]. Similar to the development of commercial DNA profiling kits, the development of commercial DNA quantitation kits has also progressed. Applied Biosystems produced the Quantiblot® human DNA quantification kit, similarly to the original technique of Walsh. Walsh *et al.* uses a probe that binds to the D17Z1 locus and has a lower detection limit of 150 pg [113, 115, 116].

An alternative method of DNA quantitation involves the use of a fluorescent dye, PicoGreen dsDNA quantitation reagent. This method was developed to allow high-throughput quantitation of multiple samples concurrently while increasing the sensitivity DNA quantitation methods. Picogreen is a Hoechst dye that can pass through the lipid membrane of live or fixed cells to form noncovalent bonds with AT-rich areas in the minor groove of double-stranded DNA [117]. The method utilizes the increased fluorescent intensity that is observed when PicoGreen binds to dsDNA. The fluorescent intensity of the PicoGreen dye is measured with a spectrofluorometer capable of producing the excitation wavelength of ~480 nm and recording at the emission wavelength of ~520 nm. The DNA is then quantified by comparison of the sample fluorescence with the fluorescence of a set of standards that are included in every sample run [118]. The biggest disadvantage to this method over the hybridization methods discussed previously is that this method is not specific for human DNA. Any animal, bacterial, or fungal DNA copurified with the human DNA of interest will contribute to the final reading and could give a falsely high DNA quantification. This method is, however, much more sensitive than the previously described methods, with a reported lower detection limit of 25 pg ml^{-1} [112].

The latest development in DNA quantitation is based on the technique of real-time PCR. The methodology was first proposed 15 years ago by Higuchi and coworkers [119] who, by including ethidium bromide in the PCR reaction, were able to continuously monitor the production of double-stranded DNA in "real time". By capturing the change in fluorescent intensity on video camera, the potential of this new development for specific DNA quantitation was realized [120]. There are several different approaches to real-time quantitation of DNA; they are all, however, based on the principle of fluorescent dye binding double-stranded DNA as it accumulates during the PCR process. As the technique is based on the PCR, DNA quantitation can be undertaken by targeting any specific region of template DNA, with many systems targeting the multicopy *Alu* sequence, which appears 500 000–1 000 000 times throughout the human genome [121–123]. This not only ensures that the technique is species-specific but also has allowed numerous variations, many designed to perform additional functions, such as independent quantitation of nuclear and mtDNA in a single reaction [124–127] and assessment of DNA quality [128, 129].

DNA Profile Interpretation

The methodologies, hardware, and software necessary for the generation of STR profiles are well established but constantly evolving. Regardless of the exact means of production, the interpretation of DNA profiles must remain constant. Using the earliest DNA profiling systems, based on RFLP analysis and Southern blot hybridization, it was proved to be difficult to standardize interpretations between different individuals and different laboratories. Significant differences in band sizing were regularly observed during collaborative exercises carried out by DNA testing

laboratories throughout Europe [10, 11]. Although measures were taken to standardize as many variables as possible including buffer system, gel running time and temperature, and DNA visualization method, there was still an uncertainty in interpretation uniformity between laboratories [10]. The breakthrough required for standardization of DNA profile interpretation of DNA profiles came with the switch to PCR amplification [1, 2] and the introduction of fluorescent labeling allowing digitization of DNA profiling results [16, 17]. These advancements paved the way for the development of universal guidelines and validation of commercially produced kits and instruments to allow for the accurate and reliable reproducibility of DNA profiling interpretation across independent laboratories worldwide. Validation experiments, designed to demonstrate the robustness of the system, must be carried out for all commercially produced forensic DNA profiling kits before they can be used for casework. Validation data for the AmpF*l*STR® SGM Plus™ Amplification kit, used throughout this article, were published in 2000 [32].

The first step toward standardization was the adoption of a common nomenclature for STR loci, including both the name of the marker and the numerical designation of each allele observed in the human populations tested. The basis for the currently used naming systems was brought about by the European DNA Profiling Group (EDNAP), which was formed in 1989 by members of the leading police organizations and universities involved in forensic stain analysis and paternity investigation at that time [130]. The next step in achieving complete interlaboratory concordance was the addition of the allelic ladder [131]. Allelic ladders are produced by creating a mixture of DNA samples containing all observed allelic variations in the human population and then amplifying this mixture with the primers used for each locus in a given STR kit. Once created, the allelic ladder is visualized in parallel with every set of samples so that a direct comparison can be made between the amplified fragment size and the fragment size of previously observed alleles for each allele at each marker. This ensures that slight variations in running conditions, such as those observed with Southern hybridization techniques, are not misclassified based on interrun fragment sizing differences [132]. The inclusion of allelic ladders also allows for the recognition of rare and novel, full and microvariant alleles that are not present in the allelic ladders for each STR kit [133]. The final leap toward accurate DNA profile interpretation was made with the switch from "home-brews" to commercially produced STR amplification kits and the use of identical or equivalent instruments in each laboratory throughout the forensic DNA typing community.

As a result of the standardization measures described above, the interpretation of single-contributor DNA profiles when sufficient template is entered into the reaction is a straightforward process. In the forensic setting, a large percentage of samples submitted for DNA analysis contain biological material from more than one source and result in mixed STR profiles. In order to produce a robust set of interpretation guidelines for mixed profiles, it is vitally important to have a thorough understanding of the amplification characteristics of biological and stochastic phenomenon in the electropherogram, from which the interpretation is made. The most commonly observed artifacts in a DNA profile are termed as *stutter bands*. They are generated during the PCR by slippage of Taq polymerase, resulting in the generation of a fragment one repeat unit shorter, or more infrequently longer, than the true allele. When the peak height of the amplified true allele is >4000 RFU in the electropherogram, stuttering is unavoidable when amplifying STR loci using Taq polymerase [134]. Each STR locus and even each allele in a STR locus has different stuttering characteristics, but stutter products are rarely observed to exceed a peak height in excess of 15% of the height of the associated allele. Any peak found in a stutter position with a height $>15\%$ of the related allele should therefore be considered as a potential true allele in a mixed DNA profile [134].

Another characteristic of the electropherogram that can mystify DNA profile interpretation is the balance of allelic amplification at heterozygotic markers. In a perfectly amplified DNA profile, the peak areas of each heterozygote allele should be equal. Differences in the amplification efficiency of each allele can, however, result in alleles being unequally amplified [134]. For a single source DNA profile produced from sufficient template (>1 ng), each shorter allele at a heterozygous locus will usually have a peak area greater than 60% of the associated larger allele [134]. Heterozygosity balance (Hb) is calculated as relative peak area differences of heterozygote alleles (Φ) for each locus. There are two similar methods used to

determine Hb for forensic DNA profile interpretation, both of which use the peak area in RFU. The first is calculated as Φ = smallest peak area/largest peak area, where a result of $0.67 > \Phi \leq 1$ indicates that the Hb is in the observed limits [134]. This calculation does not, however, provide any information as to which of the heterozygous alleles has been preferentially amplified during the PCR. This information is provided when Hb is assessed by the calculation, Φ = peak area shortest allele/peak area longest allele. In this case, a result of $0.67 > \Phi < 1.67$ indicates a balanced amplification, with Φ values < 1 indicating that the shortest allele of the pair has been preferentially amplified and Φ values > 1 indicating the opposite [135] (*see also* **Peak Height: DNA**).

The presence of additional bands in the electropherogram can also result from overamplification of template DNA causing saturation in fluorescent signal that cannot be resolved by the matrix files of the visualization software. These bands are termed as *pull-up* peaks due to their appearance in the raw data files, whereby a smaller peak is observed directly under the overamplified peak. When the dyes are separated to produce the electropherogram image, the smaller pull-up peak appears as an extra band, with excess blue signal producing green pull-ups and excess green signal producing yellow pull-ups. Pull-up peaks are easily recognized by the experienced DNA analyst due to their atypical morphology compared to true allele peaks and should not interfere with DNA profile interpretation [134]. If a pull-up peak is superimposed over a suspected true fragment, the sample can be diluted and reanalyzed to avoid fluorescent signal saturation.

Low Template DNA Profiling

The two biggest drives for developments in forensic DNA profiling are to improve sample throughput and technique sensitivity. Increases in sample throughput are achieved by automation of routine processes such as DNA extraction, sample loading, and DNA profiling interpretation coupled with effective laboratory management and sample tracking. While the sensitivity of DNA profiling reactions can also benefit from technological advances, the major advancements are achieved by optimization of reaction chemistries, investigation of alternate DNA markers, and use of additional stage processes.

One area of DNA profiling analysis that has received attention is the inclusion of a pre-PCR stage, designed to increase the total amount of template DNA available for STR analysis. A number of techniques, similar in their objective but differing in their means, can be described under the umbrella term of whole genome amplification (WGA) (*see* **Whole Genome Amplification**). As the term indicates, these techniques are designed to replicate all DNA present in a given reaction, with the intention of generating an unlimited source of template material for multiple downstream analyses. Techniques of this kind were first described in 1989, with the publication of a universal DNA amplification method, which used restriction enzyme digestion and ligation into plasmids to achieve its goal [136]. The next generation of WGA techniques appeared in the literature during the 1990s and like many other areas of molecular biology, it moved away from the cloning techniques to PCR amplification. The first PCR–WGA technique was termed as *primer extension preamplification polymerase chain reaction* (*PEP-PCR*) [137]. This technique utilizes fully degenerate 15-mer oligonucleotide primers and a range of annealing temperatures in a PCR to produce multiple copies of any template DNA present in the reaction. In his introductory article, Zhang *et al.* reported a minimum amplification of 30 copies for 78% of the human genome at a 95% confidence level from a single spermatozoon [137]. The limitations of PEP–PCR were also explored, with heterozygosity imbalance and locus dropout cited as the main concerns when the technique is applied to single-cell analysis. These observations were confirmed in the articles reporting the application of PEP–PCR to preimplantation diagnosis of single cells, whereby effects of stochastic sampling and amplification were observed for single-cell analysis [138–140]. Locus and allele dropout was also observed when PEP-PCR was applied to formalin-fixed tissues followed by microsatellite analysis [140].

Another method designed to perform the function of WGA has been termed as *multiple displacement amplification* (*MDA*). This method was first introduced in 2001 and differs from the previously described techniques by the use of an alternative enzyme, Φ29 DNA polymerase [141]. The method of replication is based on a rolling strand-displacement model that proceeds by constantly displacing the newly generated strand of DNA to form

a hyper-branched DNA replication structure, limited only by the reagents available [142]. The promise shown by this method has led to the development of several commercial WGA kits, for example, GenomiPhi WGA Amplification kit (GE Healthcare) and various publications advocating its use for human genome analysis [143–147]. All of these studies, however, have entered more than 10 ng DNA into the MDA reaction. Studies conducted using less than 10 ng template DNA have observed unequal amplification of genomic DNA when assessed by the use of forensic DNA profiling kits, similar or identical to those used in [148–150]. Of the numerous WGA method variations that have been published to date, none have yet demonstrated the unbiased amplification of microsatellite markers when very low template material is available for entry into the WGA reaction [151–153].

An alternate approach is the use of a nested PCR strategy, which involves the use of two primer sets, one designed to prime from positions within the original amplicons. Using this approach, the initial PCR performs the function of enriching the template DNA specifically in the regions of ultimate interest. The second, internal, amplification reaction can then be initiated from a greatly increased initial template amount. This approach was successfully applied to the forensic DNA analysis of charred human remains [154]. This approach, however, presents a heightened risk of laboratory-introduced sample contamination by requiring the manipulation of amplified human DNA, breaking a fundamental anticontamination principle of single direction work flow during forensic DNA analysis. A far simpler approach to increasing PCR sensitivity has, however, proven to be extremely successful and become fully integrated into forensic DNA profiling laboratories worldwide.

The term *low copy number* (*LCN*) DNA profiling was coined in a publication by Peter Gill of the Forensic Science Service (FSS) in 2000 and describes the analysis of less than 100 pg template DNA [110] (*see also* **Low Copy Number DNA**). LCN has come to be associated with the use of increased PCR cycles, but there are other techniques that claim to amplify small amounts of DNA. These can be collectively termed as *low template DNA* (*LTDNA*) analysis. Commercially available STR amplification kits, such as the AmpF*l*STR® SGM Plus™ kit, are optimized to produce complete and accurate DNA profiles when 1–2.5 μl template DNA is amplified for 28 cycles. Following this standard protocol, amplification of 100 pg DNA will typically result in severe allelic and locus dropout, or even complete failure. The sensitivity of this technique can be radically improved by simply increasing the number of PCR cycles, in the case of the AmpF*l*STR® SGM Plus™ kit, to 34 cycles [110]. The increase in sensitivity opens up the potential of forensic DNA profiling to analysis of many more sample types than had ever previously been imagined, most notably from fingerprints [155, 156], cigarette butts [101], and from the skin surface of manual strangulation victims [157–159] (*see also* **DNA: Sources of** for a further description of sources of DNA eveidence).

The improved sensitivity gained by the use of an extra six PCR cycles does, however, lead to additional issues in DNA profile interpretation. When undertaking any forensic DNA analysis procedure, the importance of anticontamination protocols cannot be underestimated. This is especially true when attempting to produce LTDNA profiles, for which the template DNA may also be of a degraded nature (*see* **DNA: Degraded Samples**). The sensitivity of the AmpF*l*STR® SGM Plus™ PCR amplification kit when employing 34 cycles is such that laboratory-derived contamination cannot actually be completely avoided [110]. The potential to acquire DNA profile information from minute template amounts can also lead to the production of partial DNA profiles, where both allele and locus dropout are observed. To combat these problems, along with an increased size of stutter bands and unbalanced amplification of heterozygous alleles, an entirely new set of DNA profile interpretation rules is required for LTDNA analysis.

The majority of the complications observed for LTDNA profiles are due to stochastic variation. For example, a stutter band generated in an early PCR cycle, unassociated contaminating alleles, or one of two heterozygous alleles can be preferentially amplified and therefore over-represented in the final DNA profile [160]. The random nature of stochastic variation means that the same stutter products or contaminating alleles will not be amplified identically in replicate analyses [110, 161] leading to the first rule of LTDNA analysis; an allele can only be reported if it is present in at least two repeated analyses [160]. Once this rule has been applied, all remaining bands in the DNA profile can now be scrutinized. For standard 28 cycle DNA profiling reactions, stutter peaks

are not observed to exceed 15% of the associated allele peak height; in LCN analysis, stutter products are much larger in the range of 20% when associated with peaks >10 000 RFU in area and 40% for associated peaks <10 000 RFU [162]. As this cannot be avoided, it must always be considered that the stutter peak could be masking an allele of a minor contributor in a mixed profile, and as such cannot be discarded simply as a stochastic artifact. Similarly for heterozygosity balance, where Hb Φ values are observed to be >0.67 (when using Φ = smallest peak area/largest peak area), under standard conditions, the minimum values observed during LCN analyses for most loci in the AmpF*l*STR® SGM Plus™ PCR kit were observed at a Hb Φ value of 0.2 [162]. It was also observed that, after repeated analysis, allelic dropout at known heterozygote loci occurred at a rate of approximately 10% per locus, but was not observed when the peak area of alleles was >10 000 RFU [162].

By the varying nature of LTDNA work, the majority of samples analyzed comprise DNA mixtures of dual or multiple contributors. Mixture analysis of standard or LTDNA profiles can be an extremely complex undertaking especially when more than two or three individuals have contributed unequally to a collected sample. In a working forensic casework environment, to avoid bias, mixture analysis is carried out without any knowledge of reference DNA profiles of potential contributors. As the work carried out for inclusion in this thesis was purely research-based, the interpretation permutations and mathematical rules required for unsighted mixture analysis are not further discussed in this introduction, especially as the interpretation of complex mixtures is today more likely to be performed by computational expert systems [163] to increase the speed at which the analysis can be completed and remove any subjectivity that may be introduced by individual DNA analysts.

The advantages of using a LTDNA protocol for DNA profile generation are clear when the analysis of minute trace evidence is required. These advantages do not, however, extend to the analysis of minor contributors to mixed DNA profiles. It has been noted that the use of LTDNA profiling does not convey any advantage over standard analysis when the minor: major DNA contributor ratio is below 1 : 10 [135]. This knowledge, combined with the additional complications of sporadic contamination, increased stutter, and heterozygote imbalance, ensure that LTDNA analysis is undertaken only when absolutely necessary.

References

[1] Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. & Arnheim N. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia, *Science* **230**, 1350–1354.

[2] Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. & Erlich, H. (1986). Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction, *Cold Spring Harbor Symposia on Quantitative Biology* **51**(Pt 1), 263–273.

[3] Mullis, K.B. (1990). The unusual origin of the polymerase chain reaction, *Scientific American* **262**, 56–61, 64, 65.

[4] Jeffreys, A.J., Wilson, V. & Thein, S.L. (1985). Hypervariable 'minisatellite' regions in human DNA, *Nature* **314**, 67–73.

[5] Gill, P., Jeffreys, A.J. & Werrett, D.J. (1985). Forensic application of DNA 'fingerprints', *Nature* **318**, 577–579.

[6] Jeffreys, A.J., Wilson, V. & Thein, S.L. (1985). Individual-specific 'fingerprints' of human DNA, *Nature* **316**, 76–79.

[7] Gill, P. & Werrett, D.J. (1987). Exclusion of a man charged with murder by DNA fingerprinting, *Forensic Science International* **35**, 145–148.

[8] Zagorski, N. (2006). Profile of Alec J. Jeffreys, *Proceedings of the National Academy of Sciences of the United States of America* **103**, 8918–8920.

[9] Gill, P., Lygo, J.E., Fowler, S.J. & Werrett, D.J. (1987). An evaluation of DNA fingerprinting for forensic purposes, *Electrophoresis* **8**, 38–44.

[10] Gill, P., Woodroffe, S., Bar, W., Brinkmann, B., Carracedo, A., Eriksen, B., Jones, S., Kloosterman, A.D., Ludes, B. & Mevag, B. (1992). A report of an international collaborative experiment to demonstrate the uniformity obtainable using DNA profiling techniques, *Forensic Science International* **53**, 29–43.

[11] Schneider, P.M., Fimmers, R., Woodroffe, S., Werrett, D.J., Bar, W., Brinkmann, B., Eriksen, B., Jones, S., Kloosterman, A.D. & Mevag, B., *et al.* (1991). Report of a European collaborative exercise comparing DNA typing results using a single locus VNTR probe, *Forensic Science International* **49**, 1–15.

[12] Evett, I.W., Werrett, D.J., Gill, P. & Buckleton, J.S. (1989). DNA fingerprinting on trial, *Nature* **340**, 435.

[13] Lander, E.S. (1989). DNA fingerprinting on trial, *Nature* **339**, 501–505.

[14] Helmuth, R., Fildes, N., Blake, E., Luce, M.C., Chimera, J., Madej, R., Gorodezky, C., Stoneking, M., Schmill, N. & Klitz, W. *et al.* (1990). HLA-DQ alpha

allele and genotype frequencies in various human populations, determined by using enzymatic amplification and oligonucleotide probes, *American Journal of Human Genetics* **47**, 515–523.

[15] Weber, J.L. & May, P.E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction, *American Journal of Human Genetics* **44**, 388–396.

[16] Ziegle, J.S., Su, Y., Corcoran, K.P., Nie, L., Mayrand, P.E., Hoff, L.B., McBride, L.J., Kronick, M.N. & Diehl, S.R. (1992). Application of automated DNA sizing technology for genotyping microsatellite loci, *Genomics* **14**, 1026–1031.

[17] Sullivan, K.M., Pope, S., Gill, P. & Robertson, J.M. (1992). Automated DNA profiling by fluorescent labeling of PCR products, *PCR Methods and Applications* **2**, 34–40.

[18] Kimpton, C.P., Gill, P., Walton, A., Urquhart, A., Millican, E.S. & Adams, M. (1993). Automated DNA profiling employing multiplex amplification of short tandem repeat loci, *PCR Methods and Applications* **3**, 13–22.

[19] Sullivan, K.M., Hopgood, R., Lang, B. & Gill, P. (1991). Automated amplification and sequencing of human mitochondrial DNA, *Electrophoresis* **12**, 17–21.

[20] Evett, I.W. & Gill, P. (1991). A discussion of the robustness of methods for assessing the evidential value of DNA single locus profiles in crime investigations, *Electrophoresis* **12**, 226–230.

[21] Evett, I.W. & Pinchin, R. (1991). DNA single locus profiles: tests for the robustness of statistical procedures within the context of forensic science, *International Journal of Legal Medicine* **104**, 267–272.

[22] Lygo, J.E., Johnson, P.E., Holdaway, D.J., Woodroffe, S., Whitaker, J.P., Clayton, T.M., Kimpton, C.P. & Gill, P. (1994). The validation of short tandem repeat (STR) loci for use in forensic casework, *International Journal of Legal Medicine* **107**, 77–89.

[23] Beckmann, J.S. & Weber, J.L. (1992). Survey of human and rat microsatellites, *Genomics* **12**, 627–631.

[24] Edwards, A., Civitello, A., Hammond, H.A. & Caskey, C.T. (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats, *American Journal of Human Genetics* **49**, 746–756.

[25] Clayton, T.M., Whitaker, J.P., Fisher, D.L., Lee, D.A., Holland, M.M., Weedn, V.W., Maguire, C.N., DiZinno, J.A., Kimpton, C.P. & Gill, P. (1995). Further validation of a quadruplex STR DNA typing system: a collaborative effort to identify victims of a mass disaster, *Forensic Science International* **76**, 17–25.

[26] Clayton, T.M., Whitaker, J.P. & Maguire, C.N. (1995). Identification of bodies from the scene of a mass disaster using DNA amplification of short tandem repeat (STR) loci, *Forensic Science International* **76**, 7–15.

[27] Kimpton, C.P., Oldroyd, N.J., Watson, S.K., Frazier, R.R., Johnson, P.E., Millican, E.S., Urquhart, A., Sparkes, B.L. & Gill, P. (1996). Validation of highly discriminating multiplex short tandem repeat amplification systems for individual identification, *Electrophoresis* **17**, 1283–1293.

[28] Gill, P., Brinkmann, B., d'Aloja, E., Andersen, J., Bar, W., Carracedo, A., Dupuy, B., Eriksen, B., Jangblad, M., Johnsson, V., Kloosterman, A.D., Lincoln, P., Morling, N., Rand, S., Sabatier, M., Scheithauer, R., Schneider, P. & Vide, M.C. (1997). Considerations from the European DNA profiling group (EDNAP) concerning STR nomenclature, *Forensic Science International* **87**, 185–192.

[29] Sullivan, K.M., Mannucci, A., Kimpton, C.P. & Gill, P. (1993). A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin, *Biotechniques* **15**, 636–638, 640, 641.

[30] Sparkes, R., Kimpton, C., Watson, S., Oldroyd, N., Clayton, T., Barnett, L., Arnold, J., Thompson, C., Hale, R., Chapman, J., Urquhart, A. & Gill, P. (1996). The validation of a 7-locus multiplex STR test for use in forensic casework. (I). Mixtures, ageing, degradation and species studies, *International Journal of Legal Medicine* **109**, 186–194.

[31] Evett, I.W., Gill, P.D., Lambert, J.A., Oldroyd, N., Frazier, R., Watson, S., Panchal, S., Connolly, A. & Kimpton, C. (1997). Statistical analysis of data for three British ethnic groups from a new STR multiplex, *International Journal of Legal Medicine* **110**, 5–9.

[32] Cotton, E.A., Allsop, R.F., Guest, J.L., Frazier, R.R., Koumi, P., Callow, I.P., Seager, A. & Sparkes, R.L. (2000). Validation of the AMPFlSTR SGM plus system for use in forensic casework, *Forensic Science International* **112**, 151–161.

[33] Foreman, L.A. & Evett, I.W. (2001). Statistical analyses to support forensic interpretation for a new ten-locus STR profiling system, *International Journal of Legal Medicine* **114**, 147–155.

[34] Li, H., Schmidt, L., Wei, M.H., Hustad, T., Lerman, M.I., Zbar, B. & Tory, K. (1993). Three tetranucleotide polymorphisms for loci: D3S1353; D3S1358; D3S1359, *Human Molecular Genetics* **2**, 1327.

[35] Kimpton, C., Walton, A. & Gill, P. (1992). A further tetranucleotide repeat polymorphism in the vWF gene, *Human Molecular Genetics* **1**, 287.

[36] Watson, A., Mazumder, A., Stewart, M. & Balasubramanian, S. (1998). Technology for microarray analysis of gene expression, *Current Opinion in Biotechnology* **9**, 609–614.

[37] Oldroyd, N.J., Urquhart, A.J. & Kimpton, C.P., *et al.* (1995). A highly discriminating octoplex short tandem repeat polymerase chain reaction system suitable for human individual identification, *Electrophoresis* **16**, 334–337.

[38] Sharma, V. & Litt, M. (1992). Tetranucleotide repeat polymorphism at the D21S11 locus, *Human Molecular Genetics* **1**, 67.

[39] Urquhart, A., Oldroyd, N.J., Kimpton, C.P. & Gill, P. (1995). Highly discriminating heptaplex short tandem

repeat PCR system for forensic identification, *Biotechniques* **18**, 116–118, 120, 121.

[40] Lareu, M.V., Barral, S., Salas, A., Rodriguez, M., Pestoni, C. & Carracedo, A. (1998). Further exploration of new STRs of interest for forensic genetic analysis, in *Progress in Forensic Genetics 7: Proceedings of the 17th International ISFH Congress*, Elsevier, Oslo, pp. 192–200.

[41] Edwards, A., Hammond, H.A., Jin, L., Caskey, C.T. & Chakraborty, R. (1992). Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups, *Genomics* **12**, 241–253.

[42] Mills, K.A., Even, D. & Murray, J.C. (1992). Tetranucleotide repeat polymorphism at the human alpha fibrinogen locus (FGA), *Human Molecular Genetics* **1**, 779.

[43] Ruitberg, C.M., Reeder, D.J. & Butler, J.M. (2001). STRBase: a short tandem repeat DNA database for the human identity testing community, *Nucleic Acids Research* **29**, 320–322.

[44] Harris, P.C. (2004). Gene hunters find cause of childhood polycystic kidney disease, *Inside iScience* **1**, 10.

[45] Frazier, R.R., Millican, E.S., Watson & S.K., *et al.* (1996). Validation of the Applied Biosystems Prism 377 automated sequencer for the forensic short tandem repeat analysis, *Electrophoresis* **17**, 1550–1552.

[46] Risch, N. & Merikangas, K. (1996). The future of genetic studies of complex human diseases, *Science* **273**, 1516–1517.

[47] Collins, F.S., Guyer, M.S. & Charkravarti, A. (1997). Variations on a theme: cataloging human DNA sequence variation, *Science* **278**, 1580–1581.

[48] Dixon, L.A., Murray, C.M., Archer, E.J., Dobbins, A.E., Koumi, P. & Gill, P. (2005). Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes, *Forensic Science International* **154**, 62–77.

[49] Graham, E.A.M. (2005). DNA reviews: mini STR's, *Forensic Science, Medicine, and Pathology* **1**, 65–68.

[50] Gill, P. (2001). An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes, *International Journal of Legal Medicine* **114**, 204–210.

[51] Strum, R.A. & Frudakis, T.N. (2004). Eye colour: portals into pigmentation genes and ancestry, *Trends in Genetics* **20**, 327–332.

[52] Sturm, R.A. & Frudakis, T.N. (2004). Eye colour: portals into pignmentation genes and ancestry, *Trends in Genetics* **20**, 327–332.

[53] Grimes, E.A., Noake, P.J., Dixon, L. & Urquhart, A. (2001). Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype, *Forensic Science International* **122**, 124–129.

[54] Rana, B.K. (1999). High polymorphism at the human melanocortin 1 receptor locus, *Genetics* **151**, 1547–1557.

[55] Sturm, R.A. (2002). Skin colour and skin cancer – MC1R, the genetic link, *Melanoma Research* **12**, 405–416.

[56] Graham, E.A.M. (2006). DNA reviews: sex determination, *Forensic Science, Medicine, and Pathology* 2.

[57] Graham, E.A.M. (2006). DNA reviews: forensic DNA on the internet, *Forensic Science, Medicine, and Pathology* **2**, 63–66.

[58] Higuchi, R., von Beroldingen, C.H., Sensabaugh, G.F. & Erlich, H.A. (1988). DNA typing from single hairs, *Nature* **332**, 543–546.

[59] Schreiber, A., Amtmann, E., Storch, V. & Sauer, G. (1988). The extraction of high-molecular-mass DNA from hair shafts, *FEBS Letters* **230**, 209–211.

[60] Hagelberg, E., Sykes, B. & Hedges, R. (1989). Ancient bone DNA amplified, *Nature* **342**, 485.

[61] Hagelberg, E. & Clegg, J.B. (1991). Isolation and characterization of DNA from archaeological bone, *Proceedings of the Royal Society of London, Series B: Biological Sciences* **244**, 45–50.

[62] Walsh, P.S., Metzger, D.A. & Higuchi, R. (1991). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material, *Biotechniques* **10**, 506–513.

[63] Greenspoon, S.A., Scarpetta, M.A., Drayton, M.L. & Turek, S.A. (1998). QIAamp spin columns as a method of DNA isolation for forensic casework, *Journal of Forensic Sciences* **43**, 1024–1030.

[64] Scherczinger, C.A., Bourke, M.T., Ladd, C. & Lee, H.C. (1997). DNA extraction from liquid blood using QIAamp, *Journal of Forensic Sciences* **42**, 893–896.

[65] Yang, D.Y., Eng, B., Waye, J.S., Dudar, J.C. & Saunders, S.R. (1998). Technical note: improved DNA extraction from ancient bones using silica-based spin columns, *American Journal of Physical Anthropology* **105**, 539–543.

[66] Ye, J., Ji, A., Parra, E.J., Zheng, X., Jiang, C., Zhao, X., Hu, L. & Tu, Z. (2004). A simple and efficient method for extracting DNA from old and burned bone, *Journal of Forensic Sciences* **49**, 754–759.

[67] Alonso, A., Andelinovic, S., Martin, P., Sutlovic, D., Erceg, I., Huffine, E., de Simon, L.F., Albarran, C., Definis-Gojanovic, M., Fernandez-Rodriguez, A., Garcia, P., Drmic, I., Rezic, B., Kuret, S., Sancho, M. & Primorac, D. (2001). DNA typing from skeletal remains: evaluation of multiplex and megaplex STR systems on DNA isolated from bone and teeth samples, *Croatian Medical Journal* **42**, 260–266.

[68] Evison, M.P., Smillie, D.M. & Chamberlain, A.T. (1997). Extraction of single-copy nuclear DNA from forensic specimens with a variety of postmortem histories, *Journal of Forensic Sciences* **42**, 1032–1038.

[69] Boom, R., Sol, C.J., Salimans, M.M., Jansen, C.L., Wertheim-van Dillen, P.M. & van der Noordaa, J. (1990). Rapid and simple method for purification of nucleic acids, *Journal of Clinical Microbiology* **28**, 495–503.

[70] Hoss, M. & Paabo, S. (1993). DNA extraction from Pleistocene bones by a silica-based purification method, *Nucleic Acids Research* **21**, 3913–3914.

[71] Kalmar, T., Bachrati, C.Z., Marcsik, A. & Rasko, I. (2000). A simple and efficient method for PCR amplifiable DNA extraction from ancient bones, *Nucleic Acids Research* **28**, E67.

[72] Hochmeister, M.N., Budowle, B., Borer, U.V., Eggmann, U., Comey, C.T. & Dirnhofer, R. (1991). Typing of deoxyribonucleic acid (DNA) extracted from compact bone from human remains, *Journal of Forensic Sciences* **36**, 1649–1661.

[73] Rennick, S.L., Fenton, T.W. & Foran, D.R. (2005). The effects of skeletal preparation techniques on DNA from human and non-human bone, *Journal of Forensic Sciences* **50**, 1016–1019.

[74] Tsuchimochi, T., Iwasa, M., Maeno, Y. Koyama, H., Inoue, H., Isobe, I., Matoba, R., Yokoi, M. & Nagao, M. (2002). Chelating resin-based extraction of DNA from dental pulp and sex determination from incinerated teeth with Y-chromosomal alphoid repeat and short tandem repeats, *The American Journal of Forensic Medicine and Pathology* **23**, 268–271.

[75] Trivedi, R., Chattopadhyay, P. & Kashyap, V.K. (2002). A new improved method for extraction of DNA from teeth for the analysis of hypervariable loci, *The American Journal of Forensic Medicine and Pathology* **23**, 191–196.

[76] Sivagami, A.V., Rao, A.R. & Varshney, U. (2000). A simple and cost-effective method for preparing DNA from the hard tooth tissue, and its use in polymerase chain reaction amplification of amelogenin gene segment for sex determination in an Indian population, *Forensic Science International* **110**, 107–115.

[77] De Leo, D., Turrina, S. & Marigo, M. (2000). Effects of individual dental factors on genomic DNA analysis, *The American Journal of Forensic Medicine and Pathology* **21**, 411–415.

[78] Malaver, P.C. & Yunis, J.J. (2003). Different dental tissues as source of DNA for human identification in forensic cases, *Croatian Medical Journal* **44**, 306–309.

[79] Pfeiffer, H., Huhne, J., Seitz, B. & Brinkmann, B. (1999). Influence of soil storage and exposure period on DNA recovery from teeth, *International Journal of Legal Medicine* **112**, 142–144.

[80] Vigilant, L. (1999). An evaluation of techniques for the extraction and amplification of DNA from naturally shed hairs, *Biological Chemistry* **380**, 1329–1331.

[81] Graffy, E.A. & Foran, D.R. (2005). A simplified method for mitochondrial DNA extraction from head hair shafts, *Journal of Forensic Sciences* **50**, 1119–1122.

[82] Pfeiffer, I., Volkel, I., Taubert, H. & Brenig, B. (2004). Forensic DNA-typing of dog hair: DNA-extraction and PCR amplification, *Forensic Science International* **141**, 149–151.

[83] Sweet, D. & Hildebrand, D. (1999). Saliva from cheese bite yields DNA profile of burglar: a case report, *International Journal of Legal Medicine* **112**, 201–203.

[84] Sweet, D., Lorente, M., Lorente, J.A., Valenzuela, A. & Villanueva, E. (1997). An improved method to recover saliva from human skin: the double swab technique, *Journal of Forensic Sciences* **42**, 320–322.

[85] Sweet, D., Lorente, M., Valenzuela, A., Lorente, J.A. & Alvarez, J.C. (1996). Increasing DNA extraction yield from saliva stains with a modified Chelex method, *Forensic Science International* **83**, 167–177.

[86] Kato, Y., Katsumata, R., Yoshimoto, T., Tanaka, M., Huang, X.L., Tamaki, K., Kumazawa, T., Sato, K. & Katsumata, Y. (1999). Large-scale preparation of high-molecular weight DNA from buccal mucosa, *Legal Medicine (Tokyo)* **1**, 6–10.

[87] London, S.J., Xia, J., Lehman, T.A., Yang, J.H., Granada, E., Chunhong, L., Dubeau, L., Li, T., David-Beabes, G.L. & Li, Y. (2001). Collection of buccal cell DNA in seventh-grade children using water and a toothbrush, *Cancer Epidemiology, Biomarkers and Prevention* **10**, 1227–1230.

[88] Hanselle, T., Otte, M., Schnibbe, T., Smythe, E. & Krieg-Schneider, F. (2003). Isolation of genomic DNA from buccal swabs for forensic analysis, using fully automated silica-membrane purification technology, *Legal Medicine (Tokyo)* **5**(Suppl 1), S145–S149.

[89] Zehner, R., Amendt, J. & Krettek, R. (2004). STR typing of human DNA from fly larvae fed on decomposing bodies, *Journal of Forensic Sciences* **49**, 337–340.

[90] Linville, J.G., Hayes, J. & Wells, J.D. (2004). Mitochondrial DNA and STR analyses of maggot crop contents: effect of specimen preservation technique, *Journal of Forensic Sciences* **49**, 341–344.

[91] Linville, J.G. & Wells, J.D. (2002). Surface sterilization of a maggot using bleach does not interfere with mitochondrial DNA analysis of crop contents, *Journal of Forensic Sciences* **47**, 1055–1059.

[92] Johnson, D.J., Martin, L.R. & Roberts, K.A. (2005). STR-typing of human DNA from human fecal matter using the QIAGEN QIAamp stool mini kit, *Journal of Forensic Sciences* **50**, 802–808.

[93] Vandenberg, N. & van Oorschot, R.A. (2002). Extraction of human nuclear DNA from feces samples using the QIAamp DNA stool mini kit, *Journal of Forensic Sciences* **47**, 993–995.

[94] Roy, R. (2003). Analysis of human fecal material for autosomal and Y chromosome STRs, *Journal of Forensic Sciences* **48**, 1035–1040.

[95] McOrist, A.L., Jackson, M. & Bird, A.R. (2002). A comparison of five methods for extraction of bacterial DNA from human faecal samples, *Journal of Microbiological Methods* **50**, 131–139.

[96] Nakazono, T., Kashimura, S., Hayashiba, Y., Hara, K. & Miyoshi, A. (2005). Successful DNA typing of urine stains using a DNA purification kit following diafiltration, *Journal of Forensic Sciences* **50**, 860–864.

[97] Yasuda, T., Iida, R., Takeshita, H. Ueki, M., Nakajima, T., Kaneko, Y., Mogi, K., Tsukahara, T. & Kishi, K. (2003). A simple method of DNA extraction and STR typing from urine samples using a commercially available DNA/RNA extraction kit, *Journal of Forensic Sciences* **48**, 108–110.

[98] Drobnic, K. (2003). Analysis of DNA evidence recovered from epithelial cells in penile swabs, *Croatian Medical Journal* **44**, 350–354.

[99] Bright, J.A. & Petricevic, S.F. (2004). Recovery of trace DNA and its application to DNA profiling of shoe insoles, *Forensic Science International* **145**, 7–12.

[100] Szibor, R., Michael, M., Plate, I. & Krause, D. (2000). Efficiency of forensic mtDNA analysis. Case examples demonstrating the identification of traces, *Forensic Science International* **113**, 71–78.

[101] Watanabe, Y., Takayama, T., Hirata, K., Yamada, S., Nagai, A., Nakamura, I., Bunai, Y. & Ohya, I. (2003). DNA typing from cigarette butts, *Legal Medicine (Tokyo)* **5**(Suppl 1), S177–S179.

[102] Esslinger, K.J., Siegel, J.A., Spillane, H. & Stallworth, S. (2004). Using STR analysis to detect human DNA from exploded pipe bomb devices, *Journal of Forensic Sciences* **49**, 481–484.

[103] Petricevic, S.F., Bright, J.A. & Cockerton, S.L. (2006). DNA profiling of trace DNA recovered from bedding, *Forensic Science International.* **159**, 21–26.

[104] Abaz, J., Walsh, S.J., Curran, J.M., Moss, D.S., Cullen, J., Bright, J.A., Crowe, G.A., Cockerton, S.L. & Power, T.E. (2002). Comparison of the variables affecting the recovery of DNA from common drinking containers, *Forensic Science International* **126**, 233–240.

[105] Cattaneo, C., Craig, O.E., James, N.T. & Sokol, R.J. (1997). Comparison of three DNA extraction methods on bone and blood stains up to 43 years old and amplification of three different gene sequences, *Journal of Forensic Sciences* **42**, 1126–1135.

[106] Greenspoon, S.A., Ban, J.D., Sykes, K., Ballard, E.J., Edler, S.S., Baisden, M. & Covington, B.L. (2004). Application of the BioMek 2000 laboratory automation workstation and the DNA IQ System to the extraction of forensic casework samples, *Journal of Forensic Sciences* **49**, 29–39.

[107] Greenspoon, S.A., Sykes, K.L., Ban, J.D., Pollard, A., Baisden, M., Farr, M., Graham, N., Collins, B.L., Green, M.M. & Christenson, C.C. (2006). Automated PCR setup for forensic casework samples using the Normalization Wizard and PCR Setup robotic methods, *Forensic Science International* **164**, 240–248.

[108] Montpetit, S.A., Fitch, I.T. & O'Donnell, P.T. (2005). A simple automated instrument for DNA extraction in forensic casework, *Journal of Forensic Sciences* **50**, 555–563.

[109] Gill, P., Sparkes, R., Fereday, L. & Werrett, D.J. (2000). Report of the European Network of Forensic Science Institutes (ENSFI): formulation and testing of principles to evaluate STR multiplexes, *Forensic Science International* **108**, 1–29.

[110] Gill, P., Whitaker, J., Flaxman, C., Brown, N. & Buckleton, J. (2000). An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA, *Forensic Science International* **112**, 17–40.

[111] Haque, K., Pfeiffer, R., Beerman, M., Struewing, J., Chanock, S. & Bergen, A. (2003). Performance of high-throughput DNA quantification methods, *BMC Biotechnology* **3**, 20.

[112] Singer, V.L., Jones, L.J., Yue, S.T. & Haugland, R.P. (1997). Characterization of PicoGreen reagent and development of a fluorescence-based solution assay for double-stranded DNA quantitation, *Analytical Biochemistry* **249**, 228–238.

[113] Walsh, P.S., Varlaro, J. & Reynolds, R. (1992). A rapid chemiluminescent method for quantitation of human DNA, *Nucleic Acids Research* **20**, 5061–5065.

[114] Budowle, B., Hudlow, W.R., Lee, S.B. & Klevan, L. (2001). Using a CCD camera imaging system as a recording device to quantify human DNA by slot blot hybridization, *Biotechniques* **30**, 680–685.

[115] Allen, R.W. & Fuller, V.M. (2006). Quantitation of human genomic DNA through amplification of the amelogenin locus, *Journal of Forensic Sciences* **51**, 76–81.

[116] Hayn, S., Wallace, M.M., Prinz, M. & Shaler, R.C. (2004). Evaluation of an automated liquid hybridization method for DNA quantitation, *Journal of Forensic Sciences* **49**, 87–91.

[117] Carrondo, M.A., Coll, M., Aymami, J., Wang, A.H., van der Marel, G.A., van Boom, J.H. & Rich, A. (1989). Binding of a Hoechst dye to d(CGCGATATCGCG) and its influence on the conformation of the DNA fragment, *Biochemistry* **28**, 7849–7859.

[118] Ahn, S., Costa, J. & Emanuel, J. (1996). PicoGreen quantitation of DNA: effective evaluation of samples pre- or post-PCR 10.1093/nar/24.13.2623, *Nucleic Acids Research* **24**, 2623–2625.

[119] Higuchi, R., Dollinger, G., Walsh, P.S. & Griffith, R. (1992). Simultaneous amplification and detection of specific DNA sequences, *Bio/technology (Nature Publishing Company)* **10**, 413–417.

[120] Higuchi, R., Fockler, C., Dollinger, G. & Watson, R. (1993). Kinetic PCR analysis: real-time monitoring of DNA amplification reactions, *Biotechnology (N Y)* **11**, 1026–1030.

[121] Nicklas, J.A. & Buel, E. (2003). Development of an Alu-based, real-time PCR method for quantitation of human DNA in forensic samples, *Journal of Forensic Sciences* **48**, 936–944.

[122] Urban, C., Gruber, F., Kundi, M., Falkner, F.G., Dorner, F. & Hammerle, T. (2000). A systematic and quantitative analysis of PCR template contamination, *Journal of Forensic Sciences* **45**, 1307–1311.

[123] Walker, J.A., Kilroy, G.E., Xing, J., Shewale, J., Sinha, S.K. & Batzer, M.A. (2003). Human DNA quantitation using Alu element-based polymerase chain reaction, *Analytical Biochemistry* **315**, 122–128.

[124] Alonso, A., Martin, P., Albarran, C., Garcia, P., Garcia, O., de Simon, L.F., Garcia-Hirschfeld, J., Sancho, M., de La Rua, C. & Fernandez-Piqueras, J. (2004). Real-time PCR designs to estimate nuclear and mitochondrial DNA copy number in forensic and ancient DNA studies, *Forensic Science International* **139**, 141–149.

[125] Andelinovic, S., Sutlovic, D., Erceg Ivkosic, I. Skaro, V., Ivkosic, A., Paic, F., Rezic, B., Definis-Gojanovic, M. & Primorac, D. (2005). Twelve-year experience in identification of skeletal remains from mass graves, *Croatian Medical Journal* **46**, 530–539.

[126] Andreasson, H., Nilsson, M., Budowle, B., Lundberg, H. & Allen, M. (2006). Nuclear and mitochondrial DNA quantification of various forensic materials, *Forensic Science International* **164**, 56–64.

[127] Andreasson, H., Gyllensten, U. & Allen, M. (2002). Real-time DNA quantification of nuclear and mitochondrial DNA in forensic analysis, *Biotechniques* **33**, 402–404, 407–411.

[128] Timken, M.D., Swango, K.L., Orrego, C. & Buoncristiani, M.R. (2005). A duplex real-time qPCR assay for the quantification of human nuclear and mitochondrial DNA in forensic samples: implications for quantifying DNA in degraded samples, *Journal of Forensic Sciences* **50**, 1044–1060.

[129] Swango, K.L., Hudlow, W.R., Timken, M.D. & Buoncristiani, M.R. (2006). Developmental validation of a multiplex qPCR assay for assessing the quantity and quality of nuclear DNA in forensic samples, *Forensic Science International* **170**, 35–45.

[130] Schneider, P.M. (1997). Basic issues in forensic DNA typing, *Forensic Science International* **88**, 17–22.

[131] Sajantila, A., Puomilahti, S., Johnsson, V. & Ehnholm, C. (1992). Amplification of reproducible allele markers for amplified fragment length polymorphism analysis, *Biotechniques* **12**, 16, 18, 20–22.

[132] Smith, R.N. (1995). Accurate size comparison of short tandem repeat alleles amplified by PCR, *Biotechniques* **18**, 122–128.

[133] Gill, P., Kimpton, C.P., Urquhart, A., Oldroyd, N., Millican, E.S., Watson, S.K. & Downes, T.J. (1995). Automated short tandem repeat (STR) analysis in forensic casework – a strategy for the future, *Electrophoresis* **16**, 1543–1552.

[134] Gill, P., Sparkes, R. & Kimpton, C. (1997). Development of guidelines to designate alleles using an STR multiplex system, *Forensic Science International* **89**, 185–197.

[135] Kloosterman, A.D. & Kersbergen, P. (2003). Efficacy and limits of genotyping low copy number (LCN) DNA samples by multiplex PCR of STR loci, *Journal de la Société de Biologie* **197**, 351–359.

[136] Ludecke, H.J., Senger, G., Claussen, U. & Horsthemke, B. (1989). Cloning defined regions of the human genome by microdissection of banded chromosomes and enzymatic amplification, *Nature* **338**, 348–350.

[137] Zhang, L., Cui, X., Schmitt, K., Hubert, R., Navidi, W. & Arnheim, N. (1992). Whole genome amplification from a single cell: implications for genetic analysis, *Proceedings of the National Academy of Sciences of the United States of America* **89**, 5847–5851.

[138] Kristjansson, K., Chong, S.S., Van den Veyver, I.B., Subramanian, S., Snabes, M.C. & Hughes, M.R. (1994). Preimplantation single cell analyses of dystrophin gene deletions using whole genome amplification, *Nature Genetics* **6**, 19–23.

[139] Snabes, M.C., Chong, S.S., Subramanian, S.B., Kristjansson, K., DiSepio, D. & Hughes, M.R. (1994). Preimplantation single-cell analysis of multiple genetic loci by whole-genome amplification, *Proceedings of the National Academy of Sciences of the United States of America* **91**, 6181–6185.

[140] Faulkner, S.W. & Leigh, D.A. (1998). Universal amplification of DNA isolated from small regions of paraffin-embedded, formalin-fixed tissue, *Biotechniques* **24**, 47–50.

[141] Dean, F.B., Nelson, J.R., Giesler, T.L. & Lasken, R.S. (2001). Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification, *Genome Research* **11**, 1095–1099.

[142] Dean, F.B., Hosono, S., Fang, L., Wu, X., Faruqi, A.F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J., Driscoll, M., Song, W., Kingsmore, S.F., Egholm, M. & Lasken, R.S. (2002). Comprehensive human genome amplification using multiple displacement amplification, *Proceedings of the National Academy of Sciences of the United States of America* **99**, 5261–5266.

[143] Hosono, S., Faruqi, A.F., Dean, F.B. Du, Y., Sun, Z., Wu, X., Du, J., Kingsmore, S.F., Egholm, M. & Lasken, R.S. (2003). Unbiased whole-genome amplification directly from clinical samples, *Genome Research* **13**, 954–964.

[144] Lovmar, L., Fredriksson, M., Liljedahl, U., Sigurdsson, S. & Syvanen, A.C. (2003). Quantitative evaluation by minisequencing and microarrays reveals accurate multiplexed SNP genotyping of whole genome amplified DNA, *Nucleic Acids Research* **31**, e129.

[145] Murthy, K.K., Mahboubi, V.S., Santiago, A., Barragan, M.T., Knoll, R., Schultheiss, H.P., O'Connor, D.T., Schork, N.J. & Rana, B.K. (2005). Assessment of multiple displacement amplification for polymorphism discovery and haplotype determination at a highly polymorphic locus, MC1R, *Human Mutation* **26**, 145–152.

[146] Paez, J.G., Lin, M., Beroukhim, R., Lee, J.C., Zhao, X., Richter, D.J., Gabriel, S., Herman, P., Sasaki, H., Altshuler, D., Li, C., Meyerson, M. & Sellers, W.R. (2004). Genome coverage and sequence fidelity of phi29 polymerase-based multiple strand displacement whole genome amplification, *Nucleic Acids Research* **32**, e71.

[147] Sun, G., Kaushal, R., Pal, P., Wolujewicz, M., Smelser, D., Cheng, H., Lu, M., Chakraborty, R., Jin, L. & Deka, R. (2005). Whole-genome amplification: relative

efficiencies of the current methods, *Legal Medicine (Tokyo)* **7**, 279–286.
[148] Barber, A.L. & Foran, D.R. (2006). The utility of whole genome amplification for typing compromised forensic samples, *Journal of Forensic Sciences* **51**, 1344–1349.
[149] Hughes, S., Arneson, N., Done, S. & Squire, J. (2005). The use of whole genome amplification in the study of human disease, *Progress in Biophysics and Molecular Biology* **88**, 173–189.
[150] Hughes, S., Yoshimoto, M., Beheshti, B., Houlston, R.S., Squire, J.A. & Evans, A. (2006). The use of whole genome amplification to study chromosomal changes in prostate cancer: insights into genome-wide signature of preneoplasia associated with cancer progression, *BMC Genomics* **7**, 65.
[151] Hanson, E.K. & Ballantyne, J. (2005). Whole genome amplification strategy for forensic genetic analysis using single or few cell equivalents of genomic DNA, *Analytical Biochemistry* **346**, 246–257.
[152] Lee, C.I., Leong, S.H., Png, A.E., Choo, K.W., Syn, C., Lim, D.T., Law, H.Y. & Kon, O.L. (2006). An isothermal primer extension method for whole genome amplification of fresh and degraded DNA: applications in comparative genomic hybridization, genotyping and mutation screening, *Nature Protocols* **1**, 2185–2194.
[153] Lee, C.I., Leong, S.H. & Png, A.E., *et al.* (2006). An isothermal method for whole genome amplification of fresh and degraded DNA for comparative genomic hybridization, genotyping and mutation detection, *DNA Research* **13**, 77–88.
[154] Strom, C.M. & Rechitsky, S. (1998). Use of nested PCR to identify charred human remains and minute amounts of blood, *Journal of Forensic Sciences* **43**, 696–700.
[155] van Oorschot, R.A. & Jones, M.K. (1997). DNA fingerprints from fingerprints, *Nature* **387**, 767.
[156] Alessandrini, F., Cecati, M., Pesaresi, M., Turchi, C., Carle, F. & Tagliabracci, A. (2003). Fingerprints as evidence for a genetic profile: morphological study on fingerprints and analysis of exogenous and individual factors affecting DNA typing, *Journal of Forensic Sciences* **48**, 586–592.
[157] Rutty, G.N. (2002). An investigation into the transference and survivability of human DNA following simulated manual strangulation with consideration of the problem of third party contamination, *International Journal of Legal Medicine* **116**, 170–173.
[158] Wiegand, P. & Kleiber, M. (1997). DNA typing of epithelial cells after strangulation, *International Journal of Legal Medicine* **110**, 181–183.
[159] Bohnert, M., Faller-Marquardt, M., Lutz, S., Amberg, R., Weisser, H.J. & Pollak, S. (2001). Transfer of biological traces in cases of hanging and ligature strangulation, *Forensic Science International* **116**, 107–115.
[160] Gill, P. (2001). Application of low copy number DNA profiling, *Croatian Medical Journal* **42**, 229–232.
[161] Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., Waits, L.P. & Bouvet, J. (1996). Reliable genotyping of samples with very low DNA quantities using PCR, *Nucleic Acids Research* **24**, 3189–3194.
[162] Whitaker, J.P., Cotton, E.A. & Gill, P. (2001). A comparison of the characteristics of profiles produced with the AMPFlSTR SGM Plus multiplex system for both standard and low copy number (LCN) STR DNA analysis, *Forensic Science International* **123**, 215–223.
[163] Graham, E.A.M. (2005). DNA Reviews: Automated DNA profile analysis, *Forensic Science, Medicine, and Pathology* **1**, 285–288.

ELEANOR A.M. GRAHAM

DNA: and Fingerprints *see* Friction Ridge Skin: Interaction between Fingerprint Detection and DNA/Biological Material

DNA: Degraded Samples

Introduction

DNA is a relatively stable macromolecule and under certain circumstances has been known to persist for tens of thousands of years (*see* **Archaeology**; **DNA**) [1–3]. Samples of human DNA that are decades [4–6] and even centuries [2] old have been amenable to genotyping for forensic purposes as well. However, the environmental conditions to which most evidentiary samples are exposed are usually much less conducive to the preservation of the information content of the DNA molecules. Exposure to UV irradiation from sunlight, as well as to warm, moist environments has been found to result in degradation of DNA within a matter of hours [7]. Evidence samples that begin with only trace amounts of DNA are particularly at risk of only being partially detected due to degradation and/or inhibition of PCR amplification.

Characteristics of Degraded DNA Samples

It has been widely observed that alleles corresponding to larger fragments of DNA typically exhibit weaker signals/intensity than smaller alleles after exposure to some environments, ostensibly because they provide a larger target for damage to be accumulated [1].

In the absence of degradation and stochastic effects due to small sample sizes, the amount of genomic template associated with any given locus in an evidence sample should be equivalent (stoichiometric). Given that the amount of product generated during PCR amplification is generally proportional to the amount of starting template in multiplex reactions [8], total peak height or area between alleles and loci should be roughly equivalent. As a result, progressively falling peak heights from small to large (left to right) DNA fragments on electropherograms are commonly considered by forensic DNA testing laboratories to be an indication of degradation (Figure 1). However, the absence of laboratory quantitative thresholds associated with these trends has made declarations of degradation subjective and commonly supported simply by an examiner's "past training and experience". Difficulties in distinguishing between the effects of: physical damage to DNA molecules, the presence of chemicals that inhibit the PCR process [9, 10] and the fact that smaller DNA fragments are more efficiently amplified than larger ones during the PCR process [8] have complicated efforts to develop objective standards of identifying degradation.

Objectively Identifying Potential Degradation/Inhibition

The downward slope characterizing degraded or inhibited single-source samples can be measured. First, a best-fit linear regression through the heights of the peaks is drawn [11]. The data collection point is the point on the x-axis and the height of the peak is the point on the y-axis (homozygotes are divided in half). A slope is then calculated for the angle of the regression line and compared to a population of non-degraded samples (a collection of positive controls whose slopes have been separately calculated). If a given sample falls outside of the population of non-degraded samples at a given level of significance (e.g., $\alpha = 0.05$), then it can be flagged as being potentially degraded or inhibited. Such samples should be evaluated to determine the potential for allelic dropout, reduced assumptions for mixture components, and other issues.

Mini-STRs

A new set of primers has been developed to address the issue of degradation [12]. The distance between the primer sequence and the STR repeats has been shortened so that the amplified DNA product is significantly smaller. Smaller amplicons are less likely to be affected by the effects of degradation and inhibition. Applied Biosystems produces the MiniFiler™ test kit, which examines eight of the STR loci (D13S317, D7S820, D2S1338, D21S11, D16S539, D18S51, CSF1PO, and FGA) and the sex-determining locus Amelogenin [13]. All of the MiniFiler™ amplicon sizes are less than 300 bp. In some instances, the new primer sets have resulted in revealing primer binding site mutations resulting in peak height imbalances or null alleles. In other instances, the reported Mini-STR allele differs from a standard STR test due to a 4 bp deletion in the region between the standard STR primer sequence and the beginning of the STR repeat region. In those cases, the reported allele from a standard STR kit would be one less than what would be reported by a Mini-STR kit.

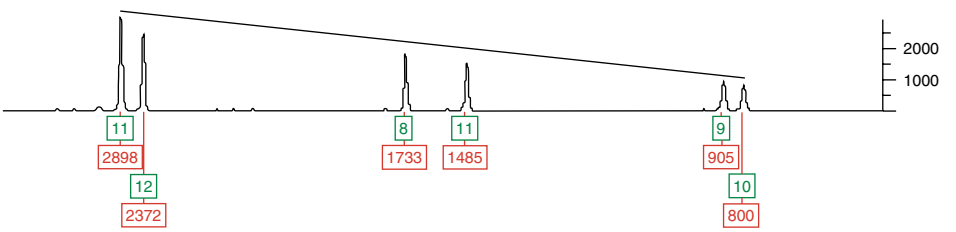


Figure 1 An electropherogram exhibiting signs of degradation or inhibition. Degradation and inhibition are marked by observing progressively falling peak heights as the size of the DNA product increases (left to right)

Y-STR Testing

Y-STR testing identifies the male contribution to a sample by targeting the Y chromosome (*see* **Y-Chromosome Short Tandem Repeats**). Y-STR testing also has the feature of producing small amplicons for a subset of the tested loci. For example, 6 of the 17 loci examined by the Applied Biosystems Yfiler® test kit produces product sizes less than 200 bp (DYS456, DYS389I, DYS458, DYS393, DYS391, and YGATAH4) [14]. Therefore, Y-STR testing may be a reasonable alternative when working with degraded samples (at the cost of less probative results).

Low Copy Number Testing

Samples containing low amounts of DNA often exhibit signs of degradation. However, it is not recommended that Mini-STRs or Y-STRs be used to alleviate the issues observed in the analysis of low copy number (LCN) DNA samples. Both the Yfiler® and MiniFiler™ test kits recommend a minimum starting template of 0.5 ng. Using less than the recommended amount of input DNA can still produce results at or below the stochastic threshold (*see* **Low Copy Number DNA**). The issues present with standard LCN DNA test results may still be observed, including allelic dropout, allelic drop-in, exaggerated stutter, and exaggerated peak height imbalance.

Addressing Inhibition

In some instances, it may be possible to distinguish between the effects of degradation and inhibition in that further purification of template DNA followed by reamplification can remedy problems associated with the presence of inhibitors. It is also possible that new extraction and amplification strategies will minimize the trends observed in peak heights for degraded/inhibited samples analyzed with currently popular methods [15].

Electropherograms can also appear degraded if the amplification mix is created improperly. Adding too little Taq polymerase can result in a profile that appears to be degraded. Adding too much EDTA in TE can sequester the magnesium ions, turning off the Taq, and again creating a profile that appears to be degraded or inhibited [8]. Further, adding too much template DNA will cause the PCR reagents to compete for template strands. The shorter fragments will be preferentially amplified, again resulting in a profile that appears to be degraded or inhibited. All of these issues can be identified by comparing the samples with the controls and references run in the same amplification. If the other samples appear to be normal, then it is likely the amplification setup is correct. In addition, reamplification can rectify the appearance of degradation or inhibition.

The Coroner's Inquest into the Death of Jaidyn Leskie

The Victoria Police Forensic Services Centre in Australia used the Profiler Plus® test kit to generate STR DNA profiles from two evidentiary samples associated with the deceased [16]. The DNA profiles that were detected were subsequently found to be consistent with the DNA profile of a rape victim associated with a distinctly separate investigation for which DNA testing was performed by the same laboratory less than one week earlier. Evidence at the inquest suggested that the rape victim could not have been involved in the death of Jaidyn Leskie. The testing laboratory suggested that the correspondence between DNA profiles of at least 7 (and as many as 12, after additional testing and review) STR loci associated with the evidence samples in the two cases may be a result of an "adventitious" (coincidental) match rather than due to contamination between the two analyses.

The small quantities of template available for PCR amplification from the investigation samples associated with the deceased, coupled with apparent degradation/inhibition, resulted in several peaks associated with the largest amplification products falling below 200 Relative Fluorescent Units (RFUs). One issue raised during the course of the Coroner's inquest was whether the most likely source of contamination from the rape investigation (an unmixed sample of the complainant on a condom) also qualified as being degraded/inhibited.

Comparison of the trends in peak height *versus* data collection point for the condom sample (Figure 2) were found to be significantly different than those of the sampling of 164 ostensibly nondegraded/inhibited positive control samples associated with this study or the positive control associated with the rape investigation. The condom falls below the threshold for variance in the sampling of positive

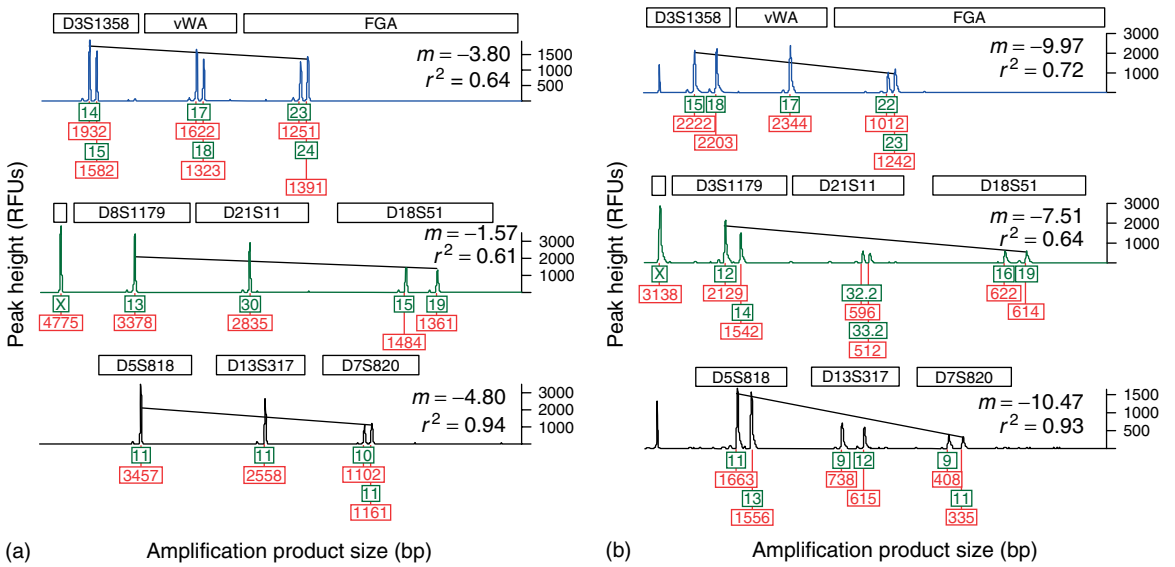


Figure 2 Electropherograms associated with a high-quality genomic DNA template (a) and with the condom sample from the rape investigation (b) that qualifies as being inconsistent with the sampling of positive controls

controls at the $\alpha = 0.01$ level. Thus, there is less than a 1% chance that a sample consistent with the sampling of positive controls (and thus presumably undegraded) would, by chance, exhibit the significant difference noted in the condom.

The Coroner ruled that the DNA testing result in the Leskie investigation was caused by contamination:

> "The match to the bib occurred as a result of contamination in the laboratory and was not an adventitious match. The samples from the two cases were examined by the same scientist within a close time frame" (G. Johnstone, Inquest into the death of Jaidyn Raymond Leskie. Coroner's Case Number: 007/98. July 31, 2006) [16].

Conclusions

Degradation and inhibition are common issues encountered during the course of forensic DNA casework review. Objective measures can be used to identify when a sample may be subject to degradation or inhibition based on linear regression slopes using the electropherogram. Inhibitors can be removed with an added purification step during the amplification process. New technologies like mini-STRs and some Y-STR profiling kits have the ability to resolve some DNA profiles that have been compromised by degradation. No matter how a degraded profile is resolved, the analyst must consider a variety of alternative interpretations (and have those additional alternatives be reflected in a lessening of the statistical weight associated with an evidence sample).

References

[1] Handt, O., Hoss, M., Krings, M. & Paabo, S. (1994). Ancient DNA: methodological challenges, *Experientia* **50**(6), 524–529.

[2] von Wurmb-Schwark, N., Harbeck, M., Wiesbrock, U., Schroeder, I., Ritz-Timme, S. & Oehmichen, M. (2003). Extraction and amplification of nuclear and mitochondrial DNA from ancient and artificially aged bones, *Legal Medicine (Tokyo)* **5**(1), S169–S172.

[3] Poinar Jr., G.O. (1994). The range of life in amber: significance and implications in DNA studies, *Experientia* **50**(6), 536–542.

[4] Kevles, D.J. (2003). Ownership and identity, *The Scientist* **17**(1), 22.

[5] Gill, P., Ivanov, P.L., Kimpton, C., Piercy, R., Benson, N., Tully, G., Evett, I., Hagelberg, E. & Sullivan, K. (1994). Identification of the remains of the Romanov family by DNA analysis, *Nature Genetics* **6**, 130–135.

[6] Ivanov, P.L., Wadhams, M.J., Roby, R.K., Holland, M.M., Weedn, V.W. & Parsons, T.J. (1996). Mitochondrial DNA sequence heteroplasmy in the Grand Duke of

Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II, *Nature Genetics.* **12**(4), 417–420.
[7] Adams, D.E., Presley, L.A., Baumstark, A.L., Hensley, K.W., Hill, A.L., Anoe, K.S., Campbell, P.A., McLaughlin, C.M., Budowle, B., Giusti, A.M., Smerick, J.B. & Baechtel, F.S. (1991). Deoxyribonucleic acid (DNA) analysis by restriction fragment length polymorphisms of blood and other body fluid stains subjected to contamination and environmental insults, *Journal of Forensic Sciences* **36**(5), 1284–1298.
[8] Walsh, P.S., Erlich, H.A. & Higuchi, R. (1992). Preferential PCR amplification of alleles: mechanisms and solutions, *PCR Methods and Applications* **1**, 241–250.
[9] DeFranchis, R., Cross, N.C.P., Foulkes, N.S. & Cox, T.M. (1988). A potent inhibitor of Taq polymerase copurifies with human genomic DNA, *Nucleic Acids Research* **16**(21), 10355.
[10] Akane, A., Matsubara, K., Nakamura, H., Takahashi, S. & Kimura, K. (1994). Identification of the heme compound copurified with deoxyribonucleic acid (DNA) from bloodstains, *a major inhibitor of polymerase chain reaction (PCR) amplification, Journal of Forensic Sciences* **39**(2), 362–372.
[11] Draper, N.R. & Smith, H. Applied Regression Analysis, *Wiley Series in Probability and Statistics* (1998).
[12] Coble, M.D. & Butler, J.M. (2005). Characterization of new MiniSTR loci to aid analysis of degraded DNA, *Journal of Forensic Sciences* **50**(1), 43–53.
[13] Hill, C.R., Kline, M.C., Mulero, J.J., Lagace, R.E., Chang, C.W., Hennessy, L.K & Butler, J.M. (2007). Concordance study between the AmpFlSTR MiniFiler PCR Amplification Kit and conventional STR typing kits, *Journal of Forensic Sciences* **52**(4), 870–873.
[14] Mulero, J.J., Chang, C.W., Calandro, L.M., Green, R.L., Li, Y., Johnson, C.L. & Hennessy, L.K. (2006). Development and validation of the AmpFlSTR® Yfiler™ PCR amplification kit: A male specific single amplification 17 Y-STR multiplex system, *Journal of Forensic Sciences* **51**(1), 64–75.
[15] Schmerer, W.M., Hummel, S. & Herrmann, B. (1999). Optimized DNA extraction to improve reproducibility of short tandem repeat genotyping with highly degraded DNA as target, *Electrophoresis* **20**(8), 1712–1716.
[16] Johnstone, G. (2006). Inquest into the death of Jaidyn Raymond Leskie. Coroner's Case Number: 007/98.

JASON R. GILDER

DNA: Electrophoretogram ***see*** Allelic Designation

DNA: Hardy-Weinberg Equilibrium ***see*** Hardy-Weinberg Equilibrium

DNA: Identifiler ***see*** Identifiler

DNA: Low Amounts ***see*** Whole Genome Amplification

DNA: Low Template ***see*** Low Copy Number DNA

DNA: Matrix ***see*** Matrix: DNA

DNA: Mixture Interpretation ***see*** Mixture Interpretation: DNA

DNA: Peak Height ***see*** Peak Height: DNA

DNA: Phenotype ***see*** Phenotype

DNA: QIAAMP *see* QiaAmp

DNA: Sources of

Introduction

In 1987, the first report appeared in the literature describing the use of DNA profiling to assist in the investigation of a crime [1]. In this now-famous case, DNA profiling using minisatellite technology [2, 3] (*see* **Microsatellites**) was used for the first time to demonstrate that semen found on the bodies of two young women was likely to be from the same man and that the young man suspected by police was excluded as the source. Further investigations, including the "blooding" of several English villages, led police to identify Colin Pitchfork, who subsequently pleaded guilty to both crimes [4].

Since then, DNA profiling has evolved through a number of stages, each one resulting in increased sensitivity or discrimination. Now, large scale DNA databases to date, where large-scale DNA databases are in use in many countries, short tandem repeat (STR, *see* **Short Tandem Repeats**) multiplexes incorporating many loci and targeting autosomal and Y chromosomes are in common use and DNA profiling is recognized as a key part of crime solution and modern policing [5].

So where does the DNA come from that produces such startling results? The answer is, almost everywhere. Today, it is well recognized that DNA profiles can be obtained from samples that are too small to be seen and sometimes difficult or impossible to attribute to a particular cell type.

The samples include the well-recognized bloodstains, semen and saliva stains, and hairs as well as less common samples such as nasal and vaginal secretions, teeth and bone, skin, including dandruff, feces, chemically treated fingerprints, plant material, and insects (for examples, see [6, 7]).

In this article, by reference to case reports and articles published over the last 20 years and by drawing on this author's personal experiences and those of others as practicing forensic scientists, attempts have been made to convey the huge variety of sources of DNA evidence that can be encountered, recognizing the infinite variety of samples, substrates, and case circumstances that exist.

Blood

Blood, splashed and smeared, is probably the most popular image conveyed of a crime scene. Blood can also be present on garments worn by individuals and on weapons such as knives and baseball bats used to inflict injury. Of course, blood can also be present as a result of innocent transfer and may have nothing to do with the alleged crime committed. The examination of the patterns of bloodstaining (*see* **Bloodstain Pattern Interpretation**) together with the analysis of DNA determining the likely source of blood can provide powerful forensic evidence and is often used to reconstruct events.

Of course, bloodstaining may not be visible. Various reports in the literature have investigated the effects of the luminol test (*see* **Luminol**) on the ability to obtain DNA profiles [8, 9]. Luminol is used to visualize possible bloodstaining that has arisen from very dilute blood, typically associated with attempts to clean up a crime scene or as a result of a trail of blood left by, for example, the shoes of a participant in a crime. Application of the luminol reagent in combination with appropriate techniques to extract and amplify the trace amounts of DNA has not been found to be detrimental to the success of the DNA analysis procedure.

The Vaginal Tract

Menstrual blood contains components that distinguish it from circulatory blood although these are currently difficult to quantify. A typical menstrual bleed contains, in addition to circulatory blood components, broken-down tissue from the endometrium and plasmin, a compound that prevents blood clotting. Blood associated with assault of a sexual nature may therefore contain blood derived from a menstrual bleed and/or circulatory blood caused by injury to the vagina and surrounding tissue. Menstrual blood is mixed with varying proportions of vaginal secretions.

Vaginal secretions are always located on genital swabs collected from the vagina and also on other objects used to penetrate the vaginal tract. These secretions contain miscellaneous proteins originating from the upper genital tract such as the

oviduct, endometrium, and cervical mucus [10]. While difficult to identify specifically, the cellular material present provides a rich source of nuclear DNA. Thus, collection of such secretions by swabbing is often a successful source of DNA.

Semen

DNA analysis of semen formed the basis of the first reported use of DNA analysis in forensic science [1] and the use of DNA in the solution of crimes of sexual assault has had a profound effect on the effectiveness of forensic evidence in the resolution of crime. Seminal fluid or semen is a mixture of cells, spermatozoa, and a variety of organic and inorganic substances, and is a commonly encountered sample type in forensic casework.

Internal genital swabs are routinely collected from rape victims in most parts of the world. These swabs may be stained with semen, saliva, and blood as well as vaginal secretions. The DNA recovered from such swabs collected properly and stored dry or frozen is usually in good condition and the success rate from this type of sample is very high. Semen stains on clothing/bedding are also common.

Semen comprises seminal fluid with or without the presence of spermatozoa. Samples containing spermatozoa are rich in DNA and DNA analysis of such samples, where sperm are visible, is nearly always successful using polymerase chain reaction (PCR) techniques. A differential lysis treatment is commonly used on samples of this type (*see* **Extraction**) in order to separate the female epithelial material from the spermatozoa, thus simplifying the interpretation of the resultant DNA profiles.

The advent of Y-STR profiling techniques (*see* **Y-Chromosome Short Tandem Repeats**) has led to an increase in the successful analysis of semen in the absence of spermatozoa [11]. In such circumstances, DNA is recovered and selectively profiled from the small amounts of epithelial material present in the seminal fluid even in the presence of large amounts of female DNA. Of course, Y-STR profiling is also useful when spermatozoa are present in small amounts.

Semen can be present in the mouth following oral intercourse, and is sometimes found mixed with saliva in stains on clothing. Once again, differential lysis separates the spermatozoa from the epithelial cells present enabling amplification of both spermatozoa and epithelial fractions separately and analysis of the DNA profiles.

DNA can also be recovered from cellular material on condoms although the success rates for this type of sample can be lower because of the presence of inhibitory substances in the spermicides and lubricants or because of the moist environment on the condom surfaces.

Case study 1: "Condom–the Critical Link in a Rape" [12] In 1993, a report appeared in the literature of the successful analysis using variable number tandem repeat (VNTR) (see **Variable Number Tandem Repeats***) analysis of material collected from a condom. The circumstances of the case were typical. A young girl was sexually assaulted by a man using a condom. The condom, discarded by the assailant was recovered and found to be stained with blood on one side and semen on the other. DNA analysis showed that the blood could have come from the young girl supporting the medical evidence and evidence of the girl. The DNA profile obtained from the semen was consistent with having come from the suspect.*

A crucial step in the DNA profiling of evidence from cases of sexual assault is the identification of semen. This can be accomplished by the microscopic identification of spermatozoa if present, which is done by histological staining of the spermatozoa. The sensitive PCR systems used in laboratories today enable the DNA profiling of the spermatozoa on such slides and this can be enhanced using techniques such as laser microdissection [13] to facilitate the recovery of the sample. Histological slide collections made during the routine examination of sexual assault cases, over many years, provide an excellent repository of biological evidence for the investigation of old "historic" cases and many laboratories around the world are now utilizing such evidence in combination with their DNA databases to solve previously unsolved cases.

An alternative source of semen for analysis are the test papers produced during the presumptive test for semen, commonly known as the *acid phosphatase (AP) test* [14]. If stored in an appropriate condition,

these papers can provide sufficient biological material to enable the development of DNA profiles.

It is our experience [15] that DNA profiles can still be obtained from semen stains on garments after laundering using a range of normal washing conditions. Thus, so long as the approximate location of the semen stain can be approximated (such as the crotch of a pair of underpants), then it is possible to obtain a DNA profile presumably from spermatozoa adhering to the fibers of the material.

Trace DNA

A report published in Nature in 1997 [16] demonstrated that DNA could be recovered from items held or touched for a time as brief as few seconds. This work stemmed from the observation that DNA could be successfully recovered from the blade of a tomahawk that had been submerged for 12 h and was seemingly unstained and clean. This demonstrated that DNA profiling was possible from samples such as keys, pens, and handles, and it was suggested that the success of obtaining a DNA profile may be related to the individual's propensity to "shed" DNA and not necessarily the length of time that the person was in contact with the item. The propensity of an individual to "shed" DNA has been examined [17, 18] and factors influencing the shedding of DNA by individuals investigated. It has been found that there is significant variation in the amount of DNA a person sheds at any one time and thus whether a person is a good or bad "shedder" at any particular time is currently unpredictable.

The work described above and the experiences of others [19] led to the realization that DNA recovered from car steering wheels, computer keyboards, and other surfaces not necessarily stained with recognizable body fluids, could be an important source of DNA evidence.

One drawback with samples collected from these sources is the prevalence of mixed profiles from multiple contributors that may not be resolved into individual sources. Another is the observation that the last person to handle the item may not be the person from whom the DNA profile originated. Hence, there may be a temporal aspect to the deposition of DNA that must be considered when interpreting the DNA profiling results.

Case study 2: "DNA Profiling from Heroin Street Packages" [20] The authors of this study suspected that it may be possible to identify the individuals responsible for sealing the edges of the plastic packets used to contain small doses of heroin for sale. The proposition was that epithelial cells from these individuals might become trapped in the burnt edges of the packets and that this might provide useful evidence for investigators. Experimental work and analysis of case material showed that the method of sealing was crucial to the successful generation of a DNA profile and that individuals who handled the packages before and after sealing might also leave traces of DNA that could be amplified. Thus, although DNA profiles can be obtained from the sealed edges, and therefore from the sealer, they can also be obtained from other individuals who may have handled the packages.

Pipe bombs are easy to make and commonly used in America and elsewhere. When exploded, they generate a large amount of heat. During experimental trials, DNA was could be recovered and analyzed from such pipes after the explosion had taken place [21]. Critical factors of success were the ability to locate sufficient pieces of pipe and the degree of shedding of the person assembling and handling the pipe. This report serves to illustrate that what might be expected to be an extreme environmental insult may, in fact, not be as detrimental to the chance of success as expected.

Of course, legitimate transfer of DNA and/or cellular material from people to such items must be considered and never has Locard's principle "every contact leaves a trace" been truer. Its just that now, every contact leaves a trace of DNA that, if it can be collected either by visual observation, good guess work, or sometimes luck, can be compared to a reference DNA profile from an individual.

Low Copy Number and Single Cells

The successful amplification of DNA from a single cell using the second generation multiplex (SGM) multiplex and elevated cycle number [22], heralding

the beginnings of low copy number (LCN) techniques (*see* **Low Copy Number DNA**) [23]. In this first report, complete DNA profiles were obtained from 50% of the single source samples analyzed using elevated PCR numbers.

Subsequent to this report, there are many examples of DNA profiles being obtained from very small numbers of cells, including the use of laser microdissection to capture small numbers of sperm cells from histochemically stained microscope slides [13].

Case study 3: "A Tale of Two Robberies" [24] Two robberies were committed in different towns in Northern Italy. In one case, traces of saliva were detected on a pillowcase used to transport the stolen goods. In the second case, a pair of sunglasses was left behind in a bank during a robbery and cellular material was recovered from the nose and ear pieces of the frames. DNA profiles were obtained using LCN techniques and the profiles entered onto the local database. Immediately, the two DNA profiles linked together and additionally, the DNA profile of an individual on the database corresponded to the crime sample profiles.

Saliva

Saliva is a secretion of the mouth that is important in digestion and comprises cells and secretions from the salivary and parotid glands. Saliva has a high proportion of water and a low level of dissolved substances and cellular material, which can make it difficult to locate visually. Saliva is commonly encountered as a source of DNA evidence.

Perhaps the first example that springs to mind is the ubiquitous cigarette butt. Cigarette butts are often left by suspects as they contemplate committing a crime, waiting for an opportunity, or possibly as a means to steady nerves. It is not uncommon to find DNA from multiple contributors on cigarettes, a result of sharing, and indeed on occasion three or four people may have shared the same cigarette [25].

Case study 4: "A New Zealand Family Affair" A car was stolen and found several days later, several hundred kilometers from home. Inside were a number of cigarette butts that should not have been there. From these butts, six different DNA profiles were obtained, five of which corresponded to five members of a "family" on a DNA database. The sixth was consistent with being from a family member not present on the database.

Perhaps a word of caution. It is not unheard of for police officers to submit numerous cigarette butts from public areas, for example, outside banks or commercial premises, in the hope of uncovering possible offenders. Such untargeted sampling approaches can lead to innocent persons being investigated and laboratories being overwhelmed with unnecessary work. As in all things, a healthy dose of common sense is helpful.

Case study 5: "A Body in the Water" [26] Saliva swabbed from human skin has been used to generate DNA profiles, enabling association between a bite mark and a suspect even when the body had been submerged in water for several hours. In this particular case, the deceased had suffered extensive head injuries and sexual assault as well as a bite on her right breast. Cellular material was collected by swabbing and a DNA profile obtained. The profile obtained was a mixed DNA profile containing DNA types that could be attributed to the victim as well as the suspect. The DNA evidence was produced at trial and proved crucial to the resolution of the trial.

Saliva is also prevalent among samples left at scenes of more minor crimes, for example, burglary, where it is not uncommon for offenders to help themselves from the cupboards and refrigerators of their victims. In such cases, saliva can be located on bottles and cans [27] on bitten pieces of chocolate, biscuits, and fruit, and on cutlery and the rims of glasses and cups [6].

Saliva can also be recovered from the back of licked stamps and envelopes [28]. In this case, care must be taken as a stamp can yield not only evidence of who might have licked the gum, but also evidence of DNA from skin cells transferred by touching the outer, exposed surface of the stamp. The scientist should be careful to distinguish between these two possibilities. Saliva trapped between stamp and envelope is however protected from contamination and reliable results should be obtained free from extraneous DNA. LCN (*see* **Low Copy Number DNA**) techniques have aided in the effectiveness of this type of evidence as the amount of cellular material transferred to a stamp or envelope is likely to be low.

Fingernails

There are two main sources of DNA associated with fingernails. The first is the nails themselves [29]. The keratin structure of the fingernail does not contain DNA itself; therefore, the analysis of fingernails themselves is limited to recovering DNA from epithelial cells adhering to the nail. This analysis may be carried out in an attempt to determine the person from whom a nail piece came or as an alternative reference sample.

Alternatively, biological material scrapped from underneath the nail can be sampled and its source identified. The material trapped under the nails could be a variety of different body fluids or tissues such as semen, blood, saliva, skin, and vaginal secretions and could have been deposited in any number of ways. For example, a victim may scratch an offender during a violent struggle, or an offender may have digitally penetrated the victim during a sexual assault.

There are a number of reports of the successful analysis of such debris under the nails [30] including from under the nails of bodies submerged in water for periods of time [31]. Indeed, it has been shown that even repeated handwashing is not always sufficient to remove the evidence [30]. In order to interpret the results of such an analysis, it is important to know the background levels of DNA expected to be found underneath the fingernails in normal situations, one such study reports that the incidence of foreign DNA under fingernails appears to be low [32].

Case study 6: "The Persistence of DNA under Fingernails Following Submersion in Water" [31] Case one. A victim of homicide was found after submersion in bath water for 2–3 h. Mixed DNA profiles corresponding to the deceased and one other were obtained from the clipped nails. The male component of this mixture was found to correspond to the suspect, identified by a database search. Case two. A woman was seen falling into the harbor with her baby and her body was recovered 3 h later. DNA profiling was attempted on the fingernails clipped at the postmortem and the major male component was found to correspond to those of a colleague with whom an altercation had previously taken place. This information corroborated the accounts of several witnesses and no charges were laid.

Fingerprints

It is commonly understood that it is difficult but not impossible to obtain DNA profiles from fingerprints after enhancement [33]. In one comprehensive study, fingermarks were made on a variety of surfaces in blood and saliva. The fingermarks were treated to a wide range of chemical enhancement methods. In most cases, STR profiles were obtained using standard DNA profiling methods, showing that the many enhancement methods used were not significantly detrimental to the DNA profiling process. This mirrors our own experience where the difficulty arises when the fingermark is made by the greasy secretions of the skin only, in which case, elevated PCR cycle numbers (*see* **Low Copy Number DNA**) may be needed because of the small amount of available biological material.

Hair

The average human head contains between 100 000 and 150 000 hair follicles [34] with hair roots present in three stages of development known as *anogen, catagen*, and *telogen* (*see* **Hair: Microscopic Analysis**). On an average, 50–100 hairs are shed everyday, naturally or as a result of force and most of these are in the telogen phase. In the anogen phase,

the hair roots are actively growing and are rich in DNA. In the catagen, or breakdown phase, nuclear DNA is still relatively abundant. Telogen hairs, however, have little or no nuclear DNA associated with them [35] and it is telogen hairs that are most often encountered in forensic casework. Unless there is additional cellular material adhering to the root structure of a telogen hair, analysis is usually unsuccessful [36], using autosomal DNA typing methods. Alternative DNA profiling methods such mitochondrial DNA typing methods [37] or LCN techniques [38] have proved to be useful for the analysis of both telogen hair roots and hair shafts, which are also known to contain very little nuclear DNA. Indeed, LCN techniques have been reported to be successful in the production of nuclear DNA profiles from hairs recovered from Siberian mummies 500 years old [38].

Teeth and Bone

The use of teeth and bone as a source of DNA for forensic work is of primary importance in the identification of human remains. Among the most challenging requirements in forensic science, in recent years, has been the requirement to identify the victims of mass disasters, commonly known as *Disaster Victim Identification* (*see* **Disaster Victim Identification**).

Where possible human remains are identified using dental records and fingerprints, this is not always possible. Early attempts of extracting DNA from bones from decaying and/or compromised bone samples were often unsuccessful due to the degraded nature of the DNA that was recovered and the VNTR DNA typing methodology employed. However, nuclear DNA can be recovered from these cells in sufficient amounts for DNA analysis by PCR [39].

With the advent of PCR techniques, it was possible to amplify degraded DNA and samples of bone and teeth that had been exposed to severe trauma became more useful sample types. The bodies for which identification using DNA is typically required may be from recently deceased individuals as a result of explosive or fire trauma such as following the Waco incident [40] or Spitsbergen aircraft crashes [41], or the bodies and body parts might be "old" such as those recovered from grave sites where notable historic figures such as the Romanov family [42] and Jesse James [43] were thought to be buried. Less newsworthy, but nonetheless important, are missing person cases where body identification is required. [44, 45].

Success has been shown to depend less on the age of the bone and more on the conditions under which it has been kept and amplification of fragments of mitochondrial DNA from ancient bone thousands of years old has been reported [46].

Recent use of short amplicons (*see* **Mini-STRs**) and single nucleotide polymorphism (SNP) technology have increased the utility of these samples and provided increased chances of successful profiles.

An advantage of using bones and teeth for forensic purposes is that stringent washing/cleaning techniques can be employed to minimize the possibility of contamination with extraneous DNA and this is of particular importance when considering the analysis of historic or ancient specimens.

Teeth, in particular, provide a rich source of nuclear DNA within the coronal pulp and radicular canals of the tooth [47], although these sources decrease with age and disease as additional dentin is deposited on the internal surfaces.

The external structures of teeth (dentin and enamel) provide an excellent physical barrier and great protection to the mtDNA and nuclear DNA found within teeth, thus conferring teeth with a great advantage and making them a reliable source of DNA for forensic purposes in situations where other samples such as bone may prove unsuccessful.

Formalin-Fixed Paraffin-Embedded Tissue Samples

Formaldehyde-fixed and paraffin-embedded tissues provide a valuable resource for forensic analysis [48, 49]. Such tissues, stored by hospital departments, can provide direct reference material for late comparison with human remains. In addition, forensic examination can be employed to determine whether or not tissue samples showing signs of disease have been wrongly attributed to patients following allegations of mislabeling or similar.

Fecal Material

Fecal material in the form of stools or stained tissue paper is found in a small but not insignificant number of associated with burglary scenes,

in particular. Fecal material contains, among other things, solid indigestible remains of consumed food, large amounts of microbial material, and other compounds known to inhibit the PCR reaction. Early attempts to recover DNA from fecal samples relied upon the use of mtDNA analysis; however, the introduction of DNA extraction methods capable of substantially removing PCR inhibitors [50, 51] has enabled nuclear DNA-based PCR methods to be utilized improving the evidentiary usefulness of this sample type.

Unusual Sources

Lice, Flies, Mosquitoes, and Maggots

Insects, particularly flies, are frequently associated with crime scenes, and the colonization of bodies by insects after death is used to determine the postmortem interval [52] (*see* **Entomology**). It has been shown that it is possible to recover sufficient human DNA from the crop contents of maggots recovered from corpses to obtain DNA profiles, using not only mtDNA techniques but also conventional autosomal STR typing methods [53]. Results were obtained from maggots recovered from corpses showing putrefaction and from maggots of several different fly families. Thus results were obtained from blow flies, flesh flies, and house flies.

Case study 7: "Body on the Beach" [55] The body of a woman was found on a beach. Inquiries led investigators to a suspect, and, upon examination of his home, a mosquito and a small bloodstain corresponding to a blood meal were found on an inside wall. The STR DNA profile obtained corresponded to that of the victim indicating that she had been in the vicinity of the address. When taken into account with other scientific evidence, a strong case was made and the suspect was convicted.

In a similar investigation, mtDNA profiling results were could be obtained from individual human crab lice that corresponded to their human source. In other words, successful mtDNA analysis has been reported of the blood meals of the human louse, thus providing an alternative avenue of investigation when such lice are found on clothing [54].

The Contact Lens

The vacuum cleaner bag has long been regarded as a source of possible fiber evidence, but due to the sensitivity of current DNA profiling techniques, trace amounts of biological material vacuumed away either deliberately or by chance can now be recovered from this most unexpected source.

Case study 8: "The Contact Lens" [56] Following a report of a sexual assault committed three days earlier, investigators searched the contents of a dust bag from a vacuum cleaner and retrieved numerous fragments of a contact lens, allegedly knocked from the complainant's eye. A DNA profile, from a single source, and corresponding to that of the complainant was obtained, confirming her presence at the address and the loss of a contact lens as described by her. Faced with this evidence, a guilty plea was entered and a new avenue for potential DNA evidence was uncovered.

Plant Material as Forensic Evidence

In recent years, the suitability of plant material as forensic evidence has begun to be recognized (*see* **Botany**; **Wood**). A number of publications have appeared in the literature describing its application.

Cannabis

A number of alternative methods have been described for the analysis of *Cannabis sativa* L. [57] including DNA sequencing of specific regions of the chloroplast DNA [58] genome suggested as a method to confirm the presence of cannabis. Other studies [59] have utilized STR typing to investigate genetic variability of cannabis plant material. While their findings were unable to fully distinguish between drug and fiber accessions, the STR markers were highly polymorphic and showed significant differentiation between individual plants that had not been clonally produced.

Grasses

Recently, PCR-based methods to distinguish between grass species have been reported. By investigating the mitochondrial sequences of 20 phylogenetically representative grass species, researchers were able to identify sufficient informative sequences (indels) in the grass mitochondrial genome to enable correct identification of 25 unknown grass samples. Thus, grass collected from items associated with crime scenes promises potential as an investigative tool [60].

Magic Mushrooms

Possession of hallucinogenic fungi is a criminal offense in many countries; however, not all closely related fungi contain the hallucinogenic ingredients and, therefore, unambiguous determination of a particular species can be of value in a criminal investigation. PCR-based DNA methods have been described [61] that in combination with phylogenetic approaches, specimens can be separated into hallucinogenic or nonhallucinogenic species.

Trees

A number of genetic approaches can be used to analyze leaf and other plant tissue from trees and this can be used to provide associative evidence in criminal casework.

Case study 9: "Disassociation with a Burial Site" [62] A pregnant woman's body was found buried within a few meters of a group of three sand live oak trees, Quercus geminata. Dried leaf material was found in the boot of a suspect's car. STR markers were developed that enabled a comparison between the dried leaf material from the car and the trees at the burial site. In this case, the DNA evidence showed that although the leaves in the suspect vehicle appeared to have come from three different trees, the results were not consistent with them having come from the three trees at the burial site.

Other groups have used different approaches to analysis. The amplified fragment length polymorphism technique has been used to investigate the potential of association of plant material from the red maple, *Acer rubrum*. [63] The results showed that within a closed set of trees, identification was possible but that interpretative difficulties arose when considering a larger group or geographical area.

Summary

Since the first reports of the use of DNA profiling to solve a forensic case, DNA profiling has evolved through a number of stages, each one resulting in increased sensitivity or discrimination, to date, where large-scale DNA databases are in use in many countries, STR multiplexes incorporating many loci, and targeting autosomal and Y chromosomes are in common use, and DNA profiling is recognized as a key part of crime solution and modern policing. The seemingly limitless nature of the sources of DNA for forensic analysis continues to impress and as forensic scientists cast their net wider into the plant and animal kingdoms, the possibilities for the future seem endless.

References

[1] Gill, P. & Werrett, D.J. (1987). The exclusion of a man charged with murder by DNA fingerprinting, *Forensic Science International* **35**, 145–148.

[2] Jeffreys, A.J., Wilson, V. & Thein, S.L. (1985). Individual-specific "fingerprints" of human DNA, *Nature* **316**, 76–79.

[3] Gill, P., Jeffreys, A.J. & Werrett, D.J. (1985). Forensic application of DNA fingerprints, *Nature* **318**, 577–579.

[4] Wambaugh, J. (1989). *The Blooding. The true story of the Narborough Village Murders*, A Perigord Press Book, New York.

[5] Butler, J.M. (2005). *Forensic DNA Typing: Biology, Technology and Genetics of STR Markers*, 2nd Edition, Elsevier Academic Press, USA.

[6] Kuperus, W.R., Hummel, K.H., Roney, J.M., Szakacs, N.A., Macmillan, C.E. & Wickenheiser, R.A., Hepworth, D., Hrycak, T.L., Fenske, B.A. & De Gouffe, M.J. (2003). Crime scene links through DNA evidence: the practical experience from saskatchewan casework, *Canadian Society of Forensic Science Journal* **36**, 19–28.

[7] Kratzer, A., Brun, C., Haas, C., Voegeli, P. & Bar, W. (2004). Evaluation of stain cases of the swiss DNA database, *International Congress Series* **1261**, 479–481.

[8] Gross, A.M., Harris, K.A. & Kaldun, G.L. (1999). The effect of luminol on presumptive tests and DNA analysis using the polymerase chain reaction, *Journal of Forensic Sciences* **44**, 837–840.

[9] Quinones, I., Sheppard, D., Harbison, S.A. & Elliot, D. (2006). Comparative analysis of luminol formulations, *Canadian Society for Forensic Science Journal* **40**, 53–63.

[10] Itoh, Y. & Manaka, M. (1988). Analysis of human vaginal secretions by SDS-polyacrylamide gel electrophoresis, *Forensic Science International* **37**, 237–242.

[11] Sibille, I., Duverneuil, C., de la Grandmaison, G.L., Guerrouache, K., Teissiere, F., Durigon, M. & de Mazancourt, P. (2002). Y STR DNA amplification as biological evidence in sexually assaulted female victims with no cytological detection of Spermatozoa, *Forensic Science International* **125**, 212–216.

[12] Brauner, P. & Gallili, N. (1993). A Condom- the critical link in a rape, *Journal of Forensic Sciences* **38**, 1233–1236.

[13] Elliot, K., Hill, D.S., Lambert, C., Burroughes, T.R. & Gill, P. (2003). Use of laser micro-dissection greatly improves the recovery of DNA from sperm on microscope slides, *Forensic Science International* **137**, 28–36.

[14] Pizzamiglio, M., Mameli, A., Maugeri, G. & Garofano, L. (2004). Identifying the culprit from LCN DNA obtained from saliva and sweat traces linked to two different robberies and the use of a database, *International Congress Series* **1261**, 443–445.

[15] Reshef, A., Barash, M., Gallili, N., Michael, A. & Brauner, P. (2005). The use of acid phosphatase test papers for DNA profiling, *Science and Justice* **45**, 97–102.

[16] Crowe, G., Moss, D. & Elliot, D. (2000). The effect of laundering on the detection of acid phosphatase and spermatozoa on cotton t shirts, *The Canadian Society of Forensic Science Journal* **33**, 1–5.

[17] Van Oorschot, R.A.H. & Jones, M.K. (1997). DNA fingerprints from fingerprints, *Nature* **387**, **767**.

[18] Lowe, A., Murray, C., Whitaker, J., Tully, G. & Gill, P. (2002). The propensity of individuals to deposit DNA and secondary transfer of low levels of DNA from individuals to inert surfaces, *Forensic Science International* **129**, 25–34.

[19] Phipps, M. & Petricevic, S.F. (2007). The tendency of individuals to transfer DNA to handled items, *Forensic Science International* **168**, 162–168.

[20] Wickenheiser, R.A. (2002). Trace DNA: A review, discussion of theory and application of the transfer of trace quantities of DNA through skin contact, *Journal of Forensic Sciences* **47**, 442–450.

[21] Zamir, A., Cohen, Y. & Azoury, M. (2007). DNA profiling from heroin street dose packages, *Journal of Forensic Sciences* **52**, 389–392.

[22] Esslinger, K.J., Siegel, J.A., Spillane, H. & Stallworth, S. (2004). Using STR analysis to detect human DNA from exploded pipe bomb devices, *Journal of Forensic Sciences* **49**, 1–4.

[23] Findlay, I., Taylor, A., Quirke, P., Farzier, R. & Urquhart, A. (1997). DNA fingerprinting from single cells, *Nature* **389**, 555–556.

[24] Whitaker, J.P., Cotton, E.A. & Gill, P. (2001). A comparison of the characteristics of profiles produced with the AMPFLSTR® SGM Plus™ multiplex system for both standard and low copy number (LCN) STR DNA analysis, *Forensic Science International* **123**, 215–223.

[25] Hochmeister, M.N., Budowle, B., Jung, J., Borer, U.V., Comey, C.T. & Dirnhofer, R. (1991). PCR based typing of DNA extracted from cigarette butts, *International Journal of Legal Medicine* **104**, 229–233.

[26] Sweet, D. & Shutler, G.G. (1999). Analysis of salivary DNA evidence from a bite mark on a body submerged in water, *Journal of Forensic Sciences* **44**, 1069–1072.

[27] Abaz, J., Walsh, S.J., Curran, J.M., Moss, D.S., Cullen, J., Bright, J., Crowe, G.A., Cockerton, S.L. & Power, T.E.B. (2002). Comparison of the variables affecting the recovery of DNA from common drinking containers, *Forensic Science International* **126**, 233–240.

[28] Fridez, F. & Coquoz, R. (1996). PCR DNA typing of stamps: evaluation of the DNA extraction, *Forensic Science International* **78**, 103–110.

[29] Kaneshige, T., Takagi, K., Nakamura, S., Hirasawa, T., Sada, M. & Uchida, K. (1992). Genetic analysis using fingernail DNA, *Nucleic Acids Research* **20**, 5489–5490.

[30] Lederer, T., Betz, P. & Seidl, S. (2001). DNA analysis of fingernail debris using different multiplex systems: a case report, *International Journal of Legal Medicine* **114**, 263–266.

[31] Harbison, S.A., Vintiner, S.K. & Petricevic, S.F. (2003). The persistence of DNA under fingernails following submersion in water, *International Congress Series* **1239**, 809–813.

[32] Cook, O. & Dixon, L. (2007). The prevalence of mixed DNA profiles in fingernail samples taken from individuals in the general population, *Forensic Science International Genetics* **1**, 62–68.

[33] Grubweiser, P., Thaler, A., Kochl, S., Teissl, R., Rabl, W. & Parson, W. (2003). Systematic study on STR profiling on blood and saliva traces after visualisation of fingerprint marks, *Journal of Forensic Sciences* **48**, 733–741.

[34] Olsen, E.S. (ed) (1994). *Disorders of Hair Growth, Diagnosis and Treatment*. McGraw-Hill, Inc, New York.

[35] Linch, C.A., Smith, S.L. & Prahlow, J.A. (1998). Evaluation of the human hair root for DNA tying subsequent to microscopic comparison, *Journal of Forensic Sciences* **43**, 305–314.

[36] Higuchi, R.G., von Beroldingen, C.H., Sensabaugh, G.F. & Ehrlich, H.A. (1988). DNA typing from single hairs, *Nature* **332**, 543–546.

[37] Wilson, M.R., DiZinno, J.A., Polensky, D., Replogie, J. & Budowle, B. (1995). Validation of mitochondrial DNA sequencing for forensic casework analysis, *International Journal of Legal Medicine* **108**, 68–74.

[38] Amory, S., Keyser, C., Crubezy, E. & Ludes, B. (2007). STR typing of ancient DNA extracted from hair shafts from siberian mummies, *Forensic Science International* **166**, 218–229.

[39] Hochmeister, M.N., Budowle, B., Borer, U., Eggmann, U., Comey, C. & Dirnhofer, M.D. (1991). Typing of deoxyribonucleic acid (DNA) extracted from compact bone from human remains, *Journal of Forensic Sciences* **36**, 1649–1661.

[40] Clayton, T.M., Whitaker, J.P., Fisher, D.L., Lee, D.A., Holland, M.M., Weedn, V.W., Maguire, C.N., DiZinno, C.P., Kimpton, C.P. & Gill, P. (1995). Further validation of a quadruplex STR DNA typing system: a collaborative effort to identify victims of a mass disaster, *Forensic Science International* **76**, 17–25.

[41] Olaisen, B., Stenersen, M. & Nevag, B. (1997). Identification by DNA analysis of the victims of the August 1996 Spitsbergen civil aircraft disaster, *Nature Genetics* **15**, 402–405.

[42] Gill, P., Ivanov, P.L., Kimpton, C., Peircy, R., Benson, N., Tully, G., Evett, I., Hagelberg, E. & Sullivan, K. (1994). Identification of the remains of the Romanov family by DNA analysis, *Nature Genetics* **6**, 130–135.

[43] Stone, A.C., Starrs, J.E. & Stoneking, M. (2001). Mitochondrial DNA analysis of the presumptive remains of Jesse James, *Journal of Forensic Sciences* **46**, 173–176.

[44] Hagelberg, E., Gray, I.C. & Jeffreys, A.J. (1991). Identification of the skeletal remains of a murder victim by DNA analysis, *Nature* **352**, 427–429.

[45] Crainic, K., Paraire, F., Leterreux, M., Durigon, M. & de Mazancourt, P. (2002). Skeletal remains presumed submerged in water for three years identified using PCR STR analysis, *Journal of Forensic Sciences* **47**, 1025–1027.

[46] Hagelberg, E., Sykes, B. & Hedges, R. (1989). Ancient bone DNA amplified, *Nature* **342**, 485.

[47] Smith, B.C., Fisher, D.L., Weedn, V.W., Warnosk, G.R. & Holland, M.M. (1993). A systematic approach to the sampling of dental DNA, *Journal of Forensic Sciences* **38**, 1194–1209.

[48] Shibata, D.K., Arnheim, N. & Martin, W.J. (1988). Detection of human papilloma virus in paraffin- embedded issue using the polymerase chain reaction, *Journal of Experimental Medicine* **167**, 225–230.

[49] Gill, P., Kimpton, C.P. & Sullivan, K. (1992). A rapid polymerase chain reaction method for identifying fixed specimens, *Electrophoresis* **13**, 173–175.

[50] Johnson, D.J., Martin, L.R. & Roberts, K.A. (2005). STR-typing of human DNA from human fecal matter using the qiagen qiaamp® stool mini kit, *Journal of Forensic Sciences* **50**, 802–808.

[51] Roy, R. (2003). Analysis of human fecal material for autosomal and y chromosome STRs, *Journal of Forensic Sciences* **48**, 1035–1040.

[52] Catts, E.P. & Goff, M.L. (1992). Forensic entomology in criminal investigations, *Annual Review of Entomology* **37**, 253–272.

[53] Zehner, R., Amendt, J. & Krettek, R. (2004). STR typing of human DNA from fly larvae fed on decomposing bodies, *Journal of Forensic Sciences* **49**, 337–340.

[54] Lord, W.D., DiZinno, J.A., Wilson, M.R., Budowle, B., Taplin, D. & Meinking, T.L. (1998). Isolation, amplification and sequencing of human mitochondrial DNA obtained form human crab louse, Pthirus Pubis (l.), blood meals, *Journal of Forensic Sciences* **43**, 1097–1100.

[55] Spitaleri, S., Romano, C., Di Luise, E., Ginestra, E. & Saravo, L. (2006). Genotyping of human DNA recovered from mosquitos found on a crime scene, *International Congress Series* **1288**, 574–576.

[56] Wickenheiser, R.A. & Jobin, R.M. (1999). Comparison of DNA recovered from a contact lens using PCR DNA typing, *The Canadian Society of Forensic Science Journal* **2/3**, 67–73.

[57] Gigliano, G.S. & Cannabis Sativa, L. (2001). Botanical problems and molecular approaches in forensic investigations, *Forensic Science Review* **13**, 1–17.

[58] Linacre, A. & Thorpe, J. (1998). Detection and identification of cannabis DNA, *Forensic Science International* **91**, 71–76.

[59] Gilmore, S., Peakall, R. & Robertson, J. (2003). Short tandem repeat (STR) DNA markers are hypervariable and informative in Cannabis sativa: implications for forensic investigations, *Forensic Science International* **131**, 65–74.

[60] Ward, J., Peakall, R., Gilmore, S.R. & Robertson, J. (2005). A molecular identification system for grasses: a novel technology for forensic botany, *Forensic Science International* **192**, 121–131.

[61] Nugent, K.G. & Saville, B.J. (2004). Forensic analysis of hallucinogenic fungi: a DNA based approach, *Forensic Science International* **140**, 147–157.

[62] Craft, K.J., Owens, J.D. & Ashley, M.V. (2007). Application of plant DNA markers in forensic botany: Genetic comparison of *Quercus* evidence leaves to crime scene trees using micro-satellites, *Forensic Science International* **165**, 64–70.

[63] Bless, C., Palmeter, H. & Wallace, M.M. (2006). Identification of *Acer rubrum* using amplified fragment length polymorphism, *Journal of Forensic Sciences* **51**, 31–38.

SALLY-ANN HARBISON

DNA: STR *see* Short Tandem Repeats

DNA: Y-STR *see* Y-Chromosome Short Tandem Repeats

DNA Databases and Evidentiary Issues

Introduction

DNA profiling and the compilation and use of forensic DNA databases are most often discussed with reference to the operational impact, the effect on intelligence-led models for forensic and policing practice, and the broader justice system ramifications such as the catalysis of legislative change, the increased socio-legal debate, and emergent aspects of jurisprudence. In this entry, the focus is refined to cover specific evidentiary issues associated with the use of DNA databases. The specific issues raised by the use of familial searches will not be treated here (see [1, 2] for the statistical issues and [3, 4] for some ethical considerations). The approach of familial speculative searches consists in searching a full profile left on a crime scene in the DNA database. No full match is returned, but a partial correspondence (on relatives sharing more of their DNA than unrelated persons, this partial match could indicate that the crime stain was left by a close relative of the person with whom the partial match was found) is obtained.

The collation into databases of DNA profiles taken from persons convicted of, or otherwise connected to, crimes has been a focal point of development efforts in the field of forensic biology for at least the past decade. The scale of database growth and use is now considerable, particularly in populous Western democracies. Databases in the United Kingdom, the United States, and Europe alone now house over 10 million profiles, and their use has resulted in almost 1 million intelligence links (Table 1).

Along with the increase in the volume of cases brought about by the introduction of DNA databases, there has also been an alteration to the types of crimes and evidence submitted for biological analysis. Initially, DNA profiling was applied primarily to serious crimes, whereas, in recent times, it has contributed to the investigation of a far broader spectrum of crimes, most notably extending to include vast numbers of property crimes [5]. An important aside to the broadened role that DNA evidence now plays in the criminal justice system, thanks largely to the introduction of DNA databases, has been the need for forensic scientists and administrators to operate in what is a more public environment with arguably greater levels of awareness and accountability.

While the investigative changes and rewards have been substantial, progress in the area of DNA databasing has also brought some measure of controversy. Much of the controversy is not associated with the investigative use of databases but rather with the areas where they intersect with evidentiary or other predominantly legal issues. For example, the enactment of specific legislation to allow broad-scale collection and storage of DNA samples from offenders, arrestees, and, in some cases, citizens otherwise associated with crimes has naturally brought comment on associated legal and ethical concerns [6]. There have been specific issues to consider in the courts also. Although the legislative regimes have assisted in easing troublesome issues in areas such as informed consent, they have brought challenges on interpretation that range from procedural issues to constitutionality. Approaches to evidence interpretation are also required to adapt to this changed environment so as to provide a framework to assist in the communication of outcomes across the interface of the forensic and legal sectors, most often in the form of expert evidence in criminal proceedings. As with other parts of the process, there have been specific issues in forensic statistics that have emerged through the use of DNA databases. The critical issues include (i) a refocusing on the likelihood of adventitious matches between profiles and robustness of interpretation models and (ii) controversy around

Table 1 Summary of the size and effectiveness of major DNA database programs

Database	Combined	United Kingdom	Europe	United States
Date	–	Jan 2007	Jan 2007	July 2007
Total profiles	10 498 199	4 103 509	1 444 916	4 949 774
Offender profiles	9 729 432	3 790 551	1 173 451	4 765 430
Crime profiles	768 767	312 958	271 465	184 344
Investigations aided	969 428	762 280	154 282	52 866

the appropriate approach for reporting a DNA match statistic in cases where the suspect was identified by a database search.

Adventitious Matches

As the discriminating power of DNA multiplexes has increased, the reasonable question that has arisen is, *when can a DNA profile match be considered proof that two DNA samples have come from the same source?* The FBI announced a policy on this in November 1997 [7] and others have supported the appeal of such a position [8]. While not declaring "uniqueness", the FBI policy encourages the use of the term "source attribution", which is intended to describe the situation where "to a high degree of scientific certainty" the source of the stain can be assigned to a particular individual [9]. This policy suggested that a match probability of less than $p = 3.9 \times 10^{-11}$ will confer a 99% confidence that the evidentiary profile is "unique" in a population of equivalent size to the United States (or $N =$ 260 million). While the simplicity is appealing, this move was controversial and has been criticized for its subjectivity [10] and the fact that the estimates were based on assumptions of independence at the population level [11]. We have previously stated our philosophical concern about scientists taking decisions such as these out of the hands of the courts [12].

While there is some effort in the advocacy of this approach to stress that source attribution applies only within the context defined by the case, and does not extend, for example, to the entire human population [13], there remain many risks with this approach. The semantics of the distinction provided describe an intent that is similar to categorization, that is, the only acceptable proposition out of a defined set of alternatives, or, one of a kind. However, the effect of the approach aligns more with individualization, itself implying the one – the only. This misapprehension is obvious in the reflections on the policy from areas of the legal community. The American Prosecutors' Research Institute announced the policy by saying; "... Jurors may be spared technical and often confusing explanations of population genetics and statistics because DNA examiners will be able to testify to a matching DNA profile without qualifying the match with complicated mathematical probabilities. Furthermore, now that the FBI policy has paved the way for definitive identifications of an individual as the source of an evidentiary DNA sample, other laboratories conducting DNA analysis may soon implement similar policies" [14]. In *Young v State of Maryland*, the appellate court concluded (at 56–57) "when the random match probability (RMP) is sufficiently miniscule, the DNA profile may be deemed unique. In such circumstances, testimony of a match is admissible without accompanying statistics. In place of the statistics, the expert may inform the jury of the meaning of the match by identifying the person whose profile matched the profile of the DNA evidence as the source of that evidence" [15]. It is incongruent to speak of uniqueness in the context of the handful of people associated with a case, and not anticipate misunderstanding.

While it is the vanishingly small random match probabilities (RMP) that are derived from direct matches between 10 and 15 locus short tandem repeat (STR) profiles that have encouraged approaches such as the assignment of source attribution, the collation of large-scale DNA databases has refocused this issue somewhat. If a database of individuals is of size N, with C entries on the crime stain database, this implies that there are $N(N-1)/2$ pairs of individual profiles, $C(C-1)/2$ pairs of crime profiles (involved in the crime-to-crime matching process), and NC person-to-crime comparisons during the matching of the two databases to each other. These numbers are not always on peoples' minds, and it is interesting to see how quickly they become large. With their increasing size comes the increasing likelihood of adventitious database matches [16].[a] In Table 2, we have estimated the number of matches expected in databases of various sizes (N) for a profile with an RMP of 1×10^{-9} or 3.9×10^{-11} (the FBI source attribution threshold).

For the purpose of this illustration there is no population substructure in the theoretical database, nor has there been any consideration given to the presence of relatives. The difference between this simple predictive model and an actual database is that a proportion of crime profiles are partial or derived from mixtures, implying that they do not have all loci scored. Such samples will have a higher RMP than full profiles and would be expected to match more frequently with other samples. Of course, there will be related people on the database and, for these people, the match probability will be larger than the match probability between unrelated people. These

Table 2 Crude demonstration of expected number of adventitious matches in databases of varying sizes

DB size	No. of comparisons	Expected number of matches	
		$RMP = 1 \times 10^{-9}$	$RMP = 3.9 \times 10^{-11}$
1	0	0	0
10	45	0	0
100	4950	0	0
1000	499 500	0	0
10 000	49 995 000	0	0
100 000	4 999 950 000	5	0
1 000 000	499 999 500 000	500	19
4 765430[(a)]	11 354 659 159 735	11 355	443
10 000 000	49 999 995 000 000	50 000	1950

[(a)] Size of combined offender DNA index system (CODIS) offender database (July 2007)

related people will increase the number of person-to-person matches but these are not considered as adventitious hits. The extent of relatedness in "offender" databases could be high. Likewise, there will be crime samples originating from people who are related to persons on the database. This increases the number of adventitious matches. A substructure is also present at the population level.

Accommodating these factors in predictive statistical models is a difficult task, largely because of the structure of databases and the fact that the match probabilities between two people, or between people and crime samples, are not constant. This matter has been tackled previously by [17] and is also addressed by Triggs and Buckleton [16]. Even without such models, it is possible to get the flavor of anticipated results (as illustrated in Table 2) and to understand clearly that adventitious matches between people on the database are expected to occur. Likewise, adventitious matches between crime samples and people are also expected. This is more likely in circumstances involving partial profiles on the database or from crime samples.

The likely event of adventitious matches seems counterintuitive for those unaccustomed to the mathematics, and we imagine it could confuse or alarm the general public. Perhaps for this reason, the issue has recently emerged in court as part of evidence questioning the validity of DNA match statistics. On occasions, these issues have been raised on the basis of observed results from databases of forensic DNA profiles. For example, reports exist in the literature relating to parentage- [18] and kinship [19, 20]-testing experiences that describe observed levels of matching or allele sharing that appear high in the context of associated DNA match statistics. There have also been anecdotal reports of coincidental matches at nine or more STR loci. All of these observations have involved comparisons with a database of some kind. One of the more noted recent examples was initiated following the observation of a single incident of an adventitious match within the Arizona State DNA Identification System (SDIS) [21]. A full search of the Arizona DNA offender database identified 144 partial matches; 122 nine-locus, 20 ten-locus, 1 eleven-locus, and 1 twelve-locus match (both of the latter two involved full siblings). At the time of the search, the database contained 65 493 profiles with an average RMP estimated to be less than 4.5×10^{-10} [22]. This finding caused alarm and led to legal requests for both State and Federal databases to be handed over in their entirety to independent researchers to undertake a full evaluation of the extent of allele and profile sharing [23]. There are many problems with using databases in this manner, mostly relating to what has to be done when a match is observed. Matching profiles may be the same person sampled twice under the same, similar, or different names, or they may originate from twins, close relatives, or unrelated people. Investigating this requires a considerable multiagency effort. The only advisable scientific safeguard is confirmation at additional STR or single nucleotide polymorphism (SNP) loci.

Some of the alarm surrounding the Arizona matches undoubtedly arose from misinterpretation of the data. As mentioned earlier, what is often overlooked in these circumstances is, first, that the critical information is the total number of comparisons

involved, and, secondly, how the observed number of matches compares with that expected from the statistical models underpinning estimates of the RMP. This comparison has been made by [24] who gives the expected figures for 9- and 10-locus matches respectively as 100 and 3, using a model that assumes no relatedness in the database. While the observed number of matches exceeds the number predicted by this simplistic model, one can see that the number of observations is not as alarmingly high as one may have intuitively thought them to be.

One of the limitations of forensic statistical approaches, and one for which they are often criticized, is the inability to empirically validate the estimates produced, and thereby, the models that underpin them. This is compounded by the fact that the numbers themselves are so large (or small) that they raise justifiable concerns about the level of accuracy. As the examples above show, large DNA databases are a pool of empirical data that, despite the limitations mentioned earlier, can provide some realistic basis by which theoretical estimates could be evaluated.

Weir [25] first addressed this issue using combined Australian profile data (~15 000 nine-locus profiles) and focussing not only on the fully matching profiles but also on those that partially match (for example, a profile where eight loci fully match and one allele or no alleles of the last locus match). The breakthrough made by Weir was to observe that the partially matching profiles do not suffer from the disadvantage that they may be the same person. Curran *et al.* [26] have recently extended this approach to include corrections in expected estimates for the common relationships, siblings, cousins, and parent/child, as well as unrelated persons. A variety of large datasets were utilized comprising diverse population groups (such as Caucasian, New Zealand Maori, and Australian Aborigine). The comparison of observed and expected numbers of partially matching profiles shows that, when subpopulation effects and relatedness are accounted for, the existing statistical models give an excellent fit to empirical observations.

From this aspect of databases, some interesting observations have emerged. The assembly of large databases of DNA profiles has brought the issue of adventitious sharing of DNA into sharp focus, which in turn raises questions about policies such as source attribution. Associated with DNA, there remains a significant issue around the expectation of the uninitiated. Most of the general public see DNA as a unique identifier and are befuddled when adventitious matches occur. These misappropriated expectations are easily exploitable by anyone wishing to denigrate the basis of the DNA match statistics, as, at first blush, the facts appear to contradict the model. Ironically, where the DNA databases themselves have generated this confusion, they also provide a resource for resolving long-held ambiguity about the validity of statistical estimates. In fact, the scientific analysis of these data that has so far been undertaken [25, 26] provides perhaps the strongest direct evidence yet as to the robustness of forensic DNA interpretation models.

Estimation of DNA Match Statistics after a Database Search

The strength of the DNA evidence resulting from an intelligence database match is often presented without any mention that the hit was obtained from a database. It is usually not in the suspect's interest to let a jury know that he or she has a sample on the DNA database. The question of whether searching a database for a match affects the strength of the evidence has been discussed extensively and forcefully in the literature. Unfortunately, there is much confused writing, and it would be very difficult for a court to make a reasoned decision based on a simple assessment of literature recommendations.

The issue was first countenanced in the first National Research Council (NRC, [27]) report, which suggested that the loci used in the matching process should not be used in assessing the weight of evidence (page 124):

> The distinction between finding a match between an evidence sample and a suspect sample and finding a match between an evidence sample and one of many entries in a DNA profile databank is important. The chance of finding a match in the second case is considerably higher ... The initial match should be used as probable cause to obtain a blood sample from the suspect, but only the statistical frequency associated with the additional loci should be presented at trial.

The second NRC report (NRC II, [28]; Recommendation 5.1, pp. 133–135) recommended that an adjustment be applied by multiplying the RMP by the number of people on the database. Using an example of an intelligence database of 100 000 and a multiplex

with an RMP of 1×10^{-6} would result in an adjusted database match probability of 0.1 being reported to the court. This approach is often referred to as the *Np* rule.[b] The conservativeness of this recommendation becomes extreme as the size of the database increases. For example, if the database were to hold the profiles of the entire world, and there was only one matching individual, logically this would provide an irrefutable association, yet, should *Np* be employed, the discriminating power of the match would be reduced by a factor of 6 billion.

In addition to the two NRC reports, [29–32] have also suggested that the match probability be multiplied by the number of people on the database. Lempert [33] suggests multiplying by the size of the suspect population and not the database. Morton [34] suggests confirmatory markers or the use of the *Np* approach. The DNA Advisory Board [13] highlights the question as being *especially important*, and favors the approach of [30], which it describes as being *more comprehensible* and *always conservative*.

The general consensus among noted authors in the field is that the evidence is slightly stronger after a database search and hence no correction is required [16, 35–47].

For a considerable portion of its history, the use of forensic DNA evidence in court fought to gain general acceptance. Having achieved a stage where its general suitability was less contentious, DNA evidence has at times struggled due to the difficulty the courts endured in interpreting its technicality. The present debate that is taking place around whether an adjusted (Np) or an unadjusted match statistic (p) should be presented impinges on both these fronts. Differences of opinion run the risk of being perceived as a lack of general acceptance across the expert community, and the technicalities of the issue could easily overwhelm lay court participants.

In the recent trial, [48] DNA evidence was excluded at trial under [49] due to the *raging debate* in the scientific literature on the issue of database match statistics. This was seen to indicate a lack of *general acceptance* in the scientific community and thereby preclude admissibility. This ruling was overturned on appeal. The appeal court ruling goes to some lengths to describe the issue. It cites three key approaches that could be used to estimate DNA statistics in the circumstance of a DNA database search that has resulted in a suspect's identification and subsequent trial. These were termed the *rarity statistic*,[c] which estimates how rare a profile is in a given society, the *database match probability*, which estimates the probability of a chance match from a particular database (calculated by *Np*), and the *Balding and Donnelly approach*, which gives a full consideration of the evidence in context (based on [38]). The Appeal court held as follows:

> there is no controversy ... as to the accuracy of the various formulas ... each approach accurately answers the question it seeks to address ... Instead, the arguments raised by each of the proponents simply state that their formulation is most probative, not more correct ... Thus the debate is one of relevancy, not methodology ... and as such there is no basis to exclude the evidence in this case ... What is and is not relevant is not appropriately decided by scientists and statisticians ... determining what is and is not relevant is a hallmark responsibility of the trial judge and that responsibility is not appropriately delegated to parties outside the court.

While the ruling in *Jenkins* is appreciative of the issues, it does not greatly assist as it gives equal legitimacy to all approaches. In fact, the debate occurring among scientists *is* on this quintessentially legal issue, as it hinges not upon the mathematics *per se* but on which statistic or statistics will help the jury appreciate the fact of a DNA match. To this end, another US appeal court ruling was more direct [50]. In dismissing the fact that the appellate was first identified as a possible suspect based on a database search as something that *simply does not matter* (at p. 15), the court also stated that "Jenkins does not convince us that the probability of obtaining a cold hit from a search of the CODIS database matters." (at p. 20).

The relevant question that must be assessed is, "Is the defendant the source of the DNA recovered from the crime scene?" For the purpose of an example, let us say that the questioned sample is a semen stain located on intimate swabs from a complainant in a sexual assault. The jury may embrace the proposition advanced by the prosecution (referred to as Hp) that the defendant is the source of the DNA on the swabs. Alternatively, it may embrace another proposition, possibly advanced by defense counsel (referred to as Hd). In general terms, this is that the defendant is not the source of the DNA on the swabs, but rather it has originated from an unrelated person who just happens to share the same profile as the defendant.

Evidence that informs the jury as to the plausibility of this alternative proposition is therefore relevant.

If the defendant is selected for reasons that have nothing to do with his DNA profile, and no one else is tested, the chance of the match under the coincidence hypothesis is the RMP, p. However, if the defendant was selected following a database search (i.e., he was not the only person considered), this requires additional analysis. The frequentist argument is that, because a match following a search is much less surprising than a confirmation match following on from other evidence of association, Hd is much more plausible than p would suggest. Np indicates how probative the match is with respect to the hypothesis that the database contains the source of the crime scene DNA as opposed to the alternative hypothesis that the crime scene DNA came from someone not represented in the database. In other words, it estimates how often searches of database's that do not include the true source generate an adventitious hit. But, as Kaye succinctly states [51]:

> the database is not on trial. Only the defendant is. The DB match is only relevant to the extent that it affects the probability that the defendant – not the database – is the source. Thus, it appears that asking how probable a search of a particular database is to generate a false cold hit is the wrong question. The jury ought to be thinking about how probable the cold hit on the defendants profile would be when the specific defendant is, and is not, the source of the crime scene DNA. The likelihood ratio formed with these probabilities is the appropriate measure of the probative value of the DB match. This LR is more complex than the figure pertaining to the simple confirmation case. It includes not only the fact that the defendant matches but also the fact that everyone else in the database has been excluded. It can be shown that when this additional information is incorporated in the LR, the resulting number can be no greater than $1/p$. Consequently, introducing p as the indicator of probative value is not prejudicial. It understates the incriminatory power of the DNA evidence ....

We have previously summarized the mathematical argument [16] and reprise that proof here. For a population of size N, we index the suspect as person 1 and the remaining members of the population as $2 \ldots N$. The first $1 \ldots M$ of these are on the database. We call the hypothesis that person i is the source of the DNA: H_i. Because the suspect is indexed person 1, the hypothesis that the suspect is, in fact, the source of the DNA is H_1. The remaining hypotheses, $H_2, \ldots, H_N$ are those hypotheses where the true offender is one of the $N-1$ "other" people from the population. Before we examine the evidence, each person has some probability of being the offender $Pr(H_i) = \Pi_i$. The suspect is on the database and has been genotyped and we call his genotype G_s. The crime sample from the swabs has been typed and found to have the genetic profile G_c. The search of the database reveals that $G_c = G_s$ and that there are no other profiles on the database match (this latter requirement is unnecessary and a generalization is easy). We now know that the $2 \ldots M$ people on the database, other than the suspect, do not match. The remaining $M+1, \ldots, N$ members of the population have genotypes $G_{M+1} \ldots G_N$, which are unknown to us. We require the probability $\Pr(H_1|G_c, G_1 \ldots G_M)$.[d]

$$\begin{aligned}
&\Pr(H_1|G_c, G_1 \ldots G_M) \\
&= \frac{\Pr(G_c|H_1, G_1 \ldots G_M)\ \Pr(H_1|G_1 \ldots G_M)}{\sum_{i=1}^{N} \Pr(G_c|H_i, G_1 \ldots G_M)\ \Pr(H_i|G_1 \ldots G_M)} \\
&= \frac{\Pr(G_c|H_1, G_1 \ldots G_M)\ \Pr(H_1|G_1 \ldots G_M)}{\left(\begin{array}{c}\Pr(G_c|H_1, G_1 \ldots G_M)\ \Pr(H_1|G_1 \ldots G_M) \\ +\sum_{i=2}^{M} \Pr(G_c|H_i, G_1 \ldots G_M)\ \Pr(H_i|G_1 \ldots G_M) \\ +\sum_{i=M+1}^{N} \Pr(G_c|H_i, G_1 \ldots G_M)\ \Pr(H_i|G_1 \ldots G_M)\end{array}\right)}
\end{aligned} \tag{1}$$

In the equation above, we have split the denominator into the suspect, the other people on the database, and those other people in the population who are not on the database. Further, we assume $\Pr(G_c|H_1, G_1 \ldots G_M) = 1$. This assumption, in words, is that the crime sample should match the suspect if it did indeed come from him. This gives

$$\begin{aligned}
&\Pr(H_1|G_c, G_1 \ldots G_M) \\
&= \frac{\Pr(H_1|G_1 \ldots G_M)}{\left(\begin{array}{c}\Pr(H_1|G_1 \ldots G_M) \\ +\sum_{i=2}^{M} \Pr(G_c|H_i, G_1 \ldots G_M)\ \Pr(H_i|G_1 \ldots G_M) \\ +\sum_{i=M+1}^{N} \Pr(G_c|H_i, G_1 \ldots G_M)\ \Pr(H_i|G_1 \ldots G_M)\end{array}\right)}
\end{aligned} \tag{2}$$

Because we know that the $2 \ldots M$ "other" people on the database do not match $\Pr(G_c|H_i, G_1, \ldots G_M) = 0$ for $i = 2$ to M. This gives

$$\Pr(H_1|Gc, G_1 \ldots G_M) = \frac{\Pr(H_1|G_1 \ldots G_M)}{\Pr(H_1|G_1 \ldots G_M) + \sum_{i=M+1}^{N} \Pr(G_c|H_i, G_1 \ldots G_M)\ \Pr(H_i|G_1 \ldots G_M)} \tag{3}$$

Under most conditions, we assume that the nonmatching people do not change our view of the match probability. However, making this assumption and

assuming that the genotypes do not affect the prior gives

$$\Pr(H_1|G_c, G_1) = \frac{\Pr(H_1)}{\Pr(H_1) + \sum_{i=M+1}^{N} \Pr(G_c|H_i, G_1)\ \Pr(H_i)} \quad (4)$$

We compare this with the equivalent probability for the same match without a database search

$$\Pr(H_1|G_c, G_1) = \frac{\Pr(H_1)}{\Pr(H_1) + \sum_{i=2}^{M} \Pr(G_c|H_i, G_1)\ \Pr(H_i) + \sum_{i=M+1}^{M} \Pr(G_c|H_i, G_1)\ \Pr(H_i)} \quad (5)$$

Under any set of priors, equation (4) gives a larger value for the posterior probability than equation (5) since the denominator is larger in equation (4). Hence, reporting the approach leading to equation (5) is always conservative. In other words, the "no correction" approach is always conservative. Up to this point, we have not made any really contentious assumptions (although we accept Balding and Donnelly's argument that the search also informs the match probability). If we now assume that every person has the same prior chance of being the true offender (this assumption is quite contentious) and we write $\Pr(G_c|H_i, G_1)$ as p in equation (4), we obtain

$$\Pr(H_1|G_c, G_1) = \frac{1}{1 + (N - M - 1)Pm} \quad (6)$$

as compared to

$$\Pr(H_1|G_c, G_1) = \frac{1}{1 + (N - 1)Pm} \quad (7)$$

Again, equation (6) gives a larger posterior probability than equation (7). The mathematical solution leads to the conclusion that the evidence is stronger after a database search and reporting the standard match statistics is always conservative. Another simple explanation of this proof is given in [39].

In summary, there is consensus on certain issues in this area. Estimating Np and p seeks to answer different questions and that one is not intended to substitute the other. There is no support for the original NRC recommendation. The weight of scientific opinion is against the use of Np, and this is supportable by mathematical proof and logic. The issue itself has not received extensive coverage in the scientific literature since the initial dispute in the late 1990s. This is not to say, however, that the debate is exhausted. It is only in the past 12–18 months that significant legal challenges have emerged in this area – which is indicative of the slower rate of technology adoption in the legal system. Duly considering issues such as these is a matter for science, and this particular issue has been researched and critiqued at the time of DNA database emergence. But the fruits of this process are only now being absorbed into jurisprudence. In addition, this is happening in such a way that the duration, extent, and weight of the scientific debate is diminished. While this is frustrating to us, it also serves as a tacit reminder of the need for prompt, robust investigation of scientific issues, and suitable disclosure through publication. It also reinforces the folly in assuming that knowledge or practice that is the scientific norm will be perceived as such by the courts, and that in the least any assessment of the fact will occur along different lines, and often for different reasons.

Conclusion

Despite DNA databases having established themselves in a practical sense, there remain some peculiarities regarding the integration of outcomes from DNA databases into evidence. The one peculiarity that has precipitated issues in court in the United States and Australia has been the occurrence or anticipated occurrence of adventitious matches. It has been shown in this article that adventitious matches occur at a rate that we can approximately predict. Most risk accrues from partial profiles and people on the database who have close relatives offending. This risk can be best ameliorated by adding discriminating power either to the database samples or at the confirmation phase, or by using all available trace evidence in addition. However, there is also a more subliminal risk associated with adventitious matches, which occurs through a lack of awareness of the likelihood of their occurrence and the subsequent influence this can bring to bear on public confidence regarding evidence interpretation models.

A residual debate is also going on regarding the correct method to interpret a database hit. The mathematics is very difficult to explain to a lay person and would certainly be difficult to explain in court. A simple census of informed opinion would show that some commentators hold an opinion favoring each side of the correction or no correction

debate; however, most seem on the "no correction side".

End Notes

[a.] Adventitious matches refer to genuine links provided through the DNA database where an individual could not have been involved in a given case but who shares a common DNA profile with the crime scene sample purely through coincidence.

[b.] Where N_p indicates that the random match probability p is multiplied by the size of the database, N.

[c.] Referred to throughout this section as the RMP.

[d.] That is, the probability that the suspect is the source of the DNA, given the genotype of the crime profile and the remaining genotypes on the database.

References

[1] Storvik, G. & Egeland, T. (2007). The DNA database search controversy revisited: bridging the Bayesian-frequentist gap, *Biometrics* **63**(3), 922–925.

[2] United States v. Jenkins, No. F320-00 (D.C. Super. Ct., Apr. 5, 2005). rev'd, 887 A.2d 1013, 1017 (D.C. 2005).

[3] Frye v. United States, 54 AppDC 46, 293Fed 1013 (1923).

[4] People v. Johnson, 43 Cal.Rptr.3d 587 (Ct. App. 2006).

[5] Curran, J.M. (2005). An introduction to Bayesian credible intervals for sampling errors in DNA profiles, *Law, Probability and Risk.* **4**(1/2), 115–126.

[6] Faigman, D.L., Kaye D.H., Saks, M.J. & Sanders, J. (2007). *Modern Scientific Evidence: The Law and Science of Expert Testimony*, 2006–2007 edn. West Thompson, Egan.

[7] Curran, J.M. & Buckleton, J.S. (2008). Effectiveness of familial searches, *Science & Justice* **84**, 164–167 .

[8] Greely, H.T., Riordan, D.P., Garrison, N.A. & Mountain, J.L. (2006). Family ties: the use of DNA offender databases to catch offenders' kin, *Journal of Law, Medicine and Ethics* **34**(2), 248–262.

[9] Haimes, E. (2006). Social and ethical issues in the use of familial searching in forensic investigations: insights from family and kinship studies, *Journal of Law, Medicine and Ethics* **34**(2), 263–276.

[10] Walsh, S.J. (2007). Current and future trends in forensic molecular biology, in R. Rapely & D. Whitehouse, eds, *Molecular Forensics*, John Wiley & Sons, London, pp. 1–20.

[11] Nuffield Council on Bioethics (2007). The forensic use of bioinformation: ethical issues, http://www.nuffieldbioethics.org/go/ourwork/bioinformationuse/publication_441.html.

[12] Holden, C. (1997). DNA fingerprinting comes of age, *Science* **278**, 1407.

[13] Devlin, K. (2007). Scientific heat about cold hits, http://www.stanford.edu/~kdevlin/papers.html.

[14] Budowle, B., Chakraborty, R., Carmody, G. & Monson, K.L. (2000). Source attribution of a forensic DNA profile, *Forensic Science Communications* **2**(3), http://www.fbi.gov/hq/lab/fsc/current/index.htm.

[15] Inman, K. & Rudin, N. (2001). *Principles and Practice of Criminalistics: The Profession of Forensic Science*, CRC Press, Boca Raton.

[16] Weir, B.S. (2001). DNA match and profile probabilities: comment on Budowle *et al.* (2000) and Fung and Hu (2000), *Forensic Science Communications* **3**(1), http://www.fbi.gov/hq/lab/fsc/current/index.htm.

[17] Buckleton, J.S., Walsh, S.J. & Harbison, S.A. (2001). The fallacy of independence testing and the use of the product rule, *Science & Justice* **41**, 81–84.

[18] DAB (2000). Statistical and population genetics issues affecting the evaluation of the frequency of occurrence of DNA profiles calculated from pertinent population database(s), *Forensic Science Communications*, **2**(3), http://www.fbi.gov/hq/lab/fsc/current/index.htm.

[19] Hart, A. (1998). New FBI policy revolutionizes DNA court testimony, *Silent Witness* **4**(1), http://www.ndaa.org/publications/newsletters/silent_witness_volume_4_number_1_1998.html; http://www.fbi.gov/hq/lab/fsc/current/index.htm.

[20] *Young v State of Maryland*. Appellate Court (2005).

[21] Walsh, S.J. & Buckleton, J.S. (2005). DNA intelligence databases, in *Forensic DNA Evidence Interpretation*, J.S. Buckleton, C.M. Triggs & S.J. Walsh, eds, CRC Press, Boca Raton, pp. 439–469.

[22] Weir, B.S. (2004). Matching and partially-matching DNA profiles, *Journal of Forensic Sciences* **49**(5), 1009–1014.

[23] Poetsch, M., Ludcke, C., Repenning, A., Fischer, L., Malyusz, V., Simeoni, E., Lignitz, E., Oehmichen, M. & von Wurmb-Schwark, N. (2006). The problem of single parent/child paternity analysis – practical results involving 336 children and 348 unrelated men, *Forensic Science International* **159**, 98–103.

[24] Birus, I., Marcikic, M., Lauc, D., Dzijan, S. & Lauc, G. (2003). How high should paternity index be for reliable identification of war victims by DNA typing, *Croatian Medical Journal* **44**, 322–326.

[25] Gornik, I., Marcikic, M., Kubat, M., Primorac, D. & Lauc, D. (2002). The identification of war victims by reverse paternity is associated with significant risks of false inclusion, *International Journal of Legal Medicine* **116**, 255–257.

[26] Troyer, K., Gilboy, T. & Koeneman, B. (2001). A nine locus match between two apparently unrelated individuals using AmpFlSTR® Profiler Plus™ and Cofiler™, *Proceedings of 12th International Symposium on Forensic Human Identification*, Nashville, www.promega.com/geneticidproc/ussymp12proc/abstracts.htm (accessed 12 May 2007).

[27] Brenner, C.H. (2007). Arizona DNA database matches, http://dna-view.com/ArizonaMatch.htm (accessed 22 Apr 2007).
[28] Budowle, B., Planz, J.V., Chakraborty, R., Callaghan, T.F. & Eisenberg, A.J. (2006). Clarification of statistical issues related to the operation of CODIS, *Proceedings of 17th International Symposium on Human Identification*, Nashville, September 2006.
[29] Myers, S.P. (2006). Felon-to-felon STR partial profile matches in the Arizona database: don't panic! *Proceedings of California Association of Criminalists and the Forensic Science Society Spring Seminar – DNA Workshop*, Concord, 9 May 2006.
[30] Weir, B.S. (2004). Matching and partially matching profiles, *Journal of Forensic Sciences* **49**(5), 1009–1014.
[31] Curran, J.M., Walsh, S.J. & Buckleton, J. (2007). Empirical testing of estimated DNA frequencies, *Forensic Science International: Genetics* **1**(3–4), 267–272.
[32] National Research Council (1992). *Report: DNA Technology in Forensic Science*, National Academy Press, Washington, DC.
[33] National Research Council (1996). *Report: The Evaluation of Forensic DNA Evidence*, United States National Academy of Sciences, Washington, DC.
[34] Devlin, B. (1993). The evidentiary value of a DNA database search, *Biometrics* **56**, 1276.
[35] Stockmarr, A. (1999). Likelihood ratios for evaluating DNA evidence when the suspect is found through a database search, *Biometrics* **55**, 671–677.
[36] Stockmarr, A. (2000). The choice of hypotheses in the evaluation of DNA profile evidence, in *Statistical Science in the Courtroom*, J.L. Gastwirth, ed, Springer-Verlag, New York, pp. 143–159.
[37] Stockmarr, A. (2001). Author's reply, *Biometrics* **57**(3), 978–980.
[38] Lempert, R. (1997). After the DNA wars: skirmishing with NRC II, *Jurimetrics* **37**, 439–468.
[39] Morton, N.E. (1997). The forensic DNA endgame, *Jurimetrics* **37**, 477–494.
[40] Aitken, C.G.G. & Taroni, F. (2004). *Statistics and the Evaluation of Evidence for Forensic Scientists*, 2nd Edition, John Wiley & Sons, Chichester.
[41] Balding, D.J. & Donnelly, P. (1995). Inference in forensic identification, *Journal of the Royal Statistical Society of America* **158**(1), 21–53.
[42] Balding, D.J. & Donnelly, P. (1995). Evaluating DNA profile evidence when the suspect is identified through a database search, *Journal of Forensic Sciences* **41**(4), 603–607.
[43] Balding, D.J., Donnelly, P. & Nichols, R.A. (1994). Some causes for concern about DNA profiles – comment, *Statistical Science* **9**(2), 248–251.
[44] Berry, D.A. (1994). Comment of Roeder, K. (1994) DNA fingerprinting: a review of the controversy, *Statistical Science* **9**(2), 252–255.
[45] Dawid, A.P. (2001). Comment on Stockmarr's "likelihood ratios for evaluating DNA evidence when the suspect is found through a database search", *Biometrics* **57**(3), 976–980.
[46] Donnelly, P. & Friedman, R.D. (1999). DNA database searches and the legal consumption of scientific evidence, *Michigan Law Review* **97**(4), 931–984.
[47] Evett, I.W. & Weir, B.S. (1998). *Interpreting DNA Evidence*, Sinauer Associates, Sunderland.
[48] Evett, I.W., Foreman, L.A. & Weir, B.S. (2000). Letter to the Editor of Biometrics – reply to Stockmarr, *Biometrics* **56**, 1274–1277.
[49] Finkelstein, M.O. & Levin, B. (2001). *Statistics for Lawyers*, Springer-Verlag, New York.
[50] Meester, R. & Sjerps, M. (2003). The evidential value in the DNA database search controversy and the two-stain problem, *Biometrics* **59**(3), 727–732.
[51] Taroni, F., Biedermann, A., Garbolino, P. & Aitken, C.G.G. (2004). A general approach to Bayesian networks for the interpretation of evidence, *Forensic Science International* **139**(1), 5–16.

Related Articles

Databases

DNA: an Overview

SIMON J. WALSH AND JOHN S. BUCKLETON

DNA mtDNA Profiling *see* Mitochondrial DNA: Profiling

Documentation *see* Crime Scene Documentation

Documentation in Mental Health Records *see* Violence Risk Assessment for Mental Health Professionals

Documents: Authentication of

Authentication, or identification, is the legal concept that before documents and physical evidence can be admitted into evidence, preliminary proof must be presented that, at least *prima facie*, the proffered evidence is what it purports to be. The requirement of authentication exists for all forms of inanimate evidence, though it has particular application to documentary evidence because of its effect on the rule against admitting hearsay evidence in common law jurisdictions.

In jurisdictions such as the United States that retain reasonably strict rules on authentication, such foundation evidence requires a testimonial sponsor. Applicable statutes, court rules, and decisional law provide the requirements for such foundational proof.[a] In some common law jurisdictions, the authentication rule is more relaxed, especially with regard to documentary evidence. In civil law countries, a variety of approaches to authentication exist and no generally recognized authentication principle is noted.

The authentication requirement is a safeguard against the introduction of unreliable evidence. It is one that is more of concern to lawyers and judges than to experts. However, because forensic experts deal with documents and other physical evidence on a daily basis, some familiarity with the rules on authentication is helpful.

Authentication of documents requires a preliminary showing to the court that a document offered in evidence is truly what it purports to be. Proof of authenticity is not required to satisfy a "preponderance of the evidence" standard. The evidentiary threshold for authentication is fairly low. The evidence must simply be sufficient to satisfy the judge controlling the procedure that it is likely what its proponent claims it to be. Any doubt raised regarding the authenticity of a document, or indeed regarding the genuineness of any other evidence, can be submitted to the trier of fact in the form of contrary evidence, or by cross-examination of the sponsoring witness.[a]

There are essentially four different ways in which an adequate foundation for the admission of documentary evidence can be presented: (i) by having the purported maker of the document take the witness stand and who is then asked for an affirmation or admission of its genuineness; (ii) by the testimony of eyewitnesses who were present when the document was executed and can testify that it is what it purports to be; (iii) by the testimony of a nonexpert person who is familiar with the handwriting of the purported maker of the document from other transactions in which the witness and purported maker have been involved in the past; or (iv) by calling a witness, duly qualified as a forensic document examiner, who can compare the questioned writing with known exemplars of the purported author's writing.[b]

Photographs and X rays are a form of documentary evidence for which special authentication rules exist. As to photographs, a testimonial sponsor must be able to state from personal knowledge that the photograph fairly and accurately depicts what it is purported to portray. This does not require that the photographer testify. What is required is that a person with a connection to the issue can testify to its accuracy. The introduction of X rays into evidence requires a preliminary showing that (i) the apparatus was properly operating and maintained; (ii) the operator was qualified; and (iii) some proof that this particular representation is connected with the case being litigated. This is generally satisfied by a showing that, at the time the picture was printed or machine-tape rolled, some notation was made as to the human subject of the X ray.

Voices on tapes or sound recordings also require authentication.[c] Such authentication can be provided in ways similar to those used for documentary evidence: either by the maker of the recording or by testimony of a person who recognizes the voice on the recording through familiarity of long standing.

Some documents may be admitted without the type of authentication proof described above.

The "ancient document" rule permits writings that are 20 or more years old at the time they are offered, and that were located in a place of natural custody and under conditions that create no suspicion as to their authenticity, to be considered *prima facie* authentic.[d] Foundation testimony is required as to both the document's age and the location where it was discovered.

In literally every jurisdiction, official documents or certified copies of public records are self-authenticating. They can be admitted without a testimonial sponsor.[e] In some jurisdictions, evidence rules make other categories of documents self-authenticating as well.

Satisfying the rules of authentication permits documents to be introduced in a legal proceeding, but that does not mean that such documents will be admitted necessarily as substantive proof of their contents. An attempt to prove the contents of documents potentially raises "hearsay" issues that are independent of authentication requirements. Because of the impact of the rule against hearsay testimony, other concepts may also need to be considered by the forensic expert seeking to utilize documents as proof of their substantive content (*see* **Hearsay Evidence**).[f]

End Notes

[a.] Article IX of the Federal Rules of Evidence set out the requirements for authentication of things and identification of persons. The federal rules are hereinafter referred to as FRE followed by the applicable rule number.

[b.] FRE 901. FRE 901(b)(3) regarding expert witness testimony also permits the trier of fact to visually make a determination of authenticity by comparing the proffered document with authenticated standard writings of the purported author.

[c.] For example, FRE 901(b)(5) covers the identification of voices, and FRE 901(b)(6), the ways in which telephone conversations can be authenticated.

[d.] See FRE 901(b)(9) on ancient documents and data compilations.

[e.] FRE 902 provides for self-authentication of domestic and foreign public documents, certified copies of public records, official publications, newspapers and periodicals, acknowledged documents, certified domestic records of regularly conducted business or activity; genuineness of trade inscriptions and marks, and several other categories. FRE901(b)(7) also deals with public reports and reports.

[f.] FRE Article X discusses the impact and requirements of the best evidence rule and its impact on proof of the contents of writings, copies of writings, summaries of writings, recordings, and photographs. FRE 1001–1006.

Related Articles

Best Evidence Rule

ANDRE MOENSSENS

Documents: Counterfeit ***see*** Forged and Counterfeit Documents

Documents: Forged ***see*** Forged and Counterfeit Documents

Documents: Handwriting and Signatures ***see*** Handwriting and Signatures, Comparison of

Documents: Interpretation of Evidence ***see*** Interpretation: Document Evidence

Documents: Photography of ***see*** Photography: Marks, Impressions, and Documents

Documents: Signatures and Handwriting, Interpretation *see* Handwriting and Signatures, Interpretation of Comparison Results

DQα

The location and function of the human leukocyte antigen (HLA) DQA1 locus has been well studied in biological sciences and population genetics. The DQA1 locus contains a region of 242 bp within which some of the bases are polymorphic from person to person. This type of variation is known as a *sequence polymorphism*, as distinct from a length-based polymorphism that characterizes a variable number of tandem repeat (VNTR) or short tandem repeat (STR) locus. This polymorphic region was the target of the first commercially available polymerase chain reaction (PCR)-based forensic DNA profiling system, known as *HLA-DQα* [1, 2]. The HLA-DQα system amplified the entire HLA-DQA1 locus and used presynthesized typing strips with immobilized sequence specific oligonucleotide (SSO) probes to detect six distinct alleles (1.1, 1.2, 1.3, 2, 3, and 4) that combined to form 21 possible genotypes. Following the hybridization of PCR products to the typing strips, genotypes were visualized by reverse dot-blot chemiluminescence. The HLA-DQα system was sensitive, rapid, and self-contained and as a result was very widely used in forensic analysis throughout the early 1990s [3].

In a later version, the three subtypes of the DQA1 4 allele (4.1, 4.2, and 4.3) were distinguishable and allowed a total resolution of 8 alleles and 28 genotypes. The enhanced test was renamed HLA-DQA1, to better reflect the naming convention of the genetic community. Adding this extended genotyping capacity increased the discriminating potential of the test – which in all other ways functioned the same as the earlier version. The technique was expanded again to the system known as AmpliType® PM+DQA1, or polymarker. Polymarker was the earliest commercially available example of a multiplex PCR system as it coamplified the HLA-DQA1 locus and five other polymorphic loci (LDLR, GYPA, HBGG, D7S8, and GC).

The HLA-DQA1 and ploymarker loci have been widely used in population studies to support forensic evidence interpretation [4–8] and examine diversity and population interrelatedness. As the polymorphic characteristics of these loci are point mutations associated with coding regions, these markers are suitable candidates for population studies. The HLA-DQA1 system was well suited to the analysis of a wide range of forensic samples [9–10]; however, it was difficult to resolve mixed profiles as the genotyping method complicated the designation of a minor contributor and background color development.

References

[1] Comey, C.T. & Budowle, B. (1991). Validation studies on the analysis of the HLA-DQ alpha locus using the polymerase chain reaction, *Journal of Forensic Sciences* **36**, 1633–1648.

[2] Blake, E., Mihalovich, J., Higuchi, R., Walsh, P.S. & Erlich, H. (1992). Polymerase chain reaction (PCR) amplification and human leukocyte antigen (HLA)-DQalpha oligonucleotide typing on biological evidence samples: casework experience, *Journal of Forensic Sciences* **37**, 700–726.

[3] Sajantila, A. & Budowle, B. (1991). Identification of individuals with DNA testing, *Annals of Medicine* **23**(6), 637–642.

[4] Tagliabracci, A., Giorgetti, R., Agostini, A., Buscemi, L., Cingolani, M. & Ferrara, S.D. (1992). Frequency of HLA DQA1 alleles in an Italian population, *International Journal of Legal Medicine* **105**(3), 161–164.

[5] Kloosterman, A.D., Budowle, B. & Riley, E.L. (1993). Population data of the HLA DQ alpha locus in Dutch Caucasians. Comparison with other population studies, *International Journal of Legal Medicine* **105**(4), 233–238.

[6] Scholl, S., Budowle, B., Radecki, K. & Salvo, M. (1996). Navajo, Pueblo, and Sioux population data on the loci HLA-DQA1, LDLR, GYPA, HBGG, D7S8, Gc, and D1S80, *Journal of Forensic Sciences* **41**, 47–51.

[7] Walkinshaw, M., Strickland, L., Hamilton, H., Denning, K. & Gayley, T. (1996). DNA profiling in two Alaskan native populations using HLA-DQA1, PM, and D1S80 loci, *Journal of Forensic Sciences* **41**, 478–484.

[8] Wolfarth, R., Nhari, L.T., Budowle, B., Kanoyangwa, S.B. & Masuka, E. (2000). Polymarker, HLA-DQA1 and D1S80 allele data in a Zimbabwean Black sample population, *International Journal of Legal Medicine* **113**, 300–301.
[9] Fujita, Y., Kubo, S.-I., Tokunaga, I., Kitamura, O., Gotohda, T. & Ishigami, A. (2004). Influence of post-mortem changes on DNA typing (D1S80, TH01, HLA DQA1, and PM typing system): case studies for personal identification, *Legal Medicine* **6**, 143–150.
[10] Hochmeister, M.N., Budowle, B., Jung, J., Borer, U.V., Comey, C.T. & Dirnhofer, R. (1991). PCR-based typing of DNA extracted from cigarette butts, *International Journal of Legal Medicine* **104**, 229–233.

Related Articles

DNA

Polymorphism: Genetic

SIMON J. WALSH

Driving: Drugs ***see*** Drug-Impaired Driving

Driving while Impaired ***see*** Drug-Impaired Driving

Driving while Intoxicated ***see*** Alcohol: Analysis

Drowning: Diatoms ***see*** Diatoms

Drug: Alcohol ***see*** Alcohol: Analysis

Drug: Amphetamine ***see*** Amphetamine

Drug: Benzodiazepines ***see*** Benzodiazepines

Drug: Cannabis ***see*** Cannabis

Drug: Cocaine ***see*** Sweat: Toxicology

Drug: Marijuana ***see*** Cannabis

Drug: Opioids ***see*** Opioids

Drug Abuse ***see*** Substance Abuse

Drug Addiction ***see*** Addictions

Drug Analysis

Introduction

Drug abuse and trafficking of controlled substances is a global problem. The purpose of analysis of illicit drugs is to examine the seized samples for the presence of scheduled drugs for the criminal justice system.

The seized drugs may be in several forms: powders, tablets and capsules, plants or dried vegetable materials, and liquids. There are a number of steps involved in the analysis of illicit drugs. The first step is a screening (or preliminary test) using basic chemical tests to establish a likely class of substance. These include color tests or precipitation reactions as well as microscopic examination. The second step is performed using thin layer chromatography (TLC) to tentatively identify the actual substance. Finally, a confirmatory test is performed using a validated instrumental method. This third step also quantifies the amount of illicit substances present in the sample and the best method for this is mass spectrometry; however, other validated techniques can be used if mass spectrometry is not available.

Techniques used in the screening and confirmation of illicit substances are described, including cannabis, amphetamine, methamphetamine, opium alkaloids, other opiates, cocaine, lysergic acid diethylamide (LSD), phenethylamines, and indoleamines (psilocybin). The analytical methods used must be appropriate for these purposes.

The sections below describe the techniques used and provide examples for a few illicit substances.

Instrumental Techniques

Thin Layer Chromatography (TLC)

TLC is one of the most widely used techniques for the separation and identification of drugs owing to its simplicity, reliability, low cost, and selectivity of detection [1]. Many compounds can be detected by examining a plate containing a fluorescent indicator under 254 and 365 nm ultraviolet (UV) light. Spray reagents such as acidified potassium iodoplatinate (platinic chloride/potassium iodide/concentrated hydrochloric acid in water), FPN reagent (ferric chloride/perchloric acid/nitric acid), fluorescamine (fluorescamine in acetone), Dragendorff spray (modified bismuth subnitrate), Marquis reagent (mixture of formaldehyde and concentrated sulfuric acid), and Mandelin's reagent (ammonium vanadate in sulfuric acid) are used to visualize substances on the plate. The distance traveled by the compound divided by the distance traveled by the developing solvent is the Rf (retention factor).

High Performance Liquid Chromatography (HPLC)

High performance liquid chromatography (HPLC) is especially useful for compounds that are thermally labile and provides a very convenient and accurate method of detection. Various types of detectors, which are nondestructive, such as an absorbance detector (ultraviolet/visible (UV/Vis) spectrophotometric detector), a photodiode array detector, a fluorescence detector, a refractive index detector, an electrochemical detector, and a conductometric detector are used [2].

Gas Chromatography (GC)

Gas chromatography (GC), which uses a gas as the mobile phase, is appropriate to analyze a wide range of chemicals that have volatility to be in the gas or vapor phase at or below 400 °C. Analytes are traditionally identified on the basis of peak retention time; however, it is dependent on the nature of the response from the detector [3]. There are various types of detectors, such as a thermal conductivity detector (TCD), a flame ionization detector (FID), an electron capture detector (ECD), a flame photometric detector (FPD), and a flame thermoionic detector (FTD) [1]. Most commonly a mass spectrometer (MS) detector is used.

Gas Chromatography Mass Spectrometry (GC/MS)

The use of GC with MS has become the routine method for the identification of most drugs, because the specificity and sensitivity of gas chromatography mass spectrometry (GC/MS) are very high [4]. Most analytes can be separated and identified by using a capillary column GC using a suitable temperature program coupled to an MS. GC/MS is applicable in the analysis of a large range of compounds including narcotics. MS produces characteristic spectra that

can be easily used to identify unknown compounds. Electron impact (EI) spectra obtained by the impact of high energy electrons on the sample is the most common form of ionization. Libraries of spectra are available to assist in the identification process.

Ultraviolet/Visible (UV/Vis) Spectrometry

The UV/Vis spectrum is recorded in a spectrophotometer. UV/Vis spectroscopy corresponds to electronic excitations between the energy levels that correspond to the molecular orbital of the systems. It is suitable for quantitative analysis for the characterization of drugs [5].

Infrared (IR) Spectroscopy

Infrared (IR) is the study of the scattering, reflection, absorption, or transmission of IR radiation in the spectral range 800–1 000 000 nm (0.8–1000 μm) [2]. Fourier transform infrared (FTIR) spectrometers are now available in most laboratories. They are faster and can work with smaller samples compared to traditional IR instruments. The difficulty with the IR analysis of drug samples is the presence of other materials that interfere with the spectrum; however, IR analysis can give valuable information on chemicals that are not normally suitable for GC/MS analysis [6].

Nuclear Magnetic Resonance (NMR) Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is the only technique that can provide a detailed solution to the structure of small proteins and polynucleotides. It has become one of the foremost methods for molecular identification. This technique is very powerful because it combines detailed information at the atomic level with the possibility of understanding whole molecule properties [7].

Capillary Electrophoresis (CE)

Capillary electrophoresis (CE) can separate a wide variety of solutes, including compounds that are highly polar, thermally labile, and/or nonvolatile, with high efficiency and selectivity. It is a complementary technique to other separation methods such as HPLC. It is especially useful in the area of drug profiling, because it is able to give information on enantiomeric drug purity.

There are micellar electrokinetic chromatography (MECC), capillary zone electrophoresis (CZE), and cyclodextrin (CD)-modified CZE in CE techniques [8–10].

Liquid Chromatography Mass Spectrometry (LC/MS)

Liquid chromatography mass spectrometry (LC/MS) has made huge advances and become powerful in drug testing. Electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI) are the most widely used interfaces between LC and MS.

Currently, LC/MS/MS is employed to detect trace amounts of drugs because MS/MS improves signal-to-noise ratio, which provides the lower limits of detection and quantification. MS/MS has become a very powerful tool in microdetermination of drugs [4].

Examples of Analyses of Illicit Drugs from Seized Materials

Cannabis

Each country has its own penalties for trafficking and growing cannabis (marijuana). In Canada, growing more than 50 plants can lead to a sentence of 14 years. Singapore imposes death penalty on anyone convicted for trafficking in more than 500 g cannabis, though cannabis is now decriminalized in some parts of Europe.

The main psychoactive compounds in cannabis and cannabis resin are Δ-9-tetrahydrocannabinol (THC), cannabinol (CBN), and cannabidiol (CBD), of which THC is the most potent component. When samples are submitted for testing, they are first weighed before analysis. To conduct a microscopic test, a small amount of the material is softened with chloral hydrate in 10% sodium hydroxide solution [11, 12]. For chemical analysis, leaves and resin are cut or pulverized into small pieces, and about 10 mg is extracted with methanol.

Cystolithic trichomes, containing a deposit of calcium carbonate, and noncystolithic trichomes secreting resin are observed under the microscope (Figure 1) [13].

The presence of cannabinoids in the sample is indicated by two color tests: the Fast Blue B salt test and the Duquenois test. For the Fast

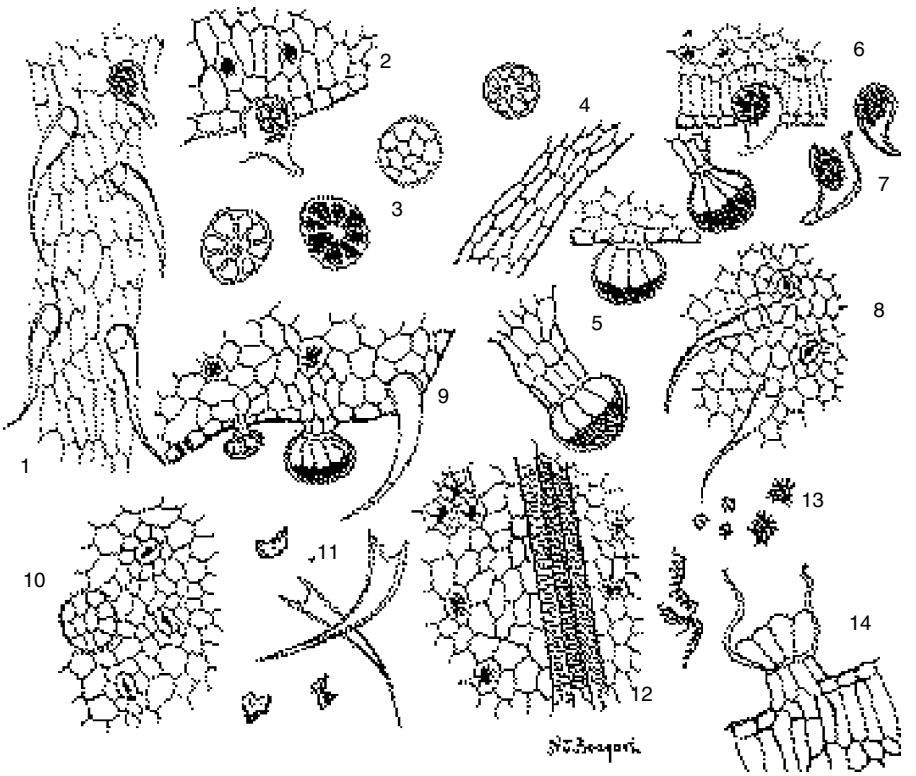


Figure 1 Cannabis powder seen through the microscope, enlarged 200 times (http://www.unodc.org/unodc/en/data-and-analysis/bulletin/bulletin_1950-01-01_4_page003.html) [13] [Reproduced with permission from UNDOC.]

(1) Fragment of bract with pointed unicellular covering hairs and hairs with cystoliths of calcium carbonate.
(2) Fragment of epidermis with a broken cystolithic hair and twin crystals of calcium oxalate.
(3) Four resin-secreting hairs seen from the front: one is still swollen with resin, two are empty, and the fourth (beside 3) still has its cuticle.
(4) Fragment of pedicel of a secretory hair.
(5) Three secretory hairs: two pedicellate, one sessile: the oleoresin is swelling and dilating the cuticle.
(6) Fragment of upper epidermis of leafstalk with cystolithic hair.
(7) Two detached cystolithic hairs: one intact, the other broken.
(8) Fragment of lower epidermis of leaf or bract with two covering hairs and two reniform stomata.
(9) Fragment of fruit bract with a covering hair and two secretory hairs in different stages of development; two twin crystals of calcium oxalate in the parenchyma; stoma on left extremity.
(10) Fragment of lower epidermis of inflorescent bract: three reniform stomata, one young sessile secretory hair with cuticle (seen from the front).
(11) Two broken unicellular covering hairs; three small pieces of solidified resin.
(12) Fragment of floral peduncle with two spiral vessels and one pitted vessel; twin crystals of calcium oxalate and a group of three sclerified cells. On the right, broken spiral vessels such as are frequently found in the preparations.
(13) Isolated and twin crystals of calcium oxalate.
(14) A pedicellate secretory hair; the cuticle is broken and has released its oleoresin

Blue B salt test, the prepared samples are shaken with chloroform. A vivid red color in lower chloroform layer is obtained. For the Duquenois test, vanillin/acetaldehyde in ethanol is added to the extract followed by concentrated hydrochloride, which results in a purple color [11, 12].

The developing solution of xylene: *n*-hexane: diethylamine (25 : 10 : 1) is commonly used for the

separation of cannabinoids. Visualization is carried out by the Fast Blue B solution, which gives a red, orange, and purple spot to THC, CBD, and CBN for each.

The quantitative analysis of cannabis samples is carried out by GC/MS and the EI spectra is compared with spectra of authentic THC. Usually, the THC content is calculated and used for drug profiling (*see* **Drug Profiling**).

Amphetamine and Methamphetamine

Penalties are different among countries for possession, trafficking, and manufacturing of amphetamines. In the United States, the law imposes a 10-year sentence for trafficking in 50 g of methamphetamine and possessing 5 g with the intention of selling. However, more severe penalties are given in some other countries: a death sentence is demanded for manufacturing more than 50 and 20 g of methamphetamine in China and Thailand, respectively, and trafficking in over 250 g in Singapore, and possessing over 50 g in Malaysia.

When samples are submitted for testing, the total amount is recorded. In the case of powders, crystals, and tablets, they are pulverized into powder and two aliquots of about 10 mg of the powder are used for analysis. One is dissolved in water for a precipitation test and the other is dissolved in methanol for TLC and instrumental analysis.

In TLC, addition of Marquis reagent (formaldehyde solution – concentrated sulfuric acid) to powder or liquid gives a light brown color followed by deep brown and finally black. Simon's reagent (solution 1, 20% aqueous sodium carbonate solution; solution 2, 50% ethanolic acetaldehyde solution; solution 3, 1% aqueous sodium nitroprusside solution) gives a deep blue color for methamphetamine and other secondary amines and pink to red color for amphetamine and other primary amines.

When 10% tetrachloroauric acid reagent is added to the prepared aqueous solution, fine needlelike crystals are formed. To check for the presence of chloride, silver nitrate reagent is added, which gives a white curdy precipitate of AgCl.

In TLC methanol/concentrated ammonia (100 : 1.5) is used as the developing solvent [12]. Amphetamine shows pink or violet color and methamphetamine shows pinkish orange color when Ninhydrin reagent is sprayed. Acidified potassium iodoplatinate reagent gives dark violet color on a pink background. Cyclohexane/toluene/diethylamine (75 : 15 : 10), methylethylketone/dimethylformamide/concentrated ammonia/isopropanol (130 : 19 : 1 : 30), and chloroform/methanol (1 : 1) can also be used as the developing solution [14].

GC/MS is considered the method of choice for confirmation. More recently, CE has become a complementary analytical tool to classical GC and HPLC [15]. Increasing scientific attention is being directed toward identification of optical isomers. The optical isomer test can be used in profiling illicit methamphetamine seizures [16]. Amphetamine and methamphetamine have one asymmetric carbon in their molecular structure and these optical isomers have different pharmacological activities. Samples submitted for analysis are in the D-, L-, or DL-forms. CE or HPLC equipped chiral column are useful for separation of optical isomers. GC/MS can be used with nonchiral column after derivatization by chiral reagents such as (*S*)-(+)-α-methoxy-(trifluoromethyl)phenylacetyl chloride ((*S*)-(+)-MTPACl), *N*-(trifluoroacetyl) prolyl chloride (TFPCl), and α-methoxy-α-(trifluoromethyl)phenylacetic acid (MTPA).

Opium Alkaloids and Other Opiates

Some countries inflict a heavy punishment for possession, trafficking, and manufacturing of heroin and morphine; for example, a death sentence for manufacturing more than 20 g heroin in Thailand, trafficking in over 15 g heroin, 30 g morphine, 1.2 kg opium in Singapore, and possessing over 15 g morphine in Malaysia.

Positive identification of opium requires confirmation of the following main alkaloids of opium: morphine (10–20%), codeine (0.2–0.8%), thebaine (0.2–1.0%), papaverine(0.5–1.0%), and noscapine (0.75–10%).

After weighing the seizures, morphological and microscopic examinations are performed. For chemical analysis, in case of poppy plants, around one-fourth of a capsule is taken and extracted with methanol. Leaves are dried and then pulverized into fine pieces and about 50 mg of the pieces is sampled. In the case of raw opium, 10 mg of the specimen is taken and extracted with methanol for the screening and confirmation test. In order to identify heroin, about 10 mg of the sample is taken and dissolved in

methanol for the screening and confirmation test. Several extraction solvents can also be used: a mixture of methanol and ammonia water; methanol; diluted hydrochloric acid; diluted sulfuric acid, etc. The pH of the extraction solution is adjusted to 9.2 with 10% ammonium water [12].

Raw opium may be distinguished from other opium preparations by the presence of plant debris, meconic acid, and the trace alkaloids. The residue will contain poppy capsule fragments and occasionally pollen grain. The poppy capsule fragments are epidermis composed of five- to six-sided cells with strongly thickened walls and sometimes with stellate lumina: infrequent anomocytic stomata, approximately 17 μm long, or sometimes circular. Occasional subspherical pollen grains with three pores are also observed [17].

Marquis, Fröhde, and Mecke reagents for visualization and the resulting colors are compared with those obtained from the reference morphine alkaloid (Table 1). To identify the opium, meconic acid color test is also required.

The developing solvent for TLC is ethylacetate/methylalcohol/concentrated ammonia (85 : 10 : 5 v/v). Acidified potassium iodoplatinate solution is used as the spray reagent [12]. This method allows separation of opiates, such as 6-acetylmorphine (6-AM), heroin, and acetylcodeine [12].

GC/MS is a reliable, simple, and sensitive method but HPLC procedures are more often used for the determination of opiates than GC/MS. HPLC is the method of choice for the quantitative determination of morphine and related alkaloids [18, 19]. However, the run times are somewhat long. CE gives a superior separation of the alkaloids found in crude morphine, poppy straw, and opium preparations and provides a faster and less costly analysis with quantitative results than HPLC [20]. MECC provides an excellent separation of the highly complex mixture of neutral and acidic impurities present in heroin [8–9].

Cocaine

Each country has its own penalties for trafficking and possession of cocaine. In the United States, the possession with intent to sell 50 g of crack cocaine would draw a 10-year mandatory minimum sentence. Singapore imposes death penalty on anyone convicted of trafficking in more than 30 g of cocaine. After weighing the samples, the powders and crystals are pulverized into powders, of which about 10 mg is taken. Cocaine crystals and powder are dissolved in methanol for chemical analysis. In the case of coca leaves, 0.5 g of leaves is ground and sonicated in methanol, or alternatively, soaked in 0.1 N hydrochloric acid. The drug is extracted with chloroform. The chloroform layer is collected and dehydrated with anhydrous sodium sulfate [12]. Immersion of coca leaves in boiling ethanol for a short period and extraction with hot methanol results in effective extraction of ecgonine-type alkaloids and minimizes the breakdown of cocaine [21]. Cocaine tablets are ground and sonicated in methanol prior to analysis.

Cobalt thiocyanate test or the modified one (Scott test) that gives a blue color indicates the presence of cocaine. A quantity of 5% methanolic sodium or potassium hydroxide is added to the test sample and warmed. It gives a characteristic odor in the presence of cocaine [11, 21].

Cocaine is classified as crack cocaine (cocaine base) and cocaine HCl. They are weighed, dissolved in water and ethanol to observe the solubility, and treated with the silver nitrate solution for the chloride test, which gives a white curdy precipitate indicating the presence of chloride. The crack form of cocaine is a free base and soluble in hexane, while the hydrochloride form is insoluble [12].

Four TLC solvent systems are used to distinguish the opiate-like drug: chloroform/dioxane/ethyl acetate/ammonia (25 : 60 : 10 : 5), 0.81; methanol/

Table 1 Results of color test for opiates

Alkaloids	Marquis reagent	Fröhde reagent	Mecke reagent
Morphine	Purple violet	Purple becoming gray/purple	Dark green
Codeine	Purple violet	Blue/green	Green/blue
Heroin	Purple violet	Purple becoming gray/purple	Dark green
6-acetylmorphine	Purple violet	Yellow/green	Dark green
Papaverine	No color	Light green	Dark blue

ammonia (100 : 1.5), 0.59; cyclohexane/toluene/diethylamine (75 : 15 : 10), 0.56; and ethylacetate/methanol/concentrated ammonia (85 : 10 : 5), 0.7. Acidified potassium iodoplatinate and Dragendorff's reagent are applied as the color spray [12, 21].

GC/MS is preferred to identify cocaine with a high degree of sensitivity and selectivity [22]. The IR method is particularly useful for identifying illicit cocaine in crystal or powder form. Using IR spectrometry, cocaine base and cocaine hydrochloride can be differentiated [1, 23].

LSD (Lysergic Acid Diethylamide)

Punishment of drug offenders for LSD differs among different countries. For the possession of LSD in Canada, if tried by indictment, an offender gets a maximum penalty of $5000 and a three-year imprisonment. LSD-related samples such as LSD tainted paper and candies are shaken in tartaric acid for tests. The pH of the filtrate is adjusted to be 8–9 with 10% ammonia water and the solution is extracted with dichloromethane. Then, the extracts are dehydrated with anhydrous sodium sulfate [12]. After the dichloromethane is removed, the residue is finally dissolved in ethanol.

The presence of LSD is indicated by a blue fluorescence [11]. Then, a drop of Ehrlich reagent is added over the spot. A blue to purple color shows the presence of LSD [18, 22].

The melting point is measured with the dried sample. The melting point of the tartrate salt of LSD is 198–200 °C and that of free base of LSD is 80–85 °C [12].

The developing solvents used and the R_f values of LSD are as follows: chloroform/methanol (9 : 1), 0.5; methanol/concentrated ammonia (100 : 1.5), 0.8; chloroform/acetone (20 : 80), 0.6. Ehrlich reagent is used as the color spray [12, 24], displaying a blue-purple color with this reagent.

Selective extraction and derivatization by silylation of the indole nitrogen of LSD is necessary for GC/MS analysis. Alternatively, HPLC separation together with fluorescence detection is applied; however, it lacks sensitivity and specificity. Consequently, LC/MS/MS can be used as the method of choice for LSD determination [25]. LSD, lysergic acid methyl propyl amide (LAMPA), and iso-LSD are resolved in less than 5 min using CE [8–10].

Phenethylamines

Singaporean law imposes a maximum penalty of 10 years in jail and a $20 000 fine for possession and up to 20 years in jail for trafficking of ecstacy. Other countries have their own punishment for possession of ecstacy.

Phenethylamines include amphetamines and designer analogs such as methylenedioxymethamphetamine (MDMA), methylenedioxyamphetamine (MD A), methoxymethylenedioxyamphetamine (MMDA), methyldimethoxyamphetamine (DOM), dimethoxybromoamphetamine (DOB), dimethoxyethylamphetamine (DOET), dimethoxyamphetamine (DMA), paramethoxyamphetamine (PMA), and trimethoxyamphetamine (TMA).

After weighing the seizures, coarse particles or crystals if any are pulverized into powders and 10 mg of powder is sampled for the screening and confirmation test. Crystals and powders are usually weighed without drying; however, dried samples are used for the measurement of melting point and FTIR analysis. Tablets are ground to a fine powder and the sample solution is prepared with methanol without any pretreatment.

The most important color tests for phenethylamines are the Marquis and Simon's tests. The results of the major color tests are shown in Table 2 [26].

The TLC plate is developed with the two types of developing solvents (1) chloroform/acetone/methanol /concentrated ammonia (15 : 12 : 3 : 0.1) and (2) methanol/concentrated ammonia (100 : 1.5)). Simon's reagent and acidified potassium iodoplatinate reagent are used for identification [12].

Even though derivatization of phenethylamines is not mandatory, the selectivity and specificity of the mass spectra of phenethylamines are improved by silylated derivatization. In the GC/MS analysis, the order of elution for underivatized phenethylamines on DB-5 and DB-1 columns is as follows: PMA<MDA<DMA<MDMA<DOM< MMDA<DOET<TMA<DOB.

Indoleamines (Psilocybin and Psilocin)

After weighing the seizures, about 20 mg of the suspected psilocybe mushroom is ground to powder and extracted with methanol, while powders, tablets, and capsules of psilocybin are pulverized into powder and dissolved in methanol.

Table 2 The results of the major color tests for phenethylamines[a]

Color test	MDMA	MDA	MMDA	DOM	DOB	DOET	DMA	PMA	TMA
Marquis	Dark blue/black	Dark blue/black	Purple	Yellow	Yellow → green	Yellow brown	Green → dark green	NR → Light green	Orange
Simon's	Deep blue	NR	NR	NR	NR	NR	NR	NR	NR

[a] Reproduced with permission from Ref. 26. © UNDOC, 2006
(NR) – no reaction

Commonly two types of TLC developing solvents are used: Solvent 1 is methanol/28% aqueous ammonia (100 : 1.5) and solvent 2 is 1-butanol/acetic acid/water (2 : 1 : 1). Ehrlich reagent and *p*-dimethylaminocinnamaldehyde are used as spray reagents [27]. When Ehrlich reagent and *p*-dimethylaminocinnamaldehyde are sprayed, psilocybin gives gray-violet to violet color while psilocin gives blue color.

For GC/MS analysis, indoleamines are analyzed with or without derivatization. It should be noted that direct injection of psilocybe mushroom extracts or psilocybin sample solutions on capillary columns without derivatization will convert psilocybin to psilocin by thermal dephosphorylation and accordingly only psilocin is detected. For the derivatization, trimethylsilyation derivatization is commonly used. The quantitative analysis for psilocybin and psilocin are performed by HPLC or liquid chromatography electrospray ionization mass spectrometry (LC/ESI/MS).

References

[1] Yukiko, M. (1998). *Manual for Identification of Abused Drugs*, 2nd Edition, Pharmaceutical and medical safety bureau, Ministry of Health and Welfare, Japan, pp. 250–267.

[2] Kupirec, T., Slawson, M., Pragst, F. & Herzler, M. (2004). *High Performance Liquid Chromatography. Clarke's Analysis of Drugs and Poisons in Pharmaceuticals, Body Fluids and Postmortem Materials*, 3rd Edition, Pharmaceutical Press, London, Chicago, pp. 500–534.

[3] Sheila, D. (2004). *Gas Chromatography. Clarke's Analysis of Drugs and Poisons in Pharmaceuticals, Body Fluids and Postmortem Materials*, 3rd Edition, Pharmaceutical Press, London, Chicago, pp. 425–499.

[4] David, W. (2004). *Mass Spectrometry. Clarke's Analysis of Drugs and Poisons in Pharmaceuticals, Body Fluids and Postmortem Materials*, 3rd Edition, Pharmaceutical Press, London, Chicago, pp. 379–391.

[5] Jan, C. & John, S. (2004). *Ultraviolet, Visible and Fluorescence Spectrometry. Clarke's Analysis of Drugs and Poisons in Pharmaceuticals, Body Fluids and Postmortem Materials*, 3rd Edition, Pharmaceutical Press, London, Chicago, pp. 313–327.

[6] Alex, D. (2004). *Infra-red Spectrometry. Clarke's Analysis of Drugs and Poisons in Pharmaceuticals, Body Fluids and Postmortem Materials*, 3rd Edition, Pharmaceutical Press, London, Chicago, pp. 328–345.

[7] John, C.L. & Jeremy, K.N. (2004). *Nuclear Magnetic Resonance Spectroscopy. Clarke's Analysis of Drugs and Poisons in Pharmaceuticals, Body Fluids and Postmortem Materials*, 3rd Edition, Pharmaceutical Press, London, Chicago, pp. 368–378.

[8] Lurie, I.S. (1998). Capillary electrophoresis of illicit drug seizures, *Forensic Science International* **92**, 125–136.

[9] Lurie, I.S., Chan, K.C., Spratley, T.K., Casale, J.F. & Issaq, H.J. (1995). Separation and detection of acid/neutral impurities in illicit heroin via capillary electrophoresis, *Journal of Chromatography* **669**, 3–13.

[10] Walker, J.A., Krueger, S.T., Lurie, I.S., Marche, H.L. & Newby, N. (1995). Analysis of heroin drug seizures by micellar electrokinetic capillary chromatography (MECC), *Journal of Forensic Science* **40**, 6–9.

[11] Les, A.K. & Sean, D.M. (2004). *Drugs of Abuse. Clarke's Analysis of Drugs and Poisons in Pharmaceuticals, Body Fluids and Postmortem Materials*, 3rd Edition, Pharmaceutical Press, London, Chicago, pp. 37–52.

[12] Yukiko, M. (1998). *Manual for Identification of Abused Drugs*, 2nd Edition, Pharmaceutical and medical safety bureau, Ministry of Health and Welfare, Japan, pp. 8–127.

[13] http://www.unodc.org/unodc/en/data-and-analysis/bulletin/bulletin_1950-01-01_4_page003.html, (accessed on, 2007).

[14] *Recommended Methods for Testing Illicit Ring-substituted Amphetamine Derivatives*, (1987). Manual for use National Laboratories, United Nations, New York, pp. 16–20.

[15] Piette, V. & Parmentier, F. (2002). Analysis of illicit amphetamine seizures by capillary zone electrophoresis, *Journal of Chromatography. A* **979**, 345–352.
[16] Bokor, I., Trenerry, V.C. & Scheelings, P. (1997). Separation and quantitation of optical isomers of methylamphetamine samples by capillary electrophoresis, *Forensic Science International* **85**(3), 177–192.
[17] *Recommended Methods for Testing Opium, Morphine, and Heroine*, (1998). Laboratory Section, United Nations, New York, pp. 20–24.
[18] Nobuhara, Y., Hirano, S., Namba, K. & Hashimoto, M. (1980). Separation and determination of opium alkaloids by high-performance liquid chromatography, *Journal of Chromatography* **190**, 251–255.
[19] Doner, L.W. & Hsu, A.F. (1982). High-performance liquid chromatographic separation of alkaloids from Papaver somniferum on a Zorbax NH2 analytical column, *Journal of Chromatography* **253**, 120–123.
[20] Trenerry, V.C., Wells, R.J. & Robertson, J. (1995). Determination of morphine and related alkaloids in crude morphine, poppy straw and opium preparations by micellar electrokinetic capillary chromatography, *Journal of Chromatography. A* **718**, 217–225.
[21] *Recommended Methods for Testing Cocaine*, (1986). Manual for use by National Laboratories, United Nations, New York, pp. 8–23.
[22] Pujadas, M., Pichini, S., Civit, E., Santamariña, E., Perez, K. & de la Torre, R. (2007). A simple and reliable procedure for the determination of psychoactive drugs in oral fluid by gas chromatography–mass spectrometry, *Journal of Pharmaceutical and Biomedical Analysis* **44**, 594–601.
[23] Elsherbini, S.H. (1998). Cocaine base identification and quantification, *Forensic Science Review* **10**, 1–12.
[24] *Recommended Methods for Testing Lysergide (LSD)*, (1989). Manual for use by National Laboratories, United Nations, New York, pp. 12–14.
[25] Johansen, S.S. & Jensen, J.L. (2005). Liquid chromatography–tandem mass spectrometry determination of LSD, ISO-LSD, and the main metabolite 2-oxo-3-hydroxy-LSD in forensic samples and application in a forensic case, *Journal of Chromatography. B* **825**, 21–28.
[26] *Recommended Method for the Identification and Analysis of Amphetamine, Methamphetamine and Their Ring-substituted Analogues in Seized Materials*, (2006). UNODC, pp. 19–20.
[27] *Analytical Methods for Drug and Toxicology and A Commentary-Analysis, Toxicity & Management*, (2006). The Pharmaceutical Society of Japan, written in Japanese, pp. 159–163.

HEESUN CHUNG, EUNYOUNG HAN, SOOYEUN LEE, SANGKIL CHOI, JUSEON LEE, HYEYOUNG CHOI, YONGHOON PARK, EUNMI KIM AND MIAE LIM

Drug Analysis: Alcohol *see* Alcohol: Analysis

Drug Impaired Driving *see* Drug-Impaired Driving

Drug Interactions *see* Alcohol: Interaction with Other Drugs

Drug Profiling

Introduction

Illicit drug profiling is the process of extracting components called *profiles* from drug seizures. The term *drug* is used as a generic term to include all illicit products controlled under the UN conventions and protocols on narcotic and psychotropic drugs, and drug traffic [1–3].

These components of the profiles can vary. They can be chemical, such as residual solvents or impurities resulting from illicit drug synthesis, or physical such as the impression of a logo in tablets such as ecstasy tablets. The detection and the recording of these "profiles" result from different analytical methods. This provides data amenable to comparisons between different illicit drug seizures for detecting similarity or identity. The resulting information may be used as objective information useful to decrypt the drug trafficking organization and distribution pattern for law enforcement purposes (intelligence) or used in a more traditional manner as evidence in court for demonstrating that two seizures are composed of the same illicit products.

Heroin, cocaine, ATS, and cannabis constitute the main substances subject to organized trafficking and

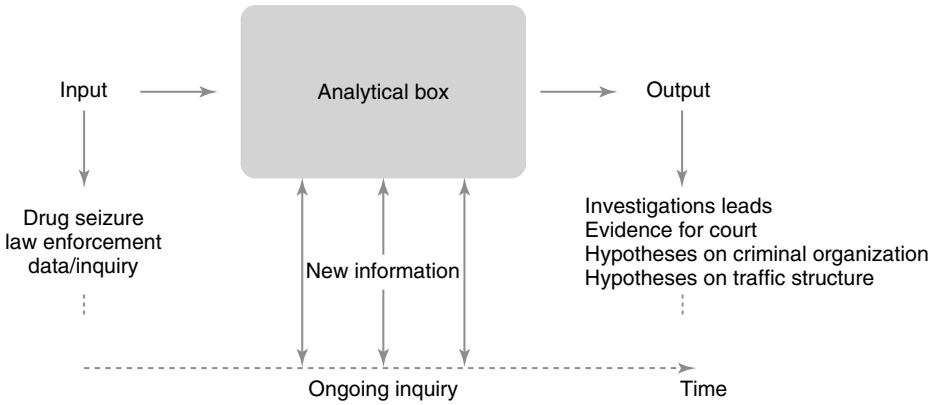


Figure 1 General concept of drug profiling process

are extensively used in the community. These illicit products are the focus of the debates both in the scientific literature and in the political arena. They also constitute what is perceived as a major health problem in most consumer countries. Therefore, from a profiling perspective, the focus is mainly on these four illicit substances.

The Drug Profiling Process

The drug profiling process is described in Figure 1.

The input starts with a drug seizure from a law enforcement activity and by gathering the circumstantial information available, i.e., the context in which the seizure occurred. The other source of information is obtained from the seizure itself, its analysis and interpretation. The combined information may be analyzed and interpreted within the analytical box (described hereafter) in order to extract pertinent outputs from this data. During this ongoing task, the inquiry remains active, may provide new information, or may use investigative leads obtained from the intelligence process as outlined in the analytical box. The output can come from various sources but, in any case, the aim is to extract useful information to pursue complex investigations, understand key elements of a criminal activity, make reasonable hypotheses on criminal organizations or on the structure of the drug trafficking and, finally, provide evidence for the court.

This analytical box is not a black box, but an organized analytical process to discover and understand phenomena of a complex nature (criminal activity, criminal organization, etc.). It uses an iterative approach that could be schematized as follows:

1. Tacit knowledge obtained from research – establishing analytical protocols [4], classifying discriminating factors [5], indicators from the chain of drug production [6], and transformation to the final product seized on the street [7], validating a database from past cases [8]. This is a key element in order to be able to extract knowledge from new cases.
2. The preliminary action for new cases consists in collecting pieces of information obtained from the drug trafficking that is known or perceived. This step is of major importance; indeed, without information, the intelligence process is stopped and cannot be achieved. These pieces of information consist of the drug seizure itself and all information (telephone call, informant data, bank transaction, etc.) gathered by different investigators.
3. Analytical data are obtained (i.e., extraction of profile for the drug seizure) and the results are organized in a database of relevant data (*see* **Cocaine**; **Opioids**; **Drug Analysis**).
4. Once this is completed, the interpretation process is initiated producing "forensic" intelligence (information extracted from the seizure profiles), which is combined to inquiry intelligence obtained from the investigative data [9]. The two sources of intelligence are combined to produce useful knowledge on drug trafficking both from an operational and a strategic perspective.

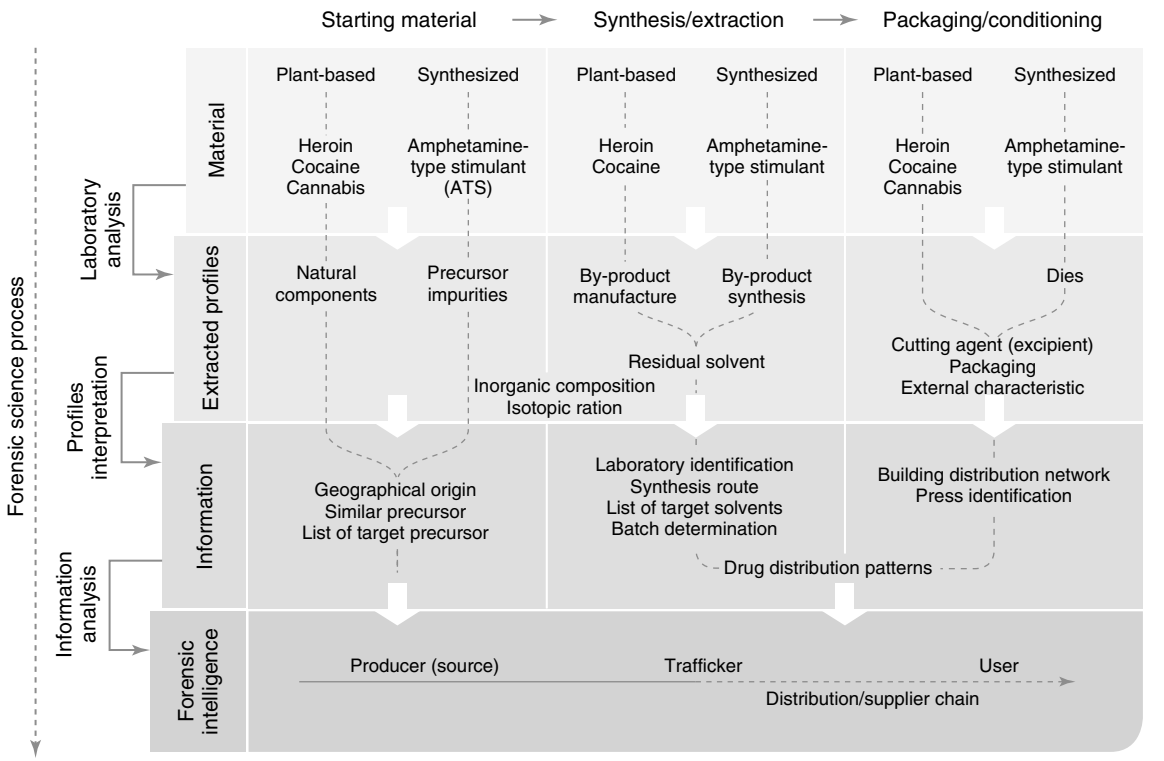


Figure 2 Forensic science processes

Within this framework of drug profiling, the focus is on the scientific process that allows the extraction of profiles from drug seizures. A major distinction concerns the nature of the starting material. If the raw material is "plant-based" (cocaine, heroin, and cannabis) or originates from synthetic precursors (ATS), the manufacturing process is drastically affected and influences the type of data that can be extracted from the seized products.

Figure 2 highlights the contribution of each production step to the chemical composition (the profile or "signature") of the final product. The raw materials such as coca leaf and opium, used for the production of cocaine and heroin, have a different composition depending on the geographical area and growing conditions, mainly because of climatic and soil differences (traces of origin). These differences could also be observed in the case of synthesized drugs if precursors are of natural origin, but the composition of the final product of the latter depends largely on the quality and the nature of the precursors. Therefore, the type of information may differ significantly and the localization of an origin is mostly unrealistic.

The extraction of the raw material and the syntheses that may follow in order to obtain the final product are crucial parts of the production process, strongly influencing the composition of the final product. Many variables such as the chemical recipes, the selected chemicals, the scale of the reaction, purification steps, etc., influence the diversity observed between production batches. Each of these variables has an impact on the formation and the proportion of so-called by-products found in the final product. The by-products are found both in the production of drugs of natural or synthetic origin. This is often referred to *traces of production* or incorrectly *production impurities*. Finally, the illicit substance is conditioned, packaged, split into batches for dealers active in the illicit market. This phase is also of major interest due to the addition of cutting agents and to the packaging characteristics that can provide additional information about the network of distribution and the dilution of the active ingredients

down to the consumer, at the end of the distribution process.

A classification of the information of this chemical/physical profile may have significantly different meanings depending for whom the forensic intelligence is intended. The first and often most sought after intelligence information takes a political dimension. It is focused on a national and international control of illicit products. This may play a significant role in geostrategic negotiations and interactions. The second type of information that is traditionally understood under the terminology "drug intelligence" is strategic and operational information about regional and local traffic structure and organization. This is mostly used for law enforcement purposes and local prevention or fight against drug-related criminality. The third, but nevertheless an important purpose is evidential. The demonstration of criminal activity, its extent, the participation in a traffic organization, and the relation to other criminals may offer sufficient strength for the court to decide on the most serious offense related to organized crime.

Support to National, International Policies (Regulatory Authorities)

Support to policy is a category of information centered on national, international, and geostrategic preoccupations. At this level, the priorities are oriented to identify or establish illicit drug distribution patterns or trafficking routes, to target the production (monitoring methods used for clandestine manufactures and location of the production), and identify the source of precursors. In order to support the authorities in their strategic choice, drug profiling research has developed dedicated methodologies. The identification of a geographical source of production is a perfect example of drug profiling research that has been directed to supporting policy initiatives.

Each manufacturing step contributes to the final chemical/physical profiles of the illicit products as demonstrated in Figure 2. In the case of plant-based materials, the natural components present in raw material have been used to determine their geographical origin. Indeed, within a region, the climatic conditions influence the proportion of natural compounds typical of the origin or source of the initial product. Furthermore, the clandestine manufacture process could also be specific of a location, producing a correlated combination of by-product compounds [10].

The determination of geographical origin necessitates extensive knowledge of localized variations through the maintenance of a database of specimens from known geographical sources. This could be done by having specimens seized directly in the production zone and in illicit laboratories. However, the maintenance of such an up-to-date database is problematic due to the difficulties of obtaining reliable specimens of known and authentic origin. Other complications arise from the fact that the growing conditions could evolve, as well as the manufacturing process inside a given region. These modifications influence the pattern of natural and by-product compounds present in an illicit drug seizure. In order to make a pertinent prognosis of the geographical source of an analyzed sample, these modifications have to be included in the database; otherwise the quality of the results may quickly become useless.

For heroin samples, the opium, the raw material, is extracted from the opium poppy, *Papaver somniferum*. The major component of opium is morphine, an alkaloid (basic natural chemical compound with a nitrogen atom in the molecule that may react with an acid to make a salt), that is transformed by acetylation into diacetylmorphine, also known as *heroin*. During the extraction of morphine from opium and its transformation into heroin, other alkaloids are also coextracted and acetylated (if they have a reactive group). Papaverine, noscapine, acetylthebaol, acetylcodeine, and 6-acetylmorphine are such compounds present in relatively high proportions and are a reflection of proportions found in a given geographical source.

The literature refers to major alkaloids that are mostly analyzed using gas or liquid chromatographic methods. The combination of the ratios of these alkaloids allows a classification of drug seizures into four different general regions of origin. Those regions are widely defined as Southwest Asia, Southeast Asia, Mexico, and South America [11, 12].

In the specific geographical location of heroin seizures, by-products of the manufacture are analyzed in addition to the usual compounds present in the material in order to determine the source of production. The strategy of combining multiple types

of profiles is often promoted to elicit more refined information.

Geographical origin determination has also been thoroughly studied for cocaine seizures. The composition of *Erythroxylum coca v. coca* whose leaves contain the cocaine alkaloid is also influenced by climatic and environmental conditions. During the extraction of cocaine, other alkaloids present in the leaves are coextracted. Ecgonine methyl ester, norcocaine, *trans*-cinnamoylcocaine, *cis*-cinnamoylcocaine, benzoylecgonine, and trimethoxycocaine are all alkaloids present in relatively high proportions in seizures of cocaine and are used in conjunction with other minor impurities to predict a geographical origin. Owing to the selectivity, high sensitivity, and reproducibility of gas chromatographic techniques, these methods are the methods of choice for the analysis of these components that are often present in low concentration in the matrix of cocaine samples [13, 14].

It must be noted that geographical source determination is a difficult task due to the underground and illicit nature of the products and the lack of true source references. Most studies use investigative information to locate probable locations but confidence levels can rarely be defined accurately.

The determination or an inference of geographical origin for synthetic drugs is not possible, since precursors and the synthetic methods used are not specific to a region. Samples having a similar chemical profile can only be related to an individual laboratory. However, such a laboratory could be located anywhere and the information has to be used differently to products obtained from natural sources. The chemical profile of an ATS sample shows impurities obtained from precursors (which may also have impurities if they are manufactured clandestinely) and from the synthetic pathways used [15, 16]. Theoretically, it is possible to determine if two seizures have originated from the same precursor and/or if they have been synthesized according to the same synthesis route by monitoring this set of impurities. This information plays a major role in helping to detect modifications in the precursors or chemicals used (according to the synthetic pathway), thus allowing a more efficient international perception of the synthetic methodology and setup needed for production. This can help in targeting production and decreasing the diversion of synthetic drug precursors.

Again, it is important to emphasize that the interpretation of the data in order to determine if the precursors come from the same clandestine source or synthetic route is difficult and must rely on an up-to-date database of so-called authentic specimen representative of the interpretation that is pursued.

Other constituents of heroin, cocaine, and ATS matrices have also been used for geotracing. Recently published methods claim successful classification of source using isotopic abundance of ^{13}C and ^{15}N. Isotopes are atoms whose nuclei contain the same number of protons, but a different number of neutrons. The exposing number placed before the chemical symbol of the element indicates the isotope, which is equal to the number of nucleons (neutrons plus protons). For example, carbon has three natural isotopes which are noted: ^{12}C, ^{13}C, ^{14}C. The isotopes ^{12}C and ^{13}C are stables, whereas the isotope ^{14}C is unstable (radioactive). On earth, the quantity of atoms of each isotope is constant. Only the distribution of the atoms within the molecules is variable and depends on the process of isotopic fractionation. The isotope ratios of compounds can be modified by a phenomenon of discrimination of mass due to biological, chemical, and physical processes. Thus, isotope ratios of an element depend on its origin and its history.

This methodology has been successfully applied to locate seized cocaine in four different regions of Colombia, Bolivia, and Peru. Isotopic ratios have also been used successfully for classifying cannabis and heroin samples [5].

Drug profiling analyses also offer the possibility to map the solvents used during the manufacturing process. Local changes in regulations may be witnessed in the illicit productions. This aspect has been studied for heroin, cocaine [17, 18], and ATS [19]. Such data may be a crucial source of information for detecting new trends in the chemicals employed and may offer the possibility to inform the concerned authorities about these evolutions.

Support to Law Enforcement Investigation

Drug profiling is an important tool for investigative purposes. The analytical results are collected through a systematic process in which each new sample analyzed is compared with existing entries of an organized memory built upon earlier cases stored in a dedicated database.

The intelligence supplied by the information extracted from the data, using previously described analytical methods, can be used operationally to try to decipher and understand the structure of the drug market. It also has the potential to detect new phenomena or new trends in the illicit drug network distribution.

The links highlighted between seizures can be directly included in the enquiry in the same way that other information collected by law enforcement agencies are being used. The chemical/physical link is not directly meant for evidential purpose but rather to help formulate reasonable and useful investigative orientations; the interpretation is not limited in determining whether the seized specimens have exactly the same history. The correspondence between two profiles is considered in a wide range of interpretations and has to be combined with investigative data. It is a piece of intelligence in the course of the inquiry.

The applications may range from a confirmation of a hypothesis of connection inferred by law enforcement officers on the basis of inquiry information or the diversion of the investigation in other, unsuspected, directions [20] or may even lead to completely new investigative hypotheses. This analysis is only possible if an extensive database recording previous cases exists and the methodology for extracting potential links has been dedicated to such a purpose. The extraction and management of such information further requires the development of a specific management structure developed through research. This management aspect includes an important inference mathematical tool regarding the grouping of similar samples into chemical/physical classes. This clustering step utilizes supervised (where previous knowledge on the distribution inside the chemical/physical classes of the samples stored in the database is needed) and nonsupervised statistical measurements. The utilization of supervised methods like the soft independent modeling of class analogy (SIMCA) or artificial neural network (ANN) type are elegant tools for assigning new analyzed samples into preexisting classes.

Different authors have investigated these kinds of methods and demonstrated their potential value [21, 22]. It needs initially the creation of a mathematical model for each class. This model is used as a reference to determine if a new candidate belongs to an existing chemical/physical class. The calculation of the model utilizes statistical fundaments regarding the determination of thresholds for evaluating whether seized specimens belong to a same batch. This threshold is set by the scientist according to the needs of the investigation. The initial hypothesis usually states that it is possible to evaluate the intra-variability of a batch of production by analyzing large seizures and, by analyzing many unrelated seizures, it is possible to determine inter-variability arising from different productions. The models could also be trained according to different hypotheses such as the determination of a geographical origin or the assignment of a synthetic route. The requirements for the training of these models are the availability of sufficient samples representative of the different classes to which the new analyzed specimens potentially belong.

It is crucial that this information is given in timely fashion to the law enforcement agencies. Indeed, if the information provided by the profiling process is not timely, or does not reach potential users, there is no reason to maintain a database and to develop a complex architecture of pattern recognition. Furthermore, it is necessary to demonstrate how useful this information can be, despite the complexity. Spectacular successes of utilization of links in a law enforcement context have shown the potentially rich dividends that can be gained using drug intelligence information. This has to be presented to convince users because it is complex and may not be directly intuitive. Diverse attempts have been done by different authors [20].

If the information gathered by chemical/physical links is not correctly integrated and understood by investigative officers the information may be used erroneously and will discredit the whole profiling process.

One major difficulty is the lack of centralized analyses due to jurisdiction boundaries that slows down the utilization of chemical links. Indeed, it is not easy to understand that a profile extracted in a laboratory is not directly comparable with a profile highlighted in another one for an end user of the information. The problem is complex because the harmonization of analytical methods in order to obtain comparable results is not an easy challenge. It would be preferable to have one single laboratory centralizing both analyses and drug intelligence routine, but this is often organizationally impossible to set up. Indeed, it is not sufficient to

apply a kit in order to obtain compatible data. The principal problem encountered is the reproducibility of the data between different partners [23]. Nevertheless, recent research has shown that harmonization of methods is not a utopia using new generation analytical instruments and could be realized under a strict control of the analytical conditions. Such an approach is actually promoted through different European and international research projects and constitutes the basic foundation for developing a common utilization of drug profiling data in an intelligence-led perspective. Ideally, internationally collected databases could have an important impact for intensifying the utilization of profiling information by law enforcement agencies on a regional or international level.

Evidence for the Court

Using drug intelligence in the investigation of drug trafficking will often lead to arrests and further seizures of substantial amounts of illicit material that is evidence in itself of criminal activities. The result of linkage is therefore rarely, if at all, needed for court purposes. But when handling specific cases, one typical question that arises from prosecutors or law enforcement officers is whether it is possible to establish a specific link between two illicit drug seizures and then linking the two persons in possession of these illicit substances in the same way. This comparison of specimens originating from selected cases, also called *case to case comparison* is at the present time privileged in the different laboratories active in drug profiling. The aim is to demonstrate that the selected samples have a common history [4]. If this hypothesis can be verified, then there is evidence of trafficking or distributing linkage. However, in order to use this data for evidentiary purposes, it is advised that the information coming from different analytical techniques is combined. For example, many authors advocate the combination of the analysis of major alkaloids compounds for the comparison of two heroin samples; followed by the analysis of minor alkaloids [24] (trace-level acidic and neutral impurities), occluded solvents, isotopic analysis [5], elemental analysis [25], as well as cutting agents [7] present in the sample. In an ideal linkage, the chemical signature of the specimens should correspond. The two compared samples share apparently an identical history from their geographical origin and manufacturing processes to the distribution network (*see* **Evidence Interpretation: a Logical Approach**; **Statistical Evidence in Court**) .

For this type of analysis and opposed to geographical determination it is not necessary to maintain a database since the samples are already preselected. In contrast, the interpretation of the results of the comparison require a complete background knowledge about the process of illicit drug manufacturing allowing experts to interpret differences and similarities observed between the "chemical signatures" of the compared samples.

The case where the features are identical is a theoretical concept and is never encountered in reality. Indeed, even in a same batch of production of heroin or cocaine there is some variation due to the nonhomogeneity of the specimen (intra-batch variation). The inter-batch variation (variation between production batch in a single laboratory) has also to be evaluated. Unfortunately, there is not a lot of information available regarding the variation of chemical profiles between different batches for a same laboratory and also the size of a batch. However, the chemical profile being highly dependant of the manufacturing method, any modification will affect the profile of the illicit drugs [6]. This is supported by evidence that the manufacturing conditions of the laboratories in Afghanistan or in Colombia are rather crafty. Also, any manipulation from the original batch chemistry will create a difference (two different dealers diluting specimens in a different way) that will be increasingly difficult to interpret without some background knowledge.

However, the notion of batch variation is more defined in the domain of ATS. Several illicit laboratories have been dismantled in Europe and important batches of production have been sequestered. This opportunity has permitted laboratories to determine that the intra-batch variation is rather small. This is mainly due to the manufacturing conditions that are more controlled with the synthesized illicit products than with plant-based drugs [4].

Furthermore, the analytical process itself introduces differences between analyses of the same specimen (this is an indication of the reproducibility or relative error of the analytical method). The role of the expert is to interpret these differences in order to conclude if the specimens came from

the same batch or not. The scientific literature contains descriptions of methodologies developed for interpreting the results. They range from visual comparisons of chromatograms to the use of advanced multivariate analysis like ANNs. Multivariate analysis is necessary for conducting these calculations since the results of illicit drug analysis are described by many variables (multivariate data). Multivariate analyses consider these variables in the form of a matrix allowing the possibility of performing complex calculations. More information is available in [26].

An attempt to determine the intra-batch variation consists in studying the homogeneity of specimens within a same seizure or, if available, coming from the same batch (for example, if one clandestine laboratory has been dismantled and batches prepared for distribution can be analyzed) and evaluate the homogeneity of the specimens. This evaluation, based generally on statistical measurements such as correlation or distance measurements [27], could be used to determine a threshold value below which it cannot be concluded that a given specimen came from a specific batch.

This threshold is set high in order to avoid false positives and to have a high level of confidence regarding the highlighted correspondence. The evaluation of any such link should be combined with other sources of information like that obtained from studying the packaging and the cutting agents [28].

Finally, when two observed profiles correspond, there is a need to determine whether the methodology was discriminating and whether the highlighted features are common or rare. This is of major importance for interpreting corresponding profiles. For example, the strength of evidence of two corresponding profiles of cutting agents will be quite different if the profiles contain only lactose or if it shows a complex and rare mixture of five different types of sugar. Cutting agent could be of two different types. The diluents (sugars, starch, ...), who have no pharmacologic effects, are substances added to the illicit drug in order to increase the mass and, by the way, the profit. In contrary adulterants (lidocaine, phenacetin, ...) have pharmacologic effects and are added both for increasing the mass and also modifying the effects of the consumption of illicit drugs.

Various methods of comparison and decision theories have been developed in the literature to determine whether two specimens come from the same batch of fabrication. A continuous approach based on likelihood ratios can be used combined with a statistical method based on thresholds [29]. This is an efficient tool for the scientists to measure the strength of the scientific evidence in forging his opinion.

Conclusions

Profiling consist in the extraction of illicit drugs signatures. These signatures are dependant of various features like the nature of the starting material, the manufacturing process and chemicals used. As presented in this article, the profiles extracted from the analysis could provide relevant information about the nature and the structure of illicit drug trafficking and solid investigative leads and reliable evidence in a recognized area of organized crime. This information is addressed at different level. Differentiation has been done between information oriented to support national, international policies (determination of geographical origin, determination of synthesis routes), law enforcement investigation and evidence for the court.

This approach, combine with a close cooperation with law enforcement agencies, has already demonstrated is potentiality to be an efficient an objective tool in the fight against illicit drugs trafficking.

Acknowledgment

The authors wish to acknowledge Romain Voisard for his support regarding the establishment of the graphical representations.

References

[1] (1961). *Single Convention on Narcotics Drugs*, United Nations.
[2] (1971). *Convention on Psychotropic Substances*, United Nations.
[3] United Nations (1988). *Convention against the Illicit Traffic in Narcotic Drugs and Psychotropic Substances*.
[4] UNODCCP (2000). *Drug Characterization/Impurity Profiling, Background and Concepts*, United Nations Publication, New York.
[5] Benson, S., Lennard, C., Maynard, P. & Roux, C. (2006). Forensic applications of isotope ratio mass

spectrometry – A review, *Forensic Science International* **157**, 1–22.
[6] Casale, J.F. & Klein, R.F.X. (1993). Illicit production of cocaine, *Forensic Science Review* **5**, 95–107.
[7] Fucci, N. & De Giovanni, N. (1998). Adulterants encountered in the illicit cocaine market, *Forensic Science International* **95**, 247–252.
[8] Esseiva, P., Dujourdy, L., Anglada, F., Taroni, F. & Margot, P. (2003). A methodology for illicit heroin seizures comparison in a drug intelligence perspective using large databases, *Forensic Science International* **132**, 139–152.
[9] Ribaux, O., Girod, A., Walsh, S.J., Margot, P., Mizrahi, S. & Clivaz, V. (2003). Forensic intelligence and crime analysis, *Law, Probability and Risk* **2**, 47–60.
[10] UNODC (2005). *United Nations International Drug Control Programme, Methods on Impurity Profiling in Heroin and Cocaine, Background and Concepts*, United Nations, New York.
[11] Dams, R., Benijts, T., Lambert, W.E., Massart, D.L. & De Leenheer, A.P. (2001). Heroin impurity profiling: trends throughout a decade of experimenting, *Forensic Science International* **123**, 81–88.
[12] Klein, R.F.X. (2003). Detection and analysis of drugs of forensic interest, 1992–2001 a literature review, *Microgram Journal* **1**(1–2), 55–153.
[13] Casale, J.F. & Waggoner, R.W. (1991). A chromatographic impurity signature profile analysis for cocaine using capillary gas chromatography, *Journal of Forensic Sciences* **36**, 1312–1330.
[14] Casale, J.F. & Moore, J.M. (1994). 3′, 4′, 5′-Trimethoxy-substituted analogs of cocaine, cis-/trans-cinnamoylcocaine and tropacocaine: characterization and quantitation of new alkaloids in coca leaf, coca paste and refined illicit cocaine, *Journal of Forensic Sciences* **39**, 462–472.
[15] Verweij, M.A. (1992). Impurities in illicit drug preparations: 3,4-(methylenedioxy) amphetamine and 3,4-(methylenedioxy)methylamphetamine, *Forensic Science Review* **137**, 138–176.
[16] Owist, M., Wilamowski, J. & Parczewski, A. (2005). Determination of synthesis method of ecstasy based on the basic impurities, *Forensic Science International* **152**, 175–184.
[17] Cartier, J., Guéniat, O. & Cole, M.D. (1997). Headspace analysis of solvents in cocaine and heroin samples, *Science and Justice* **37**, 175–181.
[18] Morello, D.R. & Meyers, R.P. (1995). "Qualitative and quantitative determination of residual solvents in illicit cocaine HCl and heroin HCl", *Journal of Forensic Sciences* **40**, 957–963.
[19] Kuwayama, K., Tsujikawa, K., Miyaguchi, H., Kanamori, T., Iwata, Y., Inoue, H., Saitoh, S. & Kishi, T. (2006). Identification of impurities and the statistical classification of methamphetamine using headspace solid phase microextraction and gas chromatography–mass spectrometry, *Forensic Science International* **160**, 44–52.
[20] Esseiva, P., Ioset, S., Anglada, F., Gasté, L., Ribaux, O., Margot, P., Gallusser, A., Biedermann, A., Specht, Y. & Ottinger, E. (2007). Forensic drug intelligence: an important tool in law enforcement, *Forensic Science International* **167**, 247–254.
[21] Esseiva, P., Anglada, F., Dujourdy, L., Taroni, F., Margot, P., Du Pasquier, E., Dawson, M., Roux, C. & Doble, P. (2005). Chemical profiling and classification of illicit heroin by principal component analysis, calculation of inter-sample correlation and artificial neural networks, *Talanta* **67**, 360–367.
[22] Casale, J.F. & Watterson, J.W. (1993). A computerized neural network method for pattern recognition of cocaine signature, *Journal of Forensic Sciences* **38**, 292–301.
[23] Strömberg, L., Lundberg, L., Neumann, H., Bobon, B., Huizer, H. & van der Stelt, N.W. (2000). Heroin impurity profiling: a harmonization study for retrospective comparisons, *Forensic Science International* **114**, 67–88.
[24] Neuman, H. & Gloger, M. (1982). Profiling of illicit heroin samples by high-resolution capillary gas chromatography for forensic application, *Chromatographia* **16**, 261–264.
[25] Wells, R.J., Skopec, S.V., Iavetz, R. & Robertson, J. (1995). Elemental analysis by ICP-MS: trace element analysis of heroin by ICP-MS, *Chemistry in Australia* **62**, 14–15.
[26] Massart, D.L., Vandeginste B.G.M., Buydens L.M.C., De Jong S., Lewi P.J. & Smeyers- Verbeke J. (1997). *Handbook of Chemometrics and Qualimetrics: Part A*, Elsevier Science, Amsterdam.
[27] Lociciro, S., Hayoz, P., Esseiva, P., Dujourdy, L., Besacier, F. & Margot, P. (2007). Cocaine profiling for strategic intelligence purposes, a cross-border project between France and Switzerland: Part I. Optimisation and harmonisation of the profiling method, *Forensic Science International* **167**, 220–222.
[28] Terrettaz-Zufferey, A.-L., Ratle, F., Ribaux, O., Esseiva, P. & Kanevski, M. (2007). Pattern detection in forensic case data using graph theory: Application to heroin cutting agents, *Forensic Science International* **167**, 242–246.
[29] Dujourdy, L., Barbati, G., Taroni, F., Guéniat, O., Esseiva, P., Anglada, F. & Margot, P. (2003). Evaluation of links in heroin seizures, *Forensic Science International* **131**, 171–183.

Related Articles

Opioids
Drug Analysis
Evidence Interpretation: a Logical Approach
Statistical Evidence in Court

PIERRE ESSEIVA AND PIERRE MARGOT

Drug Testing: Urine

Introduction

Urine testing has become the most frequently performed type of analysis in forensic toxicology, perhaps with the exception of alcohol. To standardize the testing and to provide a legally defensible procedure, a number of counties have issued standards and guidelines. These include the United States (through SAMSHA), the United Kingdom, and Australia.

Guidelines concentrate on a number of drug/drug classes including amphetamines, opiates, cannabis, cocaine, and benzodiazepines. In the United States, PCP is also included. Occasionally other drugs are tested, including barbiturates, methadone, buprenorphine, etc.

Basis of Drug Testing in Urine

Urine drugs-of-abuse testing is used essentially for two specific purposes: compliance and safety. The compliance environment would typically be correctional (custodial and community) as well as alternatives to custodial sentencing such as drug courts. In this environment, the requirement is for the individual to remain drug free and drug testing is undertaken to ensure compliance with that requirement. In the workplace environment, the emphasis is on safety and risk management. Companies undertake drug testing to reduce the risks associated with the use of impairing drugs in the workplace. This is achieved by identifying individuals whose lifestyle choices may lead them to use drugs in the workplace and possibly seriously endanger themselves and other individuals. This testing may include preemployment screening. Workplace testing definitely stresses the deterrence value of testing as well as the rehabilitative processes associated with the identification of these individuals.

Why Do We Test Urine?

Urine is a high-volume specimen, which is easily obtainable by a medically noninvasive process. Under normal conditions, it is sterile and contains high concentrations of water-soluble metabolites, which are markers for drug use. From an analytical point of view, urine is a good matrix for the laboratory to work with and given the high concentrations, sensitive sophisticated screening techniques are not generally required.

The body acts to remove drugs predominately by either biliary or urinary excretion aided in the majority of cases by metabolism. In the case of drugs, metabolism refers to chemical modification of the drug molecule mainly occurring in the liver producing a drug metabolite, which is less toxic and more water soluble. The increased water solubility aids excretion by the kidneys.

The kidneys filter blood to produce an initial filtrate rich in low-molecular-weight substances such as drug metabolites in large volumes of water. Selective reabsorption of water and essential substances such as glucose by transportation to the blood then occurs. The ability of other substances to reenter the blood stream without assistance depends upon their ability to penetrate the membrane surrounding the blood capillaries. This process is called *passive diffusion*. Fat-soluble substances such as parent drugs diffuse at much higher rates than water-soluble metabolites, with the consequence that water-soluble drug metabolites are lost to the urine and very little parent drug reaches the urine.

The kidneys also have a concentrating effect on the water-soluble metabolites. The initial filtrate is formed in an adult at 125 ml min^{-1} while urine under normal hydration conditions is formed at 1 ml min^{-1}, meaning that over 99% of the water has been reabsorbed. The metabolites that were previously in approximately 100 ml of blood are now in 1 ml of urine. Their concentration is therefore typically 100-fold higher than blood.

Collection of Urine

Testing consists of several steps: the collection of the urine specimen, dispatch to the laboratory, and laboratory analysis. Once the donor's identity has been established, the donor proceeds to provide a urine specimen under conditions of privacy; however, witnessed collections are possible in custodial or court-directed situations. While the specimen is being provided, the collector prepares the paperwork, i.e., a declaration of the donor's ownership of the specimen, consent for analysis, and chain of custody. The collector then establishes the validity of the sample (see the sections "Substitution" and "Adulteration"). Once this has been accomplished, the collector may

carry out an on-site drug test and/or an adulteration test or prepare the sample for dispatch directly to the laboratory.

If laboratory analysis is required, the specimen is usually split into two equal specimens, one of which is designated the referee specimen. Both are labeled with two unique identifiers neither of which should ideally include the name of the donor. The containers are then sealed with tamper evident tape, which is dated and initialed. The donor is required to sign the paperwork, which is marked with an identifier common to the specimens by the collector. At the time of signing, the donor is made aware of the drug/drug classes to which he/she is consenting for analysis. The specimen containers are placed in a transportation container along with their unique paperwork and dispatched to the laboratory in accord with any relevant regulations or legislation.

How is Laboratory Testing Carried Out?

The first step in the laboratory process is to inspect the specimens and accompanying paperwork for tampering and inconsistencies before completing the chain of custody section and finally entering the data for the specimens into the laboratory's tracking and registration system.

Laboratory testing is a two-step process, an initial screening test (*see* **Toxicology: Initial Testing**) followed in the case of a positive screen by a confirmatory test. The screening test is invariably an immunoassay and the confirmatory test is some form of chromatography (gas or liquid) coupled with mass spectrometry (MS) – *see also* **Confirmation Testing: Toxicology**.

Immunoassays

A drug immunoassay depends upon the interaction between an antibody raised against a specific drug/metabolite and the drug/metabolite itself. An antibody can be thought of as a molecule in an animal's blood, which seeks out a specific foreign substance and locks onto it with the purpose of disabling its action. This process is part of the normal immune response of animals. In this case, the antibody locks or binds to the drug/metabolite to which the antibody was made (raised). This is similar in concept to a lock (the antibody) and its key (the drug/metabolite). Binding of drug/metabolite is detected in various ways that may involve enzymes linked to a color reaction, radioactive atoms, gold colloidal particles, or fluorescent molecules. The response depends upon how much binding occurs.

In urine drugs-of-abuse testing cutoffs are normally applied. A cutoff is the concentration of drug/metabolite at or above which the specimen is considered to be positive for that drug/drug class. Cutoffs are set for various reasons: (i) to ensure uniformity of analytical performance between laboratories, (ii) to detect a certain population of drug users, and (iii) to exclude courtroom defenses such as passive smoking.

A lower cutoff will mean a longer detection time and cutoffs are usually set to detect a certain proportion of users within that population. Separate cutoff tables exist in standards/guidelines for both screening and confirmation. As an example, the screening cutoff concentrations for three guidelines/standards are listed in Table 1:

The response of the immunoassay varies with concentration of drug/metabolite; hence, the assay (or instrument) is set to record the presence of drug/metabolite above the response expected for the cutoff (calibrating the instrument).

Immunoassays provide a fast, cheap analysis and can rapidly eliminate the negative samples from the presumed positive samples. They can, with reasonable accuracy, identify samples within a group, which require a more probing analysis by a technique

Table 1 Screening cutoff concentrations ($ng\,ml^{-1}$)

	United States	Australia / New Zealand	United Kingdom
Amphetamines	1000	300	300
Benzodiazepines	N/A	200	200
Cannabis metabolites	50	50	50
Cocaine metabolites	300	300	300
Opiates	2000	300	300
Phencyclidine (PCP)	25	N/A	25

that is more accurate i.e., MS (see below). However, immunoassays are not entirely accurate when it comes to determining concentration (lack of sensitivity) and sometimes other closely, structurally related compounds can cause interference. This is called *cross-reactivity* and the best example would be codeine producing positive immunoassays for antibodies raised to morphine. This problem of immunoassays is called *specificity* and can cause potential misidentifications where drugs that can be obtained "over the counter" can be confused with illicit drugs. The lack of sensitivity and specificity in immunoassays also leads to false positives and false negatives.

A false positive occurs when the immunoassay produced a positive result, but the confirmation step found the sample to be negative. The failure to confirm does not mean necessarily that there was no drug/metabolite present in the sample as it may simply mean that the drug/metabolite concentration was below the cutoff (a sensitivity problem) or there is a cross-reacting substance present (a specificity problem). A false negative occurs when the immunoassay produces a negative result, but had it gone to the confirmation step, it would have been positive.

Mass Spectral Analysis (MS)

The ability to detect, identify, and determine the concentration of a drug/metabolite in a specimen by MS consists of several steps:

1. the extraction of the substance from the specimen by the use of a solvent and preparation, if necessary, of the extract for chromatography;
2. the chromatographic separation itself; and
3. the detection by MS after the separation.

Chromatography is a separation technique that uses the molecule's physiochemical properties as the basis of separation with a liquid or solid matrix in a column. Under a given set of conditions, the time to detection (retention time) is characteristic of the molecule. The MS is set to detect ions derived from the analytes. Often, the MS is only monitoring three major ions from the analyte.

The most commonly encountered ion is nominally given an abundance of 100% and all other ions are given percentages to this base ion. The MS, therefore, detects ions with definite mass-to-charge ratios and which have a definite relationship to one another in regard to how commonly they are encountered (their abundances).

MS not only unequivocally identifies the substances present but also calculates the concentration of a specific molecule or molecules of interest in the specimen (quantitation). This is achieved by preparing a number of standards of known concentration, which are measured in the same way as the unknown specimen. A calibration curve is produced and the response obtained from the specimen is used to calculate the concentration of the specimen from this response curve.

For example, if a specimen comes to the laboratory and is tested by an immunoassay for various classes of drugs and returns a positive test to the amphetamine class assay, this specimen is then subjected to confirmatory analysis for one or more of the amphetamine drugs such as amphetamine, methylamphetamine, or methylenedioxymethylamphetamine (MDMA, Ecstasy). The laboratory may only target these three amphetamines or include a number of other amphetamines known to be used in the jurisdiction. The MS is set up to monitor at least three ions for each drug. In this mode, the MS is said to be working in selected ion monitoring mode (SIM). On analysis, the ions monitored for each drug must be coincident and have the correct relative abundances and the expected retention time. The data from the standards allows the production of response curves for the separate drugs and using the data from the specimen, the computer can calculate the concentrations of any drugs found to be present. *See also* **Confirmation Testing: Toxicology**.

If, for example, methylamphetamine and its metabolite amphetamine are detected, they are only reported if their concentrations exceed the confirmatory cutoffs applied to these substances.

Defeating the Process

The main methods used to defeat or obfuscate the testing procedure can be divided into substitution and adulteration of the specimen and occur at the point of collection.

Substitution

The term is self-explanatory and is the substitution of the donor's specimen by the urine of another

person, nonhuman urine, a liquid that appears to be urine or the direct addition of water to a specimen. The main defense against substitution is temperature measurement. Provided the specimen's temperature is measured within 4 min, the measured temperature of a freshly voided urine specimen is between 33 and 38 °C or 90 and 100 °F [1]. All guidelines/standards require temperature measurement (a process called *validity testing*) and should the temperature fall outside the permitted range, a second specimen can be required of the donor. It is possible under freezing conditions for temperatures to fall below the lower limit; however, this has more to do with the temperature in the collection center and ensuring that the specimen container is not at a low temperature.

Most people who attempt substitution carry the urine in a container, usually a condom, strapped to their body; however, the heat transfer using this arrangement is usually not sufficient to ensure that the required temperature range is reached. There are, however, more extreme ways of obtaining the correct range such as women who carry the condom internally (in correctional environments) or sports people who self-catheterize with another person's urine. Commercial products on the internet include a container, which can be filled with urine and which uses a battery-operated heating unit controlled by a thermostat. Freeze-dried and certified drug-free urine can be purchased on the internet for the purpose of substitution.

Animal urine or liquids that look like urine can also be used. Temperature is again the primary check that can be done; however, particularly, in the case where the specimen is not urine, it is possible to carry out an on-site creatinine test (see the section "Substitution"), which may help assess the validity of the urine specimen.

Adulteration

Adulteration can be divided into two forms: (i) substances that when ingested can alter the concentration of drugs in the urine and (ii) substances directly added to a urine specimen in order to compromise the analysis of the specimen for drugs.

Water Loading. The most common attempt at adulteration is the deliberate ingestion of large volumes of water called *water loading* with the purpose of diluting the drugs/metabolites in the urine to a point where the concentration falls below the cutoff level and, consequently, cannot be detected. If a person were to ingest a liter of water within a short period of time, their average rate of production of urine would increase from approximately 1 to about 8 ml min^{-1}, which has the effect that the drugs/metabolites previously found in 1 ml are now in 8 ml (an eightfold dilution). The restoration of a normal rate of production of urine could take up to 3 h [2]. Drinking larger volumes of water will have even more effect; however, the process is self-limiting as one cannot exceed the theoretical excretion capacity of the kidneys (approximately 20 ml min^{-1}). The ingestion of large volumes of water can be life threatening (water intoxication), but the vast majority of individuals will become violently ill before this level is reached.

Water loading can be quite successful; however, this depends on the concentration of drug/metabolite in the urine before water loading occurred and this will depend on the amount of drug ingested, how long drug elimination has been occurring, and the type of drug involved. The practice is likely to be more successful with cannabis than with other common drugs of abuse, since the concentration ranges found in cannabis are narrower than with the other drugs and, therefore, the probability of forcing the concentration below the cutoff is greater.

The most obvious method of detecting water loading is the lack of color. While vitamin B tablets can impart some color to the urine, the measurement of creatinine concentration in the urine is an easy and objective method. Creatinine is a natural product present in all mammalian urine as a product of muscle metabolism and the excretion rate of creatinine is dependent upon muscle mass, age, exercise, and diet. A creatinine concentration below 200 mg L^{-1} is associated with possible dilution, while a concentration below 50 mg L^{-1} is not considered to be human urine. There are some difficulties, however, with the 200-mg L^{-1} limit, as certain diseases, e.g., diabetes insipidus will produce low creatinine concentrations and women, particularly of small stature, can have low creatinine levels without having what is considered an abnormal creatinine level. This occurs since women have lower muscle masses than men and people of small stature will have a lower muscle mass. Creatinine levels can be measured accurately in the laboratory or, in a less-accurate fashion, by doing an on-site adulteration screen [3]. If a collector suspects water loading, a second sample can be requested.

Ingestion of an Alkalizing Agent. Sodium bicarbonate or baking soda is commonly ingested by amphetamine users to elude detection of amphetamines in their urine. Most drugs are either weak acids or weak bases and can exist in two forms, either an electrically charged water-soluble form or an electrically neutral fat-soluble form. The form that predominates at any time depends on the pH (the degree of acidity or alkalinity) of the environment they find themselves in. By ingesting sodium bicarbonate, the user raises the pH of their urine (makes it alkaline) and this favors the electrically neutral form allowing the amphetamine to reenter the blood stream by passive diffusion (see earlier) with the consequence of a much lower amphetamine concentration in the urine [4]. There are two advantages for the user: (i) to drive the concentration below the cutoff and escape detection and (ii) to prolong the action of the drug by producing higher circulating concentrations in the blood for longer periods of time. This practice is not without danger as excessive ingestion of bicarbonate can lead to life-threatening metabolic alkalosis.

Unfortunately, pH manipulation is difficult to prove as sodium bicarbonate, even if it was to fully saturate all body cells, would only raise the pH to 8.2 or slightly above the normally accepted upper pH value of urine of 8.0. Amphetamine users need to be tested carefully as they will sometimes display a positive in a group of negative tests and such a positive may be the result of not enough bicarbonate to adequately cause the concentration to drop below the cutoff. Other common drugs of abuse form highly water-soluble metabolites and are little affected by changes in pH.

Substances Added to Urine. These substances can be further subdivided into (i) household items and (ii) commercial products, available for the express purpose of defeating a drug test.

The most commonly used household substance is bleach, the purpose of which is to chemically react with drug/metabolite and change it to a substance, which cannot be detected by the analytical procedures used [5]. Bleach does this very successfully by oxidation, which is the mode of action of many of the commercial adulteration products. Owing to its smell, bleach, either liquid or solid, is usually detected. Hydrogen peroxide is also used as an oxidant but, in household strength, appears to be less than successful.

Acids or bases are sometimes used to prevent the immunoassays working, but they can easily be detected by a pH test. Detergents and liquid soaps can be detected by cloudiness or by swirling the specimen and looking for the appearance of rainbow-colored bubbles. There is a great deal of folklore involved in this area and many products have been used including iodine, eye drops, salt, detergent, and drain cleaners with varying degrees of success [6, 7].

The first point that needs to be made regarding commercial products is that the processes are almost all directed at the prevention of detection of cannabis. The first commercial products to appear were based upon household-type adulterants and included detergents and so-called herbal diuretics. Interestingly, the herbal diuretics can include vitamin B preparations and a substance called *creatine*, a body-building supplement. Owing to the close structural similarity between creatine and creatinine, and since some creatine will be transformed to creatinine, it has been added to these preparations with the belief that it will falsely elevate the creatinine level. In fact, its contribution to the creatinine level is relatively low.

The substances that follow are all added directly to the urine and are not ingested.

Glutaraldehyde, a disinfectant, was another early favorite, which is capable of irreversibly binding the enzyme used in the laboratory immunoassay, making the assay incapable of detecting the presence of a drug/metabolite. Fortunately, laboratory immunoassay instruments return an abnormally low rate of change of absorbance, which will be flagged as an invalid result [8]. Glutaraldehyde can be detected by using a "keto" stick or an on-site adulteration test; however, the test is not specific and detects substances called *ketone bodies* in the urine. Ketone bodies are present in individuals who are undiagnosed diabetics, diabetics who are not under proper control, and people who are fasting or on high-protein diets.

Inorganic nitrites such as potassium nitrite have also been used. Nitrite and all the following adulterants have one thing in common in that they are all oxidants i.e., they chemically react with the drug/metabolite to form a new substance, which is not detected by the immunoassay or the MS. The chemical reaction induced by the addition of nitrite proceeds quickly only in acid conditions; hence, it is entirely conceivable that an immunoassay will still be positive when the specimen reaches the laboratory. The interesting part occurs when the sample

is sent for confirmation in that, during the sample preparation stages, the reaction rate speeds up destroying the metabolite, usually a cannabis metabolite, which is being searched. Again, fortunately, laboratory scientists add a set amount of a substance, which is structurally similar to the metabolite being measured so that the solvent extractions for samples and standards are independent of the extraction efficiency. This added chemical is called the *internal standard* (IS) and should be seen in every sample by the MS including the negative samples and is a component of good quality control to ensure that all laboratory procedures have been adequately carried out. Since nitrite destroys the IS as well, any loss of this substance should generate alarm signals [9].

Another oxidant now commonly used is pyridinium chlorochromate (PCC), a chemical containing the metal chromium. This adulterant, which is sold in various preparations, can be more successful in the destruction of drug/metabolite than preparations containing nitrite and the immunoassay may produce a negative result [10]. If the specimen is analyzed by MS, the IS undergoes destruction in the same manner as urines containing excess nitrite. Urines to which PCC is added have a distinctive orange tinge, which, over a period of time, turns to a green hue due to changes in the form of chromium present. If on-site drug testing is carried out, the test strips often stain yellow in the presence of PCC.

A more recent adulterant is a combination of hydrogen peroxide and an enzyme that destroys the hydrogen peroxide, namely, peroxidase. Peroxide oxidizes the metabolite and the evidence for its presence is destroyed by the enzyme peroxidase [11]. Preparations containing the oxidant potassium iodate are also being sold.

The solution to the adulteration problem is to do an on-site adulteration test at the point of collection where another specimen may be demanded and before these chemicals have the opportunity to work [12].

Interpretation

Opiates

Although most guidelines/standards contain only two opiates, morphine and codeine, it is these two substances that cause the most interpretation problems to toxicologists. The difficulty relates to morphine, which can arise from four sources, the administration of heroin, morphine itself, codeine, or poppy seeds. It can, therefore, be difficult at times to decide whether the presence of morphine in a specimen is due to illicit use.

Heroin can be found in very low concentrations and for short periods of time in urine due to its short persistence time in blood, its high fat solubility favoring reabsorption by the kidneys, and its lack of stability in water. The unique metabolite 6-acetylmorphine (6-AM) appears in the urine and can be detected for an average of 4 h (using a 10-ng ml^{-1} cutoff) after a heroin administration [13]. Its confirmed presence in a urine specimen is unequivocal proof of use of heroin. However, the major metabolite morphine (mainly as glucuronides) is present for up to three days, which means that the presence of a large concentration of morphine in the absence of 6-AM may still represent heroin or morphine administration.

Codeine administration can be problematic, since over 90% of the Caucasian population and less than 50% of the Asian population produces morphine as a metabolite. Initially, the ratio of the codeine concentration to the morphine concentration highly favors codeine; however, as the time interval increases from the administration, the ratio favors morphine since codeine is eliminated at a faster rate than morphine. In the terminal stage of a codeine elimination, it is possible to only detect morphine (typically below 1000 ng ml^{-1} of morphine) and this could lead to an erroneous conclusion that heroin/morphine was administered [14].

Poppy seed ingestion is a common defense to explain the presence of morphine in a specimen. The seeds contain both morphine and lesser amounts of codeine and their ingestion can produce a positive morphine response. The amount of morphine present is dependent on how much was ingested and what strength of morphine was present. Typically, the morphine concentrations are low, i.e., less than 2000 ng ml^{-1} and the urine will be negative within 24 h [15]. There are people, however, who abuse poppy seeds and make a boiling water infusion with large amounts of the seed to make a poppy seed tea, which produces levels far above the typical levels. Interestingly, poppy seed abuse has been associated with bakers because of their ready access to large amounts of seeds [16]. Eastern Europeans also tend to use large amounts of poppy seeds in their cooking, e.g., poppy seed cake.

As a consequence, the source of low concentrations of morphine in urine cannot be interpreted. In the United States, the guidelines were modified to raise the cutoffs to alleviate the problem but other standards persist with lower cutoffs, which makes interpretation impossible at low morphine-positive concentrations.

Cannabis

Tetrahydrocannabinol (THC), the active ingredient of cannabis, is a very fat soluble substance, which will slowly accumulate in the fatty tissues of the body with chronic use of cannabis. After cessation of cannabis use, the accumulated THC will slowly diffuse out of the fat into the blood stream where it is metabolized to form water-soluble metabolites that will appear in the urine. The most common metabolite is known as *carboxy-THC*. It is the presence of this and other metabolites that produce positive cannabis tests even though the person concerned may not be recently using the drug. In comparison with other common drugs of abuse, the time taken to produce a negative result in a chronic cannabis user is quite long and can lead to the erroneous conclusion of reuse. Conversely, chronic cannabis users can rely upon this to reuse and explain a positive in terms of the long withdrawal period (*see* **Cannabis**).

When monitoring cannabis withdrawal, it is unscientific to rely upon the concentration of the major metabolite, carboxy-THC, as the concentration of any substance in the urine is dependent on the volume of urine being produced at the time. When a person does not drink sufficient water, the kidneys will decrease their production of urine by reabsorbing water and, consequently, the urine produced will be more concentrated and the concentration of carboxy-THC will therefore increase. Conversely, if water is drunk in excess of the body's requirements, the kidneys will dump more water and the urine will be dilute resulting in a lower concentration of carboxy-THC. It is therefore possible, if the individual is close to the cutoff, to have consecutive results where a negative test will turn to a positive one even though the person has not been reusing.

To solve this problem, the carboxy-THC concentration is normalized to the creatinine concentration. This is the ratio of carboxy-THC to creatinine in the urine i.e., the concentration of carboxy-THC divided by the concentration of creatinine. Since both substances coexist in the same volume of urine, the ratio is independent of that volume and hence the ratio does not depend upon a person's fluid balance. The ratio is therefore the appropriate parameter to monitor compliance with cannabis withdrawal.

If the creatinine-normalized carboxy-THC ratio increases significantly between consecutive tests, the suspicion of reuse arises. If the ratio is equal to or greater than 1.5 times the previous ratio, there is a high probability of correctly predicting new drug use [17]. The ratio should decrease as time elapses; however, an increase of greater than 50% in the ratio is generally accepted as reuse.

Perhaps the most misunderstood aspect of cannabis testing is associated with the detection period. For a casual user (once-a-week use) the detection period is 2–4 days [18]; however, it is in the detection period of a chronic user (uses at least once per day) where the myths and distortions arise. The longest recorded consecutive immunoassay positive period in scientific literature associated with cannabis withdrawal is 46 days [19]. The test used had a 20-ng ml^{-1} cutoff, which is 2.5 times more sensitive than the usual cutoff of 50 ng ml^{-1} and with no normalization of creatinine concentration.

A more recent study [20] was conducted in 1999, using prison inmates, showed that the longest period of detection for frequent heavy users at a 15-ng ml^{-1} cutoff was 17 days. The generally accepted period for the clearance of chronic users is within 21 days; however, there will be some individuals who will be positive for longer periods due to their pharmacokinetic profile, but these individuals are in the minority. Cannabis withdrawal should be monitored by taking specimens twice a week and analyzing them by MS in a laboratory using creatinine normalization. Immunoassays should not be replied upon solely due to the contribution of cross-reacting cannabinoids.

The argument that a positive urine result for cannabis is the result of passive inhalation rather than active smoking is frequently encountered. The passive inhalation of cannabis has been extensively studied. The studies concluded that while it was possible to produce a positive result, the conditions under which this occurs are quite unrealistic [21].

On-Site Immunoassays

Increasingly, as an option, the initial immunoassay test is being carried out on-site (at point of collection) rather than in the laboratory. This has several advantages, particularly, in that it removes 95%+ of samples in workplace testing that need to be sent to a laboratory and it provides a status report on an individual's risk profile to the management, allowing them to make a determination on the work status of the individual. The former advantage decreases the amount of paperwork required, which, in turn, decreases the risk of mistakes as well and, furthermore, decreases the costs of transportation. The latter advantage allows a company to discharge its obligation in relation to a safe working environment as required under legislation or work orders. Another advantage, particularly in a custodial or a community corrections environment, is the higher admission rates of drug taking by individuals when they actually witness the result of the test.

It would be unrealistic to expect that an on-site device does not have disadvantages. This technology uses a lateral flow immunoassay (similar to a home pregnancy test in technology) and, if used properly, has almost the same accuracy as a dedicated laboratory instrument. However, equivocal results are more likely if the true result lies close to the cutoff. Fortunately, there are few specimen results that are close to the cutoff. Equivocal results are due to the slightly different sensitivities of tests conducted on-site and in the laboratory.

Lack of specificity in any type of immunoassay can only be addressed by changing the antibody used in order to lower cross-reactivity. False negatives, as they are not apparent, can only be assessed by analyzing a number of immunoassay-negative samples in the laboratory by MS.

Conclusions

Urine testing using accepted international standards is a well established, legally defensible, comparatively cheap, mature technology, which is capable particularly by use of on-site testing of providing timely results. The specimen can be collected in high volumes in a medically noninvasive manner and when it comes to obtaining drug-taking history in terms of days, it remains the specimen of choice. It is ideal for compliance and random workplace testing as an alternative to expensive laboratory tests, but it is not the specimen of choice for very recent usage. As with all specimen types, it has its strengths and weaknesses but, if conducted and interpreted properly, it is unlikely to be seriously challenged as the most popular means of drugs-of-abuse testing in the near future.

References

[1] Judson, B.A., Himmelberger, D.U. & Goldstein, A. (1979). Measurement of urine temperature as an alternative to observed urination in a narcotic treatment program, *The American Journal of Drug and Alcohol Abuse* **6**(2), 197–205.

[2] Baldes, E.J. & Smirk, F.H. (1934). The effect of water drinking, mineral starvation and salt administration on the total osmotic pressure of the blood in man, chiefly in relation to the problems of water absorption and water diuresis, *The Journal of Physiology* **82**(1), 62–74.

[3] Cook, J.D., Caplan, Y.H., LoDico, C.P. & Bush, D.M. (2000). The characterization of human urine for specimen validity determination in workplace drug testing: a review, *Journal of Analytical Toxicology* **24**, 579–588.

[4] Beckett, A.H., Rowland, M. & Turner, P. (1965). Influence of urinary pH on excretion of amphetamine, *Lancet* **1**, 303.

[5] Baiker, C., Serrano, L. & Lindner, B. (1994). Hypochlorite adulteration of urine causing decreased concentration of Δ^9-THC-COOH by GC/MS, *Journal of Analytical Toxicology* **18**, 101–103.

[6] Mikkelsen, S.L. & Ash, K.O. (1988). Adulterants causing false negatives in illicit drug testing, *Clinical Chemistry* **34**(11), 2333–2336.

[7] Warner, A. (1989). Interference of common household chemicals in immunoassay methods for drugs of abuse, *Clinical Chemistry* **35**(4), 648–651.

[8] George, S. & Braithwaite, R.A. (1996). The effect of glutaraldehyde adulteration of urine specimens on syva EMIT II drugs-of-abuse assays, *Journal of Analytical Toxicology* **20**, 195–196.

[9] Urry, F.M., Komaromy-Hiller, G., Staley, B., Crockett, D.K., Kushnir, M., Nelson, G. & Struempler, R.E. (1998). Nitrite adulteration of workplace urine drug testing specimens. I. Sources and associated concentrations of nitrite in urine and distinction between natural sources and adulteration, *Journal of Analytical Toxicology* **22**(2), 89–95.

[10] Wu, A.H., Bristol, B., Sexton, K., Cassella-McLane, G., Holtman, V. & Hill, D.W. (1999). Adulteration of urine by "urine luck", *Clinical Chemistry* **45**(7), 1051–1057.

[11] Cody, J.T. & Valtier, S. (2001). Effects of stealth™ adulterant on immunoassay testing for drugs of abuse, *Journal of Analytical Toxicology* **25**, 466–470.

[12] Peace, M.R. & Tarnai, L.D. (2002). Performance evaluation of three on-site adulterant detection devices for urine specimens, *Journal of Analytical Toxicology* **26**(7), 464–470.
[13] Cone, E.J., Welch, P., Mitchell, J.M. & Paul, B.D. (1991). Forensic drug testing for opiates: I. Detection of 6-acetylmorphine in urine as an indicator of recent heroin exposure; drug and assay considerations and detection times, *Journal of Analytical Toxicology* **15**, 1–7.
[14] Lafolie, P., Beck, O., Lin, Z., Albertioni, F. & Boreus, L. (1996). Urine and plasma pharmacokinetics of codeine in healthy volunteers: implications for drugs-of-abuse testing, *Journal of Analytical Toxicology* **20**, 541–546.
[15] Meadway, C., George, S. & Braithwaite, R. (1998). Opiate concentrations following the ingestion of poppy seed products – evidence for "the poppy seed defence", *Forensic Science International* **96**, 29–38.
[16] King, M.A., McDonough, M.A., Drummer, O.H. & Berkovic, S.F. (1997). Poppy tea and the baker's first seizure, *Lancet* **350**(9079), 715.
[17] Huestis, M.A. & Cone, E.J. (1998). Differentiating new marijuana use from residual drug excretion in occasional marijuana users, *Journal of Analytical Toxicology* **22**, 445–454.
[18] Huestis, M.A., Mitchell, J.M. & Cone, E.J. (1995). Detection times of marijuana metabolites in urine by immunoassay and GC-MS, *Journal of Analytical Toxicology* **19**, 443–449.
[19] Ellis, G.M., Mann, M.A., Judson, B.A., Schramm, M.T. & Tashchian, A. (1985). Excretion patterns of cannabinoid metabolites after last use in a group of chronic users, *Clinical Pharmacology and Therapeutics* **38**, 572–578.
[20] Smith-Kielland, A., Skuterud, B. & Morland, J. (1999). Urinary excretion of 11-nor-9-carboxy-delta9-tetra hydrocannabinol and cannabinoids in frequent and infrequent drug users, *Journal of Analytical Toxicology* **23**, 323–332.
[21] Cone, E.J., Johnson, R.E., Darwin, W.D., Yousefnejad, D., Mell, L.D., Paul, B.D. & Mitchell, J. (1987). Passive inhalation of marijuana smoke: urinalysis and room air levels of delta-9-tetrahydrocannabinol, *Journal of Analytical Toxicology* **11**, 89–96.

WARREN D. HAMILTON

Drug Tolerance and Dependency

see Alcohol: Use, Abuse, Tolerance, and Dependency

Drug-Facilitated Sexual Assault

Introduction

When a drug's pharmacological effect on a person causes incapacitation and that person then becomes the victim of a criminal act, it is often referred to as a *drug-facilitated crime* (*DFC*). While it is generally thought that the drug must be secretly administered to a person for it to be considered a DFC, any time a person commits a crime by taking advantage of the debilitating effects that a drug (or combination of drugs) has on another individual (whether that individual was surreptitiously administered the drug or voluntarily took it), a DFC has occurred. The key is that the drug(s) must assist the criminal action. Most often when we consider DFCs, we speak of drug-facilitated sexual assault (DFSA), but robberies, homicides, and drug-smuggling have also been facilitated by taking advantage of the pharmacological effects of drugs and, in essence, using the drugs as "weapons".

The prevalence of DFCs will never be fully recognized. That is because a common challenge with these cases is reporting of the crime. In many cases, the victim suffers from amnesia. In other instances, the victims of DFCs may be uncertain of sequences of events and facts, so they may delay reporting in order to try to piece together the information that they do remember. Any delay makes it much more difficult for the toxicology laboratory to identify any incapacitating agents that may have been used.

Drugs Used to Commit DFC

There are over 50 drugs known or suspected to have been used to commit DFCs. Most are fast-acting strong central nervous system (CNS) depressants that mimic ethanol intoxication. They generally cause relaxation, euphoria, and decreased inhibitions at low doses; amnesia, impaired perceptions, difficulties in walking and maintaining balance, impaired speech, and drowsiness at moderate doses; and complete loss of motor functions, vomiting, incontinence, unconsciousness, and possible death at the highest doses. At the higher doses, the effects are likened to those of general anesthetic agents. In fact, many of the drugs

that have been used are currently, or once were, used as anesthetics. Because the CNS depressant effects of these drugs generally imitate one another, it is highly unlikely that one can determine the drug used in a DFSA case simply by symptoms alone. Instead, it is vital that a thorough, sensitive toxicological analysis be performed on the best available evidence following standard forensic practices in order to improve the laboratory's chance of identifying any likely incapacitating substances hours, days, or weeks after that drug was ingested.

Many of these drugs are well known as *recreational drugs of abuse, prescription medications*, or *over-the-counter pharmaceuticals*. While surreptitious administration does occur in DFCs, in most cases the drugs are probably self-administered by the eventual victim who may not be fully aware of the effects of the drug, particularly when co-ingesting it with other strong CNS depressants. Further, it requires some preparation on the part of the criminal in order to slip a drug into the drink of a crime victim. Most drugs are in a solid, tablet formulation. Simply dropping a pill into a drink is usually not an efficient means of secretly giving the drug to someone, because the pill will not likely dissolve immediately and may fizz as it begins to dissipate into the beverage. Additionally, since most tablets contain insoluble, cellulose-based fillers, a portion of the tablet may not completely dissolve into the drink and will instead leave a grainy residue. For that reason, it is common to see pills predissolved into small volumes of ethanol-based solutions, filtered to remove the insoluble materials, and then transferred into small eyedropper-type bottles. Additionally, some drugs such as gamma hydroxybutyrate (GHB) and drugs found in gelatin capsules are already in a liquid form, thus minimizing the preparation that is required for tablets.

Ethanol

By far, ethanol (or alcohol) is the most common "drug" used to facilitate crimes. This is due to its acceptance as a social drug throughout most of the world. In most DFCs in which ethanol plays a primary or contributory role, the alcohol ingestion is voluntary; however, surreptitious administration of ethanol has been known to occur.

There are numerous factors that affect the level of intoxication reached from drinking ethanol. Some variables are easily controlled such as the type of alcoholic beverage consumed (beer, wine, or mixed-drinks), the size of the beverages, and whether ethanol is consumed with or after eating a meal as opposed to an empty stomach. All of these variables have an effect on ethanol's absorption into the bloodstream. The faster ethanol is absorbed and distributed throughout the body (including the brain), the more intoxicated the user generally becomes. Additionally, as with all CNS drugs, the more ethanol that is consumed the more severe the intoxication will become.

Because ethanol impairment has been so extensively studied in relation to operating motor vehicles, a lot is known about its behavioral effects on people. Users generally pass through different phases of intoxication beginning with a subclinical stage where there are no appreciable symptoms of intoxication. The next phase is one of euphoria where the user becomes more talkative and social as they experience diminished inhibitions. This is followed by a stage of excitement where they may begin to feel drowsiness, have indications of emotional instability, and impaired perceptions, memory, and comprehension. The more serious stages of confusion, stupor, coma, and death follow. Through these later stages, the user may experience severe disorientation, lack of coordination, slurred speech, vomiting, blackouts, unconsciousness, anesthesia, and respiratory depression.

Ethanol is generally eliminated following zero-order kinetics. This allows for the average elimination rates for both men ($0.015\,g\,dl^{-1}h^{-1}$) and women ($0.018\,g\,dl^{-1}\ h^{-1}$) to be used to predict the blood ethanol concentration of an individual based on the number, type, and size of the drinks that they consumed in a given evening. This is a useful exercise to evaluate the role that ethanol may have played in contributing to the symptoms that the DFC victim described. However, factors such as chronic alcoholism, genetic factors, and liver disease must also be considered because they can affect these elimination rates (*see* **Alcohol**).

GHB and Analogs

Of all the drugs used to facilitate sexual assault and other crimes, GHB and its metabolic precursors, gamma butyrolactone (GBL) and 1,4-butanediol (1,4-BD), are probably among the most favored. Unfortunately, this statement can only be based on circumstantial evidence. Solid proof of such is difficult to

obtain because GHB, GBL, and 1,4-BD have strong sedative, amnesiac effects and, more importantly, are very rapidly eliminated after ingestion. To further complicate the matter, GHB is a metabolite of gamma-aminobutyric acid (GABA), a natural neurotransmitter in humans. Thus measurable amounts of endogenous GHB are found in blood ($<2\,mg\,l^{-1}$) and urine specimens ($<10\,mg\,l^{-1}$) of most humans. Therefore, after ingestion, GHB is not likely to be discriminated from endogenous concentrations of GHB after 2–8 h in blood specimens or 6–12 h in urine samples.

GHB, GBL, and 1,4-BD are also very popular "club drugs". Thus, it is also common for victims of DFCs to have voluntarily ingested the drug, unaware that doing so would make them vulnerable.

Individuals exposed to GHB or one of its analogs can pass from a completely alert state to deep unconsciousness within 10–15 min after ingestion. Because of GHB's rapid clearance, the GHB-assisted sleep generally only lasts about 3–5 h, after which the individual awakes and feels unusually refreshed.

GHB, GBL, and 1,4-BD are nearly always administered as an oral solution. They are rapidly absorbed with peak plasma concentrations occurring within 30–60 min. GHB's absorption is likely to be capacity limited and may be enhanced when it is absorbed on an empty stomach.

GBL and 1,4-BD are rapidly metabolized into GHB (Figure 1). Thus their pharmacological effects mimic those of GHB. GHB is also rapidly and extensively metabolized such that less than 5% of an oral dose is excreted unchanged in the urine.

Benzodiazepines

As a drug class, the benzodiazepines are likely to be the most commonly used prescription medications used to commit DFCs. Benzodiazepines are one of the world's most widely prescribed drug classes primarily as anxiolytics, anticonvulsants, anesthetic adjuncts, hypnotics, and muscle relaxants, and for the treatment of obsessive-compulsive disorders. They are also very popular drugs to use in DFCs. This fact demonstrates the important role that drug availability has in DFCs. The perpetrator of DFCs often has personal experience with the drug effects. This may be through prescription or occupational use, but it is also very common to see benzodiazepines recreationally abused in combination with illicit drugs such as opiates and cocaine.

Benzodiazepines are administered orally, intravenously, or intramuscularly. With oral benzodiazepine ingestion, absorption into the bloodstream is nearly 100% because of their very high lipid solubility. However, the rate of absorption (and thus the rate of activity) is dependent upon the benzodiazepine itself. Generally speaking, most benzodiazepines reach their peak blood concentrations within 30 min to 6 h after ingestion. As discussed earlier, if the tablet formulation of a benzodiazepine is dissolved in an alcohol or aqueous solution to facilitate its surreptitious administration, the absorption rates should be faster and peak effects should be felt significantly earlier than if exposure occurs via a normally administered tablet.

Benzodiazepines are classified according to their elimination half-lives (Table 1). Obviously, the short-acting benzodiazepines (e.g., midazolam and triazolam) are detected for much shorter periods than long-acting benzodiazepines (e.g., diazepam, clonazepam). Additionally, the long-acting benzodiazepines tend to exhibit a longer and more severe "hangover" side effect than do their short-acting counterparts.

Metabolism of benzodiazepines is extensive by the liver via dealkylation, deamination, hydroxylation, or reduction. The hydroxyl products are further metabolized by conjugation with glucuronic acid and are the major urinary products of benzodiazepines. The cytochrome P450 3A (CYP3A) enzyme family is involved in the metabolism of many benzodiazepines, so it is important to recognize the effects of coingestion of drugs that inhibit this enzyme system

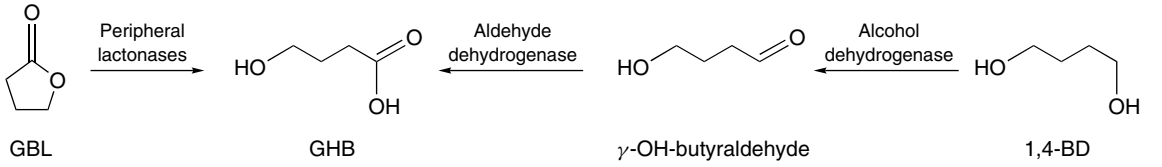


Figure 1 GHB, GBL, and 1,4-BD metabolism

Table 1 Benzodiazepines and their elimination half-lives

Benzodiazepine	Examples of trade name(s)	Half-life
Alprazolam	Xanax	Intermediate
Bromazepam	Lectopam, Lexotan	Intermediate
Chlordiazepoxide	Librium	Intermediate
Clonazepam	Clonopin, Klonopin	Long
Clorazepate	Tranxene	Short
Diazepam	Valium, Ducene	Long
Estazolam	ProSom, Eurodin	Intermediate
Flunitrazepam	Rohypnol, Hypnodorm	Long
Flurazepam	Dalmane	Long[(a)]
Halazepam	Paxipam	Long[(b)]
Lorazepam	Ativan	Intermediate
Midazolam	Versed	Short
Nitrazepam	Mogadon	Long
Oxazepam	Serax, Serepax	Intermediate
Prazepam	Centrax, Vertran	Short
Temazepam	Normison, Restoril	Intermediate
Triazolam	Halcion	Short

[(a)] Flurazepam has a very short elimination half-life, but its metabolite's half-life is long

[(b)] Halazepam has an intermediate elimination half-life, but the half-life of its metabolite is long

and their effects on benzodiazepine metabolism. For example, cimetidine, diltiazem, fluoxetine, verapamil, antifungals, antibiotics (e.g., erythromycin and clarythromycin), and protease inhibitors are all known to inhibit the CYP3A enzymes and potentially others. Grapefruit juice has also been shown to inhibit these enzymes. If these enzymes are inhibited, normal metabolism of the benzodiazepines will be altered and their resulting pharmacological effects may be enhanced (*see* **Benzodiazepines**).

Marijuana

Marijuana (or cannabis) continues to be the most widely abused illicit drug worldwide. It is typically self-administered through smoking, although there are numerous cases of its surreptitious administration to individuals in food.

The primary psychoactive analyte of marijuana is Δ^9-tetrahydrocannabinol (THC). Variability in smoking characteristics (i.e., number, duration, and spacing of puffs; hold time; and inhalation volume) strongly influence the concentration of THC that a person is exposed to.

THC can be measured in plasma within seconds after inhalation of the first puff of a marijuana cigarette. Therefore, the brain is rapidly exposed to THC following marijuana smoking. While the behavioral effects of euphoria and relaxation are generally the desired effects from marijuana use, an acute toxic effect of marijuana is CNS depression (e.g., lethargy and reduction in short-term memory). This CNS depression is enhanced when ethanol is coingested with marijuana.

The major metabolic route of THC is via hydroxylation by the hepatic cytochrome P450 enzyme system. This results in the production of the active metabolite, 11-OH-THC, which is oxidized to an inactive metabolite, THC-COOH. THC-COOH forms a conjugate with glucuronic acid, which enhances the compound's water solubility. THC-COOH, in both its free and conjugated states, is the primary metabolite seen in urine samples.

Owing to its high lipid solubility, THC will accumulate in fat where it will slowly release back into the bloodstream. This allows for extended detection times for THC and metabolites when laboratories are utilizing very sensitive analytical methods.

Generally speaking, THC-COOH excretion can occur for days after ingestion. This long elimination half-life, as well as the enterohepatic recirculation of THC, makes it very difficult to estimate the time of drug exposure when analyzing a urine sample. It

becomes even more difficult when the urine sample is from a chronic marijuana user. With some exceptions, a urine sample from a naïve user of marijuana will have measurable amounts of THC metabolites for one to seven days after smoking a single marijuana cigarette, depending upon the sensitivity of the analytical methods used to measure these metabolites (*see* **Cannabis**).

Ketamine

Ketamine is a dissociative anesthetic that is structurally similar to the more recognized recreational drug, phencyclidine (PCP). Like GHB, ketamine has also become a trendy drug of abuse and is considered a "club drug". Most users abuse ketamine for its hallucinogenic properties. Because of ketamine's popularity in veterinarian medicine, there have been numerous reports of its theft from veterinarian clinics.

Ketamine can be administered through intravenous, subcutaneous, intramuscular injections, nasal insufflation, smoking, or oral ingestion. In addition to hallucinations, it is known to produce irrational behavior and strong CNS depression resulting in impaired speech, amnesia, and "out-of-body" experiences. A particularly interesting effect of this drug is that thought processes are impaired while under its influence. It has been reported that an individual's memories of events while under the influence of this drug can be altered or shaped by an outsider telling the person what is happening to them. While an individual is under the influence of this drug, they may appear to be awake, but in a trancelike state.

Ketamine is metabolized to norketamine and dehydronorketamine. Conjugates of hydroxylated derivatives of the parent drug, as well as its metabolites, are excreted in the urine for at least 72 h after ingestion.

MDMA

Unlike most of the other drugs used to facilitated crimes, 3,4-methylenedioxymeth-amphetamine (MDMA) has more hallucinogenic and stimulant-type properties instead of strong CNS depressant effects. The desired pharmacological effect for users of MDMA, however, is its ability to heighten the user's senses. While taste, hearing, seeing, and smelling are all enhanced, the pleasure associated with touch while under MDMA's influence can increase the likelihood of sexual contact. At very high doses, symptoms may also include nausea and vomiting, profuse sweating, teeth grinding, and, on rare occasions, coma and hypotension.

MDMA is typically administered orally and is rapidly metabolized to numerous metabolites, including a related drug; methylenedioxyamphetamine (MDA). Using sensitive analytical techniques, MDMA and MDA may be detected in the urine for one to four days after oral administration (*see* **Amphetamine**).

Opiates

Like benzodiazepines, opiates represent another large drug class used to facilitate crimes (Table 2). Besides the strong CNS depression caused by opiates, they also act as, and are generally prescribed as, analgesics. The analgesic effect is believed to be particularly important in DFSA cases. Many DFSA victims report briefly waking during their assault, looking up and seeing the perpetrator assaulting them, but falling back into an unconscious state before they can fight off their attacker. These awakenings are believed to occur because of a sudden infliction of pain such as when penetration occurs, or if their head or arm is suddenly injured by banging against the floor, a headboard, or a nightstand. The analgesic effects of the opiates minimize the likelihood of this awakening occurring due to such sudden pain. However, it should be noted that the effectiveness of individual opiates will vary from one opiate to the next and from one person to the next.

Most opiates are best absorbed via parenteral administration. Gastrointestinal absorption is also generally effective, but the extent of the pharmacological effects varies depending upon the degree of first-pass metabolism of the individual opiates. Many of the opiates ultimately form glucuronide metabolites that are primarily eliminated in the urine (*see* **Opioids**).

Antihistamines

In recent years, there has been an increase in the abuse of some of the antihistamine drugs. This abuse is in large part due to their availability and the strong CNS depressant effects of some of these drugs. In particular, diphenhydramine, brompheniramine, chlorpheniramine, and doxylamine have the potential for strong sedative effects. These effects are variable

Table 2 Narcotic and nonnarcotic analgesics

Drug	Examples of trade name(s)	Half-life
Buprenorphine	Buprenex, Subutex	Short
Butorphanol	Stadol	Intermediate
Codeine	Codral forte	Short
Dextromethorphan	DM, DXM	Short
Dihydrocodeine	DHCplus, Synalgos-DC	Short
Fentanyl	Actiq, Duragesic, Sublimaze, Innovar	Short-intermediate
Heroin		Short
Hydrocodone	Anexsia, Hycodan, Lorcet, Lortab, Norco, Pancet, Vicodin, Zydone	Short
Hydromorphone	Dilaudad, Palladone	Short
Levorphanol	Dromoran, Levo-Dromoran	Intermediate
Meperidine (Pethidine)	Demerol, Mepergan	Short
Methadone	Dolophine, Physeptone	Long
Morphine	Avinza, Astramorph, Duramorph, Kadian, MS Contin, Oramorph, Roxanol	Short
Nalbuphine	Nubain	Short–intermediate
Oxycodone	Oxycontin, Oxyir, Roxicodone, Percodan, Percocet, Percolone, Roxicet, Tylox	Short
Oxymorphone	Numorphan	Short
Pentazocine	Talacen, Talwin	Short
Propoxyphene	Darvocet, Darvon, Wygesic	Intermediate

from one user to the next. In fact, children may actually exhibit a paradoxical effect in which these antihistamines act as CNS stimulants rather than depressants.

Antihistamines are readily absorbed after oral administration with the onset of action generally occurring within an hour of ingestion. The half-life of antihistamines vary considerably so detection times after administration may range from hours to days, depending on the particular antihistamine consumed.

Other drugs

There have been many other drugs used to facilitate crimes. Most recently, the drugs zolpidem and zopiclone have increased in popularity in DFCs, concurrent with their increased prevalence as prescription medications throughout the world. A number of centrally acting muscle relaxants such as cyclobenzaprine and meprobamate have also been observed in DFCs, as have barbiturates, some antipsychotics (e.g., clozapine, thioridazine, and chlorpromazine), and the sedative tricyclic antidepressants (e.g., amitriptyline and desipramine). While not technically a drug, solvents have also been used to incapacitate a person in order to commit a crime against them. The original "Mickey Finn", chloral hydrate, is still available in some countries and used to knock out victims of DFCs. As new sedative medications continue to emerge, the list of drugs that have or can be used to facilitate a crime will continue to expand.

Toxicological Specimens for DFC Investigations

It is generally accepted that urine is the most useful specimen in the majority of investigations of DFCs. Since drugs and metabolites are concentrated in urine specimens, and thus are more readily detectable, the value of urine specimens is clear when trying to determine if an individual was exposed to a drug days before the specimen was collected. Current recommendations by the Society of Forensic Toxicologists (SOFT) DFSA Committee are that urine specimens be collected as soon as possible after a DFC, not to exceed 96 h after the suspected drug exposure. If possible, 100 ml of urine should be collected to ensure that enough of the specimen is available for the laboratory to perform a sensitive and thorough analysis.

After ingestion, most drugs are below detectable levels in the blood within 24 h, therefore limiting the usefulness of this specimen. In those cases in which blood can be collected a short time after drug ingestion, the combination of blood and urine specimens can provide a clearer picture as to the window of exposure to the drug. Blood specimens should be placed into collection tubes containing sodium fluoride. In addition to urine, at least 7–10 ml of blood should be collected when it can be provided within 24 h of the suspected drug exposure.

Oftentimes a victim of a DFC does not present to medical and/or law enforcement personnel until weeks (or longer) after the alleged crime. At this point, it is no longer possible to find evidence of a drug in a blood or urine specimen. In these cases, hair specimens have shown promise as a specimen of choice; particularly when newer analytical instrumentation with superior sensitivity is used. However, hair does present additional challenges compared to blood and urine specimens. Thorough studies of the incorporation of all of the drugs that can be used for DFC into hair have not been conducted. This makes a negative result in hair very difficult to interpret.

A positive hair result can also be challenging. Head hair grows at a rate of approximately 1 cm per month. Segmental analysis of hair is necessary to pinpoint the portion of hair that corresponds to the growth period during the alleged drug exposure and to demonstrate that the drug was not ingested outside of that growth period. For example, if a month before an alleged DFC, the victim took a strong sedative, the hair test must not confuse that ingestion with any drugs ingested the night of the DFC. Only laboratory personnel skilled with handling hair specimens in this matter should be relied upon to attempt segmental analyses so as to not cause inaccurate conclusions (*see* **Hair: Toxicology**).

In some cases, another useful specimen may be vomit from the victim. Most drugs are capable of causing nausea and vomiting to some degree. If a drug is not fully absorbed before vomiting occurs, the drug may be readily detected at relatively high amounts in a vomit stain. Both ethanol and GHB, two of the most common drugs used for DFCs, are known to cause vomiting from their ingestion. Unfortunately, vomit is not generally useful for these two drugs because of the volatility of ethanol and the natural presence of GHB in humans. Most other drugs may potentially be detected in vomit stains.

Toxicological Analyses in DFCs

Over the years, specimens collected in investigations of DFC have at times been completely consumed or prematurely discarded because the specimens were sent to laboratories that did not have appropriate methods or instrumentation to adequately carry out the toxicological analyses. In general, clinical laboratories may not be able to detect subtherapeutic concentrations of these drugs. Therefore, these laboratories should not be relied upon to determine if an individual was exposed to a drug when the specimens are collected more than a few hours after the suspected drug exposure. Forensic toxicology laboratories may not be able to provide much better service, unless they have taken the steps needed to improve the sensitivities of their methods.

To improve consistency in results provided by laboratories, the SOFT DFSA Committee developed a chart of the most prevalent drugs associated with DFCs and their recommended maximum detection limits when analyzing urine specimens (Table 3). These recommended detection limits are based on published analytical methods using standard laboratory instrumentation (e.g., gas chromatography mass spectrometry (GC-MS) or liquid chromatography mass spectrometry (LC-MS)). The group's goal in developing this document was to encourage laboratories to evaluate their current capabilities and make improvements, as necessary. Furthermore, this tool provides a means of simplifying communication between analytical toxicologists and their customers, usually law enforcement personnel.

There are a number of approaches that a laboratory may take to reach the SOFT DFSA Committee maximum detection limits. Laboratories may choose to increase the specimen volume they would typically use to analyze for these drugs in their more routine cases. They should also consider a hydrolysis step in their procedures for drugs that form conjugated metabolites (e.g., benzodiazepines, opiates, etc.). Derivatization of polar metabolites combined with analysis using selective instrumental detectors (i.e., gas chromatography electron-capture detector (GC-ECD), GC-MS with negative ion chemical ionization, etc.) may also improve detection limits. Technological advances lead to improved instrumental sensitivity, so many laboratories can improve sensitivity simply by ensuring that they are using instrumentation that is less than five years old. There

Table 3 Society of Forensic Toxicologists (SOFT) DFSA recommended maximum detection limits for urine specimens

Target drug/metabolite	Recommended maximum detection limit
Ethanol	10 mg dl^{-1}
GHB and analogs	10 mg l^{-1}
Benzodiazepines	
Alprazolam	10 μg l^{-1}
Chlordiazepoxide	
Diazepam	
Hydroxyalprazolam	
Lorazepam	
Nordiazepam	
Oxazepam	
Temazepam	
7-Amino-clonazepam	5 μg l^{-1}
7-Amino-flunitrazepam	
Clonazepam	
Flunitrazepam	
THC-COOH	10 μg l^{-1}
Barbiturates	
Amobarbital	25 μg l^{-1}
Butalbital	
Pentobarbital	
Phenobarbital	
Secobarbital	20 μg l^{-1}
Antidepressants	
Amitriptyline	10 μg l^{-1}
Citalopram	
Desipramine	
Desmethylcitalopram	
Desmethyldoxepin	
Doxepin	
Fluoxetine	
Imipramine	
Norfluoxetine	
Norsertraline	
Paroxetine	
Sertraline	
Over-the-counter medications	
Brompheniramine	10 μg l^{-1}
Chlorpheniramine	
Desmethylbrompheniramine	
Desmethylchlorpheniramine	
Desmethyldoxylamine	
Dextromethorphan	
Diphenhydramine	
Doxylamine	
Opiates and nonnarcotic analgesics	
Codeine	10 μg l^{-1}
Hydrocodone	
Hydromorphone	
Meperidine	
Methadone	
Methadone Metab (EDDP)	

(*continued overleaf*)

Table 3 (*Continued*)

Target drug/metabolite	Recommended maximum detection limit
Morphine	
Normeperidine	
Norpropoxyphene	
Oxycodone	
Propoxyphene	
Miscellaneous drugs	
Carisoprodol	$50\,\mu g\,l^{-1}$
Meprobamate+	
Valproic acid	
Cyclobenzaprine	$10\,\mu g\,l^{-1}$
MDA	
MDMA	
PCP	
Scopolamine	
Zolpidem	
Clonidine	$1\,\mu g\,l^{-1}$
Ketamine	
Norketamine	

$mg\,dl^{-1}$, milligram per deciliter $mg\,l^{-1}$, milligram per liter $\mu g\,l^{-1}$, microgram per liter

is no single approach to ensure that all of the drugs that may be used to facilitate DFCs are analyzed at adequate sensitivity, but a combination of the above tactics greatly improve the laboratories chances of adequately screening for these drugs.

Conclusions

While DFCs have probably occurred for hundreds of years or longer, in recent times a concerted effort by law enforcement, medical professionals, the media, and toxicologists have raised the public's awareness of these crimes. While the true prevalence of these crimes will never be fully recognized, acknowledging the many challenges that come with these cases also provides insight as to how to improve chances of successfully investigating allegations of a DFC.

Further Reading

Elian, A.A. (2000). A novel method for GHB detection in urine and its application in drug-facilitated sexual assaults, *Forensic Science International* **109**, 183–187.

Elsohly, M.A. (2001). Drug-facilitated sexual assault, *Southern Medical Journal* **94**, 655–656.

LeBeau, M., Andollo, W., Hearn, W.L., Baselt, R., Cone, E., Finkle, B., Fraser, D., Jenkins, A., Mayer, J., Negrusz, A., Poklis, A., Walls, H.C., Raymon, L., Robertson, M. & Saady, J. (1999). Recommendations for toxicological investigations of drug-facilitated sexual assaults, *Journal of Forensic Science* **44**, 227–230.

LeBeau, M.A. & Mozayani, A. (2001). *Drug-facilitated Sexual Assault: A Forensic Handbook*, Academic Press, San Diego.

Ledray, L.E. (2001). The clinical care and documentation for victims of drug-facilitated sexual assault, *Journal of Emergency Nursing* **27**, 301–305.

McGregor, M.J., Lipowska, M., Shah, S., Du, M.J. & De Siato, C. (2003). An exploratory analysis of suspected drug-facilitated sexual assault seen in a hospital emergency department, *Women and Health* **37**, 71–80.

Schwartz, R.H., Milteer, R. & LeBeau, M.A. (2000). Drug-facilitated sexual assault ('date rape'), *Southern Medical Journal* **93**, 558–561.

Scott-Ham, M. & Burton, F.C. (2005). Toxicological findings in cases of alleged drug-facilitated sexual assault in the United Kingdom over a 3-year period, *Journal of Clinical Forensic Medicine* **12**, 175–186.

MARC A. LEBEAU

Drug-Impaired Driving

Introduction

Every year more than a million people in the world are killed in traffic crashes and many millions more are injured. Beside the heavy and tragic burden on those directly affected, there is an enormous economic impact, costing countries between 1 and 4% of their gross national product. Driving is a very complex task, requiring the cooperation of several different cognitive and psychomotor functions at once. Crashes can be the consequence of many different factors, which can be classified into three categories: the road, the vehicle, and the driver. A crash is rarely attributable to only one factor, indicating that it is very difficult to precisely determine in what percentage of crashes alcohol or drugs have contributed.

Influence of Drugs on Performance

The effects of drugs on performance can be studied by means of experimental studies, in which different doses of a certain drug are administered to volunteers, and the effects on performance are measured and compared to a placebo or a positive control. The performance of the volunteers can be evaluated by means of tests that assess the different psychomotor and cognitive functions, by means of tests in a driving simulator, or by "real" driving tests [1–3].

Alcohol

Alcohol is a central nervous system depressant. Many studies have already been performed to determine the effects of acute alcohol ingestion on cognitive functions and driving performance. These studies found that numerous driving-related skills are degraded beginning at low blood alcohol concentrations (BACs). Several skills have been shown to decrease with increasing BAC, such as prolongation of reaction time performances and lowering of coordination performance [4, 5]. (For further information *see also* **Alcohol**.)

Cannabis

A cannabis user feels euphoria, relaxation, and increased social interaction with frequent laughing and experiences changes in perception (visual, audible, sensory, or time perception). The users are aware of the effects of the drug, and this awareness increases with higher doses. Cannabis acutely reduces some cognitive and psychomotor skills such as learning, equilibrium, coordination, tracking ability, memory, perception, motor impulsivity, and vigilance, and these effects are mostly dose dependent [6, 7].

Cannabis can also have an effect on behavior. The influence of cannabis on human risk taking is unclear. The results of experiments in laboratory settings are contradictory, while in some driving studies (with rather low doses) users are aware of the impairment and often compensate their driving style by driving more slowly, overtaking less, or keeping longer distances from other vehicles. Nevertheless the driver is still unable to completely compensate for the loss of capability in some psychomotor skills [8].

Some deleterious effects of cannabis appear to be additive or even synergistic with those of alcohol, and the combination of both substances results in a prolongation as well as enhancement of their effects. Driving studies revealed that drivers under the influence of both alcohol and cannabis were less attentive to traffic approaching from side streets, while the use of either cannabis or alcohol (at low doses) had no effect [9], and that the combination of cannabis and alcohol generated an additional decrement in control of lateral deviation on top of the decrement caused by either cannabis or alcohol [10]. The detrimental effects of other drugs such as cocaine can also be reinforced by additional intake of cannabis [11]. (For further information *see also* **Cannabis**.)

Amphetamines

Amphetamine causes a strong central stimulation and euphoria. The user thinks he can do everything and will take more risks. In addition, amphetamine widens the pupils (mydriasis) and reduces sleepiness leading to insomnia, but after some time (hours or days depending on the pattern of use), the subject is exhausted and falls asleep (crash phase). Amphetamine can improve some cognitive functions such as divided attention performance and verbal interaction [12]. However, tests in driving simulators reveal that the intake of amphetamine causes a decrease in overall simulated driving by inducing problems such as incorrect signaling, failing to stop at a red traffic light, and slowing reaction times. The

decrease in simulated driving ability is only observed during daytime, which is consistent with the tunnel vision associated with amphetamine consumption [12, 13]. It is also important to note that the doses of amphetamine administered in these experimental studies were very low (10–30 mg) and thus not representative of the doses that are generally taken by abusers (100–1000 mg/day) [1].

Studies investigating the effect of amphetamine in sleep-deprived persons revealed a positive effect on psychomotor functions [14, 15]. The effects of amphetamine on cognitive functions in sleep-deprived persons are less obvious. Both positive and negative effects as well as no effects have been assessed [12, 14].

Methamphetamine like amphetamine is a central nervous system stimulant that may cause restlessness, euphoria, dizziness, dysphoria, tremor, and insomnia.

Ecstasy (methylenedioxymeth(yl)amphetamine, MDMA, XTC) is a "designer" amphetamine, indicating that it is synthesized to resemble the effects of amphetamine. It causes a weaker stimulation of the central nervous system than amphetamine, but it can also cause sensory disturbances, nausea, dizziness, ataxia, muscular rigidity, sweating, restlessness, and tremor.

Ecstasy acutely causes decreases in attention, short- and long-term memory, verbal memory, visuospatial skills, executive functioning, and prediction of object movement under divided attention [16–18]. It also leads to improved psychomotor performance on a battery of tests, such as movement speed and tracking performance in a single, as well as in a divided attention task [16]. Tests in driving simulators however revealed that the intake of ecstasy can decrease performance by increasing speed and speed variation, and inducing problems in car following, while some tasks are not influenced (reaction time, lateral control), and may even be improved (e.g., lateral control) [17, 19].

Other psychoactive substances such as alcohol can reinforce the deleterious effects of ecstasy, and cause some additional negative effects [20]. On the other hand, the use of ecstasy can diminish some, but not all detrimental effects of alcohol, while other negative effects of alcohol can be reinforced [17].

During the crash phase following the use of amphetamines, the subject feels very tired, unable to combat sleep and depressed. This phase can last for several days [21]. (For further information *see also* **Amphetamine**.)

Cocaine

Cocaine is extensively metabolized to a variety of compounds. The major metabolites are benzoylecgonine, ecgonine, and ecgonine methyl ester and they are often targeted in analyses.

The desired effects of cocaine are similar to those of the amphetamines, but the onset is slower and the duration is longer. The use of cocaine can partially reverse performance decrements in sleep-deprived persons [22]. In rested persons, some studies found no effect of the use of cocaine on psychomotor or cognitive skills [23], while other studies assessed an improvement in psychomotor performance (decreased reaction time), attention, and learning [24].

Cocaine can partially diminish performance decrements caused by alcohol consumption. The use of a combination of alcohol and cocaine decreases psychomotor impairment and improves performance on cognitive tests when compared to the use of alcohol alone. Cocaine also decreases the subjective feeling of drunkenness caused by alcohol [11, 25]. Detrimental effects of other drugs such as cannabis can be reinforced by cocaine [11].

A depressive phase follows the use of cocaine, with the subject feeling very tired, depressed, nervous, and unable to combat sleep [1]. (For further information *see also* **Cocaine**.)

Heroin

The user generally feels intense euphoria ("rush") accompanied by a warm flushing of the skin, dry mouth, and heavy extremities, and alternates between a wakeful and drowsy state. Few experimental studies have investigated the acute effects of heroin in humans. Several studies confirmed the acute effect of heroin on subjective sedation and miosis [26, 27]. One study found a trend toward a decreased performance on the circular lights task, which measures psychomotor performance [28]. In another study the administration of heroin impaired performance on a reaction time task [26]. However the doses used in these experimental studies ranged from 2 to 20 mg, while average daily doses in a chronic, tolerant user range from 300 to 500 mg of heroin [1]. (For further information *see also* **Opioids**.)

Epidemiology and Risks

Prevalence of Drugs in the General Driving Population or in a Subset of Drivers

The prevalence of drugs in the general driving population can be estimated by means of roadside surveys, in which samples of randomly stopped drivers are analyzed. Results of roadside surveys have shown that about 1–2% of drivers stopped during roadside surveys tests positive for drugs in saliva. Higher prevalence rates were found in studies using urine as sample (6–12%) or in studies where samples were only collected during weekend nights (6–15%) [29].

In some studies, samples of a subset of drivers are analyzed. Such studies have shown that drugs are prevalent in 19–50% of drivers injured by a traffic crash, 6–35% of drivers killed by a traffic crash and 55–99% of drivers suspected of driving under the influence of drugs (DUID) [29]. These figures can vary strongly, because the methodology used in the different studies can differ in many aspects. For example, some studies used blood as biological matrix, while others used urine or saliva. As these matrices have different detection times, this can have an influence on the results [30]. Furthermore, the methodology of the studies can differ in the type of drugs included in the analysis and the kind (and sensitivity) of analysis method used.

The most prevalent drug in most of these studies is cannabis. Other important drugs are amphetamines, benzodiazepines, and opiates.

Besides analyzing biological samples of drivers, the prevalence of DUID can also be assessed by interviewing persons. A disadvantage of this method is the possibility of underestimation of the prevalence. Such interviews have shown that about 3.6% of the general population, 15% of the young people, and 85% of drug users state ever to have driven after having used drugs [29]. In the United States in 2005, an estimated 10.5 million persons aged 12 or older reported driving under the influence of an illicit drug during the past year [31].

Risks Associated with DUID

By comparing the prevalence of a certain drug in the general driving population to the prevalence in drivers who were involved in a traffic accident, some studies tried to estimate the accident risk. Several studies have already shown that an increasing BAC is associated with increasing accident risks [32, 33]. Other studies have calculated the accident risks associated with other psychoactive substances. The impaired motorists, methods of roadside testing and assessment for licensing (IMMORTAL) study in the Netherlands revealed that drivers under the influence of benzodiazepines alone have a relative risk for an accident that is three times greater than the risk of a drug-free driver. The highest risk was associated with the use of drugs in combination with alcohol ($\geq$0.8 g l^{-1}), namely a 179 times higher risk [34]! A study that was performed in Canada, for example, showed that the crash risk associated with driving under the influence of cocaine, benzodiazepines, and cannabis is respectively 5, 2.5, and 2.2 times higher than that of a person who has not consumed these drugs [33].

Some studies have made a more thorough analysis than calculating the risk of being involved in a traffic crash, and have estimated the risk of being responsible for a traffic crash while DUID. The results of these types of studies indicate that increasing BACs are associated with increasing risks of being responsible for an accident [35]. They also show that driving under the influence of cannabis increases the risk of being responsible for a crash. The risk of being responsible for an accident even increases with increasing cannabis concentration in the blood, indicating a causal relationship between cannabis and crashes. The risk of being responsible for a fatal accident when driving under the influence of cannabis and alcohol is approximately equal to the multiplication of the risks when driving under the influence of cannabis or alcohol alone [35]. Responsibility analyses have also shown that benzodiazepines and cocaine are associated with increased risks of being responsible for an accident, and that the risk is higher for a combination of alcohol and benzodiazepines than for benzodiazepines alone [36, 37].

Legislation

Each developed country has its specific legislation to deal with DUID. There is a lack of uniformity in the way in which nations approach the drugged driving problem. Generally there exist two major types of DUID legislation, namely "impairment" legislation and "*per se*" legislation [38].

Table 1 Analytical cutoff limits in blood, plasma, or serum as agreed upon or proposed in different countries (all concentrations in ng ml^{-1}, except Sweden where it is in ng g^{-1})

	Germany[a]	Belgium	France	Sweden	Switzerland
Sample type	Serum	Plasma	Blood	Blood	Blood
Amphetamine	50	50	50	30	15[b]
MDMA	50	50	–	20	15
MDEA	50	50	–	20	15
MDA	–	–	–	20	–
MBDB	–	50	–	20	–
Cocaine	–	50	50	20	15
Benzoylecgonine	150	50	50	20	–
Morphine (free)	20	20	20	5	15
THC	1	2	1	0.3	1.5

MDMA, 3, 4-methylenedioxy-*N*-methylamphetamine; MDEA, 3, 4-methylenedioxy-*N*-ethylamphetamine; MDA, 3, 4-methylenedioxyamphetamine; MBDB, 3, 4-methyl-1-(3,4-methylenedioxyphenyl)-2-butanamine; THC, Δ9-tetrahydrocannabinol

[a] Lower cutoffs have been proposed, but they are not yet used everywhere

[b] For Switzerland also the cutoff is 15 ng ml^{-1} for methamphetamine; for all analytes, an error margin of 30% is added to the cutoff

Impairment Legislation

In impairment legislation, the prosecution must demonstrate that the driver was impaired, not fit to drive or "under the influence" depending on how the law is interpreted. The analysis of drugs in body fluids only provides corroborating evidence as to the cause of the impairment. This kind of legislation is subjective and requires the assessment by a medical doctor or a specially trained police officer. As a consequence many of the countries with this kind of legislation experience difficulties in obtaining convictions. Examples of countries with "impairment" legislation are Norway and the United Kingdom.

"Per se" Legislation

A "*per se*" law prohibits driving if drugs are present in blood, serum, plasma, or oral fluid above a certain threshold. Since the prosecution does not have to prove that the driver was impaired, this kind of legislation facilitates the enforcement process. The threshold concentrations, or cut-offs, used are analytical detection limits, meaning that any detectable concentration of a drug constitutes an offense. Therefore these laws are sometimes called *zero-tolerance* laws. Presently, Germany, Belgium, Sweden, France, Finland, Luxembourg, Switzerland, Denmark, a number of Australian states, and 14 states of the United States have introduced "*per se*" legislation in addition to the "impairment" legislation. The analytical cutoffs of Germany, Belgium, France, Sweden, and Switzerland are given in Table 1.

There is no consensus on the analytical cutoffs between the different countries. This lack of consensus can be partially attributed to the use of different biological matrices (serum in Germany, plasma in Belgium, and whole blood in Australia, France, Sweden, and Switzerland) and the different consequences of a positive result: for example, in Belgium there is a penal sanction that follows a positive result, while in Germany there is an administrative sanction. The effectiveness of "*per se*" legislation in increasing the number of prosecutions has already been demonstrated in some countries. For example, in Finland there was a slow increase in the number of samples that were investigated until 2002. However since the introduction of "*per se*" legislation in 2003 there was a cumulative monthly increase in the number of samples (Figure 1) [39].

In Sweden, immediately after the zero-limit law came into force, the number of cases of DUID submitted by the police for toxicological analysis increased sharply and was 10 times higher in 2005 than before the new legislation. Nevertheless it was found that Sweden's zero-concentration limit has done nothing to reduce DUID or deter the typical offender because recidivism is high in this population of individuals (40–50%). Many traffic delinquents in Sweden are criminal elements in society with previous convictions for drunk and/or drugged driving as well as other offenses. The spectrum of drugs identified in blood samples from DUID suspects has

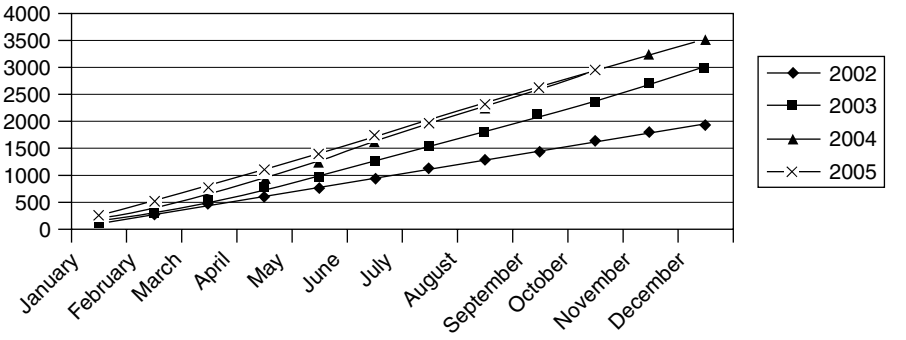


Figure 1 The cumulative monthly increase in the number of samples after introducing the zero-tolerance law in 2003 in Finland [Reproduced with permission from Ref. 39. © P. Lillsunde, 2005.]

not changed much since the zero-limit law was introduced [40].

Detection of Drugs in Drivers

Roadside Detection

Since many years, police officers involved in road safety have expressed the need for a rapid and reliable roadside screening test for drugs, similar to an alcohol breathalyzer. This would help them to determine which drivers have to provide a blood sample, or to take immediate administrative measures like confiscating the driver's license or impounding the vehicle. As illegal drugs are not released in measurable amounts in the breath, roadside drug testing must be based on other specimens (*see* **Oral Fluid Toxicology**).

Roadside detection tests are mostly immunoassays, which are read visually or by a small electronic reader. At first, urine was used for roadside drug testing because of the high drug concentrations. Unfortunately for some substances such as cannabis, the metabolites can be detected for a long time after chronic use. Consequently, the presence of drugs in urine does not necessarily indicate impairment. Another disadvantage of urine is the necessity of sufficient privacy during the sample collection. Nevertheless, roadside urine screening in Belgium significantly decreases the number of unnecessary confirmatory blood analyses for DUID [41]. An example of an on-site urine test for drugs is given in Figure 2.

In recent years, the interest in the use of oral fluid as biological matrix has increased significantly, as this matrix displays some particularly interesting properties. First of all, oral fluid can be obtained easily by nonmedical personnel in a relatively noninvasive and observable way. There is also some correlation with impairment. The results of the European project roadside testing assessment (ROSITA) indicated that for most drugs of abuse the correlation with blood is better for oral fluid than for urine. Nevertheless, the results of ROSITA and the follow-up project ROSITA 2 indicated that none of the currently available on-site oral fluid drug testing devices is reliable enough to be recommended for roadside screening for drivers [42]. However, the experience in the state of Victoria in Australia shows that random roadside oral fluid testing of drivers for methamphetamine, ecstasy, and cannabis has a deterrent effect: the level of awareness of drivers of random oral fluid testing increased from 78 to 92%, 33% of illicit drug users stated that the drug tests had influenced them (primarily to avoid taking drugs when they are going to drive) and the proportion of drug-using respondents who drove while under the influence of drugs dropped in the after period from 45 to 35% (Swann, P. 5-12-2005. Baltimore, Maryland, USA. Rosita 2 meeting. *Personal Communication*). On the other hand, the oral cavity can be contaminated by intranasal and smoked drug use, leading to extremely high concentrations in oral fluid. It is also difficult to obtain sufficient sample volume for the analysis, and the concentrations of 9-tetrahydrocannabinol (THC) and benzodiazepines in this matrix are low [43].

Evidentiary Analysis

Blood is considered to be the best matrix for confirmation analysis, because the presence of drugs

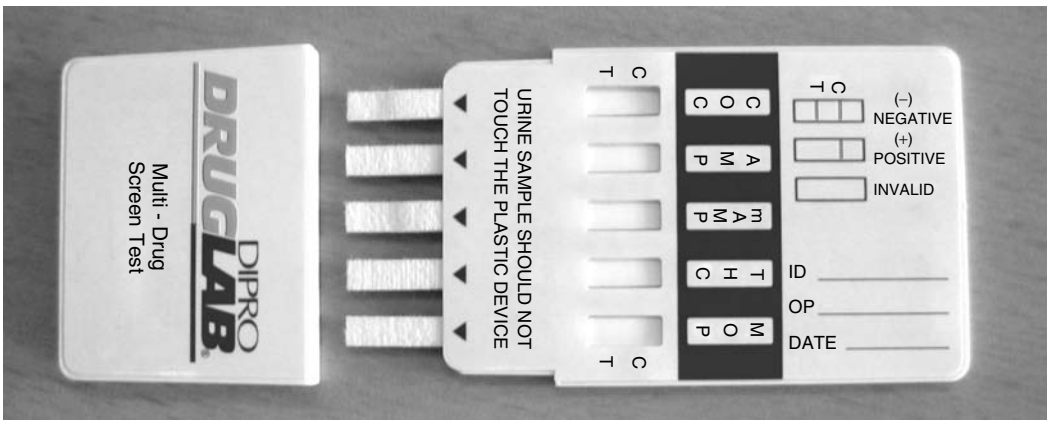


Figure 2 An example of a multidrug on-site urine test: the tips should be placed in the urine sample (urine should not touch plastic device). At the C level, a line should appear, indicating that the test is valid. At the T level, the presence of a line indicates a negative result and the absence a positive result

in blood corresponds best with recent use and impairment. In many countries with "*per se*" legislation, the only legally allowed evidence for DUID is confirmation of the presence of drugs in blood. A review on drugs of abuse monitoring in blood for control of DUID was recently published [44]. The most widely used method is gas chromatography coupled to mass spectrometry (GC-MS) because of its sensitivity and specificity. However, the procedure is labor intensive and time consuming, as solid phase extraction and derivatization are necessary for sample preparation. In addition, different methods are often needed to quantify different drug classes. Therefore, liquid chromatography (tandem) mass spectrometry (LC-MS(MS)) procedures have been introduced for different classes of drugs for confirmatory analyses or even for screening and confirmation in one step. Several laboratories are developing methods that detect a large series of different drugs in one procedure in a small sample volume [45].

Another biological specimen that is used for evidentiary analysis in the context of DUID is hair. In Germany, Italy, and other countries, hair analysis is used for license (re)granting to drug dependent persons. The hair samples are first screened by immunoassay and positives are then confirmed by high pressure liquid chromatography (HPLC), capillary electrophoresis (CE), GC-MS, or LC-MS. The major practical advantage of hair testing compared with urine and blood testing for drugs is its larger detection window, which is weeks to months, depending on the length of hair shaft analyzed. Other advantages of hair are its stability and ease-of-transportation. A disadvantage of the use of hair as biological matrix is the possible contamination by exposure to drugs in the air and the absence of correlation with recent use because of the delayed appearance of drugs in hair [46–48]. (For further information *see also* **Hair: Toxicology**.)

Conclusion

There is an increasing knowledge regarding the influence of drugs on performance, and the prevalence of drugs other than alcohol in road traffic. Experimental studies clearly show that many drugs can have a detrimental effect on driving performance. These negative effects are often even more pronounced when drugs are combined with alcohol. Results of epidemiological studies confirm the detrimental effects of drugs on driving performance. These studies show that DUID is associated with increased crash risks and increased risks of being responsible for a traffic crash. They also show that these risks increase even more when drugs are taken in combination with alcohol compared to when drugs are taken alone.

Regarding DUID there is a clear move toward "*per se*" legislation, although some countries at this time have decided to stay with impairment legislation, and some have both (e.g., Australia, Germany etc.). The detection by the police of drivers under the influence of drugs can be done by means of screening tests. In

recent years, the interest in oral fluid screening tests has grown significantly. Studies however have shown that none of the currently available on-site devices are reliable when used alone. More recent methods developed for confirmation analysis use HPLC with tandem mass spectrometry.

It is possible that future developments could lead to on-site screening of capillary blood obtained by a finger prick and confirmation of the presence/absence of drugs by analyzing dried blood spots. Possible advantages of this approach would be the easy transportation (no need for refrigeration or shipping on dry ice necessary), the stability of parent substances at ambient temperature, and less risk of loss of sample (e.g., breakage of glass tube of blood) or of infection.

References

[1] Couper, F.J. & Logan, B.K. (2004). *Drugs and Human Performance Fact Sheets*. DOT HS 809 725, NHTSA, pp. 1–100.

[2] Baselt, R.C. (2001). *Drug Effects on Psychomotor Performance*, Biomedical Publications, Foster City.

[3] Raes, E. & Verstraete, A.G. (in press) *Influence of Drugs on Performance. Workplace Drug Testing*, 1st Edition, Pharmaceutical Press, London.

[4] Liguori, A., D'Agostino, R.B., Dworkin, S.I., Edwards, D. & Robinson, J.H. (1999). Alcohol effects on mood, equilibrium, and simulated driving, *Alcoholism: Clinical and Experimental Research* **23**(5), 815–821.

[5] Grant, S.A., Millar, K. & Kenny, G.N.C. (2000). Blood alcohol concentration and psychomotor effects, *British Journal of Anaesthesia* **85**(3), 401–406.

[6] Menetrey, A., Augsburger, M., Favrat, B., Pin, M.A., Rothuizen, L.E., Appenzeller, M., Buclin, T., Mangin, P. & Giroud, C. (2005). Assessment of driving capability through the use of clinical and psychomotor tests in relation to blood cannabinoids levels following oral administration of 20 mg dronabinol or of a cannabis decoction made with 20 or 60 mg Delta9-THC, *Journal of Analytical Toxicology* **29**(5), 327–338.

[7] Ramaekers, J.G., Kauert, G., van Ruitenbeek, P., Theunissen, E.L., Schneider, E. & Moeller, M.R. (2006). High-potency marijuana impairs executive function and inhibitory motor control, *Neuropsychopharmacology* **31**(10), 2296–2303.

[8] Sexton, B.F., Turnbridge, R., Brook-carter, N., Jackson, P.G. & Wright, K. (2000). *The Influence of Cannabis on Driving*. TRL 477, Road Safety Division, Department of the Environment, Transport and the Regions, pp. 1–106.

[9] Lamers, C.T. & Ramaekers, J.G. (2001). Visual search and urban driving under the influence of marijuana and alcohol, *Hum Psychopharmacol* **16**(5), 393–401.

[10] Sexton, B.F., Tunbridge, R.J., Board, A., Jackson, P.G., Wright, K., Stark, M.M. & Englehart, K. (2002). TRL 543, *The Influence of Cannabis and Alcohol on Driving*, Road Safety Division, Department of the Environment, Transport and the Regions.

[11] Foltin, R.W., Fischman, M.W., Pippen, P.A. & Kelly, T.H. (1993). Behavioral-effects of cocaine alone and in combination with ethanol or marijuana in humans, *Drug and Alcohol Dependence* **32**(2), 93–106.

[12] Mills, K.C., Spruill, S.E., Kanne, R.W., Parkman, K.M. & Zhang, Y. (2001). The influence of stimulants, sedatives, and fatigue on tunnel vision: risk factors for driving and piloting, *Human Factors* **43**(2), 310–327.

[13] Silber, B.Y., Papafotiou, K., Croft, R.J., Ogden, E., Swann, P. & Stough, C. (2005). The effects of dexamphetamine on simulated driving performance, *Psychopharmacology* **179**(3), 536–543.

[14] Wesensten, N.J., Killgore, W.D.S. & Balkin, T.J. (2005). Performance and alertness effects of caffeine, dextro amphetamine, and modafinil during sleep deprivation, *Journal of Sleep Research* **14**(3), 255–266.

[15] Caldwell, J.A. & Caldwell, J.L. (1997). An in-flight investigation of the efficacy of dextroamphetamine for sustaining helicopter pilot performance, *Aviation Space and Environmental Medicine* **68**(12), 1073–1080.

[16] Lamers, C.T.J., Ramaekers, J.G., Muntjewerff, N.D., Sikkema, K.L., Samyn, N., Read, N.L., Brookhuis, K.A. & Riedel, W.J. (2003). Dissociable effects of a single dose of ecstasy (MDMA) on psychomotor skills and attentional performance, *Journal of Psychopharmacology* **17**(4), 379–387.

[17] Ramaekers, J.G., Kuypers, K.P.C., Wood, C.M., Hockey, G.R.J., Jamson, S., Jamson, H. & Birch, E. (2004). *IMMORTAL Deliverable D-R4.4: Experimental Studies on the Effects of Licit and Illicit Drugs on Driving Performance, Psychomotor Skills and Cognitive Function*.

[18] Smith, R.M., Tivarus, M., Campbell, H.L., Hillier, A. & Beversdorf, D.Q. (2006). Apparent transient effects of recent "ecstasy" use on cognitive performance and extrapyramidal signs in human subjects, *Cognitive and Behavioral Neurology* **19**(3), 157–164.

[19] Laurell, H. & Schlyter, F. (eds) (2000). *A Driving Simulator Study on the Effects of MDMA (Ecstasy) on Driving Performance and Traffic Safety*, The International Council on Alcohol, Drugs and Traffic Safety (ICADTS), Stockholm.

[20] Brookhuis, K.A., de Waard, D. & Samyn, N. (2004). Effects of MDMA (ecstasy), and multiple drugs use on (simulated) driving performance and traffic safety, *Psychopharmacology* **173**(3–4), 440–445.

[21] Verheyden, S.L., Henry, J.A. & Curran, H.V. (2003). Acute, sub-acute, and long-term subjective consequences of 'ecstasy' (MDMA) consumption in 430 regular users, *Human Psychopharmacology: Clinical and Experimental* **18**(7), 507–517.

[22] Fischman, M.W. & Schuster, C.R. (1980). Cocaine effects in sleep-deprived humans, *Psychopharmacology (Berlin)* **72**(1), 1–8.

[23] Hopper, J.W., Karlsgodt, K.H., Adler, C.M., Macklin, E.A., Lukas, S.E. & Elman, I. (2004). Effects of acute cortisol and cocaine administration on attention, recall and recognition task performance in individuals with cocaine dependence, *Human Psychopharmacology: Clinical and Experimental* **19**(7), 511–516.

[24] Johnson, B., Overton, D., Wells, L., Kenny, P., Abramson, D., Dhother, S., Chen, Y.R. & Bordnick, P. (1998). Effects of acute intravenous cocaine on cardiovascular function, human learning, and performance in cocaine addicts, *Psychiatry Research* **77**(1), 35–42.

[25] Farre, M., Delatorre, R., Llorente, M., Lamas, X., Ugena, B., Segura, J. & Cami, J. (1993). Alcohol and cocaine interactions in humans, *The Journal of Pharmacology and Experimental Therapeutics* **266**(3), 1364–1373.

[26] Jenkins, A.J., Keenan, R.M., Henningfield, J.E. & Cone, E.J. (1994). Pharmacokinetics and pharmacodynamics of smoked heroin, *Journal of Analytical Toxicology* **18**(6), 317–330.

[27] Martin, W.R. & Fraser, H.F. (1961). A comparative study of physiological and subjective effects of heroin and morphine administered intravenously in postaddicts, *The Journal of Pharmacology and Experimental Therapeutics* **133**, 388–399.

[28] Cone, E.J., Holicky, B.A., Grant, T.M., Darwin, W.D. & Goldberger, B.A. (1993). Pharmacokinetics and pharmacodynamics of intranasal "snorted" heroin, *Journal of Analytical Toxicology* **17**(6), 327–337.

[29] Scheers, M., Verstraete, A.G., Adriaensen, M., Raes, E. & Tant, M. (2006). *Rijden Onder Invloed van Psychoactieve Stoffen: Literatuurstudie en Evaluatie van Het Handhavingsbeleid*, Academia Press, Ghent, p. 291.

[30] Verstraete, A.G. (2004). Detection times of drugs of abuse in blood, urine, and oral fluid, *Therapeutic Drug Monitoring* **26**(2), 200–205.

[31] SAMSHA's Office of Applied Studies (OAS) (2005). *Data of the 2005 National Survey on Drug Use and Health*, http://www.drugabusestatistics.samhsa.gov/latest.htm#Prev.

[32] Kloeden, C.N. & Mc.Lean, A.J. (eds) (1995). *Grand Rapids Effects Revisited: Accidents, Alcohol, and Risk*, The International Council on Alcohol, Drugs and Traffic Safety (ICADTS), Adelaide.

[33] Mayhew, D. & Dussault, C. (eds) (2002). *The Contribution of Alcohol and other Drugs Among Fatally Injured Drivers in Québec: Some Preliminary Results*, The International Council on Alcohol, Drugs and Traffic Safety (ICADTS), Montréal.

[34] Assum, T., Mathijssen, M.P.M., Houwing, S., Buttress, S.C., Sexton, B., Turnbridge, R.J. & Oliver, J. (2005). *IMMORTAL Deliverable D-R4.2: The Prevalence of Drug Driving and Relative Risk Estimations*. A study conducted in the Netherlands, Norway.

[35] Laumon, B., Gadegbeku, B., Martin, J.L. & Biecheler, M.B. (2005). Cannabis intoxication and fatal road crashes in France: population based case-control study, *British Medical Journal* **331**(7529), 1371–1376.

[36] Soderstrom, C.A., Dischinger, P.C., Kufera, J.A., Ho, S.M. & Shepard, A. (2005). Crash culpability relative to age and sex for injured drivers using alcohol, marijuana, or cocaine, *Annual Proceedings - Association for the Advancement of Automotive Medicine* **49**, 315–330.

[37] Longo, M.C., Hunter, C.E., Lokan, R.J., White, J.M. & White, M.A. (2000). The prevalence of alcohol, cannabinoids, benzodiazepines, and stimulants amongst injured drivers and their role in driver culpability: part ii: the relationship between drug prevalence and drug concentration, and driver culpability, *Accident Analysis and Prevention* **32**(5), 623–632.

[38] Walsh, J.M., De Gier, J.J., Christophersen, A.S. & Verstraete, A.G. (2004). Drugs and driving, *Traffic Injury Prevention* **5**, 241–253.

[39] Lillsunde, P., Gunnar, T. & Seppa, H. (2005). Rosita 2 in Finland. *Rosita 2 Meeting*. 5-12-2005, Baltimore.

[40] Jones, A.W. (2005). Driving under the influence of drugs in Sweden with zero concentration limits in blood for controlled substances, *Traffic Injury Prevention* **6**(4), 317–322.

[41] Raes, E. & Verstraete, A.G. (2005). Usefulness of roadside urine drug screening in drivers suspected of driving under the influence of drugs (DUID), *Journal of Analytical Toxicology* **29**(7), 632–636.

[42] Verstraete, A.G. & Raes, E. (eds) (2006). *Rosita-2 Project: Final Report*, Academia Press, Ghent, p. 257.

[43] Drummer, O.H. (2006). Drug testing in oral fluid, *Clinical Biochemistry Reviews* **27**, 147–159.

[44] Moeller, M.R. & Kraemer, T. (2002). Drugs of abuse monitoring in blood for control of driving under the influence of drugs, *Therapeutic Drug Monitoring* **24**(2), 210–221.

[45] Oiestad, E.L., Johansen, U. & Christophersen, A.S. (2007). Drug screening of preserved oral fluid by liquid chromatography-tandem mass spectrometry, *Clinical Chemistry* **53**(2), 300–309.

[46] Nakashima, K. (2005). High-performance liquid chromatographic analysis of drugs of abuse in biological samples, *Journal of Health Science* **51**(3), 272–277.

[47] Kintz, P., Villain, M., Ciriminele, V., Janey, C. & Ludes, B. (2003). Décret n° 2003-293 du 31 mars 2003. Restitution de permis de conduire à partir d'analyses de cheveux, *Annales de Toxicologie Analytique* **15**(2), 117–122.

[48] Kintz, P. (2006). *Analytical and Practical Aspects of Drug Testing in Hair*, CRC Taylor and Francis, London, New York, p. 382.

Further Reading

Raes, E., Van den Neste, T. & Verstraete, A.G. *Drugs and Driving: Latest Developments and Findings from Research*

Since 1999, The European Monitoring Center for Drugs and Drug Addiction, Insight; in press.
Raes, E., Verstraete, A. & Wennig, R. (2008). Drugs and driving, in *Handbook of Analytical Separations, Volume 6: Forensic Science*, 2nd Edition, Elsevier, Oxford, pp. 611–651.

ELKE RAES AND ALAIN G. VERSTRAETE

Drugs: Driving ***see*** Drug-Impaired Driving

Drugs: Hair ***see*** Hair: Toxicology

Drugs: Prescription and Psychotropic ***see*** Psychopharmacology

Drugs: Prescription Psychotropic, Child and Adolescent ***see*** Psychopharmacology: Child and Adolescent

Drugs: Profiling ***see*** Drug Profiling

Drugs: Sexual Assault ***see*** Drug-Facilitated Sexual Assault

Drugs: Sweat ***see*** Sweat: Toxicology

Drugs: Urine ***see*** Drug Testing: Urine

Durham Standard of Insanity ***see*** Insanity: Defense, Behavioral Science Evidence

Duty to Protect ***see*** Duty to Warn

Duty to Warn

Introduction

Historically, a physician's paramount duty is to his or her patient. The Hippocratic oath, upon which many modern medical students still swear prior to graduation, specifically mentions the concept of confidentiality: "Whatever I see or hear in the lives of my patients, whether in connection with my professional practice or not, which ought not to be spoken of outside, I will keep secret, as considering all such things to be private" [1]. Stated otherwise, confidentiality is the obligation of a clinician to keep a patient's privacy intact. Given the personal and intimate nature of patient complaints, confidentiality of these complaints is one of the bedrock principles of the physician–patient relationship.

Mental Health and Confidentiality

Physician–patient confidentiality is nowhere as important as in the field of mental health. Mental illness is highly feared and stigmatized in most cultures. It is therefore difficult for most persons suffering symptoms of psychosis or depression to disclose their symptoms and seek help. Once they do, it is important they feel comfortable openly discussing their problems. Since most psychiatric symptoms are subjective, there are few objective tests to diagnose mental illness, so a clinician must rely upon frank patient disclosure. Lastly, most mental health treatment involves both medication and therapy, and openness and trust are critical in the establishment and maintenance of a therapeutic relationship [2]. Consider an attorney who develops significant depressed mood. If she does not believe her communications to a therapist will be kept confidential, she may well be reluctant to provide an honest recounting of symptoms, or even seek help in the first place. A clinician therefore takes the duty of patient confidentiality seriously. With few exceptions, therapists cannot reveal any patient information without the express permission of that patient.

Exceptions to Confidentiality

Although the clinician's duty to preserve confidentiality in patient communications is long established, there are exceptions where confidentially may be ethically and legally breached. In cases where there is no legal mandate to breach confidentiality, some authors recommend an ethically driven approach of using "consultation" and "carefulness" to guide the communication of patient information [3]. In other cases, for example infectious disease, communication is legally required. In 2007, a US attorney with multidrug resistant tuberculosis took several transatlantic flights. Officials in numerous countries were made aware of his name and condition, in efforts to test and treat persons he may have infected [4]. His confidentiality was breached because others were endangered.

There is precedent in medicine, therefore, for breaching patient confidentiality to protect the public, although such breaches should be ethically limited in scope to what is necessary to achieve public protection. In mental health, this situation is represented by the "dangerous patient exception" [5] to patient privilege, which recognizes that a court might require the testimony of a therapist to protect others. For example, psychiatrists in most jurisdictions can legally petition the involuntary commitment of a person, based on dangerousness to self or others. If a depressed attorney tells his psychiatrist he is going to kill himself, the psychiatrist is not ethically or legally liable for revealing patient communications about suicide in court testimony.

Although the dangerous patient exception allowed psychiatrists to testify in the court about patients who were dangerous to self or others, this exception did not include a threat of serious physical harm until *Tarasoff* (*see also* **Dangerousness: Risk of**; **Violence Risk Assessment for Mental Health Professionals**).

Confidentiality vs. Privilege

Although confidentiality and privilege are sometimes used interchangeably, they have two very different meanings. Confidentiality is the duty of the clinician to keep patient information private. The privilege of private communication, in contrast, belongs to the patient.

Tarasoff v. Regents of the University of California

Tarasoff I

In 1968, Tatiana Tarasoff and Prosenjit Poddar both were students at the University of California at Berkley. After briefly dating, Poddar assumed a more serious relationship existed, and became depressed when Tarasoff did not feel the same. In the course of therapy at the University mental health clinic, Poddar revealed thoughts of killing a woman readily identified as Tarasoff. Poddar's psychologist contacted the University police to detain Poddar, but since he appeared to be rational to police when questioned, they let him go. Poddar stopped going to therapy, and two months later fatally shot and stabbed Tarasoff.

Tarasoff's parents sued both the mental health clinic and the University of California. In 1974, in what has become known as *Tarasoff I*, the California Supreme Court ruled that since Poddar's psychologist did not warn Tarasoff, he did not sufficiently protect her. The court stated, "When a doctor or psychotherapist, in the exercise of this professional skill and knowledge, determines or should determine, that a

warning is essential to avert danger . . . he incurs a legal obligation to give a warning" [6].

Both the defendants and the American Psychiatric Association (APA) protested this decision and requested a rehearing. In an *amicus curaie* brief the APA supported the defense on the grounds that mental health professionals are unable to predict dangerousness with any special accuracy, and the Tarasoff ruling would damage confidentiality, the cornerstone of the therapeutic relationship.

Tarasoff

In 1976, the California Supreme Court reheard the Tarasoff case and established a duty to protect: "When a therapist determines, or pursuant to the standards of his profession should determine, that his patient presents a serious danger of violence to another, he incurs an obligation to use reasonable care to protect the intended victim against such danger." Duty to warn was merely one option of how to protect [7]. The court responded to the APA's concerns that the Tarasoff duty would breach patient confidentiality by recognizing the precedent of the dangerous patient exception. In a now-famous quote, Judge Tobriner commented, "the protective privilege ends where the public peril begins" [7].

Controversy and Tarasoff

Duty to Warn, or Duty to Protect?

Although the final Tarasoff ruling clearly established a duty to protect, this duty has often been misinterpreted by both courts and therapists as a duty to warn [8]. This confusion is clearly reflected in the subsequent history of Tarasoff in California, in which the courts emphasized warning to the exclusion of any other type of protection [9].

After the second Tarasoff decision in 1976, which created a duty to protect, several court cases found therapists at fault for failing to predict dangerousness. For example, therapists were found liable for failure to warn third parties of a patient's inability to drive [10]. Although some clinicians felt that therapists had only themselves to blame due to "overstated claims of capacities to predict and treat aggressive behavior" [11], California mental health professionals sought an "immunity statute" to provide some protection from the liability for failure to predict dangerousness. In 1986, the California Civil Code was enacted as a "Tarasoff immunity statute". This statue "made clear that victims had to be readily identifiable and was intended to establish warning as one way . . . (to) discharge the duty to warn or protect" [9].

Despite this statute, the Tarasoff duty continued to be misunderstood as a duty to warn. This misinterpretation was evident in two California Court of Appeals cases involving former police officer G. Colello, who killed first his exgirlfriend's lover (K. Ewing), then himself. Mr Colello had communicated his intent to harm Mr Ewing to his father, who reported this to Dr Goldstein, Mr Colello's therapist. Mr Colello was hospitalized but soon discharged, despite communication of the threatening statements Mr Colello made to his father [9].

Mr Ewing's parents sued both Dr Goldstein and the hospital. The court ruled that therapists had a duty to protect solely by warning, regardless of whether they judged this to be the best way of protecting. A therapist could therefore be held liable if he decided to protect potential victims of his patient by, for example, involuntary hospitalization, but not warning. Moreover, this duty would be established whenever a therapist believed a threat to be credible, without any risk assessment or evaluation of dangerousness. A therapist could therefore breach patient confidentiality and issue a warning, based merely on a patient's verbalization of threat (or communication of threat to family members), regardless of whether the threat was credible. As a result of persistent lobbying by California mental health groups, a 2006 amendment to the California Civil Code returned the Tarasoff duty to a duty to protect [9]. At present, therapists in California can now individualize their responses to assessed risks, a recommended approach [12, 13].

The Therapeutic Relationship

Many therapists worried that if they were unable to guarantee confidentiality under Tarasoff, patients would either avoid honest disclosure in therapy or eschew therapy altogether. A variety of studies, however, have showed that issuing a Tarasoff warning has had a minimal effect on the therapeutic relationship [14].

Criminalization of Tarasoff

After *Tarasoff*, some argued that the decision allowed for "criminalization" of the doctor–patient relationship. In subsequent California cases, the courts mandated the testimony of a therapist if that therapist

had given a warning under *Tarasoff*. The therapist had to testify not only about the warning but also about other communications that might provide justification for the warning. One author noted in 1992 "the continuing erosion of confidentiality has resulted in psychiatrists or other mental health professionals becoming prosecution witnesses at the criminal trials of their own patients" [8]. In one instance a psychiatrist even surreptitiously gathered evidence that was ultimately used in testimony against his patient [15], an action that seems to go against the root of medicine's precept, "do no harm".

Outside of California, court rulings seem to leave open the possibility of continued "criminalization". In 1996, the US Supreme Court found in *Jaffee v. Redmond* a patient–therapist privilege, but acknowledged exceptions to this privilege. In the line of duty, police officer Redmond shot and killed Allen, a man she believed was brandishing a weapon toward a third party. Allen's estate (Jaffee) disputed this and sued for damages. The prosecution discovered that Redmond sought counseling after the shooting, and demanded the therapist's notes for cross-examination. When the therapist refused, the judge instructed the jury that the notes had likely been "unfavorable" to Redmond. The jury found for Jaffee.

This decision was overturned by the court of appeals (seventh circuit), which cited that "reason and experience" under Rule 501 of the Federal Rules of Evidence compelled the trial court to recognize and protect a therapist–patient privilege. The US court of appeals agreed, but found that this privilege could not be uniformly protected. In a footnote to the decision, the court stated, "... we do not doubt that there are situations in which the privilege must give way, for example, if a serious threat of harm to the patient or to others can be averted only by means of a disclosure by the therapist" [16]. This decision seems to leave room for courts mandating disclosure by a therapist in court proceedings, including criminal [17].

The Legacy of Tarasoff

The United States

In the past 30 years, 27 states have adopted *Tarasoff*-defining statues [9]. In adopting these statutes, courts have used a variety of approaches, and reached a variety of conclusions. Given the variety of court reactions, and the ever-changing nature of *Tarasoff*, one author concludes, "even in states that have *Tarasoff* statutes, clinicians must continue to rely on their clinical and ethical judgment, rather than statutory guidance, when considering potential protective disclosures or future drafts of protective disclosure statutes" [18].

Canada

In Canada, the courts have recognized a common law exception to doctor–patient confidentiality in the area of public safety [19], but in general have not legislated a duty to protect. The province of Ontario has probably come closest to such legislation. In 1998, the province organized a medical expert panel on duty to inform, "to consider the duty of Ontario physicians in circumstances where a patient threatens to kill or cause serious bodily harm to a third party". After reviewing the *Tarasoff* case and literature on risk assessment, the panel concluded that although the principle of confidentiality should be closely guarded, it should be superseded in the cases of suspected serious harm. The panel recommended the following: physicians have a duty to inform a third person when a patient threatens to cause serious harm to another person, and "it is more likely than not the threat will be carried out". The panel concluded that "the need to protect the public from likely risk of serious harm supersedes a physician's duty to keep patient information confidential", and recommended "a standard of practice for the assessment of risk when a threat is made, and this standard should be enforced" [20].

In 1999 the Supreme Court of Canada heard the case, *Smith v. Jones*, and determined that three factors should be considered before confidentiality should be breached in favor of public safety: (i) is there a clear risk to an identifiable person or persons; (ii) is there a risk of serious harm or death; and (iii) is the danger imminent? Although this case did not establish a legislative duty to protect, it did clarify questions that Canadian mental health professionals should ask when assessing risk of violence [19].

Europe

Most European courts, while recognizing that psychiatrists might need to breach confidentiality in cases of patient dangerousness, have not issued *Tarasoff*-type statuary duties. At the time of this writing, however, some speculate that the European Court of Human

Rights' Ruling in *Osman v U.K.* (2007) may bring about such statues [21].

Australia

Although there is no case law in Australia that specifically protects clinician–patient confidentiality, the criminal courts have exercised discretion in requesting the disclosure of privileged information. In the civil courts, however, confidentiality is specifically protected by statute in Victoria, Tasmania, and the Northern Territory. In New South Wales, the privilege of clinician–patient confidentiality is suggested by general law [19].

In 1999, the Supreme Court heard the case *Clifford v. Victorian Institute of Forensic Mental Health and Anor*. In this case, local police refused to provide a search warrant to obtain a psychiatric file that might contain information relating to a homicide. The court concurred, finding that "public interest immunity" applied to the files [19].

The Duty

What then is the duty of a therapist when he or she is worried about a potentially dangerous patient? Historically, it has long been recognized that breaching confidentiality in such cases has merit. State and legislative action since the Tarasoff rulings, however, suggest or mandate a duty to protect. Today's mental health practitioner must, therefore, balance the duty of patient confidentiality, with a duty to protect others if that patient is a threat.

Most agree that the best way of determining threat is an assessment of patient dangerousness. When Tarasoff was first heard it appeared that mental health professionals had no ability to predict dangerousness [22]. In the past three decades, however, researchers and clinicians have developed successive "generations" of risk assessment tools that show some predictive accuracy [23]. Despite this increased accuracy, however, it is still up to individual clinicians to judge when a patient seems sufficiently dangerous such that duty shifts from the patient to the public. One study showed that clinical judgment varies widely on this issue [24]. It is possible that incorporating risk assessment into training would help standardize clinical judgment [25] (*see* **Risk Assessment: Patient and Detainee**; **Violence Risk Assessment for Mental Health Professionals**).

The Future

The jurisprudence on the duty to protect continues to evolve. Prudent clinicians should become familiar with laws and statues in jurisdictions in which they practice. Professional associations also regularly address issues such as duty to protect, and issue helpful guidelines.

References

[1] Michael, N. (ed) (2002). *Hippocratic Oath*, National Library of Medicine, http://www.nlm.nih.gov/hmd/greek/greek_oath.html.

[2] Gabbard, G. (2007). Psychotherapy in Psychiatry, *International Review of Psychiatry* **19**(1), 5–12.

[3] Liegeois, A., Haekens, A. & Eneman, M. (2006). Sharing information among carers involved in mental health care: ethical advice, *Tijdschrift voor Psychiatrie* **48**(10), 787–795.

[4] http//:www.webmd.com/news/20070703/andrew-speakers-tb reclassified, (accessed on, 2007).

[5] Harris, G. (1999). The dangerous patient exception to the psychotherapist-patient privilege: the Tarasoff duty and the Jaffee footnote, *Washington Law Review* **74**(1), 33–68. Jan.

[6] Tarasoff v. Regents of University of California, 551 P.2d 553 (Cal. 1974).

[7] Tarasoff v. Regents of University of California, 551 P.2d 334 (Cal. 1976).

[8] Leong, G., Eth, S. & Silva, J. (1992). The psychotherapist as witness for the prosecution: the criminalization of Tarasoff, *American Journal of Psychiatry* **149**(8), 1011–1015.

[9] Weinstock, R., Vari, G., Leong, G. & Silva, J. (2006). Back to the past in California: a temporary retreat to a Tarasoff duty to warn, *Journal of the American Academy of Psychiatry and Law* **34**, 523–528.

[10] Pettis, R. (1992). Tarasoff and the dangerous driver: a look at the driving cases, *Bulletin of the American Academy of Psychiatry and the Law* **20**(4), 427–437.

[11] Miller, R., Doren, D., Van Rybroek, G. & Maier, G. (1988). Emerging problems for staff associated with the release of potentially dangerous forensic patients, *Bulletin of the American Academy of Psychiatry and the Law* **16**(4), 309–320.

[12] Gutheil, T. (2001). Moral justification for Tarasoff-type warnings and breach of confidentiality: a clinician's perspective, *Behavioral Science and the Law* **19**(3), 345–353.

[13] Kachigian, C. & Felthous, A. (2004). Court response to Tarasoff statues, *Journal of the American Academy of Psychiatry and the Law* **32**(3), 63–73.

[14] Binder, R. & McNiel, D. (1996). Application of the Tarasoff ruling and its effect on the victim and the therapeutic relationship, *Psychiatric Services* **47**(11), 1212–1215.

[15] Herbert, P. (2004). Psychotherapy as law enforcement, *Journal of the American Academy of Psychiatry and Law* **32**(1), 91–95.
[16] Jaffee v. Redmond (95–266), 518 U.S. 1 (1996).
[17] Weinstock, R., Leong, G. & Silva, J. (2001). Potential erosion of psychotherapist-patient privilege beyond California: dangers of "criminalizing" Tarasoff, *Behavioral Science and The Law* **19**(3), 437–449.
[18] Kachigian, C. & Felthous, A. (2004). Court responses to Tarasoff statues, *Journal of the American Academy of Psychiatry and The Law* **32**(3), 63–73.
[19] McSherry, B. (2004). Risk assessment by mental health professionals and the prevention of future violent behavior, *Trends and Issues in Crime and Criminal Justice* (paper number 281), 1–6.
[20] Ferris, L., Barkun, H., Carlisle, J., Hoffman, B., Katz, C. & Silverman, M. (1998). Defining the physician's duty to warn: consensus statement of Ontario's Medical Expert Panel on Duty to Inform, *Canadian Medical Association Journal* **15**(11), 1473–1479.
[21] Gavaghan, C. (2007). A Tarasoff for Europe? A European human rights perspective on the duty to protect, *International Journal of Law and Psychiatry* **30**(3), 255–267. Epub 2007 Apr 24.
[22] Monahan, J. (1992). The clinical prediction of violent behavior: perceptions and evidence, *The American Psychologist* **47**, 511–512.
[23] Mossman, D. (1994). Assessing predictions of violence: being accurate about accuracy, *Journal of Consulting and Clinical Psychology* **62**, 783–792.
[24] Mossman, D. (2004). How a Rabbi's sermon resolved my Tarasoff conflict, *The Journal of The American Academy of Psychiatry and The Law* **32**, 359–363.
[25] Tolman, A. (2001). Clinical training and the duty to protect, *Behavioral Sciences and the Law* **19**(3), 387–404.

Related Articles

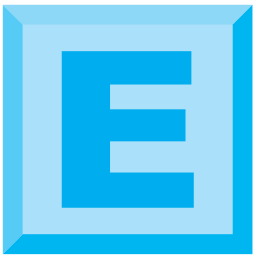
Dangerousness: Risk of

JOY E. STANKOWSKI

DVI ***see*** Disaster Victim Identification

Dyadic Death ***see*** Suicide (Behavior)

Dynamics: Fire ***see*** Fire: Dynamics and Pattern Production



Earprints: Interpretation of

The Use of Earprints for Person identification

Burglars may listen at doors or windows before breaking and entering. Oils and waxes on their ears will then leave prints that can be made visible using techniques similar to those used when lifting fingerprints. Such prints appear to be characteristic of the ears that made them and may be used for person identification or to link various cases. In May 2006, suspected burglars were convicted in the Netherlands for a large series of burglaries that were linked by the presence of similar earprints, as well as toolmarks of what was believed to be the same adjustable wrench.[a] In June 2008, a similar success story was recorded from Belgium when a number of suspected burglars were convicted after earprints provided part of the evidence.[b] In the Dutch case, the earprint evidence was challenged and subsequently accepted in appellate court.[c]

Success stories such as these spark enthusiasm for the use of earprints among police officers and prosecutors. The application of earprints as evidence in court has, however, also frequently been the subject of criticism on the grounds that the process of individualization was considered to be subjective (e.g., [1–3]). Indeed, formal protocols for collecting earprints have not yet been implemented. There are also no generally accepted methods for analyzing and comparing earprints. In the United States, a conviction for murder was overturned in 1999 after an appellate court ruled that the earprint evidence was not generally accepted by the forensic science community under the Frye standard [4]. Prosecutors later dropped the charges. In the United Kingdom, the public image of earprints reached a low point with the case of Mark Dallagher, who was accused of murdering an elderly woman and convicted in 1998 on the basis of earprint evidence. In 2004, when it was found that his DNA did not match the DNA that was recovered from the original earmark, all charges against him were formally dropped.

The possibility that the DNA evidence was contaminated should, however, not be ignored. The print was secured by officers who did not intend to perform a DNA analysis, therefore not taking precautions to prevent contamination. The equipment that was used could easily have been contaminated from the previous cases. Also, the original mark had been stored for years on a nonsterile surface before DNA was recovered, as Kieckhoefer *et al.* [5] pointed out.

Some restraint on the use of earprints for person identification is, however, advised. An earmark is usually left in a publicly accessible area before the crime itself is committed (e.g., on the *outside* of a building before a break-in). Furthermore, no accurate indication of the time of its formation may usually be inferred.

With respect to the Dallagher case, it was reported that the window on which the earmark was discovered had been cleaned three or four weeks before the murder took place [6]. The presence of his earprint at the window of the murdered woman could be

considered suspicious, and reason for questioning. Without additional evidence, however, it would not provide sufficient grounds for a conviction for murder as all it implied was that Dallagher listened at the house of the murdered woman within the three to four weeks period leading up to the murder.

A more fundamental point of criticism may be that the study of earprints as a quantitative and rigorous discipline is immature. Standard tools and methods for analyzing earprints are not yet available. As there is no fully automated system for comparison available, the process of individualization at this moment involves visual comparison of the crime scene print with control prints of a given suspect. In this process, one records similarities and differences. Van der lugt [7], in explaining the procedure, quotes Tuthill [8] in saying that one should form an opinion on whether or not the similarities are of such number and significance as to preclude the possibility of their having occurred by mere coincidence. One should further make sure that there are no differences, except "those that can be accounted for" [7]. But when are differences sufficiently small or insignificant as to claim that we can account for them? And when is the degree of similarity of such significance as to preclude the possibility of it having occurred by mere coincidence? Moreover, when can corresponding individual characteristics be said to *agree*? One should keep in mind that not even characteristics in two prints from the same ear will ever be identical. In practice, judging agreement means judging the degree of similarity between various characteristics. As degrees of similarity usually vary continuously, the term *agreement* is difficult to define objectively.

Indeed, subjectivity is an important factor that can diminish the evidential value of earprints. Human experts are required to make the final judgment about individualization, even when automatic systems are used to analyze the forensic data. On the issue of subjectivity in fingerprint individualization, Stoney [9] commented: "The modern image processing techniques used to classify fingerprints may provide an illusion of complete objectivity, yet only a list of most likely matches from a database are provided, and the expert will have to compare and make conclusions." This would also be the case for earprint individualization. Some level of subjectivity would therefore have to be accepted. It may, however, be mitigated using an appropriate program of training and proficiency testing. Progress has been made in developing quantitative methods for comparing large samples of earprints. The search is on for operator-invariant procedures that can be used to calculate the match probability for a combination of two prints. It is hoped that, by combining image-processing techniques and large, representative samples, tools can be developed that will assist experts in making objective statements on individualization. Finally, there is a growing consensus that this should preferably be done using probabilistic terms [10, 11].

Automatic Matching and the Probability of a Match

A number of initiatives toward (semi-)automated classification or matching of earprints have been undertaken. Image-processing algorithms to extract features from the anthelix area were presented by Valvoda [12], quoted from [11]. In 2005, Rutty *et al.* [13] presented their concept of a more elaborate "computerized earprint identification system". Their research, however, was based on a database containing 800 prints of 800 different ears, and therefore did not allow the possibility to verify if selected parameters offered, besides a high interindividual variability, also a sufficiently small intraindividual variability. In line with Ingleby *et al.* [14], Rutty *et al.* proposed to calculate centroids of imprinted areas. Parameters generated using the position of these centroids offer clues for individualization. In both the studies, it was assumed that the effect of variation in applied pressure would be overcome by the use of centroids. As we showed in Meijerman *et al.* [15], however, the imprints of morphological structures do not only narrow or widen due to a chance in applied pressure. To some extent, features can change position in relation to each other as well.

Partners in "FearID", an international research project aimed at the individualization of earprints, have also attempted to design a (semi-)automatic matching system for earprints. For this purpose, multiple earprints of 1231 different persons were collected and analyzed. To facilitate the detection of the various imprinted anatomical structures for analyses, the shape and position of the imprinted ear surface in each analyzed print was manually indicated. This was done by drawing a spiral-shaped polyline from the apex of the anthelix to the lowest point of the imprinted earlobe (Figure 1). The use of

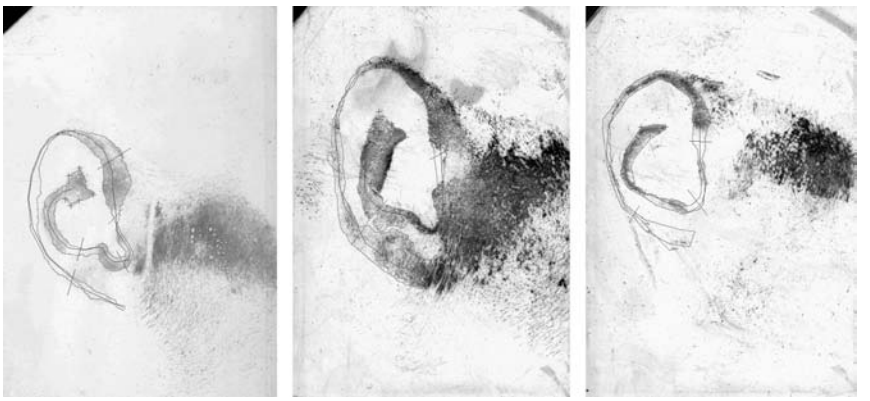


Figure 1 Imprinted earprint features captured in a spiral shape. The longitudinal center line is the indicator line, transverse lines indicate transitions between gross anatomical feature zones [Reproduced with permission from Ref. 23. © Barge's Anthropologica, 2006.]

image-processing techniques to find the imprinted ear surface meant that the manual labeling did not need to be extremely precise. We refer to Van Munster *et al.* [16] for further reading on image processing for earprint analysis.

A number of approaches to automatic classification and matching were then explored. Contours of the imprinted ear surface were used to calculate the width and curvature of the various parts in the spiral shape. This data was then compared among prints ("weighted width comparison" and "angular comparison"). Additional information for classification was provided by the variation in relative intensity along the spiral shape. The operator further annotated minutiae, landmarks, and other characteristics in each earprint, and indicated transitions between the imprinted gross anatomical features in the spiral shape. This generated additional data for comparison that was analyzed using a method called *vector template matching* [5]. The performance of the weighted width comparison of earprints appeared to be operator independent [17, 18]. Achieving interoperator objectivity in vector template matching based on manual annotations, however, appeared more problematic. Performance of the method appeared to vary significantly when prints were annotated by different operators [19].

In addition to the described semiautomated approaches, we have applied a method to match earprints fully automatically [20]. As the method requires no human input, it allows objective and repeatable earprint comparisons that are free from observer bias. Furthermore, it can handle tasks that are very time consuming for a human expert to consider. In this method, referred to as *keypoint matching*, algorithms automatically detect and describe the salient regions – "keypoints" – in an earprint and use them to match the corresponding anatomical features. Keypoints can be detected and described using various methods, but the experiments conducted so far have been conducted using the methods described by Lowe [21].

The method was initially tested using a sample of 36 right-earprints from six pairs of identical twins [20] and is shown in Figure 2. For each earprint in the sample, keypoints were located and described. The appearance and constellation of the described keypoints in each earprint could then be compared with those found in all the other prints of the sample. The number of matching keypoints between any two prints was chosen to indicate the level of similarity between these prints. These results could then be used to construct a ranking list, or "hit-list", for potential matches. In this setting, keypoint matching provides a tool for recovering matching earprints from a large database. The possibility to assess the evidential value of an assumed match between two earprints is another potentially valuable application of this method. There is a need for methods to describe the weight of evidence – in this case an earprint comparison – as a likelihood ratio, the numerator of the ratio being the probability of obtaining the evidence,

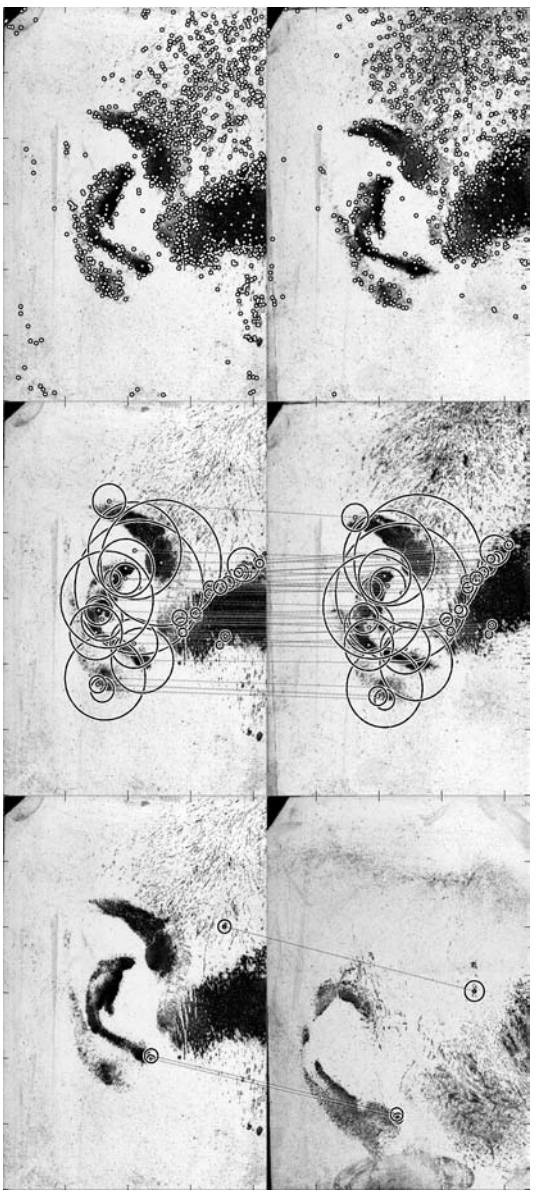


Figure 2 Keypoint matching illustrated. For reach pair of illustrations, the print on the left is the same. The upper pair shows all detected keypoints in two different prints of the same ear. For the middle pair dotted lines join matching keypoints. In the lower pair three keypoints matches to a print from a different ear are shown (note that prints do not need to be aligned) [Reproduced with permission from Ref. 23. © Barge's Anthropologica, 2006.]

given that the earmark was made by a certain person (usually the prosecution hypothesis), and the denominator being the probability of obtaining the evidence, given that the earmark was made by someone else (usually the defense hypothesis) [1]. The development of fully automatic methods for comparing earprints is an important step toward this goal. Firstly, because they are not the subject to observer bias and secondly because they make the task of analyzing large samples tractable and open the possibility of making reliable probabilistic statements based on large-number statistics. In the case of keypoint matching, similarity scores between earprints can be generated automatically and by using control samples it is possible to measure the frequency with which certain similarity scores are obtained when comparing the prints obtained from the same individual or from different individuals. Thus, the probability of obtaining a certain earprint given either the defense or prosecution hypothesis can be represented by the probability of obtaining a given similarity score given either hypothesis. The sample containing 36 prints from 6 pairs of identical twins is a good testbed for developing automatic methods because the earprints from twins show strong similarities and distinguishing individuals is challenging. However, it is not an appropriate sample for estimating likelihood ratios for routine earprint comparison. Firstly, it is by no means representative of earprint population in general and secondly, it only contains 12 individuals. Making more reliable estimates requires work on a large, well-defined sample. As a result of the FearID research project, a working sample containing three to four prints per ear of both ears of 772 persons is available for research. An additional validation sample further contains a similar number of prints per individual from another 459 persons. In a preliminary and unpublished study, using keypoint matching on a sample of 1552 ears showed promising results; when these prints are compared to a single print from a different, randomly chosen donor none show more than 10 keypoint matches, whereas when a different print from the same donor is used over 60% showed more than 10 keypoint matches. Further work is, however, required to refine the technique and gather rigorous statistics.

Intraindividual Variation

At this point, we would like to emphasize the importance of studying intraindividual variation in

addition to interindividual variation in prints. For an earprint to have evidential value in a forensic setting it needs to posses a set of features for which not only the interindividual rate of occurrence is low but the intraindividual rate of occurrence is high. In order to be able to test the latter, the research database of earprints should be compiled in such manner that realistic intraindividual variation is taken into account. In the following section, we briefly summarize the results from a number of studies exploring possible sources for variation in different prints from a single ear.

Changes in the amount of oils and waxes present on the ear surface due to, for instance, variation in outside temperature or whether the ear was recently cleaned or not could in theory influence the dimensions and/or intensity of the imprinted area. In turn, this might affect the area in which characteristics can be found. It may also affect the visibility of such details. We therefore collected multiple earprints from a number of ears. From each ear, prints were collected before and after the ear was cleaned. We then compared print-mass, which is a measure for the size and intensity of the imprinted area. No evidence for a significant decrease in the mass of prints created by cleaned ears was observed [22]. To determine if the imprinted details are of equal quality, it would require further investigations into the stability of characteristic features (i.e., valuable for individualization) in prints of recently cleaned ears.

Another – more often explored – source of intraindividual variation in earprints is the variation in pressure that is applied by the ear to the surface during listening [23–25]. We have tried to induce changes in applied force between different listening efforts of individuals. We varied the level of ambient noise while recording the force that was applied during listening, but no significant effect from the level of ambient noise on applied force was observed [26]. Listening to either sound or silence did also not appear to significantly affect applied force [27]. We did, however, note a significant effect from changes in the level of the target sound. When the sound level was reduced in between listening efforts, it appeared to cause the listeners to apply more force [26]. In addition, we observed that listeners generally applied less force during their first listening effort. We assumed that – unfamiliar with the procedure – our subjects were more cautious during first listening efforts.

It should be noted that equal variation in applied force did not necessarily lead to equal intraindividual variation in the prints. For some ears, small changes in force appeared to have a relatively large effect on the prints, while for other ears relatively large changes in force seemed to have little effect on the appearance of the prints [15, 26]. Prints of the same ear may further be affected by a change in pressure distribution [15, 26, 28]. The duration of listening also appeared to affect the appearance of an earprint. We found that the mass of prints significantly increased with length of listening [29]. Kieckhoefer *et al.* [28] showed that fidgeting of the ear during listening increased the amount of imprinted surface.

When compiling a research database, or collecting a number of control prints from a suspect, experience gained about the factors that affect the appearance of a print may be exploited to induce a broad range of realistic intraindividual variation. For example, applied listening force appears to be generally lower during a first listening effort. A loud target source during this attempt, or a relatively short duration of ear-surface contact, will increase the chance of yielding a print with a relatively low print-mass.

Finally, during various studies into the force applied while listening, we have noted that intraindividual variation in applied force is comparatively small when compared with the interindividual variation [26, 27]. This observation was confirmed by Kieckhoefer *et al.* [28]. To prevent the introduction of unrealistic intraindividual variation in a set of prints, which would likely impede the search for diagnostic features and lead to misrepresentations of the probability scores for a match between prints, we would advise that, if possible, reference prints of the suspects are "functional earprints", i.e., taken after actual efforts at listening.

End Notes

[a] LJN: BA3372, Gerechtshof 's-Gravenhage, 22-003548-06. http://ljn.rechtspraak.nl/BA3372 (http://zoeken.rechtspraak.nl/resultpage.aspx?snelzoeken=true&searchtype=ljn&ljn=BA3372&u_ljn=BA3372).

[b] Het laatste nieuws, Belgie. http://www.hln.be/hln/nl/957/Belgie/article/detail/301774/2008/06/05/Tot-zes-jaar-cel-voor-226-inbraken.dhtml.

[c] LJN: BC9536, Hoge Raad, 07/10104. http://ljn.rechtspraak.nl/BC9536. (http://zoeken.rechtspraak.nl/

resultpage.aspx?snelzoeken=true&searchtype=ljn&ljn=BC9536&u_ljn=BC9536).

References

[1] Champod, C., Evett, I.W. & Kuchler, B. (2001). Earmarks as evidence: a critical review, *Journal of Forensic Sciences* **46**(6), 1275–1284.

[2] Egan, T. (1999). *Are Dutch Ears Different from American Ears? A Comparison of Evidence Standards*, UMKC School of Law, Kansas City, http://www.forensic-evidence.com/site/ID/ID00004$_1$.html.

[3] Moenssens, A.A. (1999). Identifying individuals by ear photographs, and earprints – is this process reliable enough to justify expert opinion of identity in court? *Proceedings of the 1st International Conference on Forensic Human Identification in the Next Millennium*, London, October 24–26, 1999.

[4] State v. Kunze. Court of Appeals of Washington, Division 2. 97 Wash. App. 832, 988 P.2d 977 (1999). http://www.forensic-evidence.com/site/ID/ID_Kunze.html.

[5] Kieckhoefer, H., Ingleby, M. & Alberink, I. (2005). *Vector Template Matching of Earprints*, Research Report series (RR05 13d), School of Computing and Engineering, University of Huddersfield, Huddersfield, http://forensic.to/fearid/VTMfinal.doc.

[6] *Crownv Mark Anthony Dallagher*, In the Supreme Court of Judicature – Court of Appeal (Criminal Division), Neutral Citation No. (2002) EWCA Crim 1903, Case No. 2000/5024/Z2, London (2002).

[7] Van der Lugt, C. (2001). *Earprint Identification*, Elsevier Bedrijfsinformatie, The Hague.

[8] Tuthill, H. (1994). *Individualization: Principles and Procedures in Criminalistics*, Lightning Powder Company, Salem.

[9] Stoney, D.A. (1991). What made us ever think we could individualize using statistics? *Journal of Forensic Science Society* **31**(2), 197–199.

[10] Broeders, A.P.A. (2003). *Op zoek naar de bron. Over de grondslagen van de criminalistiek en de waardering van het forensisch bewijs*, Kluwer, Deventer.

[11] Champod, C. & Evett, I.W. (2001). A probabilistic approach to fingerprint evidence, *Journal of Forensic Identification* **51**(2), 101–122.

[12] Valvoda, J.T. (1999). *Otolobe – Earprint Recognition*, Ecole Polytechnique Fédérale de Lausanne (EPFL), Semester project report.

[13] Rutty, G.N., Abbas, A. & Crossling, D. (2005). Could earprint identification be computerised? An illustrated proof of concept paper, *International Journal of Legal Medicine* **119**(6), 335–343.

[14] Ingleby, M., *et al.* (2000). *Ear-prints: A Neglected Forensic Resource? Internal Report*, School of Computing and Engineering, University of Huddersfield, Huddersfield.

[15] Meijerman, L., Sholl, S., De Conti, F., Giacon, M., Van der Lugt, C., Drusini, A., Vanezis, P. & Maat, G.J.R (2004). Exploratory study on classification and individualization of earprints, *Forensic Science International* **140**, 91–99.

[16] Munster van, R.J., Antwerpen van, G. & Thean, A. (2005). *FearID Image Processing for Earprint Analysis: Final Report TNO Contribution to WP5, Internal FearID Report*, Netherlands Organization for Applied Scientific Research, Delft.

[17] Alberink, I.B. & Ruifrok, A.C.C. *Inter-Operator Test for the Clicking of Polylines in Earprints, Internal FearID Report*, Netherlands Forensic Institute, The Hague.

[18] Alberink, I.B. & Ruifrok, A.C.C. (2007). Performance of the FearID earprint identification system, *Forensic Science International* **166**, 145–154.

[19] Alberink, I.B., Ruifrok, A.C.C. & Kieckhoefer, H. (2006). Inter-operator test for anatomical annotation of earprints, *Journal of Forensic Science* **51**(6), 1246–1254.

[20] Meijerman, L., Thean, A., Van der Lugt, C., Van Munster, R.J., Van Antwerpen, G. & Maat, G.J.R. (2006). Individualization of earprints: variation in earprints of monozygotic twins, in *Inter- and Intra-Individual Variation in Earprints*, L. Meijerman, ed, Barge's Anthropologica, Leiden, pp. 139–159.

[21] Lowe, D. (2004). Distinctive image features from scale-invariant keypoint, *International Journal of Computer Vision* **60**(2), 91–110.

[22] Meijerman, L., Van der Lugt, C., Van Antwerpen , G., Van Munster, R.J. & Maat, G.J.R. (2006). Preliminary comparison of earprints that were made before and after cleaning the ear, in *Inter- and Intra-individual Variation in Earprints*, L. Meijerman, ed, Barge's Anthropologica, Leiden, pp. 111–117.

[23] Hammer, H.J. & Neubert, F. (1989). Experimentelle Untersuchungen zur Auswertung von Ohrabdrückspuren, *Kriminalistik und Forensische Wissenschaften* **73–74**, 136–139.

[24] Neubert, F. (1985). *Die Bedeutung der Täteridentifizierung durch Ohrabdrücke*, MSc thesis. Karl-Marx University, Leipzig.

[25] Saddler, K. (1996). *The Establishment And Evaluation of An Ear-Print Database At The National Training Centre for Scientific Support to Crime Investigation*, M.Sc. Thesis. University of Strathclyde, Glasgow.

[26] Meijerman, L., Nagelkerke, N., Van Basten, R., Van der Lugt, C., De Conti, F., Drusini, A., Giacon, M., Shell, S. Vanezis, P. & Maat, G.J.R (2006). Inter- and intra-individual variation in applied force when listening at a surface, and resulting variation in earprints, *Medicine, Science, and the Law* **46**(2), 141–151.

[27] Meijerman, L., Nagelkerke, N., Brand, R., Van der Lugt, C., Van Basten, R., De Conti, F., *et al.* (2005). Exploring the effect of occurrence of sound on force applied by the ear when listening at a surface, *Forensic Science, Medicine and Pathology* **1**(3), 187–192.

[28] Kieckhoefer, H., Ingleby, M. & Lucas, G. (2006). Monitoring the physical formation of earprints: optical and pressure mapping evidence, *Measurement* **39**(10), 918–935.
[29] Meijerman, L., Van Antwerpen, G., Van Munster, R.J. & Maat, G.J.R. (2006). Exploring the effect of duration of listening on earprints, in *Inter- And Intra-Individual Variation in Earprints*, L. Meijerman, ed, Barge's Anthropologica, Leiden, pp. 99–109.

L. Meijerman, A. Thean and G.J.R. Maat

Ecology *see* Environmental Science

Education and Accreditation in Forensic Science

Introduction

Forensic science is the quintessential multidisciplinary field. It incorporates methods, theories, and concepts from practically every science including chemistry, physics, biology, mathematics, statistics, and many disciplines of medicine. If a science has a public or legal application, it is a forensic science. On one level, education in forensic science could be accomplished by giving students a strong scientific background and then either allowing them to learn forensic science by applying this knowledge on the job, or by adding classes that teach the student how to apply scientific knowledge to the sometimes peculiar circumstances of scientific evidence analysis. For much of its recent history, especially in the United States, that has been the way the forensic science is taught. Many early "forensic science programs" were little more than a chemistry or biology degree with an internship in a crime laboratory. In some cases, the program was largely based on criminal justice, with little "hard" science coupled with a crime lab internship. These types of programs still exist but are becoming discredited.

Since forensic science is mainly applied science, what is wrong with teaching it at the college level this way; a science education with an internship? Actually there are several things wrong with this approach. First, and most important, it assumes that forensic science is "only" the application of science to criminal matters. In fact, there is much more to forensic science. It is part of the criminal justice system and the ultimate consumers of forensic science are judges, juries, prosecutors, defense attorneys, victims, and defendants. In addition to the science itself there are considerations of expert testimony, law and jurisprudence, ethics and quality assurance and control including accreditation of laboratories, certification of forensic scientists and proficiency and competence testing as well as development of standard methods of analysis for various types of evidence. One of the major shortcomings of forensic science has been and continues to be a lack of attention to the forensic part of forensic science education. Over the years, this has caused significant problems within the field. Many crime lab directors (the main consumers of many forensic science graduates) mistrust forensic science education because it does not deliver what it promises and they rely instead on graduates from "real" science programs and then train them on the job.

Another problem with the "hard science plus internship or on the job experience" model of forensic science education is that it denies the existence of a body of knowledge that is unique to forensic science. There are several disciplines that are unique to forensic science. They have few or no applications outside the criminal or civil law context. These include fingerprints and other friction ridges, handwriting and printed word analysis including inks and papers, and firearms and other tool marks. They are also (incorrectly) disparaged because they involve only physical comparisons and do not use the methods of scientific inquiry. It is no coincidence that there are few courses in the US forensic science education programs in these areas of forensic science.

Finally, the traditional method of forensic science education inhibits its recognition as a profession in its own right. Among the hallmarks of a profession are that it has a body of specialized knowledge and a strong educational foundation. If education programs do not recognize this and offer the forensic as well as

the science part of forensic science, then it will not be regarded or respected as a profession.

Fortunately this situation has been changing. There are an increasing number of forensic science education programs in the United States and other countries that offer both the hard science and forensic aspects of forensic science in their degree programs. There is now a movement to accredit forensic science education programs using consensus standards that contain a broad spectrum of topics that help define forensic science as a profession (see below). There are a number of reasons for these changes. One of the important catalysts has been the development of DNA typing. The stringent requirements of DNA methodologies and their history in medical science require that attention be paid to the forensic aspects of biological applications. The increase in public awareness of forensic science from TV shows, books, and movies has put more attention on the forensic sciences and increased interest among prospective students. This, in turn, has caused explosive growth in the number of forensic science education programs. Prospective students, crime lab directors, and criminal justice officials have demanded that there be some order and rigor in such programs. As will be demonstrated later, this has given rise to the forensic science education program accreditation movement that is now taking place in the United States.

Structure of Forensic Science Programs

The website of the American Academy of Forensic Sciences (AAFS) (www.aafs.org) currently lists more than 140 forensic science education programs in the United States. Some estimates place the number of programs in the United Kingdom at more than 300. There are both undergraduate and graduate forensic science programs worldwide. In the latter, the vast majority of programs are at the master's degree level. There are few doctoral degrees in forensic science worldwide.

Undergraduate Programs

There are many models for undergraduate forensic science degree programs. One of the more popular ones is to have a "concentration in forensic science" within a Bachelor of Arts or Science degree. Most commonly the degree is in chemistry, although some are in biology. A few are in criminal justice, but these are only a small fraction. A number of colleges and universities have a forensic science department or program and offer a "Bachelor of Science degree in Forensic Science". Some of these programs are called *Criminalistics*, others are *Forensic and Investigative Sciences,* others are *Police sciences* and there are other variations. These programs are characterized by their strong background in chemistry and biology with supporting forensic science lecture and lab courses as well as criminal investigation, law, criminal justice, etc. Internships and a research component are required or strongly encouraged in most or all of these programs. At this writing, the great majority of forensic science programs in the United States are on the undergraduate level as are most of the new programs. This is probably due to the lower cost and higher student count in undergraduate programs. They take advantage of existing science and criminal justice courses and, because research is not as high a priority as it would be in a graduate program, lab facilities tend to be less expensive.

Undergraduate forensic science curricula vary widely but most have an introductory course that may be open to all students, not just those in the major. It is not uncommon for such classes to attract hundreds of students. There are also usually a few forensic science laboratory courses that teach students how to apply scientific concepts to evidentiary materials and how to interpret the results. Many programs also have courses in crime scene investigation and those that cover ethics, quality assurance, crime lab culture, the law, etc.

Master's Degrees

Like bachelor's programs, master's degrees are also offered in a number of venues. Most of the master's programs today offer "Master of Science in Forensic Science". These programs may offer concentrations in "forensic chemistry", "biology", "computer forensics", "anthropology", and "other disciplines". The most common are chemistry and biology. Other institutions offer a concentration in forensic science within a science department such as a "Master of Science in chemistry with a concentration in forensic science".

Master's degree programs generally span 30–40 credit hours. Requirements for admission are usually

a bachelor's degree in a science or forensic science. There are relatively few graduate programs and competition for spots can be fierce. It is not unusual for students to have grade point averages (GPAs) (US GPAs) of over 3.5/4.0 in order to gain admission. Some programs attract more than one hundred applications for a dozen or so spots.

Since students in graduate programs have a strong science background when admitted, most of the coursework in the program is in forensic science and related areas. Since most of the students will not have a forensic science background there is at least one graduate-level introductory course in forensic science and at least one law and forensic science class. There are also advanced lab courses in the various science areas. Some programs have courses in one or more areas of criminalistics, such as firearms and tool marks or questioned documents. Most programs have a research component and it is common for a master's thesis to be required although, exceptions are often made for students who are already working in a forensic science lab. Internships are also available in most programs.

PhD Programs

There are very few doctorate programs worldwide at this writing. This may be due to a combination of factors including cost to the university, cost to the students, lack of a reliable source of research funding, lack of student fellowship and assistantship money, and scarcity of qualified faculty. Most crime lab systems do not recognize the value of having a PhD for a bench scientist job and do not pay graduates any more money than they would a qualified candidate with a BS or MS degree.

There is an increasing need for PhD graduates in forensic science. They are needed to staff the mushrooming forensic science education programs, to perform research in industry and government and to take on administrative roles in forensic science. Some universities are attempting to address these needs by offering a forensic science concentration within a science PhD program. At least one university offers qualified students the opportunity to obtain a PhD in a science and a master's degree in forensic science at the same time. The dual program takes no longer than the PhD alone because the two departments share elective courses and the master's thesis may be waived if the doctoral dissertation has a forensic science component.

Accreditation of Forensic Science Education Programs

In recent years, public attention to forensic science has increased dramatically owing in part to the proliferation of forensic science related television shows, movies, and books. This has resulted in unprecedented demand for forensic science education programs. The number of students who have expressed a desire to major in forensic science has exploded, prompting many colleges and universities to respond by creating numerous forensic science programs. At some institutions, forensic science is the most popular major on the campus. Many of these new forensic science programs were started hastily, by people who have no significant background in forensic science education or practice, and often without sufficient resources. The result has been a veritable hodgepodge of offerings; some very good and some deplorable. With more than one hundred offerings in the United States, it is difficult or impossible for students or crime lab directors to discern which programs are scientifically and forensically sound. In response to this crisis, the AAFS and the National Institute of Justice (NIJ) teamed up to formulate a solution: development of standards for undergraduate and graduate forensic science programs and creation of a mechanism of accreditation of programs that would measure them against these standards. In the United Kingdom, the Forensic Science Society has launched an accreditation scheme for higher education courses in forensic science. This scheme accredits only three main areas; crime scene investigation, laboratory analysis, and evaluation and presentation of evidence.

The Forensic Science Education Programs Accreditation Commission (FEPAC) was formed in 2002. Its policies and procedures manual, which can be found at www.aafs.org, describe its beginning and development as follows:

> The American Academy of Forensic Sciences (AAFS) was established in 1948 to promote education for and research in the forensic sciences; to encourage the study, improve the practice, elevate the standards, and advance the cause of the forensic sciences; to promote interdisciplinary communications; and to plan, organize, and administer meetings, reports, and other projects for the stimulation and advancement of these and related purposes.

An assessment of forensic sciences published in 1999 by the National Institute of Justice (NIJ) entitled *Forensic Science: Review of Status and Needs*, described the educational and training needs of the forensic science community as "immense." Among the recommendations contained in the report was the establishment of the following:

- National standards for education in forensic sciences,
- An independent, community-wide, consensus-building, standard-setting body such as a technical working group for education in forensic sciences, and
- An accreditation system for forensic science education programs.

The NIJ established a technical working group for education and training in forensic sciences (TWGED) in 2001 for the purpose of recommending sample curricular guidelines for educational programs in forensic sciences. The results of TWGED's deliberations were delineated in a research report published in 2003, entitled Education and Training in Forensic Sciences: A Guide for Forensic Science Laboratories, Educational Institutions, and Students.

Acknowledging the importance of an accreditation system for academic programs built on the foundation of TWGED, the AAFS in 2002 established an ad hoc committee, called Forensic Education Programs Accreditation Committee, to explore issues related to the development of such an accreditation system. In 2004, the committee became an official standing committee of the AAFS at which time its name was changed to Forensic Science Education Programs Accreditation Commission (FEPAC).

The Policies and Procedures manual of FEPAC also describes the Commission's role, mission, and composition as follows:

Authority. FEPAC is responsible for, and has final authority over, the standards used for the evaluation of college and university forensic science programs for purposes of accreditation, judgments involving compliance with those standards, and accreditation decisions.

Composition. FEPAC is composed of eleven Commissioners. Five Commissioners are forensic science educators, five are forensic science practitioners, and one is a representative of the public. All Commissioners (except the public member) must be either a Member or Fellow of the AAFS. To serve as a forensic science educator, an individual must be a faculty member or an administrator at a college or university that offers a FEPAC-accredited forensic science program. At least three of the forensic science practitioners must be members of the American Society of Crime Lab Directors (ASCLD). No two educators may be from the same institution, nor may any two practitioners be from the same agency.

The representative of the public may not be (1) an AAFS member; (2) an employee, member of the governing board, owner, or shareholder of, or consultant to, either a forensic science program or the institution in which the forensic science program is located; (3) a member of any trade association or membership organization that is related to, affiliated with, or associated with FEPAC; or (4) a spouse, parent, child, or sibling of any individual identified in (1), (2), or (3).

The FEPAC Chair may appoint committees to assist the Commission with its operational responsibilities, and these committees may include nonvoting members to the Commission.

Selection of Commissioners. FEPAC solicits nominations for positions on the Commission from its relevant constituencies, including forensic science programs, ASCLD, the Criminalistics Section of AAFS, regional forensic science organizations, forensic science certification bodies, and government agencies such as the NIJ, the National Institute of Standards and Technology (NIST), and the U.S. Department of Education. FEPAC reviews the nominations it receives and prepares from among the list of nominees a slate of individuals to serve on the Commission. FEPAC then recommends the slate to the President of AAFS, who appoints individuals to serve on the Commission after reviewing FEPAC's recommendation.

Chair of the Commission. The Chair of the Commission is elected annually by the members of the Commission and must be either a forensic science educator or a forensic science practitioner. The new Chair shall take office on the last day of the annual AAFS meeting. The Chair may be reelected.

Terms of Office. The term of office for all Commissioners is three years, with appointments made on a staggered basis. After serving one term, a Commissioner may be reappointed to a second three-year

term. However, no Commissioner may serve more than two consecutive three-year terms.

The ten individuals appointed by the President of AAFS in 2002 to be members of the Forensic Science Education Programs Committee will serve as the charter commissioners of FEPAC from February 2002 until February 2005. A public member will be appointed to the Commission by the President of AAFS during that period.

Beginning in February 2005 and continuing in February 2006 and February 2007, the President of AAFS will, after reviewing a slate recommended by FEPAC, appoint new Commissioners to replace the charter members of the Commission, according to the following schedule: In February 2005, one charter forensic science educator and one charter forensic science practitioner will be replaced. In February 2006 and again in February 2007, two of the charter forensic science educators and two of the charter forensic science practitioners will be replaced. All new appointees will serve three-year terms and may be reappointed to one additional term.

At this writing, 16 institutions have had their forensic science programs accredited by FEPAC and 10 more are in process. The Commission receives about a half dozen applications each year. A list of the accredited programs is maintained on the AAFS website. TWGED's publication on standards; "Education and Training in Forensic Sciences: A Guide for Forensic Science Laboratories, Educational Institutions, and Students", is also on this website.

The Future of Forensic Science Education

Assuming that the resources are available, the future trend in forensic science education will undoubtedly be toward more graduate-level programs. The demand for master's degree programs is very strong and expected to stay that way. The job market for graduates of all forensic science programs is strong and is especially so for master's graduates. There is also increased activity toward the formation of PhD programs and several should debut in the next few years. There are however, significant structural impediments that have been and will continue to hamper progress in forensic science education:

1. Qualified faculty

Since there are few PhD programs in place, there are very few qualified PhD forensic scientists being produced that could staff forensic science programs. There are only a handful of postdoctoral programs available for scientists who want to get some practical experience in forensic science. The FBI offers a visiting scientist program and the NIST offer postdoctoral opportunities. The Armed Forces Institute of Pathology and a few private labs offer postdoctoral training for biologists and medically trained scientists. The great demand for faculty is causing administrators to hire chemists and biologists with no forensic science experience and then getting them an "externship" in a crime lab to gain some experience. This is better than nothing but cannot adequately substitute for faculty with real crime lab experience.

2. Funding for graduate students and research

There has been a general trend toward disinvestment in scientific research by the Federal Government in recent years. The amount of funding for forensic science has always been extremely limited. It has gotten a bit better in recent years when the Government, through the NIJ, has been funding DNA technologies. Most of this money has gone to increasing capacity of crime labs to perform DNA typing, although some has also gone to research. The NIJ also provides some limited funds for general forensic science research and student support. This effort is hampered by the increased tendency to earmark the NIJ budget for pet Congressional projects and by competition from Federal laboratories for these scarce funds. A much larger investment in forensic science research and student support will be necessary if the quality of forensic science programs is to continue to increase.

3. Continued growth in forensic science education programs

At some point the number of graduates from forensic science programs will outstrip the demand for them in the job market. This is already occurring among graduates from bachelor's programs, who are becoming less competitive in the job market compared to master's degree graduates. FEPAC's accreditation efforts will hopefully help identify the best programs. There are also signs of Federal links developing that would give some monetary preferences to accredited institutions. This may help weed out the poorer quality programs over time.

Summary

Forensic science education is experiencing explosive growth today because of heightened public awareness brought on by increased attention by the media and by the publicity generated by real cases such as O.J. Simpson. This has caused a proliferation in educational degree offerings. Information from the United Kingdom suggests that some programs are undoubtedly a consequence of the need to attract students to science faculties with failing traditional scientific programs. Many of these programs are not legitimate forensic science degrees and do not serve the forensic science community well. Development of reputable programs is hampered by the lack of faculty resources and research and student support. An effort by the NIJ and the AAFS has resulted in the formation of a Commission that accredits forensic science programs against strict educational standards. This will give crime lab directors and prospective students some tools to use to evaluate forensic science educational programs for quality.

Further Reading

Atasoy, S., Coluglu, A.S., Abaci-Kalfoglu, E. & Polat, O. (1996). Postgraduate forensic science education in Turkey, *Journal of Forensic Sciences* **41**(2), 206–208.

Furton, K.G., Hus, Y.L. & Cole, M.D. (1999). What educational background do crime laboratory directors require from applicants? *Journal of Forensic Sciences* **44**(1), 128–132.

Higgins, K.M. & Selavka, C.M. (1988). Do forensic science graduate programs fulfill the needs of the forensic science community? *Journal of Forensic Sciences* **33**(4), 1015–1021.

Kobilinsky, L. & Sheehan, F.X. (1984). The desirability of a PhD program in forensic science, *Journal of Forensic Sciences* **29**(3), 706–710.

Lee, H.C. & Gaensslen, R.E. (1988). Forensic science laboratory/forensic science program cooperation and relationships: the view from the forensic science laboratory, *Journal of Forensic Sciences* **33**(4), 1071–1073.

Midkiff, C.R. (1988). College and University forensic science programs-graduate and undergraduate, *Journal of Forensic Sciences* **33**(3), 726.

Peterson, J.L. & DeForest, P.R. (1977). The status of forensic science degree programs in the United States, *Journal of Forensic Science* **22**(1).

Stoney, D.A. (1984). A medical model for criminalistics education, *Journal of Forensic Sciences* **33**(4), 1084–1086.

JAY A. SIEGEL

EEG *see* Seizures: Behavioral

Eighth Amendment *see* Death Penalty and Age

Elder Abuse: Policy

- Steven, 52 years old and unemployed, convinces his elderly, bed-bound aunt who has severe Arthritis to fire her home care attendant and hire him instead. While under Steven's care, she develops severe pressure ulcers.
- Edith, now 78 years old, has repeatedly been beaten and threatened by her husband throughout their 48-year marriage.
- Sheila persuades her elderly father, who has an advanced dementia, to sign a power of attorney, which he does not understand. She uses it to raid his savings account.
- Posing as a government worker, a man in Canada contacts elderly Americans claiming that he needs information to update files. He uses the information to empty their bank accounts.
- Vivian, who lives in a skilled nursing home, is sexually molested by a nursing assistant.
- A 92-year-old woman provides care to her bed-bound, 103-year-old sister. Years earlier, the sisters had made a pact, each promising never to let the other be institutionalized. When the younger sister runs errands, she locks the older sister in their apartment.
- Janet, who is in her 50s, applies for a volunteer job visiting frail elders. Two months later, she marries one of the clients, Rod, who is 87. Janet quits her job, replaces Rod's children as the beneficiaries of all his investments, and prevents them from visiting or talking to him by phone.

These scenarios demonstrate the varied ways that elders suffer harm by others. But do they all constitute "elder abuse"? Surprisingly, the answer depends on whom you ask.

Although elder abuse has been recognized for over 25 years, there is still widespread disagreement

about what it is and is not. Consider the following definition of elder mistreatment created by a panel of experts convened by the National Academy of Sciences (NAS) in 2002 to review and evaluate the research on elder abuse:

> *(a) Intentional actions that cause harm or create a serious risk of harm (whether or not harm is intended) to a vulnerable elder by a caregiver or other person who stands in a trust relationship to the elder, or (b) failure by a caregiver to satisfy the elder's basic needs or to protect the elder from harm* [1].

As definitions of elder abuse go, the Council's is relatively narrow. It requires that elderly victims be "vulnerable", which means that they have cognitive, physical, or communication deficits that compromise their independence and judgment. Applying the requirement, therefore, eliminates acts of abuse against "well elders" who have no such impairments. Some researchers, agencies, and governments include the vulnerability requirement; others do not. The Council's definition further limits mistreatment to situations in which abusers are caregivers or in "trust relationships" with their victims, which includes family members, friends, acquaintances, paid caregivers, and fiduciaries. This requirement excludes abuse by strangers.

Unlike the Council's definition, which requires that the abuse be intentional, other definitions cover situations in which harm results from recklessness, ignorance, or lack of resources, knowledge, or skills. Some definitions require that the mistreatment be ongoing as opposed to single acts, and some require that the conduct result in significant physical, financial, or emotional injury.

Variations in how elder abuse is defined account in part for the widely different incidence and prevalence estimates found in the literature as well as the divergent and sometimes conflicting profiles of victims, abusers, and risk factors. The National Council estimated that between one and two million Americans aged 65 or older have been injured, exploited, or otherwise mistreated [1].

The definitions used to classify the various subcategories of elder abuse also vary widely. They too may or may not require that victims have impairments, that perpetrators be in positions of trust, that the abuse be intentional and ongoing, or that it have significant consequences.

- *Financial abuse* ranges from simple theft to complex financial manipulations [2]. Much of the literature on financial abuse focuses on situations in which persons in positions of trust coerce, manipulate, or trick elders with diminished capacity into surrendering assets. Perpetrators may, for example, persuade these elders to sign checks, deeds, wills, or powers of attorney that benefit themselves. Perpetrators may also gain access to elders' assets through guardianship, marriage, or adoption. When the "vulnerability" and "trust relationship" requirements are absent, financial elder abuse may include theft or fraud by strangers, including predatory lending, telemarketing scams, or identity theft against well elders [3].
- *Physical abuse* generally refers to intentionally or recklessly causing bodily injury, pain, or impairment. Examples include striking, pushing, burning, and strangling elders, and using physical or chemical restraints. Several specific forms of physical elder abuse have been the focus of recent attention, including domestic violence, homicide/murders, and suicide/homicides. An emerging body of research suggests that older women are likely to experience domestic violence, which may have begun earlier in life or may begin or escalate in old age. For others, the violence begins when they enter into new relationships [4, 5]. In some cases, the violence is linked to such age-related factors as retirement and heightened dependency.
- *Elder homicides and murders* are likely to be concealed or staged to look like deaths by natural causes, suicides, or accidents [6, 7]. The cause of death may be suffocation, strangulation, starvation, neglect, overmedication, undermedication, drowning, causing someone to fall, poisoning, or arson. *Elder homicide-suicides* typically involve elderly men killing their spouses or intimate partners and subsequently committing suicide [8]. Because these killings are likely to be prompted by the physical decline, hospitalization, or institutionalization of either partner, they may be mistaken to be "double suicides" or "mercy killings" whereas closer scrutiny reveals that one partner was not a willing participant.
- *Sexual abuse* is nonconsensual sexual contact of any kind with an older person. It includes rape; molestation; lewd or lascivious conduct; coercion

through force, trickery, or threats; or sexual contact with any person who lacks sufficient decision-making capacity to give consent [9–11]. Elderly victims of sexual abuse are most often females and are likely to have impairments that make them dependent on others. Abusers include spouses and intimate partners, other relatives, paid caregivers, and acquaintances.

- *Verbal or psychological abuse* includes the use of words, acts, or other means to cause fear, humiliation, emotional stress, or anguish. Victims may be threatened with violence, deprivation, or institutionalization; or they may be berated, infantalized, humiliated, ridiculed, cursed, ignored, or isolated.
- *Neglect* is the failure of any person who has responsibility for an elder to provide the level of care that a reasonable person in a like position would provide. It includes failure to provide medical, health, or mental health care; to assist in personal hygiene; to prevent malnutrition or dehydration; or to protect against health and safety hazards. Because victims of neglect depend on others for care, it can logically be assumed that they have mental or physical disabilities. Perpetrators include family caregivers, paid attendants, long-term care facilities, and others who have a "duty" to provide care. Some researchers and practitioners distinguish between unintentional or intentional neglect, with the former resulting from such factors as caregivers' lack of resources, physical strength or stamina, emotional stability, maturity or skills to meet elders' needs. Intentional neglect is when perpetrators withhold needed care out of malevolence or malice or for financial gain (e.g., perpetrators want to hasten elders' death because they stand to inherit).

Other forms of elder abuse include the *violation of basic human rights, abduction,* and *abandonment*. Human rights include the right to privacy, to confidentiality, to associate with whomever one chooses, to exercise choice, and to refuse psychotropic medications or involuntary confinement. *Abduction* includes taking elders from their residences and preventing them from returning through force, coercion, or undue influence. *Abandonment* is when caregivers willfully desert or forsake elders under circumstances in which reasonable people would continue to provide care. Elders may be left unattended in public settings or hospital emergency rooms. Caregivers may leave elders alone without adequate provisions, quit, or move away without arranging for substitutes. Multiple forms of abuse are often found to coexist.

Abuse in Long-Term Care Facilities

Although most discussions of elder abuse focus on abuse in domestic settings, abuse also occurs in long-term care facilities [12]. It includes acts of violence, neglect, psychological abuse, or financial abuse by employees, visitors, and other residents; but often refers to management practices that endanger residents. These include facilities' failure to provide adequate staff, to screen or supervise employees, or to protect residents from abusers. In these situations, culpability may rest with supervisors, management, or corporate entities. Facility-related abandonment or abduction includes the discharge of patients into unsafe situations.

Detecting and Evaluating Abuse

Detecting and evaluating abuse and neglect can be extremely complex, and each type poses specific challenges. A common barrier is victims' unwillingness to disclose what has happened to them out of fear, shame, or loyalty to their abusers. Others are unable to describe what has happened as a result of physical, cognitive, or communication deficits. There are few reliable diagnostic or forensic markers. For these reasons, examiners typically look for constellations of factors, including physical evidence, victims' and abusers' behaviors and interactions, and the plausibility of explanations that are offered.

Many of the frequently cited indicators of physical abuse have been extrapolated from research on general populations. They include fractures to the head, neck, and spine [13]; injuries consistent with being grasped, squeezed, or restrained [14]; and signs of strangulation, including neck pain, soreness, petechiae, raspy voice, difficulty in swallowing, lightheadedness or head rushes, fainting and unconsciousness, red eyes, ligature marks, and loss of control of bodily functions [15]. Other commonly cited indicators include abrasions, defensive injuries, multiple bruises in various stages of healing, multiple skin

tears in locations other than arms and legs, and spiral fractures with a rotational component and in areas other than wrists, hips, or vertebrae [16].

Because elders bruise more easily than young people, fall more often, and are more likely to have illnesses and conditions that mimic abuse and neglect, there has been mounting attention in recent years to exploring abuse markers and evidence-based data specifically for elders.

A study that documents the occurrence, progression, and resolution of "natural" bruises, conducted by the University of California, Irvine (UCI), provides baseline data that can serve as the basis for evaluating nonnatural bruising [17]. The study revealed that bruises on the palms, soles of feet, face, neck, chest, abdomen, or buttocks are unlikely to occur accidentally among elders and that the color of bruises, previously thought to be helpful in "dating" injuries, has not been found to be consistent with elders. A study of burns in the elderly, handicapped, and disabled populations suggests that the markers of inflicted burns are similar to those found in child abuse and include characteristic patterns, a story that does not fit the injury, and a delay in seeking medical care [18].

Another study to identify potential markers of abuse examined coroners' reports of elderly nursing home residents in Arkansas after the state passed a rule requiring nursing homes to report all deaths to local coroners. The study identified four categories of markers that pertained to (i) patients' physical condition and quality of care (e.g., untreated injuries; multiple and untreated pressure ulcers, malnutrition; (ii) facilities' characteristics' (e.g., unchanged linens, strong odors; (iii) inconsistencies (e.g., between documented and observed care; and (iv) staff behaviors (e.g., staff who follow investigators too closely, lack of knowledge and concern about residents [19].

Victims' and perpetrators' behaviors are also important factors in assessing physical abuse (*see* **Elder Abuse: Risk**). Physical abuse may, for example, be revealed by frequent emergency room visits or implausible or conflicting explanations for how injuries were sustained. In the case of elder homicides, for examples, perpetrators may arrange for the hasty cremation of decedents to avoid detection. Investigations of homicide-suicides may include psychological autopsies (*see* **Psychological Autopsy**) to disclose whether both partners had suicidal tendencies.

Evaluating neglect is complicated by the fact that the symptoms of natural diseases may resemble abuse and neglect. Pressure ulcers (also known as *bedsores* or *decubitus ulcers*), which are lesions caused by pressure that results in damage to the underlying tissue, are a key indicator in neglect cases but there is disagreement about whether they can be prevented even in the best of circumstances [20]. In evaluating ulcers, investigators therefore often consider how they are cared for, and review medical histories and records, which is critical.

Evaluating neglect by caregivers may also be complicated by legal considerations. As described earlier, neglect is by persons who have a duty to provide care. That duty may be created by contractual arrangement (the caregiver is paid) or by virtue of a relationship (the parties are married) and persons who do not have a "duty" to provide care cannot be held civilly or criminally liable for neglect. However, the extent to which family members and others are legally responsible has not been clearly established. Courts have determined that people have a duty to provide care to spouses who are incapable of deciding whether they need help [21] and, in a precedent-setting case, the California Supreme Court[a] determined that an adult child is only criminally liable for a parent's care where the duty is affirmatively accepted.

In the case of verbal and psychological abuse, assessment is complicated by the fact that actions or words that can be extremely traumatic to some, may not be damaging to others. While these differences stem from personal differences and different norms within families, it appears that culture also plays a role. For example, a study comparing cultural attitudes toward various forms of abuse found that Caucasian elders were more tolerant of verbal abuse than elderly Korean Americans [22].

Financial elder abuse cases can be extremely complicated to detect and investigate [23, 24]. Victims may not be aware that they have been abused or refuse to cooperate in legal proceedings. They may not discover abuse until long after it has occurred. Determining whether or not elders have the requisite capacity to make financial or legal transactions can also be extremely complicated, as described later in this chapter. This may be especially difficult if it requires recreating what happened in the past and proving what an elder with diminished capacity understood at an earlier period of time. Critical

documents may be in the possession of abusers or financial institutions [25].

Evaluating whether or not abusers are in positions of trust can also be complex. Some, who are strangers to begin with, befriend elders intending to exploit them. For example, "sweetheart scams" are crimes in which offenders use deceptive romantic overtures to gain access to older peoples' assets. Even perpetrators who contact their victims by phone and never meet face-to-face may establish trust and confidence through repeated contacts, by playing upon their victims' loneliness or sympathy, and by endearing or ingratiating themselves [3].

Causes and Risk Factors

The reasons abusers commit abuse are as varied as the definitions and examples offered earlier (*see* **Elder Abuse: Risk**). They include the following:

- **Financial motives**

In many abuse cases, the underlying motive is greed, a sense of entitlement, or financial exigency (perceived or real). Offenders may perceive elders, particularly elders with diminished capacity, as easy targets for exploitation, with some experts even suggesting that certain perpetrators who are not criminally inclined to begin with decide to abuse when they gain access to elders' assets and realize that the abuse is unlikely to be discovered [26]. Caregivers may feel that they deserve more compensation than they receive, and family members who stand to inherit may believe that an elder's assets are "almost" theirs anyway. Financial motives are not limited to financial abuse [27]. Perpetrators may physically abuse or intimidate elders to get them to surrender assets, or neglect those in their care to render them compliant or hasten their decline or death if they stand to inherit victims' estates.

- **Interpersonal conflict**

The role of interpersonal conflict in elder abuse is not well understood. There is evidence to suggest that some abuse by caregivers stems from unresolved conflicts that started before the onset of disability (see "Caregiver issues"). In some cases, the motivation may be revenge or "pay back" for previous abuse [28, 29].

- **Perpetrators' dysfunction**

Perpetrators may have mental illnesses; personality disorders; or addictions to drugs, alcohol, or gambling. Sexual abusers include "gerontophiles" with specific sexual inclinations toward the elderly [30], some of whom seek out employment in "victim-rich" environments like long-term care facilities [31]. Sexual predators are also increasingly being admitted to nursing homes as residents [31, 32].

- **Caregiving issues**

Caregivers may abuse or neglect because they are under severe stress, exhausted, inexperienced, or reluctant to perform the caregiving role. They may lack skills, empathy, or understanding [33]. Although it was previously assumed that stress was associated with the level of care provided, many now believe that other factors play a greater role in predicting abuse. These include poor "premorbid relationships" between caregivers and care receivers (the relationship was poor prior to the onset of disability) [34, 35], verbal or physical abuse by care receivers, carers' level of anxiety, and carers' perception that they are not receiving adequate help [34, 36].

- **Power and control**

The drive for power and control has long been recognized as a primary motive for domestic violence. Perpetrators, whose power stems from physical dominance and social privilege, use violence or the threat of violence to subjugate their intimate partners. In recent years, it has been recognized that elders also experience domestic violence for these reasons [37]. Some suicide-homicides appear to be motivated by perpetrators' need to exercise power and control, and, in one-third of cases, there was a history of domestic violence [8].

Perpetrators of elder abuse may also exercise power and control for financial gain. "Undue influence" is when individuals who are stronger or more powerful get weaker people to do things they would not have done otherwise, using various techniques or manipulations over time. They may isolate the weaker person, promote dependency, or induce fear and distrust of others. The abuser tries to convince the vulnerable person that friends, family members, or caregivers have malevolent motives and cannot be trusted. This is sometimes referred to as creating a "siege mentality" [38, 39].

- **Dependency**

Perpetrators are likely to be dependent on the people they abuse [1, 40]. Adult children who abuse their elderly parents are likely to be unemployed, unmarried, and dependent on their victims. Some studies have found that victims who are dependent on their abusers are also at heightened risk [1] with some proposing that a "mutual web of dependency ' may exist wherein victims depend on their abusers for care, and abusers depend on their victims for financial or emotional support and housing [41].

The National Academy group reviewed factors that heighten the risk of abuse and neglect. They divided them into categories on the basis of the extent of supporting evidence. Risk factors validated by substantial evidence include a shared living arrangement, social isolation, victims' dementia, and pathological characteristics of abusers, including mental illness, hostility, alcohol abuse, and abuser dependency [1]. "Possible" risk factors include gender (agency samples universally find that most victims are women although it is not clear whether this is accounted for by the fact that there are more women in the elderly population or that they are actually at greater risk), and being Black [42, 43]. Contested risk factors include victims' physical impairment and dependence on their abusers, and intergenerational violence within families [1]. The risk factors associated with institutional abuse by staff include burnout, aggression from residents, negative attitudes toward residents, and age (perpetrators tend to be younger than nonabusive employees). Other contextual factors include such factors as staff shortages [44].

Consequences

Elder abuse may affect victims' psychological, physical, and financial well-being. The consequences of physical abuse include fractures, depression, dementia, malnutrition, and death. Early studies suggested that nearly 50% of all reported incidents of abuse resulted in physically apparent trauma [45]. A pioneering study of the impact of abuse and neglect on victims' physical status suggests that victims are more likely to die prematurely [46]. Some states have acknowledged that the impact of violence against elders is likely to be greater than for younger persons by enhancing penalties for violence against elders.

Although neglect is sometimes viewed as less serious than acts of commission, its impact can be extremely serious, even leading to death. Failure to provide adequate nourishment to frail individuals can quickly lead to dehydration or malnutrition. When left untreated, pressure ulcers may cause sepsis. Failure to provide adequate assistance or medication can lead to accidents or illnesses becoming worse.

Financial abuse can have devastating consequences because replacing lost assets is generally not a viable option for retired individuals or those with physical or mental disabilities [47] and the depletion of assets may result in elders becoming dependent upon family members or public assistance. Psychological consequences include the loss of trust in others or in one's own abilities, stress, isolation, depression, hopelessness, or even suicide [48].

Emotional reactions to abuse include anger, helplessness, reduced coping, sleep disturbances, eating problems, denial, fear, anxiety, feelings of learned helplessness, alienation, guilt, shame, and posttraumatic stress syndrome [49]. Although abused elders are more likely to experience depression or psychological distress, there is no way of knowing whether these conditions were present before the abuse started or were the result of the abuse [50–52]. In comparing older and younger female trauma victims, however, Acierno and his associates [53] found a lower incidence of depression and posttraumatic psychopathology among older victims.

Capacity and Consent

Victims' mental capacity and ability to give consent are critical considerations in elder abuse cases (*see* **Capacity to Consent to Medical Treatment**; **Capacity for Independent Living**; and **Guardianships of Adults**). In particular, determining whether clients have decision-making capacity may be a critical consideration in evaluating and proving abuse. For example, proving financial abuse may involve demonstrating that a perpetrator induced an elder to sign a contract for which he or she lacked the requisite capacity. Victims' capacity further dictates the extent to which they can participate in planning for their own care and the interventions that are needed and available to them.

Current thinking acknowledges that mental capacity is a cluster of mental abilities that people use

in their everyday lives, and that different skills are needed for different mental tasks. Decision-making capacity refers to people's ability to make and communicate decisions, understand the consequences of their actions, and act in their own self-interest. It is also task-specific. Determining whether someone has decision-making capacity requires looking at the decision in question and whether the person has the mental skills that are needed to make it.

Professionals are generally in agreement about how to evaluate decision-making capacity for certain decisions but not others. For example, there is general agreement among lawyers about "testamentary capacity", which is the capacity needed to execute wills: people must understand what a will is, have a basic plan for distributing their assets to heirs, know the nature and extent of their assets, and be able to identify or recognize potential heirs and beneficiaries. There is less agreement, however, about how to evaluate other decisions that are commonly questioned in elder abuse cases. These include the capacity to get married, give gifts, consent to sexual relations, and select and supervise home care workers.

One form of decision-making capacity is "consent", which is agreeing to actions, transactions, or services proposed by others. It requires that the person understands the act or transaction and is acting freely and voluntarily, and is not under the influence of threats, force, or duress. In the case of informed consent for financial decision–making, for example, the person making a transaction must be provided with information and have the mental capacity to understand and appreciate it. Furthermore, they must be acting voluntarily and be free from coercion [38].

Services and Interventions

The primary response to elder abuse in the United States was patterned after the response to child abuse. Overwhelmingly, states passed laws requiring those who are likely to discover elder abuse to report to adult protective services (APS), which investigate and substantiate reports, and either provide follow-up services or refer clients to other agencies that can.

While the number of elder abuse cases reported to APS program has risen substantially, suggesting that the system is successful in providing victims with access to help, there have been problems associated with it. Unlike their colleagues in CPS, APS workers can only intervene with the consent of clients unless one of the two conditions apply: (i) when abuse constitutes criminal conduct, it must be reported to the police; and (ii) when clients lack capacity to consent to services and the potential negative consequences of failure to act are high, involuntary interventions may be needed. While these exceptions may seem straightforward, they are not. As described in the last section, there are no universally accepted standards for how to determine when someone has or lacks capacity to consent to services. As more cases are handled criminally, many service providers are beginning to recognize when abuse constitutes criminal conduct and they must, therefore report to police.

Most abuse cases, however, do not fall within these two categories, which means that victims must voluntarily agree to services. Many refuse help out of fear, shame, or loyalty toward their abusers. In the last 15 years, as practitioners have come to better understand the formidable social, physical, cultural, and financial barriers that victims face in seeking help, they have become more successful in offering help. Drawing from the fields of domestic violence, substance abuse, victim advocacy, and psychology; they have become more adept at engaging clients who were previously dismissed as "resistant" or "reluctant". Clinicians and advocates have come to better understand how perpetrators exercise power, control, and undue influence; and how to counter these forces. Recognizing that many victims choose to remain in violent relationships, practitioners in the field of elder abuse prevention have borrowed approaches from the field of domestic violence such as safety planning, options counseling, shelters, and support [52, 54–56].

Common services needed by victims are listed below. Their availability varies significantly across the country [57].

- **Shelters**

Elderly victims may need shelter in a variety of circumstances. Some require safe haven to avoid further victimization. Others need shelter when they have been evicted from homes or apartments, abandoned by caregivers, when abusive caregivers have been fired or arrested, when essential utilities have been discontinued, or when their homes are unsafe or unhealthy as a result of abuse or neglect. A variety of shelter options have been designed, including rooms in battered women's shelters that have been adapted

for women with disabilities, temporary stays in residential care facilities or senior apartment houses, and free-standing shelters that have been designed specifically for elderly victims.

- **Services for caregivers**

The risk of abuse by caregivers can be reduced by enhancing caregivers' skills, providing them with information about disease progression and how to manage difficult behaviors, and reducing stress through services like respite care, support groups, counseling, and financial relief.

- **Counseling**

Group or individual counseling may be needed to alleviate the immediate and long-term traumatic stress associated with abuse, provide emotional support, assist victims explore their options, and address such issues as codependence, depression, and diminished self-esteem.

- **Emergency funds**

These funds may be needed for food, emergency caregivers, mortgage payments, transportation, utilities, locks to secure victims' homes, court filing fees, and repairs and relocation costs.

- **Legal assistance**

Victims may need help to secure orders of protection, annul bogus marriages and adoptions, sue for civil recoveries, create or revoke powers of attorney that have been misused, handle guardianships (a legal proceeding in which courts appoint individuals or agencies to manage the personal and/or financial affairs of vulnerable people who lack sufficient mental capacity to manage on their own (*see* **Guardianships of Adults**).

- **Victim witness assistance programs**

These programs help victims whose cases are in the criminal justice system. They provide information to victims about the court process and the status of their cases; court accompaniment; and assistance in securing compensation, restitution, and community services.

- **Mental health assessments**

These assessments are often needed to determine if elders are capable of meeting their basic needs, making decisions about services, entering into contracts, offering testimony, and protecting themselves against abuse. Assessments of alleged abusers' mental status are sometimes needed to determine if they pose a danger to others and need treatment. Assessments range from simple "shorthand screening tools" to comprehensive batteries of tests.

- **Support services**

These services decrease vulnerability to abuse and neglect by enhancing the independence of elders with physical and cognitive limitations and reducing their reliance on others. Examples include daily money management, meals, attendant care, adult day centers, friendly visitors, and telephone reassurance programs.

Multidisciplinary Teams

The diversity and complexity of abuse cases makes it unlikely that any single agency has all the resources, services, or expertise needed to handle all situations. For that reason, many communities have organized multidisciplinary teams (MDTs), which provide a forum for professionals from diverse disciplines and agencies to discuss difficult abuse cases; learn what services, approaches, and resources are available from other agencies and disciplines; share information and expertise; identify and respond to systemic problems; and ensure offender accountability.

Typically, teams include health and social service providers, law enforcement personnel, Ombudsmen, mental health care providers, physicians, advocates for persons with developmental disabilities, lawyers, domestic violence advocates, financial institutions, money managers, case managers, and many others.

In the last decade, specialized teams have emerged in response to the increasingly diverse and complex types of abuse being reported [58, 59]. Several communities have formed financial abuse specialist teams (FASTs), which include members with expertise in such areas as real estate, insurance, banking practices, investments, trusts, estate and financial planning, and other financial matters. Elder fatality review teams, which were patterned after child and domestic violence fatality review teams, evaluate injuries and causes of death, attempt to distinguish accidental from nonaccidental deaths, shed light on events leading up to deaths, identify systemic problems and aid in prosecutions. Teams include coroners/medical examiners, law enforcement (local, state and, in some situations, federal), prosecutors, state agencies that oversee long-term care facilities, and others. A few

communities have developed teams with a medical focus.

The Legal System Responds

Although elder abuse was traditionally viewed as a social service concern, there has been a trend in recent years to hold perpetrators accountable by strengthening and extending the use of civil and criminal remedies and penalties. States have enhanced penalties for crimes involving elders, passed laws that make abuse cases more financially viable for attorneys, and extended domestic violence laws to cover some situations of elder abuse. Police departments and prosecutors' offices have developed specialized elder abuse units. More cases are also being seen in civil courts.

To improve elders' access and safety, courts have designed safe and comfortable waiting areas, instituted special calendars, and assigned court staff to provide extra assistance to elders. The Alameda County (California) Elder Protection Court, established in 2002, features a separate calendar (in the late morning) for elders seeking restraining orders. In some situations, when elders are unable to appear in court, the proceedings are conducted in chambers, with a judge issuing orders by telephone. Self help clinics and services provide access to courts by people who cannot afford lawyers. San Diego's Family Justice Center, which houses police, prosecutors, forensic experts, and victim advocates from public and private nonprofit agencies in the same building has in recent years begun to serve abused elders. The first forensic center designed exclusively for elder abuse was launched in 2003 at the University of California, Irvine's College of Medicine [60]. The center has been replicated in other communities.

Conclusion

Since elder abuse first came to light in the late 1970s, researchers, policy makers, advocates, and practitioners have struggled to respond to this highly complex and multifaceted problem. Drawing from the fields of child abuse, domestic violence, family caregiving, and victim advocacy; communities have fashioned a variety of new services, policy, and practices. As new research emerges and practitioners gain insight and expertise in handling cases, the challenge now is to mold these disparate elements into a comprehensive, coherent, and cohesive service response system.

End Notes

[a.] People v. Heitzman (1994) 9 Cal. 4th 189.

References

[1] National Research Council (2003). *Elder Mistreatment: Abuse, Neglect, and Exploitation in an Aging America*, National Academies Press, Washington, DC.

[2] Rabiner, D.J., Brown, D. & O'Keeffe, J. (2004). Financial exploitation of older persons: policy issues and recommendations for addressing them, *Journal of Elder Abuse and Neglect* **16**(1), 65–84.

[3] Deem, D., Nerenberg, L. & Titus, R. (2007). Victims of financial crime, in *Victims of Crime*, 3rd Edition, R.C. Davis, A.J. Lurigio & S. Herman, eds, Sage Publications, Thousand Oaks.

[4] Bonomi, A.E., Anderson, M.L., Reid, R.J., Carrell, D., Fishman, P.A., Rivara, F.P. & Thompson, R.S. (2007). Intimate partner violence in older women, *Gerontologist* **47**(1), 34–41.

[5] Seaver, C. (1996). Muted lives: older battered women, *Journal of Elder Abuse and Neglect* **8**(2), 3–21.

[6] Collins, K.A. & Presnell, S.E. (2006). Elder homicide: a 20-year study, *American Journal of Forensic Medicine and Pathology* **27**(2), 183–187.

[7] Falzon, A.L. & Davis, G.G. (1998). A 15 year retrospective review of homicide in the elderly, *Journal of Forensic Science* **43**(2), 371–374.

[8] Cohen, D. (2000). An update on homicide-suicide in older persons: 1995–2000, *Journal of Mental Health and Aging* **6**(3), 195–197.

[9] Burgess, A.W. & Phillips, S.L. (2006). Sexual abuse, trauma and dementia in the elderly: a retrospective study of 284 cases, *Victims and Offenders* **1**(2), 193–204.

[10] Jeary, K. (2004). Sexual abuse of elderly people: would we rather not know the details? *Journal of Adult Protection* **6**(2), 21–30.

[11] Teaster, P.B. & Roberto, K.A. (2004). Sexual abuse of older adults: APS cases and outcomes, *Gerontologist* **44**(6), 788–796.

[12] Hawes, C. (2003). Elder abuse in residential care settings: what is known and what information is needed? in *Elder Mistreatment: Abuse Neglect, And Exploitation in an Aging America*, R.J. Bonnie & R.B. Wallace, eds, National Academies Press, Washington, DC, pp. 446–500.

[13] Fanslow, J., Norton, R. & Spinola, C. (1998). Indicators of assault-related injuries among women presenting to the emergency department, *Annals of Emergency Medicine* **32**, 341–348.

[14] Brogdon, B.G. (1998). *Forensic Radiology*, FL: CRC Press Boca Raton.
[15] Strack, G.B., McClane, G.E. & Hawley, D. (2001). A review of 300 attempted strangulation cases, part I: criminal legal issues, *The Journal of Emergency Medicine* **21**(3), 303–309.
[16] Dyer, C.B., Connolly, M.T. & McFeeley, P. (2003). The clinical and medical forensics of elder abuse and neglect in *Elder mistreatment: Abuse, Neglect, and Exploitation in an Aging America*, R.J. Bonnie & R.B. Wallace, eds, National Academies Press, Washington, DC, pp. 339–381.
[17] Mosqueda, L., Burnight, K. & Liao, S. (2005). The life cycle of bruises in older adults, *Journal of the American Geriatrics Society* **53**, 1339–1343.
[18] Bowden, M.L., Grant, S.T., Vogel, B. & Prasad, J.K. (1988). The elderly, disabled and handicapped adult burned through abuse and neglect, *Burns, Including Thermal Injury* **14**(6), 447–450.
[19] McNamee, C.C. & Murphy, M.B. (2006). Elder abuse in the United States, *NIJ Journal* 225. Retrieved October 19, 2008, from http://www.ojp.usdoj.gov/nij/journals/255/elder_abuse.html.
[20] Brandeis, G.H., Berlowitz, D.R. & Katz, P. (2001). Are pressure ulcers preventable? A survey of experts, *Advances in Skin and Wound Care* **14**(5), 244, 245–248.
[21] Payne, B. (2003). Preventing elder abuse requires an integrated approach, *Quest* **6**(2), 1–4.
[22] Hudson, M.F. & Carlson, J.R. (1999). Elder abuse: its meaning to Caucasians, African Americans, and Native Americans, in *Understanding Elder Abuse in Minority Populations*, T. Tatara, ed, Brunner/Mazel, Philadelphia, pp. 187–204.
[23] Hafemeister, T.L. (2003). Financial abuse of the elderly in domestic settings, in *Elder Mistreatment: Abuse Neglect and Exploitation in an Aging America*, R.J. Bonnie & R.B. Wallace, eds, National Academies Press, Washington, DC, pp. 382–445.
[24] Rabiner, D.J., O'Keeffe, J. & Brown, D. (2004). A conceptual framework of financial exploitation of older persons, *Journal of Elder Abuse and Neglect* **16**(2), 53–73.
[25] Nerenberg, L. (1999). *Forgotten Victims of Elder Financial Crime and Abuse: A Report and Recommendations*, Retrieved October 18, 2008, from http://www.ncea.aoa.gov/NCEAroot/Main_Site/Library/Publications/Publications3.aspx.
[26] Wilber, K.H. & Reynolds, S.L. (1996). Introducing a framework for defining financial abuse of the elderly, *Journal of Elder Abuse & Neglect* **8**(2), 61–80.
[27] Choi, N.G., Kulick, D.B. & Mayer, J.B. (1999). Financial exploitation of elders: analysis of risk factors based on county adult protective services data, *Journal of Elder Abuse & Neglect* **10**(3/4), 39–62.
[28] Grafstrom, M., Nordberg, A. & Winblad, B. (1992). Abuse is in the eye of the beholder. Report by family members about abuse of demented persons in home care. A total population-based study, *Scandinavian Journal of Social Medicine* **21**(4), 247–255.
[29] Kosberg, J. (1998). Abuse of elderly men, *Journal of Elder Abuse and Neglect* **9**(3), 69–88.
[30] Kaul, A. & Duffy, S. (1991). Gerontophilia: a case report, *Medicine Science and the Law* **31**(2), 110–114.
[31] Burgess, A.W., Prentky, R.A. & Dowdell, E.B. (2000). Sexual predators in nursing homes, *Journal of Psychosocial Nursing and Mental Health Services* **38**(8), 26–35.
[32] A Perfect Cause (2005). *Predators in America's Nursing Homes: 2005 Report*, http://www.aperfectcause.com/PDF_APC/APerfectCause-PredatorsinAmericasNursingHomes-2005Report.pdf (retrieved Jan 7 2007).
[33] Reis, M. & Nahmiash, D. (1998). Validation of the indicators of abuse (IOA) screen, *Gerontologist* **38**(4), 471–480.
[34] Compton, S.A., Flanagan, P. & Gregg, W. (1997). Elder abuse in people with dementia in Northern Ireland: prevalence and predictors in cases referred to a psychiatry of old age service, *International Journal of Geriatric Psychiatry* **12**(6), 632–635.
[35] Cooney, C. & Mortimer, A. (1995). Elder abuse and dementia: a pilot study, *International Journal of Social Psychiatry* **41**, 276–283.
[36] Anetzberger, G.J. (1987). *The Etiology of Elder Abuse by Adult Offspring*, Charles C. Thomas, Springfield.
[37] Aitken, L. & Griffin, G. (1996). *Gender Issues in Elder Abuse*, SAGE, London.
[38] Naimark, D. (2001). Financial exploitation of the elderly: the evaluation of mental capacity and undue influence, *American Journal of Forensic Psychiatry* **22**(3), 5–19.
[39] Nerenberg, L. (1996). *Hornswoggled? An interview with Margaret Singer on Undue Influence*, http://www.preventelderabuse.org/nexus/singer.html (retrieved Dec 6 2006).
[40] Lachs, M.S. & Pillemer, K. (1995). Abuse and neglect of elderly persons, *New England Journal of Medicine* **332**(7), 437–443.
[41] Wolf, R.S. & Pillemer, K.A. (1989). *Helping Elderly Victims: The Reality of Elder Abuse*, Columbia University Press, New York.
[42] Lachs, M.S., Berkman, L., Fulmer, T. & Horwitz, R.I. (1994). A prospective community-based pilot study of risk factors for the investigation of elder mistreatment, *Journal of the American Geriatrics Society*, **42**(2), 169–173.
[43] Lachs, M.S., Williams, C., O'Brien, S., Hurst, L. & Horwitz, R. (1997). Risk factors for reported elder abuse and neglect: a nine-year observational cohort study, *Gerontologist* **37**(4), 469–474.
[44] Pillemer, K. (1988). Maltreatment of patients in nursing homes: overview and research agenda, *Journal of Health and Social Behavior* **29**(3), 227–238.
[45] Floyd, J. (1984). Collecting data on abuse of the elderly, *Journal of Gerontological Nursing* **10**(12), 11–15.

[46] Lachs, M.S., Williams, C.S., O'Brien, S., Pillemer, K.A. & Charlson, M.E. (1998). The mortality of elder mistreatment, *JAMA: The Journal of the American Medical Association* **280**(5), 428–432.
[47] Coker, J. & Little, B. (1997). Investing in the future: protecting the elderly from financial abuse, *FBI Law Enforcement Bulletin* **66**(12), 1–5.
[48] Deem, D., Nerenberg, L. & Titus, R. (2007). Victims of financial crime in *Victims of Crime (3rd Edition)* R.C. Davis, A.J. Lurigio & S. Herman, eds, Sage Publications, Thousand Oaks, CA.
[49] Goldstein, M.Z. (1995). Maltreatment of elderly persons, *Psychiatric Services* **46**(12), 1219–1221, 1225.
[50] Comijs, H.C., Penninx, B.W., Knipscheer, K.P. & van Tilburg, W. (1999). Psychological distress in victims of elder mistreatment: the effects of social support and coping, *Journal of Gerontology: Psychological Sciences* **54**(4), P240–P245.
[51] Dyer, C.B., Pavlik, V.N., Murphy, K.P. & Hyman, D.J. (2000). The high prevalence of depression and dementia in elder abuse or neglect, *Journal of the American Geriatrics Society* **48**(2), 205–208.
[52] Wolf, R. (1997). Elder abuse and neglect: causes and consequences, *Journal of Geriatric Psychiatry* **30**(1), 153–174.
[53] Acierno, R., Gray, M., Best, C., Resnick, H., Kilpatrick, D., Saunders, B. & Brady, K. (2001). Rape and physical violence: comparison of assault characteristics in older and younger adults in the National Women's Study, *Journal of Traumatic Stress* **14**, 685–695.
[54] Brandl, B., Hebert, M., Rozwadowski, J. & Spangler, D. (2003). Feeling safe, feeling strong: support groups for older abused women, *Violence Against Women* **9**(12), 1490–1503.
[55] Vinton, L. (1998). A nationwide survey of domestic violence shelters' programming for older women, *Violence Against Women* **4**, 559–571.
[56] Wolf, R.S. (2001). Support groups for older victims of domestic violence, *Journal of Women and Aging* **13**(4), 71–83.
[57] Nerenberg, L. (2007). *Elder Abuse Preventing: Emerging Trends and Promising Strategies*, Springer, New York.
[58] Teaster, P.B. & Nerenberg, L. (2003). *A National Look at Elder Abuse Multidisciplinary Teams*, National Center on Elder Abuse, Washington, DC.
[59] Nerenberg, L. (2004). *Multidisciplinary Elder Abuse Prevention Teams: A New Generation*, National Center on Elder Abuse, Washington, DC.
[60] Wiglesworth, A., Mosqueda, L., Burnight, K., Younglove, T. & Jeske, D. (2006). Findings from an elder abuse forensic center, *Gerontologist* **46**, 277–283.

Further Reading

Burgess, A.W., Dowdell, E.B. & Prentky, R.A. (2000). Sexual abuse of nursing home residents, *Journal of Psychosocial Nursing and Mental Health Services* **38**(6), 10–18.
Chen, A.L. & Koval, K.J. (2002). Elder abuse: the role of the orthopedic surgeon in diagnosis and management, *Journal of the American Academy of Orthopedic Surgeons* **10**(1), 25–31.
Chew, D.J. & Edmondson, H.D. (1996). A study of maxillofacial injuries in the elderly resulting from falls, *Journal of Oral Rehabilitation* **23**(7), 505–509.
Hawley, D. (2003). *Elder Physical and Sexual Abuse: The Medical Piece Parts 1 and 2*. An educational documentary film produced jointly for the California District Attorneys Association and the United States Department of Justice with a grant from the Office on Violence Against Women (2002-EW-BX-0004), 2003.
Ingram, E. (2003). Expert panel recommendations on elder mistreatment using a public health framework, *Journal of Elder Abuse and Neglect* **15**(2), 45–65.
Mitchell, R.A., Hasbrouck, L., Ingram, E., Dunaway, C. & Annest, J.L. (2003). Public health and aging: nonfatal physical assault-related injuries among persons aged < 60 years treated in hospital emergency departments-United States, 2001, *Morbidity and Mortality Weekly Report* **52**(34), 812–816.
Moskowitz, S. (2003). Golden age in the golden state: contemporary legal developments in elder abuse and neglect, *Loyola Law Review* **36**(2), 589–666.
Myers, J.E.B. (2005). *Myers on Evidence in Child, Domestic, and Elder Abuse Cases*, Successor Edition, Aspen Publishers, New York.
Shields, L.B., Hunsaker, D.M. & Hunsaker III, J.C. (2004). Abuse and neglect: a ten-year review of mortality and morbidity in our elders in a large metropolitan area, *Journal of Forensic Sciences* **49**(1), 122–127.
Wolf, R. (1999). *Elder Shelters: U.S., Canada, and Japan*, National Center on Elder Abuse, Washington, DC.

LISA NERENBERG

Elder Abuse: Risk

Stan and Eugene

Stan could not remember a time when he did not hate his grandfather, Eugene. After all, this was the man who virtually abandoned him when his beloved grandmother died, eight years after his mother left Stan with her parents. Not yet a teenager and Stan had to manage on his own. Eugene provided housing and sometimes brought in food, but that was the limit of his support.

Like Stan's mother had done before him, Eugene did what he wanted. It did not matter who or what might be harmed in the wake. Eugene loved hanging out with his friends at bars or the races. Occasionally he would travel to Las Vegas and spend days, sometimes weeks, gambling and socializing, until he was broke. All this left little money for household expenses and no time for Stan.

Stan is in his late 20s now. Frequently unemployed and still living in his grandfather's house, he has a well-deserved reputation for binge drinking, a quick temper, and violent behavior. Stan admits, "It just happens. Someone ticks me off and winds up in intensive care". A school counselor once diagnosed Stan as having a personality disorder.

Eugene is in his early 70s and suffers from dementia along with a host of other chronic conditions, largely resulting from long-term alcoholism and neglect of health. Eugene is frail, no longer the massive figure that brought fear to Stan, both for his size and for the brutal discipline he imposed. Eugene also is forgetful and occasionally disoriented. He increasingly becomes anxious leaving home.

Brushes with the law and other authorities are common for Stan. Before quitting school, he spent more time truant or in detention than attending class. His violent temper and general disregard for people and property resulted in a three-year prison sentence. Recently he went to jail for so severely beating Eugene that the older man was hospitalized for nearly a month, recovering from multiple injuries, including broken hip, jaw, and ribs. The trigger for the abuse was when Stan came home and discovered that Eugene forgot to light the pilot for the gas stove and fumes filled the dwelling. "How could he be so stupid? He drives me mad", explains Stan. Over the years Eugene and Stan frequently fought, usually both verbally and physically. However, the balance of power shifted as Eugene aged and his health deteriorated, with Stan becoming the victor in any dispute, although never before with such serious consequences.

Causes versus Risk Factors

The causes of elder abuse and mistreatment are complex and challenging (*see* **Elder Abuse: Policy**). The US Congress in 1981 provided an early summary of proposed theories for why elder abuse occurs, including caregiver stress and burden; insufficient community resources for caregivers; retaliation for past child abuse; ageism; acceptance of violence for expressing frustration and anger; lack of close family ties; resentment of the victim's dependency; and stress or life crises experienced by the perpetrator, such as mental health problems, alcoholism, or unemployment [1]. Only a handful of these theories have survived scrutiny more than a quarter century later, although most remain entrenched beliefs of the public and even many professionals.

Lacking evidence for causation, we tend to use the term *risk factors* for those commonly accepted conditions found through research to be closely linked to the occurrence of elder abuse. Unfortunately, there have been relatively few studies focused on elder abuse risk factors. Moreover, many existing studies are methodologically flawed, failing to use standardized definitions, control groups, and reliable and valid measures.

Risk Factor Applications

The inadequacies of elder abuse risk factor research should not be construed to mean that such investigations are regarded as useless or unimportant. Rather, knowledge of risk factors is seen as critical to problem detection, assessment, and prevention.

Understanding risk factors helps detect elder abuse by identifying conditions that may contribute to its occurrence. This means that whenever we uncover established elder abuse risk factors in circumstances involving older adults and trusted others, we should be watchful for specific examples or signs of the problem. Under the provisions of most state elder abuse reporting or adult protective service laws, examples and signs are the basis on which reports or referrals are made [2].

Risk factors also indicate the need for more thorough elder abuse assessment. It is only after completely evaluating the person and situation of victim and perpetrator that we can know what interventions are required and when. Finally, risk factors are key to elder abuse prevention. They offer a framework for determining what strategies (e.g., anger management, counseling, substance abuse treatment) may deter the reoccurrence of elder abuse or inhibit it from happening in the first place; for example, anger management, counseling, or substance abuse treatment.

Risk Factor Qualities

Risk factors increase the likelihood that elder abuse will occur. In addition, the greater the number of risk factors evident in a situation, the higher the possibility of abuse occurrence. Similarly, longer exposure to risk factors makes elder abuse more likely.

Some elder abuse research suggests that risk factors vary by form and setting. For instance, a pioneering examination of 328 elder abuse cases found different perpetrator and victim profiles for physical abuse, material abuse, and active neglect, among maltreatment forms studied. Physical abuse perpetrators were characterized by a history of mental illness, alcohol abuse, recent decline in mental and physical health status, increased dependency, and poor relations with the victim. Physical abuse victims tended to have poor emotional health, more independent functioning, and a stable social network.

In contrast, active neglect perpetrators found their victims a source of stress, and victims had deficits in memory, orientation, and performing daily living tasks. Material abuse perpetrators were alcoholic, lacked family support, were financially dependent but did not live with the victim, and experienced both long-term and recent financial problems. Their victims had difficulty with financial management and transportation as well as recent loss of social supports [3].

Elder abuse in institutional settings, like nursing homes, is distinguished from that in domestic settings, like single-family homes, in such ways as potential perpetrators and likely risk factors. More specifically, perpetrators in institutional settings include staff, visitors, and other residents; most perpetrators in domestic settings are family members. The National Center on Elder Abuse recently reviewed available studies to develop a list of elder abuse risk factors in residential care facilities. The following were among those named: no abuse prevention policy, inadequate staff education and training, insufficient staff screening, staff stress and burnout, staff shortages and high staff turnover, history of deficiencies and complaints, a closed structure with reactive problem-solving and little attention given to resident concerns, few visitors, and poor resident/staff interaction [4].

Some researchers have found that certain risk factors are particularly important for predicting elder abuse in institutional settings, such as stressful work conditions (largely resulting from staff shortages), staff burnout, staff viewing residents as childlike, staff frequently thinking about quitting their jobs, the combination of resident aggression or conflict, and poor staff training in managing challenging behaviors [5, 6].

Conceptual models exist that integrate elder abuse risk factors to better explain the problem. None have been rigorously tested. For example, Anetzberger proposes a model for domestic elder abuse where the problem is a function of characteristics on the part of the perpetrator primarily and characteristics of the victim secondarily. These come together and provide the etiology for abuse occurrence. Context is important in the model, initially as the vehicle for bringing the perpetrator and victim in contact and later for triggering the actual mistreatment [7].

The National Research Council proposes a model for institutional elder abuse. In it, risk factors begin in a sociocultural context (exemplified by the state regulatory environment and urban-rural location of the facility). These are combined with risk factors related to the social embeddedness of the facility in providing resident care (considering characteristics like supervision/staff ratios and average resident age) and those related to individual characteristics of employees or volunteers who provide resident care (including relevant work experience and stress) [8].

Elder Abuse Risk Factors

Risk factors for elder abuse can be found in characteristics of the perpetrator, victim, perpetrator/victim environment, and cultural milieu. Research is minimal on all four types of characteristics, but it is nearly absent on the last. Still, culture provides the basis for each of us becoming members of society, and specific groups within society. In the process, we learn about and adopt commonly accepted values, attitudes, and normative responses. Youth, individualism, and material acquisition are American values thought to contribute to elder abuse. Ageism, stereotyping of people with disabilities, and glorification of aggression are attitudes believed to foster prejudice, discrimination, and sometimes mistreatment of elderly and other vulnerable adults.

Perpetrator Risk Factors

Pathology can predispose individuals to be abusive through distortion of reality, inability to empathize,

lack of impulse control, and inability to handle frustration. Evidence for perpetrator pathology as an elder abuse risk factor exists for mental illness e.g., [9, 10], alcoholism e.g., [11, 12], and hostility e.g., [13, 14].

Dependency can be related to pathology in that persons with mental illness or alcoholism may have difficulty holding jobs and therefore come to rely on others for support. Dependent adults can feel resentment and powerlessness. This can result in abusive behavior as a means to compensate for or retaliate against perceived imbalances in resources and power. There is considerable evidence for perpetrator dependency as an elder abuse risk factor [15, 16]. However, two popular explanations for the problem largely have been discredited through research, namely caregiver stress and transgenerational abuse.

Victim Risk Factors

Only a couple of victim risk factors have held up to scrutiny. There is a substantial body of research identifying dementia as a victim risk factor [17, 18]. The effect of Alzheimer's disease and other dementias is: (i) to increase the vulnerability of sufferers through decreased functional capacity, rendering them dependent on others for care, not all of whom are willing or able to provide it and (ii) to limit the ability of sufferers to protect themselves through self-defense, escape, or sometimes even recognition of elder abuse when it happens.

Victim problem behaviors also represent a validated risk factor for elder abuse [19, 20]. Since persons with dementia can exhibit problem behaviors, the two risk factors can be related. Victim problem behaviors identified in the literature include being complaining, critical, demanding, disagreeable, passive, aggressive, or uncooperative. The evidence for other suggested victim risk factors, like gender or physical impairment, is either insufficient or contradictory.

Perpetrator/Victim Environment Risk Factors

While there is ample evidence for shared living arrangements [13, 15] and social isolation [21, 22] as elder abuse risk factors, there is considerably less evidence for other proposed environmental characteristics, like financial difficulties or family disharmony.

References

[1] US House Select Committee on Aging (1981). *Elder Abuse (An Examination of a Hidden Problem)*, Government Printing Office, Washington, DC. p. 39.

[2] Moskowitz, S. (1998). Saving granny from the wolf: elder abuse and neglect – the legal framework, *Connecticut Law Journal* **31**, 77–203.

[3] Wolf, R.S., Godkin, M.A. & Pillemer, K.A. (1986). Maltreatment of the elderly: a comparative analysis, *Pride Institute Journal of Long Term Home Health Care* **5**(4), 10–17.

[4] National Center on Elder Abuse (2005). *Nursing Home Abuse Risk Prevention Profile and Checklist*, National Association of State Units on Aging, Washington, DC.

[5] Pillemer, K. & Moore, D.W. (1989). Abuse of patients in nursing homes: findings from a survey of staff, *The Gerontologist* **29**(3), 314–320.

[6] Hawes, C. (2003). Elder abuse in residential long-term care settings: what is known and what information is needed? in *Elder Mistreatment: Abuse, Neglect, and Exploitation in an Aging America*, National Research Council, National Academies Press, Washington, DC, pp. 446–500.

[7] Anetzberger, G.J. (2000). Caregiving: primary cause of elder abuse? *Generations* **24**(2), 46–51.

[8] National Research Council (2003). *Elder Mistreatment: Abuse, Neglect, and Exploitation in an Aging America*, National Academies Press, Washington, DC.

[9] Pillemer, K.A. & Finkelhor, D. (1988). The prevalence of elder abuse: a random sample survey, *The Gerontologist* **28**(1), 51–57.

[10] Brownell, P., Berman, J. & Salamone, A. (1999). Mental health and criminal justice issues among perpetrators of elder abuse, *Journal of Elder Abuse and Neglect* **11**(4), 81–94.

[11] Anetzberger, G.J., Korbin, J. & Austin, C. (1994). Alcoholism and elder abuse, *Journal of Interpersonal Violence* **9**(2), 184–193.

[12] Reay, A.C. & Browne, K. (2001). Risk factor characteristics in carers who physically abuse or neglect their elderly dependents, *Aging and Mental Health* **5**(1), 56–61.

[13] Anetzberger, G.J. (1987). *The Etiology of Elder Abuse by Adult Offspring*, Charles C Thomas, Springfield.

[14] Quayhagen, M., Quayhagen, M.P., Patterson, T.L., Irwin, M., Hauger, R.L. & Grant, I. (1997). Coping with dementia: family caregiver burnout and abuse, *Journal of Mental Health and Aging* **3**, 357–364.

[15] Pillemer, K.A. (1986). Risk factors in elder abuse: results from a case-control study, in *Elder Abuse: Conflict in the Family*, K.A. Pillemer & R.S. Wolf, eds, Auburn House, Dover, pp. 239–263.

[16] Greenberg, J.R., McKibben, M. & Raymond, J.A. (1990). Dependent adult children and elder abuse, *Journal of Elder Abuse and Neglect* **2**(1/2), 73–86.

[17] Paveza, G.J., Cohen, D., Eisdorfor, C., Freels, S., Semla, T., Ashford, J.W., Gorelick, P., Hirschman, R., Luchins, D. & Levy, P. (1992). Severe family violence and Alzheimer's disease: prevalence and risk factors, *The Gerontologist* **32**, 493–497.
[18] Dyer, C.B., Pavlik, V.N., Murphy, K.P. & Hyman, D.J. (2000). The high prevalence of depression and dementia in elder abuse or neglect, *Journal of the American Geriatrics Society* **48**(2), 205–208.
[19] Pillemer, K. & Suitor, J.J. (1992). Violence and violent feelings: what causes them among family caregivers? *Journal of Gerontology* **47**, S165–S172.
[20] Compton, S.A., Flanagan, P. & Gregg, W. (1997). Elder abuse in people with dementia in Northern Ireland: prevalence and predictors in cases referred to a psychiatry of old age service, *International Journal of Geriatric Psychiatry* **12**(6), 632–635.
[21] Phillips, L.R. (1983). Abuse and neglect of the frail elderly at home: an exploration of theoretical relationships, *Journal of Advanced Nursing* **8**, 379–392.
[22] Lachs, M.S., Berkman, L., Fulmer, T. & Horowitz, R. (1994). A prospective community-based pilot study of risk factors for the investigation of elder mistreatment, *Journal of the American Geriatrics Society* **42**(2), 169–173.

GEORGIA J. ANETZBERGER

Elderly in Court

Older Adults as Witnesses in Court

Aging and Changes in Memory

There is no uniformly agreed upon definition of what age constitutes old age. Most gerontological research uses the threshold of 65 years (or sometimes 60 years), and this is also the threshold adopted in this entry. Certain aspects of memory performance decline with age, although not all parts of memory are equally affected [1]. One of the most important memory areas related to eyewitness performance is long-term episodic memory, which is particularly affected by the aging process [1].

Another relevant aspect of memory that declines with age is source–memory, or the ability to identify the source of a memory. In general, it becomes more difficult to remember the contexts in which a particular piece of information was learned and this can lead to source misattribution (i.e., misremembering where familiar information comes from, such as direct perception, suggestion, dreams, etc.). Compared to young adults, seniors have been shown to have impaired memory when having to decide if they know about a certain fact because they heard it or saw it [2] or if they heard a statement from one person as opposed to another. This confusion about which one of two people said something is especially evident if the sources are similar, such as both of them being women [3]. In the legal context, the source of a memory can be of importance in several ways.

In some cases it may be crucial to know which individual made a particular statement. In many cases, however, source memory is more important in a more indirect way: source misattribution is one on the avenues in which misinformation can taint the testimony of a witness. Witnesses to a crime or an accident may learn additional information about the observed incident by reading a related article in the local newspaper or by watching coverage of it in the news.

Witnesses may be questioned by interviewers who mistakenly introduce false information in some of their questions. If source monitoring is impaired, then witnesses may come to believe that the additional information stems from the event itself, as opposed to originating from the subsequent informational sources. Source misattributions are therefore important for the explanation of false memories and susceptibility to suggestions [4–6]. Mistakes in remembering in which exact circumstances a familiar looking person was encountered can also lead to the false identification of an innocent bystander to a crime as the perpetrator [7, 8].

Aging can also impair the ability to accurately recall the chronological order of events and actions. Older adults make more mistakes in the temporal ordering of actions than young adults [9, 10]. Similarly, spatial memory declines with age. Older adults have been found to be less good in describing the spatial layout of places visited than young adults [11]. Thus, in situations when it can be crucial to describe the layout of a crime scene or the temporal order of events, seniors may perform less accurately than young adults.

How many and which details a person remembers is affected by the retrieval format used, e.g., the types of questions asked of a witness to obtain a memory account. A witness can be asked to provide a free

recall account of what happened, answer open or specific questions (also known as *cued recall*, e.g., "What did the perpetrator look like?") or be presented with recognition questions (e.g., "Is this person the person who attacked you?"). Memory research has shown that age differences between younger and older witnesses are largest for free recall, smaller for cued recall, and smallest for recognition questions [1]. This distinction is important for eyewitness testimony: A witness statement usually contains some information obtained in free recall and cued recall. A person identification decision made in a lineup is a memory recognition task.

Older Adults' Person Identification Performance

As would be expected from the fact that person identification is a recognition task, and that recognition is less affected by age, older and younger adults do not differ very much in their rates of correctly identifying a target from a target-present lineup [12], i.e., a lineup where the suspect is indeed the perpetrator, although older adults are more affected by delays between exposure and the lineup [13] (*see* **Eyewitness Lineups: Identification from**). Reliable age differences, however, occur when the task is to identify a person from a lineup in which the actual perpetrator is not present, i.e., where the suspect is not the perpetrator of the alleged crime. Here older adults have higher false alarms rates, which means that they are more likely to choose a foil than young adults [14–16]. In situations in which the witness has looked through a mugbook that did not contain the perpetrator, and where the subsequent lineup contains a person seen in the mugbook but not the actual perpetrator, older adults are more likely than young adults to pick out this person previously encountered in the mugbook falsely as being the perpetrator [8].

Eyewitness Memory in Older Adults

When giving witness statements, older adults tend to give fewer details than young adults [17–19]. In particular, older adults seem to be less good in describing in free recall person-related details, such as what an offender looked like [18, 20]. For other details, age-related decreases in memory seem to occur later than the usual research cutoff age of 60 or 65 years: Witnesses over the age of 75 have been found to give fewer correct details for actions, objects, or the surroundings of an incident than those aged 74 and below [18].

When questioned, older adults also recall fewer correct details than young adults, both when questioned in techniques used in investigative interviews (e.g., Structured Interview or Cognitive Interview, see below) [18], or in simulated direct examinations and cross-examinations [19].

Suggestibility and Older Adults. Another issue impacted by aging may be suggestibility, also called *susceptibility to misinformation*. Most people can be influenced by misinformation, depending among other issues on one's memory of an event and the skill of the person trying to plant misinformation (*see* **Eyewitness: Suggestibility of**; **Interrogative Suggestibility**), but older adults may be more susceptible to misinformation than younger adults. Most research to date on suggestibility and the elderly has employed the standard misinformation paradigm, in which participants watch a film, receive misinformation about some details of the film a short time later (typically after 10–30-min intervals), and then, after another brief delay, they are tested for acceptance of the misinformation in a questioning phase. Results of these studies on seniors have been mixed.

While some researchers have reported higher rates of susceptibility to misinformation in older adults than in young adults [21–24], others have not reported reliable age differences [25, 26]. A study using a live event in which participants were actively involved [17] found that seniors were overall more suggestible than younger adults. However they were more susceptible for certain types of misinformation (central details, such as whether or not they had been touched during the event, or concerning actions they had performed themselves) but not for other types (e.g., peripheral details, such as whether a person they had interacted with had been wearing a skirt or trousers).

Another path to the investigation of interrogative suggestibility is the Gudjonsson Suggestibility Scale, a standardized assessment tool. An application of this to older adults demonstrated that while they were more suggestible than younger adults overall, they were more influenced by the suggestive wording of the questions than young adults, but not more easily moved to change their answers through social pressure [27]. These findings that older adults can be more suggestible than young adults become more

pertinent when one considers that older adults are less likely to use self-initiated remembering and thus need to be asked more questions than young adults to obtain a comparable amount of information [17, 18]. Whenever questions are asked in an interview, this opens the possibility that (unintentional) misinformation is introduced. Therefore, particular care needs to be taken when questioning older adults to ensure that no suggestive question or misinformation is introduced into the interview.

Improving the Testimony of Older Adults: the Cognitive Interview. Given that older witnesses provide fewer details in free recall and when questioned than young adults, efforts have been made to enhance their interviewing performance. The most promising to date is an application of the cognitive interview. The cognitive interview [28] relies on two main assumptions: that contextual reinstatement improves memory performance by increasing the accessibility of encoded information and that memories are encoded in a network of associations, which implies that they can be triggered by multiple cues.

The components of the original cognitive interview include: mental reinstatement of the context, complete reporting including details believed to be trivial, varied retrieval routes, and witness compatible questioning. Varied retrieval includes changes in retrieval order and perspective, e.g., starting at the beginning, then starting at the end, then reporting from the perspective of a bystander. Witness compatible questioning includes tailoring the interview questions to the mental representation of the witness instead of using the same standardized checklist for all witnesses. The cognitive interview is also sensitive to the fact that witnesses' resources are limited and therefore questions are not asked while the witness is still thinking about the last question.

Applications of the cognitive interview to the interviewing of older adults have shown substantial increases in the number of correct details reported when compared to standard interviews [18, 29], importantly without a concomitant increase in number of incorrect details. The increase in correct details is mainly due to increasing the details reported in the free recall part of the interview. This is an interesting finding because seniors do not usually do well on free recall tasks, which suggests that the cognitive interview provides the contextual scaffolding that facilitates free recall [30].

Moreover, as older witnesses remember more in free recall, they will have to be asked fewer questions, thus decreasing the likelihood of unintentionally exposing them to misinformation. The cognitive interview can even be successfully used to interview older adults who are cognitively impaired [31]. To date, there is very limited research investigating to what extent older persons with cognitive impairment, e.g., in early stages of Alzheimer's disease, can accurately report events. This is an important field in need of more attention because persons with cognitive impairments may be particularly likely to be subjected to maltreatment and abuse [32] (*see also* **Elder Abuse: Risk**).

Credibility of Older Witnesses

When keeping the performance of younger and older witnesses constant, e.g., in experimental mock-juror studies where participants read a fictitious witness statement and cross-examination presented to be either from a young or older witness, older adults are consistently rated as more honest than young witnesses. In such research, older male witnesses seem particularly highly regarded. The credibility assigned to older witness in such mock-juror studies has been shown to be related to what aging stereotypes jurors hold. Persons who have more favorable attitudes tend to rate older adults' credibility more highly than persons who have more ageist attitudes [33].

In reality, of course, older witnesses often perform less well than younger witnesses. Studies looking at the credibility of actual memory accounts given by younger and older witnesses found that older adults' statements were perceived to be less accurate and less credible than those of young adults [19]. However, credibility was not affected by the age of the witness itself but by the quality and characteristics of the statements. It has been shown for younger witnesses, that they are perceived as more credible if their accounts include a high amount of peripheral details in their descriptions of the observed crime [34] and when their statements are uttered with confidence. As seniors typically provide fewer peripheral details in their accounts, and as a group more frequently use negative qualifiers (e.g., "I am not sure"), it is plausible that this signals to mock jurors that older adults are comparatively less accurate witnesses.

In all these considerations, it is important to emphasize that not all persons are equally affected

by the aging process. Some maintain high levels of cognitive functioning until very old age, while others experience an earlier decline. Consequently, there is large variability in the memory performance for this age group. When assessing the performance of a particular older witness it is important to consider this person as an individual and not merely as a member of the group of "the elderly".

References

[1] Balota, D.A., Dolan, P.O. & Duchek, J.M. (2000). Memory changes in healthy older adults, in *The Oxford Handbook of Memory*, E. Tulving & F.I.M. Craik, eds, Oxford University Press, Oxford, pp. 395–409.

[2] McIntyre, J.S. & Craik, F.I.M. (1987). Age differences for item and source information, *Canadian Journal of Psychology* **42**, 175–192.

[3] Ferguson, S.A., Hashtroudi, S. & Johnson, M.K. (1992). Age differences in using source-relevant cues, *Psychology and Aging* **7**, 443–452.

[4] Multhaup, K.S., De Leonardis, D.M. & Johnson, M.K. (1999). Source memory and eyewitness suggestibility in older adults, *The Journal of General Psychology* **126**(1), 74–84.

[5] Belli, R.F., Lindsay, D.S., Gales, M.S. & McCarthy, T.T. (1994). Memory impairment and source misattribution in postevent misinformation experiments with short retention intervals, *Memory and Cognition* **22**, 40–54.

[6] Johnson, M.K., Hashtroudi, S. & Lindsay, D.S. (1993). Source monitoring, *Psychological Bulletin* **144**, 3–128.

[7] Perfect, T.J. & Harris, L.J. (2003). Adult age differences in unconscious transference: source confusion or identity blending? *Memory and Cognition* **31**(3), 570–580.

[8] Memon, A., Hope, L., Bartlett, J.C. & Bull, R.H.C. (2002). Eyewitness recognition errors: the effect of mugshot viewing and choosing in young and old adults, *Memory and Cognition* **30**, 1219–1227.

[9] Krausler, D.H. & Phillips, P.L. (1988). Instructional variation and adult age differences in activity memory, *Experimental Aging Research* **14**(4), 195–199.

[10] Krausler, D.H. & Wiley, J.G. (1990). Temporal memory and content memory for actions: adult age differences in acquisition and retention, *Experimental Aging Research* **16**(3), 147–150.

[11] Uttl, B. & Graf, P. (1993). Episodic spatial memory in adulthood, *Psychology and Aging* **8**(2), 257–273.

[12] Yarmey, A.D. & Kent, J. (1980). Eyewitness identification by elderly and young adults, *Law and Human Behavior* **4**, 359–371.

[13] Memon, A., Bartlett, J., Rose, R. & Gray, C. (2003). The aging eyewitness: effects of age on face, delay, and source-memory ability, *Journal of Gerontology* **58B**(6), 338–345.

[14] Memon, A. & Bartlett, J.C. (2002). The effect of verbalisation on face recognition in young and old adults, *Applied Cognitive Psychology* **16**, 635–650.

[15] Memon, A. & Gabbert, F. (2003). Improving the identification accuracy of senior witnesses: do prelineup questions and sequential testing help? *Journal of Applied Psychology* **2**, 341–347.

[16] Searcy, J.H., Bartlett, J.C. & Memon, A. (1999). Age differences in accuracy and choosing in eyewitness identification and face recognition, *Memory and Cognition* **27**, 538–552.

[17] Mueller-Johnson, K. & Ceci, S.J. (2004). Memory and suggestibility in older adults: live event participation and repeated interview, *Applied Cognitive Psychology* **18**, 1109–1127.

[18] Wright, A.M. & Holliday, R.E. (2007). Enhancing the recall of young, young-old an old-old adults with cognitive interviews, *Applied Cognitive Psychology* **21**, 19–43.

[19] Brimacombe, C.A.E., Quinton, N., Nance, N. & Garrioch, L. (1997). Is age irrelevant? Perceptions of young and old adult eyewitnesses, *Law and Human Behavior* **21**, 619–634.

[20] Yarmey, A.D. (2001). The older eyewitness, in *Elders, Crime and the Criminal Justice System*, M.B. Rothman & P. Entzel, eds, Springer, New York, pp. 127–148.

[21] Karpel, M.E., Hoyer, W.J. & Toglia, M.P. (2001). Accuracy and qualities of real and suggested memories: nonspecific age differences, *Journal of Gerontology* **56B**(2), 103–110.

[22] Mitchell, K.J., Johnson, M.K. & Mather, M. (2003). Source monitoring and suggestibility to misinformation: adult age-related differences, *Applied Cognitive Psychology* **17**, 107–119.

[23] Loftus, E.F., Levidow, B. & Duensing, S. (1992). Who remembers best? Individual differences in memory for events that occurred in a science museum, *Applied Cognitive Psychology* **6**, 93–107.

[24] Cohen, G. & Faulkner, D. (1989). Age differences in source forgetting: effects on reality monitoring and on eyewitness testimony, *Psychology and Aging* **4**, 10–17.

[25] Bornstein, B.H., Witt, C.J., Cherry, K.E. & Greene, E. (2000). The suggestibility of older witnesses, in *Elders, Crime and The Criminal Justice System- Myth, Perceptions, and Reality In The 21st Century*, M.B. Rothman, B.D. Dunlop & P. Entzel, eds, Springer, New York, pp. 149–161.

[26] Coxon, P. & Valentine, T. (1997). The effect of the age of eyewitnesses on the accuracy and suggestibility of their testimony, *Applied Cognitive Psychology* **11**, 415–430.

[27] Polczyk, R., Wesolowska, B., Gabarczyk, A., Minakowska, I., Supska, M. & Bomba, E. (2004). Age differences in interrogative suggestibility: a comparison between young and older adults, *Applied Cognitive Psychology* **18**(8), 1097–1107.

[28] Fisher, R.P. & Geiselman, R.E. (1992). *Memory-Enhancing Techniques for Investigative Interviewing – The Cognitive Interview*, Charles C. Thomas, Springfield.
[29] Mello, E.M. & Fisher, R.P. (1996). Enhancing older adult eyewitness memory with the cognitive interview, *Applied Cognitive Psychology* **10**, 403–417.
[30] Wright, A.M. & Holliday, R.E. (2003). Interviewing elderly witnesses and victims, *Forensic Update* **73**, 20–24.
[31] Wright, A.M. & Holliday, R.E. (2007). Interviewing cognitively impaired older adults: how useful is a cognitive interview? *Memory* **15**(1), 17–33.
[32] Lachs, M.S., Williams, C., O'Brien, S., Hurst, L. & Horwitz, R. (1997). Risk factors for reported elder abuse and neglect: a nine-year observational cohort study, *The Gerontologist* **37**(4), 496–474.
[33] Mueller-Johnson, K., Toglia, M.P., Sweeney, C.D. & Ceci, S.J. (2007). The perceived credibility of older adults as witnesses and its relation to ageism, *Behavioral Sciences and the Law* **25**, 355–375.
[34] Wells, G.L. & Leippe, M.R. (1981). How do triers of fact infer the accuracy of eyewitness identifications? Using memory for peripheral detail can be misleading, *Journal of Applied Psychology* **66**(6), 682–687.

KATRIN MUELLER-JOHNSON

Elderly Offenders *see* Death Penalty and Age

Electrical Engineering

Introduction

Forensic electrical engineering is the practical application of electrical engineering knowledge to legal questions about electrical phenomena. Practical electrical engineering knowledge is obtained from experience in designing, installing, maintaining and repairing electrical devices, appliances, and equipment. Reports, demonstrations, depositions, and court testimony are used to explain electrical phenomena to insurers, attorneys, arbitrators, judges, and juries. The area of practice extends from software for computers to the generation and distribution of electrical power, which might be controlled by software, and to consumer products.

The electrical engineer explains how the electrical software, equipment, or device functions normally and why it malfunctioned, violated a copyright, or failed in this instance causing damage, financial loss, injury, or death. In some instances, the electrical engineer might be retained by a client to verify that electricity was not involved with the cause of a fire, damage, or injury. Quite frequently, electrical engineers must use mechanical, thermodynamic, and optical knowledge to answer questions since the generation, distribution, and utilization of electrical power involves mechanical components, which can produce heat and light.

History

The investigation of electrical phenomena began in Europe in the seventeenth century [1]. The first electrical investigators were called *electricians*. At the time electricity was believed to be fire, "fulmen fulminis" or fire of fire and lightning was fire from the sky, the wrath of God].[a] St Paul was struck by lightning in 33 AD while traveling to Damascus to suppress Christianity. On regaining his sight, St Paul was baptized and immediately began preaching Christian theology.[b] On July 2, 1505, lightning struck near a happy-go-lucky law student, Martin Luther. Fifteen days later, Martin Luther entered the Black Monastery in Erfurt, Germany [2]. Church bells often were inscribed with a Latin phrase, Fulgura frango, which when translated means, "I break up the lightning". A treatise on the subject by a medieval scholar titled "Proof that the ringing of bells during thunderstorms may be more dangerous than useful", found that over a 33-year period in Germany, a total of 386 lightning strikes on church towers killed 103 bell ringers [3, 4]. In 1767, a church in Venice storing 100 t of gunpowder was struck by lightning, 3000 people were killed and a large section of the city destroyed [5]. Benjamin Franklin perceived and proved that lightning is an electrical phenomenon. He devised the famous sentry box experiment in 1749, which clearly displayed the electrical nature of lightning [6]. When he reported his findings, the King of France, Louis XV, who was an

electrical fanatic, had a sentry box built in 1752, and subsequently Franklin's experimental findings were verified in Europe [7, 8]. Benjamin Franklin made on-site inspections of churches, which were struck by lightning, [9] wrote newspaper articles requesting information about lightning damage.[c] and read articles about lightning damage to churches to determine the effectiveness of bell ringing [10].[d] Franklin found that bell ringing during a thunderstorm is a very dangerous occupation. If metal wire was used to ring the bells, the lightning would vaporize it but not damage anything along its path but killed the bell ringer. However, if rope was used to ring the bells, the lightning would cause severe damage to the steeple but probably not kill the bell ringer. This is where Franklin got the idea to attach metal wire to his rod. Franklin rods were opposed by church clergy on their structures because it seemed a sacrilegious act and an expression of no confidence in the mercy of God in striking a sacred structure. However, a Franklin rod was installed on a building in Venice (Campanile of San Marco) that had been struck and severely damaged by lightning nine times from 1388 to 1762. Since the installation of a Franklin rod in 1766, it has not been damaged by lightning [3, 4]. Franklin inspected buildings protected with a Franklin rod, which resulted in Franklin writing the first specifications or standards improving lightning rods [11].

Electrical engineering emerged as a discipline in 1864 when the Scottish physicist and mathematician James Clerk Maxwell summarized the basic laws of electricity in mathematical form and predicted that radiation of electromagnetic energy would occur in a form that later became known as *radio waves*.

The first practical application of electricity was the telegraph, invented by Samuel B. F. Morse in 1837. However, there was not a great need for electrical engineers until the inventions of Alexander Graham Bell and Thomas A. Edison approximately 40 years later; Bell's telephone in 1876 and Edison's incandescent lamp in 1878 and the first central electric generating plant in New York in 1882. Overnight a large demand was created for men educated and trained to work with electricity [1]. During the 1890s, engineering schools all over the world added studies in electrical engineering, often as optional courses within the mechanical engineering field. Previously, the study of electricity was considered to be a subfield of physics. Academic departments of electrical engineering emerged around the turn of the century [12]. In 1883, Cornell University introduced the world's first course of study in electrical engineering and, in 1885, the University College, London, founded the first chair of electrical engineering in the United Kingdom [13]. The sudden explosion of the electrical industry generated a need for electrical experts to answer legal questions about electrical designs, patents, contracts, personal injury, and damage. The adversarial legal system permitted each party involved in litigation to retain their own electrical expert as opposed to other legal systems where the judge appoints the expert to assist with technical questions. In addition, the ability of attorneys in the United States to pursue litigation on a contingency basis and the subrogation laws of the United States increased the demand for forensic electrical engineers.

Electrical litigation started with patent disputes and Lewis Howard Latimer was probably the first electrical engineer to testify in a patent dispute [14]. It is possible that there were earlier trials where someone was qualified in electrical engineering, but without an appeal that there is no documentation.

Education, Training, and Certification

There is presently no formal program for a Bachelor of Science degree in forensic electrical engineering. In the United States, the basic educational requirement for a forensic electrical engineer is normally a Bachelor of Science degree in electrical engineering from an accredited university. A few universities in the United States offer graduate courses in forensic engineering. The University of Iowa has a Biomedical Engineering Program. The University of Texas at Tyler has "Advanced Topics in Engineering (introduction to forensic engineering)". The University of Colorado at Denver offers "Failure Analysis and Condition Assessment in Civil and Mechanical Engineering". Purdue University's Weldon School offers a course to seniors and graduate students titled "Medical Device Accidents and their Engineering Analysis".

In addition, the individual should be a licensed professional engineer since state laws in the United States require engineers who work independently to be licensed. The requirements for licensing differ among the states.

In Canada, the engineering students normally take their professional engineering test in the senior year of their educational program.

Typically, engineers get involved with forensic work by joining a consulting firm, which specializes in forensic engineering, or associating with practicing forensic engineers. The senior members of the consulting firm or practitioners provide the training needed by the degreed and experienced engineer in the aspects of forensic engineering.

In 1997, the National Academy of Forensic Engineers (NAFE) appointed a committee, the Forensic Engineering Curriculum (FEC) Committee, and supplied funding to develop guidelines for university courses in forensic engineering. The FEC Committee determined that any forensic engineering degree program should begin at the masters degree level because of the specialized nature of forensic engineering.

Education and training for forensic electrical engineers is also available through various technical, legal, and forensic societies. The Institute of Electrical and Electronic Engineers (IEEE), conducts seminars and lectures on electrical equipment and the problems associated with the equipment. NAFE has seminars twice a year where technical papers are presented about forensic engineering investigations. The academy also publishes a journal of the papers that were presented at the seminars. Many of the presentations and published papers involve electrical engineering. The American Academy of Forensic Sciences (AAFS) has an engineering section. The engineering section of AAFS presents papers about forensic engineering investigations at their annual meeting each February. The AAFS has workshops at their annual meeting and sometimes the workshops involve some aspects of electrical engineering.

Certification in forensic electrical engineering in the United States is provided by NAFE through the Council of Engineering and Scientific Specialty Boards. Subsequently, members of NAFE can refer to themselves as a Board Certified Diplomate in forensic engineering by NAFE, program accredited by Council of Engineering and Scientific Specialty Boards (CESB). Another certification program is through the International Institute of Forensic Engineering Sciences, Inc., (IIFES). The IIFES maintains its certification program through the Forensic Specialties Accreditation Board. Forensic engineers who meet the requirements of IIFES are awarded a Certificate of Qualification as a Diplomate, IIFES. They are then entitled to represent themselves as being board certified by the IIFES.

Societies, CPD's and Sections

The largest forensic engineering society is NAFE. NAFE has approximately 400 members and 52 are degreed electrical engineers. NAFE is a chartered affinity group of the National Society of Professional Engineers of the United States and all members are licensed/registered professional engineers. All members of NAFE must provide forms each year reflecting that they have accumulated 100 continuing professional development (CPD) credits during the past 5 years.

The AAFS established an engineering sciences section in 1981. The AAFS engineering sciences section has approximately 170 members and 20 members are electrical engineers.

Casework

Electrical engineers are asked to answer legal questions about electrical phenomena from charged particles to the appliances in our homes and the electrical system that supplies them with power. The questions might be asked verbally or in written statements by a client to determine if further investigation or litigation should be pursued. Most of the casework involves civil litigation and the majority of the clients are insurance companies and their attorneys. The engineer is contacted by telephone, email, and fax machine by claim/loss adjusters, attorneys, public defenders, manufacturers, private investigators, military investigators, law enforcement, and individuals who are handling a matter involving financial loss, violations of law, or personal injury/death. The matter is discussed with the potential client and the engineer informs them if they can provide any assistance in the matter. The client usually requires a curriculum vita (CV) (resume) to check if the engineering education, training, and work experience would qualify them to be declared an expert witness in a court of law involving the matter. Rate sheets (charges for services) are almost always requested. If the client decides to retain the engineer, a contract should be signed between them, which details what service the engineer will provide to the client: inspection, investigation, testing, analysis, reports, and/or testimony.

The subrogation laws in the United States permit insurance companies to recoup the money they paid for fire damage from individuals, companies, or manufacturers for improper/defective installation, use, or manufacture of a product. Subsequently, most of the forensic electrical engineer's casework involves fire investigation. Electrical engineers who investigate the cause of fires must also have training in fire investigation. Fire investigation training is necessary because to determine the cause of a fire, the individual must be able to locate the origin of the fire (where it started). Examination of the electrical components, if any, at the point of fire origin will usually reveal if electricity was involved with the fire's cause. Training in fire investigation can be obtained through the International Association of Arson Investigators (IAAI). The IAAI holds annual fire investigation training seminars. It has state chapters throughout the United States and in other countries, which conduct fire investigation training seminars.

Electrical injury and electrocution cases are normally large loss matters exceeding several million dollars but there are very few of them compared to the number of fires.

Engineers evaluate electrical damage to determine if it is covered by an insurance policy or which policy (fire insurance, boilermaker, or umbrella). If the verbal statement of the engineer is not sufficient to answer the legal technical questions, then a written report might be requested from the engineer to pursue the matter further. Normally, the report would include the scope of work to be performed, details of the investigation, any testing, research, analysis, and conclusions. The analysis might include an evaluation of different versions of events in the matter and their affect on the cause of the damage or injury. The majority of the cases settle after a report is submitted. However, a few go on to depositions where the engineer is asked questions under oath by the attorneys involved in litigating the case. Before the deposition, the engineer must educate his client attorney about electricity and/or fire investigation. The client should understand how electricity caused the injury, damage, or fire and what laws, codes, or standards apply to the case and if any were violated by any parties involved. At the deposition, the engineer is asked for their CV by the deposing attorney and they are questioned under oath about the CV to evaluate their qualifications to testify about the matter at hand. They are asked details about their case file, investigation, testing, and analysis. Hypothetical questions are asked to evaluate potentially different conclusions. Sometimes, an estimate of the percentage of reliability of the conclusions is asked of the engineer.

Very few cases go to trial because of the cost of presenting a case in court. Some cases go to trial because the various parties have experts with drastically different opinions, personal grievances, or ask the court to determine how the case should be settled. Before trial, the engineer might be asked to submit a list of casework, authored articles, depositions, and trial testimony. At trial, the engineer is asked about their education, training, and professional work experience by attorneys and sometimes judges to determine if they are qualified to give expert witness testimony at the trial and on what areas/subjects they can testify about. The judge determines if the engineer qualifies as an expert witness. If qualified, the engineer answers technical questions posed by his client attorney, opposing attorneys, and rarely the judge and jury to aid the jury and court in their deliberations. During questioning by his client, the engineer must educate the judge and jury about electricity, fire investigation, computer software, or whatever is the subject of the litigation. The normal operation of the electrical device should be explained and also how it malfunctioned in this instance causing the financial loss, injury, or death. Next, the engineer is questioned by the opposing attorneys who attempt to find errors in the expert's testimony.

An example electrical injury case involved the loss of arms and a liability of \$10–15 million. An electrical engineer was contacted by an independent adjuster retained by a supermarket company to assist him or her with investigating the cause of the injury. Normally, property owners in the United States are liable for electrical damage or injury on their property and the injury occurred on supermarket property. A group of children had been playing behind a closed supermarket where a 13 kV switch was located near a loading dock (Figure 1). The engineer was asked to determine why the injury occurred; if the electrical equipment was properly installed, maintained, or manufactured and if not, what parties were at fault. At the injury location, the engineer saw that the switch did not have a fence or additional enclosure around it, as required by electrical standards [15]. This indicates that the switch was improperly installed by the supermarket's contractors. The observation opening on the



Figure 1 Figure showing the front doors on the high-voltage disconnect switch. Arrow 1 indicates the missing cover over the observation window of the switch. Arrow 2 indicates a lock with a key broken off in it, which is shown in more detail in Figure 2. Arrow 3 indicates holes in the fuse compartment door, which is shown in more detail in Figure 3

front of the switch was missing a cover to keep water out, arrow 1 in Figure 1. The switch operating handle was in the closed position, arrow 1 in Figure 2. A key was broken off in a lock that was mechanically linked with the position of the switch, arrow 2 in Figure 2. Normally, the key can only be removed from the lock when the switch is in the open position. When the switch is open, the key is removed from the lock and utilized to unlock another lock, which is securing the lower door of the switch where fuses are located. Additional unsealed openings were found around the switch handle, arrow 3 in Figure 2. The switch's lower fuse compartment door had holes in it where a lock that was controlled by the broken key should have been located, arrow 3 in Figures 1 and 3. This indicates that the switch was not properly maintained by the supermarket. Subsequently, the fuse compartment of the switch could be opened and entered while the switch was energized with 13 kV. Evidence of burning, arcing, and melted metal was found at the top of the left fuse holder in the fuse compartment (Figure 4). This is where the child's right hand made contact with 13 kV. Fingerprints in metal were found on the left side of the fuse compartment opening (Figure 5). This is where the electrical current exited the boy's left hand. The electrical current flowed from the top of the fuse holder through the boy's arms to the side of the switch, which is grounded as required by electrical codes. This injury was caused by the improper design, installation, and maintenance of the high-voltage disconnect switch, which violated electrical codes and standards. It was not designed for outdoor use. The openings in its enclosure permitted water and debris to enter into the switch. A single-door design covering the fuse compartment and switch operating handle would have eliminated the problem with the missing interlock lock on the fuse compartment door. The fuse compartment is required to be secured from entry when the switch is closed. The switch was designed for installation in a locked electrical room. However, it was incorrectly listed in a utility company's book of acceptable switches for the location. A scheduled maintenance program would have detected the switch's deficiencies and corrected them. This case settled out of court without depositions or reports. The engineer educated the supermarket attorneys verbally and with printed documents about the improper installation, design, and maintenance of the switch and the codes/standards that were violated. The education and documents enabled the supermarket attorneys to bring their contractors and the utility company to the bargaining table and share in the liability for the injury.

An example of a fire case involves a severe residential fire with a large property loss. A fire investigator brought a box of items from the fire's area of origin to this engineer's office. The fire investigator asked the engineer to inspect and photograph the items to determine if they were involved with the cause of a fire. He also emailed photographs of the fire scene and his fire determination report to the engineer. Among the items was a severely heat-damaged steam generating tank (Figure 6). The steam generating tank was X-rayed (Figure 7). The X-ray showed that the sheath (exterior enclosure) of the unit's heating element was open (arrow 1 in Figure 7) and a thermal over-temperature device had blown apart (arrow 2 in Figure 7). In addition, the nichrome wire

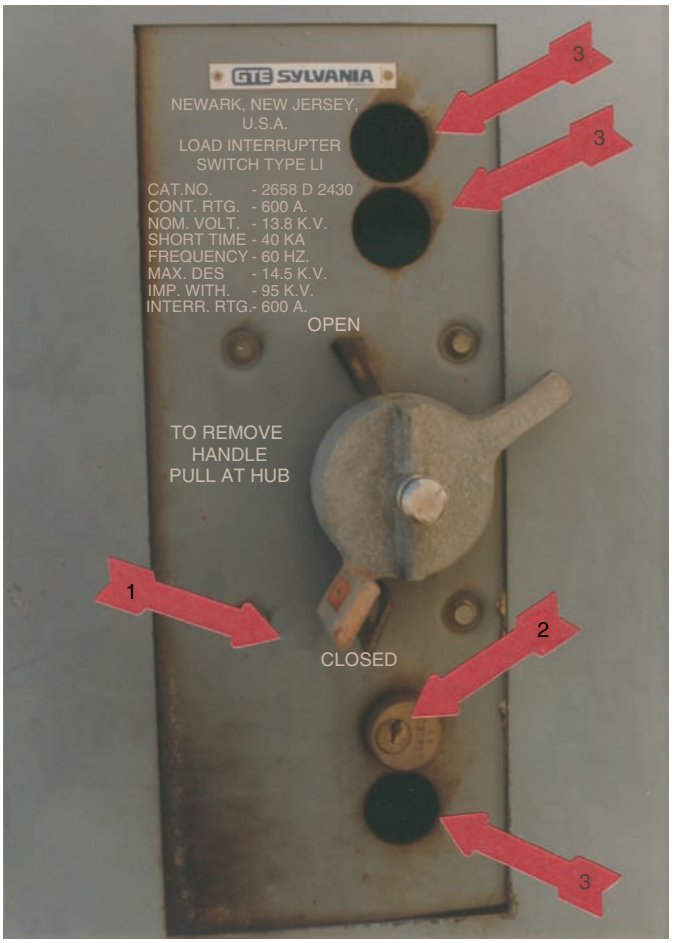


Figure 2 Figure showing the operating handle of the switch. Arrow 1 indicates that the switch is in the closed position. Arrow 2 indicates the broken key in the lock. Arrow 3 indicates unsealed openings

inside the heating element was separated and disintegrated in many locations (Figure 8). The nichrome wire is the device that generates heat inside the heating element when electric current passes through it. Nichrome wire is a nickel-base alloy, containing chromium and iron. Nichrome wire has a melting point of 1020–1452°C depending upon the chemical composition of the alloy [16]. The fire temperatures in a residential building normally do not exceed 800°C. Subsequently, the disintegration of the nichrome wire inside the heating element indicates that the heating element malfunctioned and caused the fire. The engineer verbally reported his conclusions about the inspection and X-rays to the fire investigator and the insurance company that retained him. An exemplar steam generating tank by the same manufacturer was purchased, inspected, and tested by the engineer. The exemplar was found to be defective in design and manufacture. An engineering report was requested and submitted with photographs and X-rays. The report stated that the subject steam generating tank was defective in design, manufacture, and caused the fire. The manufacturer retained experts with different opinions even after examining the evidence, X-rays, and exemplar. As a result, the engineer was deposed by the manufacturer's attorneys and he educated them about the appliance's defects. The case settled after the deposition.

Another case involved a large residential building that sustained fire damage while $500 000 worth of

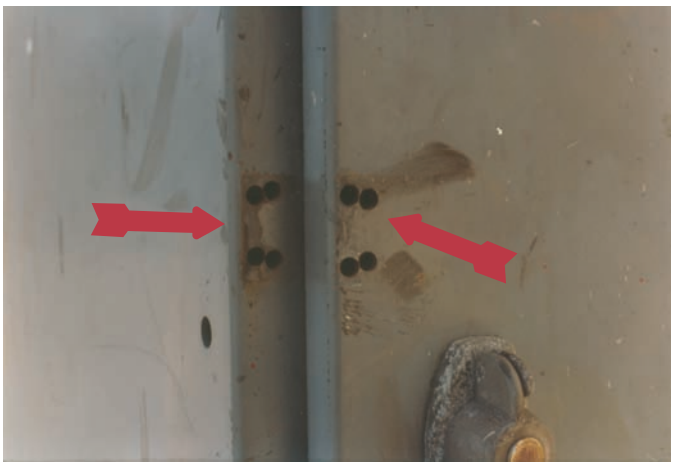


Figure 3 Close-up figure showing the holes in the door of the fuse compartment and frame of the switch, as indicated by the arrows in the Figure

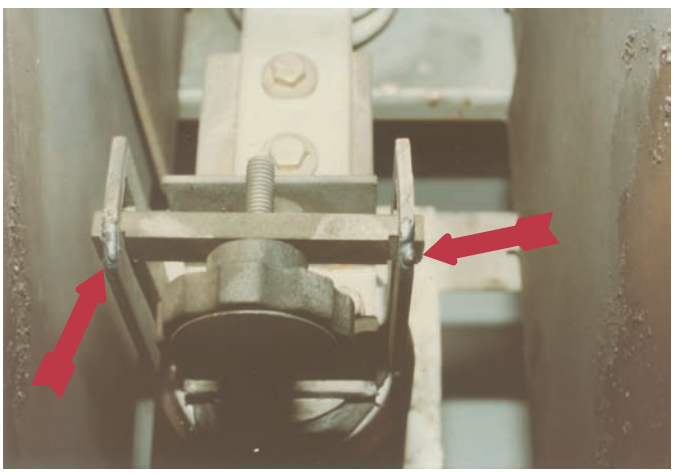


Figure 4 Close-up figure showing the melted metal at the top of the fuse holder, as indicated by the arrows in the Figure

renovations was being performed. An electrical engineer was retained to investigate the origin and cause of a fire along with a fire investigator and mechanical engineer for subrogation purposes. The building had experienced electrical problems and the fire was discovered just after the building's new fireplace was lit for the first time. The fire originated within the walls of the building where electrical wiring was located. The origin was in the concealed space above the new fireplace. Electrical wiring was located there to supply electrical power to an outlet on the right side of the fireplace. The wiring was solid copper and evidence of melting was found <2 cm apart on two wires (Figure 9). Melted copper between two electrical wires quite frequently indicates that the two wires short-circuited together. When copper wiring short circuits, it generates temperatures exceeding 1200°C with sufficient energy to ignite most common combustibles. However, the evidence of melting is normally at the same location along the length of the wire. Microscopic inspection of the melted areas found that they were bite marks (Figures 10 and 11). Subsequently, an animal bit the wiring and produced the short circuit, which ignited the building. The fire

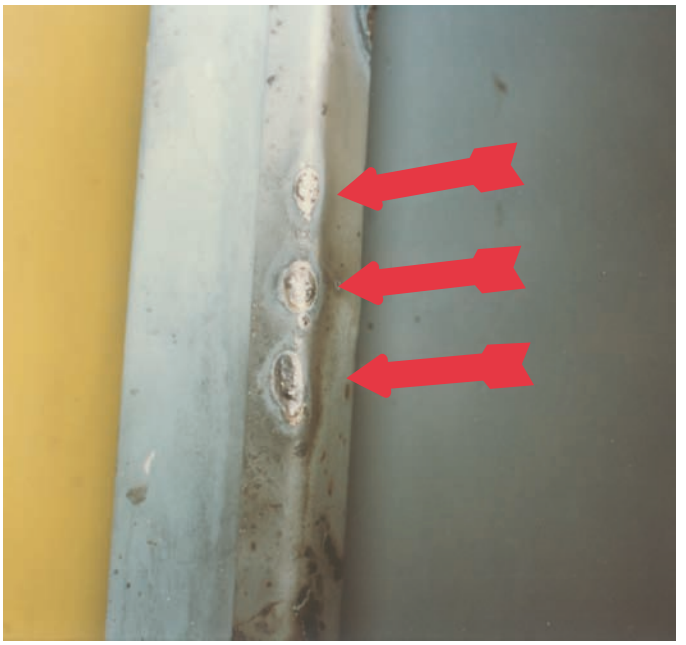


Figure 5 Figure showing fingerprints in metal, as indicated by the arrows in the Figure



Figure 6 Figure showing the bottom of the steam generating tank

was the result of the renovation contractor leaving openings to the interior of the walls unsecured during construction after electrical power was supplied to the building. The engineer determined the origin and cause of the fire and the party at fault, the renovation contractor. However, no litigation followed because the building owner had signed a contract with the renovator that eliminated their liability for damages that occurred during the renovation.

Another example of an electrical injury case involves an electrical explosion in an electrical distribution room that caused electrical injury and severe

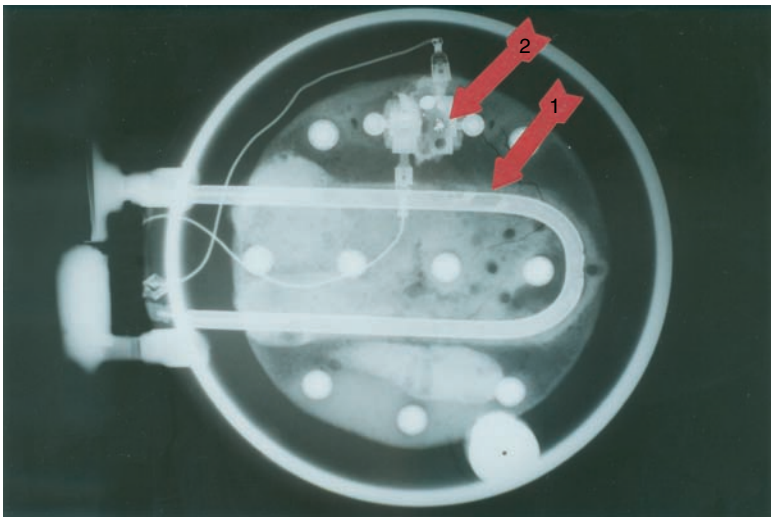


Figure 7 Figure showing an X-ray of the steam generating tank. Arrow 1 indicates an opening in the sheath of the heating element. Arrow 2 indicates a thermal control device

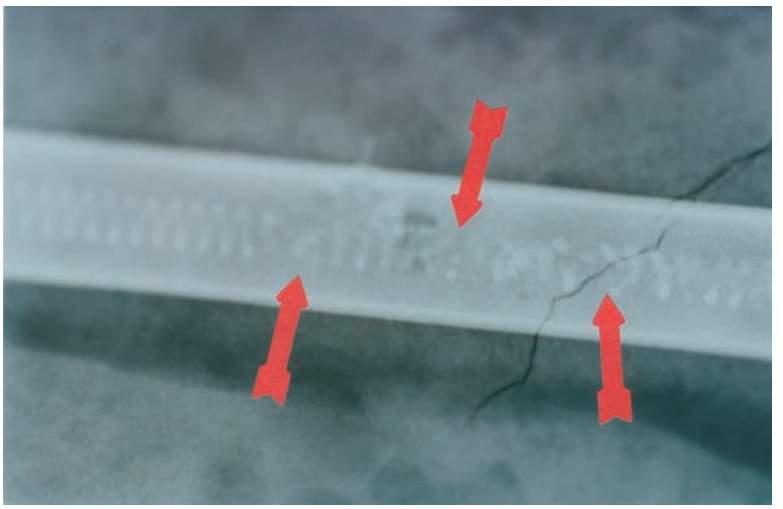


Figure 8 Close-up figure showing the nichrome wire inside the heating element, as indicated by the arrows in the Figure

property damage (Figure 12). The engineer was asked by a law firm to inspect the scene of the explosion, evidence retained from the scene and research documents to determine the cause of the explosion and the parties at fault for causing the personal injuries of workers. The oil within a 5000-V oil circuit breaker had exploded and ruptured its tank (Figure 13). The explosion occurred when the oil circuit breaker was closed after maintenance was performed on it. The maintenance personnel were only permitted 30 min to work on the circuit breaker. Their maintenance consisted of draining the oil out of the bottom of circuit breaker and pumping new oil in. However, proper maintenance on this type of oil circuit breaker requires a minimum of two men working 4 h. The oil should have been put back into the tank from the top of the tank after its interior was cleaned, inspected, and mechanically tested. Furthermore, the electrical

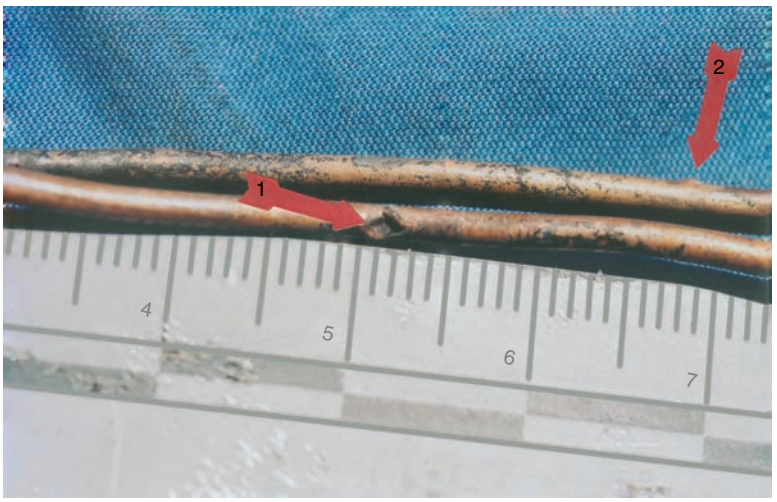


Figure 9 Figure showing copper wires with melted areas. Arrow 1 indicates a melted area shown in more detail in Figure 10. Arrow 2 indicates a melted area shown in more detail in Figure 11

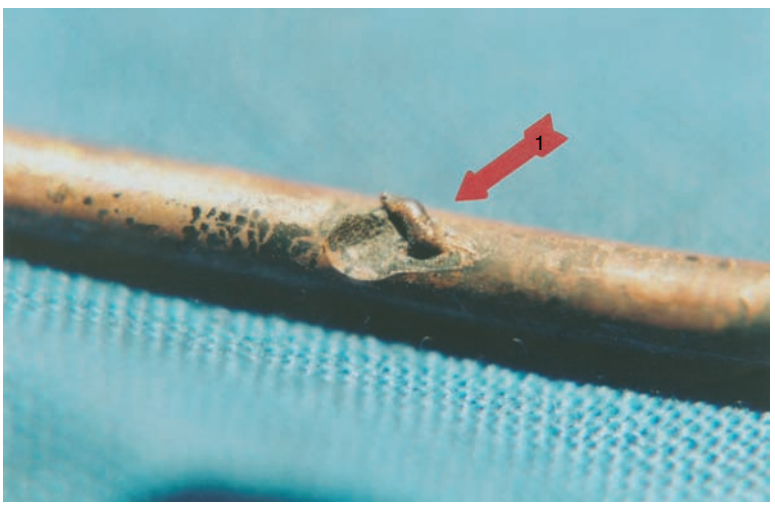


Figure 10 Close-up figure showing the melted area indicated by arrow 1 in Figure 9

distribution room should have been cleared of personnel before the circuit breaker was closed after maintenance. Consequentially, the property damage was caused by bad maintenance. The personnel injuries were the result of not following proper electrical protocol in closing the switch after maintenance without clearing the area of all unnecessary personnel. The engineer reported his conclusions to his attorney client who did not request a formal written report. Litigation against the maintenance people was started and during the discovery process the engineer was deposed by the maintenance attorneys. The engineer educated the maintenance attorneys while testifying about their own electrical protocol, which required the injury area to be cleared of all unnecessary personnel (the injured workers) before a high-voltage switch is closed after maintenance. After the deposition, the case settled out of court.

The last example involves a severe fire in a commercial building with a large property loss. The fire



Figure 11 Close-up figure showing the melted area indicated by arrow 2 in Figure 9



Figure 12 Figure showing where an explosion occurred in an electrical room

originated in an electrical distribution panel in a room of one of the commercial businesses in the building. An electrical engineer was retained by the insurance company who provided liability coverage to the tenant of the commercial business where the fire originated. The engineer was asked to determine the origin, cause, and parties at fault for the fire damage. The owner of the building had filed a million-dollar complaint against the tenant. Most of the contents of the fire building had been removed prior to this engineer's arrival at the fire scene except for the main circuit breaker, which protected the electrical conductors (wires) located within metal conduits (pipes) that feed electrical power to distribution panel previously located in the area of the fire's origin. The burn pattern in the area of fire origin indicated that the fire started in the ceiling area above the distribution panel where a conduit containing the service conductors feeding the panel was located. The service conductors were fused to the bottom inside of the conduit

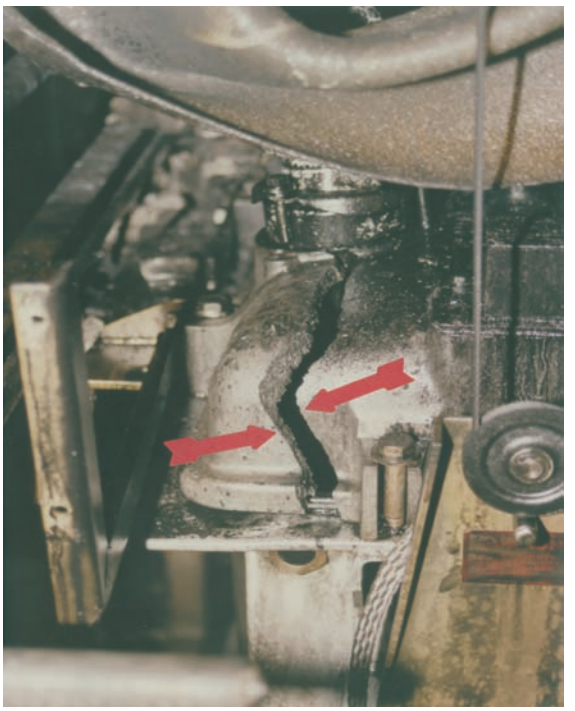


Figure 13 Figure showing where the tank of the circuit breaker ruptured



Figure 14 Figure showing the exterior of the circuit breaker

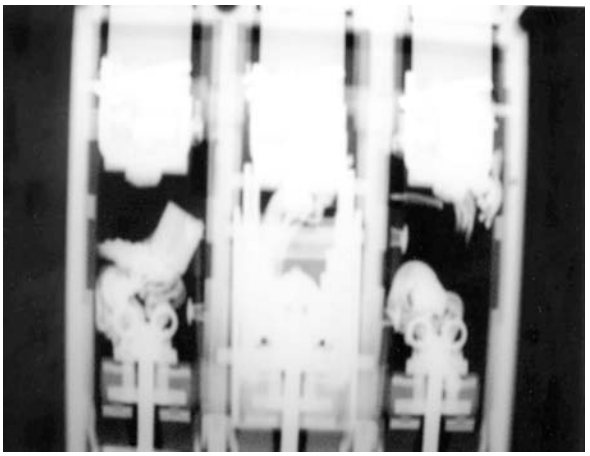


Figure 15 Figure showing the X-ray of the circuit breaker

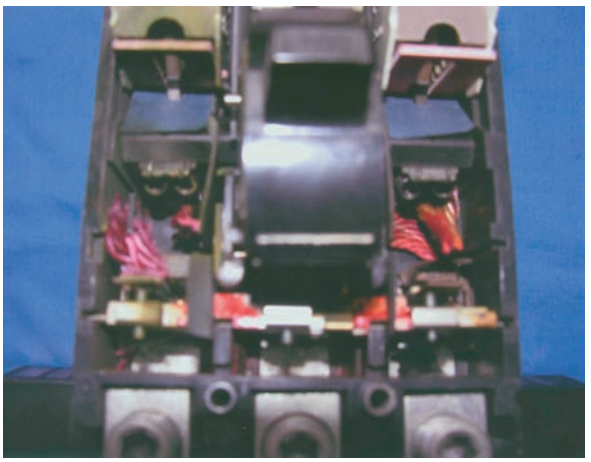


Figure 16 Figure showing the interior of the circuit breaker after disassembly

that was in contact with the structural wood components of the ceiling. The conductors were separated from the conduit to perform tests on the circuit. The tests indicated that the conductors had short-circuited to the interior of the conduit all the way back to the main circuit breaker. The main circuit breaker's operating handle was found in the closed position even though the conductors were not energized. Tests found no voltage present on the load terminals of the circuit breaker. Therefore, the local utility company was contacted to shut off power to the building so that the main circuit breaker could be removed from the main panelboard for further testing. The main circuit breaker had three phases and was rated at 200 A. No evidence of fire damage or overheating was found on the exterior of the circuit breaker (Figure 14). Preliminary tests of the circuit breaker indicated that two of the phases (circuits) inside the circuit breaker were open even when its operating handle was in the closed position. The circuit breaker was X-rayed. The X-ray showed that internal components of two phases of the circuit breaker had separated and melted (Figure 15). The engineer reported the results of the inspection, testing and X-rays to the insurance company adjuster. A report was requested with Figures and X-rays. The manufacturer of the defective circuit



Figure 17 Figure showing the internal components of an exemplar circuit breaker

breaker was put on notice of the financial loss. Subsequently, the circuit breaker was disassembled with representatives of its manufacturer present. The internal examination of the circuit breaker confirmed what the X-rays showed (Figure 16). An exemplar circuit breaker was disassembled to compare internal components. The components that melted and separated in the other circuit breaker were much lighter in color (Figure 17). The difference in color indicates that the separated components had overheated. The circuit breaker should have opened before its components melted and separated. A circuit breaker is "a device designed to open and close a circuit by nonautomatic means and to open the circuit automatically on a predetermined overcurrent without damage to itself when properly applied within its rating". The circuit breaker "is provided to open the circuit if the current reaches a value that will cause an excessive or dangerous temperature in conductors or conductor insulation" [17]. The purpose of the main circuit breaker is to open and disconnect electrical power from an electrical distribution panel and the circuit feeding it before the malfunction produces sufficient heat to ignite a fire in a building. Consequentially, the fire was caused by a defective main circuit breaker. The engineer answered the technical questions about the origin, cause, and party at fault for the fire damage, the circuit breaker manufacturer. The case was settled after disassembly of the circuit breaker with representatives of the circuit breaker manufacturer present, engineers, and attorneys.

End Notes

[a.] Revelations 8: 3–5 and http://www.energycite.com/ben%20franklin.htm.

[b.] Acts of the Apostles 8 : 1–3; 9 : 1–30; 22 : 3–21; 26 : 9–23; Galatians 1 : 12–15.

[c.] Franklin, B. Request for information on lightning, Pennsylvania Gazette, June 1753.

[d.] Effect of lightning on Captain Waddel's Compass, and on the Dutch Church in New York, from James Bowdoin to Benjamin Franklin, Boston, 2 March 1752.

References

[1] (1978). *Encyclopedia Britannica*, 15th Edition, Vol. 6, p. 536 ISBN: 0-85229-330-5.

[2] (2008). http://www.Luther.de/en/ and http://chi.gosel com.net.

[3] Matthews, J. (1998). Fear and lightning, *Chain Reaction Magazine*, **1**, 22–23, http://chainreaction.asu.edu.

[4] Matthews, J. (2008). Fear of lightning grounded in myth, *Arizona State University Research Magazine*, http://researchmag.asu.edu/stories/lightning.html.

[5] Sechel, A.L. & Edwards, J. (2008). *Franklin's Unholy Lightning Rod*, http://evolvefish.com/freewrite/franklgt.htm.

[6] Golde, R.H. (1977). *Lightning*, Academic Press, Vol. 1.

[7] Krider, E.P. & Franklin, B. (2005). in *Search of a Better World*, P. Talbott, ed, Yale University Press, New Haven, Chapter 5.

[8] Cohen, I.B. (1990). *Benjamin Franklin's Science*, Harvard University Press, Cambridge, Chapter 6.

[9] Golde, R.H. (1977). *Lightning*, Academic Press, Vol. 1, pp. 170–172.
[10] Morse, R.A. (2004). *Benjamin Franklin: Papers on Electricity*, pp. 276–277, http://www.franklinpapers.org.
[11] Labaree, L.W. *et al. The Papers of Benjamin Franklin*, Yale University Press, New Haven, Vol. 10, p. 52, http://www.franklinpapers.org.
[12] (1978). *Encyclopaedia Britannica*, Encyclopedia Britannica, Inc., 15th edition, Vol. 6, p. 537, ISBN:0-85229-330-5.
[13] (2008). http://enwikipedia.org/wiki/History_of_electrical_engineering, p. 2.
[14] (2008). Edison Papers, http://edison.rutgers.edu/images/fa/fa0631.jpg.
[15] ANSI C2-1981, *National Electrical Safety Code*, (1981). Edition, Section 381-G, p. 304.
[16] (2008). http://www.wiretron.com.
[17] National Fire Prevention Association (2005). Standard 70 *National Electrical Code*, pp. 27, 34, 81.

THOMAS P. SHEFCHICK

Electron Microscopy ***see*** Microscopy: Scanning Electron Microscopy

Emergency Triage ***see*** Threat Assessment: Workplace

Engineering: Biomedical in Vehicle Trauma ***see*** Trauma Causation: Analysis of Automotive

Engineering: Electrical ***see*** Electrical Engineering

Entomology

Introduction

Definition

Entomology is the study of insects and related arthropods. Forensic entomology is the use of such knowledge in legal investigations including cases of medicolegal, stored product or urban circumstances. Although entomological knowledge is often employed for investigations involving stored products (e.g., a grasshopper in a can of green beans, grain beetles in cereal) or urban damage (e.g., termite infestations), we focus on the aspects of entomology used in medicolegal investigations. Such investigations are broadly defined as felonious violent crimes that include murder, suicide and rape, but can also involve cases of extreme neglect and abuse. In most instances, forensically relevant inferences are made by the collection, identification, and study of arthropods associated with a decomposing body. The arthropods usually consist of the true flies and beetles, but may also include incidental organisms such as assassin bugs, cockroaches, bees, wasps, ants, spiders, mites, and lice, among others (Table 1).

History

Although forensic entomology has only recently seen increased interest and acceptance in the forensic sciences and in courts of law (since the early 1990s), the initial contributions of insects in legal investigations dates back to centuries and has an established history of routine use for about 150 years. Regular use of insects in criminal proceedings began in Europe in the later half of the nineteenth century with techniques that are similar to what are used today. The publication of better insect identification keys during the early to mid-1900s inspired interest in forensic entomology in North America, but its acceptance within the forensic community was meager until the early 1980s. It was at this time that there were publications of developmental rates and other biological aspects on forensically important species of flies by investigators such as R.D. Hall and B. Greenberg. Recent interest has been facilitated by a surge of research into the biology and ecology of forensically

Table 1 Decomposition stages of a body along with associated arthropods[a]

Stage of decomposition	Accumulated time[b] (days)	Common arthropods	Relative degree of PMI accuracy
Fresh			
Begins at the moment of death and ends with bloating caused by the breakdown and metabolism of proteins, lipids and carbohydrates.	0–3	Adult blow flies and eggs, muscid flies, yellow jackets, ants, daddy longlegs	Fair
Bloated			
Putrefication begins during this stage and trapped metabolic gases from anaerobic bacteria inflate the body.	4–7	Adult and larval blow flies, flesh flies, muscid flies, rove beetles, hister beetles, carrion beetles, ants, assassin bugs, yellow jackets	Good
Decay			
Gases rupture the abdomen and the body deflates. This stage is very odiferous.	8–18	Adult and larval blow flies, muscid flies, rove beetles, hister beetles, carrion beetles, dermestid beetles, scarab beetles, cockroaches, red-legged ham beetles	Best
Postdecay			
In dry habitats the remains are composed of dry skin, cartilage and bones. In wet habitats, the remains are wet and viscous as byproducts of decomposition.	19–30	Dermestid beetles, hister beetles, fungus beetles, springtails, mites, fungus gnats, fruit flies, cheese skipper flies, phorid flies	Fair
Dry			
Only bones, cartilage, hair and dry skin remain. Most odor is gone.	≥31	Dermestid beetles, ants, cheese skipper flies, sow bugs, solder flies, ground beetles	Poor

[a] The relative degree of PMI accuracy provides an estimated accuracy of entomological evidence between each stage of decomposition when a body can be discovered. A forensic entomologist will generally provide an estimated time range (e.g., 1–7 days) when death may have occurred, and the accuracy increases with a narrower time range

[b] Accumulated time estimated from studies on pig carcasses in southern Michigan during the summer and early autumn

important arthropods from around the world, and the use of arthropods in highly publicized cases bringing attention to the usefulness of "bugs" or "maggots" in criminal investigations. The entertainment industry also has publicized the use of arthropods in legal proceedings through popular TV programs (e.g., New Detectives, Cold Case Files, CSI).

Forensic entomology is now a recognized discipline of the American Academy of Forensic Sciences and has gained international acceptance with the American Board of Forensic Entomology (ABFE), the North American Forensic Entomological Association (NAFEA), and the European Association for Forensic Entomology (EAFE).

Background

Because insects and other arthropods have predictable life histories, habitats, known distributions, and developmental rates, the presence/absence and

size of specific species at a crime scene, such as a homicide, can provide important information about when, where, and even how a particular crime occurred. Insects and other arthropods play a natural role in the decomposition of carrion in the environment, consuming the decomposing organic material and recycling the energy and nutrients as part of their life cycle. When an organism dies, bacteria that were once held in equilibrium by the immune system immediately begin to digest the body's proteins, lipids and carbohydrates as energy sources, creating both gaseous and liquid byproducts that act as an olfactory cue (i.e., a smell) for colonization by multicellular organisms such as flies and beetles. In most instances the initial arthropod colonizers are adult blow flies that will feed and lay eggs (i.e., oviposition) on the remains. Within a few hours the decomposing body acts as a food source for newly hatched larvae, and they grow and develop through life stages at temperature-dependent rates. The presence of blow fly (Diptera: Calliphoridae) larvae attract roaming predators and parasites such as beetles, mites, ants, wasps, and spiders that then feed on or parasitize the eggs, larvae, or pupae of the flies. This is followed by other insect species that come to feed on previously eaten or conditioned (e.g., dry skin) remains in a predictable succession (sequence) of arthropod species that colonize and ultimately decompose the carrion to dry bones and hair. Some representative insect groups associated with carrion in nature are shown in Figures 1–4.

Forensic entomologists use data on insect development rates and the natural and predictable species succession to estimate the time of initial insect colonization. Because blow flies can oviposit within minutes to hours after death, this estimate of initial insect colonization is a reasonable surrogate of the amount of time since death, or what is known as the *postmortem interval* (*PMI*). However, many factors can influence oviposition, larval development rates and species succession, so only expert forensic entomologists should be consulted when insects are collected from a crime scene.

Assumptions of Entomology-Based PMI Estimates

On the basis of understanding the natural history of various insect species, a forensic entomologist

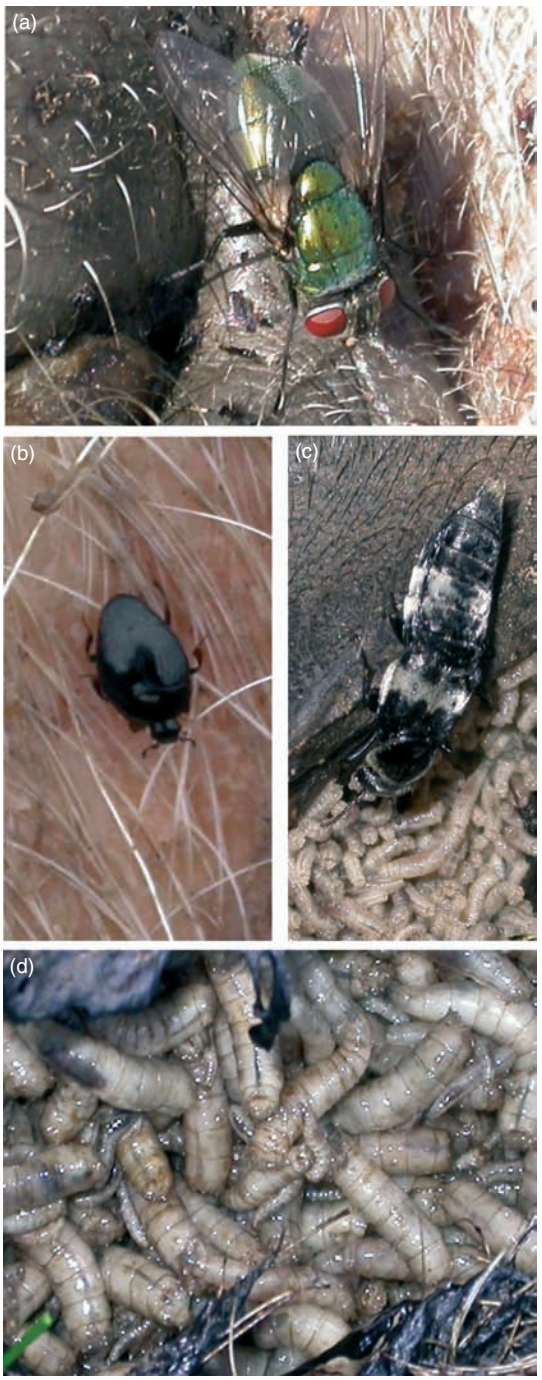


Figure 1 Representatives of some forensically important insects: (a) blow fly (Diptera: Calliphoridae); (b) histerid beetle (Coleoptera: Histeridae); (c) rove beetle (Coleoptera: Staphylinidae); and (d) blow fly maggots (Diptera: Calliphoridae)

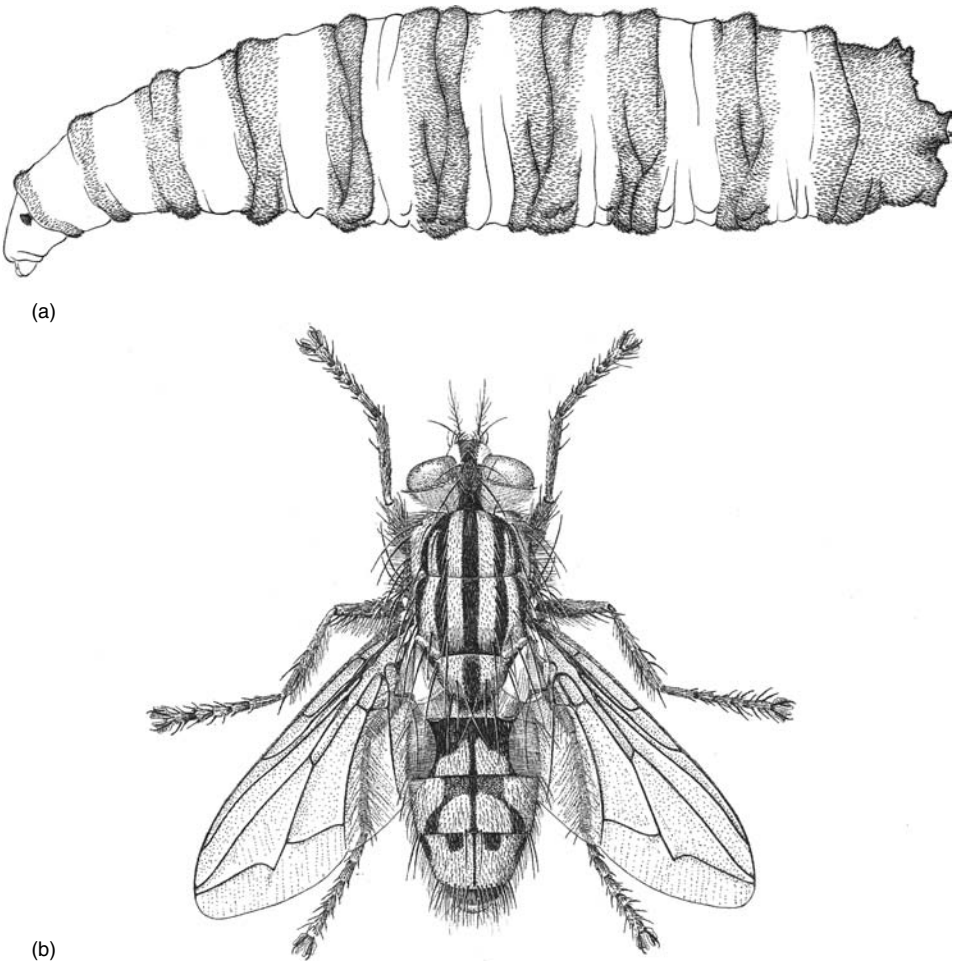


Figure 2 A flesh fly (Sarcophagidae): (a) larva and (b) adult [Reproduced from Gorman JR (ed.) (1987). *Insect and Mite Pests in Food: an Illustrated Key,* US Department of Agriculture, Handbook 655 [1].]

can make estimates of an entomology-based PMI, which represents a range of time for initial insect colonization. As with any forensic discipline there are basic assumptions that require evaluation when making determinations and inferences during a crime scene investigation. We provide several points here that are important to forensic entomology:

1. Oviposition usually does not take place at night. Most homicides occur at night, so there may be a delay in oviposition until the following morning.
2. Adult flies will generally oviposit as soon as they find a body. This is the general assumption that associates the PMI with initial insect colonization; however, any barrier to initial oviposition should be considered during an investigation.
3. Insect succession follows a predictable order. This is generally true, but can vary depending on the season, habitat and regional location, as well as other environmental conditions such as periods of extreme rainfall or drought, and whether the body is inside a house or outside exposed to the elements.
4. Ambient air temperature is the major variable influencing larval growth and development rates. Although temperature is the overriding factor affecting development, it does not always apply equally throughout the life cycle and

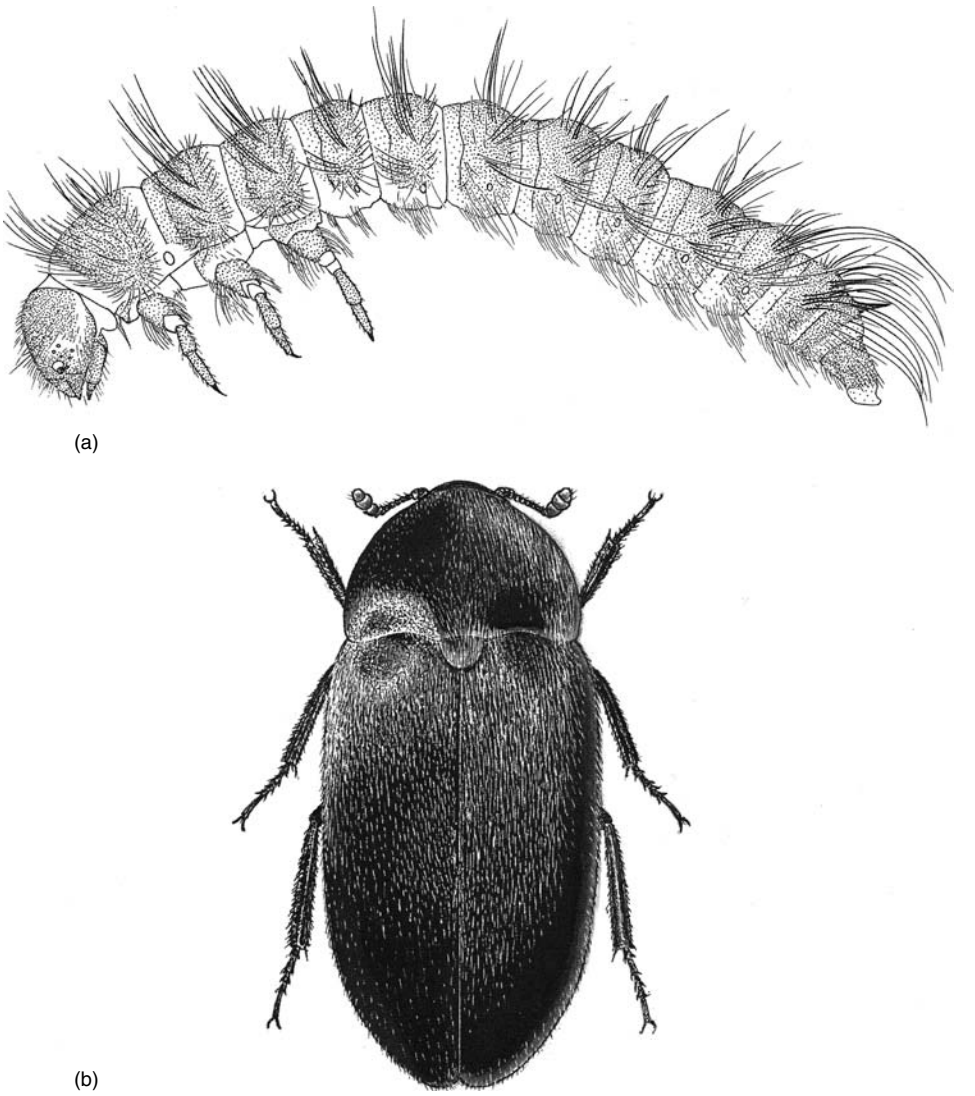


Figure 3 A dermestid beetle (Dermestidae): (a) larva and (b) adult [Reproduced from Gorman JR (ed.) (1987). *Insect and Mite Pests in Food: an Illustrated Key,* US Department of Agriculture, Handbook 655 [1].]

other variables can work in a synergistic fashion. When the larvae are in their early, small stages and when there are only a few individuals, ambient air temperature has its most profound effects on development. The effect becomes more variable as the larvae grow bigger and/or become more numerous. Hundreds to thousands of larvae can form masses (i.e., maggot masses) on a body (Figure 5). The thermogenic heat created by the metabolic and movement activities of the mass can raise the temperature of the mass and the body above the ambient temperatures. This effect can persist under much cooler ambient conditions and even when a body is placed into a refrigerator. Any larval masses should be noted, the size estimated and a temperature measurement made in one or two places within the mass. Other factors such as humidity, rainfall and sometimes chemicals or drugs consumed by the deceased

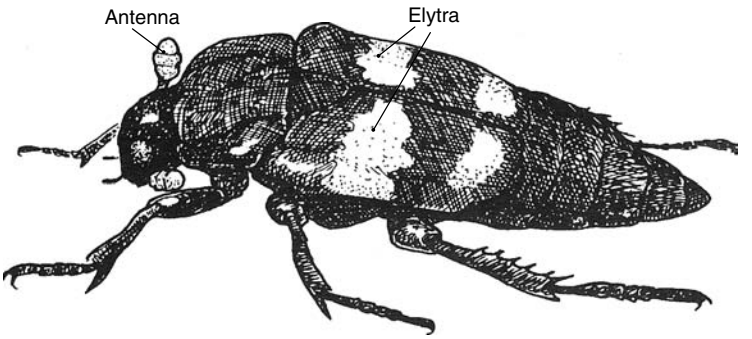


Figure 4 Adult carrion beetle (Silphidae) *Nicrophorus* sp. (20–25 mm in length) [Reproduced with permission from Catts and Haskell [2].]



Figure 5 A large mass of larvae (i.e., maggot mass) will sometimes form on a decomposing body, influencing the temperature at which they are developing. The white matter in this photo is the maggot mass

can influence development (see the following data).

5. Climate data from a weather station distant from a crime scene represents the ambient conditions of that scene. This is usually the weakest input to decision making in forensic entomology. Correlations between the crime scene and distant weather station data can vary substantially. Many variables can influence this relationship, including ground cover, topography, wind speed, and direction, and other microhabitat conditions that influence temperature conditions at a scene. It is recommended that crime scene temperature readings be taken for several days to weeks and compared to weather station data in order to check the validity of such assumptions; but this is not always done. Indoor crime scenes are much less variable compared to those found outdoors, and often a thermostat can be checked for ambient conditions.

Here we expand on these assumptions and discuss in greater detail the biology behind, and application of, using insect evidence for making entomology-based PMI estimates.

Entomological Succession

Ecological succession is defined as the predictable and sequential colonization and replacement of specific communities of organisms over time in, or on, an open area. This same principle is used in forensic

entomology with the open area being a body, and the communities represented by arthropods that are primarily insects. When an organism dies, it goes through five stages of decomposition that coincide with the activity of bacteria and insects (Table 1). The duration of each stage is directly related to the air temperature and the degree of insect activity associated with the body. While the length of each stage varies with temperature, the series of stages is very predictable, and a forensic entomologist can estimate a PMI based on the stage of decomposition, the insects present, and known temperature data.

Adult blow flies, flesh flies and other species are generally the first insects to arrive at a decomposing body, appearing within minutes or hours after death to lay eggs or live larvae in some instances (Figures 1 and 2). From a few hundreds to several thousands of eggs can be laid on a body and the resulting larvae are the most obvious and important group of insects, for they will eventually consume most of the decaying flesh. Larval activity, and thus decomposition, is greatest through the decay stage (Table 1). It is during this stage of decomposition that an entomologist can estimate the PMI with a high degree of confidence by evaluating the various larval developmental stages. The largest larvae are most important since they represent the earliest time of oviposition. After the decay stage, when most of the flesh is gone, changes in the decomposing remains and the associated insect fauna occur at a much slower rate, making the accuracy of the entomology-based PMI more variable. The later stages are dominated by dermestid and carrion beetles and other insects that feed on the remaining dried flesh, skin, and hair (Figures 3 and 4; Table 1).

Fly Life Cycle and Development

Life Cycle

The true flies (order Diptera) are the most important group of insects that are used in making entomology-based PMI estimates. In particular, two families of flies have evolved to specialize on carrion: blow flies (Calliphoridae: Figure 1a,d) and the flesh flies (Sarcophagidae: Figure 2). Blow flies are generally medium- to large-sized insects (usually <1 cm in length) with a metallic green or blue color and are often referred to as *greenbottle* and *bluebottle* flies (Figure 1a). Adult females are known to oviposit on a body immediately after death. Oviposition is primarily a daylight activity, and as such, may be delayed for several hours if death occurs at night. Eggs are generally equal to or <2 mm and are initially deposited around naturally moist open areas (mucous membranes) or orifices such as eyes, nose, ears, and mouth, but also are laid in and around wounds. A single adult female blow fly may lay up to several hundred eggs in a short period of time, and many females oviposit on the same body; the results can be thousands of larvae and the formation of larval masses, commonly referred to as *maggot masses* (Figure 5). Flesh flies are usually gray with black longitudinal stripes (Figure 2). Female flesh flies deposit small live larvae on a body in the same areas as blow flies. The larvae of flesh flies are usually far fewer in number compared to the more dense blow fly larvae. The larval body size of both types of flies can vary between 5 and 15 mm, depending on the species and developmental stage.

The life cycle of true flies is described as *holometabolous*, meaning there is a "complete change" in body form during growth and development (Figure 6). In this type of life cycle there is complete metamorphosis from the larvae to pupae, where the larvae are structurally different from both pupae and adults (Figure 6). In other life cycles, such as the hemimetabolous life cycle, the larvae resemble the adults throughout development; they are just smaller versions of the adults and there is no metamorphosis or pupal stage.

Whether deposited as eggs (blow flies) or as small larvae (flesh flies), there are three larval developmental life stages, or instars. During development the larva body size increases dramatically during each instar before molting into the next instar (Figure 6). The soft-bodied larvae feed on decaying flesh with hard, hook-like mouthparts and secrete enzymes that aid in digestion. As a larva progresses from the first, to the second and into the third instar, its size increases tenfold from continuous feeding. It is when the larvae have reached the third instar that most of the decomposing biomass is consumed leading to the beginning of the dry stage of decomposition (Table 1). At the end of the third instar, the larvae go through a brief stage of postfeeding, prepupal migration: the larvae move away from the decomposing remains for up to 10–30 m to burrow into soil or leaf litter to undergo metamorphosis. During this prepupal stage, the larval body begins to shorten and turn darker, while mobility is reduced.

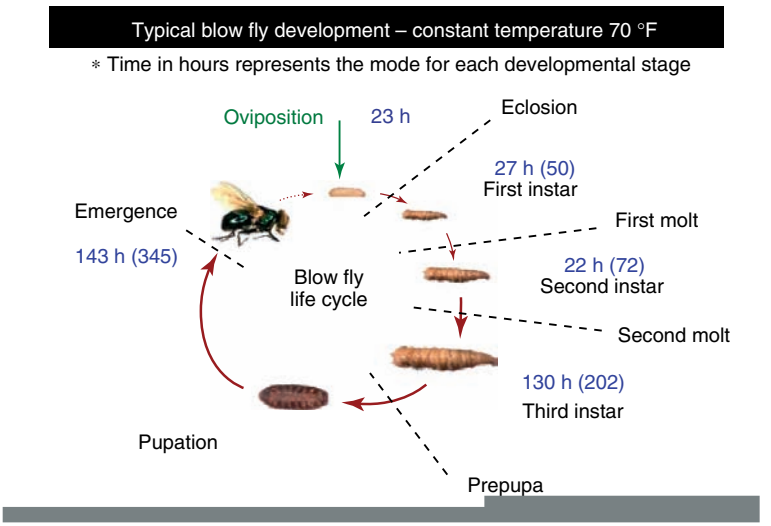


Figure 6 A schematic of the life cycle of a representative blow fly, indicating the approximate time necessary for each stage of development

These changes mark the beginning of metamorphosis into the pupal stage. It is the exoskeleton of the third instar that forms the outer hardened casing known as the *puparium*, and within this structure the pupa is formed (Figure 6). Inside each puparium a pupa slowly metamorphoses into an adult fly, which in several days will emerge as a flying adult, beginning the next life cycle.

Effects of Temperature on Development

Insects are poikilothermic, meaning that their body temperature is dependent on ambient temperature conditions of the macro- and microenvironment. Because of this growth and development are closely related to temperature: as ambient temperature increases so does the rate of development (Figure 7). This relationship holds for most ambient temperatures; however, development slows or ceases at extreme cold and hot temperatures (e.g., 2 °C and 100 °C, respectively) (Figure 7). These thermal thresholds are not universal for all fly species, but rather, they can vary between species and between geographically distant populations of the same species (e.g., temperate *vs.* tropical populations) (Figure 8). For instance in Figure 7, development was undetectable at 12 °C.

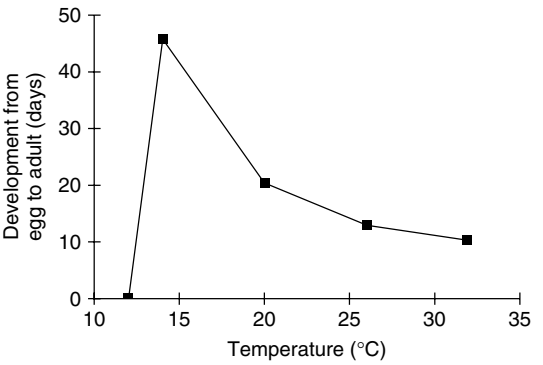


Figure 7 A representative relationship between blow fly development and temperature. Development is the time it takes to grow from egg to adult. Data are for laboratory reared *Phormia regina* and plotted from Table 1 of Nabity *et al.* [3]. At 12 °C there was no detectable development

Further, insect metabolism increases with temperature, therefore, feeding and the resulting decomposition increases throughout larval development. A body exposed to direct sunlight on a hot summer day may be reduced to bones in a little over a week or less, while an exposed body in the late autumn or winter may show little change for months. Because of this relationship between development and temperature, it is possible to predict how long it takes for an insect to develop from egg to adult by measuring temperature

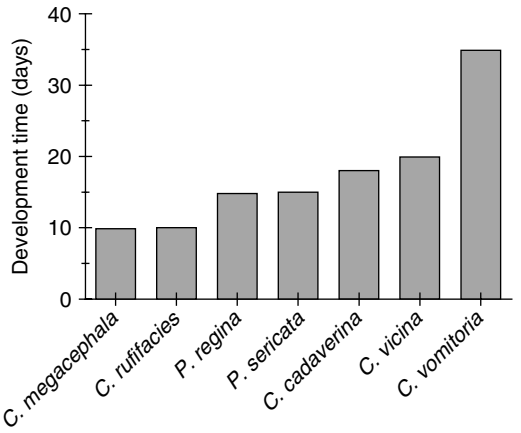


Figure 8 Development time for various blow fly species from different regions of the United States: two tropical species (*Chrysomya megacephala* and *Chrysomya ruffifacies*) reared at 28 °C; two summer species (*Phormia regina* and *Phaenicia sericata*) reared at 26.5 °C; and three cool-weather species (*Cynomyopsis cadaverina*, *Calliphora vicina* and *Calliphora vomitoria*) reared at 26.5 °C [Modified from Haskell *et al.* [4]]

over time and calculating species-specific accumulated thermal units.

Accumulated Degree-Days and Degree-Hours

Entomologists have carefully studied the development of many species of carrion-feeding flies, determining their thermal requirements for completing each stage of development from egg to adult (Figure 6). These thermal requirements are expressed in temperature-time units called *accumulated degree-days* (*ADD*) or *accumulated degree-hours* (*ADH*). It may be easier to think of ADD or ADH as the accumulation of heat units over time. A hot summer day will accumulate more hourly heat units compared to a cool autumn day. For instance, if a species of fly requires 400 degree-hours to reach the third instar, it may accumulate 100 of those on a single, warm summer day, while it may accumulate only 10 degree-hours on a cold autumn day. As an overly simplified example of ADD, if a fly species requires 200 degree-days to develop from eggs to pupae, it may acquire those in 10 days if the average daily temperature is 20 °C, or only 5 days if each averages 40 °C. In addition, there is always a species-specific threshold temperature where no development takes place (Figure 7), and below which adult flies are inactive. For most species, this threshold temperature is around 10 °C but can be as low as or lower than 4 °C, but varies substantially by species and geographic location. In general though, a body exposed to air temperatures below 10 °C will generally not be colonized by flies until the temperature rises.

An exception to this ambient temperature to developmental time relationship occurs when large numbers of second and third instars feed in a concentrated "maggot mass" (Figure 5), thereby generating additional heat within the mass that can influence activity and development. The activity of several hundred or thousand fly larvae in a small area can raise the temperature within this mass up to 30 °C above the ambient air temperature, and larval development may proceed at a much faster rate than might be expected at a given temperature. The masses become so hot that larvae sometimes need to move from the middle of the mass to the outside in order to cool. If a maggot mass is present on a body, a forensic entomologist may consider recording the temperature within the mass when computing an estimate of the entomology-based PMI.

Estimating an Entomology-Based PMI

A forensic entomologist can estimate the PMI by two methods. The first and usually more precise method is the development-driven PMI and involves identifying the species of flies (larvae and/or pupae) collected from the body at the time of discovery, determining the growth stage, and calculating how long it would take them to reach this stage being given the environmental temperatures at the scene. This provides information for making an estimate of fly age. Because adult flies will usually oviposit on a body within minutes or hours after death, knowing the age of the larvae gives a close approximation of the time of death.

In order to estimate an accurate PMI, the forensic entomologist should be able to identify the species present on the body; should be working with the oldest, and hopefully the largest larvae at the scene; must have access to nearby temperature records, and; must know the thermal requirement (i.e., ADD or ADH) of the identified species. Identification of

the fly larvae is difficult, but an entomologist can generally use several structures including the shape and structure of the spiracles (Figure 9), both of which are species-specific and change with instar stage. Once the species is identified and measured the ambient temperatures of the scene are acquired and annotated. Local temperature data usually comes from records of the nearest weather station; however, weather station temperatures may not always accurately reflect the local crime scene conditions. In these situations, scene temperatures are often compared to weather station data in order to assess their relationship and make any necessary corrections. From the temperature data and working backward from the time the body was discovered, an entomologist calculates the accumulated thermal units associated with the instar of the fly species collected at the scene. The entomology-based PMI is determined when the calculated scene-specific accumulated thermal units are compared to those of the same species developed under temperature controlled laboratory conditions.

The second method is the succession-driven PMI and involves examining the composition of the insect community on the body associated with the appropriate stage of decay (Table 1). This method is usually employed in cases of advanced decay, after most of the larvae have gone through metamorphosis and emerged as adults. By this time most of the flesh has been removed by larvae, so much of what remains is hair, skin, and bone. Changes in the insect community and the body itself during this stage occur very slowly, making estimates of the PMI much more variable. The entomologist uses the presence of various beetle species and other insects to estimate the PMI, and usually must present a possible time range rather than an exact date.

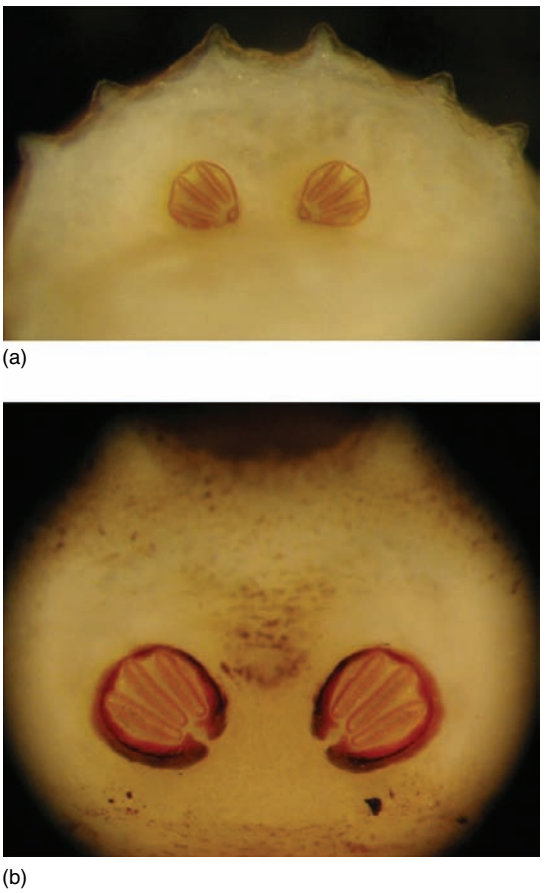


Figure 9 The posterior end of a third instar showing the structure of the spiracles for (a) *Phaenicia sericata* and (b) *Chrysomya rufifacies*. The spiracles are the respiratory openings for the larvae and are one of the most distinctive morphological features used to identify the species of fly, and aids in determining the developmental stage (instar). The shape and size of the posterior tubercles (fleshy projections) also assist in making identifications

Other Factors Affecting an Entomology-Based PMI

Any factor that makes a body inaccessible to insect colonization can affect the accuracy of estimating a PMI. Night (darkness) and cold temperatures, as well as rain, delay adult oviposition. In addition, factors that physically prevent insects from reaching a corpse can delay colonization for potentially long periods of time. Such factors include submergence under water, burial underground, concealment in sealed containers, caskets, or wrapping, and enclosure in a freezer, closed vehicle, or building. While these factors will often delay insect colonization, rarely will they totally prevent it. Flies are very adept at locating bodies, and will usually find access into a closed vehicle or building. A specific group of flies (Phoridae) have been found 1–2 m below the ground in association with coffined bodies, hence their common name "coffin-flies". Various studies have been conducted to determine how long such factors listed earlier will delay insect colonization, and the forensic entomologist accounts for such factors when estimating a PMI.

Collecting Entomological Evidence

Collecting at the Scene

When insects are found associated with a decomposing body a trained forensic entomologist should be called to collect specimens. If a forensic entomologist is not available, it is imperative that collections be made at the scene as soon as possible. If larvae are present on the body, the largest specimens should be collected, along with representatives of other sizes. The largest larvae are generally the oldest, and thus represent the earliest colonizers. The oldest larvae allow for the most accurate estimate of PMI. The investigator should search the surface surrounding the body for larvae crawling away to pupate, and also dig into the surrounding soil (up to 2–3 m away from the body) for fly puparia. If no larvae are visible, search the body, particularly around the eyes, ears, nose, and mouth for fly eggs. The eggs look like small grains of rice. Care should also be taken to collect insects from underneath the body once it has been removed. Detailed field notes such as climatic conditions, location and condition of the body (e.g., rainfall, full sun, shade, fog, position of the body, wounds, etc.), and visual observations should be made of any insect activity in and around the remains. All of this information can be important at a later time in the investigation. In addition, photographs and/or video of the crime scene should include close-ups of entomological evidence and any potential barriers to insect colonization.

Many times at a scene there will be numerous flying insects on and around the body. If flying insects are present over the body, an aerial sweep net can be used to collect fast flying or fast crawling adult insects. These specimens are then placed in a wide-mouth killing jar containing ethyl acetate and later pinned or placed in ethanol for subsequent identification by an entomologist.

Specimen Preservation

The sample of insects collected from the body itself should be placed in a vial or jar containing alcohol (ethanol or isopropyl) along with a label (written in pencil) that includes the following: case number, date, time, collector, the presence and location of a maggot mass, ambient temperature conditions and the location on the body where specimens were collected. It is preferred to have both a label inside the container and one adhered to the outside. If possible, larvae first should be killed in boiling water and then transferred to ethanol to prevent shrinkage in larval size. This is the most appropriate method to preserve larvae for the best identifications. However, this may not be practical, thus, specimens should be placed directly into alcohol. Larvae collected away from the body should be preserved separately, as should any fly puparia and this information should be noted on the label and a death scene form.

If possible, live specimens of eggs, larvae and puparia also should be collected. An entomologist may wish to rear these specimens to adults to aid in identification. Separately, eggs and larvae should be placed in a jar and transferred as soon as possible to an aluminum foil envelope containing a small amount of fresh beef liver. Puparia can be placed in a jar with some of the surrounding soil, vermiculite or sand. Live specimens should be kept in a warm (about room temperature) but not too hot location, where temperatures can be continuously recorded (e.g., about every two to three hours). All specimens, living and preserved, should be transferred to a qualified forensic entomologist as soon as possible for analysis. Copies of crime scene notes and photographs also should be provided.

Additional Value of Insects at Crime Scenes

While assisting in an investigation by estimating the PMI is the most common job of the forensic entomologist, insects at a crime scene also can reveal additional information important to an investigation. For instance, larvae and pupae can be analyzed for toxicological trace elements when more traditional sources such as blood and tissue samples are no longer available due to decomposition. Traces of cocaine and other drugs have been obtained from insects on a corpse, and have been used to determine the probable cause of death. In addition, research has shown that larvae feeding on tissue containing cocaine experience accelerated growth. The presence of unusually large larvae, or "super maggots" as they have been called, particularly those feeding around the nasal cavity, may indicate cocaine use by the deceased, and such specimens should be preserved for toxicological analysis.

The presence of larval activity on areas of the body other than the eyes, nose, ears, and mouth prior to the decay stage of decomposition (Table 1) may indicate possible trauma sites. Adult flies are attracted to wounds and will lay eggs in these areas. Larval activity in the female urogenital and anal areas during the early stages of decomposition may be indicative of rape.

Entomological analysis also can be useful in nonhomicide cases, particularly those dealing with neglect and accidental death due to bites or stings. Neglect can often be identified by the presence of larvae in living tissue, such as one might find in a nursing home for the elderly. Certain fly species lay eggs only in living tissue, and their presence on a body can indicate invasion while alive or prior to an immediate death.

Deaths resulting from the bites or stings of insects, spiders, and scorpions are fairly common. The marks left by these bites or stings are usually very small and may be overlooked during an investigation. For instance, death may be attributed to a heart attack rather than an accident, a conclusion that may be important for insurance purposes. A forensic entomologist, or a physician familiar with such cases, can often recognize evidence of bites and stings.

Acknowledgment

The authors would like to thank R. Kimbirauskas for the use of his photographs in Figures 1 and 6.

References

[1] Gorman, J.R. (ed) (1987). *Insect and Mite Pests in Food: An Illustrated Key*, Handbook 655, US Department of Agriculture.

[2] Catts, E.P. & Haskell, N.H. (1990). *Entomology and Death: A Procedural Guide*, Joyce's Print Shop, Clemson.

[3] Haskell, N.H., Hall, R.D., Cervenka, V.J. & Clark, M.A. (1997). On the body: Insects' life stage presence and their postmortem artifacts, in *Forensic Taphonomy: The Postmortem Fate of Human Remains*, W.D. Haglund & M.H. Sorg, eds, CRC Press, Boca Raton, FL, pp. 415–448.

[4] Nabity, P.D., Higley, L.G. & Heng-Moss, T.M. (2006). Effects of temperature on development of *Phormia regina* (Diptera: Calliphoridae) and use of developmental data in determining time intervals in forensic entomology, *Journal of Medical Entomology* **43**, 1276–1286.

Further Reading

Amendt, J., Campobasso, C.P., Gaudry, E., Reiter, C., LeBlanc, H.N. & Hall, M.J.R. (2007). Best practice in forensic entomology – standards and guidelines, *International Journal of Legal Medicine* **121**, 90–104.

Arnaldos, M.I., Garcia, M.D., Romera, E., Presa, J.J. & Luna, A. (2005). Estimation of postmortem interval in real cases based on experimentally obtained entomological evidence, *Forensic Science International* **149**, 57–65.

Byrd, J.H. & Castner, J.L. (2008). *Entomological Evidence: The Utility of Arthropods in Forensic Investigations*, 2nd Edition, CRC Press, Boca Raton.

Catts, E.P. & Goff, M.L. (1992). Forensic entomology in criminal investigations, *Annual Review of Entomology* **37**, 253–272.

Goff, M.L. (1993). Estimation of postmortem interval using arthropod development and successional patterns. *Forensic Science Reviews* **5**, 81–94.

Greenberg, B. (1991). Flies as forensic indicators, *Journal of Medical Entomology* **28**, 565–577.

Hall, R.D. & Haskell, N.H. (1995). Forensic entomology – applications in medicolegal investigations, in *Forensic Sciences*, C. Wecht, ed, Matthew Bender, New York.

Keh, B. (1985). Scope and applications of forensic entomology, *Annual Review of Entomology* **30**, 137–154.

Rodriquez, W.C. & Bass, W.M. (1983). Insect activity and its relationship to decay rates of human cadavers in East Tennessee, *Journal of Forensic Sciences* **28**, 423–432.

Smith, D.G.V. (1986). *A Manual of Forensic Entomology*, Comstock University Press, Ithaca.

Related Articles

Botany

Crime Scene Investigation

Environmental Science

Evidence Collection and Preservation: Casting

Soil: Forensic Analysis

Time of Death Determinations

Toxicology: Forensic Applications of Wildlife

RICHARD W. MERRITT AND M. ERIC BENBOW

Environmental Science

Introduction

Forensic ecology is an emerging discipline concerned with complex threats to environmental systems and the human populations that depend on them. It provides evidence on environmental issues in civil, commercial, and criminal legal actions, and on compliance with environmental legal frameworks. There are three main areas in which ecologists are employed as experts.

Compliance Analysis

The increasing scale of development projects places a considerable demand for full compliance of projects and developments with the multiple legal frameworks and the ethical codes of practice of governments, international development agencies (IDAs) and professional consultancy firms. Paradoxically, despite the proliferation and increasing complexity of environmental regulatory frameworks, project preparation, and compliance analysis still rarely employ legal specialists. Consequently, the task of auditing compliance with all of these disparate codes has moved out of the sphere formerly controlled by the planners and engineers into the domain of the forensic ecologist.

Claim Management

Ecologists are employed by insurers assessing claims for environmental pollution damage. Marine salvage firms may obtain very significant bonuses if they can show that their work resulted in the removal of a significant threat of environmental damage, so the role of ecologists is central to both compiling and settling salvors' claims. The complexity of ecosystems and the pathways whereby pollutants travel through them and affect different plants and animals – sometimes thousands of miles away – requires special skills and experience to identify the full potential range of impacts of environmental contamination, and of the inevitable claims that even relatively minor local incidents may generate.

Acting for the Defense in Environmental Prosecutions

A third, less obvious area in which ecology is in increasing demand is assisting defendants faced with actions by regulators. Regulators are usually funded by public finance, and have comparatively greater access to scientific services than the defendant. Consequently, courts tend to assume that the prosecution has superior expertise. Defendants may have to cope with evidence that they find incomprehensible, and need independent expertise that enables them to understand the reliability and interpretation of the evidence against them.

The Role of Forensic Ecology

In the past, ecologists have taken only minor roles in legal actions and the resolution of criminal charges. But as environmental issues take center stage in the debate on climate change and sustainable management of endangered natural resources, legislation is proliferating at both national and international levels. A bewildering array of statutes, directives, conventions, agreements, and treaties is aimed at ensuring continued access to essential resources and environmental stability for the future; all need to be consistent in their approach and application (*see* **Alcohol: Behavioral and Medical Effects**).

The methods of strategic environmental assessment (SEA) [1] underpin the assessment of project compliance, and ecologists with an extensive knowledge of the dynamics of ecosystems are able to act as moderators in the process of harmonizing the environmental legal framework under which national and international policy-making and development operate.

This escalation of environmental legislation has not been matched by an adequate increase in forensic training capacity to improve the skills of the regulators' field staff. Prosecutors all too often rely upon inadequately prepared material and inadmissible or flawed evidence. The ecologist is required to identify anomalies in regulatory legislation, uncertainties over scientific issues, and a lack of enforcement capacity. Even if his or her skills are used to defend against such system weaknesses, the exposure of these defects help to ensure that, in the long run, the system is strengthened and justice served.

Dealing with issues in local pollution actions in the face of ill-informed or defective prosecution

evidence is a relatively simple matter, but the greatest risks of noncompliance are frequently caused by the most powerful corporate and institutional entities. When these offend or are about to do so, they can appear to enjoy immunity to all but the most sophisticated legal challenges. The ecologist may therefore be placed in the unusual position of not only providing evidence of an offense but of also deciding exactly how to use this information to bring about adequate restitution. This is particularly the case when the regulators themselves misbehave, and it is this aspect of forensic ecology that poses the greatest challenges, both professional and physical, to the ecologist working in the field of compliance auditing. In this article, it is described how unilateral direct action by the ecologist can be used to resolve particularly difficult issues of enforcement.

Principles

Beyond Reasonable Doubt, or on the Balance of Probability?

In many ecological assignments, the questions in issue relate to an incident that occurred months, or even years, before. In such cases, the reliability and provenance of environmental evidence may be weak; pollution cases can take years to come to trial, while the medical effects of an exposure to an environmental contaminant may emerge only decades later. Much of the primary evidence is long gone, and post-incident contamination or interference soon renders the interpretation of what residual data can be salvaged difficult, if not impossible (*see* **Interpretation: Legal Perspective**).

Although the regulators may have secured formal samples close to the time of the incident, the vital contextual information that turns evidence into understanding in court could well be irretrievably lost. Supporting evidence for a specific interpretation of the issues may even have been deliberately cherry-picked, and any that is inconvenient to the prosecution's case deliberately eliminated – this is by no means a rare occurrence. The ecologist has therefore to rely on data salvaged from obscure sources, often by untrained members of the public. The provenance and admissibility of such evidence is inevitably uncertain, and only useful as part of a collection of contextual evidence – material that is collectively consistent and unlikely to exist purely by chance.

It is often only possible to secure secondary circumstantial evidence about an old event, but if a wide enough range can be obtained, then the case can be presented on the basis of a coherent contextual analysis that could establish causation, but only on the basis of balance of probability. In practice, much of the workload of the ecologist is settled by out-of-court agreements. Large corporate organizations are often reluctant to see their deficiencies exposed publicly in the courts and the media, and prefer to settle privately rather than risk a formal guilty verdict being reached, thus setting a precedent that would be embarrassing, were it to be disclosed publicly.

In issues of severe human injury, however, a court is much more likely to require evidence that establishes the charges or claims beyond reasonable doubt. In practice, unless there is clear evidence of toxic damage from a very characteristic and unusual chemical, causation can be difficult to establish to this standard. This is especially so if the same or similar pathology occurs naturally in equivalent-sized populations, even if only sporadically. In such cases, the statistical approach of the epidemiologist must be relied on to estimate the probability of an excessive incidence of the condition, but this can offer only presumption evidence of a possible causative relationship. Unless it is supported by clear toxicological and pathological evidence, in such cases class actions are more likely to succeed than one dealing with damage to a single individual.

A Matter of Scale – Sampling the Ecosystem Rather than the Incident Location

In cases of unusual environmental contamination by toxic substances, the ecologist seeks to obtain as wide a range of corroborative evidence as possible. If the nature of the contamination is known or suspected, then it may be possible to predict what unusual medical or veterinary conditions are likely to be found close to the release location. But the situation is often reversed; symptoms of ecological, veterinary or medical damage may appear with no indication of their probable origin or causation. Such incidents should always be treated as ecological in scope, and not confined to purely medical investigations (*see* **Sampling Trace Evidence**).

Initially, evidence is collected from humans, animals, and the environment at or near to the location of the incident, in an attempt to pinpoint the source.

But sometimes it may be possible to collect more revealing samples farther from the site of an incident. Traces of release of a toxic substance from an unknown location may be transferred along food chains in an environment, becoming more concentrated as they move from one species to the next. It is often possible to detect high concentrations of otherwise unsuspected or undetectable contaminants in species at the summit of a food chain, or at the distal end of a linear environment – for example, in the estuary of a river.

Discovering such traces allows the ecologist to work back up the transfer pathway of the substance through the ecosystem, until evidence of the presence of the substance disappears. At that point, the entry to the environment is presumed to have occurred. Many estuarine and marine shellfish passively accumulate extraordinarily high concentrations of persistent organic contaminants in their flesh, especially oil fractions, polychlorinated biphenyls (PCBs), and pesticides. Concentrations of these fat-soluble compounds may be many times more than those that are present in the surrounding water or sediments [2].

The "Keumdong No 5" Oil Spill – Identifying the Source of Lethal Oil Contamination in Shellfish

After a small but remarkably lethal oil spill in Kwangyang Bay in South Korea, I collected samples of the dispersants that were used to clean oil from beaches, oil from tidal zone sediments, and smears of oil from the hulls of fishing boats soiled in the incident. Live shellfish were also obtained from the contaminated areas and the oil residues and shellfish tissues were analyzed at an iccredited forensic analytical laboratory. Using fast atom bombardment mass spectrometry (FABMS) and gas chromatography/mass spectrometry (GC/MS), detailed chemical descriptions of the oil-related contaminants present were obtained (*see* **Toxicology: Initial Testing**; **Confirmation Testing: Toxicology**).

The oil "fingerprints" from the shellfish corresponded with those from the oil residues, and revealed that the type of oil spilled was not a heavy fuel oil, as was claimed by the oil industry pollution claim managers. It was a much lighter fuel oil, with a higher concentration of those components that are able to cause oil narcosis and delayed death in shellfish. This information allowed the legal team acting for the fishermen to estimate the significance of the incident to the recorded shellfish mortalities, and present verifiable compensation claims to a Tribunal in London, and to the Seoul District Court 4 years later.

Ecological Processes – the Key to Forensic Investigation

The most important factor in understanding how an ecosystem works is an appreciation of the processes that operate in its normal, undisturbed or uncontaminated state. Only then it is possible to identify existing problems or predict future ones, and devise cost-appropriate mitigation strategies. Although modern development programs emphasize the importance of avoiding adverse impacts, some very large existing projects have evolved piecemeal over many years. The compliance of some (or even all) of the original proposal components may be defective according to current standards and procedures. When this happens the ecologist needs to be aware of both the underlying engineering constraints and of exactly how the ecosystem that is affected actually works, because this is the key to identifying least-cost but effective remedies.

Using Circumstantial Evidence to Direct Targeted Evidence Collection

Carefully targeted follow-up investigations can only be planned once the most likely locations for an effective data salvage operation have been suggested by the initially collected circumstantial evidence. If human victims are involved, the aim is to establish whether or not there is an unexpectedly large group of people with specific injuries that could have been caused by their exposure. Patterns of related effects may emerge, and these can be used to predict specific conditions that a monitoring program should concentrate upon. The aim is to use as wide a range of evidence as possible to ensure that nothing is missed that might persuade a court that a causative link exists between the incident and identifiable damage.

In some environmental contamination incidents the examination of exposed individuals is often seriously delayed; in others adverse medical conditions may develop only long after their initial exposure. In both cases, a high level of discrimination and vigilance is necessary to distinguish genuinely damaged individuals from those who may develop physical or

medical conditions that are similar to those that might naturally occur in the population as it ages.

The Camelford Poisoning – Watching for Delayed Alzheimer-Like Neurodegenerative Conditions. A classic illustration of the need for increased vigilance in pos-tincident monitoring is the severe contamination of the public water supply by aluminum sulfate in a quiet UK town in 1988. As a result of a mistaken delivery to the local water treatment works at Camelford, concentrations of aluminum in the town's water supplies reached 620 mg l^{-1}; this was 3100 times the permissible maximum in drinking water under both European community and British legislation. On entering the treated water supply, the chemical released sulfuric acid that turned the normally neutral water acidic. This corroded domestic copper plumbing systems, and the water in some hot water tanks contained several grams of copper per liter. Many of the 20,000 people exposed became extremely ill, and a large number of pets and livestock were either severely damaged or died (*see* **Reconstruction: Accident** and (**Poisons: Detection of Naturally Occurring Poisons**).

After the incident, the public sector health authorities refused to carry out any detailed clinical investigations on human victims of the poisoning. Although they knew that some individuals in the population would be "superabsorbers", they perversely claimed that aluminum cannot be absorbed in any significant quantity from the gut into the bloodstream, so no adverse medical effects could possibly occur. The possibility of a link between aluminum and Alzheimer's disease (AD) [3] led to public speculation about a possible epidemic of AD after the event, and this was vigorously denied by the health authorities.

The acute effects in the human population included severe arthritic-type pains in joints as well as eruptions of skin conditions. But these can, and do, occur normally in any population, and no definitive decision about their significance could be made. However, information on the sudden deaths of many pets and livestock in the area indicated that there could be potentially serious long-term medical risks to the human population.

Analyses of tissue samples taken from sows at a local pig farm revealed significant increases in aluminum loadings in some organs, and implied that severe aluminum-related medical conditions might develop in some of the human victims after a relatively long period. The effects of an exposure on animals may therefore provide an "early warning system" that can warn of possible future adverse effects in the human population when direct invasive medical examinations of human victims are either unethical or are obstructed by authorities [4].

Limitations of Ecological Evidence

Evidence presented in most pollution prosecution trials relies heavily on the chemical analysis of samples taken from the site of an incident. Biological sampling of populations tends to be less revealing despite the existence of techniques that can suggest the presence of polluting materials. This is because the abundance of species and individuals, and the complexity of community structures, naturally vary unpredictably in both space and time. Experts in the chemical and physical sciences are able to rely upon legally enforceable standard parameters which, if exceeded, are taken as firm evidence of contamination. No such standard populations exist for the ecologist to use as a legal baseline, so it is more difficult impossible to quantify biological changes caused by an incident with a high degree of certainty.

Practice

The Individual Expert versus the Institutional Laboratory

In the cases of environmental contamination brought quickly before a court, individual experts in chemical and physical analysis are usually members of teams working in public sector (and occasionally private) laboratories. In contrast, the ecologist working for the defense on such contamination incidents normally works alone, with much of his or her efforts directed to highlighting defects in the prosecution evidence. When it is necessary to obtain direct private sector evidence, the ecologist must arrange for samples to be taken by a specialist from an accredited analytical laboratory, to ensure that the resultant data are robust and reliable enough to be admissible as evidence. But this does not preclude the collection of sample evidence by less conventional methods.

In the Camelford incident, a water sample collected by a family was rejected by the water authority

responsible for the contamination of the water supply, but was eventually analyzed six months later by a BBC team making a documentary on the incident. The aluminum content of this sample, revealed by analysis at a public health laboratory, was far in excess of that in any sample taken officially by the authority, and forced the reluctant British government to set up the formal inquiry that the public had been demanding for 6 months. Part of the job of the ecologist is to seek out such hidden evidence, regardless of its type and location, and ensure that it is professionally assessed in case it proves critical to an otherwise hopeless case.

When carrying out compliance audits, ecologists rely on, or even manage, teams of experts in a wide range of social, environmental, and engineering disciplines. They collect the contextual data upon which an ecological analysis depends. There is no formal training in forensic ecology (which should not be confused with environmental forensics, for which there are established training courses available) – the rare specialist in the field tends to be an experienced loner who has drifted into the discipline after a varied and busy career as a team leader or environmental impact assessment (EIA) manager.

Ecological Techniques

Engineering and Hydraulic Modeling – Part of the Ecologist's Toolkit. Ecological systems are extremely complex, and identifying damage (or the threat of damage) to them demands considerable expertise and experience. For the forensic practitioner, the most important skill is understanding the core processes that govern the stability of the systems under examination. An innovative approach to field data collection is needed, to apply this experience to the collection and interpretation of evidence that may be elusive, obscure, or even initially quite incomprehensible.

A working understanding of civil engineering technology is essential, since identifying practical methods of damage avoidance or limitation demands an appreciation of what can, and what cannot, be done. In river systems, hydraulic modeling is emerging as a critical technique for assessing options, and familiarity with the strengths and weaknesses of the techniques employed by modelers often enables the ecologist to predict future risks to the environment that no other expert has been able to foresee (*see* **Computer Animation and Simulation Evidence**; **Reconstruction: Three Dimensional**).

Elegant Experimentation or Inspired Guesswork? Many standard ecological field research practices rely on repeated long-term studies involving complex statistical analysis. This is because ecological processes tend to be either cyclical or progressive, varying on a seasonal basis or over much longer periods as habitats pass through a succession of stages toward a stable climax community. In the case of compliance audits, this is an important consideration when assessing the potential risks of unacceptable future outcomes. But when the ecologist takes on a new assignment involving a severe environmental incident, many of the standard techniques have to be abandoned in favor of an idiosyncratic or innovative approach that may be required.

Working with Complexity Frequently, the most important questions to ask at the start of a mission are "How does this system work? What am I not seeing? Who has something to hide?" The ecologist needs to be able to predict impacts in not only the three spatial dimensions, but to extend predictions in time. Ecosystem processes operate on many different timescales, from the 30-min cycle of reproduction in many species of bacteria, to the millions of years of erosion and the movements of continental plates. So the results of an incident or intervention may become evident almost instantly, or may be delayed for centuries or even longer. The ecologist looks for disturbances in the rhythms of ecosystems that might imply a link with a past incident or intervention.

Ecological versus Political Boundaries – Complexity at its Most Formidable

A large part of the work load of the ecologist tends to revolve around project compliance auditing. Historically, this developed as one of the tools used in EIA. Since the mid-1970s, ecologists have been responsible for identifying environmental and social threats in every conceivable environment and social regime. This involves the correlation of disparate data across an often dauntingly wide range of disciplines. Because it is essential to devise solutions that reduce or avoid specific predictable social and environmental problems, the ecologist has to be aware of what is, and especially what is not, permissible. From this

has emerged the mantra that project development is limited not by what is technically feasible, but by what is legally permissible (*see* **Expert Opinion in Court: a Comparison of Approaches**).

But on occasion, a development project spans several national boundaries. Political policies, religious codes, and administrative regulations often differ widely, even across single borders, and the resolution of conflicting transboundary attitudes and approaches to environmental problems can present the ecologist with enormous challenges, demanding quite remarkable powers of concentration and flexibility in resolving issues in such locations. All too often, what may be accepted as scientific fact or political correctness in one country may be denied in an adjacent one. Increasing competing demands for water in rivers that cross national boundaries are frequently cited as the most probable cause of future "water wars". The migrating fish that travel across these boundaries regardless of political restrictions are the unwitting hostages to fortune in all of the world's great rivers, and many are now in imminent danger of extinction.

Eradicating the Fruit Fly in South America

An excellent example of transboundary political confusion was the Fruit Fly Eradication Project in South America. The larvae of the Malaysian Fruit Fly living in Carambola (Starfruit – *Averrhoa carambola*) were accidentally introduced into Guyana, Suriname, French Guiana, and a small area in northeastern Brazil. After a period when the Fly appeared to be relatively contained, and only occurred locally in orchards of exotic fruits, it suddenly developed the ability to use native fruits as larval food, threatening to colonize the entire Brazilian and adjacent rain forests. The resulting risk of a complete ban on exporting fruit from the entire continent because of its presence in commercial fruit would have been an economic catastrophe.

The most effective eradication technique available involved the destruction of the male flies, by luring them to chemical baits that attract only the male flies. The active ingredients were the pseudo pheromone methyl eugenol (a natural component of clove oil), and the pesticide malathion. Unfortunately, the project was bogged down in arguments over incompatible legislation in each of the four infested countries, as well as the laws of France (French Guiana is an external department of France), the European Union, and California. In addition any solution had to be compatible with the internal rules of the donor organization, based in Italy.

The resulting conflicting framework of six disparate legal codes, printed in four languages, threatened to prevent the project from being completed just as success was in sight. The tangled legal issues were resolved with an analysis of the legislation and the available options for compromise, and provided new toxicological risk assessments on both substances, allowing the eradication program to be resumed [5].

Statutory Environmental Audit Procedures

Many governments, and all leading IDAs, have standard procedures that dictate the methodology to be used in EIA and in the auditing of project compliance and performance. In recent years these have been augmented by techniques that extend this work considerably. Both nongovernmental organizations (NGOs) and individual ecologists are now involved in assessing issues relevant to the judicial process and in advising legal practitioners in personal injury, quantum of loss of assets and resources, and the provision of indemnity against environmental and other losses and claims.

Technical developments in the practical methodology of ecological analysis are rapidly adopted into the standard armory of skills used by ecologists. They enhance the efficacy of ecological services provided to project proponents, donors, lawyers, and insurers. They also help to minimize the need for compensation when projects fail to deal with regulatory constraints, or cause damage that could and should have been avoided.

Hydropower Conflicts in the Mekong Basin

In planning the expansion of hydropower in the Mekong Basin, it was realized that new hydropower schemes scattered almost randomly throughout the entire Mekong Basin would cause significant changes in the timing of floods on which migrating fish depend. The development program would place fish stocks in the main and tributary rivers under enormous environmental pressure, threatening the viability of the entire indigenous river and delta fisheries, worth an estimated US$2 billion a year (*see* **Learned Treatises as Evidence**).

By ranking the 39 proposed schemes according to their socioenvironmental impacts and their technical and economic reliability, and showing how hydraulic changes in the rivers affect the biological mechanism governing fish migration, it was then possible to predict the disruption that would be caused by the cumulative impacts of any specific combination of projects. This provided a basis for estimating the consequential costs for expensive mitigation options and reducing compensation claims.

"Kitchen Sink" Chemistry and Toxicology

In the real world of dirty environments and sometimes corrupt operators, it is frequently been necessary to collect whatever marginally relevant historical data or contaminated materials remained on site, and submit them to analysis in the hope that new insights could be obtained and new approaches to interpretation developed. When there has been a release of a toxic substance, such as when fish are killed by water pollution incidents or public drinking water supplies have been contaminated, "quick and dirty" experiments may need to be designed that help to clarify and answer the questions that were thrown up by the initial situation analysis (*see* **Toxicology: Initial Testing**).

Many contaminants have never been formally assessed for their toxicity, or may undergo unexpected reactions in the environment that may completely alter the interpretation of the event. Even if the ecologist does not have access to a sophisticated chemical analysis laboratory, or to a toxicological research facility, simple "kitchen sink" experiments may reveal crucial new facts from very simple procedures.

Blue Baths and Foaming Torrents – Resolving Difficult Problems with Simple Science

In the Camelford poisoning incident, investigators were baffled by unexplained reports that bathwater turned blue when it came into contact with soap. Extremely high concentrations of copper had unexpectedly appeared in the contaminated water, but none of the samples of water collected after the incident contained enough to produce this effect. In fact, although this was not appreciated at the time, the extreme acidity of the water supply caused domestic copper plumbing pipes and storage tanks to dissolve. The crucial question was simply, just how much copper does it take to produce a "blue bath"?

Small pieces of cotton cloth were soaked with different concentrations of copper sulfate solutions and soap solution, and compared the intensity of the colors that developed with that of blue-stained clothing retained for 12 years by a few local people. The concentration of copper needed – at least 2000 mg l^{-1} – was 100 times greater than the highest concentration detected by conventional analysis of supposedly contemporary water samples taken by public sector agencies. The officials had been too slow to respond to the incident, and had failed to detect the worst of the induced secondary contamination by the copper.

Both aluminum and copper are now regarded as possible environmental agents in the development of conditions similar to Alzheimer's disease and some forms of Parkinsonism. This simple "kitchen sink chemistry" had alerted our community to be vigilant for unusual neurological conditions amongst those who had complained of particularly severe effects of their exposure; 15 years after the event we began to see just such delayed medical developments (*see* **Postmortem Toxicology: Laboratory Analysis**).

Community Epidemiology – Evidence Collection by Members of the Community

When public sector authorities seem to be obstructing the understanding of an environmental incident, the new field of community epidemiology provides a powerful tool with which local communities can counter public sector subversion of evidence collection. Some of the most inspiring examples of forensic ecology have involved independent experts working with communities to collect highly sensitive data and releasing their data and conclusions to the media. On many occasions the resulting publicity has forced reluctant authorities to accept the evidence now in full public view, and carry out their statutory regulatory duties and enforcement responsibilities.

In this field, "communities" can be functional groups with a common interest, or conventional geographic communities. They generally develop spontaneously out of frustration that the regulatory or health authorities persistently refuse to admit that a problem exists. The stimulus is usually public frustration that the "Establishment" refuses to accept that there has been a significant event that has resulted in serious damage; the accusation that there has been an official "cover-up" is often correct, even if hard to prove.

A global association of airline crews established the "Aerotoxic Association" in London on June 18, 2007. Its aim was to force official recognition of the potentially lethal contamination of cabin air by tricresyl phosphate, originating from burnt oil leaking from jet engines. Members organized a covert program of collecting swab samples from aircraft cabins for a year, and this provided clear evidence of universal contamination of commercial aircraft by this dangerous organophosphate (*see* **Confirmation Testing: Toxicology**). The evidence was passed to the media, stimulating a series of television programs and newspaper articles exposing the risks to aircrew and public alike [6].

Problems

Corruption

A very common problem for the ecologist is that evidence of illegal activities may be deliberately suppressed, often by corrupt authorities and the regulators themselves. Some of the examples given earlier show how political or commercial vested interests can lead to the deliberate blocking of comprehensive studies. Researchers who receive funding from industry are also very vulnerable to the suppression of potentially damaging evidence, and if forced to face an inquiry they may be accompanied by "minders", to make sure that no compromising material is disclosed.

The Future

Forensic ecology is at an early stage in its development, but it promises to become increasingly important in conflict reduction and environmental dispute resolution. Already their skills in compliance auditing make ecologists crucial specialists in project management and incident investigation. They are in increasing demand as team leaders, supplanting the role traditionally occupied by civil engineers, but they are also in demand for investigating serious incidents and the assessment of the quantum of damage in some of the most spectacular contamination incidents such as the Exxon Valdez oil spill in Alaska and the catastrophic reactor meltdowns at Chernobyl and Three Mile Island.

Universities are increasingly aware of the need for trained specialists in fields such as wildlife and marine forensics, but entry to the ecological field is still reliant on the migration of experienced field specialists into the less disciplined and often openly aggressive arena of adversarial law enforcement, international conflicts in access to depleting natural resources, and the increasing involvement of organized crime in toxic waste disposal. Few traditional ecologists realize that their skills are appropriate in the forensic environment. But as environmental conflicts intensify and regulatory frameworks proliferate, the shortage of skilled experts in this field will force universities to develop courses to provide the expertise demanded by an increasingly environmentally aware international, indeed, global legislature.

References

[1] Verheem, & Tonk, (2000). Strategic environmental assessment: one concept, multiple forms, *Impact Assessment and Project Appraisal* **18**(3), 177–182.

[2] Widdows, J. & Donkin, P. (1992). Mussels and environmental contamination: bioaccumulation and physiological aspects, in E. Gosling, ed, *The Mussel Mytilus: Ecology, Physiology, Genetics and Culture*, *Developments in Aquaculture and Fisheries Science*, Elsevier, Vol. 25, Chapter 8.

[3] Exley, C. & Esiri, M.M. (2006). Severe cerebral congophilic angiopathy coincident with increased brain aluminium in a resident of Camelford, Cornwall, UK, *Journal of Neurology, Neurosurgery, and Psychiatry* **77**, 877–879.

[4] Cross, D. (1990). The politics of poisoning -the Camelford aluminium sulphate scandal, *The Ecologist* **20**(6), 228–233.

[5] Mathur, S. (2003). *Using ICTs to Eradicate the Carambola Fruit Fly*, ICT Update, Issue 11: Pest management. May.

[6] Booker, C. (2007). *Pilots Disabled by Poisoned Air*, Christopher Booker's Notebook, Sunday Telegraph 24th June.

Further Reading

Williams, G. & Popay, J. (1994). *Lay Knowledge and the Privilege of Experience*, J. Gabe, D. Kelleher & G. Williams, eds, Challenging Medicine, Routledge London.

Murphy, B. & Morrison, R.D. (2007). *Introduction to Environmental Forensics: A Forensic Approach*, Academic Press, ISBN 0123695228, 9780123695222.

Wang, Z. & Stout, S. (2007). *Oil Spill Environmental Forensics: Fingerprinting and Source Identification*, Academic Press, ISBN 0123695236, 9780123695239.

DOUGLAS W. CROSS

Enzymes

Enzymes are biomolecules, most commonly proteins, that catalyze biochemical reactions in the body. Catalysts are substances that facilitate specific chemical reactions, but which are not themselves consumed or changed in those reactions. Usually these reactions involve the conversion of a molecule from its substrate form to its product form. Enzymes are extremely selective for their substrates and only catalyze specific reactions. The set of enzymes available to a cell determines the metabolic pathways that can occur within that cell.

Enzyme activity can be affected by other molecules that act as either activators or inhibitors. Some enzymes also require nonprotein molecules called *cofactors* to be bound in order to promote enzyme activity. Cofactors can be either inorganic (e.g., metal ions) or organic compounds (e.g., heme). Temperature is also an important regulator of enzyme activity with most reactions having an optimal temperature range. Excess temperature can also denature enzyme structure, thereby reducing or preventing its activity.

There are many enzymes that have a critical influence on the forensic biology process. Some naturally occur in the template material available for testing, whereas others are introduced by forensic scientists to drive specific reactions in the laboratory (a selection of relevant enzymes are mentioned in Table 1).

Table 1 Enzyme classes and examples of their use in forensic biology

Enzyme	Properties	Importance to forensic biology process
Nucleases	These enzymes are capable of cleaving the phosphodiester bonds between the nucleotide subunits of nucleic acids, such as DNA	Endogenous nucleases can cleave the DNA in the cell rendering it unable to be analyzed. Some DNA isolation methods (such as the Chelex 100™ resin) function by attempting to disable the nuclease activity through heat and chelation of the inorganic cofactor required for nuclease activity (Mg^{2+})
DNA polymerases	A DNA polymerase enzyme catalyzes the addition of doexynucleotides during DNA replication. This thermostable DNA polymerase carries the name of its bacterial host, *Thermus aquaticus*	Taq DNA polymerase is able to withstand high temperatures and has an optimal temperature for activity of ~72 °C. This property makes it suitable for use in the PCR to catalyze the addition of nucleotides during the final extension step
Proteinases	A proteinase enzyme cleaves peptide bonds in proteins, destroying their tertiary structure. Proteinase K is a serine pronteinase originally isolated from fungus.	Proteinase K is commonly used in DNA extraction techniques to digest cellular protein, particularly those in cell membranes, thereby promoting cell lysis
Acid phosphatase	A phosphatase enzyme cleaves phosphate groups from other molecules and is active around pH 4.9–5.5	Pancreatic acid phosphatase is present in high concentrations in seminal fluid. The presence of acid phosphatase is used as a presumptive indicator of semen. The test is simple, rapid, and sensitive and is a cornerstone test in forensic bioscreening
α-Amylase	An amylase enzyme digests long-chain hydrocarbons. The α-amylase enzyme hydrolyzes α-(1,4) glycosidic bonds of glucose polymers such as glycogen or starch	α-Amylase is present in high concentrations in saliva and is used as a presumptive test for this body fluid. The common test is applied through reaction with dyed starch substrates (e.g. Phadebas test). In the presence of amylase the starch is digested and the dye released into solution

Related Articles

Acid Phosphatase
Extraction

SIMON J. WALSH

Epilepsy *see* Automatism as a Defense to Crime; Seizures: Behavioral

Erasures in Documents: Detection of *see* Alterations: Erasures and Obliterations of Documents

Error Rates in Forensic Methods

Introducing Error in the Forensic Process

Errors committed during forensic examinations can be the result of a myriad of situations and involve individuals other than laboratory examiners.

The procedures at crime scenes are of special concern regarding acts of omission or commission.

Acts of omission include failing to

- protect the scene from contamination (*see* **Crime Scene Investigation**; **Crime Scene Management**);
- record the scene (*see* **Rape Trauma Syndrome**; **Crime Scene Photography: US Perspective**);
- collect the evidence (*see* **Evidence Collection and Preservation: Casting**); and
- properly package the evidence (*see* **Packaging and Transport**).

Acts of commission include:

- allowing access to untrained personnel;
- introducing potentially contaminating material e.g., *via* clothes or equipment; and
- moving DNA around the scene on people or equipment (e.g., fingerprint brushes).

Analytical Error

In analysis, there are numerous error types. In an ideal testing system, there would be no errors. However, most testing regimes have some level of error associated with them. These error rates, and if possible how to identify and mitigate them, should be measured as part of the validation process of the test method.

Errors can be binary, off/on, yes/no, true/false classification errors, or errors whose magnitude may vary with the amount of material being tested or be a systemic bias in the system.

Binary errors are of two types.

Type 1 Error

This is more usefully, probably, and usually called a "*false positive*" error. In a Type 1 error, a material that should be classed as "negative" produces a "positive" result. For example, crime scene examiners may use a presumptive test for blood (see Km test), which also produces a positive result with some fruit juices. In such instances, of course it is not possible to know the error "rate", but simply the types of material that may produce an erroneous result can be identified.

Type 2 Error

This is the corollary of the false positive error: a false negative. A false negative result is produced when the material under test contains the analyte being sought, but fails to identify it.

Using Controls

It is a normal scientific practice to ensure that methods are working properly by running simultaneous tests on materials that are known to contain the analyte (known as a positive control sample) and that

do not (known as a "negative control" or, in some chemical tests, a "reagent blank").

Type 1 and Type 2 errors inform another two important parameters of testing; the specificity and sensitivity of the test. These errors are primarily qualitative; i.e., whether the material has some quality (such as being blood) or not. There are more sophisticated errors that occur in quantitative analysis (*see* **Case Assessment and Interpretation**).

Other Types of Error

The measurement of error is usually performed for the purpose of determining the reliability of a result. This area can be particularly confusing to the nonscientist, as many terms, such as "reliability", have a very specific meaning when used in a scientific context. In scientific parlance, there are various terms that have been developed to define the variation that may be seen when quantitatively testing things. In considering the reliability of a test result, it is important to know and understand these parameters, which should have been derived during the validation of the procedure. Each of these parameters can be a source of error.

Many laboratories have quality assurance systems that require recording of all errors, howsoever they have been caused. Some laboratories partake in schemes that involve the testing of materials of known provenance as a test of the laboratories' system. These proficiency tests may be "blind" or "declared". In blind trials, the laboratory staffs are ostensibly unaware that the sample is a test, although this can frequently be fairly easy for the staff to spot. In a declared trial, the staffs are aware that they are being tested. The former are considered better measures of laboratory performance although it can be difficult or impossible to ensure truly blind trials with some evidence types.

These proficiency tests have occasionally been cited as evidence of the reliability of the laboratory or system. For such purposes, it is important to consider the type of recorded error and when it occurred. One of the main reasons that a laboratory will undertake such exercises is to ensure that they provide reliable results. Nonconformances, as they are called, are identified as part of the normal practices of most laboratories and are specifically documented and dealt with in almost every form of laboratory accreditation. These records should be maintained and available. Clearly, if an error has been identified and dealt with, then it is inappropriate to use the precorrective action error rate as a measure of the post-correction system. To fully understand laboratory error rates, a consideration of not only the overall rate but also the rate for individual methods, equipment, and staff is required.

Measurement of Error

An additional complication in the use of error rates is the fact that a rate is one unit divided by another (e.g., miles per gallon and dollars per test). An error rate could be expressed as, *inter alia*, per test, per person, per sample, and per laboratory (*see* **Sampling and Estimation of Quantities**). The choice of the appropriate rate depends on the purpose for which the information is being used. The fact that one analyst is a poor performer does not necessarily reflect the reliability of the test, method, team, laboratory, or organization.

Another consideration is whether any error in the multitude of processes that occur in a modern laboratory is manifest in the final report. If an analytical process has a checking process, for example, the inclusion of positive and negative controls, or the manual checking of results, this minimizes or eliminates the risk that these errors will not be detected in the formation of an opinion or evaluation. In that event, the error rate for the process may be irrelevant with respect to that opinion.

A sophisticated analysis is normally required to understand the factors affecting a specific case.

Error Rate Calculation

Thus, the words "error rates" also have multiple meanings. The term can denote the statistical likelihood that a value calculated on the basis of multiple repeated occurrences of data collected in a databank corresponds to its "true" value in a particular instance.

In the law, the words "error rate" have acquired a special meaning in those jurisdictions that follow the *Daubert* (*see* ***Daubert v. Merrell Dow Pharmaceuticals***) rules on admissibility of expert testimony.[a] In the *Daubert* decision, the US Supreme Court delineated a number of different factors by which trial judges are asked to decide whether proffered

expert testimony will qualify as "scientific knowledge" shown to be sufficiently reliable to be admitted at trial.

One of these factors is whether the results of the underlying theory, technique, or methodology have been subjected to an error rate determination. The Court stated: " ... Additionally, in the case of a particular technique, the court ordinarily should consider the known or potential rate of error." [1]

Conclusion

In the hundreds of decisions following *Daubert*, the courts have been at pains to elucidate exactly what the Supreme Court's opinion requires in this and in other regards. Judge Alex Kozinski of the Ninth Circuit Court of Appeal whose decision in the same case while sitting on the intermediate appellate court was reversed by the later statement of Supreme Court, when discussing the *Daubert* decision at a legal conference, stated "In addition, you have to consider the error rate ... What is the error rate? I have no idea what this means but there it is." [2]

End Notes

[a.] See **Computer Animation and Simulation Evidence** – Daubert v. Merrell Dow Pharmaceuticals. Also, **Judicial Notice of Scientific Principles and Facts** – Kumho Tire v. Carmichael.

References

[1] Daubert v. (1993). Merrell Dow Pharmaceuticals, 509 U.S. 579, 594, 113 S.Ct. 2786, 2797.
[2] Kozinski, A. (1997). *Brave New World*. 30 U. Call, Davis L. Rev. 997, 999.

ANDRE MOENSSENS

Ethanol *see* Alcohol: Use, Abuse, Tolerance, and Dependency, Alcohol: Analysis

Ethics: Codes of Conduct for Expert Witnesses

Introduction

It is universally understood that, in the practice of their professions, forensic scientists are required to refrain from unethical practices, but what is "unethical" is often difficult to determine. By what standard is the ethical nature of one's conduct to be measured? The very first source of information of what is considered unacceptable professional behavior is the code of ethics of the professional association to which one belongs. Most forensic membership groups maintain a code of ethics [1]. Since many scientists belong to more than one society, the person's conduct may be governed by several different codes of ethics that may not necessarily look similar. All of them, however, make engaging in ethical conduct a requirement of belonging to the association and provide sanctions for violating the professional standards of conduct.

In addition to defining unprofessional conduct, codes of ethics also typically contain procedural mechanisms whereby purported unethical conduct is to be investigated and, if appropriate, sanctioned. Such sanctions can run the gamut from a mere reprimand to expulsion from membership in the organization. This may, at times, be accompanied with loss of license, accreditation, or board certification. Since the law in common law countries tends to view membership in a professional association as a valuable property right, the mechanisms whereby a person is stripped of that right must comport with procedural safeguards against abusive and otherwise improper exercise of the power to sanction. Because of the great variety of mechanisms, this article does not concern itself with the "due process" aspects of procedures that seek to enforce the ethical injunctions imposed upon an association's membership. This article is confined to a general overview of what conduct may be regarded unethical.

The conflicts between the codes of various groups are at times *de minimis*, but one must, nevertheless, remain sensitive to the fact that wide differences in approaches to ethics can exist. Some scientists who must possess a license to practice their profession granted by governmental authorities in some countries – i.e., medical doctors, dentists, psychiatrists,

psychologists, etc. – may see their licenses jeopardized if found to have been sanctioned for unethical professional behavior.

Professionals may also belong to specialty boards with different ethical codes, such as that of the generalist group,[a] as well as the codes of ethics promulgated by specialty boards. The same is true of the members of virtually every specialty group within the forensic sciences.

Finally, it should be remembered that, in addition to the sanctions that can be imposed by a professional society upon its own members, there is a "common law" of ethical conduct as well. Violating these common law prohibitions against unethical conduct, usually involving professional conduct that includes deception, fraud, overreaching, perjury, or behavior merely considered to be dishonest or inequitable, whereby either the public or society, in general, is injured – as by engaging in "negligent" professional conduct – may expose a forensic scientist to civil legal sequelae in damage actions as well as potential criminal liability for perjury or fraud. These are sanctions that professional societies cannot impose. Even the expert witness who does not belong to an organization that has a mandatory code of conduct may be subject to such legal actions. Potential liability of this sort falls within the area of "professional malpractice" for either intentional or unintentional (negligent) conduct – tortious conduct – as defined by applicable law.[b]

Defining Ethics

The existing ethics provisions in forensic science vary greatly in how the terms "ethics" or "unethical conduct" are defined. Some list specific prohibitions, others frame ethical principles in terms of aspirations. How these rules can be applied in specific fact settings is not always clear. Outside the legal profession, most professional societies do not publish compilations of ethics complaints that were decided by its boards. Additionally, while general ethical principles may be listed, the manner in which ethics codes have been applied in illustrative cases may remain obscure to the membership.

In general, the professional codes of conduct devote much more space to outlining the procedure to be followed when an accusation of questionable ethics has been made, than to the definition of ethics itself. Perhaps that is understandable because notions of fair play and clear advance notice of what conduct is complained of, require provisions that outline the process of *enforcing* codes in more elaborate terms. The notion of "due process" that pervades the forensic sciences, because of its close connection to civil and criminal litigation, of necessity requires it to exemplify fairness in the application of its rules of conduct to its members.

Defining Prohibited Conduct

The definition of what constitutes unacceptable expert behavior is stated, in some ethics codes, only in general terms. Other codes go into great detail about the type of conduct that it requires of its members. Perhaps the majority of codes of conduct define ethics in terms of prohibitions. An example of the broad general approach is that of the American Academy of Forensic Sciences (AAFS), which contains just four general prohibitions:

> "a. Every member ... shall refrain from exercising professional or personal conduct adverse to the best interests and purposes of the Academy. The objectives stated in the Preamble to these By-Laws include: promoting education for and research in the forensic sciences, encourage the study, improving the practice, elevating the standards and advancing the cause of the forensic sciences;
> "b. No member or affiliate ... shall materially misrepresent his or her education, training, experience, area of expertise, or membership status within the Academy.
> "c. No member or affiliate ... shall materially misrepresent data or scientific principles upon which his or her conclusion or professional opinion is based;
> "d. No member or affiliate ... shall issue public statements that appear to represent the position of the Academy without specific authority first obtained from the Board of Directors".[c]

It is understandable that required conduct provisions of a society such as the AAFS, which gathers under its umbrella members who are engaged in a wide variety of different disciplines, be fairly general. It is impossible to describe specific instances of prohibited conduct when the organization contains a broad spectrum of specialties, each of which may also be subject to ethics codes within its own specialty.

An example of a code of professional conduct that is more explicit in terms of prohibited professional

behavior is the American Board of Criminalistics (ABC) Rules of Professional Conduct which, in Article IV.5.d of the bylaws, enumerate 18 specific mandatory rules of behavior that its members "shall" obey. Unlike the AAFS code, it enjoins members to report to its Board any violation of the rules by another applicant or diplomate.[d] The ABC rules place a great emphasis on impartiality and integrity when dealing with evidence,[e] and prohibit members from using "techniques and methods that are known to be inaccurate and/or unreliable".[f] Considering that users of a technique will invariably advocate its reliability, there may, nevertheless, be an important segment within a profession that frowns upon use of some methods of analysis, which they consider to lack reliability. One wonders how controversies on those issues can ever be effectively resolved.

Defining Ethical "Aspirations"

Rather than defining the prohibited conduct, the approach within some professions is to define appropriate professional conduct in terms of the aspirations toward which its members should strive. The Code of Conduct of The Forensic Science Society in the United Kingdom is an example. It states that members have a duty to

> "- . . . conduct themselves honourably in the practice of their profession
> - promote to the utmost of their power the interests of the Society * * *
> - have special regard at all times to the public interest and to the maintenance of the highest standards or competence and integrity * * *
> - only undertake any forensic activity commensurate with and in the field within which they are registered or accredited by the Society".[g]

Some groups eschew categorizing any specific instances of unprofessional conduct or the categorization of "standards" of behavior. An example of such a group is the American Psychological Association (APA), which describes the "Ethical Principles of Psychologists and Code of Conduct" into five rather wordy aspirations, wherein a full paragraph of text each is devoted to Beneficence and Nonmaleficence; Fidelity and Responsibility; Integrity; Justice: and Respect for People's Rights and Dignity. The APA code then elaborates on these aspirations in more than a dozen pages of ethical standards stated in broad, generalized concepts.[h]

While flexibility in terms of expected professional aspirations are the norm in the behavioral sciences, similar approaches are taken also in some organizations that deal with the physical or comparative sciences. The International Association for Identification (IAI) Code of Ethics is not worded in terms of prohibited conduct but is, instead, worded in the form of a personal pledge of good behavior that its members publicly affirm. Its true "code" of prohibited conduct then follows as "Standards of Professional Conduct".[i]

Similarly, the code of ethics of the Northwest Association of Forensic Scientists[a] lists series of ethical considerations relating to five different categories: those relating to the scientific method, to opinions and conclusions, to aspects of court presentation, to the general practice of forensics, and to the profession.

The importance of a thorough familiarity with a profession's ethics rules is particularly crucial because potential legal sequelae may follow from proven unethical conduct. When legal actions for damages are brought against forensic scientists based on either negligent or willful conduct, or when a prosecution for criminal conduct is initiated, whether the activity that is the subject of the complaint is considered "ethical" or in accord with "professional standards" may well be taken by courts as depending on what conduct the applicable professional society expects of its members.

Fact Settings Presenting Potential Ethical Problems

Fact settings in professional practice that may involve ethical problems fall within certain categories. While clear authority for each fact setting may be difficult to provide, the experience of persons serving on ethics committees may serve as a guide.

A first type, and perhaps the most frequent complaint lodged against experts is in regard to inaccuracies in the way they represent their training, education, and experience. Misrepresentation of this sort may occur in statements made in a public forum (in speeches, on websites, or in correspondence), or in drafting curriculum vitae that are disseminated to potential clients and courts. If the inaccuracies simply amount to exaggeration that can be justified as "puffing," the perceived inaccuracy will not necessarily be considered an unethical practice. More often, the assertions are made in support of establishing

the expert's competence and qualifications in depositions or in court testimony. If the misrepresentation is deemed significant and occurs at a time the expert is giving sworn information to a court, whether in a deposition or testimony, the expert may also be subject to criminal prosecution for perjury, or in damages in a civil suit by parties injured or aggrieved as a result of willful misstatement of one's background and experience [2].

Examples of misstatements that are normally considered material and significant would be to assert having received credentials (diplomas, academic degrees, board certifications, and honors), which were not, in fact, obtained. In this same category, fall claims of membership or of a category of membership status for which the expert has not been qualified.

A second category of ethical complaints, and one that is more fraught with enforcement difficulties, encompasses assertions that a person was incompetent when he engaged in a certain examination. In many disciplines, it may be disputed whether a particular methodology leads to reliable or repeatable results. When a method is fairly novel, there exists a danger in prematurely using an unproven method as the basis for expert opinions in court. Influential people in a discipline may assert that the method was not properly validated and that, consequently, reliance on test results produced by such methods is evidence of incompetency. Because of the difficulty of defining "competency," some associations refrain from entertaining complaints based on incompetency altogether, especially if the organization has fairly rigorous standards for admission to membership and the person complained of has met these criteria.

Lack of established reliability or inability to establish method validation has become a more frequently litigated issue in courts as a result of court decisions and other legal provisions, which place restrictions on opinion testimony based on unreliable processes or technologies[j].

A third category of potential problems has to do with fraud and deceit in such areas as misrepresenting data examined, results obtained, or in testifying to conclusions that are not supported by the examinations conducted. While fraud and deceit typically connote willful conduct, certain acts of omission may be placed in this same category – allegations that a person did not use recommended or standard methods which the profession recognizes as valid and routine for a specific purpose.

Finally, there exists a category of potential ethical conflicts that do not fall within the above categories, such as those arising from conflicts of interest or violation of confidentiality rules. Alleged unethical conduct in these areas often depends on existing law on conflicts of interest and confidentiality in the jurisdiction where the conduct arose.

Sanctions Against Experts for Unethical Conduct

Most, but not all, ethics codes provide for a gamut of potential consequences that befall a member found to have violated its ethical provisions. The sanctions range from measures that may be seen by the public as a simple "slap on the wrist," such as a reprimand (whether oral or in writing), to the more serious ones, which encompass public censure, suspension of membership for a stated period of time, or expulsion. Where applicable, the most severe sentence of expulsion may also be accompanied by a revocation of credentials or professional certifications.

The imposition of severe sentences against professionals such as membership suspension or expulsion can be imposed only if a procedural code exists which affords due process to the accused member. Due process typically requires notice of the specific charges brought, presentation of evidence at a hearing at which the accused is permitted to attend, confront the charges against him, and present evidence to rebut them. It also typically requires the complaining association to carry the burden of proving that the violation occurred, though the quantum of proof is by no means uniform.

In some codes, the quantum of required proof of unprofessional conduct must be "beyond a reasonable doubt," though, more typically, ethics provisions require that unprofessional conduct be established by either the greater weight of the evidence or, at most, by clear and convincing evidence. Some ethics codes permit the member against whom the complaint is filed to be assisted by legal counsel at the hearing. Some codes also make provisions for a right to appeal an ethics board decision to the entire membership of the organization.

Since the imposition of a severe sanction may impair the ability of the member to engage in his profession, especially when an expulsion also has the

effect of revoking the member's certification, courts whose power was thereafter invoked by the expelled member have insisted that the procedure, whereby the expulsion was effectuated, comported with due process of law. In deciding such lawsuits, when the procedural aspects of the ethics codes were found to satisfy notions of fairness and fair play, courts tend to be deferential to the professional society's factual determination of unprofessional conduct.

Conclusion

If professional societies intend to retain the power to sanction conduct of their members that is deemed unethical, the rules of professional conduct must be described with particularity so as to provide adequate notice to its members of the type of conduct that is deemed unacceptable. The process by which violations of the codes of professional conduct are to be enforced must also comport with basic notions of fair play and due process.

End Notes

[a.] For example, Medical examiners will belong to generalist countrywide, state or local medical societies, such as the American Medical Association (AMA), and further be subject the professional rules of conduct of the National Association of Medical Examiners (NAME), the American Board of Forensic Pathology, and the Pathology-Biology Section of the American Academy of Forensic Sciences.

[b.] See, in this regard the related article on Malpractice of Experts.

[c.] The Code of Ethics and Conduct is contained in Article II of the group's Bylaws. See, 2007 DIRECTORY OF MEMBERS AND AFFILIATES – AMERICAN ACADEMY OF FORENSIC SCIENCES, p. 222. The version printed herein was as modified at the Annual Meeting of the membership in Washington, D.C. on February 20, 2008. It will be noted that subsection a. appears to be a catch-all provision that may be difficult to enforce for lack of specificity. Both subsections b. and c. deal with misrepresentations, but the addition of the word "material" may also make it difficult to know in advance when a misrepresentation might not rise to the level of materiality.

[d.] http://www.criminalistics.org (accessed on, 2008).

[e.] Subsections 3–5 require members to treat objects of potential evidential value with care so as to ensure their integrity, provide appropriate technical analysis while following the necessary standards and controls.

[f.] Subsection 6. Emphasis is placed elsewhere on the duty not to "exaggerate, embellish or otherwise misrepresent qualifications," (sub. 11) and "maintain an attitude of independence and impartiality in order to ensure an unbiased analysis of the evidence" (in sub. 14).

[g.] http://www.forensic-science-society.org.uk (accessed on, 2008) See also the provisions of The International Institute of Forensic Engineering Sciences, Inc. (IIFES), which, in defining its Rules of Professional Conduct, lists 11 aspirations in positive terms rather than in prohibitions. The 11 provisions are: "1. Be aware of ones [sic.] own professional and technical qualifications in dealing with each case, and address only those factors that are within ones [sic.] own expertise and competence. Seek the assistance of other qualified experts whenever necessary. "2. Treat ones [sic.] own personal integrity with great respect, and never do or say anything that might compromise that personal integrity. This entails not only saying the truth art all times, but also giving due weight to all pertinent observations and facts. "3. Treat all information from a client, agency, or any other exclusive source with the confidentiality required. "4. Treat every object or specimen of potential evidential value with the case and control necessary to preserve its integrity. "5. Utilize the appropriate standards and controls in conducting examinations and analysis. "6. Render opinions and conclusions strictly in accordance with the evidence in the case and only to the extent justified by that evidence. "7. Maintain an attitude of independence and impartiality in order to ensure an unbiased analysis and presentation of the evidence. "8. Carry out the duties of this profession in such a manner as to inspire the confidence of the public. "9. Respect ones [sic.] peers in this profession and regard them with the same standards that one holds for oneself. "10. Report to the IIFES board any violation of these rules by any other person who is accredited by the board. "11. Refuse to accept, on a contingency fee basis, any assignment involving opinions, conclusions, or expert testimony intended for use in arbitration and/or litigation." See, http://www.iifes.org/ethics.htm (Last visited, Feb. 11, 2008).

[h.] The current version of the Ethics Code of the American Psychological Association went into effect on June 1, (2003) See, http://www.apa.org/ethics/code

2002.html; To the same effect, though not as wordy, are the Ethical Guidelines for the Practice of Forensic Psychiatry of the American Academy of Psychiatry & the Law. See, http://www.aapl.org/ethics.htm (accessed on 2008).

[i] The Code of Ethics of the International Association for Identification is printed on page 1 of the 2008 IAI Membership Directory. It provides: "As a member of the International Association for Identification, and being actively engaged in the profession of Scientific Identification and Investigation, I dedicate myself to the efficient and scientific administration thereof in the interest of Justice and the betterment of Law Enforcement". "To cooperate and others in the profession, promote improvement through research, and disseminate such advancement in my effort to make more effective the analysis of the expert". "To employ my technical knowledge factually, with zeal and determination, to protect the ethical standards of the profession of Scientific Identification and Investigation". "I humbly accept my responsibility to Public Trust and seek Divine guidance that I may keep inviolate the Profession of Law Enforcement." This "Code of Ethics" is then followed, on page 2, by a series of nineteen specific instances of mandatory conduct that are titled "Standards of Professional Conduct".

[j] See, in this regard, the related articles on Admissibility of Expert Opinion Evidence (**Expert Opinion in Court: a Comparison of Approaches**; **Expert Opinion: United States**; **Expert Opinion in Court: Civil Law Jurisdictions (France, Germany, Italy, and Spain)**; **Expert Opinion: United Kingdom, Canada, and Australia**) and on the influential court decisions in Daubert v. Merrell Dow Pharmaceuticals ***Daubert v. Merrell Dow Pharmaceuticals*** and in Kumho Tire v. Carmichael [514].

References

[1] For a sample of the codes prescribing the required ethical provisions for its members of just a few of the dozens of professional societies in forensic sciences not specifically mentioned in this article, see, e.g.: -Code of Ethics for Forensic Artists, http:///.theiai.org/certifications /artist/ethics.php -Academy of Behavioral Profiling: http: //www.profiling.org/abp_conduct.html;Piette M.J. (1991). Codes of professional ethics for forensic economists: problems and prospects, *Journal of Forensic Economics* **4**(3), 269–276; National Guild of Hypnotists, www.canadianhypnosiscentre.com/docs/NGH ethics.pdf; Australian and New Zealand Forensic Science Society, http://anzfss-vic.blodspot.com/2006/10/code-of-ethics. html; International Association of Forensic Nurses, http: //www.iafn.org.membership/membershipEthics.cfm; Forensic Consultants Association, http://www.sdfca.org. about.php;Northwest Association of Forensic Scientists, http://www.nwfs.org/Documents/Code%20of%20Ethics. pdf;American Society of Crime Lab Directors Code of Ethics, http://www.ascld.org/ethics.html. (accessed on, 2008.

[2] A different, but related, form of "unethical" conduct of this nature exists when experts, in their student days, are shown to have resorted to falsification of data, plagiarism, or other fraudulent practices in academic pursuits. When these practices are discovered years later, experts may suffer the added consequence of seeing their earned credentials revoked for fraud. See, Johnston, R.G. & Oswald, J.D. (1998) Academic dishonesty: revoking academic credentials, *Journal of Marshall Law Reviews* **32**, 67.

ANDRE MOENSSENS

Evidence: Best, Rule ***see*** Best Evidence Rule

Evidence: Complex ***see*** Jury Dynamics

Evidence: Hearsay ***see*** Hearsay Evidence

Evidence: Interpretation ***see*** Evidence Interpretation: a Logical Approach

Evidence: Novel Science ***see*** General Acceptance Test for Novel Expert Evidence

Evidence: Oral ***see*** Statistical Evidence in Court

Evidence: Packaging ***see*** Packaging and Transport

Evidence: Presentation ***see*** Statistical Evidence in Court

Evidence: Publications as ***see*** Learned Treatises as Evidence

Evidence: Rules of

There exist compilations of rules, codified in statutes, court decisions, or other legal literature, that determine how courts operate, what types of evidence tribunals may consider in reaching decisions, and how other aspects of the litigation process are conducted. Most of these law rules are highly technical and are not particularly of interest to forensic experts. Some legal evidentiary terms, however, interface with expert testimony so often or in such significant ways that forensic examiners who are likely to have contact with the courts should be aware of their meaning, importance, and applicability.

The rules that pertain particularly to expert opinion testimony are discussed in this Encyclopedia in many different articles. The length of each article varies with the importance or complexity of its subject matter and the frequency of its interface with expert testimony. Thus, forensic experts can consult the articles listed below covering rules of evidentiary law on the following:

- **Expert Opinion in Court: a Comparison of Approaches**
- **Expert Opinion in Court: Civil Law Jurisdictions (France, Germany, Italy, and Spain)**
- **Expert Opinion: United Kingdom, Canada, and Australia**
- **Chain of Possession of Tangible Evidence**
- **Cross-Examination of Experts**
- **Demonstrative Evidence**
- **Direct Examination of Experts**
- **Foundation Testimony**
- **Hearsay Evidence**
- **Hypothetical Question**
- ***In Limine* Motions and Hearings**
- **Therapeutic Jurisprudence**
- **Judicial Notice of Scientific Principles and Facts**
- **Jury Instructions on Expert Testimony**
- **Learned Treatises as Evidence**
- **Chemical, Biological, Radiological, and Nuclear Investigations**

ANDRE MOENSSENS

Evidence Collection and Preservation: Casting

Casting – The First Steps

Whether it is a tire track, footwear, or tool mark impressions, the first step is to take photographs, including a scale (Figure 1). That way should something go awry during the casting process, the crime scene investigator (CSI) will at least have the



Figure 1 Scale photograph of footwear impression

ability to produce a full-size scale photograph of the impression.

Once photographs have been taken, the casting can begin. With the introduction of dental stone as a casting material, Bodziak [1] rightly states that the compound can be poured directly into the indented impression. If the impression is on even a slight slope, a frame or form should be placed around the impression. This could be a commercially produced frame for the purpose. Some CSIs prefer to use a section of flexible plastic lawn or path edging (Figure 2). If neither is available, cardboard is a good substitute. Whichever system is used, press the frame into the ground to ensure no liquid compound can escape. On uneven ground use some soil to block gaps at the bottom of frame.

A fine spray of shellac paint or hairspray should be applied to the impression [2] as a fixative. The CSI must apply this carefully and allow it to form a haze over the impression by spraying horizontally downwind. Never spray it directly on to the impression, to avoid damaging the fine details. In addition, practice in the United Kingdom and recommended by Bodziak [1] is the use of a release agent. This allows the soil to come away from the cast easier. Release agents can be frying pan sprays, WD40® or even a baby powder in a puff spray.

Casting the Impression

The best compound to use when casting impressions in soft earth is Class I dental stone [3], a gypsum powder product. Different gypsum compounds have different properties and the results can vary depending on the product used. Generally a water-to-powder ratio of around 40 parts water to 100 parts powder will provide the correct consistency and strength [3].

Water can be added to a premeasured bag of dental stone or use a plastic bowl to provide a suitable mixture of the powder and water. The gypsum industry recommends that the powder is added to the



Figure 2 Form constructed from lawn edging

water; hence, using a bowl is a better method for the less experienced CSI.

Pour a measured amount of water into the bowl and add some powder. Continue to stir the mixture, while adding the water. During the mixing process, one may see the mixture darken around the edge and appear much thinner. This is caused by water separating out from the mixture. The mixture must be well combined as separation will adversely affect the cast. It is important that the CSI make enough of the mix to pour into the impression and surrounding area in one operation. Attempting to add a further mixture may result in a loss of evidence and a broken cast. It is better to have a surplus of the mixture. The dental stone is relatively inexpensive.

Practitioners will differ on the consistency of the mixture; some describe it as similar to double cream or thick emulsion paint. When pouring the mixture from either a bag or a bowl, the CSI should hold the mixture as close to the ground as possible and pour it slowly into the impression (Figure 3). When complete, the liquid cast should be gently tamped down with a small stick. This will knock out most of the air bubbles that may have formed. To give the finished cast some extra strength, one can also add a small section of wire mesh. Small twigs are an effective substitute for the mesh, although the CSI should exercise care as the wood may expand causing the cast to crack while drying (Figure 4).

Drying time depends on the air temperature. In relatively warm weather, the time needed for drying can be 20–30 min, increasing time as the temperature decreases. Mixing gypsum and water produces an exothermic reaction. As the mixture sets, the heat generated is apparent when placing a hand over the cast. As it sets the gloss may fade and the cast will become matte in appearance. As the mixture begins to set, it is good practice to etch identification markings onto the top of the cast. These will vary according to agency protocols and standard operating procedures.

Impressions Covered by Water

During inclement weather, impressions, especially tire tracks, may be found in standing water. Bodziak [1] advises that the CSI should not drain the water from the impression area. Any items floating on the water but not touching the impression should be removed.

Place a frame or dam around the impression. The technician must be careful that they do not damage the impression by placing them too close to the impression itself. Bodziak [1] recommends about 2-in. separation (5 cm) as a minimum.

Sift the powder into the water until it has covered the impression to a depth of about 1 in. (2.5 cm) and



Figure 3 Dental stone mix being poured into impression



Figure 4 Strengthening the cast

then mix enough compound to cover the framed area to a depth of 2 in. (5 cm). The cast should be left for a minimum of 1 h.

Recovering the Cast

To remove the cast, once it has set, it is best to use a small trowel to ensure the CSI can release some earth from around and under the cast. This will prevent any thin parts of the cast from breaking. When recovered, the CSI should not attempt to remove any of the soil from the footwear impression (Figure 5). The cast needs to be air dried for about 48 h [3]. The soil will be removed by the examiner at the laboratory. This also allows the CSI to have a soil sample from exactly where the tire of footwear impression was.



Figure 5 Recovering the cast

This may provide additional valuable evidence in some cases.

Place the cast soil side upwards in a suitable box (*see* **Packaging and Transport**).

Casting in Snow

Footwear and tire impressions in snow can be problematic as attempting to photograph impressions will be extremely difficult because light may reflect off the snow. In addition, mixing any material used as a casting agent creates an exothermic reaction that can destroy the detail of the impression.

To overcome this, a fine coat of matte black spray paint should be applied gently over the impression prior to photography. Where this is unavailable, some CSIs gently sprinkle some black powder over the impression. Both the methods increase the contrast, enabling a better photograph to be produced.

Prior to casting, a product called "Snow Print Wax™" is gently sprayed on the snow. This is a spray wax solution that creates a thin membrane on the surface and reduces the effects of the exothermic reaction on the snow.

When casting only little water is added to the powder to start the reaction and this is then supplemented by the addition of snow, which melts but reduces the temperature rise.

Once set, the resultant cast can be removed and packaged in the same manner as one cast in soil.

Casting Instrument Marks

Instrument marks, commonly referred to as *tool marks* are cast in a similar way.

Prior to casting, a photograph is taken showing the position of the mark. Then a close-up photograph of the mark with a scale.

Casting materials vary greatly but they are normally proprietary silicone-based casting compounds of the flexible type used by dentists when providing dentures.

When cured, they can be peeled away from the mark or impression. Prior to removal, it is good practice to show the vertical position of the mark by means of an arrow, indicating whether this is pointing up or down.

The resultant cast is then placed between two sheets of acetate or plastic to prevent damage. This should be sealed all around with adhesive tape and details written on the plastic sheet and on the attached chain of custody document (*see* **Crime Scene Documentation**).

References

[1] Bodziak, W.J. (2000). *Footwear Impression Evidence*, 2nd Edition, CRC Press, Boca Raton, pp. 59–97.

[2] Gardner, R.M. (2005). *Practical Crime Scene Processing*, CRC Press, Boca Raton, pp. 252–258.

[3] Hildebrand, D.S. & Miller, M. (1995). Casting materials – which one to use? *Journal of Forensic Identification* **45**, 618.

ALLAN MATHIESON SCOTT

Evidence Evaluation *see* Bayesian Networks, Case Assessment and Interpretation

Evidence Interpretation: a Logical Approach

Introduction

A criminal trial is suffused with uncertainty, and everyday court language reflects this: words such as *probability, likelihood, likely, unlikely* are regularly invoked, as well as phrases such as *balance of probabilities, beyond reasonable doubt, can't exclude the possibility of*, etc. among others. It is important for us to gain a good understanding of the role of the scientist in all this. No one pretends that a jury reasons logically, and few lawyers and judges have a grasp of the logic of forensic inference to match that of Professor Kaye (**Interpretation: Legal Perspective**). However, here we are concerned with the role of the forensic scientist, and it seems desirable that if scientists must reason in the face of uncertainty, then they should do so within a robust framework of logic.

This article describes such a framework. It would hardly be compatible with a scientific viewpoint to claim that it is the only framework that could ever exist, but at the time of writing it is the one that the overwhelming majority of students of evidence interpretation espouse.

Many people equate inference with statistics, but that is a mistake. Statistics is about data and we do not discuss data analysis in this article. We will be concerned with probability.

In the following section we give a brief discussion of the nature of probability. We then go on to look at how a fundamental theorem of probability theory – Bayes' theorem – provides a model for inference in the context of a criminal trial. We then see how this approach leads to the formulation of three principles of interpretation. We discuss issues of how probabilities can be addressed – with or without data. It is necessary to talk about a common fallacy encountered in the courts – the prosecutor's fallacy. And finally, we discuss the presentation of evidence at the court.

Probability

Three Examples

Consider the following statement:

> The card revealed when I cut this deck of playing cards will be a spade.

This seems a fairly simple statement[a] – but is it true? Well, the answer to that depends on a number of things. You would, presumably, like to know whether or not this is a conventional deck of cards; whether it has been prepared in any particular way – e.g., by shuffling; whether the person making the statement has any facility for party tricks; and so on. If you are told that it is indeed a conventional deck, that it has been shuffled, and that there is no question of a trick, then you might feel that the probability of the statement being true is one quarter.

Now consider:

> England will beat Switzerland in their next soccer match.

This is another simple statement (again making some assumptions about the knowledge of the reader), but, again, deciding on whether it is true or not depends on many things. Indeed, on the days leading up to the match the popular press will furnish you with masses of information about the previous encounters between the sides, the composition of the teams, the recent form of individual team members, the expected strategies of the two sides, and so on. Sporting pundits will make varying predictions and, no doubt, individual fans will be swayed by emotive issues such as national pride. Hence, resolving your uncertainty with regard to the truth of this second statement is far more difficult than it was for the first. Indeed, we should not be at all surprised when different people

express different probabilities. If, in turn, you wish to make your own rational assessment (perhaps because you are considering a small wager on the outcome!), then you will be more swayed by those who appear to be knowledgeable, rather than those who are more inspired by emotion.

Finally, consider the following:

The defendant murdered Mr Y.

This is a statement made by a prosecuting counsel to you, a juror in a criminal trial. Is the statement true? You might form an initial view from the physical appearance of the man in the dock: physiognomy, clothing, body language, etc. The prosecutor calls witnesses to show how the defendant threatened to shoot Mr Y with his shotgun, how Mr Y was indeed shot by a shotgun, how the defendant initially confessed to the police but later retracted, and so on. The defense attorney leads evidence from witnesses to the effect that Mr Y did not actually own a shotgun, he was maltreated by the investigators, he was 30 miles away at a prayer meeting at the time of the incident, and so on. Subject to the direction of the judge, each of these items of evidence will have a bearing on your uncertainty with regard to the truth of the prosecutor's statement.

Basic Concepts

The first thing to notice about the three examples these is that each of them may or may not be true. The extent to which you believe each of the three to be true depends on many things. Let us consider them one at a time.

There are many differences in detail between the three examples but there are some important common elements. First, in each case we consider a *proposition* that may, or may not, be true. Next, your uncertainty about its truth is influenced by what you know, what you are told, and how effectively you employ the information that is available to you. These ideas stimulate a form of mathematical notation that we will use throughout this article. For a particular proposition, we write:

$$\Pr(H|I)$$

where Pr denotes probability, H is the proposition whose truth is uncertain,[b]| is shorthand for "given", and I represents everything that we know, or assume, that is relevant to the truth of H.

Our shorthand notation is convenient because in the space of a few letters we summarize:

The probability that the proposition H is true given everything that we know that is relevant.

A few remarks about the nature of probability are desirable here. First, probability is a means of encapsulating our uncertainty about the truth of a proposition. Second, we have emphasized that our probability for the truth of a proposition depends entirely upon what we know (or assume). Probability is then conditional. Third, it necessarily follows that probability is *personal*; apart from relatively simple situations – such as tossing coins, rolling dice, etc. – different people considering the same proposition will have different levels of knowledge.

Laws of Probability

There is a rigorous body of mathematical theory that shows that provided that probability is constrained to obey certain laws, then it is the most effective means of dealing with uncertainty.

The first law states that probabilities must lie between 0 and 1. The other two laws deal with addition and multiplication of probabilities. There are so many excellent texts that explain probability theory that we give no more detail here. References in the forensic field include Aitken and Taroni [1] and Evett and Weir [2].

There is one aspect of probability theory that we do need to explain here – the concept of odds. The odds in favor of a proposition being true are given by the probability of it being true divided by the probability of it being false. In mathematical terms,

$$O(H|I) = \frac{\Pr(H|I)}{\Pr(\overline{H}|I)} \qquad (1)$$

where $\overline{H}$ denotes "the proposition H is false".

Inference in a Criminal Trial

The analysis that we discuss in this section is perfectly general, but it will help if we present it in the form of an example.

Eyewitnesses, outside a nightclub, saw two men fighting and grappling with each other. Eventually, one of the two men drew a knife and fatally stabbed the other. The man with the knife ran from the scene. After interviewing eyewitnesses and taking account

of all that was known about the circumstances of the crime, police arrested a suspect shortly after the incident. The suspect denied all knowledge of the crime and said that he had never been in the company of the deceased individual. In this example, we consider the interpretation of the results of the examination of the clothing of the deceased and the suspect for fibers. A number of fibers were found on the clothing of the suspect that matched samples from the upper clothing of the deceased; and fibers were found on the clothing of the deceased that matched samples from the clothing of the suspect.[c] We consider the case at court; we refer to the deceased as the "victim" and the suspect as the "defendant".

The defendant has pleaded not guilty to the charge of murder and the jury hears evidence about the circumstances of the incident from the eyewitnesses and from the investigators. We will summarize all of this nonscientific evidence by I. The prosecution position is that the following proposition is true:

H_p: the defendant is the person who stabbed the victim.

The defense position will, of course, be completely different from this. The defense proposition will depend to a considerable extent on what, if anything, the defendant wishes to say in his own defense. For the sake of the present discussion, the defense proposition will be taken to be:

H_d: the defendant had nothing to do with the incident.

The evidence of the eyewitnesses, investigators, and what the defendant himself says, if he chooses to give evidence, will together influence the jurors' uncertainty in relation to the two propositions. We do not study this process, which is undoubtedly extremely complex. We content ourselves with recognizing that the jurors' belief may be expressed in terms of odds. Because our interest is in considering the impact of the scientific evidence, we call these the *prior odds* and, using our notation, we can write them as follows:

$$\text{Prior odds} = \frac{\Pr(H_p|I)}{\Pr(H_d|I)} \tag{2}$$

Now the scientist gives evidence with regard to the results of the examination for fibers – we summarize this evidence by E – and we are interested in what effect this has on the jurors' belief with regard to the truth of H_p and H_d. We express this as the *posterior odds*:

$$\text{Posterior odds} = \frac{\Pr(H_p|E, I)}{\Pr(H_d|E, I)} \tag{3}$$

Hence, we recognize that the jurors' uncertainty has changed because it is now *conditioned* by more evidence than was available before the scientist presented his results. This notion of updating uncertainty in the light of new evidence is not just confined to the court of law – it is an integral part of everyday life and there is a standard result of probability theory that enables us to write out a formal, logical relationship between the prior and posterior odds. It is called *Bayes' theorem* and, in the present situation, it states that in odds form:

$$\frac{\Pr(H_p|E, I)}{\Pr(H_d|E, I)} = \frac{\Pr(E|H_p, I)}{\Pr(E|H_d, I)} \cdot \frac{\Pr(H_p|I)}{\Pr(H_d|I)} \tag{4}$$

Nonmathematical readers may, by now, be feeling somewhat apprehensive – but be reassured that (i) there will be little more in the way of demands on your patience in this article and (ii) the insights that come from this somewhat forbidding expression are profound, yet very simple. Let us unpick it slowly.

The equation tells us that the posterior odds can be derived from the prior odds if we multiply them by a ratio (the middle ratio in the equation). This ratio has a special name that will be recurring throughout this encyclopedia; hence, it will be worth spending some time in grasping it. It is known (for reasons that we do not need to go into) as the *likelihood ratio* and practically everybody denotes this by its short form – LR. It is worth writing it out:

$$LR = \frac{\Pr(E|H_p, I)}{\Pr(E|H_d, I)} \tag{5}$$

Hence, our forbidding bit of mathematics may be written in words as follows:

The *posterior odds* are equal to the *prior odds* multiplied by the LR.

Now, one reaction to this is that it is a pretty unrealistic model for the way in which real-world jurors behave. Indeed, that would appear to be the case. In *R v Denis J Adams* [3], defense attempted to encourage the jurors to assign probabilities to all of the relevant aspects of the nonscientific evidence: something that proved to be an uncomfortably complicated process. To cut a long story short, the Appeal Court took a firm line against this kind of approach to nonscientific evidence, and quite rightly so. But what we are looking

for here is not something to model the jurors' thinking – but that of the *scientist* and the insights that follow from this analysis have profound implications for scientific evidence.

This is what the jury considers (however imperfectly):

$\Pr(H_p|E, I)$: the probability that the prosecution proposition is true, given the scientific evidence and all the other evidence;
$\Pr(H_d|E, I)$: the probability that the defense proposition is true, given the scientific evidence and all the other evidence.

This is what the scientist must consider:

$\Pr(E|H_p, I)$: the probability of the scientific evidence, given that the prosecution proposition is true and given all the other evidence;
$\Pr(E|H_d, I)$: the probability of the scientific evidence, given that the defense proposition is true and given all the other evidence.

Let us consider these two questions in the example that we have outlined. The model directs that the scientist should address two questions:

$\Pr(E|H_p, I)$: "what is the probability that I would find a number of matching fibers if the defendant was the man who stabbed the victim, given the circumstances of the incident?"
$\Pr(E|H_p, I)$: "what is the probability that I would find a number of matching fibers if the defendant had nothing to do with the incident, given the circumstances of the incident?"

We do not consider in this article the issues that the scientist must consider to address these questions. A few remarks are desirable, however. We have been deliberately vague in talking about "a number of" matching fibers: in practice, the scientist would know the number, at least to a broad order of magnitude. In relation to the first of the two probabilities – the numerator of the LR – the scientist needs to take into account the circumstances of the incident as far as they relate to the nature and extent of contact between the examined garments if the prosecution proposition were true (transfer), and the time interval between the incident and the taking of the exhibits for examination (persistence). In relation to the second probability – the denominator of the LR – the scientist will refer to whatever data are available relating to the presence of fibers of a given type on the clothing of persons who are unconnected with particular criminal incidents. The circumstances of the case are relevant when the scientist considers the issue of the extent to which the data are representative of the conditions that pertain in this particular case.

Three Principles of Interpretation

The analysis we have presented here suggests three principles that a scientist should adhere to when interpreting evidence.

1. Scientific evidence must be interpreted within a framework of circumstances.
We have seen that the conditioning of the probabilities that constitute the LR includes I, which summarizes the nonscientific evidence. Much of this, of course, will not have a bearing on the scientific evidence but it is essential to recognize those aspects that do. In the example that we have discussed, there are aspects of I that influence the scientist's deliberations relating to transfer and persistence and also any background data that might be relevant to the probability of finding matching fibers on the clothing of the defendant were the defense proposition true.

As another example, evidence about the ethnic appearance, if known, of a person who left DNA at a crime scene is highly relevant to the choice of a database for interpreting a matching DNA profile.

We have seen much confusion among scientists as to the extent of the I information impacting on the assessment of the evidence (see [4] for example). The forensic scientist should be made aware only of the relevant information (nature of the alleged activities, timing, nature of the garments and surfaces, etc.) that may impact on his/her expectations as to the forensic material recovered and the comparative results obtained. Other information regarding the case that will impact on the decision of the ultimate issue should neither be considered nor assessed by the forensic scientist.

2. Scientific evidence can only be interpreted by considering two[d] propositions.
As we have seen, it is necessary for the scientist to address the probability of the evidence, given both the prosecution and defense propositions. In more complex cases, such as where there are multiple defendants, there may be more than two: these are much more difficult to deal with because the simple

form of Bayes' theorem that we have used here is only applicable when there are two propositions. We do not discuss such cases here, but content ourselves with observing that whatever the conditions the scientist cannot present a balanced interpretation unless both the prosecution and defense positions are considered.

3. It is necessary for the scientist to address questions of the kind "what is the probability of the evidence, given the proposition?"

Whereas the jury necessarily addresses questions of the kind "what is the probability of the prosecution proposition given the evidence?" and "what is the probability of the defense proposition given the evidence?", we have seen that the scientist must consider the probability of the evidence, given the prosecution proposition and the probability of the evidence given the defense proposition. The ratio of the answers to these two questions – the LR – is of crucial importance.

Statistics and Qualitative Opinions

A common reaction to the foregoing treatment is "that's all very well, but it doesn't work in my area of forensic science because we don't have any data to assign the probabilities that are required." But this is a misconception that arises from a confusion between probability and statistics. Certainly, if data are available to the scientist, then the logical approach leads to the effective and proper use of those data and, no doubt, we can cite DNA profiling as a field where an enormous work has been done on the proper use of DNA statistics. But the data are secondary: the logic of inference comes first, data come second. The strength of the logical approach is that it helps the scientist to identify the most appropriate *questions*. If we have data, then we may be able to provide quantitative answers to those questions. If there are no data, or if the data that are available defy simple analysis, then qualitative answers can provide powerful guidance to courts of law. A fine example of this is the field of documents examination and insightful illustrations, which are given in the chapter by Steve Day in this encyclopedia (*see* **Handwriting and Signatures, Interpretation of Comparison Results**).

There is a growing concern expressed in the literature when forensic findings are expressed by qualitative opinion and not substantiated by statistical numbers [5]. Opposing systematically human judgment to statistical value translates a misconception of the nature of science [6]. There is nothing wrong in applying informed judgment and expressing opinions in the form of qualified opinions such as in the following examples (still in relation to our above fiber case):

- The probability of observing such a number of fibers having been transferred, persisted, and recovered under the alleged conditions is high.
- It is very rare to observe this type and number of fibers, recovered by mere coincidence on the garments of a victim.
- The findings are what I would expect to find if the defendant had nothing to do with the incident, given the circumstances of the incident.

Such qualitative opinions amount to subjective probabilities [7] based on the training and experience of the examiners. As long as they are addressing relevant questions (as defined above) and that there are mechanisms in place to assess the reliability of them (either by a calibration with known data or through collaborative or proficiency tests), they have a crucial role to play in the interpretation of evidence.

The Case Assessment and Interpretation approach

In 1991 the Forensic Science Service (FSS) became a government agency: still a government department, still a nonprofit-making service, but with a strong emphasis on being run on business-like lines. One of the early developments was to introduce direct charging for casework, and so the FSS came to be financed purely through the sale of its services. An immediate consequence of this was that the costs of forensic science were no longer invisible to operational police officers – who were now referred to by the FSS as "customers". The aim of direct charging was to enable customers to make better decisions based on a greater sense of value for money. In the mid-1990s a project was initiated within the FSS to achieve improvement in value for money by encouraging greater participation on the part of the customer in decisions about what work was done in the laboratory. The Case Assessment and

Interpretation (CAI) model was developed with the objective:

> To enable decisions to be made which will deliver a value for money service meeting the needs of our direct customers and the Criminal Justice System.

Several FSS scientists participated in the CAI project and a number of them are contributors to this encyclopedia. The most important aspects of the work were published in a series of papers [8–12]; they included some key advances in thinking about evidence interpretation, which we now briefly discuss.

The Hierarchy of Propositions

Consider the following pairs of propositions:

H_p: the glass fragments came from window X.
H_d: the glass fragments came from some other broken glass object.

H_p: the semen came from Mr B.
H_d: the semen came from some other man.

H_p: the blood on Mr C's clothing came from Mr Z.
H_d: the blood on Mr C's clothing came from an unknown person.

These three have an important feature in common in that each pair relates to the *source* of some recovered material. These are examples of *source level* propositions and they represent level I in what the CAI project defined as the *hierarchy of propositions*. With the advent of more sensitive DNA techniques, it has become necessary to propose a further level, called *sublevel 1*, or *subsource level* [13] in the form of:

H_p: the DNA profile obtained is that of Mr Z.
H_d: the DNA profile obtained is that of an unknown person.

If we assume that a full, unmixed DNA profile (at 10 loci, for example) is obtained in some of the above cases, then the LR at subsource level will be in the order of a billion, assuming that the unknown contributor is unrelated to the suspected contributor and that no laboratory error has occurred. The LR, in such a case, will simply be the inverse of the match probability (*see* **Short Tandem Repeats: Interpretation**). To progress from the subsource to the source level, in addition to the match probability, we will need to consider uncertainties surrounding the relationship between the DNA profile obtained and the underlying body fluid assumed. It will depend on the type of presumptive tests conducted in an attempt to identify the body fluid, their specificity and sensitivity, the microscopic examination carried out, the efficiency of obtaining DNA profiles from targeted fluids, etc.

Consider the next set of examples:

H_p: Mr A is the man who smashed window X.
H_d: Mr A was not present when window X was smashed.

H_p: Mr B had sexual intercourse with Ms Y.
H_d: Ms Y had sexual intercourse some man other than Mr Y.

H_p: Mr C is the man who kicked Mr Z in the head.
H_d: Mr C was not present when Mr Z was kicked in the head.

Each of these three pairs relates to *activities* and they represent level II of the hierarchy. As with source level propositions, these will be addressed on the basis of the observations, measurements, and analysis that the scientist has carried out. But addressing propositions at the activity level, in general, requires more background information than for those at source level. In particular, the scientist will need to address issues of evidential transfer and persistence; and more may be needed in the way of background surveys. In relation to the first pair, for example, access to data on glass found on the clothing of people unconnected with particular crimes is highly desirable: much more can be read on issues relating to activity level propositions in glass cases (*see* **Glass Evidence: Bayesian Approach to**).

Now consider the following:

H_p: Mr A committed the burglary.
H_d: Mr A had nothing to do with the burglary.

H_p: Mr B raped Ms Y.
H_d: some other man raped Ms Y.

H_p: Mr C assaulted Mr Z.
H_d: Mr C had nothing to do with the assault on Mr Z.

These represent the highest level – level III – of the hierarchy and can be termed *offense level*

propositions. If the scientist can address propositions at this level, and this will depend to a large extent on the amount of information that is available to him/her, then he/she will be providing the jury with the greatest possible assistance in their task. Retreating from offense level propositions down the hierarchy leaves more of the interpretative issues to the jury.

Preassessment

The CAI project did much to develop the notion of preassessing a case in order that there should be a meaningful discussion between scientist and investigator with a view to optimizing value for money. The scientist is encouraged to consider what propositions may be addressed *before* attempting any examination and considering what magnitude of the LR might be expected if the prosecution proposition were true or if the defense proposition were true. This process is described in more detail, with a worked example elsewhere (*see* **Case Assessment and Interpretation**).

Structure of Statements

The three principles of interpretation lead naturally to a structure for a formal statement for use in criminal proceedings [12]. We concentrate here on the section having interpretative elements (leaving aside *technical issues* and *examination and results* sections).

Framework of circumstances
The scientist will have been provided with information about the circumstances of the alleged crime, the suspect, and what the suspect says about the incident. These circumstances will have been directly relevant to the formulation of the propositions that have been addressed and also to the interpretation of the observations that have been made. It is therefore important that the framework of circumstances, as understood by the scientist at the time of writing the statement, should be clearly expressed in the statement. Of course, it is necessary to recognize that the circumstances are rarely based on incontrovertible facts, but depend on evidence from other witnesses; as such, they may change as the case approaches trial. It is necessary then for the scientist to point this out and to emphasize that in the event of changing circumstances it may be necessary for him to revise his opinion.

Purpose
This would normally be a short section that explains the reasons for carrying out whatever scientific examination was undertaken.

Interpretation
This involves a clear statement of the propositions (and their level in the hierarchy) that are addressed and an assessment of the extent to which the findings are explained or not under one or other of the propositions.

Conclusion
The conclusion must help to address the questions raised in the purpose and should be expressed following the principles of interpretation. The statements must then be focused on the assessment of the findings in the light of the relevant propositions at hand. That implies the concept of various degrees of "support" for one proposition versus others depending on the magnitude of the LR.

To ease communication with the criminal justice system and especially when qualified opinions dominate the assessment of the LR, scientists may want to express their findings using a verbal scale of strength. For example, a common scale is summarized in Table 1.

The table is given for LRs in excess of 1, which therefore represent support for the prosecution proposition. The scale works in a directly comparable way for LRs less than 1, when the evidence represents support for the defense proposition. Such a verbal scale reflects a consensus on the correspondence between LRs and verbal equivalents describing the degree of support for one proposition versus another.

Table 1 Common scale showing the relationship between verbal equivalent and range of likelihood ratios (here for LRs in excess of 1)

Likelihood ratio	Verbal equivalent
>1–10	Limited evidence to support
10–100	Moderate evidence to support
100–1000	Moderately strong evidence to support
1000–10 000	Strong evidence to support
>10 000	Very strong evidence to support

We feel strongly that the scale should reflect the logical framework for interpreting evidence and encapsulate the concept of support for propositions. Any other verbal scales using terms such "possible", "probable", or "very likely" when referring the propositions at hand are logically flawed and ought to be avoided.

The Transposed Conditional

Dubbed "the prosecutor's fallacy" by Thompson and Schumann [14], this common error may conveniently be illustrated by an example from DNA profiling. Assume that the genotype of a crime sample is the same as that of a defendant; then the scientist's evidence might include a sentence of the kind "the probability of observing this genotype if the crime sample came from someone else is one in a million". The prosecutor's fallacy is to say "the probability that the crime sample came from someone else is one in a million". It is easy for the layman to confuse these two statements – and journalists do it all of the time – but, armed with the rigor of the previous sections, we can see the problem. The scientist is making a statement about the denominator of the LR: $\Pr(E|H_{\mathrm{d}}, I) = 10^{-6}$. The lawyer is making a statement about the posterior probability of the defense proposition: $\Pr(H_{\mathrm{d}}|E, I) = 10^{-6}$. The interchange between E and H_{d} is why this is known more widely among statisticians as *the fallacy of the transposed conditional*.

In spite of rulings in the British Central court of Criminal Appeal, particularly in *R v Doheny and Adams* [15], there is little doubt that the fallacy is not widely understood. In particular, consider the following extract from another Appeal Court judgment (*R v Bates* [16]), set down some 10 years after *Doheny and Adams*:

> Data drawn from empirical research is available to enable analysts to calculate the statistical likelihood of any person within the population having a particular allele at a particular locus. Using that data it is possible to estimate the *statistical likelihood that a particular sample of DNA originated from the person whose profile is being used for comparison*. This is usually referred to as the "*match probability*". (Italicized for emphasis.)

Moving away from DNA into fields where weight of evidence cannot be quantified, it has been common practice for experts to express opinions of the kind "the questioned handwriting was probably written by the suspect." We can now see that this kind of opinion is, in effect, a transposed conditional and is not logically sustainable.

Identity and the Logical Approach

Interpretation of some evidence types – most notably fingerprints, footwear, and toolmarks – are characterized by a willingness among experts to express categorical opinions of identity of source. The issues of identity, uniqueness, and individualization are discussed in **Identification and Individualization**; for the present we content ourselves by remarking that it is hard to reconcile such opinions with the logical approach to evidence evaluation. Typically, a conclusion of individualization amounts to claiming an infinite LR. Such an extreme value is difficult to justify satisfactorily [17].

In British Courts at present we have a paradoxical situation. The judgment in *Doheny and Adams* said, in relation to a DNA expert witness [15]:

> He will properly explain to the Jury the nature of the match [. . .] between the DNA in the crime stain and the DNA . . . from the Defendant. He will properly, on the basis of empirical statistical data, give the Jury the random occurrence ratio – the frequency with which the matching DNA characteristics are likely to be found in the population at large [. . .].
>
> The scientist should not be asked his opinion on the likelihood that it was the Defendant who left the crime stain [. . .].

This, as we have seen, was a response to the prosecutor's fallacy, which had occurred in both of the original trials and, although its wording could be considerably improved, it is difficult to take exception to what their lordships said. This judgment means that a DNA scientist may not give an opinion as to whether or not a defendant left a crime stain. We have seen from the logical model that this is right – the scientist should address the probability of the evidence given the proposition, and not the probability of the proposition given the evidence.

However, in relation to a fingerprint comparison, it has been accepted in courts throughout the world for some 100 years that it is perfectly reasonable for a fingerprint expert to give an opinion of the form "in my opinion, this mark was made by that person."

That is undoubtedly an opinion about the truth of a proposition.

Conclusion

Understanding the basic concept of probabilities is essential in developing a logical framework for interpreting forensic findings in court. The breadth of probabilistic knowledge required to tackle and understand these issues is reasonable (within the capability of any scientist or member of the judiciary) and allows us to identify clearly the role of the forensic scientist in the interpretation process.

We hope that the above section will allow scientists to identify clearly the nature of answers they are inclined to bring to the decision maker and how these assessments may be dependent on the circumstances of the case and the level of the propositions addressed. Having a clear underpinning logic is a decisive mechanism that creates greater transparency in the way scientific findings are presented in the court.

End Notes

[a.] It seems that every statement we make invokes some assumptions. Here, of course, we are assuming that the reader knows what a deck of playing cards is and is familiar with the idea of cutting the deck.
[b.] We use H for "proposition" because P is so often associated with "probability". I might be seen as standing for "information" in its broadest sense.
[c.] Of course, in a real case, the scientist would also search the clothing of the deceased to see if there were any fibers that match the clothing of the suspect. However, to keep the example simple, we will consider just the one direction of transfer.
[d.] In certain circumstances, there may be more than two – the point is that there cannot be only one proposition.

References

[1] Aitken, C.G.G. & Taroni, F. (2004). *Statistics and the Evaluation of Evidence for Forensic Scientists*, 2nd Edition, V. Barnett, ed, John Wiley & Sons, Chichester.
[2] Evett, I.W. & Weir, B.S. (1998). *Interpreting DNA Evidence – Statistical Genetics for Forensic Scientists*, Sinauer Associates, Sunderland.
[3] R. V. *Denis John Adams (No.2)*, Court of Appeal – Criminal Division, [1997] EWCA Crim 2474 (16th October, 1997).
[4] Wiersema, S.J. (2001). Is the Bayesian approach for you? *Proceedings of the Fourth European Meeting for Shoeprint/Toolmark Examiners*, Berlin, Germany, May 15–18 2001, pp. 149–153.
[5] Saks, M.J. & Koehler, J.J. (2005). The coming paradigm shift in forensic identification science, *Science* **309**, 892–895.
[6] Evett, I.W. (1996). Expert evidence and forensic misconceptions of the nature of exact science, *Science and Justice* **36**(2), 118–122.
[7] Taroni, F., Aitken, C.G.G. & Garbolino, P. (2001). De Finetti's subjectivism, the assessment of probabilities and the evaluation of evidence: a commentary for forensic scientists, *Science and Justice* **41**(3), 145–150.
[8] Cook, R., Evett, I.W., Jackson, G., Jones, P.J. & Lambert, J.A. (1998). A hierarchy of propositions: deciding which level to address in casework, *Science and Justice* **38**(4), 231–240.
[9] Cook, R., Evett, I.W., Jackson, G., Jones, P.J. & Lambert, J.A. (1998). A model for case assessment and interpretation, *Science and Justice* **38**(3), 151–156.
[10] Cook, R., Evett, I.W., Jackson, G., Jones, P.J. & Lambert, J.A. (1999). Case pre-assessment and review in a two-way transfer case, *Science and Justice* **39**(2), 103–111.
[11] Evett, I.W., Jackson, G. & Lambert, J.A. (2000). More in the hierarchy of propositions: exploring the distinction between explanations and propositions, *Science and Justice* **40**(1), 3–10.
[12] Evett, I.W., Jackson, G., Lambert, J.A. & McCrossan, S. (2000). The impact of the principles of evidence interpretation on the structure and content of statements, *Science and Justice* **40**(4), 233–239.
[13] Evett, I.W., Gill, P.D., Jackson, G., Whitaker, J. & Champod, C. (2002). Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks, *Journal of Forensic Sciences* **47**(3), 520–530.
[14] Thompson, W.C. & Schumann, E.L. (1987). Interpretation of statistical evidence in criminal trials: the prosecutor's fallacy and the defence attorney's fallacy, *Law and Human Behavior* **11**(3), 167–187.
[15] R v. Alan James Doheny, R v Gary Adams, Court of Appeal – Criminal Division, [1996] EWCA Crim 728 (31st July, 1996).
[16] R v. Richard Bates, Court of Appeal – Criminal Division, [2006] EWCA Crim 1395 (7th of July 2006).
[17] Saks, M.J. & Koehler, J.J. (2008). The individualization fallacy in forensic science evidence, *Vanderbilt Law Review* **61**, 199–219.

CHRISTOPHE CHAMPOD AND IAN WEBBER EVETT

Evidential Chain ***see*** Chain of Possession of Tangible Evidence

Evidentiary Reliability of Confession ***see*** Confessions: Evidentiary Reliability of

Evil: Illusion of

> The belief in a supernatural source of evil is not necessary; men alone are quite capable of every wickedness.
>
> -Joseph Conrad

> Nature has no morality. She makes no distinction between good and evil.
>
> -Anatole France

Interest in "evil" is growing. The psychological and psychiatric literature has reflected increasing attention to the concept of "evil" over the past two decades. While most of the relevant publications are from the field of social psychology, there has also been a growing interest in the fields of psychiatry and forensic mental health. Articles by prominent forensic psychiatrists have debated the issue of whether forensic psychiatrists should "define" and testify about evil [1, 2]. Nevertheless, attempts by behavioral science to define evil as though it were an objective and quantifiable concept will be inherently flawed. Because evil is a subjective moral concept with inextricable ties to religious thought, it cannot be measured by psychiatric science. Moreover, there does not appear to be any significant need to define or use the term *evil*, as forensic mental health already has working concepts describing deviant behavior that is harmful to others. This article will address why evil is an entirely subjective, nonscientific concept. Relevant historical issues will be discussed, as well as the apparent surge of interest in evil by forensic clinicians. The need to accept the term *evil* into the forensic mental health lexicon will be examined, as well as the potential costs of doing so.

The Illusion of "Evil"

Evil is an entirely subjective concept created by *Homo sapiens*. Evil cannot be detected or measured in nature or the universe. Primitive cultures believed that natural calamities were manifestations of evil. It was in this way that humanity first began to personify adverse circumstances or tragedy so that they could attempt to master attendant anxiety. Yet in all of evolutionary theory and natural selection, there is no designation for evil [3]. The relentless, brutal manner of natural selection may dispose us to a belief in so-called natural evil, while the reality is that this is nothing more than our own subjective interpretation and projections. Further, our own interpretations will invariably be ambiguous, culture-bound, and likely to evolve over time. The "evils" of the past are often today's sources of literary amusement.

The word "evil" has very ancient origins. It is "emotionally loaded, morally judgmental, full of brimstone and fire" [4]. When evil is used to label an individual, it has a strongly damning consequence. The word "evil" inescapably invokes religious and mythological mind-sets, which were responsible for it's origination [5]. It summons the supernatural, the mystic, and the esoteric. Labeling someone as evil suggests that they are "beyond redemption". Concluding that someone is evil also suggests that they are permanently beyond human understanding, a sentiment which is contrary to scientific principles. A question about the material universe may be currently beyond our understanding, but saying it is permanently outside of our grasp suggests the issue lies in the realm of the supernatural. Perhaps the most objective conclusion one could reach about "evil" is that it is a term associated with considerable linguistic ambiguity, and has variable meanings to different people [6].

At best, a term such as "evil" is mere subjective abstraction [7]. Further confounding the concept of evil is the conundrum: Evil from whose perspective? The victim's perspective? The perpetrator? Laypersons? All will be different, and all will simply consist of that individual's subjective conception of how evil is defined. Biases and distortions can be expected to flourish. While the victim's perspective is essential

for the *moral* evaluation of evil acts, it is ruinous for a causal understanding of them [8]. Ultimately, viewing evil as a distinct or quantifiable concept is an illusion. The real causes of violent or harmful behavior will always be different from the way people think of evil. This phenomenon has been called the *myth of pure evil*. According to the myth of pure evil, "the face of evil is no one's real face – it is always a false image that is imposed or projected on the opponent (p.62)" [8]. In contrast, what is not illusory is our history of feeling justified in committing atrocities against individuals who are labeled as "evil". Herein lies one of the strongest cautions against embracing a subjective moral concept as though it were a scientific concept.

To the best of our current and limited knowledge, what leads people to commit acts of intentional harm involves a complex interaction of biological, psychological, and social forces in concert with situational variables. One set of factors affects and is affected by the others, and cannot likely stand on its own. Behavioral science has made efforts to objectively study each of these factors, mostly in a reductionistic approach. In contrast, it is important to consider whether accepting an emotionally laden and morally judgmental term like evil will advance our understanding of deviant or violent behavior. We must also consider whether forensic mental health professionals will be able to remove the biased moral connotations of the term, particularly in the courtroom, so that ethical and objective testimony is proffered. This will be a difficult, if not impossible challenge, given the inherent predisposition of some courts, as noted by Gilligan:

> "There were times in the courtroom and prisons in which I did my work when I felt as though I had somehow been transported . . . back into the Middle Ages, when people still thought that evil (like its mythical embodiment and namesake, the devil) was an objective thing that actually existed independently of our subjective feelings and thoughts, rather than a word we all to often use to rationalize, justify, and conceal, from ourselves and others, our own violence toward those we hate and wish to punish." (p.14) [9]

Learning from History: The Witches Hammer

In biblical times, mental illness was often seen as evidence of demonic forces and thus, the opposite of what was "good". During the middle ages, progress in medical science was severely squelched by the Christian church. During the Renaissance period, an obsession with evil in the form of witches became prominent. The official "practice guidelines" on detecting witches, the *Malleus Maleficarum* (1486), assisted Inquisitors in finding evil lurking amidst women, the socially disenfranchised and the mentally ill [10].

The witch-hunting of the fifteenth and sixteenth centuries serves as a fascinating, sobering example of an "official recognition of a hitherto unknown form of deviance" [11]. Once the crime of witchcraft was officially recognized, serious problems developed in terms of providing "proof" and legal restraints to the hysteria. The powerful legal and religious emphasis on the reality of witchcraft helped to reinforce the legitimacy of the trials, in addition to the public's belief that there was "evil" afoot. It has been theorized that the English government's systematic efforts for dealing with witchcraft served as a form of repressive deviance-management. In addition, one of the benefits to church and state of the witch-hunting hysteria was that it effectively shifted public attention away from growing demands for more equitable redistribution of wealth [12].

In retrospect, "evil" (in form of witches) was nothing more than what the English legal system claimed that it was. Those who were found to be witches were often ill-equipped and powerless to fend off this creative label of deviance. Once the definition of witchcraft was officially accepted, very little could be done to prevent or limit the system's abuse of the term. As a result, large populations of "deviant" witches were discovered, particularly among vulnerable lower-class groups. This in turn, fostered the growth of an "industry" revolving around the detection, prosecution, and punishment of witches. The industry included the proliferation of "rackets", and entrepreneurs seeking to profit from its operation [13]. The development of a profit-making "deviance" industry was perpetuated in cyclic fashion. The more rigorous the detection efforts, the higher the rates of deviance appeared to be, which then justified the use of more extreme measures of detection. However, forces other than economic had vested interests in defining and controlling deviance. Political, religious, and psychological interests have also been cited as playing significant roles [13]. One of lessons from the witchcraft hysteria in England is that once a definition of "evil" has been officially

sanctioned, the potential for abuse becomes virtually unlimited.

While the example of witchcraft is an entirely invented form of deviance, it is the process of stigmatization and repressive control that merits present-day consideration. It requires little stretch of the imagination to consider how more modern notions of "evil" might be creatively imputed to those who are unable to ward off its powerful moralistic connotations. Indeed, it is hubris to conclude that we are beyond such societal dynamics today. Given the right setting and circumstances, a regressive return to a variety of analogous behaviors is distinctly within our repertoire of responses. Consider the example of present-day Russia. The unstable environment of post-Soviet society has been characterized by drastic social changes and societal insecurity. A therapist working in a boarding school for teens reported a startling return to the practices of various superstitions and witch persecution [14]. The witch persecution was described as providing a socially sanctioned outlet for repressed anger, anxiety, and frustration. Exposing a witch among their peers helped them to explain daily misfortunes, and reaffirmed the boundaries between good and bad parts of the group.

When considering the present-day implications of legitimizing evil as forensic science concept, it should be noted in the present day and age, over 40% of Americans believe in demons, devils, and other superstitious concepts [13]. Beliefs in "evil" as an objective force can be observed among many ordinary citizens. For example, individuals are quite ready to believe that Hitler's personality or aura of "evil" can spread into his sweater, causing them to refuse to wear it [15, 16]. This is but one example of the tendency toward magical thinking in which material objects come to be seen as symbolic representations. Indeed, the distinction for many between the laws of magical thinking and reality is dangerously ambiguous [17].

Forensic Interest in "Evil": The Recurrence of an Illusion

Over the past two decades, an explicit emphasis on evil has been developed by several respected social psychologists [9, 18–21]. However, even in these scientific contexts, the term is used inconsistently. More importantly, use of the term does not escape vagueness and biased connotations. Over the past several decades, there have been quite reasonable advances in the areas of neuroscience, psychology, and sociology that begin to address, in a scientific manner, the problem of violent and/or deviant behavior. Thus, one may be inclined to wonder why some forensic mental health professionals are "raising the spectre of some demonic force at work", despite its regressive implications [22]. Indeed, might it not be irresponsible, given the fixed connotations of the term, its implications of untreatability and, dare I say it – the need for extermination? Then why do we find ourselves conjuring evil and summoning demons?

Both the legal and forensic mental health literature inform us that we are in the midst of a "punitive era" of criminal justice [23, 24]. Rehabilitative efforts have been pruned away like so much dead wood. The number of incarcerated individuals in this country at the end of 2005 reached a record high at approximately 2.2 million [25], and there are no signs that this trend will reverse itself. Keeping the prevailing emphasis on punishment in mind, it is possible to discern *some* of the hypothetical "pressures" on forensic mental health professionals, *vis-à-vis* the justice system, to identify and root out "evil". It has been suggested that the United States may be in the midst of a "moral panic", where radical measures are seen as reasonable and reassuring options [26, 27]. This is of little surprise during a period in which the "politics of crime" have been driven by fear inducing appeals to "common sense" punitiveness [28]. Such appeals have the allure of reducing the complex to a simple battle between good and evil. Thus, anxieties about "moral relativism" are concretely allayed.

In times of trouble, societies tend to stiffen and enforce conformity. Typically, strenuous efforts are made to root out the elements of "sin" and "vice". Encouraged by leadership, society is transformed into metaphorical "Spartan fist", as it prepares to steel itself against threat or chaos [29]. Fear and anxiety further drive the contraction of societal attitudes, and a return to earlier, more familiar practices. In such times, if an illusion of a "handle" by which to control a problem is produced, it is often grasped with intensity and passion. At the base of such a handle, it is not uncommon to find the illusion of an "evil" enemy. When threats to safety and survival become a prominent feature in society, the attendant terror is managed by reinforcing well-worn cultural values. The uncertainty produced by existential anxiety is mastered by reifying a system (or individual) which promises

order and permanence [30]. Thus, the latent meaning underneath confronting the "problem of evil" is the fundamental "struggle to make meaning out of a universe that appears cruel and wicked" [31]. However, the gain of coping with existential anxiety in this way must be carefully balanced against the loss that comes with abandoning some measure of rationalism.

The Need to Keep "Evil" Obscured

Part of the drive to catalog and illustrate evil may be a growing awareness that its face is rather ordinary, and provides no prospective cues for recognition. From Eichmann to a long list of serial killers, the phrase "terrifyingly normal" is sometimes used. The absence of a "look" or a mark of Cain is highly disturbing to most. It momentarily turns a rational, ordered view of the world on its head. How can someone who appears normal commit such deeds? A journalist covering the Jeffrey Dahmer trial could not get over how ordinary Dahmer appeared, remarking that "there was nothing to him" [32]. It would seem that we need to give evil a face. If we can somehow learn to recognize it, we may be able to bring it under our control. Otherwise, we will continually be threatened by unknown forces cloaked in the guise of normality. "Moral monsters", unbranded by physical stigmata, will be able to commit atrocities undetected [33].

Giving evil a distinct face also gratifies us by putting distance between us and the "others", lending certainty to the idea that we are worlds apart in our differences. Thus, there is a "social virtue" to outlining the face of evil – society is exonerated and bears no responsibility. When a bright line separating good from evil has been identified, society may "take comfort in the illusion that such a line constrains crossovers in either direction" [34]. But the desire to know the "face of evil" is in *direct* conflict with our desire to keep it partially obscured. This is necessary for keeping it the repository of noxious projections. It has been commonly observed that "humanizing" a monster "makes him less compelling as the embodiment of evil" [35]. In other words, very clear view of evil's face undermines the effectiveness of projection. The "celebrity status" of serial killers in American culture reflects precisely such contradictory desires [36]. The public is repulsed by their deviance, yet attracted by the vicarious thrill and tension. It is the tension between the killer's outer normalcy and inner deviance that elicits the fascination. Finally, there is the inevitable, comforting denouement that the killer is an "alien" aberration, whose detection has made society a much safer place. This societal morality play requires "powerful" evildoers, as society appears uninterested in the "mundane", ubiquitous purveyors of violence [34]. Crimes must be truly deviant or monstrous for them to be easily disavowed.

The claim that "We do not want to look at evil" [37] is difficult to refute for obvious reasons. Further, the tendency to keep evil obscured is consistent with lay notions that evil is *beyond* comprehension, and that those who commit evil acts lie *outside* the demarcation of being human [38]. Thus, the psychological need to keep the "face of evil" at least partially obscured will present yet another impasse to any serious scientific attempts to "standardize" or quantify evil. While attempting to study human cruelty under the auspices of a morally judgmental term may ease the investigator's subjective discomfort, it will do little to permit unbiased observations.

"Evil" in the Courtroom

It is not difficult to imagine a scenario in which the results of a legal adjudication of "evil" include discrimination against poor or disadvantaged individuals. Strong emotional and psychological forces are at play, particularly during capital trials, which are potentially biasing. It is well-known that much more than legal fact is communicated in the courtroom, and that this "much more" has a direct and powerful effect on a jury's punishment decision. For example, a defendant's *appearance* significantly influenced whether jurors imposed the death sentence [39]. Thus, one may be inclined to wonder: If jurors are unable to discount the physical appearance of a defendant in their deliberations, what is the likelihood that they will remain objective when a word steeped in religious morality is introduced by "experts" as an allegedly "scientific" construct?

When testifying about forensic mental health issues in court, it is not uncommon for the expert to encounter issues of moral conflict. A very helpful guideline for dealing with such a dilemma is to make use of the "touchstone" of asking oneself "what the ideal forensic *pathologist* would do in a similar situation" [40]. This ultimately has the effect of focusing one's testimony on the technical matters of forensic mental health. At the same time, it steers one clear of biased inferences or impressions. When the expert

strays too far from the touchstone of forensic scientist, she delves even deeper into the problem of the fact–value distinction [41]. It is difficult to argue that mental health professionals are not constantly making value judgments, either implicitly or explicitly. However, as an expert witness one must take care not to move away from clinical science, and toward illusory moral concepts, where the line between fact and value becomes hopelessly blurred.

Here it must be acknowledged that even certain clinical terms have, in fact, come to be associated with moral judgments. Diagnoses such as psychopath, personality disorder, and conduct disorder may be used by some as more of a moral judgment than a clinical diagnosis [42]. Rather than take this as proof that such practice is acceptable, forensic mental health professionals should remain cautious about the ease with which we inject our moral judgments into situations where they are not called for. It is critical that attempts via forensic mental health to understand such behavior be accompanied by a vigilant suppression of moral judgment. Achieving a state of absolute moral neutrality may indeed be unrealistic; nevertheless, "the *ideal* of being value-free is important because values prevent one from seeing the facts (p.386)" [9].

"Evil" Unnecessary

Forensic mental health already has working concepts describing deviant behavior. Certainly, these concepts and related nomenclature are not without their imperfections. However, none of them have blatantly recognizable origins in notions of religious morality. Efforts to understand violent human behavior have already provided us with steps toward an initial objective approach. Forensic mental health theories supported by research and case studies abound on topics such as malignant narcissism, psychopathy, sexual sadism, and serial sexual homicide [43–47]. From the biological perspective, although deviant or violent behavior is not well understood, genetic and neurochemical studies are beginning to proliferate [48, 49]. Studies on the neurobiology of maternal and pair bonding are being used to understand the origins of human violence [50]. Although still in the very early stages, neuroimaging studies have suggested dysfunction in certain neural systems of violent and psychopathic individuals [40, 51]. Tangible neuropsychiatric deficits have been found among serial and sexual murderers, individuals considered by many to represent the extreme of human cruelty and narcissism [52–54].

Thus, forensic mental health already has a tradition of at least attempting to avoid moralistic bias by focusing on concepts such as violence, aggression, or sexual disorders. Terms with value-laden or pejorative connotations are either limited or avoided. This is a tradition that places value on the struggle for neutrality and objectivity. One potential result of using terms that have a tendency to shut down objective thought is that the complexity and sophistication of scientific dialogue may be reduced. Additionally, sensationalized terms may serve to distract from underlying problems that are too difficult or complicated to acknowledge. Many violent crimes involve significant social issues that cannot be reduced to a dichotomous equation of "good vs. evil". Providing a finding of "evil" or "not evil", reduces the likelihood that more complex issues will receive consideration.

The Cost of Legitimizing an Illusion

> "When the natural consequences of a deed are no longer "natural," but thought of as caused by the conceptual specters of superstition ... then the presupposition of knowledge has been destroyed – *then the greatest crime against humanity has been committed*."[55]
>
> - Friedrich Nietzsche

Substantial skepticism already exists in the courts about the reliability of mental disease evidence [56]. Doubts about the objective reality of mental illnesses are less likely to be resolved by the courts, should forensic mental health pursue scientific answers under the shadow of a religious concept. There would also be the added risk that forensic expert witnesses might veer further into a moral and quasi-religious morass. It is the path of least resistance to find the offensive party "evil", and thus beyond the need for further understanding. The matter is settled, and no one need dispute or question the presence of evilness. As one nationally known criminal justice expert put it, "Some people are just evil" [49].

In essence, a finding of "evil" is an expedient substitute for thought [57]. But the business of dismissing someone as evil is really too easy, and "merely begs the question of how they became that way" in the first place [42]. Lay concepts of evil often fuse with the professional ethics of

mental illness, and threaten to confound each other's ideologies [58]. There remains a deeply ingrained societal prejudice that persons with mental illness are "ticking time bombs, ready to explode into violence" [59]. A significant portion of the lay public requires little persuasion to associate mental illness with mystical and potentially evil forces. Arguably, this association is strengthened by Hollywood movie portrayals of mental illness (e.g., *The Exorcist*), which perpetuate the myth that evil and mental illness are overlapping, related phenomena [60]. Thus, the primitive association between mental disorder and moral depravity has yet to be completely dissolved. The archaic myth that it is ultimately evil which gives birth to mental disease lingers in the shadows and waits to be resurrected.

In addition to the potential for further stigmatization of mental illness, legitimizing evil as a forensic mental health concept will have serious detrimental effects on forensic clinical practice. Sobering evidence of this comes from a study of forensic psychiatric nurses' approaches to treatment on a high-security psychiatric hospital in the United Kingdom [48]. Nurses were given a series of vignettes describing themes such as child killing, serial rape, and interpersonal violence. The nurses' discourse in semistructured interviews was analyzed, and compared to data collected from actual care plans of forensic patients on the nurses' wards. When a patient was judged to be "evil", staff abandoned medical discourse and reverted lay notions of badness.

Further, "evil" patients were viewed as being beyond help, which was then reflected in their care plans. The authors stressed that forensic psychiatric services may be actively limited in the face of socialized values and lay concepts of evil. The semantics of lay discourse proved to be of more than academic interest, particularly where forensic terminology is translated into practice and treatment. Those forensic patients who were labeled as evil by nursing staff were, in effect, excluded from the usual medical, symptom-centered approach [61]. In other words, *a finding of "evil" caused a withdrawal of treatment.* The implications of using the term *evil* as a form of punitive sanctions by the staff are being considered in future research. It is this worrisome effect – the withdrawal of treatment efforts – that should be among the strongest objections to the use of the term *evil* in forensic mental health. Given the implications of these early studies on forensic treatment efforts, it may be difficult to reconcile the use of the term *evil* with Hippocratic ethics or the medical ideal of *primum non nocere.*

Withdrawal of treatment is not the only possible outcome of declaring someone evil. Consider a scenario in which a defendant has been "scientifically" determined to be "evil". If it is a capital trial, he will either be given the death penalty (presumably if rated "high" in evilness) or life imprisonment (perhaps if rated as only moderately evil). Already, the specter of the witches' hammer has been resurrected – an essentially moral judgment has been cloaked in science in order to put someone to death. While in prison, on or off death row, the label of "evil" will doubtlessly stick. In a correctional setting, inmates and staff are only too aware of charges and findings in disturbing cases. Thus, it is not difficult to imagine a scenario in which "evil" inmates will be targeted for a special brand of hysteria-induced harassment and denial of services. To doubt that this could occur is to be unaware of the way in which sex offenders are severely harassed in prisons today. If the humanitarian aspects of this scenario are insufficient, there is also the troubling adverse economic outcome to consider. When the "evil" inmate is no longer able to tolerate steady doses of harassment and other forms of sadistic projection, he will ultimately seek escape through suicide. Some of these cases will result in costly litigation for corrections, and ultimately for taxpayers.

Views of evil as the adversary of good work well for increasing hostilities between rival nations, ethnic groups, and other social units [2]. Having strong beliefs in an active and tangible evil (i.e., "Satan") is significantly correlated with intolerance toward others [62], thus setting the stage for aggressive tensions. In addition, depicting an enemy as evil helps to foster an obligation to oppose and dispose of him. Because he is evil, there is little need to concern oneself with his health, welfare, or gaining a better understanding of him. All of this can be done free of guilt, for those who are "evil" bring about their own "just desserts". Perhaps the most self-satisfying achievement of perpetuating the myth of evil is that it allows us to be reassured of our own "goodness". It confers a moral superiority that is itself a setup for perpetrating all manner of atrocities. Most importantly, none of this does anything appreciable to stop people from committing terrible violence against others.

Finally, another untoward effect of legitimizing, and perhaps sensationalizing evil is the risk that it will present an attractive identity to those who feel disenfranchised by society. Poor social bonding has been found to increase the chances that adolescents will bond with deviant peers and adopt similar identities [63]. Vulnerable adolescents who have endured abusive developmental experiences have been observed to identify themselves as "evil", and search for identity in satanic cults [64]. Should evil be legitimized, particularly with the assistance of media sensationalism, there is the chance that it will serve as an orienting influence for the generally dissatisfied.

Conclusions

"Evil" can never be scientifically defined because (i) it is an illusory moral concept, (ii) it does not exist in nature, and (iii) its origins and connotations are inextricably linked to religion and mythology. The term *evil* will be very unlikely to escape religious and unscientific biases that reach back over the millennia. Any attempt to study violent or deviant behavior under this rubric will be fraught with bias and moralistic judgments. Embracing the term *evil* as though it were a legitimate scientific concept will contribute to the stigma of mental illness, diminish the credibility of the forensic clinicians, and corrupt forensic treatment efforts.

Preoccupation with evil has always been a part of human affairs. Indeed, the notion of "good vs. evil" has served as a catalyst for many positive human achievements. However, the field of forensic mental health should not succumb to a manner of thought that is regressive and distinctly unscientific. While further defining evil within the fields of philosophy or theology may provide greater precision for discourse on ethics, this purpose does not proclaim itself evidenced-based science or suitable for expert testimony. The fact that no philosophical, religious, or ethical debate over the past several thousand years has conclusively "defined" evil must inform us of something. Well-designed studies which help us to understand, and ultimately prevent violent, harmful, and deviant behaviors are sorely needed. In contrast, efforts to legitimize moral judgments by forensic scientists who should strive for moral neutrality should be avoided. Should forensic mental health accept the view of evil as an objective, quantifiable concept, a sacrifice of rational thought for expediency will have been made. The price of this sacrifice is likely to involve an insidious creep toward shutting down scientific thought and toward illusory moral judgments.

References

[1] Simon, R. (2003). Should forensic psychiatrists testify about evil? *The Journal of the American Academy of Psychiatry and Law* **31**(4), 413–416.

[2] Welner, M. (2003). Response to Simon: legal relevance demands that evil be defined and standardized, *The Journal of the American Academy of Psychiatry and Law* **31**(4), 417–421.

[3] Duntley, J. & Buss, D. (2004). The evolution of evil, in *The Social Psychology of Good and Evil*, A. Miller, ed, The Guilford Press, New York.

[4] Tangney, J. & Stuewig, J. (2004). A moral-emotional perspective on evil persons and evil deeds, in *The Social Psychology of Good and Evil*, A. Miller, ed, The Guilford Press, New York.

[5] Carus, P. (1996). *The History of the Devil and the Idea of Evil*, Gramercy, New York.

[6] Alford, C. (1997). The political psychology of evil, *Political Psychology* **18**, 1–17.

[7] Beck, A. (1997). Foreword, in *Evil: Inside Human Violence and Cruelty*, R. Baumeister, Henry Holt and Company, New York.

[8] Baumeister, R. (1997). *Evil: Inside Human Violence and Cruelty*, Henry Holt and Company, New York.

[9] Gilligan, J. (2001). *Preventing Violence*, Thames & Hudson, New York.

[10] Kramer, H. & Sprenger, J. (1971). *The Malleus Maleficarum*, Translated by M. Summers, ed, Dover Publications, New York.

[11] Currie, E. (1968). Crimes without criminals: witchcraft and its control in renaissance Europe, *Law and Society Review*, **3**, 7–32.

[12] Harris, M. (1974). *Cows, Pigs, Wars and Witches*, Random House, New York.

[13] Rice, T. (2003). Believe it or not: religious and other paranormal beliefs in the United States, *Journal for the Scientific Study of Religion* **42**, 95–106.

[14] Tartakovsky, E. (2001). Group psychodynamics of witchcraft and witch-hunting: immigrant adolescents from the former Soviet Union in Israel, *Group Analysis* **34**(1), 129–142.

[15] Nemeroff, C. (1995). Magical thinking about illness virulence: conception of germs from "safe" versus "dangerous" others, *Health Psychology* **14**, 147–151.

[16] Rozin, P., Millman, L. & Nemeroff, C. (1986). Operation of the laws of sympathetic magic in disgust and other domains, *Journal of Personality and Social Psychology* **50**, 703–712.

[17] Lindeman, M. & Aarinio, K. (2007). Superstitious, magical, and paranormal beliefs: an integrative model, *Journal of Research in Personality* **41**, 731–744.

[18] Miller, A. (1999). Harming other people: perspectives on evil and violence, *Personality and Social Psychology Review* **3**, 176–178.
[19] Darley, J. (1992). Social organization for the production of evil, *Psychological Inquiry* **3**, 199–218.
[20] Miller, A., Gordon, A. & Buddie, M. (1999). Accounting for evil and cruelty: is to explain to condone? *Personality and Social Psychology Review* **3**, 254–268.
[21] Wilson, P. (2003). The concept of evil and the forensic psychologist, *International Journal of Forensic Psychology* **1**(1), 1–9.
[22] Cooklin, R. (1995). Letter to the Editor re: psychiatry and the concept of evil, *The British Journal of Psychiatry* **166**, 114–115.
[23] Roberts, J. (1997). *Reform and Retribution: An Illustrated History of American Prisons*, American Correctional Association, Lanham.
[24] Kupers, T. (1999). *Prison Madness: The Mental Health Crisis Behind Bars and What We Must Do About It*, Josey-Bass, San Francisco.
[25] Harrison, P. & Beck, A. (2005). *Prisoners in 2005: Bureau of Justice Statistic Bulletin*, U.S. Department of Justice, November 2005. http://www.ojp.usdoj.gov/bjs/pub/pdf/p05.pdf.
[26] Appelbaum, P. (2003). Dangerous persons, moral panic, and the uses of psychiatry, *Psychiatric Services* **54**(4), 441–442.
[27] Grieg, D. (2002). *Neither Bad Nor Mad: The Competing Discourse of Psychiatry, Law, and Politics*, Kingsley, London.
[28] Kennedy, J. (2006). Facing evil, *Michigan Law Review* **104**, 1287–1304.
[29] Bloom, H. (2000). *Global Brain: The Evolution of Mass Mind From the Big Bang to the 21st Century*, John Wiley & Sons, New York.
[30] Landau, M., Landau, M.J., Solomon, S., Greenberg, J., Cohen, F., Pyszczynski, T., Arndt, J., Miller, C.H., Ogilvie, D.M. & Cook, A. (2004). Deliver us from evil: the effects of mortality salience and reminders of 9/11 on support for President George W. Bush, *Personality and Social Psychology Bulletin* **30**(9), 1136–1150.
[31] Bloom, S. (1996). Trauma and the nature of evil. *12th Annual Meeting of the International Society for Traumatic Stress Studies*, November 1996. http://www.sanctuaryweb.com/Documents/Trauma%20and%20the%20Nature%20of%20Evil.pdf.
[32] Schwartz, A. (1993). *The Man Who Could Not Kill Enough*, Birch Lane, New York.
[33] Douard, J. (2007). Loathing the sinner, medicalizing the sin: why sexually violent predator statutes are unjust, *International Journal of Law and Psychiatry* **30**, 36–48.
[34] Zimbardo, P. (2004). A situationist perspective on the psychology of evil: understanding how good people are transformed into perpetrators, in *The Social Psychology of Good and Evil*, A. Miller, ed, The Guilford Press, New York, p. 25.
[35] Gabbard, G. (1999). Book review of: Hannibal, *The American Journal of Psychiatry* **156**(11), 2001.
[36] Schmid, D. (2005). *Natural Born Celebrities: Serial Killers in American Culture*, University of Chicago Press, Chicago.
[37] Tucker, N. (2007). *Giving Evil the Eye: Juries Don't Always Know Heinous Crimes When They See Them, But This Might Help*, Washington Post.
[38] Mason, T., Richman, J. & Mercer, D. (2002). The influence of evil on forensic clinical practice, *International Journal of Mental Health Nursing* **11**, 80–93.
[39] Antonio, M. (2006). Arbitrariness and the death penalty: how the defendant's appearance during trial influences capital jurors' punishment decision, *Behavioral Sciences and the Law* **24**, 215–234.
[40] Hoptman, M. (2003). Neuroimaging studies of violence and antisocial behavior, *Journal of Psychiatric Practice* **9**(4), 265–278.
[41] Stone, A. (1984). The ethical boundaries of forensic psychiatry: a view from the Ivory Tower, *The Bulletin of the American Academy of Psychiatry and the Law* **12**(3), 209–219.
[42] Woolley, E. & Muncey, T. (2004). Demons or diamonds: a study to ascertain health professionals' attitudes toward children with conduct disorder, *Journal of Child and Adolescent Psychiatric Nursing* **17**(4), 151–160.
[43] Turco, R. (2001). Child serial murder-psychodynamics: closely watched shadows, *The Journal of the American Academy of Psychoanalysis* **29**(2), 331–338.
[44] Hare, R. (2006). Psychopathy: a clinical and forensic overview, *The Psychiatric Clinics of North America* **29**(3), 709–724.
[45] Stone, M. (2001). Serial sexual homicide: biological, psychological, and sociological aspects, *Journal of Personality Disorders* **15**(1), 1–18.
[46] Myers, W.C., Husted, D.S., Safarik, M.E. & O'Toole, M.E. (2006). The motivation behind serial sexual homicide: is it sex, power, and control, or anger? *Journal of Forensic Sciences* **51**(4), 900–907.
[47] Hill, A., Habermann, N., Berner, W. & Briken, P. (2006). Sexual sadism and sadistic personality disorder in sexual homicide, *Journal of Personality Disorders* **20**(6), 671–684.
[48] Vovlaka, J. (1999). The neurobiology of violence: an update, *The Journal of Neuropsychiatry and Clinical Neurosciences* **11**(3), 307–314.
[49] Meyer-Lindenberg, A., Buckholtz, J.W., Kolachana, B., Hariri, A.R., Pezawas, L., Blasi, G., Wabnitz, A., Honea, R., Verchinski, B., Callicott, J.H., Egan, M., Mattay, V. & Weinberger, D.R. (2006). Neural mechanisms of genetic risk for impulsivity and violence in humans, *Proceedings of the National Academy of Sciences of the United States of America* **103**(16), 6269–6274.
[50] Pedersen, C. (2004). Biological aspects of social bonding and the roots of human violence, *Annals of the New York Academy of Sciences* **1036**, 106–127.
[51] Raine, A., Lencz, T., Taylor, K., Hellige, J.B., Bihrle, S., Lacasse, L., Lee, M., Ishikawa, S. & Colletti, P. (2003).

Corpus callosum abnormalities in psychopathic antisocial individuals, *Archives of General Psychiatry* **60**(11), 1134–1142.
[52] Silva, J., Ferrari, M. & Leong, G. (2002). The case of Jeffrey Dahmer: sexual homicide from a neuropsychiatric developmental perspective, *Journal of Forensic Sciences* **47**(6), 1347–1359.
[53] Silva, J., Leong, G. & Ferrari, M. (2004). A neuropsychiatric developmental model of serial homicidal behavior, *Behavioral Sciences and the Law* **22**(6), 787–799.
[54] Briken, P., Habermann, N., Wolfgang, B. & Hill, A. (2006). XYY chromosome abnormality in sexual homicide perpetrators, *American Journal of Medical Genetics. Part B (Neuropsychiatric Genetics)* **141B**, 198–200.
[55] Neitchze, F. (1982). *The Portable Neitchze*, W. Kaufmann, ed, Penguin Books, New York.
[56] Appelbaum, P. (2006). Insanity, guilty minds, and psychiatric testimony, *Psychiatric Services* **57**(10), 1370–1372.
[57] Masters, B. (1997). *The Evil That Men Do: From Saints to Serial Killers*, Black Swan, London.
[58] Billig, M., Condor, S., Edwards, D., Gane, M., Meddleton, D. & Radley, A. (1988). *Ideological Dilemmas: A Social Psychology of Everyday Thinking*, Sage, London.
[59] Appelbaum, P. (2004). One madman keeping loaded guns: misconceptions of mental illness and their legal consequences, *Psychiatric Services* **55**(10), 1105–1106.
[60] Butler, J. & Hyler, S. (2005). Hollywood portrayals of child and adolescent mental health treatment: implications for clinical practice, *Child and Adolescent Psychiatric Clinics of North America* **14**, 509–522.
[61] Mercer, D., Mason, T. & Richman, J. (1999). Good & evil in the crusade of care: social constructions of mental disorders, *Journal of Psychosocial Nursing* **37**(9), 13–17.
[62] Wilson, K. & Huff, J. (2001). Scaling Satan, *The Journal of Psychology* **135**(3), 292–300.
[63] Oetting, E. & Donnermeyer, J. (1998). Primary socialization theory: the etiology of drug use and deviance. I, *Substance Use and Misuse* **33**(4), 995–1026.
[64] Belitz, J. & Schacht, A. (1992). Satanism as a response to abuse: the dynamics and treatment of satanic involvement in male youths, *Adolescence* **27**(108), 855–872.

Related Articles

Death Penalty and Age

Expert Opinion: United States

Ethics: Codes of Conduct for Expert Witnesses

Jury Dynamics

JAMES L. KNOLL, IV AND PHILLIP J. RESNICK

Examination of Fibers and Textiles

Introduction

Edmond Locard had already postulated in the early twentieth century that an exchange of fragmentary evidence takes place at the scene when two or more items come in contact [1]. Looking at textiles in more detail helps to confirm this statement from the point of view of fibers. Textiles and clothes are constructed of tiny fibers. These fibers and their fragments readily break free from the textile surface and are left behind making contact with another textile or a different item. On the basis of this principle [2–4], fibers from scenes can be recovered and compared with target garments to identify possible links between scenes or between persons. To probe the role of fibers in trace evidence, this article reviews the following:

recovery of fiber evidence including issues related to contamination;
fiber/textile analysis;
textile intelligence work.

Recovery of Fiber Evidence

Textile fibers are mostly tiny particles that are invisible to the unaided eye. Therefore special precautions must be taken to recover the fibers.

At the time of trace recovery at the scene, it is not always possible to predict the type of evidence that is relevant to that specific case. The strategy of how to progress with the case depends on the surrounding information; therefore, it is vital to get as much information as possible. Only in this way is it possible to get a fairly good idea of where the case-relevant fibers can be expected to be present. The taping of an item or body is an appropriate means to recover textile fibers or possibly other micro traces. Altering objects or moving the objects or the victim before taping should be avoided as far as possible. This is due to the type of evidence involved, since all micro traces, including fibers, are extremely sensitive to movement. Items collected at the scene for trace recovery in the laboratory should be handled to the minimal extent possible to prevent loss or contamination. They also need to be packaged and

sealed in a way that prevents contamination or loss and minimizes alterations. The material needs to be labeled in such a way that it is absolutely clear by whom, when, and where exactly the material was recovered. If movement or any other changes cannot be avoided, it is necessary to document these in detail.

To prevent contamination in the laboratory, it is important to recover traces from victims, offenders, or other people involved, using different rooms and via different personnel. The collection of reference samples from the immediate surrounding of the scene, from living areas, and other surroundings of victims and offenders, and the clothes involved, are as important as the trace recovery itself.

Techniques of Recovery

Picking-off fibers from a surface or a sharp edge is a suitable method for visible textile traces (e.g., fiber tufts). It is also a suitable means to recover fibers under a low-power microscope from smaller surfaces where other traces are present (e.g., blood or fingerprints) and where a tape lifting might destroy other evidence. *Tape lifting* is a highly recommended technique because it involves a detailed and overall collection of the fiber population on a surface [5]. *Additive tape lifting* is a procedure in which a defined area or zone is taped with the multiple use of only one tape. It represents, to a certain extent, the fiber distribution in a defined area if the chosen area for the multiple tape lifting is relatively small. Another highly specialised technique for an exhaustive collection of fibers and also other trace evidence is *1 : 1 taping*. The technique offers more comprehensive information since one tape is used only once, and recovered fibers can be associated with a defined spot on a surface; this allows a detailed mapping of fibers. It is mainly used in high profile cases (e.g., murder), whereby information about density, position, and distribution of certain fiber types can be obtained [6].

For all tape lifting techniques, it is essential to choose a clear tape with glue that will not affect the fibers over time.

There are other different techniques available like combing [7], vacuuming, and scraping, which are just briefly mentioned here. These techniques are suitable only for certain aspects of fiber recovery. The type of fibre recovery to be used has to be decided individually and depends on the situation at the scene. Special situations might require creative techniques for fiber recovery.

Textile Fibers

The analysis of fibers and fiber fragments requires special measures to adapt routine industrial methods for forensic analytical purposes.

Single textile fibers are the microscopic building materials of clothing and other textile products. Huge numbers of these tiny fibers are carefully constructed to form a thread or fabric. In many different ways, they are "built in" and therefore show a number of different characteristics. Single fibers can occur as staple fibers (fibers of a more or less defined length) or filaments (fibers of indefinite length). Besides morphological features, the most important characteristic of fibers is color. During production and during wear and tear, a textile and its fibers are exposed to different extraneous influences. These influences may alter, damage, and fragment the fibers and its color, and influence the shedding abilities of the textile. In combination with the intensity of contact and the donor qualities of a garment, this dictates the number, distribution, and appearance of fibers on the surface of contact.

Classification of Textile Fibers

Fiber classification starts with two basic groups – *natural* and *manmade fibers*. The first group comes about in nature and refers to the origin of the material. The second group does not occur naturally and contains only manufactured fibers. The term *classification* refers to the production process.

There are numerous classification systems available, which differ slightly from each other [8, 9]. For fiber identification purposes in forensic applications, generic names and their subtypes should be used.

For the forensic analyst, it might be essential to identify even the generic subtype of a fiber in order to verify a match or to gain further information about production and use.

Generic names and codes are used according to regulations of the International Organization for Standardization (ISO), the German Institute for Standardization Deutsches Institut für Normung (DIN) and The International Bureau for the Standardization of Man-Made Fibres (BISFA).

Care must be taken not to mix up generic names and brand names of fibers. The latter only refer to fiber products of certain companies.

Natural Fibers. Natural fibers are split into three main groups (animal, vegetable, and mineral), which are divided into subgroups according to the origin of the fiber material [10, 11]. Natural fibers are listed below in Table 1.

Man-made Fibers. Man-made fibers are split into two groups – organic and inorganic fibers, which are divided into subgroups depending on the origin [12, 13].

Table 1 Natural fibers

Type	Group -Generic name examples	Code
Animal	silk	
	-mulberry silk	SE
	-tussah silk	TS
	wool	
	-wool	WO
	hair from sheep only, different breeds	
	hair	
	-angora	WA
	-alpaca	WP
	-camel	WK
	-cashmere	WS
	-lama	WL
	-mohair	WM
	-vicuna	WG
Vegetable	seed	
	-coir	CC
	-cotton	CO
	-kapok	KP
	bast	
	-flax	LI
	-hemp	HA
	-jute	JU
	-kenaf	KE
	-ramie	RA
	leaf	
	-abaca	AB
	-henequen	HE
	-phormium	NF
	-sisal	SI
Mineral	mineral	
	-asbestos	AS

According to ISO 6938 and Din 60001 -1

Organic fibers are either produced by transformation of natural polymers or from synthetic polymers. Inorganic fibers are produced by a transformation of state (Table 2).

Table 2 Man-made fibers

Type	Group -Generic name examples -subtype examples	Code
Organic from natural polymers	alginate	
	-alginate	AL
	cellulose ester	
	-acetate	CA
	-triacetate	CTA
	cellulose, regenerated	
	-cupro	CUP
	-lyocell	CLY
	-modal	CM
	-viscose	CV
	elastodiene	
	-rubber	ED
	protein, regenerated	PR
	-kasein animal	
	-arachin vegetable	
	-soy vegetable	
	-zein vegetable	
Organic from synthetic polymers	aramid	
	-aramid	AR
	polyamide	
	-polyamide	PA
	-polyamid 6	PA 6
	-polyamid 6.6	PA 6.6
	-polyamid 6.10	PA 6.10
	-polyamid 6.11	PA 6.11
	-polyamid 6.12	PA 6.12
	polyester	
	-polyester	PES
	-polyethylene terephthalate	PET
	-polybutylene terephthalate	PBT
	-polytrimethylene terephthalate	PTT
	polyolefin	
	-polyethylene	PE
	-polypropylene	PP
	polyurethane	
	-polyurthan non segmented PU	PU

continued overleaf

Table 2 (*continued*)

Type	Group -Generic name examples -subtype examples	Code
	-elastane segmented PU	EL
	polyvinyl derivates	
	-acrylic	PAN
	-acrylic/ methylacrylate	PAN/MA
	-acrylic/methy- metacrylate	PAN/MMA
	-acrylic/vinyl- acetate	PAN/VA
	-chlorofiber	CLF PTFE
	-fluorofiber	MAC
	-modacrylic	PVAL
	-vinylal	
Inorganic	carbon	
	-carbon	CF
	ceramic	
	-ceramic	CEF
	glass	
	-glass	GF
	metal	
	-metal	MTF

According to ISO 2076 and BISFA

High-performance fibers with added values have surfaced in the market and more will show up in the near future. "Smart Clothes" and "Intelligent Textiles" are catch phrases but these textiles are highly specialized products that are a challenge for the forensic textile expert. The identification of nanoparticles and nanocomposites, especially, is a task for the forensic scientist, since these particles are at the limit of perception in brightfield microscopy.

New developments in this area need more refined techniques to identify these specialized textiles and fiber types and also offer possibilities to enhance textile intelligence work. The question for the near future is how to develop refined procedures to identify these specialized textiles and fiber types in cases where routine examinations are not conclusive.

Fiber Analysis

The methods of fiber analysis are divided into microscopy, analysis by other instrumentation, which is very often carried out with the additional help of a microscope, and microchemical techniques for analysis.

The norm, as with other trace evidence types, is to characterise the donor fiber type and then use the most discriminating techniques to check whether the recovered fibers match. The microscope will always be the first tool as the level of discrimination is very high.

The range of tests may be limited by the size of the recovered fibers. Most instrumental techniques in common use in fiber examination are discussed below.

Microscopy

Microscopy is the basis for all further examination of fibers. It is a highly discriminative technique with regard to the morphological features of fibers. The microscopy of textile fibers allows identification and discrimination of the physical characteristics (morphological features) of natural and manufactured fibers. It also allows discrimination of the different optical characteristics.

The discriminative power of microscopy makes it the starting point of every case of fiber examination [14, 15].

Essential for all microscopic investigation is the correct set up of the microscope.

The adjustment of all brightfield microscopes for Köhler illumination should be a basic routine. Calibration and performance checks need to be carried out on a regular basis for the different microscopic systems in use.

Generally, the longitudinal view of fibers or fiber fragments is observed, but sometimes the additional examination of cross-sectional shapes is required to gather further information (e.g., regarding fibers structures, delustrants, and pigment particles).

Low-Power Microscopy (Reflected Light Stereomicroscopy). In general, fiber investigation starts with the examination of tape lifts or smaller items under a low-power microscope (*see* **Microscopy: Low Power**; **Microscopy: Light Microscopes**). The length, outer shape, and general behavior, which partly are due to the production process, can be detected. Color and color distribution already play an important role in detecting differences or similarities. Fibers showing notable differences from the known sample will be eliminated from further investigation

at this stage. Visually similar fibers need to be marked on the tape lifts. The position of a specific fiber on a tape lift is thus documented. For further detailed examination, the marked fibers need to be removed from the tapes. The removal of fibers is carried out with the help of a low-power microscope to prevent loss.

Subsequently, fibers should be mounted on microscope slides. The choice of mounting media differs from laboratory to laboratory, but it should certainly meet the requirements that known and recovered fibers are mounted in the same media; fibers should be so mounted that they can be removed from the slides without any damaging effects.

Brightfield Microscopy. The microscopic examination of textile fibers provides a nondestructive means to describe the morphological features and helps to discriminate fibers. Natural fibers can be distinguished on the basis of their characteristic features. Manmade fibers offer also a wide range of different morphological features (e.g., diameter, cross-sectional shape, delustrant particles, channels or voids, inclusions, and pigments) to help in distinguishing them. Besides morphological features, the type of coloration (e.g., print and type of print, spin dye, fiber dye, yarn or fabric dye), can be determined using brightfield microscopy.

Some information may already have been obtained under a low-power microscope but further information can be obtained only with the help of a brightfield microscope (*see* **Microscopy: High Power**; **Microscopy: Light Microscopes**).

Polarized Light Microscopy. Examination between crossed polars can prove helpful in identifying the fiber type. With natural fibers, polarized light is a helpful means to identify mercerized cotton or wool. Polarized light also helps to perceive morphological features more clearly (e.g., crossover marks in silk) under certain circumstances. Otherwise, polarized light does not play a dominant role in the identification of natural fibers. Regarding the nature of synthetic fibers, the alleged polymer type can be detected via optical characteristics by using polarized light [16]. An authentic collection of fiber types is useful to differentiate the material. Otherwise, a table showing the behavior of most generic types should be consulted. For confirmation and further detailed classification into fiber subtypes, other methods need to be employed (*see* **Microscopy: Light Microscopes**).

Polarized light microscopy also offers the possibility to observe whether dichroism is present in a fiber; this is observed when the fiber is orientated at different angles.

In combination with a hot stage, polarized light is also a helpful means to determine softening range and melting point of a fiber more easily for further classification [17].

Fluorescence Microscopy. Fibers fluoresce when excited by light of a shorter wavelength and emit energy usually at a longer wavelength. Therefore combinations of excitation and barrier filters are used for fluorescence microscopy [18, 19]. Generally, a range of broadband excitation filters is needed to cover the UV, blue and green ranges. A good light source is required to produce sufficient results (*see* **Microscopy: Light Microscopes**).

Besides the presence or absence of fluorescence, the color and intensity should also be noted.

Variations may occur already within the known material. Therefore it is essential to choose a wide range of fiber material from a potential source to make sure that all the variations are detected.

Fluorescence may arise from the fibers themselves or from additives (e.g., dye, washing powder, and optical brighteners).

Microspectrophotometers (MSP), Diode Array Detectors (DAD), or Charged Coupled Devices (CCD) can be equipped to measure emission fluorescence spectra (see below). Spectra might also prove helpful if needed.

Scanning Electron Microscopy/Energy Dispersive X-Ray (SEM/EDX). A very fine beam of electrons is focused on a fiber sample. From the interaction of electrons with the molecular composition of the fibers, a pseudo three-dimensional image is produced; this represents the unique elements present in the fiber. Magnification up to 250 000 times is possible [20, 21] (*see* **Microscopy: Scanning Electron Microscopy**).

Scanning electron microscopy (SEM) is a helpful tool for further morphological characterisation of fiber surfaces. It is a useful means for the detail examination of wool or hair scales. Besides the detailed examination of fiber characteristics, it provides the possibility to detect debris on the surface. Furthermore, the fiber edges may show specific characteristics originating from cuts (knife, scissor) or tears. The

easy characterisation of fiber cross sections is also a possible benefit of this technique [22].

Scanning electron microscopy/energy dispersive X-ray (SEM/EDX) may be used for the identification of pigments and delustrants or other additives used as flame retardants (e.g., halogens, antimony, and phosphorus). Moreover, it is possible to apply this method for the identification of other additives used in production, chemical residues, or debris on the surface of the fiber.

Other Microscopic Techniques. Besides the microscopic basic methods described, other techniques may be of further help and are used for fiber discrimination mostly by determination of the refractive index. Birefringence, interference microscopy, and dispersion staining offer a range of possible examination techniques that might add to the discriminative power in fiber examination, but these are not discussed in further detail since they are not routine techniques in every laboratory. The use of different methods highly depends on the specific equipment of each laboratory and its availability on the market.

Analysis by Other Instrumentation

UV/VIS Spectroscopy. Color is one of the most important characteristics to distinguish fibers. MSP, DAD- and CCD-spectroscopy offer highly discriminating methods for comparison and analysis [23].

Dyes are conjugated systems of excitable electrons and both principles explained beneath offer a method of measuring the interaction of the radiation with the fiber sample [24]. The region used for fiber spectroscopy is the UV/VIS (ultraviolet/visible) region. Usually a range from 240–760 nm is covered. If spectroscopy is restricted to the visible range, differences in dye components may be undetected. For UV/VIS detection, fibers need to be mounted on a quartz slide with a fluorescent-free media.

Generally, there are two different methods for UV/VIS spectroscopy available.

The first one is used with MSP where a step monochromator allows only a defined area of light (with a certain bandwidth) to pass through the sample at a time (2-nm steps are often used). In general, the light source used is an Xenon Short Arc Lamp (XBO) lamp. After passing the fiber sample a Photo Multiplier Tube (PMT) detects the signals and transfers the photons of the light into electrical energy. Creating spectra step by step is a time-consuming but exact process when the equipments setup works correctly.

The second method is used with a DAD or CCD spectrometer where a single beam of light passes the optical path and the sample simultaneously and is split for detection only after passing the sample. This is called multichannel spectroscopy and all results are detected simultaneously. The resolution for detection depends critically on the number of diodes used in a DAD or CCD system. Older systems sometimes lacked a sufficient number of diodes with a small bandwidth. Newer systems, meanwhile, offer a high number of diodes so that the quality of spectra is comparable to MSP. The method is less time consuming and equally exact if the setup works correctly. Applied to these systems, a deuterium/halogen combination has proven to produce good results.

With both methods, spectra can be taken either in absorption or transmittance. Most systems offer software to switch spectra between absorption and transmission. The first derivative of the absorbance spectra should also be taken into account for evaluation. First derivative spectra sometimes prove to be helpful for the interpretation of spectra with a broad and featureless appearance [25].

In addition, fluorescence emission spectra can be measured if fluorescence is observed and systems are suitably equipped.

FTIR Spectroscopy. IR microscopes coupled with a Fourier Transform Infrared (FTIR) spectrometer, covering the wavelength of about 4000–600 cm^{-1}, can be used to determine the composition of single fibers. Spectra can be obtained using transmittance or reflectance methods (*see* **Microscopy: FTIR**).

As with a DAD or CCD spectrometer, all wavelengths pass through the optical path and the sample simultaneously, and the light is only split for detection after passing the sample. This is called *multichannel spectroscopy* where all results are detected simultaneously. The resolution for detection depends extremely on the number of diodes used in an FTIR system. Older systems sometimes lacked a sufficient number of diodes with a small bandwidth. Newer systems, meanwhile, offer a high number of diodes so that the resolution and quality of spectra is satisfying.

With the help of the microscope it is possible to define very small areas for analysis on single fiber or fiber fragments and it is even possible to determine the different materials of core/sheath fibers after flattening the fibers slightly.

FTIR spectroscopy can be used either for the purpose of identification or comparison of fibers [26]. It is a technique that allows the identification of the polymeric composition of the fibers and it is even possible to determine the subtypes of the fibers, in addition to the generic classes [27, 28].

Optical information, melting point determination, or solubility tests cannot always give satisfactory information regarding the fiber subtypes, therefore FTIR might be used as an additional tool to get further information or as a replacement for other techniques.

It is to be noted that, despite the potential of gaining additional information regarding the fiber composition, FTIR spectroscopy can never be used as a replacement for microscopic examination! Microscopic examination is the basis for all further analytical work.

A very important step for using FTIR as a tool for further examination and information is the preparation of the fiber material. To prevent loss, it is suggested that a low-power microscope is used for preparation purposes. To get satisfactory results, some of the fibers need to be flattened before spectra can be taken. This needs some experience and training because the flattening of fibers might produce interference fringes if too much pressure is applied. Diamond cells are often used for this purpose. With attenuated total reflection (ATR) and microinternal reflection two surface techniques are available which are also used for FTIR spectroscopy. The sample is in close contact with the crystal but for the same reasons as above, it is also essential not to apply too much pressure. A variation of other techniques is available and probably useful for special purposes.

Regarding the use of libraries, it is highly recommended to produce in-house spectra libraries for a specific instrument, using known reference material of a reliable source.

Raman Spectroscopy. Raman spectroscopy (named after Sir C. V. Raman) involves the detection of the inelastic scattering of light depending on the vibrational energy level of molecular nonpolar bonds. Recently, various instruments, with different lasers for excitation, have become available for analysis. Different lasers for excitation are necessary to choose the lasers most suitable to gather sufficient information. Therefore the instrument should be equipped with different lasers covering the range from Ultraviolet (UV) to Near Infrared (NIR). Usually, instruments are fitted with three or four lasers for excitation, a sensitive CCD detector is required [29]. Additionally, a light microscope of good quality is essential for fiber examination [30].

Spectra are generally taken in the range $1800-200\,cm^{-1}$. The fibers should be mounted on an aluminum foil that can be fixed on a glass slide. The technique is not widely used as a routine method in fiber analysis, its predominant focus being on the detection of the major dye components [31] and under certain circumstances it may offer valuable information regarding the detection and/or comparison of fiber dyes. However, it must be taken into account that there is a detection limit for minute dye components.

Raman spectroscopy cannot be used as an alternative for UV/VIS spectroscopy. It is a technique that might prove useful in conjunction with UV/VIS spectroscopy and can play a complementary role.

The discrimination power depends on the substrate, the dye, the intensity of the dye, the sensitivity of the instrument, the choice of lasers used for excitation, and the careful interpretation of results. Fluorescence of fibers has an effect on the results too.

The interpretation of results is crucial since small relative intensity differences seem to have an impact. The dye identification on single fibers proves to be difficult at present since no comprehensive spectral libraries of commercial dyes are available in the market. Another difficulty is the limit of detection for minute dye components.

A subgroup within the European Fibres Group (EFG) evaluates the use of Raman to analyze the dye components on single fibers. Projects are ongoing and helpful information is being gathered [32].

Other Instrumentation. There is a wide range of other analytical techniques available. These techniques are used occasionally, depending on circumstances. Pyrolysis–Gas chromatography mass spectroscopy (GC(MS)) and Pyrolysis–Mass Spectrometry are two techniques that are briefly mentioned [33]; these are destructive analytical methods. In general, these are sensitive methods for identification by comparison but the applicability for fiber examination depends highly on the amount of substrate present in a case. This is also true for high performance liquid chromatography (HPLC) [34], which

is also a fast and sensitive technique for dye analysis. The basis for HPLC analysis of fiber dyes is the use of standard dyes for verification.

Numerous other analytical techniques showed up in fiber examination and vanished again, or only single prototypes of the instruments were on the market (e.g., laser microprobe mass analysis (LAMMA)). Specific instrumentation might prove helpful for certain applications but, at present, there is no other instrumentation or technique used in routine case work. Therefore these instruments and applications are not discussed.

Microchemical Techniques for Analysis

Thin Layer Chromatography. Thin layer chromatography (TLC) is one method to detect additional dye components. TLC might add valuable information to visible spectroscopy and is often used if UV spectroscopy is not possible, but since the method is destructive, it should be used only after all other possible paths of examination are exhausted. Prior to TLC analysis, it is necessary to determine the generic class of fibers [35] and the dye class [36]. Depending on the generic class of the fibers and the dye class, it is more or less difficult to extract the dye. Especially, reactive dyes on wool cause problems to extract the dye [37]. It is necessary to choose specific extraction solutions to remove the dye. This can be achieved with the help of dye extraction/classification schemes.

Generally, it has to be noted that it is easier to get sufficient information from single fibers when the dye is dark and a sufficient amount of dye can be extracted. When single fibers only show a very light coloring, most likely more then one fiber is necessary to gain sufficient results.

TLC of optical brighteners and other fluorescent agents proves to be extremely difficult and might cause problems, therefore extreme care must be taken if the method is applied. Under visible and ultraviolet light bands, positions and colors can be detected and documented. In general, TLC results tend to fade. Therefore the plates should be documented immediately (e.g., via photo documentation) to prevent the loss of results. To avoid fading processes, the TLC plates should be stored accordingly.

Solubility Tests. Solubility testing is a destructive method that was often used in the beginning of forensic fiber analysis. Solubility tests provide reliable information regarding the composition of material in terms of generic classes and even of subtypes in some cases [38]. There are several schemes available that are of interest for fiber identification or the analysis of other fiber or dye characteristics [39].

Using specific solvents the effects can be observed more easily under a microscope. Small fragments of the fiber can be used for the different steps of identification.

Although the techniques used in fiber examination are not unique, their application in the fiber area is very specialized and therefore good laboratory systems and training schemes are needed to ensure reliable results. The European Fibres Group (EFG) within the European Network of Forensic Science Institutes (ENFSI) is an excellent platform to exchange information and learn from each other. Organizing suitable collaborative exercises [40] poses particular challenges in fiber work. EFG has organized one such exercise every year since 1994.

Fibers and Textiles Intelligence Work

Textile investigation in forensics covers a wide range of different aspects. Samples range from very small and "invisible" textile traces to those that cannot be detected without the help of a microscope to whole pieces of textiles or clothing, which may offer clues for the investigation.

All information fibers and textile examination can offer that supports the investigator in the investigative process is described as fiber and textile intelligence work and can also be used to enhance the evidential value of the findings. Fiber and textile examination can play a key role in police investigations and can be used as an investigative tool if the forensic scientist and the investigator work in close contact. The continuous exchange of objective information between both sides is a necessary method to increase the efficiency and evidential value of examinations.

Textile Contact Traces and Their Advantages

Textile microtraces in forensic investigation are very often related to contact traces in the best sense of Locard's exchange principle, that is, that which may occur during the contact of one or more textiles with

a person or another item/textile. Therefore, traces can be found in the contact areas between two or more people or objects. The number of transferred fibers depends on intensity and shedding abilities of the donor garment. Additionally, the number of recovered fibers depends on the recipient's surface qualities and on the activities and changes taking place between the times of contact and recovery of the garment.

This characteristic is useful in setting time limits on interactions to be examined. For example, DNA may provide a useful link between a wearer and a balaclava. Fiber links in the hair combing of the suspected wearer suggests recent contact. The literature has many references to target fiber studies [41, 42] and survey data that have been carried out to help assess the significance of matching fibers [24, 43, 44].

Fiber findings can be used as investigative lead, to link garments to a scenario or to the living areas where they originate from. Contact traces in a certain context might also enable the examiner to put them into a relative time frame within a scenario and to enlighten the interactions at the scene (e.g., in the case of car accidents). Furthermore, it is possible to link different crime scenes via the combination of fiber populations found at the scenes. Fiber examination can be a powerful tool if carried out with adequate care to avoid other contamination.

Nonpermanent Contact Traces. This sort of contact causes a nonpermanent reflection of a textile contact, and fiber traces are transient. During the interaction of two or more different textiles/objects, particles originating from these textiles/objects may be transferred (primary transfer). It is most likely that fibers on the surface of a donor garment, which originate from other textile sources, are also transferred in a close symbiosis with the primary transferred fibers (secondary transfer). Mostly, those surface fibers originate from the immediate surrounding of the person wearing the donor garment.

Without available comparison material, directly or primary transferred fibers may be used as valuable lead for further comparison with the living area or certain textiles of a suspect [45]. They are easy to perceive if they are prominent either by number, color or distribution. The technique of using populations or collectives of these prominent fibers as an investigative lead is also known as the *Leitspur Principle* [46, 47] and depends on an extremely detailed fibers recovery and even more detailed information about the case. This is a very labor-intensive examination and is not routinely performed in all laboratories and not suitable for every type of casework.

The donor garment is the intermediate between a suspect's living area and the case-related contact traces and represents the living area of the owner via its surface fibers. These so called secondary fibers occur always in the near vicinity of primary contact traces on the recipient and the combination of primary and secondary fibers in the supposed contact area heightens the evidential value of findings enormously.

The position of fiber findings also can play a prominent role. A detailed trace recovery not only allows to state a contact between two items/textiles it is also possible to link the findings to an alleged scenario. In addition fibers offer the opportunity to link different crime scenes via found fiber populations and the combination of different fiber types and colors heightens the evidential value enormously. Under defined conditions, it is possible to link a person via fiber examination to a certain area at the scene and subsequently to a certain event within a relative time frame (e.g., via 1 : 1 taping in murder cases or via fiber fusion marks in car accidents).

Textile microtraces are not visible to the unaided eye. The invention of taping techniques in the early 1950s meant a revolution in trace recovery. It is now the method of choice in Europe and is very effective in the collection of surface fibers. The use of 1 : 1 taping in murder cases is a time-consuming but high-end technique that offers the opportunity to locate the exact position of the found fibers on a taped surface. Fiber mapping enables the forensic scientist to evaluate the findings and connect them to the crime scenario and the actions taking place on the scene. For this method, it is necessary to place each tape only once for tape lifting. This kind of extensive trace recovery is very useful in cases in which no significant change of the crime scenario took place and trace alterations or a loss of traces were prevented. Cases of murder with the immediate trace recovery on the scene are ideal for the technique of 1 : 1 taping.

The technique applied in murder cases allows to distinguish fiber collectives on a victim by evaluating

the fibers by distribution, density, and number, which, in consequence, offers the opportunity to associate them to case-relevant activities.

A zonal approach [48] might be less time consuming and might offer enough information in some cases, but with this additive method, fiber distribution schemes are not as detailed as for 1 : 1 taping. Questions regarding the interactions at the scene probably cannot be answered in great detail via a zonal mapping, whereas, a 1 : 1 distribution chart enables the examiner to visualize the detailed imprint of the contact. The contact transfers have to be regarded as an imprint of the events taking place on the scene.

As the fiber transfer mainly takes place in the intense contact areas between victim and offender the advantages can be seen in the possibility of the exact location of the fiber's position on a body and/or an object. This way, a fiber map can be produced and the findings can be related to the scenario and the alleged activities on the scene. It is of eminent value to compare the findings with the results of other evidence. As an example, the correlation of information coming from different sources like forensic medicine and fiber investigation (e.g., broken ribs in correlation with fiber findings in the specific area) is of added value in the interpretation of results. An additive collection of fibers would not enable the forensic expert to interpret the findings in the same way.

The foreign fibers found on a textile or another object may assist the examiner and also the investigator in identifying the textile source of the trace evidence [49]. The more detailed the trace recovery is at the scene and the less the situation is untouched and unaltered, the more evidence can be gathered to gain as much information as possible.

In case of 1 : 1 taping (e.g., in murder cases) single fibers can be traced back to a certain location on the victim. This offers the possibility for a fiber mapping that helps to evaluate whether the found fibers not only match but are in correspondence with the position of a suspect's clothes and accordingly fit into the alleged scenario or not. It is a powerful tool to link the fibers found by way of assessing the scenario and building a chain of plausibility that might stay within the fibers' discipline or even might involve other disciplines.

Car accidents offer another approach to link nonpermanent fibers on a surface to a certain incident. Under routine conditions, an airbag sits secured and folded in its compartment and it only inflates during a crash of relatively high energy. This way, a massive fiber transfer can only occur at the time of an accident and might lead very quickly to the person driving the car at the time of the accident. Again, the number, distribution, and relative position of fibers from different worn garments can be detected by using a very detailed recovery technique.

Permanent Fixed Contact Traces. Textile fibers can also occur as permanently fixed traces. This offers the opportunity to find traces a long time after the actual crime or incident took place. Those traces can be detected, for example, as fiber fusion marks in car accidents and may help state the contact between a car and a person involved in an accident and/or to state the position of a person in a car at the time of the accident [50, 51]. This type of trace evidence enables also the reconstruction of individual positions of different people inside a car during a crash.

Textile fusion marks are traces that develop in combination with high kinetic energy. A car crash, for example, is an incident involving a very high energy that enables the plastic surface of a car interior or exterior to melt and to fasten the fibers from a textile at the very moment of the accident. Very often, the direction of the forces can be observed via the position and direction of the fibers in the molten plastic. Under these circumstances, the fiber findings can be related to the case scenario again. This is especially important and helpful in cases where the potential driver of a damaged car is also the authorized owner and/or user of the car. Other then loose fibers on the surface of the car seats, fiber fusion marks can relate a garment of the official user to the accident.

This kind of trace evidence is very resistant against change after the incident and trace recovery can take place a long time after the accident took place. An important factor that has to be taken into account is the influence of UV light on single fibers, especially when fusion marks are located on the car exterior. Another aspect that needs consideration is the thermal influence on fibers [52].

The advantage of this kind of evidence is, in general, the restricted number of persons and often, the limited number of textiles involved, which allows this trace to be highly valued. The cross transfer

of fibers and plastic fusion marks can also be observed in certain cases and allows a very specific interpretation.

Damage

Information about the type of damage can be gained by examining a textiles surface, especially the edges of damaged areas. This way, it is possible to determine the type of damage (mechanical, thermal, or chemical/biological) and its origin [53]. Cuts (slash cuts, stab cuts, scissors cuts, and punctures) [54], tears, shot holes, abrasion, and heat marks, bite marks, degradation effects [55], and effects from wear and tear can be detected. This way, valuable clues for the investigator can be offered by the examiner (e.g., the type of weapon or tool responsible).

Damages in textiles or clothes result often from a device. In other cases, damages are due to interactions between the offender and the victim and result in tears. To determine the cause of the damage the entire outward appearance (e.g., shape and effects of stretching), dimensions (size, length, diameter, etc.) course of the damage (e.g., vertical, horizontal, diagonal, and zigzag) need examination. The surface of a fabric, and the edges of threads and fibers offer additional and more detailed information.

By investigating these damages very carefully, sometimes, the scientist is able to draw conclusions about the crime, the scenario and the subsequent order of actions taking place on the scene. This is very often only possible if other trace evidence is taken into account (e.g., textile damages and blood stains, textile damages and soil) and an overall investigation and interpretation is accomplished.

The textile fiber examiner may have to consult other experts to reach a conclusion when considering other types of trace evidence.

Physical fits of two textile parts in question allow a definite conclusion that these two parts of a textile originate from the same source. Physical matches indicate that the two or more pieces of material belonged to a continuous piece of textile. If no physical match is possible, a complete fiber and textile comparison, including construction (as far as possible) is the basis for further conclusions. In addition, the foreign fibers from different textiles surfaces can be considered to corroborate the other findings.

Reconstruction of Textiles and Further Information

The reconstruction of textiles helps to assess the clothing of a victim at the time of the crime, which might lead to an identification of a person. Accessories, zippers, buttons, labels, and seams are very often the remains that are encountered. These residues of a textile offer the opportunity to identify a certain textile and its model characteristics. Textile imprints left at a scene can help provide information about a textile source and knots might indicate an occupation if special techniques are applied.

Textile labels might also offer valuable information concerning the manufacturer, the type of garment, the number of garments produced, the time frame of production, and the distribution of textiles and clothes in certain countries or areas [56]. If a number of textiles is left behind, labels enable the textile examiner to identify the origin of the textiles and arising from this, possibly the region or country of a person involved.

Video Imaging

In addition, video imaging (comparison between video images and textiles or garments) can play an important role in forensic textile investigation and offers the possibility to gain information about the worn clothes via patterns and logos and offers the possibility for an individual characterization of textiles. Video imaging is the comparison of images and textiles or garments. The images may occur from bank robberies or other scenarios where images related to a crime were taken at the scene. Very often, the physical appearance of a person is not clearly seen on such images. The comparison of textile characteristics from the clothes to be seen can be of important value.

Clothes depend on different influences due to design, construction, and manufacture. Textiles and clothes are influenced and characterized at large by numerous influences during production, which enable the forensic scientist to individually characterize a certain garment [57].

Visible details like the general cut (e.g., arms, bodies, and collars) are responsible for the appearance of a textile. The observation of a textile cut can only give some information about a certain type of clothing and is the basis for all further investigation and comparison.

Certain characteristics can arise by chance in patterned clothes due to production. In general, the patterned fabrics are not designed to fit a certain part of a garment. Resulting from this, a completely unintentionally combination of textile pattern parts results. In different clothes from the same type and model, numerous variations of the general theme can be observed and lead to an individualization of the textiles.

During sewing procedures, unintentionally formed folds and creases can occur. The characteristics of the chosen textile materials can increase these effects. This offers also a possibility to individualize a certain textile, if unintended.

Clothes are characterized by the effects from wear and tear. Folds and crimps and color abrasion, for example, offer a lot of characteristics to individualize a textile.

Another possibility to individualize a textile can be observed when the basic textile underwent intentional or unintentional changes (e.g., different buttons, addition of applications, wrong colored seam material, and additional seams). Through these alterations, a single object is also individually characterized.

A basis for this type of investigation is a wide knowledge of garment production and side effects on garments. In general, every garment has the potential to be individualized, but this is dependent on the quality of the provided images and the possibility to visualize certain details on the images and compare them to the original garment. Conclusions can be drawn if a garment seen on an image is comparable to the garment from an individual.

Every textile, in general, has the potential to be individualized – this is certainly due to the quality of the provided images and the resulting abilities to visualize certain details from the images and relate them to the original textile [58].

Conclusion

Fiber and textile examination rely heavily on the basic analysis of fibers. Fiber persistence and population studies already underline the value of findings at this point of investigation.

In combination with other disciplines, fibers can close missing links in the sequence of events, and despite being generally not visible to the unaided eye, offer lots of valuable information. In addition to pure fibers analysis, there is much more the examiner has to offer that might help the investigator in the process of investigation or in heightening the evidential value of findings. Fiber and textile analysis together with intelligence-led information can be thus used as powerful tool [59].

References

[1] Locard, E. (1928). Dust and its analysis, *Police Journal* **1**, 177–192.

[2] Locard, E. (1930). The analysis of dust traces, part I, *American Journal of Police Science* **1**, 276–298.

[3] Locard, E. (1930). The analysis of dust traces, part II, *American Journal of Police Science* **1**, 401–418.

[4] Locard, E. (1930). The analysis of dust traces, part III, *American Journal of Police Science* **1**, 496–514.

[5] Frei-Sulzer, M. (1951). Die Sicherung von Mikrospuren mit Klebeband, *Kriminalistik* **10/51**, 190–194.

[6] Kück, A. (2004). Different Techniques of taping, CD-ROM, *Proceedings of the 1st Young Scientists Workshop*, European Fibres Group.

[7] Griffin, R.M.E. & Crawford, C. (1997). An improved method for the preparations of combs for use in hair combing kits, *Science & Justice* **37**, 109–113.

[8] David, S.K. & Pailthorpe, M.T. (1999). Classification of textile fibres: production structure and properties, in *Forensic Examination of Fibres*, J. Robertson & M. Grieve, eds, Taylor & Francis, pp. 1–31.

[9] Eberle, H., Hermeling, H., Hornberger, M., Kilgus, R., Menzer, D. & Ring, W. (2008). *Clothing Technology, from Fibre to Fashion*, Europa Verlag.

[10] McIntrye, J.E. & Daniels, P.N. (eds) (1995). *Textile Terms and Definitions*, The Textile Institute, Manchester.

[11] Gordon, C.J. (1984). *Handbook of Textile Fibres, Natural Fibres*, Woodhead.

[12] BISFA (2006). *Terminology of Man Made Fibres*, BISFA, pp. 5–93.

[13] Koslowki, H.J. (1998). *Dictionary of Man Made Fibres – Terms Figures and Trademarks*, International Business Press.

[14] Palenik, S.J. (1999). Microscopical examination of fibres, in *Forensic Examination of Fibres*, J. Robertson & M. Grieve, eds, Taylor & Francis, pp. 311–325.

[15] McCrone, W.C. (1991). Light microscopy, in *Physical Methods of Chemistry*, B.W. Rossiter & J.F. Hamilton, John Wiley & Sons.

[16] Stoeffler, S. (1996). A flowchart system for the identification of common synthetic fibres by polarised light microscopy, *Journal of Forensic Science* **41**, 297–299.

[17] Grieve, M.C. (1983). The use of melting point and refractive index determination to distinguish between colourless polyester fibres, *Forensic Science International* **22**, 31–48.

[18] Herman, B. (1998). *Fluorescence Microscopy*, BIOS Scientific Publishers, Oxford.

[19] Bradbury, S. & Evennett, P. (1996). *Fluorescence Microscopy. Contrast Techniques in Light Microscopy*, BIOS Scientific Publishers, Oxford.
[20] Lynch, B. (1981). Investigation of single fibres by X-ray diffraction in forensic analysis, *X-Ray Spectroscopy* **10**, 196–197.
[21] Roux, C. (1999). Scanning electron microscopy and elemental analysis, in *Forensic Examination of Fibres*, J. Robertson & M. Grieve, eds, Taylor & Francis, pp. 239–250.
[22] Pelton, W.R. (1995). Distinguishing the cause of textile fibre damage using the scanning electron microscope (SEM), *Journal of Forensic Science* **40**, 874–882.
[23] Adolf, F.P. (1999). Microspectrophotometry/colour measurement in forensic examination of fibres, in *Forensic Examination of Fibres*, J. Robertson & M. Grieve, eds, Taylor & Francis, pp. 251–289.
[24] Biermann, T.W. (2007). Blocks of colour IV: the evidential value of blue and red cotton fibres, *Science & Justice* **47**(2), 68–87.
[25] Biermann, T.W. & Wiggins, K. (2006). UV/VIS spectroscopy, *Young Scientists Workshop*, Prague.
[26] Tungol, M., Bartick, E.G. & Montaser, A. Forensic examination of synthetic textile fibres by microscopic infrared spectroscopy, in *Practical Spectroscopy, Practical Guide to Infrared Microspectroscopy*, H.J. Humecki, ed, Vol. 19, Marcel Dekker, pp. 245–286.
[27] Grieve, M.C. (1995). Another look at the classification of acrylic fibres, using FTIR microscopy, *Science & Justice* **35**, 179–190.
[28] Grieve, M.C., Griffin, R.M.E. & Malone, R. (1998). Characteristic dye absorption peaks found in the FTIR spectra of coloured acrylic fibres, *Science & Justice* **38**(1), 27–37.
[29] Thomas, J., Buzzini, P., Massonnet, G., Reedy, B. & Roux, C. (2005). Raman spectroscopy and the forensic analysis of black/grey and blue cotton fibres. Part 1. Investigation of the effects of varying laser wavelength, *Forensic Science International* **152**(2–3), 189–197.
[30] Bergin, F.J. (1990). A microscope for Fourier transform Raman spectroscopy, *Spectrochimica Acta* **46A**(2), 153–159.
[31] Bourgeois, D. & Church, S.P. (1990). Studies of dyestuffs in fibres by Fourier transform Raman spectroscopy, *Spectrochimica Acta* **46A**(2), 295–301.
[32] Massonnet, G., Buzzini, P., Jochem, G., Stauber, M., Coyle, T., Roux, C., Thomas, J., Leijenhorst, H., van Zanten, Z., Wiggins, K., Russell, C., Chabli, S. & Rosengarten, A. (2005). Evaluation of Raman spectroscopy for the analysis of colored fibers: a collaborative study, *Journal of Forensic Sciences* **50**(5), 1028–1038.
[33] Almer, J. (1991). Subclassification of polyacrylonitrile fibres by pyrolysis capillary gas chromatography, *Canadian Society of Forensic Science Journal* **24**, 51–64.
[34] Griffin, R. & Speers, J. (1999). High-performance liquid chromatography, in *Forensic Examination of Fibres*, J. Robertson & M. Grieve, eds, Taylor & Francis, pp. 311–325.
[35] Rendle, D.F. & Wiggins, K. (1995). Forensic analysis of textile fibre dyes, *Review of Progress in Coloration* **25**, 29–34.
[36] Wiggins, K.G. (1999). Thin layer chromatography for fibre dyes, in *Forensic Examination of Fibres*, J. Robertson & M. Grieve, eds, Taylor & Francis, 291–301.
[37] Wiggins, K.G., Crabtree,S.R. & March, B.M. (1996). The importance of thin layer chromatography in the analysis of reactive dyes released from wool fibres, *Journal of Forensic Science* **41**(6), 1042–1045.
[38] Stratmann, M. (1987). Identification of textile fibres, in *Applied Polymer Analysis and Characterization*, J.R. Mitchell, ed, McMillan Publishing, pp. 387–411.
[39] Perry, D.R. (ed) (1985). *Identification of Textile Materials*, The Textile Institute Manchester.
[40] European Network of Forensic Science Institutes (1998–99). *Guide on the Conduct of Proficiency Tests and Collaborative Exercises within ENFSI*, Yearbook, European Network of Forensic Science Institutes.
[41] Palmer, R. & Chinherende, V. (1996). A target fibre study using cinema and car seats as recipient items, *Journal of Forensic Science* **41**, 802–803.
[42] Brüschweiler, W. & Grieve, M.C. (1997). A study on the random distribution of a red acrylic target fibre, *Science & Justice* **37**, 85–89.
[43] Grieve, M.C. (1999). Interpretation of fibre evidence, in *Forensic Examination of Fibres*, J. Robertson & M. Grieve, eds, Taylor & Francis, pp. 343–361.
[44] Grieve, M.C. (2001). The evidential value of black cotton fibres, *Science & Justice* **41**, 245–260.
[45] Nehse, K. (2004). Using 1:1 taping to reconstruct a source, in *Trace Evidence Analysis*, M. Houck, ed, Elsevier, pp. 191–210.
[46] Neubert-Kirffel, D. (2000). Das Leitspurenkonzept, *Kriminalistik* **6**, 398–404.
[47] Decke, U. (2000). Nochmals: Das Leitspurenkonzept, *Kriminalistik* **7**, 467–472.
[48] Palmer, R. (2006). A zonal approach, CD-ROM, *Proceedings of the 14th European Fibres Group meeting*, European Fibres Group.
[49] Palmer, R. (2004). An intelligence led investigation using trace evidence, in *Trace Evidence Analysis*, M. Houck, ed, Elsevier, pp. 89–104.
[50] Jochem, G. (2001). Rekonstruktion der Insassen-Sitzverteilung in Unfallfahrzeugen, *Kriminalistik* **5**, 341–345.
[51] Jochem, G. (2004). Fiber-plastic fusions and related trace material in traffic accident investigation, in *Trace Evidence Analysis*, M. Houck, ed, Elsevier, pp. 53–88.
[52] Schiller, W.R. (1956). Textilfasern in anschmelzspuren – thermisch bedingte farbänderungen faseranschmelzspuren, *Kriminalistik* **49**, 728–730.
[53] Taupin, J.M., Adolf, F.P. & Robertson, J. (1999). Examination of damage to textiles, in *Forensic Examination of Fibres*, J. Robertson & M. Grieve, eds, Taylor & Francis, pp. 65–87.

[54] Adolf, F.P. (1996). Examination of damage to fabric caused by sharp objects, *Proceedings of the 4th European Fibres Group Meeting* European Fibres Group, pp. 82–92.
[55] Evans, E. & McCarthy, B. (1998). Biodeterioration of natural fibres, *Journal of the Society of Dyers and Colorists* **114**(4), 114–116.
[56] Dillinger, S. (2005). A review of textile labels in forensic science, CD-ROM, *Proceedings of the 13th European Fibres Group Meeting* European Fibres Group.
[57] Nehse, K. & Wendt, C. (2002). Wie Individuell Sind Textilien? *Kriminalistik* **6**, 391–395.
[58] vor der Bruegge, R.W. (1999). Photographic identification of denim trousers from bank surveillance film, *Journal of Forensic Science* **44**(3), 613–622.
[59] Biermann, T.W. & Grieve, M. (2001). Die Zukunft der forensischen Faseranalyse. *Kriminalistik* **5**, 337–340.

KORNELIA NEHSE

Expert Malpractice *see* Malpractice Actions against Experts

Expert Opinion: Admissibility, France, Germany, Italy, Spain *see* Expert Opinion in Court: Civil Law Jurisdictions (France, Germany, Italy, and Spain)

Expert Opinion: Admissibility, UK, Canada, Australia *see* Expert Opinion: United Kingdom, Canada, and Australia

Expert Opinion: Appeal v. Trial

When Does the Presentation of Expert Testimony Occur

The presentation of expert testimony in court is (*see* **Expert Opinion in Court: a Comparison of Approaches**; **Expert Opinion: United States**; **Expert Opinion: United Kingdom, Canada, and Australia**;) ordinarily restricted to the giving of evidence at trials. Seldom is expert testimony needed when the losing party appeals the decision of the trial court to a tribunal with appellate power. In order to explore the differences between various juridical approaches, two different jurisdictional systems are examined. They are (i) the process followed in the United States of America and (ii) that in use in the United Kingdom.

The United States of America

The American system of trial and appeals procedure is still very much based on the common law model (*see* **Adversary Systems of Justice**) that developed in England centuries ago, modified only insofar as US constitutional requirements of the Bill of Rights have mandated some alteration. It permits the taking of appropriate and admissible expert testimony at the trial, and at certain postconviction proceedings. No expert testimony, indeed, no lay evidence either, is admitted in appellate proceedings.

The bulwark of the system of justice, in particular that of criminal justice, is that all factual disputes are to be settled at the trial. The common law jury (*see* **Jury Instructions on Expert Testimony**) system, which retains its validity for civil as well as criminal trials, is the place where all factual disputes are to be settled. In theory, if not in absolute terms, nothing is permitted to interfere with the jury's fact-finding authority. Therefore, experts may testify only at trials, and not on appeal.

In the United States, plaintiff as well as defendant in a civil dispute has a right to appeal the determination of the trial. In criminal cases, however, an appeal can be taken only by a defendant following conviction. Where the jury returns a verdict of not guilty, the prosecution cannot appeal even if flagrant

jury error can be shown to have occurred. The constitutional protection against being placed in jeopardy twice for the same act makes impossible prosecutor appeals from not guilty verdicts.

There are some rare situations where prosecution appeals are possible from adverse rulings on pretrial motions excluding evidence, if a statutory provision authorizes such appeals. In the absence of such statutory authority, the common law reigns supreme and appeals can only be taken by an accused who has been convicted.

When a convicted person appeals (the appellant), the only issues before the appellate court will be matters of law. All factual disagreements in the evidence as presented by either the prosecution or the defense will be deemed to have been resolved by the jury. When the defendant appeals, he or she will prepare a petition for appeal, wherein all legal arguments for reversal of the judgment of conviction must be enumerated. Defendant's counsel will also secure the printing of the verbatim transcript of so much of the trial testimony and argument as is deemed to be appropriate for a proper consideration of the issues on appeal. Ordinarily, no argument can be made by the appellant that the jury erred in its fact finding.

Following the preparation of the petition for appeal, each party will file briefs with the reviewing tribunal, which may also schedule arguments where each side may orally argue its position. At these oral arguments, only counsels speak. No witness, lay or expert, is permitted to give testimony on appeal.

The United States also adheres to a fairly strict "waiver" of rights concept. This means that if a convicted person knows of any reason why a trial court is committing error during the trial, the accused is required to object to the error during the trial. If that is not done, the point is deemed waived forever. If the reviewing tribunal, after briefing and oral argument, hands down a decision to the effect that nothing prejudicial to the defendant has happened at trial, the judgment of conviction will be affirmed.

If legal errors were committed, the conviction will be reversed and the case will typically be remanded for a new trial. In rare cases, the appellate tribunal may find that the evidentiary record is so defective that it cannot, as a matter of law, support a judgment of conviction. If that is the case, an outright reversal without retrial will be mandated and the accused is ordered to be released.

Upon reversal and remand, the court will treat the earlier conviction as if it had never occurred. At the subsequent new trial before a different jury, if a jury trial is selected, all the evidence presented at the first trial, including expert testimony, can now again be presented, including additional evidence that has been discovered since the end of the first trial.

In order to enable higher courts to correct egregious errors in court decisions and prevent injustices from occurring, especially when newly discovered evidence of potential innocence that was not available at the first trial has been discovered, there exist separate statutory postconviction remedies in the United States that permit further review even after a conviction has already been affirmed by a reviewing court on direct appeal.

To initiate such a postconviction action, the defendant files a petition for a writ of habeas corpus in a state or federal trial or appellate court. Such a petition asserts that the petitioner is being unlawfully detained and requests release and/or retrial. If, upon a study of the petition itself, the judge believes that it likely has merit, a hearing on the specific issues that are properly cognizable on habeas corpus will be scheduled. At such a hearing, expert or lay testimony on the newly discovered evidence may be heard.

If, at the conclusion of the habeas corpus hearing, the judge is convinced that there is no ground supporting the convicted person's petition, the habeas petition will be dismissed. If the judge decides the petition was well founded, the petition for habeas corpus will be granted and the conviction reversed and remanded for a new trial. The prosecution may also appeal the granting of the writ before retrial can be held.

The United Kingdom

In the United Kingdom, the use of expert testimony (*see* **Expert Opinion: United Kingdom, Canada, and Australia**) at trial proceeds generally in the same manner as in the United States. When it comes to the appellate process, some notable differences can be observed.

While the United Kingdom's system has the same common law origin as that of the United States, there has been a more extensive change in the UK posttrial procedure than is the case in the United States. In

accord with common law traditions, before 1907, the legal system of England and Wales allowed appeals to be made on issues of law only. Such an appeal was perfected by the filing of a writ of error and the reservation of legal questions. Issues of fact could not be relitigated on appeal, because an appeal against a conviction would undermine the role of the jury. Thus, as in the United States, expert testimony was typically admitted at trials, but not during appellate proceedings.

The enactment of the Criminal Appeal Act in 1907 (amended several times subsequently) essentially safeguarded the traditional common law role of the trial jury, but gave the new Court of Criminal Appeal powers that allowed it to rectify certain miscarriages of justice. According to this act, an appeal could be granted on any ground of fact, or mixed fact and law. If the Court of Appeal allowed an appeal against the conviction on the ground the conviction was "unsafe", the conviction could be set aside and a new trial ordered.

The Act's extension through the passage of the Criminal Justice Act of 1988 gave the court discretion under its Section 23 to do whatever in its judgment was necessary and expedient in the interests of justice. As a result of this further enactment, Section 23(1) permits the bringing of new evidence before the Court of Criminal Appeal. Such evidence may include real, documentary, and oral testimony before the court. The law provides that "To that end it may order the production of a document or other object or the attendance or examination of a witness, whether on request or of its own motion, and may receive evidence not adduced at trial".

Unlike in the United States of America, in the United Kingdom, expert testimony may be freely presented before the Court of Criminal Appeal and the common law prohibition restricting criminal appeals solely to legal issues has been removed.

R. v. Pendleton [2002] 1 W.L.R. 72, a decision of the House of Lords, settled some significant legal issues as a result of these appellate process changes in the United Kingdom. In paragraph 17 of the opinion, it began by stressing that the central role of the jury in a trial by indictment was to be maintained:

> "Trial by jury does not mean trial by jury in the first instance and trial by judges of the Court of Appeal in the second. The Court of Appeal is entrusted with a power of review to guard against the possibility of injustice but it is a power to be exercised with caution, mindful that the Court of Appeal is not privy to the jury's deliberation and must not intrude into territory which properly belongs to the jury".

Whether a retrial was desirable is an issue the reviewing court determined should not be reached until, on the evidence presented in the Court of Appeal, the earlier conviction was deemed to be "unsafe". If so, the statutory duty of the court requires it to allow an appeal and to quash the conviction. A mere risk that a conviction was unsafe, however, does not suffice; the appellant (the convicted person) has the burden to persuade the Court of Appeal that the conviction is, in fact, unsafe. If, exercising its own judgment, the Court of Appeal finds that the evidence supports a finding that the conviction was unsafe, only then can the issue of whether a retrial is to occur be reached.

Related Articles

Discovery: Depositions
Discovery: Discovery Motions
Discovery in the United States: Civil Cases
Discovery in the United States: Criminal Cases
Daubert v. Merrell Dow Pharmaceuticals
Discovery of Expert Findings
Expert Opinion in Court: a Comparison of Approaches
Expert Opinion: United States
Expert Opinion in Court: Civil Law Jurisdictions (France, Germany, Italy, and Spain)
Expert Opinion: United Kingdom, Canada, and Australia
Frye v. United States
General Electric v. Joiner
***In Limine* Motions and Hearings**
Weisgram v. Marley

ANDRE MOENSSENS

Expert Opinion: Communicating in Court *see* Statistical Evidence in Court

Expert Opinion: Peer Review *see* Peer Review as Affecting Opinion Evidence

Expert Opinion: Statistical *see* Statistical Evidence in Court

Expert Opinion: United States

American courts have admitted expert testimony for nearly two centuries. Before binding court precedents or legislation existed regulating such admission, the decision to permit a person to offer opinion testimony was made under poorly articulated common law principles. Essentially, this meant that if the trial judge believed the witness's expertise was believable, and available information showed he was at least marginally qualified, the witness would be permitted to offer opinion evidence.

The first important legal development occurred when, in 1923, the Court of Appeals for the District of Columbia handed down its relatively brief, but extremely significant, decision in the case of *Frye v. United States,* 193 F. 1013 (D.C. Cir. 1923) (*see* ***Frye v. United States*** for the facts and full text of this brief, though still significant, precedent-setting court opinion) [1]. In the *Frye* case, the court created what became known as the *general acceptance* test, requiring the specialized subject matter of the evidence to have been generally accepted as reliable in the field to which it belongs.

Since neither Congress nor the US Supreme Court had enacted rules on the subject, each federal circuit court of appeals was free to decide what the admissibility test ought to be in its own jurisdiction. Individual state courts, too, were free to adopt their own rule of admissibility. Perhaps by default of other workable models, *Frye*'s "general acceptance test" became, over the years, the polestar that generally guided admissibility decisions of all types of expert opinion testimony in criminal cases, in state and federal American courts.

In 1975, Congress, for the first time, enacted legislation dealing with a wide variety of evidential issues when it passed the Federal Rules of Evidence. Rule 702 had applicability to expert witnesses in civil and criminal cases. The Rule, as it currently exists, provides the following:

> "If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case".

The language that follows "if (1) . . . " was added after the Supreme Court's decision in *Daubert*, next to be discussed.

Following 1975, state and federal courts were divided alike on whether the passage of the Rules of Evidence, which were also adopted as state rules in a number of American states, signified that *Frye* was no longer the controlling precedent. Most of the courts that faced this issue concluded that *Frye* retained its validity, though a minority of jurisdictions fashioned rules that hinged the admissibility decision on "mere relevance" – whether the evidence, if believed, was more likely to be helpful than not.

Frye also had never become widely used as a guide for admitting expert evidence in civil cases until the next significant legal development, which occurred in 1993, when the United States Supreme Court applied Federal Rule of Evidence 702 (on expert witnesses) in an important civil case. The decision was *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579 (1993), and the Court overturned federal precedent of more than half a century by replacing the "general acceptance test" with a loose standard that conditioned admissibility of challenged expert opinion testimony on its reliability [2]. The Court also articulated a series of guidelines for lower courts to use in making the reliability decision (s*ee* ***Daubert v. Merrell Dow Pharmaceuticals*** for a discussion of the facts in *Daubert* and the record trial courts are

expected to create in deciding whether a particular technique was shown to be reliable or not).

The *Daubert* opinion revolutionized the way in which courts had to think before admitting challenged expert opinion evidence. The meaning of the Court's new test remains fiercely debated by academicians. Perhaps that was inevitable since *Daubert* sought to fashion a test for admissibility in a way that attempted to combine irreconcilable principles. Proceeding from the thought that the reliability of expert evidence had to depend on whether it rested on genuinely "scientific" principles, the Court then seemed to embrace two incompatible philosophies of science. On one hand, it seemed to embrace Karl Popper's "falsifiability" test, and, on the other hand, it sought to accommodate Carl Hempel's "verification" approach to validity. These two philosophies are incompatible. The result: *Daubert* is an opinion that lacks the very thing it sought to require of expert testimony: scientific validity. No wonder courts and legal commentators have been at odds to apply or explain *Daubert* since 1993. The Supreme Court was quite clear, however, in stating that the new approach was less restrictive than *Frye*'s principle, which it saw as too conservative.

A whole host of new issues surfaced in the wake of *Daubert*. Since the Court had fashioned its rule in terms of "scientific" expert testimony, the question soon arose whether it would also apply to nonscientific, experience-based expert evidence. In applying *Daubert* in such cases, lower courts came to inconsistent results, as they did in attempting to unravel whether a discipline's endeavors were based on "science".

Those questions were resolved in 1999 when the US Supreme Court handed down another momentous decision in *Kumho Tire v. Carmichael*, 526 U.S. 137 (1999), another civil case (*see* ***Kumho Tire v. Carmichael*** for a discussion of its facts and holding) [3]. Recognizing that *Daubert's* rationale was based on an interpretation of Federal Rule of Evidence 702, which rule lumps all types of expert opinion testimony in the same category, the Court said that the *Daubert* principle, which required proof of reliability, applied to all forms of expert evidence, though the Court recognized that it might not be possible to literally apply each of the *Daubert* factors to all forms of experience-based opinion evidence. In recognition, the *Kumho Tire* Court reiterated its admonishment of *Daubert* that the 1993 decision offered flexible guidelines, instead of rigorous litmus tests, for courts to apply.

There are perhaps more court decisions in the United States dealing with the admissibility of forensic evidence in the years that followed 1993, than there were in all the years preceding *Daubert.* The interpretations of courts given to the Supreme Court precedents, often inconsistent with one another, continue to appear monthly in the American jurisprudential literature.

References

[1] Frye v. United States, 193 F. 1013 (D.C. Cir. 1923).
[2] Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993).
[3] Kumho Tire v. Carmichael, 526 U.S. 137 (1999).

Related Articles

Daubert v. Merrell Dow Pharmaceuticals
Discovery: Depositions
Discovery: Discovery Motions
Discovery of Expert Findings
Discovery in the United States: Civil Cases
Discovery in the United States: Criminal Cases
Expert Opinion in Court: Civil Law Jurisdictions (France, Germany, Italy, and Spain)
Expert Opinion: United Kingdom, Canada, and Australia
Expert Opinion in Court: a Comparison of Approaches
Frye v. United States
General Electric v. Joiner
***In Limine* Motions and Hearings**
Scientific Method Compared to Legal Method
Weisgram v. Marley

ANDRE MOENSSENS

Expert Opinion: Ultimate Issue

see Ultimate Issue Evidence by Experts

Expert Opinion: United Kingdom, Canada, and Australia

England, Wales, and Scotland

The role of the expert in UK courts is to assist the court itself in deciding the reliability of evidence. There is no such thing as a claimant's or a defendant's expert and, by Rule 35.3 of the Civil Procedure Rules 1998 [1], it is the duty of the expert to assist the court on matters within their expertise. This duty overrides any contractual obligation that may have been formed between the expert and the party instructing and paying them.

It is the court that decides the validity of the expert witness and it is not necessary, in most cases, that the person called is formally qualified. In *R v. Silverlock* [1894] 2 QB 766 [3], an amateur graphologist was allowed to act as an expert on handwriting and a number of cases since have maintained the position that expert status can be obtained by personal experience as well as by formal training. The decision in *Liddle v. Middleton* [1894] 2 QB 766 [4] would seem to agree that the expert need only have expertise beyond that of a layman, not that the expert be qualified beyond that of a layman.

In the civil courts, the admissibility of expert evidence is governed by section 3 of the Civil Evidence Act 1972 and rules of the court [2]. "An expert was only qualified to give evidence that was relevant if his knowledge and expertise was beyond that of the layman and such evidence had to relate to a factual issue in the case".

There has long been some difficulty in determining which matters require an expert opinion but, in general, where the evidence to be given is outside the experience of the court, there is a *prima facie* requirement for an expert. If indeed an expert is required, then evidence of a nonexpert is not admissible. *R v. Inch* (1989) 91 Cr App R 51 [8].

Where the tribunal of fact has sufficient knowledge and understanding of the subject, it is not necessary to call an expert and this renders expert testimony inadmissible, *R v. Turner* [1975] QB 834 [6]. It is also the case that an expert's views on subjects within the understanding of the tribunal are inadmissible as explained poignantly by Lawton LJ:

> "The fact that an expert witness has impressive scientific qualifications does not by that fact alone make his opinion on matters of human nature and behaviour within the limits of normality any more helpful than that of the jurors but there is a danger that they may think it does."

By Rule 24 Criminal Procedure Rules in criminal trials the evidence of the expert must be, as soon as practicable, furnished to the other side along with the expert's observations, calculations, and other procedures. Failure to do so will render expert opinion inadmissible.

One further feature of the expert evidence is that an expert may rely on information, which in other circumstances, would be regarded as hearsay. A scientific expert may, for instance, rely on publications in journals, databases, and textbooks as source material for the opinions they give in evidence. *R v. Abadom* [1983] 1 WLR 126 [7].

An expert cannot be asked to state an opinion on the ultimate issue (guilt or innocence). *R v. Holmes* [1953] 1 WLR 686 [5]. The ultimate issue is a question for the jury, but it has been said that so long as the question is worded differently it may be allowed. *R v. Stockwell* (1993) 97 Cr App R 260 [9].

Australia

In Australia, the common law rules on evidence have been largely superseded by the Evidence Act 1995, although some residual use of common law remains in some state jurisdictions. Under section 56 of The Evidence Act, expert testimony is only admissible where it is relevant. The "Opinion Rule" under section 76 of the Act states that

> "Evidence of an opinion is not admissible to prove the existence of a fact about the existence of which the opinion was expressed."

The Evidence Act 1995 makes provision for exceptions to this rule where the evidence is given by experts by section 79 of the Act.

> "If a person has specialised knowledge based on the persons training, study or experience, the opinion rule does not apply to evidence of an opinion of that person that is wholly or substantially based on that knowledge."

The meaning and scope of this provision can be found in the ruling of Justice Heydon in *Makita (Australia) Pty Ltd. v. Sprowles* (2001) 52 NSWLR 705 [11].

This equates with the position in the United Kingdom where, in general terms, opinion evidence is not admissible unless the person tendering it demonstrate that he have this specialized knowledge. It is to be noted in this codified version of the rule that no formal qualifications are required and that "study and experience" have as much bearing as training. There is no specific definition of what is meant by "specialized knowledge" and it largely coincides with the position in the United Kingdom that the knowledge must be beyond the experience of ordinary people.

In civil cases, in Australia, the Uniform Civil Procedure Rule 2005 provide the code of conduct for experts and, as in the United Kingdom, the position is clear as to the role of the expert in that they are there ostensibly to assist in the decision making of the court rather than as an advocate for either party. The expert in Australia is not a hired gun.

Canada

The Canadian Courts traditionally followed the position established in the United Kingdom that expert evidence was admissible if the evidence assisted the court in reaching a decision. The admissibility relating to reliability and quality was assessed by the judge and, in 1994, the Supreme Court of Canada decided that the standard was too low. The Supreme Court stated that while the judge was appropriately the "gatekeeper" for expert evidence that novel science should be subjected to special scrutiny, *R v. J.-L J* [2000] 2 SCR 600 [10]. This position placed the Canadian courts in a similar position to that in the United States and courts there are now encouraged to use the US model known as the *Daubert* test of reliability, *William Daubert et al. v. Merrell Dow Pharmaceuticals Inc.* 509 U.S. 579 (1993) [12] (*see* **Expert Opinion: United States**). Courts in Canada following *R v. J.-L J* may strike out expert witness evidence if it does not come up to scrutiny under the *Daubert* test.

References

[1] *Civil Procedure Rules* 1998 Rule 35.3 (UK). Uniform Civil Procedure Rule 2005 (Australia).
[2] The Evidence Act s.56, s.76, s.79. (Australia) Civil Evidence Act 1972.
[3] R v. Silverlock [1894] 2 QB 766.
[4] Liddle v. Middleton [1894] 2 QB 766.
[5] R v. Holmes [1953] 1 WLR 686.
[6] R v. Turner [1975] QB 834.
[7] R v. Abadom [1983] 1 WLR 126.
[8] R v. Inch (1989) 91 Cr App R 51.
[9] R v. Stockwell (1993) 97 Cr App R 260.
[10] R v. J.-LJ [2000] 2 SCR 600.
[11] Makita (Australia) pty Ltd. v Sprowles (2001) 52 NSWLR 705.
[12] William Daubert *et al.* v. Merrell Dow Pharmaceuticals Inc. 509 U.S. 579 (1993).

Related Articles

Expert Opinion in Court: Civil Law Jurisdictions (France, Germany, Italy, and Spain)

Expert Opinion in Court: a Comparison of Approaches

BARRY TURNER

Expert Opinion in Court: a Comparison of Approaches

Before commencing on a jurisdiction-by-jurisdiction examination of the systems for employing expert testimony, a brief look at the main jurisprudential traditions lays a proper foundation for the subject matter.

There are many commentaries on the differences between traditions of legal systems of Europe, the United States, and elsewhere. Most juridical systems fall into one of two groups, each exhibiting distinct legal philosophies and characteristics. On one hand, there is the tradition employed in the English-speaking world, which is typically described as the Anglo–Saxon jurisprudence, commonly referred to as employing an *adversarial approach*. On the other hand, there is the civil law system that is based on the Napoleonic Code, which in turn was modified from

the Roman Law model. The civil law system is often called the *inquisitorial* or *investigative system*.

An effective comparison of the two main legal philosophies is fraught with difficulties. Both systems share some terminology, but the terms used have significantly different meanings in each. The rules of how the tribunals function, what the roles of various law professionals are, and how decisions are made, including the uses of juries and panels of judges, also differ between these two systems.

The inquisitorial system is broadly based on a search for the truth and is structured around an exhaustive analysis of the evidence by a specialist judge or, as this official is called in the French legal system, the *juge d'instruction* or *examining magistrate*. One who is familiar with the inquisitorial system might be tempted to compare the examining magistrate to the prosecutor of the American system, but such comparison is inappropriate. An examining magistrate is a member of the judicial branch, whereas a prosecutor is typically in the service of the executive branch.

The adversarial system is grounded on a system of rules of evidence designed to limit the "factual" evidence that may be presented to the court. The purpose of these rules of evidence is to avoid injecting into the lay jury's fact-finding process, any proof that is of little relevance to resolving disputed issues. Further, judges in adversarial system jurisdictions have the ability to also exclude evidence of doubtful reliability or that would unfairly prejudice the jury to an accused in a criminal case. This does not occur in civil law systems where all evidence is effectively analyzed by the examining magistrate, although it must be added that the examining magistrate does not have any role to play in any subsequent trial.

In the adversary system, questions of fact are decided by a jury, or, if no jury is present (called a *bench* trial), by the judge. In a jury trial, the judge decides the issues of law and instructs the jurors in the legal principles that they must apply to the facts, as they have found them to be established by evidence that has been presented. The most common type of evidence that is excluded from the trier of fact in adversary system courts is evidence of bad character. The rationale behind excluding, what in many cases can be directly relevant evidence, is that the evidence of the bad character of the accused may have a prejudicial effect on the minds of jurors that outweighs its probative value, thus leading to a verdict based on prejudice rather than "fact". Counsel for the accused would, at the Voir Dire, which is the preliminary hearing before the court and conducted without a jury, make an application (or present a "motion") that certain evidence should not be admitted. In England and Wales, the rules on bad character evidence established for over 100 years in the Criminal Evidence Act of 1898, have now been fundamentally altered by the passage of the Criminal Justice Act of 2003, which increased the circumstances where evidence of character is admissible despite its potential prejudicial effect.

Other evidence that has increasingly come under scrutiny for exclusion is autoptic or forensic evidence. The use of forensic evidence is of powerful probative value, but it can also be inherently misleading in that the reliability of scientific evidence often falls below the trier of fact's faith in it. In recent years, a number of high profile cases regarding expert testimony have drawn attention to the often fallible nature of such evidence. Forensic evidence is now much more likely to be questioned at the Voir Dire as to not only its relevance but also its accuracy.

One common misconception held in the adversarial world is that, in civil law systems, the accused is "guilty until proved innocent". That is incorrect. All European civil law systems are intimately tied to the European Convention on Human Rights, which recognizes the presumption of innocence. The misconception is fueled by a lack of understanding of the processes of the courts in the civil law system, and also to some extent on the subtleties of translation of legal language. In this regard, the French text of the European Convention of Human Rights at Article Six reads as follows:

> "Toute personne accusée d'une infraction est presumée innocente jusqu'à ce que sa culpabilité ai été établie."

The article refers to the defendant's presumption of innocence, which persists until the facts *establish* his guilt. This compares to the English text where the presumption of innocence remains until guilt is *proved* according to law. It is in the arrival at *proof* or *establishment* that the systems differ. The English text presumes that the "proof" is obtained during a trial and that *proved according to law* refers to a verdict arrived at after the trier of fact has examined all evidence adduced before the court. The French version talks of guilt being "established". This means

that a *prima facie* case of guilt can be demonstrated in advance of the trial by the evidence accumulated in the *dossier* during the inquiry by the examining magistrate. The purpose of the trial in the French civil law system is to allow the trier of fact to judge the accused as a person against the material in the *dossier* or case file. Any lack of *prima facie* evidence will almost certainly result in the case not reaching trial.

In reality, the two systems do not exist in opposition or in isolation. There is a great deal of overlap or convergence, and the assessment of factual evidence in both differs more in style than substance. The judges in the adversarial system now often play a more "inquiring" role, while in some civil law systems, such as Italy, there is a move toward adversarial-type cross examination of witnesses. In Australia, in particular, ostensibly an adversarial jurisdiction, there is a major debate as to the evolution toward a more active and inquisitorial participatory role played by judges.

In Britain, the Crown Prosecution Service has a role, which could be compared to that of an examining magistrate, in that a Crown Prosecutor will examine evidence for relevance and probative value and will apply an "evidence test" before the case can be brought to the bar. In practical terms, the Crown Prosecution Service will not send a case to trial unless the totality of the evidence, what in France would be a *dossier*, contains sufficient evidence to make a guilty verdict, at least, very likely. While there is still the possibility for evidence to be skillfully unraveled or silenced by a sharp defense counsel at the actual trial, there will be in existence at the opening of the trial a situation that, to all intent and purposes, could be equated to the *prima facie* case seen in the civil law process.

As to the position regarding expertise of witnesses, there will be opportunities prior to the trial for this to be examined, and if a case for ureliability or the underlying science or technology or inapplicability of its conclusions to the disputed facts can be made, such evidence may be ruled inadmissible. While the partially codified system established by the *Frye* and *Daubert* rulings in the United States are not directly applicable in the majority of adversarial courts, and do not apply at all in the inquisitorial system, comparable mechanisms for exploring issues related to forensic evidence use are present in all jurisdictions.

Nonexpert, or lay witnesses testifying in courts that follow the Anglo-Saxon model – the "adversary system" – are permitted, with some exceptions, to testify only to perceptions made with the senses: hearing, sight, and smell. Under court rules that decide what evidence is admissible, such witnesses are prohibited from offering opinions on the significance of their observations and perceptions, if knowledge of the perceptions alone will give the fact finder sufficient information to resolve the disputed facts. If the sense perceptions of a witness alone do not alert the fact finder to their significance in the case at bar, a witness who qualifies as an expert may be permitted to explain the meaning of the factual testimony and also testify to the conclusions that the expert has drawn from those facts or data.

An expert is a person who, by education, training, knowledge, or experience has gained a special skill, ability, or knowledge in a field and who is, by virtue of that status, permitted by a court to offer opinion testimony. Thus, in the application of forensic science to the purposes of the law, being an "expert" is not a job title, but rather the designation of a type of witness and the form that his testimony may take.

Admissibility of expert opinion testimony requires, as a foundation for the testimony, that there be a showing that the witness is qualified in the particular field in which he proposes to testify. In the Anglo–Saxon system of justice, the determination whether a proffered witness qualifies as an expert is a question of law for the trial judge to decide.

Courts worldwide have traditionally possessed a great deal of latitude in deciding (1) whether a discipline is a proper one for the offering of expert testimony and (2) whether the witness before the tribunal is qualified as an expert. It is in the United States that courts have developed, during the twentieth century, a series of specialized rules that purport to regulate the judge's exercise of discretion in making an admissibility decision on expert opinions. These legal developments have been studied, and are followed, in many other countries as well. It is common, today, to see Canadian, United Kingdom, or even continental European courts dealing with novel forensic issues to refer to United States court precedents relating to these topics. Therefore, it is important for forensic scientists everywhere to be familiar with these tests of admissibility.

Further Reading

Criminal Evidence Act 1898. Criminal Justice Act (2003). European Convention on Human Rights, Art 6.
Frye vs. United States 193 F. 1013 (DC Cir 1923).
William Daubert *et al.* v. Merrell Dow Pharmaceuticals Inc. 509 U.S. 579 (1993).

Related Articles

Discovery: Depositions
Discovery: Discovery Motions
Discovery of Expert Findings
Discovery in the United States: Civil Cases
Discovery in the United States: Criminal Cases
Daubert v. Merrell Dow Pharmaceuticals
Expert Opinion in Court: Civil Law Jurisdictions (France, Germany, Italy, and Spain)
Expert Opinion: United Kingdom, Canada, and Australia
Expert Opinion: United States
Frye v. United States
General Electric v. Joiner
***In Limine* Motions and Hearings**
Weisgram v. Marley

BARRY TURNER

Expert Opinion in Court: Civil Law Jurisdictions (France, Germany, Italy, and Spain)

Jurisdictions: Regulatory Similarities and Differences

While civil law jurisdictions differ from country to country in the manner in which they regulate the participation of experts in litigation, they also share many procedural characteristics. The discussion of the approaches in the following five countries can be taken as representative of both the similarities and differences in the utilization of expert evidence.

France

The purpose of the expert witness in France in the inquisitorial system of justice is ostensibly the same as in other jurisdictions. They are there to effectively explain matters, which are beyond the experience of the triers of fact, and to help them in their decision making. As in other legal systems, the expert is considered to possess knowledge outside the range of the layman.

The inquisitorial system differs from that of the common law adversarial jurisdictions in that the trial commences following an exhaustive analysis of evidence during the investigation by a judicial official referred to as the *juge d'instruction* (examining magistrate). This is the process followed when serious crimes are investigated. In less serious cases, evidence of the offences will be gathered and assessed by the law enforcement agencies under the guidance of a *procureur* or state prosecutor. The evidence is subjected to analysis and is examined as to its relevance and admissibility. The decision-making processes are guided by a search for proof of what is asserted and while the legal burden lies with the party affirming rather than the party denying, as in common law jurisdictions there may be a greater reliance on presumptions to support the assertions. It may not always be necessary for witnesses to discharge the burden of proof.

There are a number of facts that may be presumed. They may be assumed because of their inherent probabilities (*presumptio judicis*). In such a case, the judges, using their own intellect and knowledge, determine the probability of a fact being true. There are presumptions of law where the law presumes unless evidence to the contrary is proved (*presumptio juris*) and presumptions required by law, where the law disallows any attempt to prove the contrary. In the common law tradition, this would be the irrebuttable presumption (*presumptio juris et de jure*).

In civil procedure, the evidence is presented to the court in the form of a *dossier* or file, which contains all the material collected to form the basis of the judgment. This material gathered in the investigatory stage of the process includes all the evidence in the form of summaries of oral testimony, expert reports,

and pleadings of the parties. This is also the case in criminal trials as in commercial and administrative cases. The expert's report is there to assist the trier of fact in his decision making.

The French procedural codes allow for the adducing of oral testimony, but it is unusual to do so except in criminal cases, where an expert may be required to give such testimony in court. Ordinarily, though, the judges will rely on written submissions. Expert evidence may be requested by either of the parties to the case or requested directly by the judge. The expert will provide the court with a provisional report, which may be later amended into a final report, for inclusion in the dossier once it has been amended by reference to facts or legal necessities. Where expert evidence is requested by either of the parties, it may be contested by other evidence. In the case of a court-appointed expert, the evidence is considered to be conclusive and is not open to challenge.

The position of the expert in France underwent a significant change in 2004 with the adoption of statute no. 2004-130 of 11 February 2004 and its subsequent implementing regulations, no. 2004-1463 of 23 December 2004.

This legislation impacted upon two different areas of the status of the expert. In the first instance, it was designed to codify the conditions for registration as a court expert and secondly examined practical matters dealing with the courts' ability to exercise quality control when appointing experts.

Experts are now usually appointed from official lists but the judge retains the discretion to appoint any person he considers to be the best person.

The expert is registered on a list for the area in which he lives or works and may also be listed on the final appeal court list (*Cour de Cassation*). These provisions do not apply in the administrative courts.

Under the old statute no. 71-498 of 29 June 1971 and its implementing regulations no. 74-1184 of 31 December 1974, there was no "quality control" or adequate monitoring of the abilities of experts. It was the 2004 statute that introduced these measures.

Germany

While German courts are not strictly civil courts in structure or function, they more easily lend themselves to this philosophy than to the one pertaining to adversarial courts.

In the German criminal process, the function of a witness is to prove factual evidence, which may be relevant for a conviction. Witnesses may be asked to provide evidence on matters they have personal knowledge of or even their observations on matters. This is a striking difference to the role of a witness in the adversarial system where conjecture and speculation is inadmissible and opinion evidence is confined to expert testimony to the greatest extent. The witness in the German criminal justice system has duties codified in the *Strafprozeßordnung* (*StPO or* Code of Criminal Procedure). He is obliged to appear in court before the public prosecutor, must give evidence under oath and must tell the truth subject to the purpose of the examination. Giving false testimony is punishable under the Code of Criminal Procedure. The law on requiring the appearance of witnesses is much broader than that employed in common law courts, but a potential witness has a range of circumstances under which he may refuse to give evidence.

In Germany, the *Sachverständige* or expert witness is appointed by the court under §§ *et seq* 72 StPO. They are, as in most systems, required when some evidence needs explaining to the court because the subject material or content is beyond its understanding. The expert's duty is similar to that of the lay expert and, in their case, has a duty to prove facts and circumstances by reference to their particular expertise. As in other jurisdictions, the expert is allowed to make reference to scientific or other scholarly protocols in order to carry out the necessary research or experimentation, and may refer to published material in the area of expertise. As with ordinary witnesses, the expert can refuse to give evidence or to provide an opinion. Experts in German courts can be refused leave to give their evidence on the grounds of prejudice where prejudicial effect may outweigh the probative value of the expert opinion. This prevents the expert from overly influencing the final decision and reserves the ultimate issue to the trier of fact.

Under the *Zivilprozeßordnung ZPO* (Code of Civil Process), the expert's role is to provide the court with a nonpartisan assessment of the facts at issue based on their expertise on the subject matter. While the expert's opinion is classed as formal evidence by §§ 402 *et seq* ZPO and will be assessed as to its merit by the judge in making an ultimate decision, the evidence is not binding on the court. The court may

take into account all evidence submitted by all parties and witnesses and retains the discretion in finding the facts, including those that are the subject of the expert's evidence. The expert is not to be understood to be a substitute for the court's decision-making powers. Appointment by the court of an expert has, as its primary purpose, to allow the court access to a person who possesses specialist knowledge. Such appointment does not constitute a delegation of the court's decision-making duty on the ultimate issue.

Experts are appointed by the courts or they may be appointed by consent of the parties. Where evidence is adduced by an expert appointed by one party only, it is seen simply as an assertion by that party and does not possess any special merit. Expert evidence can be the domain of a public authority as well as individuals. In the case of an individual, the evidence may be regarded as "expert" or documentary. The opinions of a public authority are always considered as expert opinions, as is the opinion of a body of experts such as those working for a forensic or medicolegal institute. § *et seq* 404 ZPO requires that, where available, public authority experts are to be given priority.

German courts have a preference for expert testimony to be in the form of a report but they may be called to give oral testimony where required.

The courts may not interfere with or interpret partially the expert's evidence, since the expert is neutral in the proceedings. However, under § 404a ZPO, the court is allowed to define the parameters within which the evidence may be adduced and does, of course, retain final decision on whether an expert may or may not give evidence. Experts may be rejected on grounds of bias or prejudice and may not testify, if they are connected with either of the parties by § 41 and 42 ZPO.

Italy

The expert in Italy can be called to give technical advice to either the judge or the parties. They give their evidence during the inquisitorial stage of the proceedings.

There are two kinds of experts in civil proceedings. The parties may appoint experts to assist them or the judge may appoint an expert as an assistant of the judge. In the latter role, the expert will be directed by the judge to engage in an analysis of material where the litigation requires specialized knowledge or expertise. Experts are commissioned by the judge to carry out inquiries and to present to the judge the report, clarifying any technical issues on which the judge needed assistance. When testifying the judge will delineate specific questions, which will be answered in the experts report.

The expert remains in an advisory role rather than as a witness proper. Their purpose is to provide a report on technical evidence to assist in the analysis of the evidence where, without special expertise, the evidence would be difficult to understand. The report provided by the expert is not binding on the judicial decisions.

The court-appointed expert, although strictly an officer of the court rather than a witness, gives evidence under oath.

When the court appoints an expert, the parties may appoint their own experts to assist them in understanding the court-appointed expert's report. Experts retained by the parties may also file a report allowing them to comment or criticize the court-appointed expert's findings. It is unnecessary for the parties to require the court to formally appoint their experts and all they need to do is give notice to the court clerks.

Under Article 61 of the Italian Code of Civil Procedure, the court-appointed expert must be chosen from a register categorizing the specific scientific, technical, or professional expertise of prospective appointees. Each of the courts has its own list known as an *Albo dei Periti* and the individuals on it are managed by the President of the Court.

Strict rules apply to ensure a court-appointed expert is impartial and unconnected in any way with any of the parties. Similarly, the expert may not be involved in any other proceedings, which would, by their nature, involve a conflict of interest or litigation on the same issue. The Supreme Court has stated that the expert is guaranteed neutral because they are appointed in the capacity of assistant to the judge who is naturally neutral in the proceedings. The Supreme Court stressed this position by also referring to a court-appointed expert's discretion permitting him to abstain or object under the provisions of the Italian Code of Civil Procedure.

Although an expert has discretion to abstain from the appointment, a solid foundation for doing so must exist, such as conflict of interest or unsuitability for the inquiries necessary. It is then up to the judge to decide whether the abstention will be accepted.

After a court-appointed expert has been designated, the parties may still object to the court-appointed expert if they have reason to believe the expert lacks impartiality or that a conflict of interest is present.

The Netherlands

In Netherlands, the expert witness opinion is initially considered a part of the court inquiry rather than as *prima facie* evidence of fact. This does not exclude the expert's evidence being eventually admitted as evidence of fact but acts to protect the reservation of the ultimate issue as remaining in the exclusive domain of the trier of fact. However, if a court decides to disregard an expert's evidence, it must incorporate the reasons for doing so, into the final judgment.

By the Netherlands Code of Civil Procedure, experts may be appointed either at the request of the parties or on the court's own initiative. The court enjoys considerable discretion in the appointment of an expert and, where the judge believes that his or her own expertise is sufficient to an understanding of the evidence, he may refuse leave to appoint. This follows the standard pattern seen in many jurisdictions where unnecessary experts may be seen as actually prejudicial to the proceedings similar to the position held in the English case of *R v. Turner* [1975] QB 834. Where a court decides to appoint an expert, it is by interlocutory proceedings.

When the court agrees to the appointment of experts, the parties may choose from a list of potential experts. The court will then exercise its discretion again and appoints one suitable to both parties. This is also done by interlocutory order, which delineates the matters that the expert is to comment on. The interlocutory nature of this appointment and the appointment itself are not open to appeal by the parties.

Upon acceptance of the appointment, the expert is required to carry out an impartial analysis of the evidence. Changes to the rules from 1 January 2002 allows for the hearing of noncourt appointed experts with agreement of the court. In order to maintain equality of arms where this is permitted, the other party may also call a nonappointed expert.

The interlocutory orders direct the expert to produce an opinion in writing and to prepare for further oral examination, if the court needs clarification on anything contained within the written report. Where an expert is required to supplement the report with oral testimony, the court registrar will make a written record of the testimony. Under these circumstances, the first report is treated as a draft, which the parties may accept, rebut, or ask for further and better particulars. The expert will then issue a final report to which the parties may respond.

Parties to proceedings may themselves ask for a preliminary expert report. This may be used to assist them in formulating their case. Since this takes place prior to the commencement of proceedings, the court has no discretion. However, the court may dismiss such a request for a preliminary report where it can be shown that to allow it would involve an abuse of process or waiver of rights.

Spain

In Spain, evidence of experts is categorized by regulation, which came into force on 8 January 2001, as opinion offered by a person with special scientific or other specialized knowledge, not commonly understood by laypersons. The expert's contribution to the trial process is advisory and is designed that the court may be helped to understand complex issues in order to evaluate their value as evidence or relevant facts.

Experts may be appointed by the judge or by the parties. In the case of party-appointed experts, they are referred to as *expert witnesses* (*testigos expertos*). Experts appointed by the judge are described as judicial experts (*perito judicial*). There are fundamental legal differences between the two types of expert; these affect the way they are nominated (*nombramiento*) in the first instance and the way they may participate (*participación*) in the proceedings.

Experts for the Parties

Both parties may instruct an appropriate expert with the special scientific or technical knowledge required to assist the judge in understanding the relevance of issues to the dispute. In Spain, as in many other jurisdictions, formal qualifications are not necessary to establish expertise; however, in many areas, it is likely that a lack of formal qualifications such as degrees in scientific subjects may reduce the credibility of any evidence given. Judges are more likely to give weight to an expert's evidence where they are in possession of formal qualifications. From the perspective of the parties, they may also need to consider

the experts' experience in evaluating complex scientific issues within the context of judicial evidence.

Experts may be discredited within the proceedings or disqualified where conflicts of interest are discovered. Either party may raise opposition to the expert called by the other where their participation would materially affect the neutrality required in the litigation process. If a party intends to object to the nomination of an expert, it must be done before the hearings of the preliminary directions. The objecting party has the burden to show evidence that the expert is unsuitable.

The expert will produce a report, which must be filed accompanying the preliminary documents. This will be the statement of case or statement in defense. Only exceptionally will it be permitted to file the expert report after the parties serve their respective documentation. In such cases, the party is required to inform the judge that it intends to adduce expert evidence after commencement of proceedings. This may be the case where an expert report is required following examination of the contents of the statement of case or statement of defense. There are time limitations on the introduction of evidence after commencement of the proceedings.

Where the judge has given leave to an expert witness, following a request in the preliminary documents, they may participate in the trial by answering questions of the parties in order to fully explain their opinions. In the alternative, they may simply give an opinion on one single issue or may be required to expand on their explanation. They may also expect to have their opinions subjected to criticism. The judge may also ask questions of the expert witness but may not require them to expand on the statements in their preliminary submissions. There is no legal direction on cross examination of expert witnesses but it will be allowed on request.

The expert's fees will be paid by the party losing the litigation and these form part of the judicial assessment of costs.

The Judicial Expert

The parties may request that the judge nominate an expert, which they must do in their preliminary submissions. If the judge consents, then this nomination should take place within five days of the request.

The parties are invited to agree on experts taken from available lists or they may even draw straws to settle the nomination. The experts are to be found on various lists of professionals, academics, and institutional experts. Unlike the parties' expert witnesses, judicial experts must possess a recognized academic degree or appropriate professional qualification demonstrating specialized knowledge in the area from which the opinion is sought. Inevitably, this results in the judicial expert having greater credibility than experts called by the parties.

Once appointed, the parties have no control over the expert and they are advised to be diligent in selecting an appropriate person from the lists. The expert will apply his own understanding of the frequently complex issues and is under no obligation to take instruction from the parties.

Judges may object to the nominations of an expert. They may also object where there is a conflict, such as a professional relationship between one party and the judicial expert. It is possible to object to the expert either at the preliminary nomination stage or during the trial. The judicial expert's participation in the trial constitutes the production of a written opinion within a period determined by the judge. It is often the cases that the report is not complete within the time frame, and cases often need relisting.

The judicial expert appears at the trial when requested by the parties. Another notable difference between this judicial expert and experts appointed by the parties is that the judicial expert may be requested by the judge to expand on his or her original report.

The cost of the judicial expert is usually borne by the losing party but the judges have discretion to vary this and, where no specific direction is made, it is borne equally by the parties.

Related Articles

Expert Opinion in Court: a Comparison of Approaches

Expert Opinion: United Kingdom, Canada, and Australia

BARRY TURNER

Expert Testimony: Admissibility, Comparison *see* Expert Opinion in Court: a Comparison of Approaches

Expert Testimony: Jury Instruction ***see*** Jury Instructions on Expert Testimony

Expert Testimony: Mitigating Factors in Capital Crimes ***see*** Mitigation Testimony

Expert Witness: Cross Examination ***see*** Cross-Examination of Experts; Direct Examination of Experts

Expert Witness: Testimony ***see*** Direct Examination of Experts

Expert Witness: Who Is?

In adversary systems (*see* **Adversary Systems of Justice**), that is, jurisdictions that follow the Anglo-Saxon model (United Kingdom, Commonwealth countries, Canada, and the United States), an expert is a witness who possesses those qualifications that permits him to offer to the court not only observational information but also formulate opinions and draw conclusions on the basis of an examination of forensic data using knowledge 'beyond that of the average juror' (*see* **Expert Opinion: United States**). Thus, being an expert is not an occupational title, but rather a special category of witness in legal proceedings who is permitted to offer information to the court that ordinary fact witnesses are not permitted to express.

> ''The opinion expressed by an expert witness in any branch of technical science depends for its effect for on, inter alia, his qualifications, skill and experience in that science. If it appears to be based on sufficiency of research directed accurately and relevantly to a particular issue and to be so supported as to convince a court of its fundamental soundness and applicability to the particular issue, a court is entitled, although not obliged, to accept it even if unsupported by any corroborative expert opinion." (Davie v Magistrates of Edinburgh, 1953, SLT 54; 1953 SC 34.)

Qualifying the Witness Is the First Step

To be considered an expert authorized to offer opinion testimony to a judicial tribunal, the witness must first be examined as to his qualifications. This involves demonstrating that the witness possesses the special abilities to engage in the work pertinent to a relevant forensic discipline. The witness's abilities may have been obtained through formal and/or informal study, work-related experience, pursuit of continuing education programs designed to maintain and advance one's skill and knowledge, and the participation in professional societies (*see* **Federal Rule of Evidence 702**).

When a particular discipline or profession has specific academic or licensing requirements, possessing the necessary diplomas or degrees, in addition to the required certification if such is the norm, are all factors that the court will take into consideration in determining of the witness qualifies as an expert. Although many bodies have been created to enhance or endorse experts' professional qualifications, most courts will reserve to themselves the decision as to who or what will qualify as an expert or expertise (*see* **General Acceptance Test for Novel Expert Evidence**; ***Kumho Tire v. Carmichael***).

The Judge Decides When Witness Is Expert

When, after the preliminary qualifying process, the trial judge becomes convinced that a witness satisfies the legal requirements of skill and knowledge, the witness will be recognized as an expert. Whether a proffered witness is qualified is a legal decision for a judge to make. The judge has a great deal of discretion in making this decision and, on appeal, will not be reversed on the basis of a finding that a witness either was or was not qualified as an expert, unless a serious abuse of discretion is shown to have occurred.

Once a witness has been qualified to give expert testimony, his qualifications may still be challenged by the opposing litigant during cross-examination (*see* **Direct Examination of Experts**, **Cross-Examination of Experts**). The background and qualification of the witness affect his credibility. The believability of any witness, including that of an expert witness, remains an issue that the jury must determine. The jurors will indeed be instructed in the law to that effect (*see* **Jury Instructions on Expert Testimony**).

The extent of qualifications required for a particular specialty is highly dependent on those required for entry and advancement in each chosen forensic profession.[a]

End Notes

[a.] See each separate discipline for the competency and certification requirements of its practitioners and examiners, if any such requirements exist.

ANDRE MOENSSENS

Expert Witness Testimony: Admissibility *see* *Daubert v. Merrell Dow Pharmaceuticals*

Expert Witnesses: Selection and Investigation of Credentials

Introduction

Expert testimony is increasingly important to attorneys. The field of scientific interpretation of evidence and its portrayal in popular media has progressed to the point that expert testimony at trial is not only commonly accepted by judge and jury alike, but expected [1]. This article addresses the factors that influence the selection of an expert, including the importance of investigating the expert's credentials and making an informed assessment of the credibility and expertise that the expert's qualifications will project to a judge and jurors.

While this article is written from a US perspective, many of the issues discussed here are applicable to expert witness testimony worldwide. This article will not, however, address the legal standards for the admissibility of expert testimony [2], or attorneys' ethics in dealing with experts [3].

Selecting an Expert

Many variables should be considered in selecting an expert witness, including the expert's availability, cost, experience, and reputation. When an expert serves as a consultant or in the pretrial phases of litigation, the criteria for selecting that expert may be limited to the expert's competency in the field. As a trial witness, however, the expert's integrity, charisma, and overall effectiveness as a witness must also be considered.

Thus, consideration should be given not only to the expert's formal training but also to the expert's personality, demeanor, and capacity to organize, express, and interpret complex concepts for the jury. The weight accorded to the expert's opinion by the judge or jury will be determined in large part by the expert's perceived character, objectivity, and impartiality.

Of course, the quality of the expert's credentials remains an important factor to consider as well.

A thorough evaluation of an expert should take into account such matters as (i) the membership requirements of the associations to which the expert belongs; (ii) how the expert's credentials compare to those of the opposing expert; (iii) whether the journals in which the expert's articles appear are held in high regard in the field; and (iv) whether the conclusions in those articles were subject to peer review. Studies of jurors' perceptions of experts can be particularly helpful in guiding this evaluation.

Care should always be taken to verify the credentials of one's own expert, as well as those of an adversary's expert, for although it is unlikely that an expert has faked credentials, it is not unheard of. Indeed, experts have come under increased scrutiny in recent years for fabricating or inflating their qualifications.

Qualifications

The court must determine whether a proffered witness is qualified to testify as an expert, and that determination will not be overturned except in the case of an abuse of discretion [4].[a] Federal Rule of Evidence 702 states that a witness may qualify as an expert on the basis of knowledge, skill, training, experience, or education. An expert witness must possess only one of these traits for the judge to find the expert qualified to give an opinion. In making this evaluation, the judge may consider the expert's educational background, work experience, publications, awards, teaching, speaking, or other professional engagements, prior expert witness testimony, and membership in professional associations.

Often, the expert may have to educate the attorney proffering the expert regarding the significance of particular experience, achievements, and certifications to ensure that they are appropriately emphasized to the judge. An expert must be prepared to explain board certification and licensure requirements to the judge in detail.

Experience as an Expert Witness

Experience and training are often more significant than academic background and are accorded more weight by jurors, according to a jury study evaluating fingerprint experts [5]. However, experience as an expert witness, by itself, does not qualify someone as an expert in later cases. For example, in *Bogosian v. Mercedes-Benz of North America Inc.* [6], the court rejected the opinion of a witness who had testified as an expert 126 times. Another court noted that, "it would be absurd to conclude that one can become an expert by accumulating experience in testifying" [7]. Conversely, a lack of previous experience as an expert witness does not disqualify one from testifying as an expert [8].

Education and Training

An expert may be qualified on the basis of academic credentials, including the expert's undergraduate, graduate, and postgraduate work. An expert's academic credentials should only be issued by accredited educational institutions and programs [9], because the proliferation of the Internet, while laudable for so many reasons, has also rekindled the old-fashioned diploma mill. One such business, Deplomas 4U, once provided bachelors, masters, MBA, or PhD degrees in its customers' field of choice; advertisements assured that no one would be turned down and that there would be no bothersome tests, classes, books, or interviews.

An expert should continuously perform research and publish in the expert's field, preferably in peer-reviewed publications. Teaching experience is another of the qualifications that judges will evaluate: all forms of teaching – regular, speciality, guest lecturing, visiting professorships, continuing education, and short courses – weigh in as credentials. An expert should also keep up to date with developments in his or her field of expertise by reading the current literature, enrolling in continuing education seminars, joining professional societies, and attending professional meetings.

Membership in Professional Associations

A study published by the US Department of Justice in 1987 found that jurors perceived those fingerprint experts who belonged to professional associations to be more credible than other experts, and presumed experts would belong to such groups.[a] It is therefore important for an expert to remain active and participate in professional societies; the expert's credibility is diminished if the expert has not recently attended a professional meeting. Professional associations that only require annual dues payment to become a member are not as prestigious as associations that are

joined by special invitation only, by approval of special referees, or by passing an examination.

Thus, an expert should be selective about which professional associations to join. The *Wall Street Journal* once noted that the American College of Forensic Examiners (ACFE), for example, had "mail-order" credentialing, for which applicants needed to pay a fee and pass an ethics exam; there were no examinations for board certifications in various specialties during ACFE's "waiver of examination" periods – all one needed to do was verify possession of 200 "experience points" and send in a fee [10].

A 2001 *New York Times* article discussed the case of a doctor who was board-certified in forensic medicine by the ACFE's sister organization, the American Board of Forensic Medicine (ABFM), while serving time in prison for the manslaughter of his wife [11]. In his application, he stated that he had never been convicted of a felony and that he was on sabbatical from his job. The business address he provided was that of a state prison. After learning of the doctor's conviction from the *Times* reporter, the ABFM stated that it would strip the doctor of his membership.

One expert, Dr. Steve K. D. Eichel, exposed the ease with which some organizations grant questionable credentials by obtaining board certification by the American Psychotherapy Association for his cat Zoe [12]. Despite the fact that Dr. Eichel planted numerous hints about Zoe's true identity in the curriculum vitae he sent to credentialing bodies, Zoe's tale was not publicly exposed until a reporter requested an interview with Dr. Zoe D. Katze regarding an article on hypnosis: "Dr. Katze" was listed on the American Association of Professional Hypnotherapists' website.

Increased Scrutiny of Experts

Experts have come under increased scrutiny for either fabricating or inflating their qualifications. For instance, in *State v. Ruybal* [13], the state's expert – an FBI agent who specialized in forensic serology – stated that he held a master's degree in biology when in fact he held only a bachelor's degree. Upon further examination, it was discovered that the expert had falsified other reports and lied under oath concerning tests in another case.

In Florida, in 1998, a person who had been testifying as an expert in toxicology for three years for both the prosecution and defense in criminal cases was prosecuted for perjury for testifying with fraudulent credentials [14]. The expert claimed to possess masters and doctorate degrees from Florida Atlantic University, but when a prosecutor sought to confirm the claims, he discovered that the registrar's office had no record of the expert attending or receiving a degree from the university. In fact, the university did not even offer a PhD in the program from which the expert claimed to graduate. The expert had also distributed a copy of his purported master's degree dated 1971 and signed by "Lawton Chiles, Governor" Unfortunately, Chiles was a US Senator in 1971, and did not serve as Florida's governor until 1991. The "expert" eventually pled guilty to three counts of second-degree perjury (one count for each first-degree murder case in which the expert testified as a defense witness), and received a three-year prison sentence.

In addition to perjury prosecutions for false qualifications, some jurisdictions also prosecute for academic fraud. For example, in Florida, a person who misrepresents association with, or academic standing at, a postsecondary educational institution is guilty of a first-degree misdemeanor [15].

In another case, a Harvard medical professor was sued for trademark infringement for falsely claiming to be board-certified by the American Board of Psychiatry and Neurology (ABPN) in five trials [16]. The board sought to seize the expert's witness fees and treble damages, but the court denied that relief because it believed the expert was unlikely to infringe in the future.

In *In re Vioxx Products* [17], the court granted the plaintiff a new trial in her product liability action when it was discovered that the pharmaceutical company's cardiology expert had misrepresented his credentials by testifying that he was board certified in internal medicine and cardiovascular disease when in fact those certifications had expired.

Courts have also overturned convictions where the experts testified outside their field of expertise. For example, in *Gilliam v. State* [18], the court held that the medical examiner was not qualified as an expert in shoe-pattern analysis; therefore, it was error for the trial court to allow her to testify that the defendant's sneaker left marks on the decedent's body.

There is evidence to suggest that, since the US Supreme Court's decisions in *Daubert* [19], and *Kuhmo Tire Co.* [20], courts have been more willing

to exclude expert testimony. The Federal Judicial Center compared a 1998 survey of 303 federal judges with a 1991 survey [21]. In 1998, 41% of the judges claimed to have excluded expert testimony, whereas only 25% of the judges did so in 1991. A 2001 RAND study similarly concluded that judges were becoming more vigilant gatekeepers; for example, in the US Third Circuit Court of Appeals, the exclusion rate in products liability cases rose from 53 to 70% [22]. This contradicts most of the reported case law following *Daubert*, which seems to indicate that the exclusion of expert testimony remains the exception, not the rule [23].

Weight of the Evidence

Once a judge decides that an expert may testify, the jury must then decide the weight to accord the expert's opinion. Expert witnesses and attorneys should be aware of the studies regarding jurors' perceptions of expert witnesses, which have shown that jurors give great weight to expert testimony, and reveal jurors' evaluations of the experts' qualifications, appearance, demeanor, and communication skills.

In 1978, Rich Tanton published a study in the *Journal of Forensic Sciences* in which he found that jurors held stereotypical views of experts [24]. For example, the male expert was expected to be 44 years old, upper-middle class, white, neat, intelligent, wearing a dark suit, and glasses, professional, calm, and serious; the female expert was expected to be white, 37 years old, dressed conservatively, neat, pleasant, and honest.

In 1984, Saks and Wissler conducted a telephone survey of potential jurors [25], in which respondents were asked the following: (i) whether they would believe testimony from the experts; (ii) whether they perceive the testimony to be honest; and (iii) whether they perceive the witnesses from different fields to have adequate experience to testify. The study concluded that physicians, chemists, and firearms experts ranked the highest in believability, honesty, and experience. Next-highest ranked were accountants, psychiatrists, psychologists, and eyewitnesses. Lowest-ranked were police officers, handwriting experts, and polygraph examiners.

Dr. Joseph Peterson conducted three studies on forensic evidence and the courts from 1984 to 1987. His studies concluded that jurors accorded great weight to expert testimony and found that in criminal cases in which expert witnesses testified for the prosecution, jurors were more likely to render guilty verdicts [26–28].

In 1992, the *National Law Journal* conducted a survey of jurors and their views of the jury system, and concluded that jurors were influenced by expert witnesses and accorded their opinions great weight [29]. In the civil and criminal cases surveyed, 89% of the jurors thought the experts were believable, and 71% said the experts' testimony made a difference in their verdict.

A 1994 study of preconceived notions of experts consisted of an extensive survey of American lawyers, judges, jurors, and expert witnesses [30]. It revealed that the characteristics of experts that were most important to jurors in determining the experts' credibility were (i) the expert's willingness to draw firm conclusions and (ii) the expert's ability to convey technical information in plain language that a layperson could understand.

Another study concluded that an expert's believability is linked to the expert's qualifications, familiarity with the facts of the case, good reasoning, and perceived impartiality [31]. Jurors were also influenced by independent research that corresponded with the expert's opinion.

A 1997 study of jurors' perceptions of expert witnesses in death penalty cases by Scott F. Sundby found that defenses based solely on expert testimony are likely to fail, but defenses that integrate expert testimony with persuasive lay testimony are more likely to prevail [32].

A 1998 *National Law Journal*/DecisionQuest Study exposed jurors as a more skeptical, cynical group [33]. Among the findings, the study concluded that 50% of those surveyed thought that expert witnesses say only what they are paid to say; 33% did not believe police testimony; and 75% said they would set aside what a judge says the law requires and reach a verdict the jurors felt was right.

A 2001 study demonstrated that mock jurors almost always sided with court-appointed experts except when the experts' opinion favored corporate defendants [34].

A 2007 study concluded that using expert testimony to counter the prosecution's expert in criminal cases caused jurors to be skeptical of all expert testimony, rather than simply sensitizing them to flaws in the prosecution expert's testimony [35]. In fact,

jurors rendered more guilty verdicts when they heard defense expert testimony than when they did not. This study throws into question the Supreme Court's assumption in *Daubert* that opposing expert testimony effectively safeguards against "junk" science in the courtroom.

Increasing awareness of errant experts and exonerations of the wrongly accused has influenced how jurors perceive scientific evidence. For example, background beliefs about the possibility of laboratory errors and intentional tampering affect the weight jurors afford a DNA report [36], and jurors with such beliefs gave probability estimates less weight. A separate poll regarding forensic fraud and its impact on potential jurors found that 32% think wrongful convictions happen frequently; 23% said that wrongful convictions are rarely an accident [37].

Today, 40% of the US jury pool consists of "Generation X" and "Generation Y" jurors. GenXers were born between 1966 and 1981; Generation Y members, or "Millennials" as they are sometimes known, were born after 1981.

In order to communicate most effectively with these generations of jurors, it is important to understand defining experiences that have influenced their lives. GenXers grew up with the television as a babysitter, no institutional stability, high divorce rates, gangs, and the specter of acquired immunodeficiency syndrome (AIDS). They grew up fending for themselves and therefore tend to be cautious, skeptical, and practical; they do not readily trust institutions and show little deference to authority. Generation X is also the first generation to grow up computer-literate; therefore, GenXers are media-and technology-savvy and expect to be entertained in the courtroom [38].

Millennials were also raised with cable television, high-speed internet, and increasingly elaborate and complex video games, and consider such phenomena as mobile phones, instant text messaging, and the 24-hour news cycle as natural parts of everyday life [39]. As a result, they expect "all information available all the time in a constant and instant stream," yet have increasingly short attention spans [40]. Millennials are accustomed to getting their information visually, graphically, and in 10–30s sound bites. Unlike GenXers, Millennials tended to grow up in an over-scheduled, heavily structured environment with "helicopter" parents who hovered about, guiding and supervising the Millennials' social and educational development. As a result, Millennials tend to be less independent than their GenX counterparts, and expect more personalized attention.

Experts must understand that effective communication with these jurors requires organized content, and the effective use of visual presentation techniques including, whenever possible, demonstrative exhibits that incorporate large, user-friendly data presentation monitors and systems for use both by the court and individual jurors, as well as interactive electronic timelines and e-documents that allow jurors to feel they are in control of and have access to all information regarding the facts of the trial. Data also suggests that when testifying to these jurors, experts should attempt to associate themselves with a more collaborative, personalized role such as a teacher, rather than a more hierarchal and impersonal profession such as a scientist [41].

National Law Journal/DecisionQuest juror outlook surveys confirm these conclusions distinguishing jurors from Generations X and Y from past generations [42]. For example, while 64% of jurors overall believe the police tell the truth when they testify, only 51% of jurors aged 18–24 years old share that belief. 60% overall and 72% of those aged 18–25 viewed presentations using videos, simulations, and computers positively.

Conclusion

Expert testimony will continue to play an important role in the future. Expert witnesses have been facing increased scrutiny in the US and worldwide. For more effective expert testimony, lawyers and experts must be aware of the factors that the courts will evaluate in order to determine whether an expert is qualified or not, as well as jurors' changing perceptions of experts.

End Notes

[a.] Juries, Fingerprints, and the Expert Fingerprint Witness, supra.

References

[1] 'CSI Effect' Has Juries Wanting More Evidence, USA Today A1 (Aug. 5, 2004).
[2] Fed. R. Evid. 702; Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).
[3] ABA Model R. Prof. Conduct 3.3, 5.3. (2004).
[4] Kumho Tire Co. v. Carmichael, 526 U.S. 137 (1999); but see, e.g., Radlein v. Holiday Inns, Inc., 971 So.2d

1200 (La. App. 4 Cir. 2007). (Holding that the trial court's decision will not be reversed unless there is a clear showing of error.).

[5] Illsley, C. (1987). Juries, fingerprints, and the expert fingerprint witness. *Presentation at the International Symposium on Latent Prints*, at the FBI Academy, (July 1987).

[6] Bogosian v. Mercedes-Benz of North America Inc., 104 F.3d 472, 477 (1st Cir. 1997).

[7] Thomas J. Kline, Inc. v. Lonillard, Inc., 878 F.2d 791, 800 (4th Cir. 1989),. cert. denied, 493 U.S. 1073 (1990).

[8] U.S. v. Locascio, 6 F.3d 924, 937 (2d Cir. 1993). ("even the most qualified expert must have his first day in court"), cert. denied, 511 U.S. 1070 (1994).

[9] The Technical Working Group on Education and Training in Forensic Science (2004). *Education and Training in Forensic Science: A Guide for Forensic Science Laboratories, Educational Institutions and Students*, National Institutes of Justice Special Report (June, 2004). (This guide is a valuable resource for evaluating forensic science academic programs as well as training and continuing education programs. The Forensic Science Education Programs Accreditation Commission (FEPAC) is a committee of the American Academy of Forensic Sciences. It develops and maintains standards and administers an accreditation program that recognizes and distinguishes high quality undergraduate and graduate forensic science programs.).

[10] McDonald, E. (1999). The making of an expert witness: it's in the credentials, *Wall Street Journal* B1.

[11] Hansen, M. (2000). Expertise to go, *ABA Journal* (Feb.) **86**, 44.

[12] Hansen, M. (2002). See the cat? See the credentials? (Oct.), *ABA E-Journal* **1**, 41.

[13] *State v. Ruybal*, 408 A.2d 1284 (Me. 1979); *see also People v. Cornille*, 448 N.E.2d 857 (1983). (New trial ordered where defense discovered that prosecution arson expert gave false credentials.).

[14] *See* Fitzgerald, H. Jr (1998). *Phony "Expert" Jailed for 3 Years*, Sun-Sentinel 3D (Dec. 1, 1998).

[15] Fla. Stat. § 817.566 (2004).

[16] ABPN v. Johnson-Powell, 129 F.3d 1 (1st Cir. 1997).

[17] In re Vioxx Products, 489 F. Supp. 2d 587 (E.D. La. 2007).

[18] Gilliam v. State, 514 So.2d 1098 (Fla. 1987);. *see* also Kelvin v. State, 610 So.2d 1359 (Fla. App. 1 Dist. 1992). (Finding evidence technician not qualified to give expert testimony about the trajectory of bullets depicted by dowels stuck into bullet holes in a sofa at a crime scene, as he was not a crime-scene reconstructionist and had no training in ballistics.).

[19] Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).

[20] Kuhmo Tire Co. v. Carmichael, 526 U.S. 137 (1999).

[21] Hansen, M. (2001). Admission tests: fewer post-daubert judges allow experts to testify without limitations in civil trials, study finds, *ABA Journal* (Feb.), **87**, 28.

[22] Dixon, L. & Gill, B. (2001). *Changes in the Standards in Admitting Expert Evidence In Federal Civil Cases Since Daubert Decision*, RAND Monograph.

[23] See, Fed. R. Evid. 702, Adv. Cmte. Note to the 2000 Amendment. (2000).

[24] Tanton, R. (1978). Jury preconception and their effect on expert scientific testimony, *Journal of Forensic Science* **24**, 681.

[25] Michael Saks & Roselle Wissler (1984). Legal and psychological bases of expert testimony: surveys of the law and jurors, *Behavioral Science and Law* **2**, 435.

[26] Petersen, J. *et al.* (1984). *Forensic Evidence and the Police: The Effects of Scientific Evidence on Criminal Investigations*, National Institute of Justice Research Report (Oct. 1984).

[27] Petersen, J. *et al.* (1986). *Forensic Science and the Courts: The Uses and Effects of Scientific Evidence in Criminal Case Processing*, Chicago Center for Research in Law & Justice, University of Illinois, Chicago.

[28] Petersen, J. (1987). *Use of Forensic Evidence by the Police and the Courts*, National Institutes of Justice Research in Brief.

[29] Cheever, J. & Naiman, J. (1993). The view from the jury box, *National Law Journal* **15**, S1–S16.

[30] Shuman, D.W. *et al.* (1994). An empirical examination of the use of expert witnesses in the courts – part II: a three city study, *Jurimetrics* **35**, 193.

[31] Shuman, D.W. *et al.* (1996). Assessing the believability of expert witnesses: science in the jurybox, *Jurimetrics* **37**, 23.

[32] Sundby, S.F. (1997). The jury as critic: an empirical look at how capital juries perceive expert and lay testimony, *Virginia Law Review* **83**, 1109.

[33] Aronson, P. *et al.* (1998). Jurors: a biased, independent lot, *National Law Journal* (Nov. 2) **21**, A1.

[34] Cooper, J. & Hall, J. (2001). Reaction of mock jurors to testimony of a court appointed expert, *Behavioral Science and Law* **18**, 719.

[35] Levett, L.M. & Kovera, M.B. (2007). The effectiveness of opposing expert witnesses for educating jurors about unreliable expert evidence, *Journal of Law and Human Behavior* (Pub. online Oct. 17, 2007) **32**, 363–374.

[36] Schklar, J. & Seidman, S. (1999). Juror reactions to DNA evidence: errors and expectancies, *Journal of Law and Human Behavior* **23**, 159.

[37] Godfrey, E. (2001). Poll shows oklahomans distrust system, *The Daily Oklahoman* A1 (May 27, 2001).

[38] LeFevre, A. (2000). Understanding generation X, *Trial* 58 (June 2000).

[39] Brennan, L. (2004). Pitching the generation X jury, *National Law Journal* **26**, (June 2004).

[40] Anthony, P.K. (2008). *Meeting Juror Expectations in the Twenty-First Century*, DecisionQuest, http://www.decisionquest.com/litigation_library.php?NewsID=323.

[41] Hamlin, S. (2000). Who are today's jurors and how do you reach them? *Litigation* Spring **27**, 9.

[42] Voris, B.V. (2000). Jurors to lawyers: dare to be dull, *National Law Journal* (Oct.) **23**, A1.

Related Articles

Daubert v. Merrell Dow Pharmaceuticals

Expert Opinion in Court: a Comparison of Approaches

Expert Opinion: United States

Expert Opinion in Court: Civil Law Jurisdictions (France, Germany, Italy, and Spain)

Expert Opinion: United Kingdom, Canada, and Australia

Foundation Testimony

Federal Rule of Evidence 702

Kumho Tire v. Carmichael

Web Resources

CAROL HENDERSON AND KURT W. LENZ

Experts in Trial Courts: Experts before Appellate Tribunals *see* Expert Opinion: Appeal v. Trial

Explosions: Scene Investigation

Improvised Explosive Devices

Improvised explosive devices (IEDs) or "homemade" bombs have been thrust upon the world stage of international terrorism and brought into our homes through daily newscasts.

An IED is essentially a combination of components not originally designed to be combined with each other; some explosives and some not, which when combined, form an improvised device that has a capability of violently exploding. In its simplest form, an IED consists of a main charge explosive and a fusing or initiation system that provides the stimulus to cause the explosive to detonate. Some IEDs have a few components while the others have many. As such, the number and type of components fabricated into an IED are generally dependant upon a number of factors. The first is how and where the device is to be used. It may have a timing mechanism built into the initiation system to afford a time-delayed initiation of the explosive. Examples of timing-mechanism components found after a bombing are shown in Figures 1 and 2. Usually the builder "hides" this type of device to prevent its discovery prior to explosion. There are several additional factors that affect the construction of an IED, including the following:

- the ingenuity of the builder;
- the skill of the builder;
- access to nonexplosive components; and
- access to explosives.

Although usually extensively fragmented, most of the components, other than the actual explosives, at least partially survive the explosion of the IED. This very basic point of component survivability is not apparent when the bomb explodes and most people incorrectly believe that nothing can or does survive. However, it is a fact that bomb components do survive an explosion and the recovery of those fragmented components constitutes the processes commonly referred to as *Bombing Crime Scene* or *Postblast Investigation*. Figures 3 and 4 show the postblast remains of two pipe bombs: one that was held together with threaded caps and the other held together by welded metal plates. Figure 5 depicts the fragmented remains of electrical detonators recovered following their use to initiate high explosives.

Bomb scene investigation involves a series of challenges to the investigator, which require not only basic crime scene investigation skills but also knowledge in a number of specialized disciplines. These disciplines include the following:

- explosion dynamics;
- identification features of main charge explosives and the accessories used to initiate them (detonators, detonation cord, safety fuse, shock tube, etc.);
- pre- and postblast identification features of nonexplosive components used to fabricate the IED;
- understanding of the capabilities of the laboratory with regard to the types of examinations that can

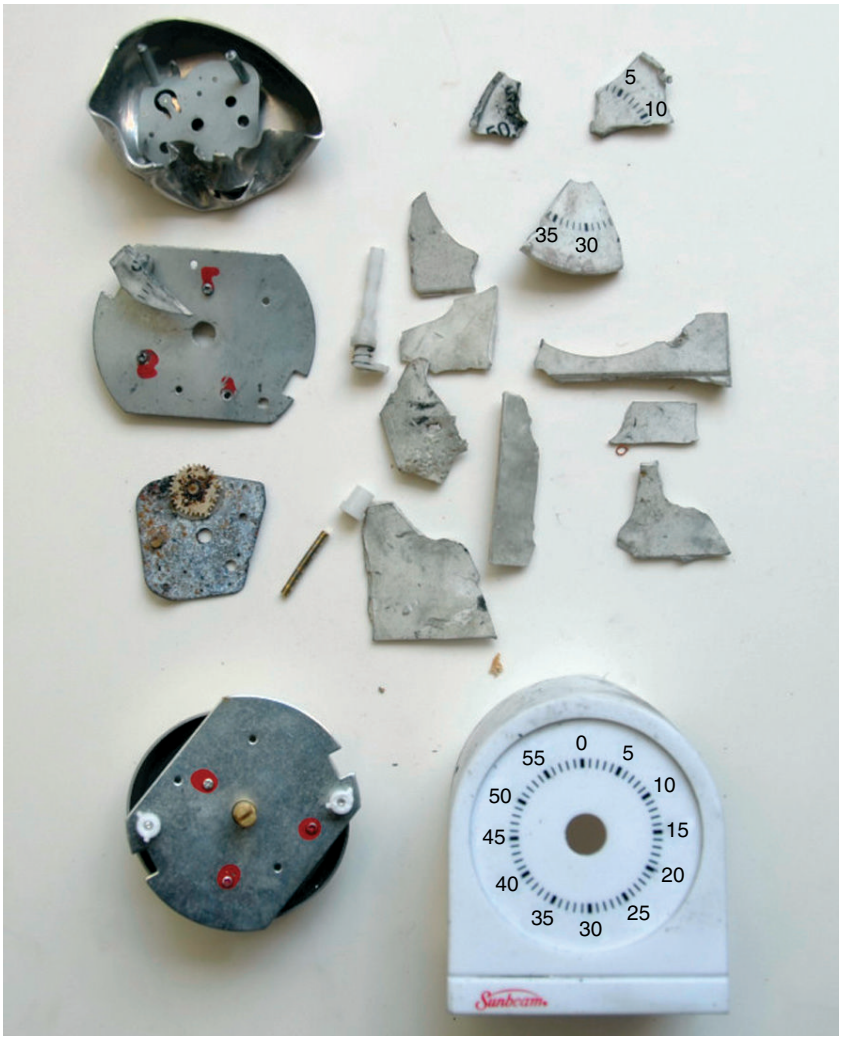


Figure 1 Fragments of a mechanical timing device compared with undamaged components

be conducted on the debris collected from the blast scene (*see* **Explosion Debris: Laboratory Analysis of**); and

- understanding and the identification of features associated with military explosive ordnance.

Additionally, the postblast investigator should be aware of the types of and the indicators of weapons of mass destruction (WMD), specifically, chemical, biological, and nuclear materials. This is especially important in understanding that the explosion of an IED may be a precursor to the delivery of WMD material.

The investigation into a reported explosion involves two stages: the scene investigation and the field investigation. The scene investigation involves site documentation and collection of postblast debris and components. The field investigation is that part of the investigation that is separate from the actual scene.

Scene Investigation

The following investigative steps are used to determine whether an explosion scene actually exists and



Figure 2 Remains of a mechanical clock used as a timing device, found after a bomb blast



Figure 3 Pipe bomb fragments. This bomb was constructed with a pipe nipple and two threaded end caps to form an enclosure for the explosive

then, the identification, collection, and documentation of bomb components:

- initial response;
- evaluation of the explosion scene;
- entering the scene;
- documentation of the scene;
- where to find evidence;
- how to find evidence;
- final survey; and
- release of scene [1].

The initial response to a reported incident may or may not include a postblast investigator or a bomb



Figure 4 Pipe bomb fragments. This pipe bomb was constructed with a pipe nipple and two welded end plates that formed an enclosure for the explosive

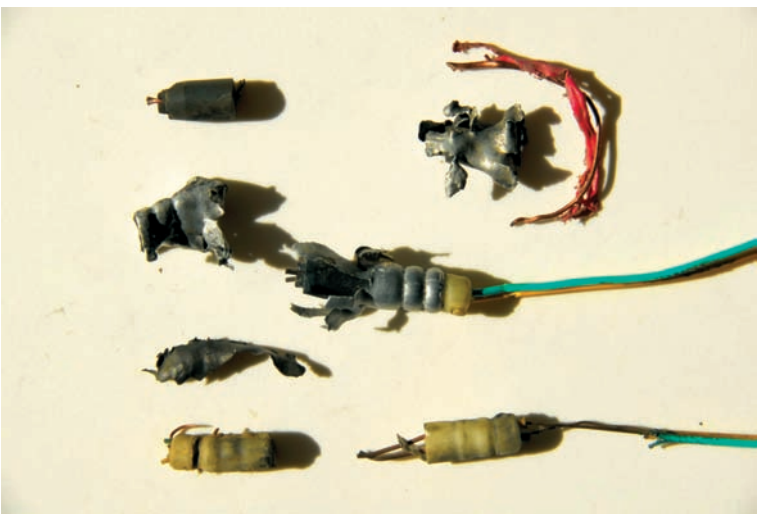


Figure 5 Fragmented remains of electrical detonators recovered following their use to initiate high explosives

technician, but it certainly includes the response by a uniformed police officer. The first responding officer evaluates whether an explosion has occurred and the degree of additional resources (additional police, fire fighters, and/or medical support) required. Should there be indications of an explosion, it is not the duty of the first responding officer to make a cause determination of the explosion, only that the scene "appears" to be from the result of an explosion. The actual cause determination is made later by the postblast investigator, which may be in conjunction with a bomb technician. It is the responsibility of the first responders, regardless of the cause of the explosion, to immediately assess the safety of the scene, establish exclusion perimeters around the scene, remove personnel from the scene, and assist medical support personnel in the removal and care of the injured.

The safety assessment should include an evaluation of the following list of possible explosion scene hazards:

- blood-borne pathogens;
- metal-glass debris;
- fire;
- structural integrity;
- utilities, especially electrical;
- hazardous materials; and
- secondary explosive devices, if the incident was caused by a bomb.

The evaluation of the explosion scene is conducted by an investigator knowledgeable to make the determination of whether the explosion was caused by the *detonation* of an explosive (a chemical compound or mixture of chemicals designed to explode), the detonation or *deflagration* of a diffuse fuel explosion consistent with air explosion involving a gas (e.g., natural gas or propane) or vapors from pooled flammable liquids or dusts, or some other source. Detonations are explosions in which the combustion zone propagates at a rate in excess of the speed of sound, while a deflagration's combustion zone propagates at a rate less than the speed of sound [2]. Mechanical explosions, such as those caused by ruptures of boilers or other pressure vessels, are usually not difficult to recognize. The following is a list of characteristics that an investigator may use to help distinguish between an event caused by an explosive material and an event involving the deflagration or detonation of a diffuse fuel Table 1.

The evaluation of the scene that is formed by the explosion of an IED requires additional steps to ensure that the scene is safely secured and that potential evidence is protected until it can be collected. These steps include locating the seat of the explosion or crater, and from that conclusion the demarcation of more specific inner and outer perimeters. Once this is accomplished, an evidence collection command post should be established. The person in charge of the evidence command post should document all personnel entering and leaving the scene, establish evidence documentation procedures, determine the level of investigative assistance required to locate the evidence, and obtain the legal authority required to search and collect evidence. This involves determining whether a consent to search is sufficient or a search warrant is required. Additionally, at this time, the principal scene investigator assembles the post-blast investigative team that has the responsibility of locating and securing the bomb remains.

Table 1 Investigative Characteristics of Explosion Scenes

Explosive material	Gas, vapor, or dust [3]
Crater	No crater
Localized damage	Entire walls blown out
Defined epicenter	No epicenter
Presence of bomb components	No bomb components
Explosives or residues present	No explosives present
No gas, vapor, or dust fuels on site	Gas, vapor, or dust fuel on site
Location of incident	Location of incident
Witness statements	Witness statements

The number of team members depends on the size and complexity of the scene, but can be from as few as one to dozens of investigators. Their team members play the following roles [4]:

- team leader;
- bomb disposal technician;
- photographer and assistant;
- evidence recovery personnel (searchers);
- evidence custodian;
- sketch artist; and
- specialists (e.g., chemist and structural engineer).

Additionally, the types of evidence collection equipment and supplies vary depending on the complexity of the scene. At a minimum, the team requires various types of containers (plastic and nylon bags of many sizes, metal cans), paper, writing instruments for evidence documentation, evidence labels, rulers and measuring tapes, hand tools, sifting screens, and residue collections kits.

Prior to the actual investigation of the bombing scene, it should be clearly understood by all involved that the investigative process has at least three main goals. First, to establish a legal basis that an explosion of malicious intent occurred. Second, to investigate the scene to recover IED components. And third, to link a subject with the bomb scene and/or the fabrication of the bomb. All three tasks

must be accomplished in a manner that is of sufficient integrity and credibility that the evidence developed is acceptable in a court of law.

Establishing malicious intent is not a trivial exercise in many cases. While it is always necessary to eliminate the possibility of an accidental explosion, scenes like the one shown in Figure 6, where a government building was destroyed by an exterior blast, leave little doubt about the malice of the bomber. Establishing the cause of aircraft explosions (particularly those that explode over water) is far more difficult. TWA flight 800 is an example of an explosion determined to have been caused by the ignition of jet fuel in the center fuel tank by what is believed to have been an electrical spark from deteriorated wiring. (There are still some skeptics.) The evidence that proved Pan American Flight 103 was the victim of a malicious act was postblast components, including a tiny fragment of a circuit board from the electronic timer, shown in Figure 7. It was possible to trace the source of the timer and determine that the timer was custom manufactured for Libya.

Upon entering the scene to conduct the investigation and collection of evidence, one of the first responsibilities is to begin the process of documenting the explosion scene. The documentation involves three separate forms that are used to supplement and not replace one over the other. These are as follows:

- written notes and logs;
- sketches and diagrams; and
- photographs [5].



Figure 6 The Alfred P. Murrah Federal Building destroyed by American terrorists Timothy McVeigh and Terry Nichols on April 19, 1995

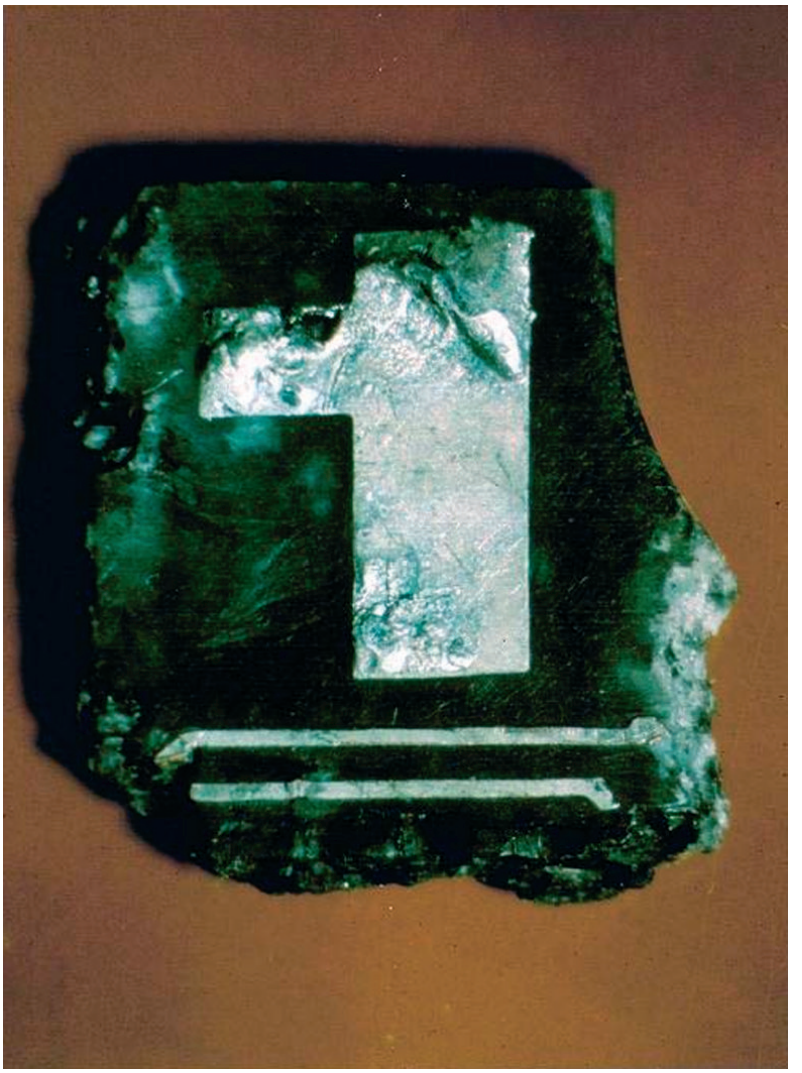


Figure 7 Small section of a circuit board that was part of the MST 13 electronic timer for the bomb that destroyed Pan American Flight 103 over Lockerbie, Scotland on December 21, 1988

These types of documentation provide a written and visual record of all the actions and conclusions of the investigative team, a systematic depiction of the scene, and the inventory and collection of evidence. Additionally, the investigators continually monitor the scene for any previously unidentified hazards, reevaluate the scene perimeters to determine whether they are appropriate and begin their efforts to locate evidence.

Once inside the scene, the investigative team needs to determine where to search for evidence. There are three general areas that contain evidence from the explosion. These are [6] as follows:

- the target of the bomb;
- victims of the bombing; and
- areas outside of the immediate target.

Obviously, the target of the bombing is the area that was impacted by the blast of the explosion. Depending on the quantity and type of explosive used, the target can be small (such as a postal

box) or quite large such as an office building. Within the target, evidence can usually be found in the crater or epicenter of the explosion and in the area and materials immediately surrounding the crater. Depending on where the explosion took place, the materials around the crater may be household furniture, vehicle components, building materials, and earthen materials. These materials are referred to as *witness materials*. The bodies of bombing victims, particularly those that were close to the blast, are excellent receptors of bomb debris. As such, all victims should be examined for the presence of bomb components, both construction components and explosive residues. The third area of search should be outside of the epicenter of the explosion. The areas usually are the places where bomb components have been projected away from the immediate scene along flight paths that can be identified by careful examination of the scene with an understanding of explosion dynamics.

Once a determination has been made regarding where to look for evidence of the bombing, the investigators need to determine how they search for evidence. The physical search for evidence can involve four types of organized methods. These are as follows [7]:

- explosive residue swabbing;
- organized search;
- sifting; and
- vacuuming.

Swabbing is a systematic collection procedure used to collect organic and inorganic residues of explosives resulting from the explosion of the main charge explosive. It is best used on nonporous items too large to be taken to the laboratory. The swabs are taken at the scene, entered into evidence, and then shipped to the laboratory with the other items of evidence recovered from the scene.

The organized search typically employs one of three types of search patterns: grid, line (or strip), and the spiral search. These patterns can be used independently or in conjunction with each other and involve the visual search of a given area for bomb debris. Of these patterns, the grid offers the most reliable method for locating evidence.

Evidence screening or sifting involves the separation of evidence from background materials by the use of a set of screens. These screens, much like those used on archaeological sites, have been found to be very useful when searching for evidence among soil, sand, or finally divided materials.

The last method is vacuuming that is used on porous surfaces too large to send to the laboratory for examination. This method can not only recover explosive residues but also minute bomb components that may be embedded in carpets, floors, and bomb fabrication workbenches. Vacuuming should be conducted after the bomb scene has been cleared of the larger debris and items of evidence; therefore, it should be the final search conducted.

Once evidence has been located, it needs to be collected and packaged in order to afford sufficient protection for it to be transported to the laboratory or used in any subsequent use in legal proceedings. The usual means is by "bagging and tagging". This operation provides a seamless chain of custody for each item recovered at the scene until final disposition. The "bagging" refers to the placement of the evidence into the appropriate type of container that protects it from being contaminated, altered, or lost. "Tagging" is the process of marking each evidence container with the required information that identifies where the evidence was found, by whom, the date it was found, and an inventory of the container by a numerical description that, in turn, corresponds to the evidence log. Guidance on the appropriate means of labeling evidence may be found in ASTM E1459, *Guide for Physical Evidence Labeling and Related Documentation* [8] and the Federal Bureau of Investigation, *Handbook of Forensic Services* [9].

The following is a listing of the types of evidence that may be found at the bomb scene and the type of container that is recommended for its collection Table 2:

Following the collection of the evidence and scene documentation, a *final survey* should be conducted prior to leaving the scene. The final survey is a review of all aspects of the scene investigation to ensure that the objectives of the investigative team have been met, and that all identifiable evidence has been collected, packaged, marked, inventoried, and entered into the evidence log. One of the last acts of scene documentation should be to depict the postinspection appearance through photography. This provides a visual record regarding the appearance of the scene after the scene investigation, should any questions arise as to the appropriateness of the

Table 2 Recommended Evidence Collection Containers

Type of evidence	Type of container
Components not requiring residue testing	Polyethylene zip-top bag
Materials requiring residue testing	Nylon zip-top or heat sealed bag
Material requiring residue testing	Metal can or glass container
Liquids, unconsumed explosives	Metal can or glass container
Unconsumed low explosives	Antistatic polyethylene bag
Dried clothing or cloth containing blood	Paper or Tyvek® bag

collection process and what the investigators did or did not do at the scene [10].

Finally, the team leader releases the scene to an authorized entity, if available, such as the property owner or public health official. In many instances, this may not be possible because of the victim's injuries or death. In the case where no actual person is available to take "custody", all reasonable efforts should be made to secure the scene prior to departure.

Following the completion of the scene investigation, the next phase of the postblast investigation process begins, the field investigation. The field investigation endeavors to establish a link between the person(s) responsible for the bombing and the evidence recovered from the bomb scene by establishing who had the motive, opportunity, and the means to commit the bombing.

To provide answers to these questions, various lines of investigation need to be considered. Depending on the circumstances of the bombing incident, these include detailed interviews of personnel associated with the following:

1. the neighborhood of the bombing for witnesses who may have observed suspicious activity prior to the bombing;
2. medical treatment facilities for the recovery of evidence from victims;
3. the first responders for what they saw upon arrival at the bomb scene and what the victims may have told them;
4. victim – subject associates and relatives for information that would lead to establishing an actual subject; and
5. review of logical sources of the components of the bomb in attempts to associate the bomb builder with the purchase of the components used to construct the IED. Retail outlets where bomb components are often purchased include hardware or home improvement stores, electrical supply houses, hobby shops, sporting goods stores, distributors of low and high explosives, farm supply stores, drug stores, mail order, and internet suppliers of componentry [11].

In conclusion, it is the responsibility of the postblast investigator and the investigative team to conduct as thorough and imaginative scene investigation as possible to determine and document the actual cause of the explosion and, if it is determined to have been the result of a criminal act, to assist the search for the person(s) responsible by locating, collecting, and securing the physical evidence.

References

[1] Thurman, J.T. (2006). *Practical Bomb Scene Investigation*, CRC Press, Boca Raton, p. 175.

[2] NFPA 921 (2004). *Guide for Fire and Explosion Investigation Investigations*, National Fire Protection Association, Quincy, p. 10.

[3] NFPA 921 (2004). *Guide for Fire and Explosion Investigations*, National Fire Protection Association, Quincy, p. 153.

[4] Federal Bureau of Investigation (2000). *Suggested Guidelines for Establishing Evidence Response Teams*, Federal Bureau of Investigation, Washington, DC, p. 5.

[5] Federal Bureau of Investigation (2000). *Suggested Guidelines for Establishing Evidence Response Teams*, p. 11.

[6] Thurman, J.T. (2006). *Practical Bomb Scene Investigation*, CRC Press, Boca Raton, p. 237.

[7] Thurman, J.T. (2006). *Practical Bomb Scene Investigation*, CRC Press, Boca Raton, p. 240.

[8] ASTM E 1459-92 (2005). *Standard Guide for Physical Evidence Labeling and Related Documentation*, American Society for Testing and Materials, West Conshohocken.

[9] Federal Bureau of Investigation (2003). *Handbook of Forensic Services*, Federal Bureau of Investigation, Laboratory Division, Quantico, p. 47.

[10] (2000). *A Guide for Explosion and Bombing Scene Investigation*, U.S. Department of Justice, National Institute of Justice, Washington, DC, p. 33.

[11] Thurman, J.T., (2006). *Practical Bomb Scene Investigation*, CRC Press, Boca Raton, p. 267.

JAMES T. THURMAN

Explosion Debris: Laboratory Analysis of

Introduction

In the United Kingdom, the United States, and most other countries, very strict regulations apply to the possession, storage, and handling of explosives. Apart from the stringent legal requirements for appropriate licenses for handling explosives, safety dictates that examination of explosives items should only be undertaken in properly equipped facilities by trained personnel.

Explosions, like dogs, are things that we all think we can recognize even if we find it hard to give a precise definition. Most people's perception of an explosion is formed from seeing special effects in cinema films and television programs. This has serious deficiencies. Typically, an explosion is characterized by a violent and exceedingly rapid release of energy, often accompanied by a bright flash of light. Nearby objects may be destroyed, damaged, or moved. For the forensic investigator presented with evidence from the scene of an event, the following questions need to be addressed:

- Does the evidence from the scene show that an explosion took place?
- If it was an explosion, what sort?
- Dispersed or condensed phase?
- Detonation or deflagration?
- Was it a deliberate act (a crime) or an accident?
- Does the evidence link to a suspect?

In addition, will be the following questions:

- What was the material of the bomb?
- How large was it?

At the Scene

If evidence collection is bungled at the scene, nothing done at the laboratory can fix it. Moreover, expert viewing of the scene can often establish whether, in fact, it was an explosion and indicate the order of magnitude and the general type of explosion, i.e., detonation or deflagration. Such observations can often enable subsequent laboratory examinations to be more efficiently and effectively focused. Ideally, the scientist would attend every scene. However, in many circumstances this is impractical or superfluous. If the scientist cannot attend the scene, if at all possible, a comprehensive account of the scene and circumstances, including, not merely descriptive text, but plans and photographs, should be provided.

Different approaches to investigate the scene of an explosion may be adopted depending on the local circumstances. For example, if an explosion scene is remote from the laboratory, it may be necessary to undertake more of the sifting and searching of potential evidential material at the scene. In the extreme, it may be necessary to deploy portable laboratory facilities. In more favorable circumstances, it may be practicable to undertake only limited searching of material at the scene, transporting potentially interesting material back to the laboratory for detailed forensic examination under controlled conditions. Investigation of the Bali bombing in 2002 is a good example of the dilemma faced by the forensic units. Teams were deployed from a number of nations, including Australia, the United Kingdom, and the United States. The Australians set up a clean room in a local hotel for on-site analysis and at the same time split samples in three ways for on-site analysis and analyses at the forensic laboratories in Australia and the United Kingdom.

Given the large quantities of debris that may contain material of potential evidential significance at the scene of an explosion, the investigation can be expedited by a degree of screening at the scene. A number of instruments have been designed for field use. Examples, as of 2008, are the hand-held ion mobility spectrometers (IMSs), e.g., Smiths, GE, Implant Science, or hand-held detectors based on other technologies, e.g., Nomadics Fido, Scintrex EVD-3500, Fido PaxPoint, Smith Travel IR, and Ahura Raman [1]. There are also colorimetric tests, which are acceptable presumptive tests, but destructive in nature, which can preclude subsequent confirmation in the laboratory. The use of field instrumentation at the scene may introduce difficult issues of calibration, maintenance, reliability, and quality assurance. Field conditions make the control of contamination and cross-contamination particularly challenging. It is essential that properly documented logs be kept to demonstrate the controls, which have been applied at the scene and ensure the subsequent integrity of evidence. Field analyses, in general, lack selectivity and are prone to contamination; thus, results should always be

checked under the rigorous conditions of the accredited laboratory. Unfortunately, analysis of a sample in the field often precludes further laboratory examination since the sample has effectively been destroyed. The risk that dirt may mask explosive traces in the absence of a laboratory cleanup treatment should not be overlooked when reviewing the results of field analyses; false negatives can occur in such circumstances. Another factor that needs to be considered is that there are often a number of social and economic pressures to return a bomb scene back to normal as quickly as possible. Although the recovery of evidence should not be compromised by these pressures, the possible benefits of using field instrumentation at the scene need to be carefully weighed up against the possible disadvantages, including time and resource implications.

The material of potential evidential value at an explosion scene can amount to a substantial quantity, sometimes measured in tons from large explosions. Dealing with such situations efficiently demands a well-designed and well-planned system. The scene investigator should have available large nylon and plastic bags and sheeting capable of safely containing large items, e.g., door and window frames, light fixtures, street furniture, ducting, and even automobiles. In addition, he needs collection containers for small items and a kit for swabbing immovable objects. The investigator should also be equipped with a supply of dustpans and brooms and a vacuum cleaner with replaceable filters and bags. Items, which obviously contain biological materials, such as blood-stained clothing, should be packaged separately with appropriate warning labels.

The items used to collect and package evidence must be verified as explosive-free prior to use. This may be achieved either by quality assurance before use or by the examination of unused control samples in parallel with the relevant evidential items. In the case of swabs, this may involve prior preparation and analysis at a forensic laboratory. For plastic packaging materials, dustpans, and brooms, this may involve use of new materials at every scene. Packaging and transport of evidential material is an important issue; without preplanning inappropriate makeshift containers such as unused garbage bins are likely to be used. Unfortunately, such containers once filled are too heavy and unwieldy for convenient handling. In the United Kingdom, a system has been developed using a standardized type of rectangular plastic box holding about 351 of debris and having a lid that can be sealed with an evidence tag. Many thousands of such debris collection boxes are deposited at strategic locations around the country ready for use by police officers charged with management of bomb scenes. Once filled the boxes are returned to the laboratory for examination. The potential evidential material should be divided into separate streams before it is sent to the laboratory. The basic division depends on the type of laboratory examinations expected. Swabs and objects for trace analysis require special packaging to prevent contamination, while items for physical examination only or bulk material to be searched for bomb fragments require only basic packaging to preserve evidential integrity and chain-of-custody. Given that it is not always readily apparent which items have the best potential for the recovery of explosives traces, it is wise to err on the side of caution in choosing the type of packaging to be employed.

A general misconception is that everything in the immediate vicinity of an explosion is completely destroyed. This is incorrect. Explosions shatter and scatter. The size of the fragments and the degree of dispersion depends on the nature of the items being fragmented and also on the proximity, size, and violence of the explosion. Thus, big powerful explosions tend to produce smaller fragments; and conversely, small weak explosions tend to leave larger fragments at or near the center (seat) of the explosion. Depending on the bomb design and explosive train, explosive material is spalled (thrown) from the outer surface of the bomb and projected clear of the high pressure, high temperature zone at the explosion seat [2–4]. In a well-built explosive device, it is only the spalled explosive material that is likely to survive detonation. Fortunately, most improvised explosive devices (IEDs) are far from ideal in either design or construction. Consequently, it is not uncommon to find visible explosive residue. This is especially true in the cases where a low explosive, e.g., black powder, smokeless powder, or a pyrotechnic has been employed.

The first step of the explosion scene investigator is to locate the seat of the explosion, i.e., the place where the explosive device was placed. This location will usually be marked by a crater. The presence of a crater is good evidence that a condensed-phase (concentrated in one small area) explosive was used. If no crater is found, further investigation is needed.

This may point toward a dispersed-phase explosion, such as a natural gas explosion. The type of explosion, i.e., dispersed phase or condensed phase, is likely to be quite relevant to the question: crime or accident? Accidents caused by leakage of fuel gases and the subsequent explosion of gas–air mixtures are not uncommon, although the possibility of deliberate act, rather than accident, still needs to be eliminated. A condensed-phase explosion, in the absence of legitimate storage of explosives or volatile and reactive chemicals, is much more likely to indicate a criminal act. Dispersed-phase events, such as gas–air explosions, tend to produce low pressures and moderate effects over a large area without any clear origin. In contrast, condensed-phase explosions usually give very severe damage at the original position of the explosive (the "seat"), with a rapid decrease in effect with distance. The size of the crater and the distance from it at which damage is found are the normal ways that the size of the bomb is judged, although it should be noted that damage can be quite asymmetric. In some cases, information obtained from a suspect or from the suspected bomb factory may be the best indication of the amount of material used to create the bomb, e.g., the 1993 World Trade Center bombing.

Sampling in or near the crater often yields useful chemical and physical evidence. Unfortunately, the crater is often full of water due to breakage of sewer or water lines and fire-fighting activities. Surfaces near the seat of an explosion can be good sources of chemical residues from explosions. The most promising are those surfaces that exhibit some degree of visible disturbance from the effects of the explosion, for example, sooting, pitting, cratering, bluing, or gas wash. Figure 1 shows an explosively damaged metal strut from the scene of the bomb attack on the London Stock Exchange on July 20, 1990. The metal has been twisted out of its original shape, and the surface is peppered by tiny craters.

In our experience, metal surfaces often are a good place to find explosive residue; however, porous media and soil samples should not be overlooked. Particularly volatile residues, such as accelerants used in arson or triacetone triperoxide (TATP), a homemade initiating explosive, are often quickly lost from smooth surfaces but can be found embedded in porous surfaces. Of course, swabbing of porous surfaces is not generally fruitful; such evidence is best collected and returned to the laboratory. In addition, experimental trials with very large bombs showed that some explosive residues were deposited from the smoke cloud formed after an explosion; thus, residue collection may also be more promising downwind of the explosion seat for very large events [5–7].

Although small items of debris are readily collected, it is worth considering the value of large items of explosive debris. Large items retained should be identified as to location and then bagged and transported to the laboratory for examination. For

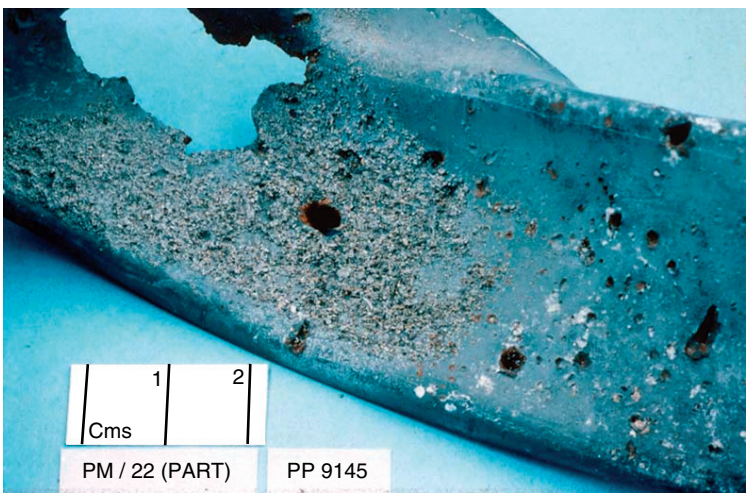
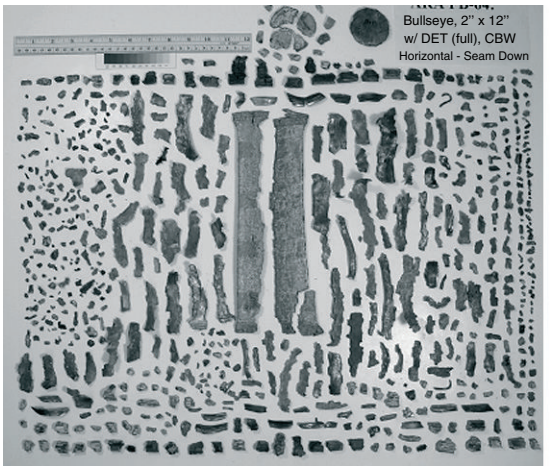


Figure 1 Explosively damaged metal strut from the bomb attack on the London Stock Exchange on July 20, 1990



(a)



(b)

Figure 2 Fragments of 2 in. × 12 in. steel pipes initiated with match and filled with Bullseye (a); commercial black powder (b)

example, in the case of an explosion inside a building fixtures and furnishings can act as useful witness pieces, potentially collecting bomb casing fragments and chemical residues depending on their position. Figure 2 illustrates the difference in a pipe bomb filled with black powder, a low explosive, and a smokeless powder capable of detonating. Note the difference not only in the number and size of fragments but in the manner in which the edges fractured. A complete discussion can be found in references [8–11].

In the case of an explosion in a street, nearby motor vehicles are potentially fruitful sources of evidence. Although there are numerous makes and models of motorcars, in practice, many of the basic components are standardized and shared among manufacturers. The condition and position of motor vehicles at the scene should be carefully documented. Subsequent reference to a car damage chart, such as that shown in Figure 3, can help estimate the approximate scaled distance of the car from the seat of the explosion. Since "scaled distance" is the distance divided by the cube root of the mass of the charge, using this chart in combination with the distance the car was from the seat of the explosion allows approximation of the size of the explosive charge. Great caution should be exercised in estimating charge sizes. Allowance must be made for the numerous unknown and unknowable factors involved in a bombing incident, and only rough order of magnitude estimates of charge size should be produced. In addition, motor vehicles can act as both excellent collectors of explosives residues and fragment interceptors at a bomb scene. It is common to find useful fragments of the bomb mechanism or container embedded in nearby vehicles. In addition to searching the bodywork and upholstery, the tires should be removed and examined for possible evidential fragments. Consideration should be given to sampling the exterior surfaces of vehicle facing the blast, either at the scene or back at the laboratory.

A particularly difficult concept to convey to scene responders, as opposed to laboratory investigators, is the need for background sampling. Determination of traces of nitrates in evidence is often a meaningless result, given that nitrates are ubiquitous in modern society [12, 13], but determination of particularly high levels of nitrates may be significant, particularly if taking into account the presence of other significant species such as ammonium, potassium, or uronium. It should be noted that for components of low explosives, the presence of invisible traces is rarely significant on their own due to their prevalence in the general environment. In these cases, the scientists would typically be looking for bulk (i.e., visible) quantities of unconsumed material or their combustion products. An example comes from the World Trade Center bombing in New York in

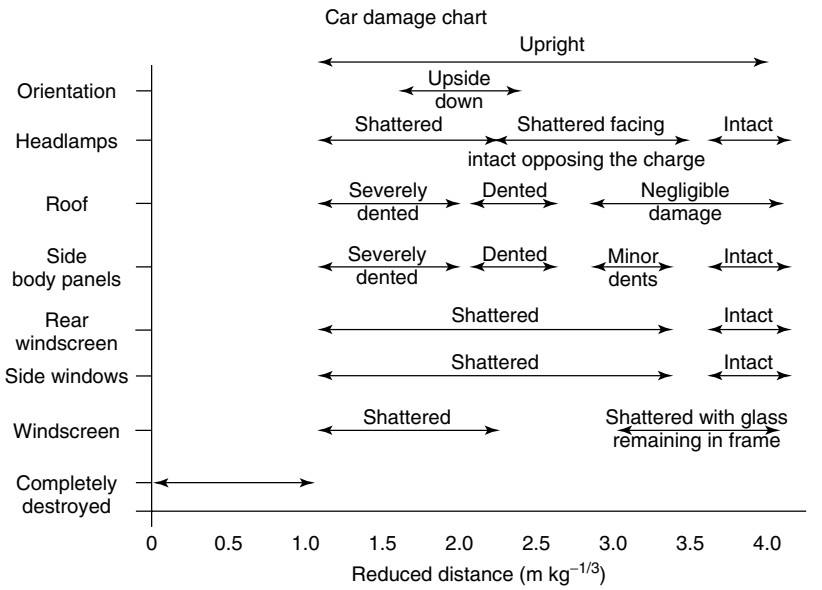


Figure 3 Car damage chart

February 1993. In that case identification of urea residue at the scene was of no evidential value, given that bags of urea were stored in the building for deicing of slippery ramps. The determination that urea nitrate was the main charge in that bombing came from examination of the contents of a storage locker rented by the perpetrators. Their purchase records and the marks on the floor indicating missing barrels led to the inference of the type and amount of material used. That being said, the presence of invisible traces of components of inorganic explosives may still have some evidential value from postexplosion scenes when taking into account other associative evidence such as the physical damage, other materials recovered, the combination of species, and the background environment.

In the Laboratory

There are several types of evidence that arrive at the forensic explosive laboratory: postexplosion scene debris, including fragments of possible device, fixed objects at the scene, and cars, parked or moving. The handling of large volumes of debris and large pieces of debris requires much preplanning. Sorting, cataloguing, and storing will require extensive efforts.

For items recovered from a premises search, whether the search is for a clandestine bomb-making labortatory or for evidence that the location had been used to store explosives, chemical analysis will be the main activity required. Even though the materials may appear to be in their original container, their chemical composition will need to be confirmed. Unlabeled materials, waste bins, and liquid mixtures may be particularly hazardous for the scene investigator. Hand-held detection devices may or may not provide the guidance needed for safe disposal of unlabeled chemicals. Rapid analysis of these becomes critical. Investigators will look to the laboratory forensic scientist to provide potential leads concerning where materials were obtained and even where the perpetrator(s) came from.

Samples from a suspect's person or property: his hands, his clothes, and his car – will again call for chemical analyses and may strongly influence court decisions as to whether a suspect should be remanded in custody, or granted bail, pending trial.

Physical Examination

Bins of debris from the scene of a large explosion can easily overwhelm a forensic laboratory. The first task of the forensic scientist is to select what to process first. Often, rain- or fire-fighting efforts at the scene of an explosion have left the debris wet, and, consequently, the first step in examination of

debris is to dry it. One approach that has been found satisfactory is a provision for a large room equipped with a dehumidifier. Material to be dried is spread out on a large clean plastic sheet on the floor and left overnight to dry. This makes subsequent processing of the material much easier.

Most explosives possess low vapor pressures (Table 1). Thus, if the debris is kept cool and dry the explosive will remain on the residue for years. A notable exception is the peroxide explosive TATP, which is extremely volatile and this can make its recovery more challenging. If this explosive is suspected or if there is sufficient evidence, some of the evidence should be tightly sealed until analysis, and analysis should be as soon as possible. The decomposition of 2, 4, 6-trinitrotoluene (TNT) in soil has been examined, and although it does decompose with time due to biological activity, some TNT remains detectable [14, 15]. Explosives lying between these two extremes in vapor pressure (TATP and TNT) in Table 1 may be in danger of disappearing over time. Cooled, sealed, and dry storage of debris is the only way to preserve evidence. Explosives traces can persist for much longer on adsorbent substrates and there are examples of traces of nitroglycerine (NG) being recovered after months or even years [16–18].

Table 1 Explosive properties including vapor pressure[a]

	Molecular weight	Melting point °C	Vapor pressure Pa at 25 °C
HMX	296	280 d	4×10^{-7} [b]
Picric acid	229	122	1×10^{-7}
RDX	222	204 d	6×10^{-7}
PETN	316	141	2×10^{-6}
Tetryl	287	129	5×10^{-3} [b]
UN	123	160	9×10^{-5}
TNT	227	81	1×10^{-3}
AN	80	169	1×10^{-3}
NG	227	13	6×10^{-2}
DMNB	176	210–214	0.3
2,4-DNT	182	69	0.7
4-NT	137	55	5
TATP	222	98	6
EGDN	152	−23	6
DADP	148	133	17
2-NT	137	−3	20
NM	61	−29	5×10^{3}

[a] Data compiled and adapted from References [19–26];
[b] vapor pressure at 100 °C.; Abbreviations: d, decomposes; HMX, cyclotetramethylene tetranitramine; Picric acid, 1-hydroxy-2,4,6-trinitrobenzene; RDX, cyclotrimethylene trinitramine; PETN, pentaerythritol tetranitrate; Tetryl, trinitro-2,4,6-phenylmethylnitramine; UN, urea nitrate; TNT, 2,4,6-trinitrotoluene; AN, ammonium nitrate; NG, nitroglycerine; DMNB, 2,3-dimethyl-2,3-dinitrobutane; 2,4 DNT, 2,4-dinitrotoluene; 4-NT, 4-nitrotoluene; TATP, triacetone triperoxide; EGDN, ethylene glycol dinitrate; DADP, diacetone diperoxide; 2-NT, 2-nitrotoluene; NM, nitromethane

Biological residue, such as blood-soaked garments, will also require immediate appropriate handling. In particular, biohazard materials must be packaged separately from other evidence and appropriately labeled and sealed. To minimize decay, such items are likely to need drying before being packaged; however, this needs to be balanced against the possible loss of volatile explosive traces. Refrigerated storage is likely to be advantageous in some circumstances. Staff handling biohazard materials require an ongoing medical surveillance program. This should include vaccination against such diseases as hepatitis and others as advised by the supervising medical practitioner. Referring to Figure 4, it will be obvious from the investigator's dress that the evidence must be protected from him and him from the evidence.

In the case of large explosions, the laboratory investigator is often under tremendous pressure to produce a result: was it an explosive? If so what was used? It is imperative to resist a premature judgment that such pressure attempts to elicit; however, in any case, the investigator will want to examine the most potentially fruitful evidence first. Forensic trace analysis is time consuming and costly. In the event that large volumes of potential trace evidence need to be screened to select the most promising items for analysis, one technique which has been used at times in the United Kingdom has been to segregate items in sealed nylon bags, warm these in an oven, and then sniff the vapor inside with a portable explosives detector. This technique, somewhat reminiscent of arson debris identification, can enable the best items to be selected and given priority for analysis. However, it should be realized that the most portable explosive detectors are configured for detection of a limited set of explosives, and there is a risk of missing compounds outside this select group.

Explosion debris can range from invisible traces of explosive, fragments of the bomb mechanism, bomb damaged vehicles such as motorcars, street furniture, and many tons of building rubble as well as

unexploded, disrupted, or partially exploded bombs. The laboratory system must be able to cater for all these disparate types of evidence. Sorting is tedious, time consuming, filthy, and potentially hazardous to health; however, in our experience, in roughly 40% of the cases, a useful object was uncovered.

For bulk material, visual examination is the first step. This is aimed at establishing if an explosion actually occurred and identifying evidential material from the immediate proximity of the explosion. Next, the debris is sieved through a series of screens to remove dust (typically about 25% of the mass) and divided into a set of different size fractions of 3, 6, and 12 mm. Practical experience has shown that it is slightly easier to search debris which is all of a similar size. What is being sought is that which does not belong; thus, elimination of the size variable makes the task slightly more manageable. The sieved material is spread out a little at a time on a well-illuminated search bench. First, it is swept with a magnet to pick out any magnetic material; then, it is searched by eye. Having a choice of lighting and different colored backgrounds for the search benches can be of significant practical help in this task (Figure 4). Staff engaged in such searching should be carefully monitored and given frequent short breaks to ensure that concentration is maintained. Training and skills revalidation should include searching presalted mock debris to ensure that a high proportion of test items are recovered. Health and safety issues need to be carefully controlled. Bomb debris may contain sharp items such as broken glass, dangerous materials such as asbestos, and biohazards, such as body fluids, body parts, and sewage.

Recovery of fragments of the bomb mechanism and container often provides valuable investigative leads and evidence. For example, if it transpires that a bomb was carried in a distinctive kind of bag, then witnesses can be asked if they saw such a bag and surveillance camera tapes can be reviewed to see if the bag can be identified. Recovery of fragments of the bomb mechanism, e.g., parts of a timer, can be exceptionally useful to investigators; and such evidence often provides links to the person or persons responsible. In addition, it is often possible to recover fragments of metal, plastic, or textiles that were close to the seat of the event. Such items must be examined for characteristic damage indicative of proximity to a detonation. Thus, metal and plastic fragments may exhibit what is known as *gas wash* caused by the scouring action of the very hot fast moving gases close to an explosion. In addition, surfaces may show craters caused by the impact of tiny particles of the unreacted explosive; these are typically visible under a low power microscope. Some synthetic textiles may also show characteristic effects, for example, the formation of "toffee apples" on the ends of fibers which have been exposed to the hot high velocity gases from a nearby explosion. The scale of the surface disturbances, such as cratering and pitting, can range from that clearly visible to the naked eye, as illustrated in Figure 1, to effects which require either a low power lens or a microscope. The scanning electron microscope (SEM) is particularly valuable for this type of examination because it provides not only a wide range of magnification but also a much



Figure 4 Drying cabinet for wet evidence (a); screening evidence (b); and examining evidence on colored background (c)

greater depth of field than a conventional optical microscope operating at the same power. Figure 5 shows SEM pictures of some explosive damage effects. The top left micrograph shows clubbing on nylon fibers; the other three show effects at different scales on aluminum alloy. The nature of such explosive damage to surfaces will depend on the target material as well as the nature and proximity of the explosion. The forensic scientist may well choose to generate reference photographs to illustrate the point being made in a particular case.

A bomb, typically, consists of an initiating system (detonator, match, squib, and flash bulb), a main charge, perhaps a booster, and a container. Timers may also be found, but these are not essential, e.g., suicide bomb, victim activated bomb, and tripwire bombs (Figure 6). Examination of unexploded or disrupted bombs involves identification of all these components. Moreover, these physical components can also be a rich source of trace evidence, for example, fingerprints, fibers, and DNA [27, 28] as well as toolmarks. Visual examination of any detonator and comparison against reference collections will often allow its identification. Detonators should never be broken down or opened to examine their internal components except in facilities which have been specially designed for the purpose. Such operations are extremely dangerous.

Chemical Analysis

Often the forensic scientist will have some background information which enables him to partly short circuit the full analytical scheme required for complete unknowns. As a simple example, recovered

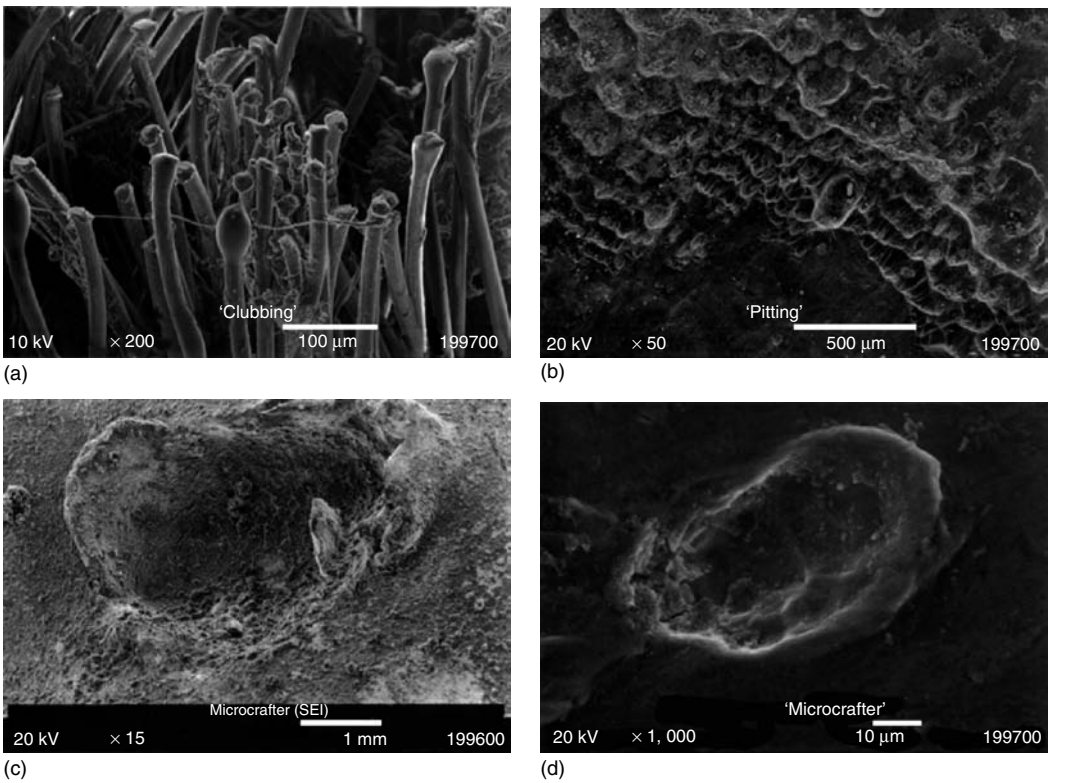
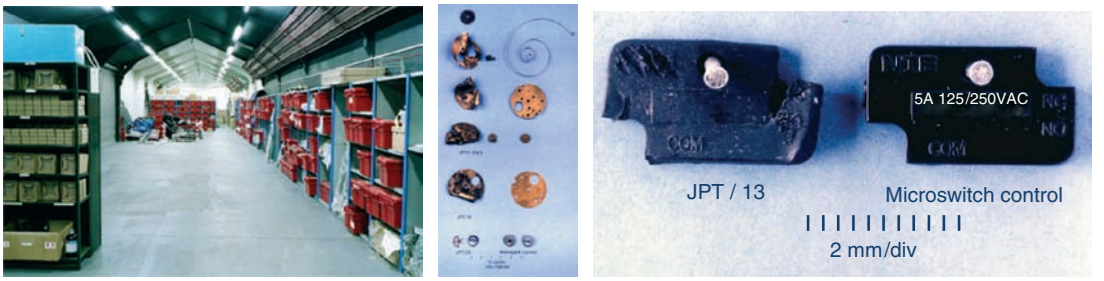


Figure 5 SEM pictures of some explosive damage effects Visual examination of bomb debris can also be supplemented by detailed photography to document the evidence. Photography of bomb debris can be technically challenging; for example, photographing features on soot blackened surfaces needs careful lighting if the necessary degree of contrast and extreme dynamic range is to be achieved in the final image



(a) (b) (c)

Figure 6 Storage of explosive debris in bins (a) and components (b and c)

material may already have a label giving its identity. In such a case, it would be reasonable to start by only doing the tests needed to confirm, or not, the correctness of the label. In other cases, the nature may be fairly obvious to the trained eye. For instance, an exploded pipe bomb with large amounts of residue may well suggest the use of a propellant or pyrotechnic and the visual appearance of the residue may be sufficiently distinctive that it can be recognized. In such circumstances, an abbreviated analysis scheme may suffice. Significant residue is often obtained from pipe bombs filled with low explosives because such materials require confinement for the reaction to proceed; and when the confining vessel bursts, the confinement is lost. Unconsumed particles can often be recovered by sweeping around the seat of the explosion. These particles can often be recognized by visual comparison with standard materials.

Bulk explosive, i.e., quantities visible to the naked eye should be collected and subjected to chemical analysis to establish their composition. This can then be compared with data in the reference libraries and, when possible, to reference samples. The approach to chemical analysis of residue varies from laboratory to laboratory. As might be expected, the analytical approach depends on information known about the sample, instrumentation available, and laboratory tradition. A consortium of North American forensic examiners, under the organization of Technical Working Group in Fire and Explosives (TWGFEX), published a list of suggested instrumental methods for identifying bulk explosives, including gas chromatography (GC)/mass selective (MS); liquid chromatography (LC)/MS; infrared (IR); energy dispersive X-ray (EDX); and Raman, GC/thermal energy analyzer (TEA), GC, LC/TEA, ion chromatography (IC), capillary electrophoresis (CE), thin layer chromatography (TLC), polarized light microscopy (PLM), spot test, and flame and burn tests [29].

Examination of intact explosive (visible residue or bulk) starts with a flame test to determine if the material is, indeed, energetic. This is a simple test to do and requires only a few milligrams of explosives. However, it is imperative that *only* a few milligrams be used and that the test be done a safe distance from the bulk of the material or any other explosive. Laboratories equipped with a differential scanning calorimeter (DSC) may find it useful to scan the sample. Not only will this indicate whether or not the sample is energetic, but it will provide information as to how energetic and whether it melts. For those familiar with the DSC trace of explosives, it may even provide a tentative identification (Table 2). The next examination for intact explosive is usually determining the elements present using EDX. This is particularly useful for identifying black powder and fuel-oxidizer or pyrotechnic mixtures. EDX can be a stand-alone instrument or attached to a SEM. An SEM provides a better depth of field than a traditional optical microscope. The SEM can be used to detect the presence of most elements of interest, although it is not usually suitable for determination of oxygen or nitrogen and lighter elements. SEM gives a starting point for subsequent chemical analysis. Furthermore, the combination of visual and element mapping is particularly powerful. For example, if EDX has detected the presence of aluminum, it is unclear whether that aluminum is in the elemental state or in a salt. SEM will clearly make that distinction. Furthermore, when an element map is performed by the SEM, it will be obvious whether the aluminum is dispersed or discrete. SEM or, indeed, optical

Table 2 Table of DSC peaks and IR spectrum

	Density ($g\,cm^{-3}$)	Exotherm maximum ($^{\circ}C$ at 20 $^{\circ}C\,min^{-1}$)	Average DSC exo heat (cal. g^{-1})	Major infrared peaks (cm^{-1}) by ATR				
				CH, OH, NH	Peak 1	Peak 2	Peak 3	Peak 4	Ranking of intensity
NG	1.6	211	500		1620	1260	1020	820	2, 3, 4, 1
PETN	1.7	215	800		1630	1260 (d)	1000	830	3, 2, 4, 1
RDX	1.8	253	900	3060	1560 (d)	1260 (m)	900 (d)	780	1, 3, 4, 2
TNAZ	1.8	275	750		1580	1530	1320	1270	4, 3, 1, 2
HMX	1.9	277	1000	3040	1520	1260	1200	940	3, 1, 4, 2
TNT	1.6	320	600	3120	1530	1350	900	720	1, 2, 4, 3
NC (13% N)	1.7	217	670		2460	1360	1020 (d)	820	3, 1, 2, 4
TATP	1.3	229	800	3000, 2940	1360	1180	880	780	4, 1, 2, 3
AN	1.7	328	350		3240	3060	1420	1300	3, 4, 2, 1
AP	1.2	360	450		3260	1410	1020		3, 2, 1
DADP	1.3	250	330		1200	890	820	700	1, 3, 4, 2
HMTD	1.6	166	700	3000, 2950	1360	1220	940	880	4, 2, 1, 3

microscopy will reveal whether the sample is heterogeneous or homogeneous. Inorganic elements, such as aluminum, magnesium, sulfur, barium, strontium, etc., can be confirmed and quantified by spectroscopic methods: inductively coupled plasma (ICP) emission, flame atomic absorption spectroscopy (AAS), or x-ray fluorescence [30–36].

Often the next step, with visible amounts of explosive, is IR spectroscopy. The nitro groups (NO_2) of most military explosives, the nitrate (NO_3^-) of ammonium nitrate (AN) or black powder mixtures, and the chlorates (ClO_3^-) of pyrotechnics are clearly visible in an uncluttered region of the IR spectrum. Thus, the single IR scan may allow classification of the material as organic or inorganic and possibly even a tentative identification of the explosive [37–39]. Table 2 lists the principal peaks observed in the IR spectrum of explosives commonly used illicitly. These spectra were collected using a portable, attenuated total reflectance (ATR) IR. (We have safely used ATR IR on explosives and even primary explosives; however, samples were always small and pressure was applied slowly.) This technique offers the advantage that the only IR peaks observed are attributable to the sample; no solvent or mulling material appears in the spectrum. The appendix shows the IR collected in their entirety, along with the corresponding Raman spectrum. The Raman spectrum can be used on white materials, but it should not be used on dark colored energetic materials as it may heat them too hot for safety. Because of safety problems, it is difficult to completely free the peroxide explosives hexamethylene triperoxide diamine (HMTD) and TATP from final traces of water and solvent used in their preparation. Thus, the presence of impurities can cause the appearance of the spectra to vary with the method of preparation. We include spectra of various different preparations of diacetone diperoxide (DADP), HMTD, and TATP in the appendix to illustrate the point.

After the entire sample has been examined, an attempt is made to separate the sample into organic and inorganic fractions. This is done by differential solubilities: inorganics are water soluble; organics dissolve in nonaqueous solvent. If trace samples (invisible explosive) are being examined, this is the point where relevance to their examination begins. Figure 7 shows the scheme used by the UK Forensic Explosives Laboratory for separation, cleanup, and recovery of traces of common organic and inorganic explosives.

Although this scheme is good for the recovery of the conventional explosives shown it is not satisfactory for the recovery of the peroxide explosives such as TATP, DADP, and HMTD.

Aqueous solutions are examined for the anions NO_3^-, NO_2^-, Cl^-, ClO_4^-, ClO_3^-, Br^-, PO_4^{2-}, F^-, SO_4^-, and SCN^- and cations Na^+, NH_4^+, Ba^{+2}, Sr^{+2}, K^+, Mg^{+2}, and Ca^{+2} of interest by IC or CE

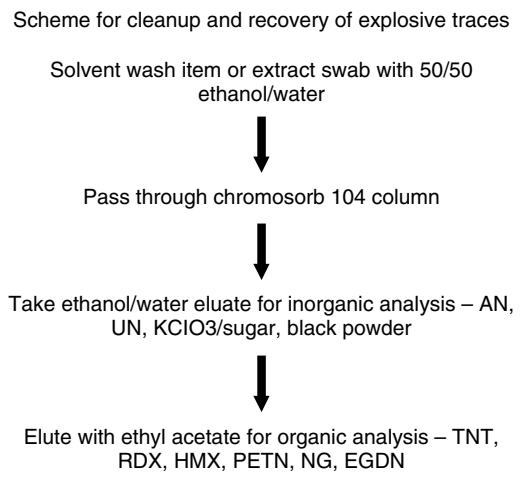


Figure 7 Scheme for separating trace explosives at the UK Forensic Explosive Laboratory

(Figure 8) [40–42], or colorimetric tests. Colorimetric tests are well established but not as sensitive or as definitive as chromatographic methods [43, 44]. Nevertheless, colorimetric tests are good for tentative field identification and commercial field kits, e.g., explosive test kit (ETK), are sold [45].

A number of homemade explosives are based on urea nitrate or on AN with a fuel. It should be noted that determination of the presence of nitrate and ammonium or nitrate and urea are insufficient to prove the presence of these specific compounds, as opposed to their component ions, which may be associated with a different species. Evidence as to these specific compounds can be found in the IR spectrum [39], X-ray diffraction, LC/MS, or PLM [46]. The order of analysis (organic versus inorganic) may be guided by field intelligence or observations made to this point.

Organic explosives (hexogen (RDX), TNT, pentaerythritol tetranitrate (PETN), NG, ethylene

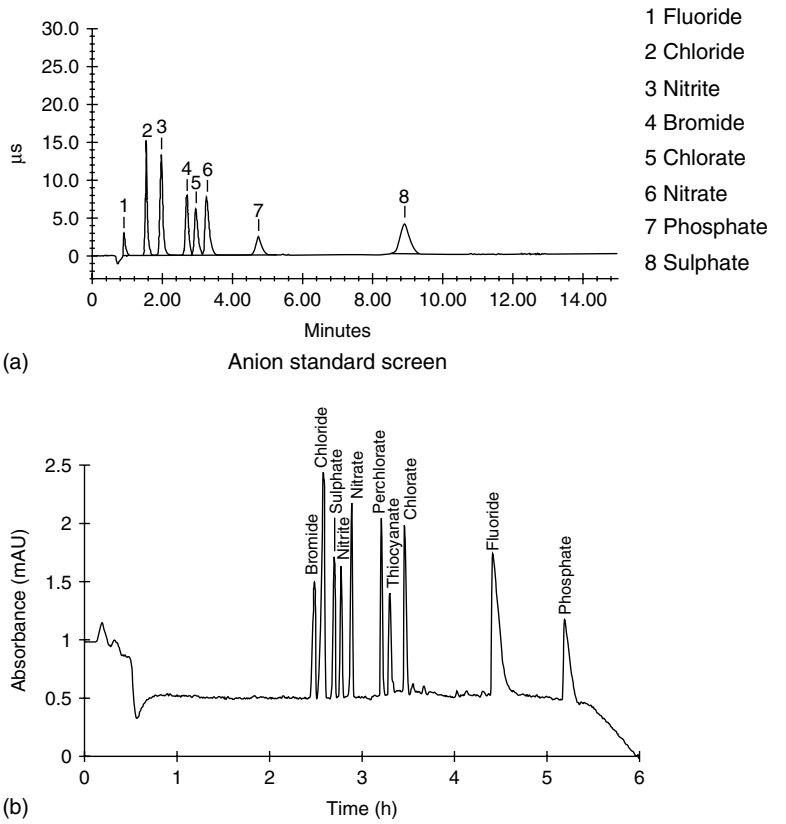


Figure 8 Ion chromatograph (a) and capillary electrophoresis (b) of anions

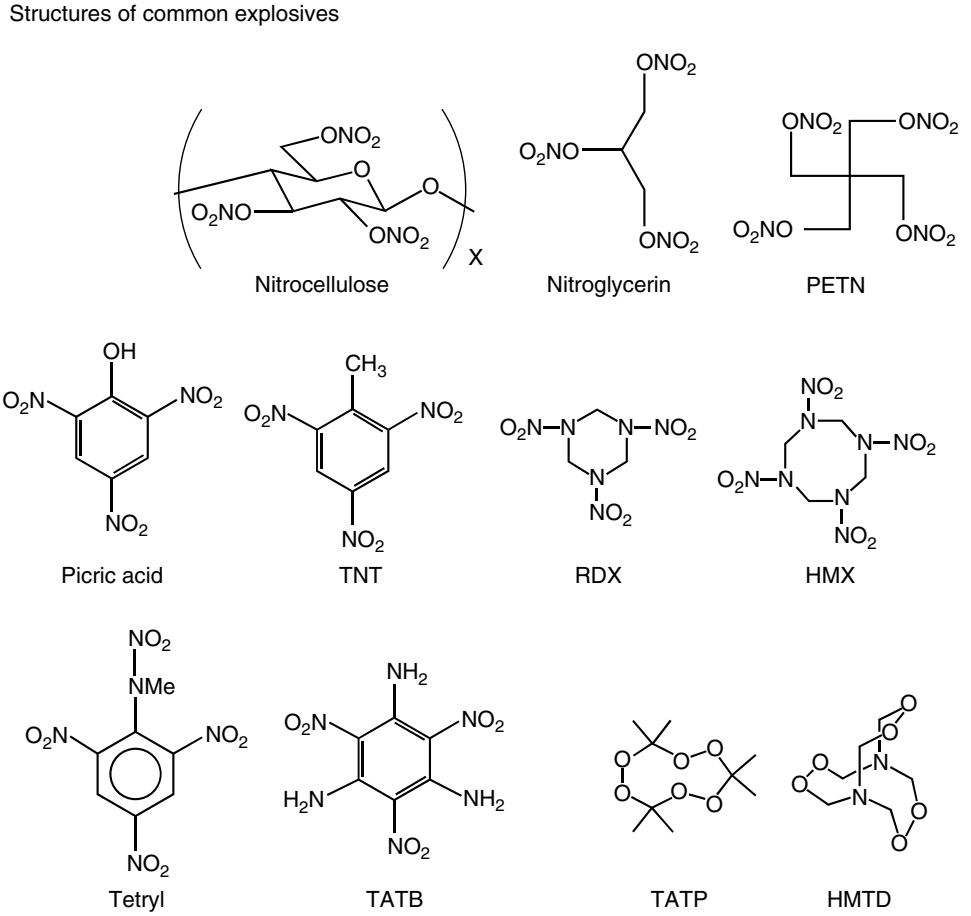


Figure 9 Structures of common explosives

glycol dinitrate (EGDN), octogen (HMX), and TATP, HMTD) can be identified by GC, LC, or TLC. Figure 9 shows the chemical structures of some of the common explosives. The availability of instrumentation, specifically GC or LC with MS detectors, has largely replaced the use of TLC, although it is still a valuable technique for screening and for meeting identification criteria requirements. TLC is a well established method for compound identification with surprisingly good sensitivity. The reader interested in this technique can consult Tables 3 and 4 and the suggested references [33, 47–49]. Although R_f values (ratio of distance traveled by analyte to solvent) are given for various explosives, the reader is advised to use this simply as guidance when developing his own library.

An organic solvent is used to extract swabs or physical samples to determine if the traces of organic explosives can be found. Various solvents could be used. In addition to the ethanol/water scheme described above in Figure 7, we have also found that HPLC-grade acetonitrile is well suited to extract the swabs and physical objects. Swabs can be immersed in solvent in sealed containers and vigorously shaken or sonicated with minimum volume of acetonitrile. After some minutes the solvent can be removed and filtered via a syringe filter, e.g., Millex-FG filter unit (0.20 μm, PTFE membrane for the removal of fine particles from organic solvents). Swabs should be rinsed or reextracted with solvent. It is often convenient to concentrate the solvent to less than 1 ml by blowing a stream of dry prepurified nitrogen

Table 3 Solvents for TLC

System	Eluent composition	(%) Volume	Comments
1	Toluene	100	NG and PETN not resolved
2	Petroleum ether 40–60 °C	40	NG and PETN resolved. RDX and HMX have low R_f's
	Petroleum ether 60–80 °C	40	
	Ethyl acetate	20	
3	Chloroform	90	RDX, HMX, and NC all resolved. Especially good for HMX in presence of high RDX
	Methanol	10	
4	Toluene	90	Recommended general screen. NG, PETN, RDX, HMX, and NC all resolved
	Ethyl acetate	10	

Table 4 TLC R_f values for explosives

Explosive	R_f in solvent system				Behavior under ultraviolet light	Color after spraying with NaOH solution and heating	Griess test color
	1	2	3	4			
NC	0.00	0.00	0.00	0.00	No effect	No effect	Magenta
HMX	0.01	0.02	0.25	0.06	Absorbs	No effect	Magenta
RDX	0.07	0.03	0.53	0.13	Absorbs	No effect	Magenta
Tetryl	0.43	0.18	0.83	0.61	Absorbs	Orange, then yellow after heating	Magenta
NG	0.61	0.32	0.84	0.66	No effect	No effect	Magenta
PETN	0.64	0.44	0.88	0.78	No effect	No effect	Magenta
TNT	0.74	0.52	0.93	0.88	Absorbs	Brown	Magenta

gas by or through the solvent extracts. This solvent concentrate is adjusted to an exact volume (e.g., 1 ml) [50].

A reference standard should be prepared containing the explosive species for which the test is screened (RDX, TNT, PETN, NG, EGDN, and HMX). Standard commercial solutions are available. When preparing reference standards for use in a trace analysis laboratory, it is good practice to only use mixed standards so that any contamination of samples by laboratory standards is immediately obvious from the improbable mixture of compounds detected. An electron capture detector (ECD) is one of the most sensitive detectors that can be coupled with GC. A TEA detector is also extremely sensitive to explosives, but it is less versatile being selective only for nitro-containing explosives (although this can also be an advantage). If the detector used is flame ionization (flame ionization detector (FID)), ECD, or TEA, (i.e., not MS), then a positive match to a particular explosive requires that the retention time of the sample from the explosion scene must match that of the standard, and this match must be confirmed using at least one, if not two, different chromatograph columns.

With TEA or ECD it is not unreasonable to expect limits of explosive detection in the parts per billion (ppb) range. Most mass spectrometers are less sensitive than ECD or TEA; however, the technique is constantly improving. A number of good gas chromatographic columns are sold. We have found that shortening the column to 15 m or less improves detection, depending upon the film thickness and temperature program. DB-1 or DB-5 or RTX-200 capillary columns have given fine results, as have BP-1, BP-5, BP-20, and CPSIL-19. Mass spectroscopy offers the advantage of a second mode of identification. MS fragmentation patterns may be collected and stored as part of the library of the instrument; a few can be found

on the National Institute of Standards and Testing) (NIST) and on the University of Rhode Island websites (http://expdb.chm.uri.edu). GC-ECD identification protocols are found on environmental protection agency (EPA) method 8095, "Explosives by GC [51]".

Choice of GC or LC may depend on availability, volatility of sample, and amount of sample [52]. Explosives such as PETN, tetryl (CE), HMTD, and HMX can be difficult to identify by GC. These explosives are prone to decomposition either in the inlet or on column. This makes the choice of inlet and column temperatures as well as column length and type critical. Unless these parameters are optimized the parent peak may not be observed. For these particular explosives LC/MS may be more suitable. Until his death in May 2008, Yinon published many papers discussing identification of explosives by mass spectroscopy [33, 34, 53].

The peroxide explosive TATP can readily be identified by MS assuming that necessary care has been taken during the recovery and packaging of the explosion debris. Figure 10 shows the mass spectrum of TATP obtained on a bench-top GC/MS, and Figure 11 shows the mass spectrum of the other common peroxide explosive, HMTD, obtained on the same instrument. Because analytical techniques for handling TATP are relatively new, an exemplar method is given below and in references [22, 54–60]. Analytical methods for analysis of other homemade explosives are in the references: HMTD [58–60], urea nitrate [61–65], AN [66–69], ammonium perchlorate (AP) [40–43], and conventional explosives [52, 70–77].

Figure 12 is a chromatogram of a mixed TATP and DADP standard in acetonitrile analyzed on the GC-ECD. A Hewlett Packard 5890 Series II GC-ECD was equipped with a fused silica DB1 column (J & W Scientific $-8\,\text{m} \times 0.53\,\text{mm}$ i.d. $\times\ 0.5\mu\text{m}\ d_f$). The injection port temperature was set at 175 °C (2 : 1 split) with an injection volume of 1 μ. The carrier gas was ultra high purity (UHP) hydrogen (6.6 ml min^{-1} measured at 65 °C) and the makeup gas was UHP nitrogen (45 ml min^{-1}), with a detector temperature of 300 °C. The oven temperature program was as follows: 65 °C for 2 min, 20 °C min^{-1} to 250 °C with a 2-min final hold. The peak for TATP had a retention time of 2.49 min and the limit of detection was circa 0.01 ppm. The peak for DADP had a retention time of 0.57 min and the limit of detection was also circa 0.01 ppm. The TATP peak consisted of two fully resolved peaks,

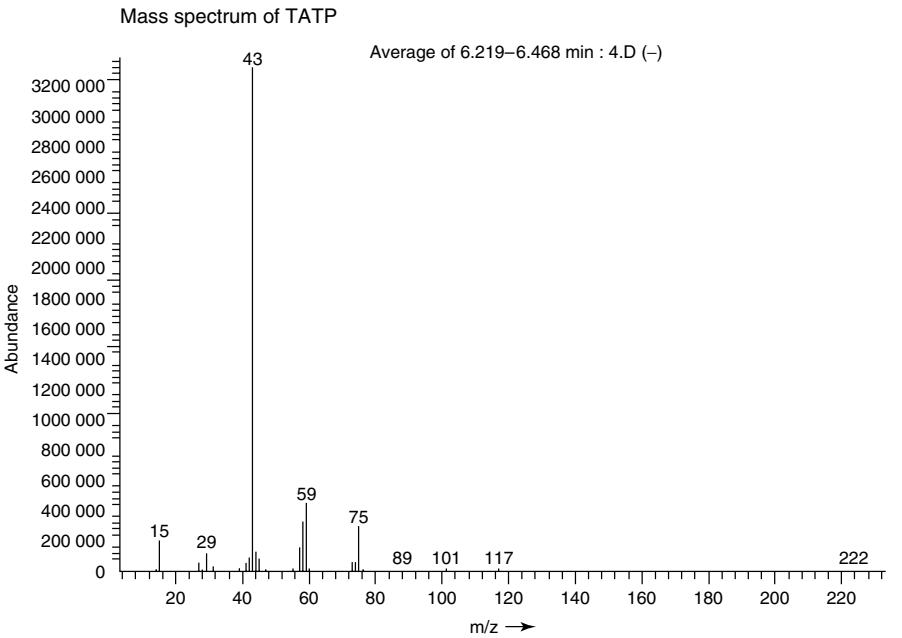


Figure 10 Mass spectrum of TATP

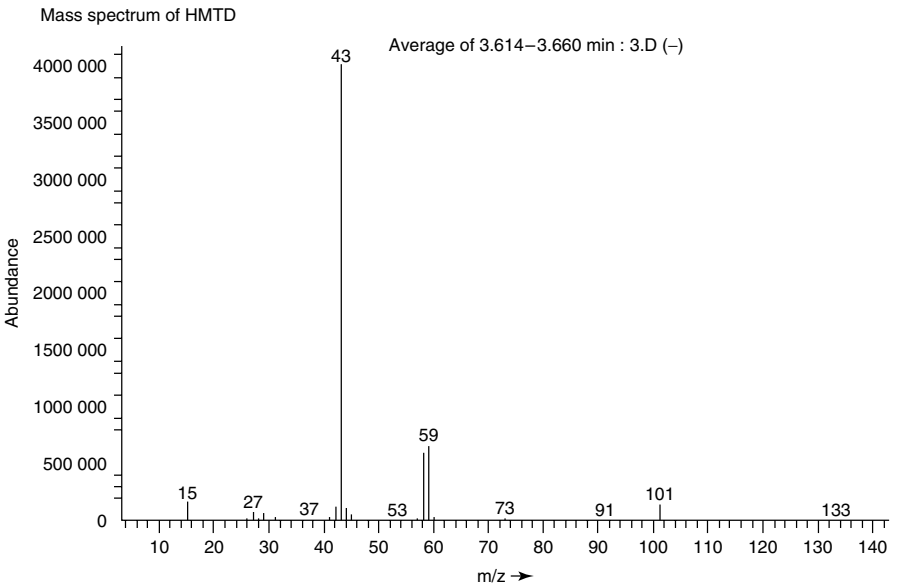


Figure 11 Mass spectrum of HMTD

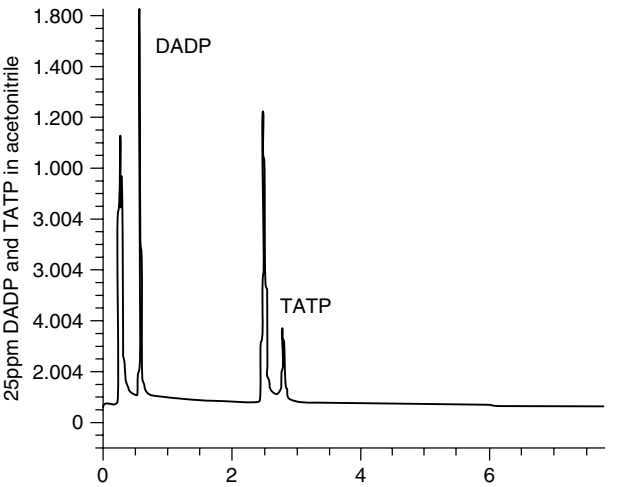
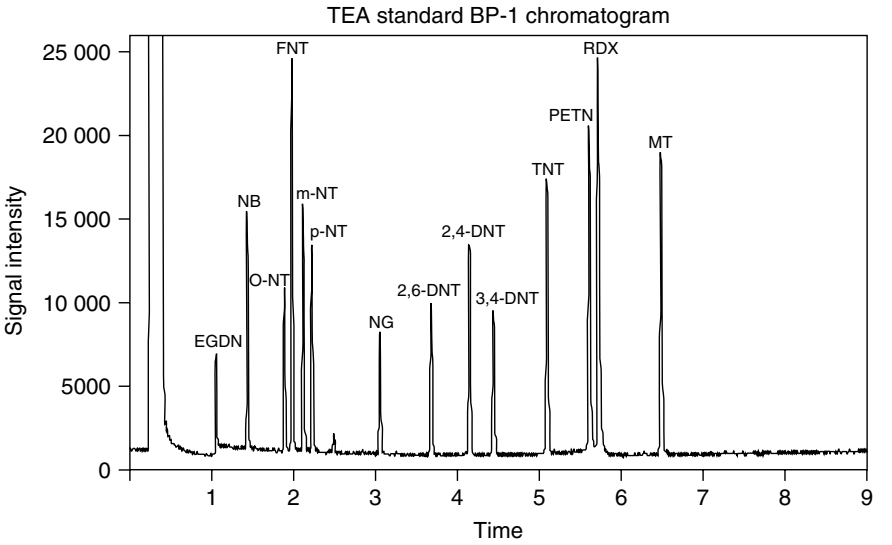


Figure 12 GC-ECD of 25 ppm DADP and TATP in acetonitrile

although on other chromatographic columns the second peak might appear as a shoulder. Both peaks had the same mass spectrum and have been shown in scientific literature to be two stable conformers of the nine-membered TATP ring. The larger peak corresponds to the conformation with the more stable symmetry (D_3) and the smaller peak corresponds to the conformation with the less stable symmetry (C_2). These kinetically stable conformers can interconvert through a high barrier, in the gas phase.

Typical chromatograms for conventional explosives are shown in Figures 13 and 14. The former shows a GC-TEA analysis of trace amounts, while the latter shows the use of GC-FID for analysis of bulk quantities.



SGE type 12QC2/BP1 0.22 mm 12-m polyimide clad silica, 0.22 mm i.d., 0.33 mm o.d., coated with bonded dimethylsiloxane 0.25 μm film thickness

Figure 13 GC-TEA analysis of trace explosives using BP-1 column

GC-FID analysis of bulk explosives:

SGE type 12QC2/BP5 0.25. 12-m polyimide clad silica, 0.22 mm i.d., 0.33 mm o.d., coated with bonded 5% diphenyl-dimethylsiloxane 0.25-μm film thickness

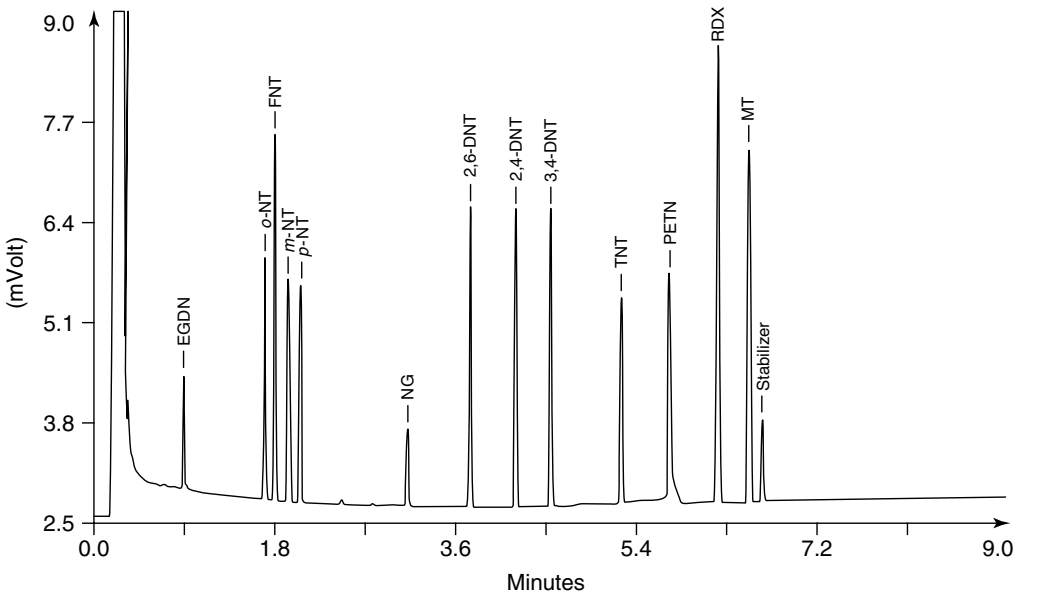


Figure 14 GC-FID analysis of bulk explosives

The Questions and the Court

Was it a Bomb?

If military explosives or peroxide explosives such as TATP and HMTD were identified among the explosive debris and if this corresponds with the other observations and evidence recovered, then the answer is probably "yes" providing the explosion was not at a location where explosives would normally be found, e.g., an explosive factory or storage facility and providing all sources of contamination were eliminated. Potential sources of contamination at the scene are the bomb squad and law enforcement who may have attended previous incidents. Contamination at the scene can be avoided by insisting on disposable materials and fresh outer garments. Written procedures, tight controls, and an access log should be enforced at scenes to avoid claims of contamination. In the absence of relevant environmental survey data, background samples should be taken "near" the scene to look for ambient contamination. In the laboratory, working in bulk explosives environments prior to carrying out trace work should be avoided. The laboratory environment should be controlled and monitored. We have found the best way to prevent contamination is to overdress with disposable outerwear, including hair net and booties [78]. The work environment should be screened prior to handling scene samples. In running chromatographs or similar analytical tools, appropriate control samples, both blanks and spiked samples, should be added.

Finding trace levels of explosives needs to be set in context for investigators and the Court. For example, finding trace levels of TNT on a person could be highly significant. However, the possibility of an innocent source must be considered and evaluated. Thus, a trace of TNT on a soldier probably has a very different significance than a trace on a member of the general public. The investigator needs to take into account the location where the trace was found, whether there were potential sources of the explosives in the vicinity, whether the compounds in question are widely used commercially or industrially or, indeed, occur naturally in the environment. Finding chlorate traces at the seat of explosion could be highly significant; however, it must be remembered that chlorate and sodium chlorate, in particular, are used as weedkiller in some countries so that this potential source must be considered and eliminated if significance is to be attached to the finding. Perchlorate is generally uncommon in United Kingdom and United States public environments though one would expect it to be much more common in Chile where it occurs naturally. However, even in the United Kingdom and United States perchlorates are widely used in fireworks and rocket propellants, and traces have been found in both the cow's milk and mother's milk [13, 41, 42]. If significance is to be attached to traces of perchlorate, potential sources of innocent contamination must be considered.

Environmental surveys are particularly valuable as this enables the levels found in evidential samples to be compared with those found in the general public environment. Clearly, the greater amount of explosive in an evidential sample compared to the background samples, the higher the likely significance.

If the Explosion was a Bomb, who did it?

If a suspect is apprehended, it is good evidence if traces of the material found on the bomb can be found on the suspect's person. Timothy McVey (the Oklahoma City bomber) had traces of PETN on his tee shirt. Ahmad Ressam, the would-be millennium bomber, had acid holes in his jeans. Trace explosive residue should be sought on hands, clothing, jewelry, and glasses. The template pieces of glasses and watch bands often collect residue that trip up the bomber as well as the forensic scientist trying to prevent contamination. We have found that some explosives stay in hair for days, even after washing. That explosive can be collected by entwining clean swabbing material in the teeth of a comb and combing that through hair. Favorite head wear, e.g., baseball caps or hair bands, often become contaminated [79, 80]. Nonetheless, explosive trace evidence alone would not usually be considered sufficient to prove guilt; generally courts would look for other corroborative evidence. A jigsaw of many pieces all fitting together to produce a consistent picture is more

reliable than a single item of apparently conclusive evidence which might later prove to have some undetected flaw.

Scene investigators will likely be the ones to take samples from the suspect's vehicle or premise. Contamination can be avoided if the laboratory investigator prepares clean swabbing, vacuuming, and other collection tools. Laboratories with large, clean garage bays can request that the scene investigators simply wrap vehicles in plastic sheet and take to laboratory for careful sampling. The importance of avoiding contamination and instituting a rigorous protocol for quality control cannot be overemphasized if evidence is to be acceptable in a court of law [78, 81–83].

It may be many years before a case goes to court. Meticulous notes should be taken, and the scientist should sign and date them and either have a cross-check signature or an effective laboratory quality assurance system. Since evidence may be required in court years after the event, the critical pieces of evidence should be resealed and labeled and stored according to laboratory protocol.

In court, the forensic scientist will be asked if he identified a device. Optimally, he should be able to identify the components of the device, the explosive(s) used, the means of initiation, and the container and to discuss its construction. He may be asked whether it was an actual explosive device or an elaborate hoax. This will lead to a discussion of whether the device could have functioned and with what effect. Presumably the investigator has been able to identify the explosive or pyrotechnic content of the bomb. However, in some cases, e.g., smokeless powder, he may only be able to say that it was single-base or double-base and was or was not consistent with the powder found in the defendant's premises. In other cases, particularly those involving high explosives, it may not always be possible to positively identify traces of the actual explosive(s) used or all of the components of the device. In such cases, the forensic scientist may be able to make an assessment regarding the type of explosive used and the construction of the device by the nature of the damage, both at the scene and on individual fragments, together with other associative evidence. These observations and assessments can also be considered against materials found at a defendant's premises, etc.

It is important to determine if there has been a history of similar devices. The Unabomber eventually started marking his devices with the initials "FC" (standing for freedom club) so that he would be sure to be credited. However, his completely homemade devices were already so unique that even after a hiatus of many years, his devices were easily recognized. Today, at the federal bureau of investigation (FBI) there is a multiagency group devoted to characterizing the IEDs coming of Iraq.

In the case of the break-up of a so-called bomb factory, the investigator requires considerable care and objectivity. He needs to look at the whole picture, and not just the chemicals found. What hardware was found in that location? Was it timers, clocks, batteries, or matches? These are common electronics and may not be sufficient evidence of a crime. The investigator should look for evidence of manufacture, e.g., mixing bowls with traces of a homemade explosive, clocks with attached wiring, collections of explosives literature, or plans. If a bomb has already exploded, it will be important to look for correlations between unused items and remains found at bomb scenes.

Summary

Reliable laboratory analysis requires much preparation and training before the explosion has even occurred. Scene and laboratory investigators need to communicate as to how and where the sampling will be performed. Laboratory protocols for analysis should be written and tested. The laboratory, itself, should be checked for possible explosive contamination. And, of course, the investigators themselves need sufficient training and experience, which may be obtained to some extent, e.g., by practicing on mock materials so that when the actual even occurs, all will be in readiness.

Appendix

Infrared and Raman spectra of commonly abused explosives are shown in Figures A1–A27.

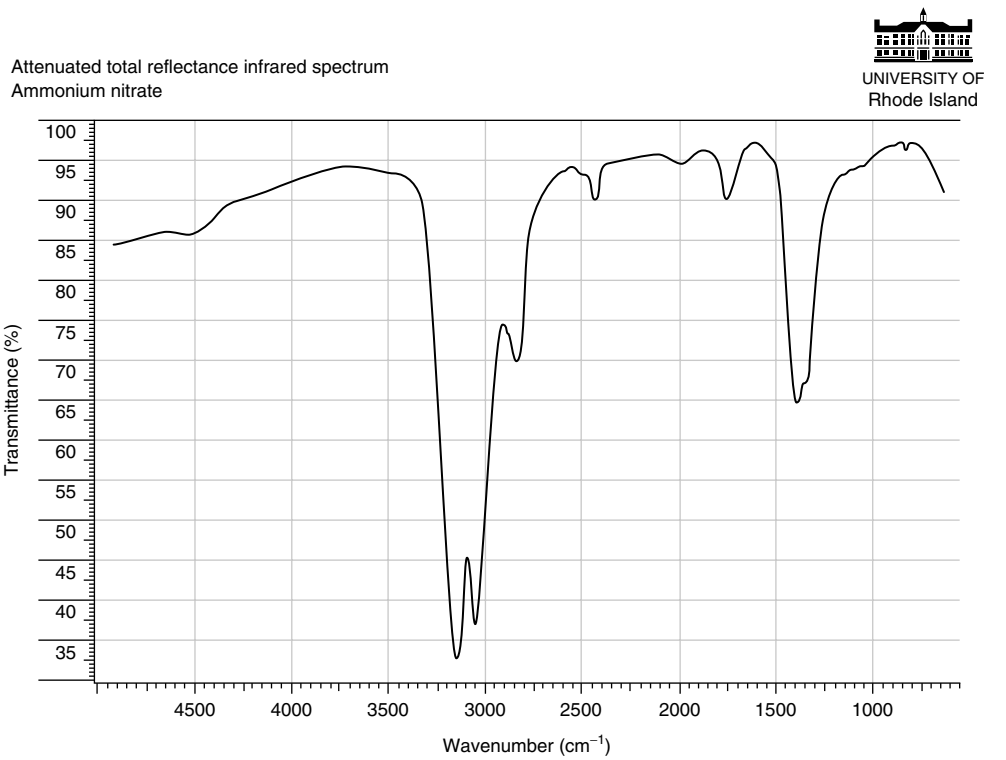


Figure A1 Infrared spectrum of ammonium nitrate (AN)

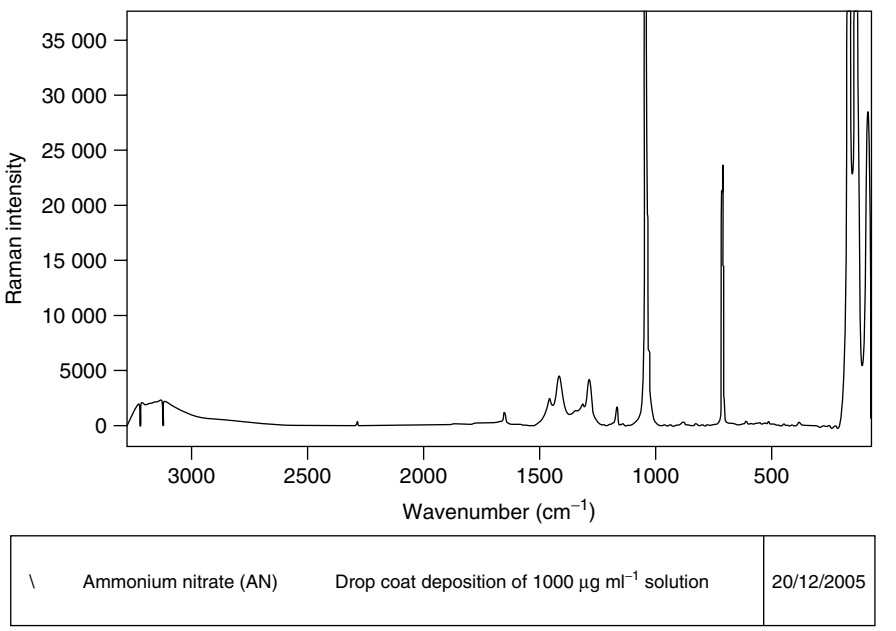


Figure A2 Raman spectrum of AN

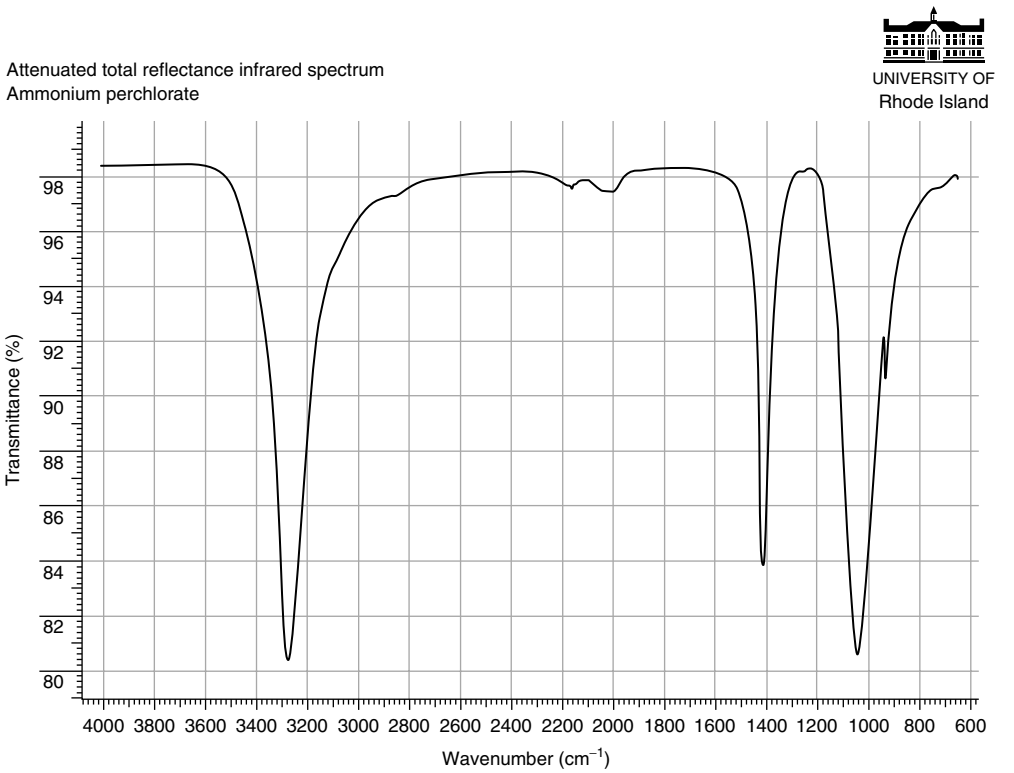


Figure A3 Infrared spectrum of ammonium perchlorate (AP)

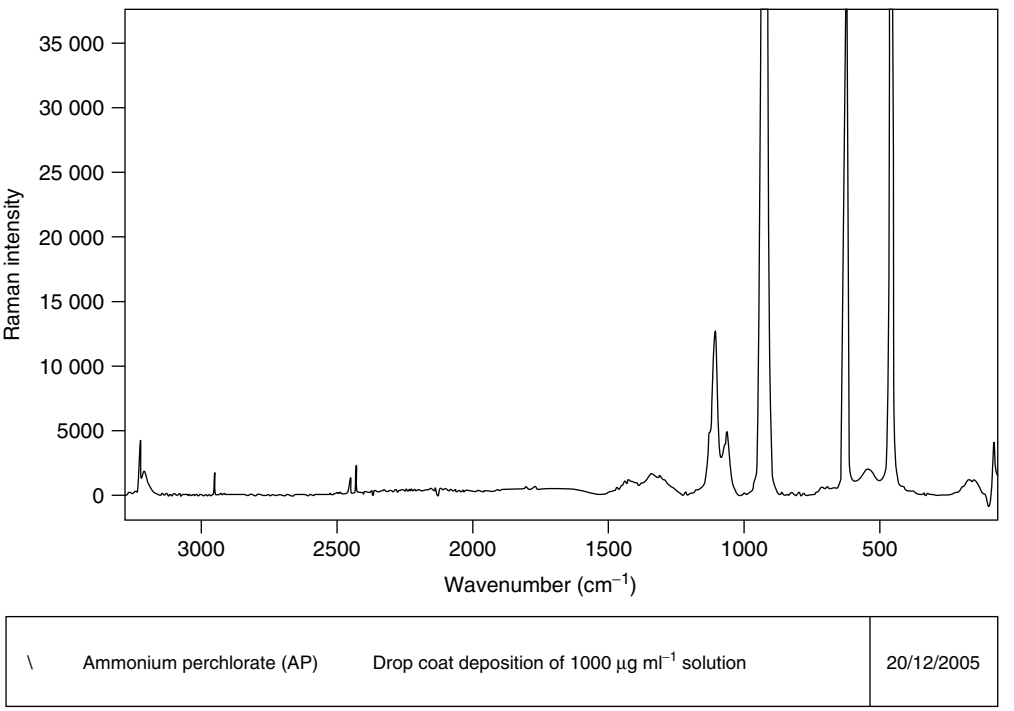


Figure A4 Raman spectrum of AP

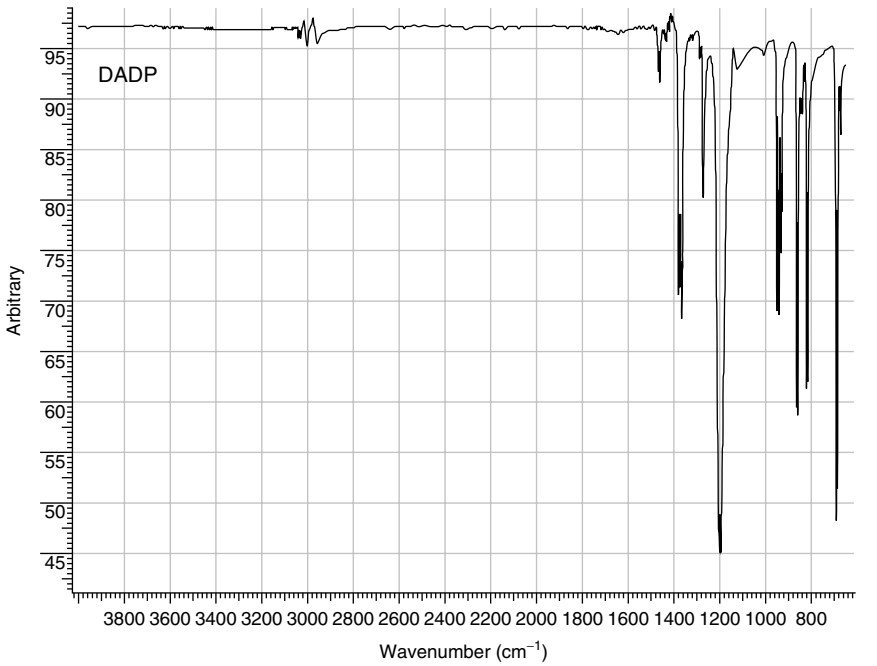


Figure A5 Infrared spectrum of diacetone diperoxide (DADP)

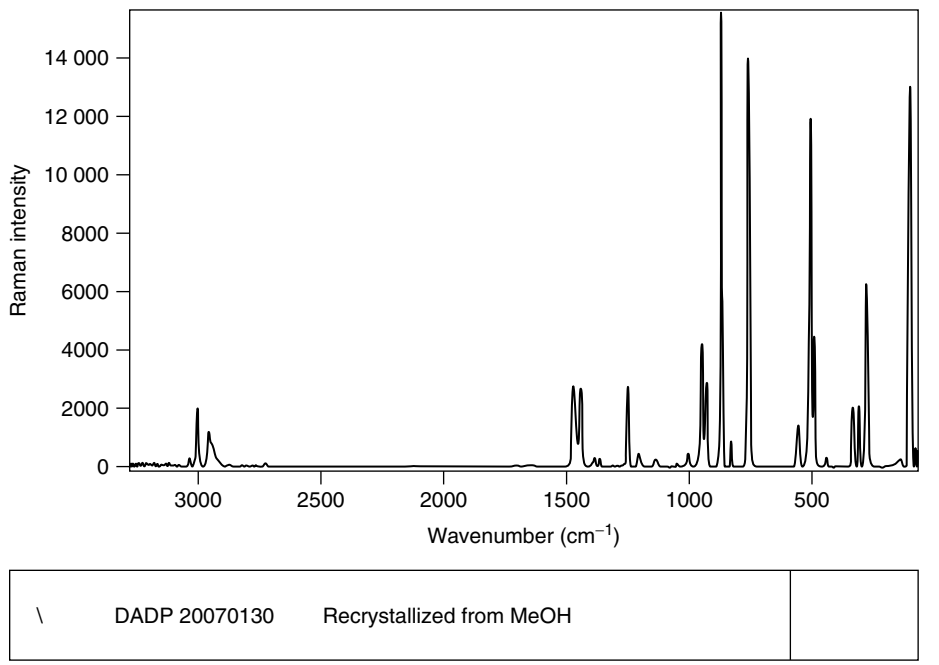


Figure A6 Raman spectrum of DADP

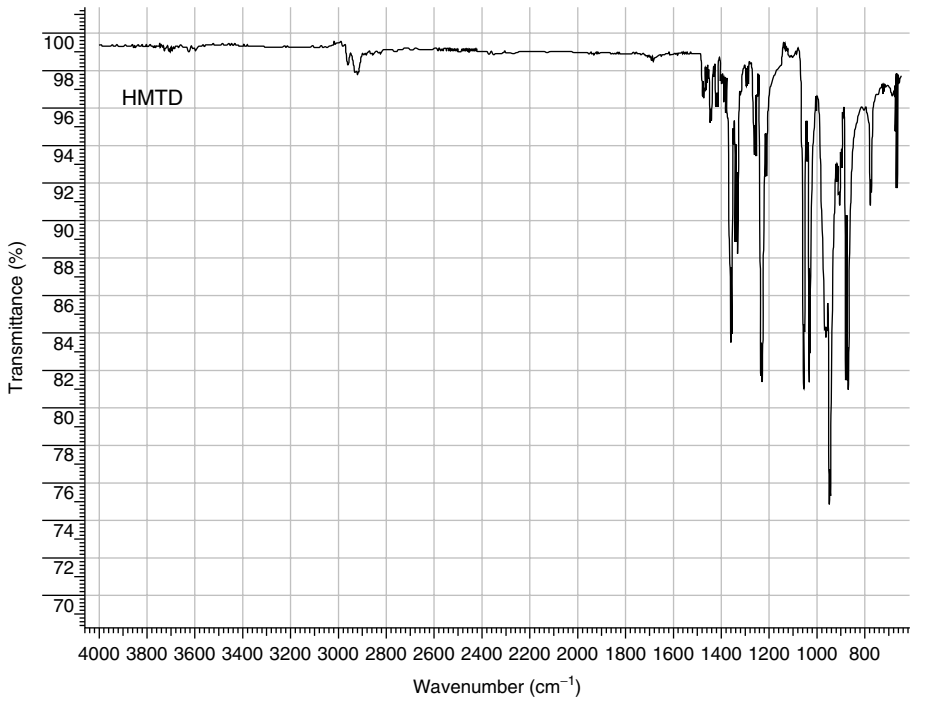


Figure A7 Infrared spectrum of hexamethylene triperoxide diamine (HMTD)

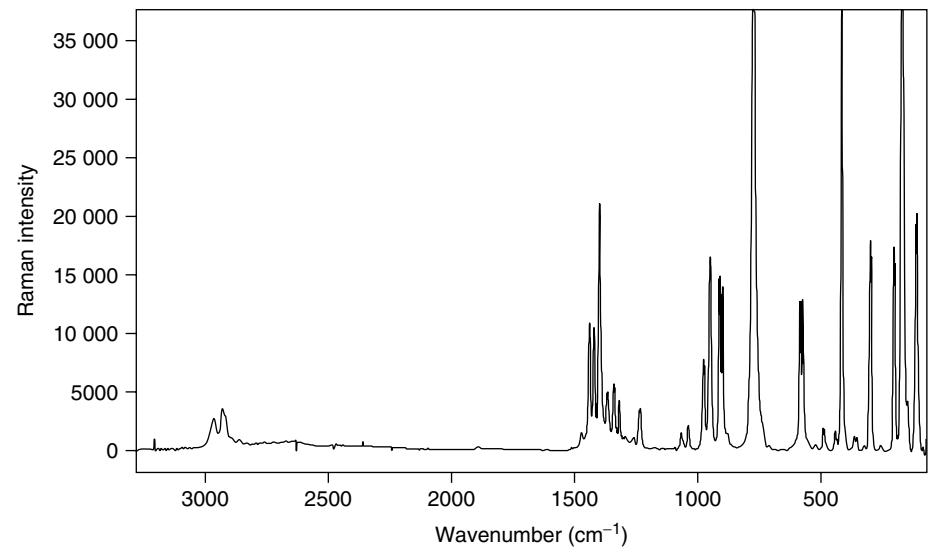


Figure A8 Raman spectrum of HMTD

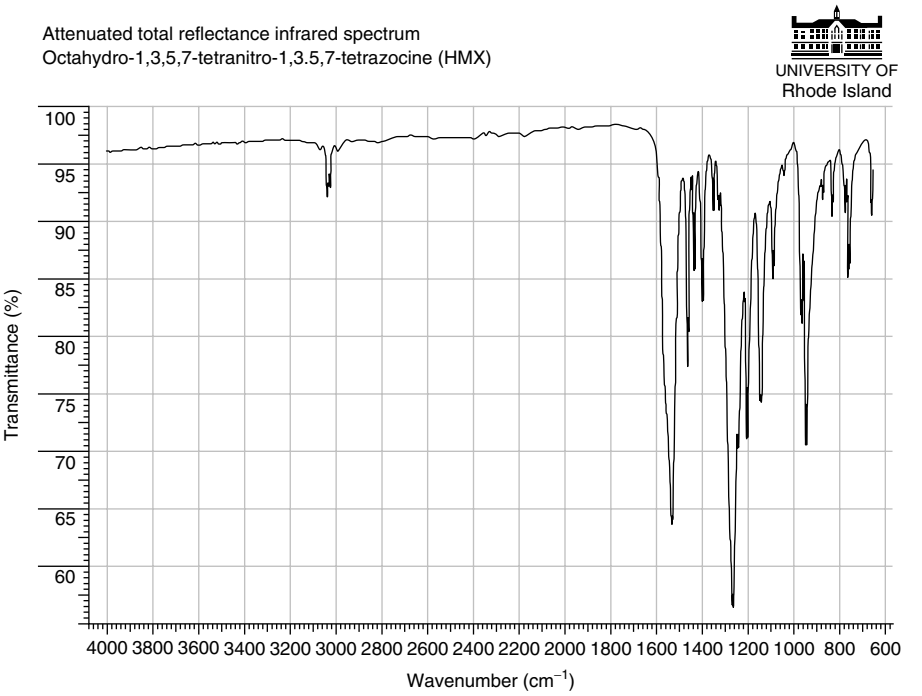


Figure A9 Infrared spectrum of octogen (HMX)

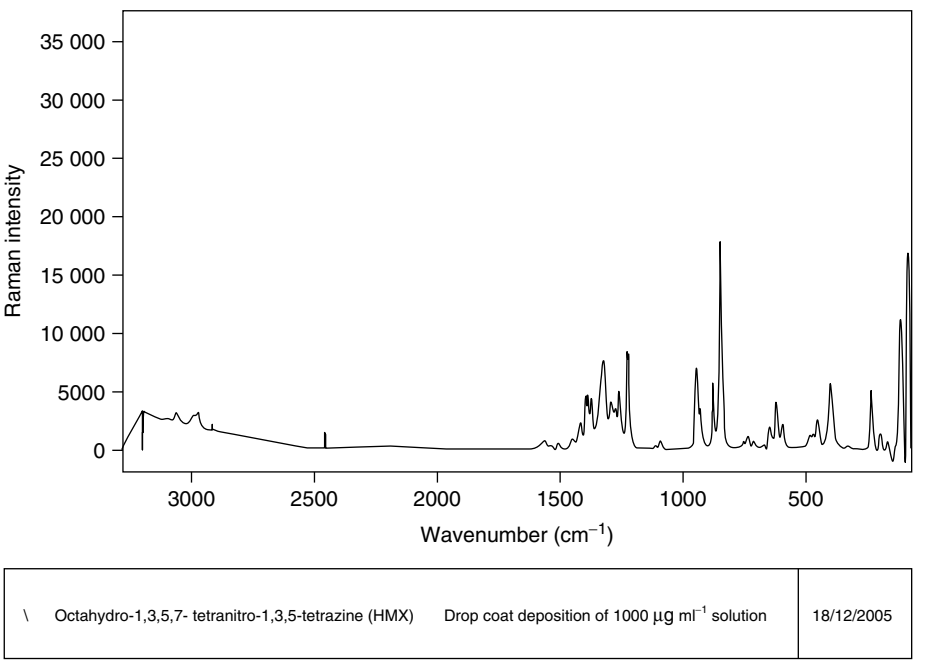


Figure A10 Raman spectrum of HMX

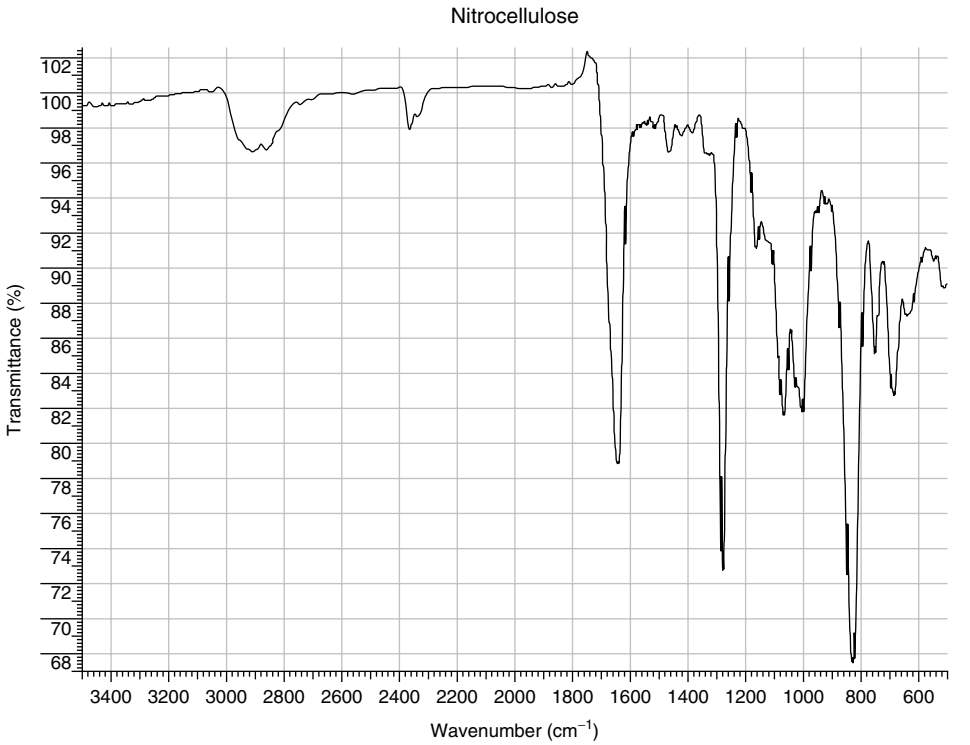


Figure A11 Infrared spectrum of nitrocellulose (NC)

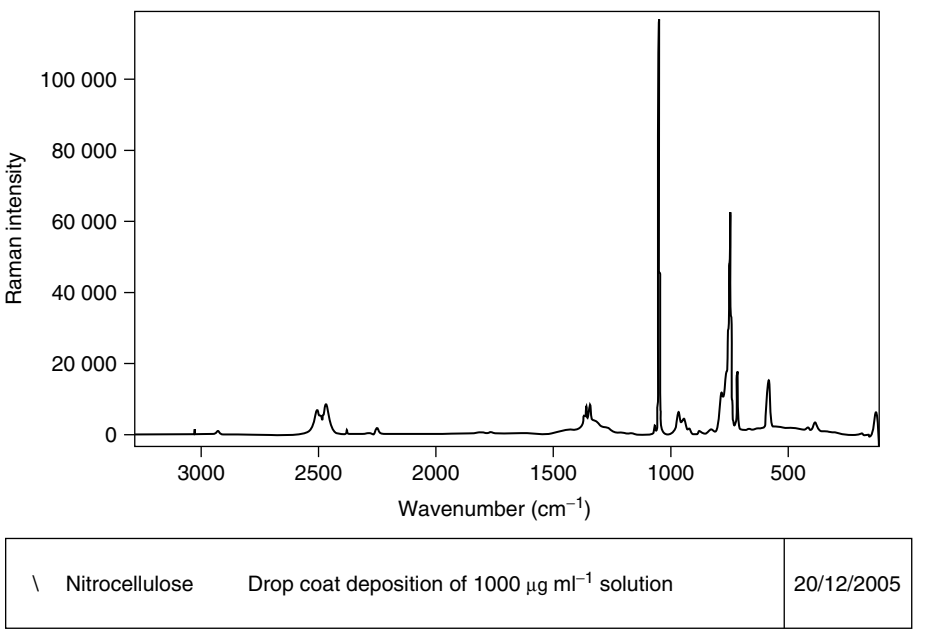


Figure A12 Raman spectrum of NC

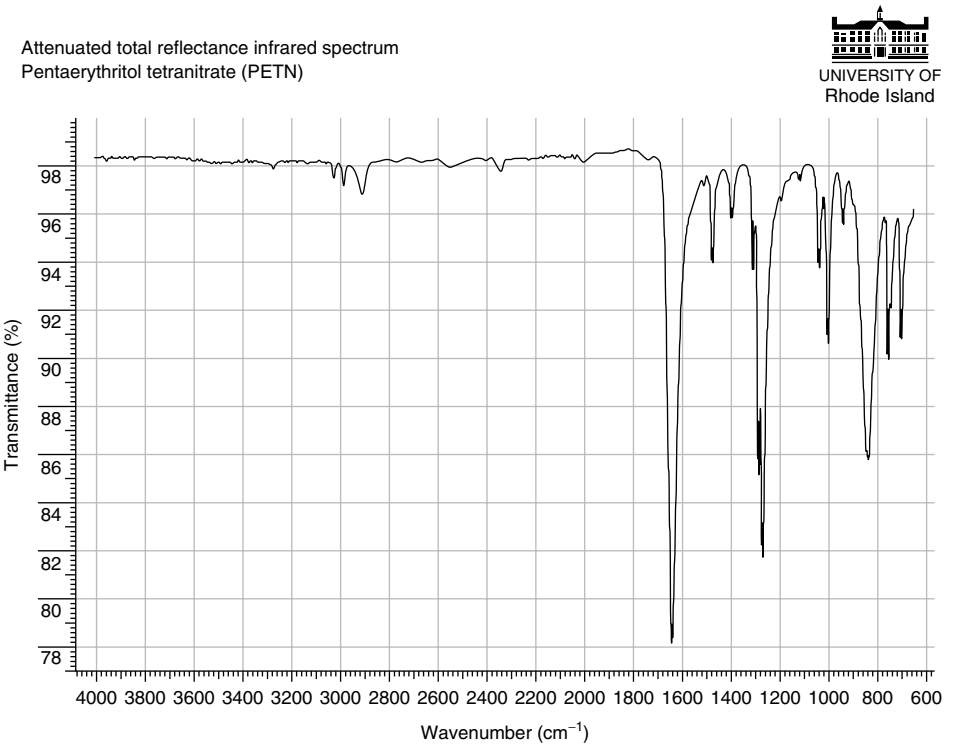


Figure A13 Infrared spectrum of pentaerythritol tetranitrate (PETN)

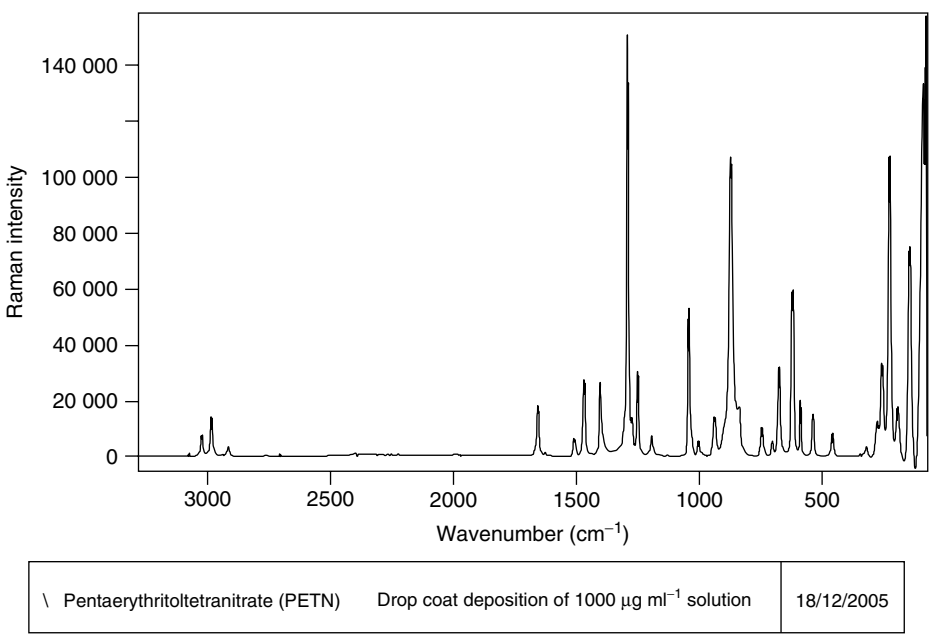


Figure A14 Raman spectrum of PETN

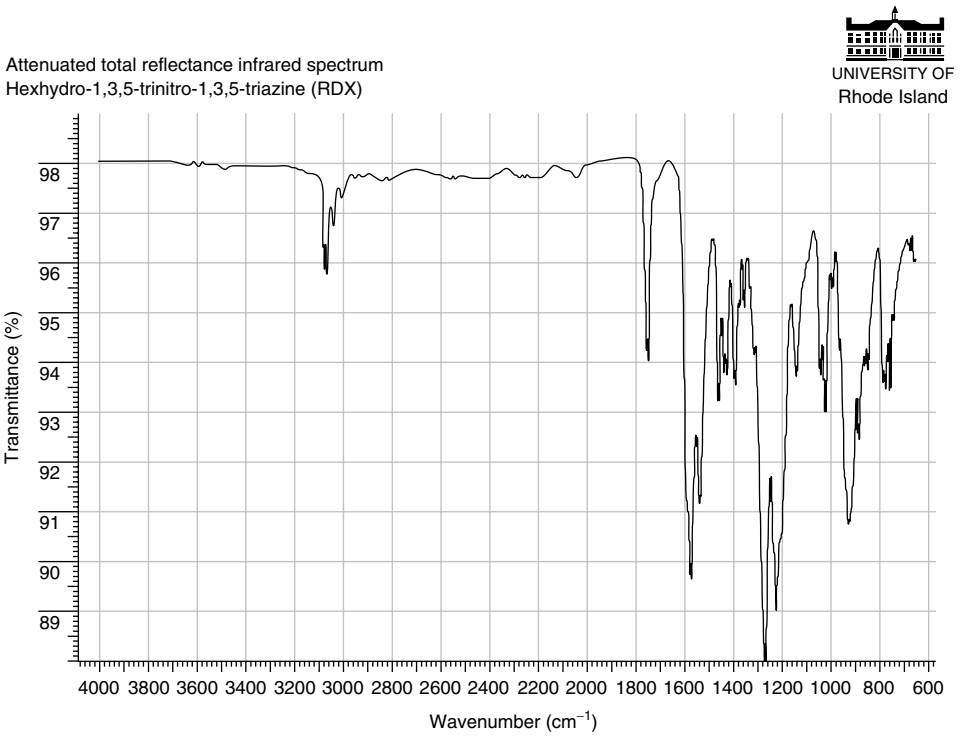


Figure A15 Infrared spectrum of hexogen (RDX)

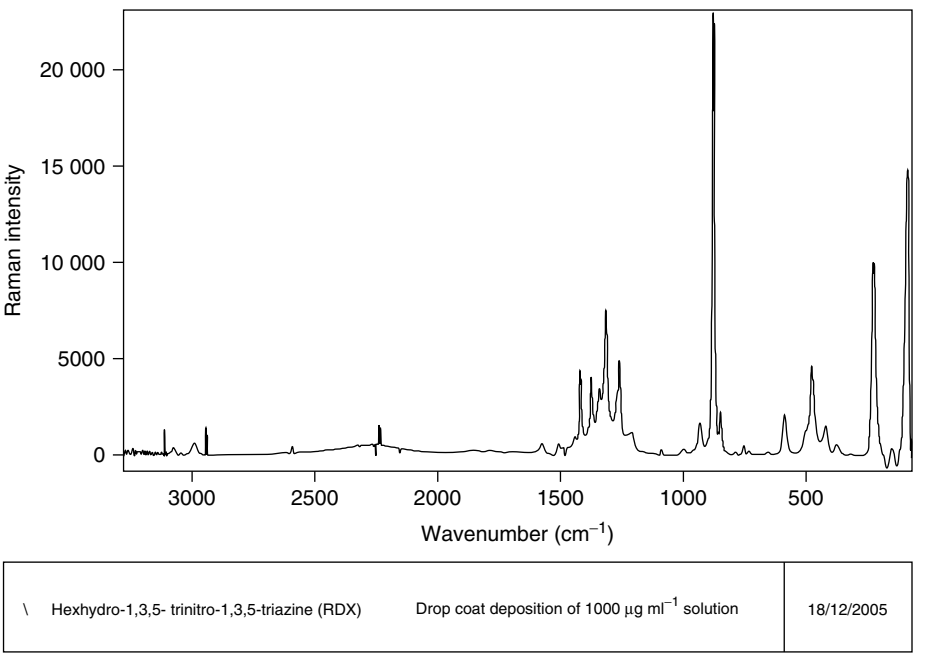


Figure A16 Raman spectrum of RDX

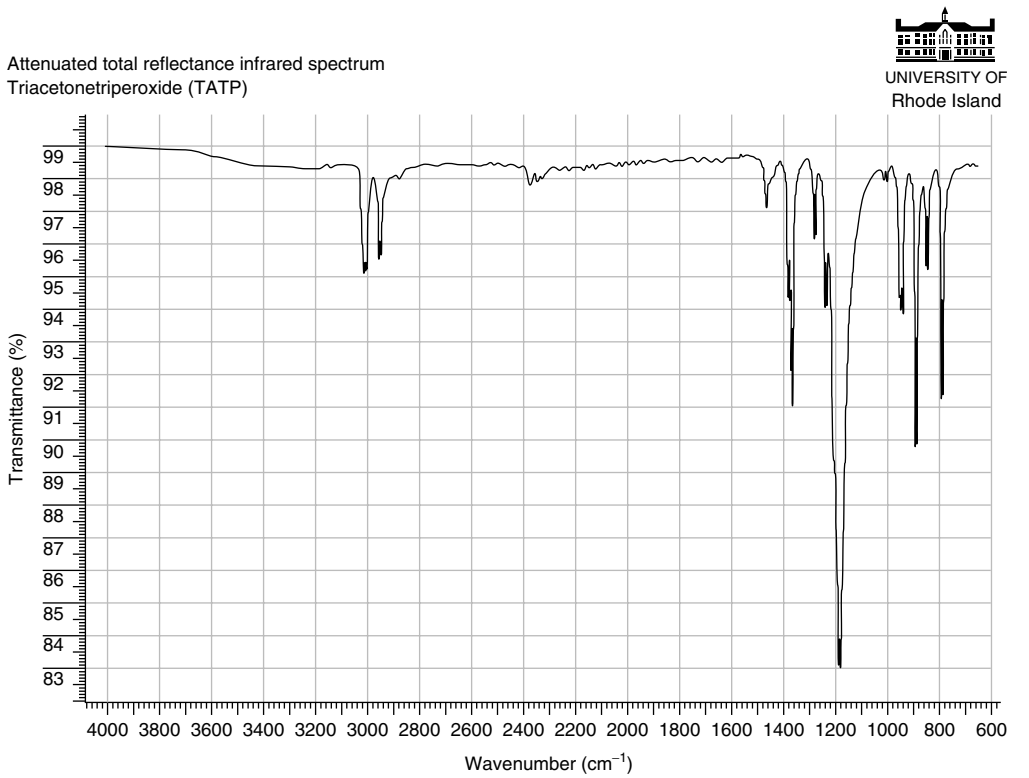


Figure A17 Infrared spectrum of triacetone triperoxide (TATP)

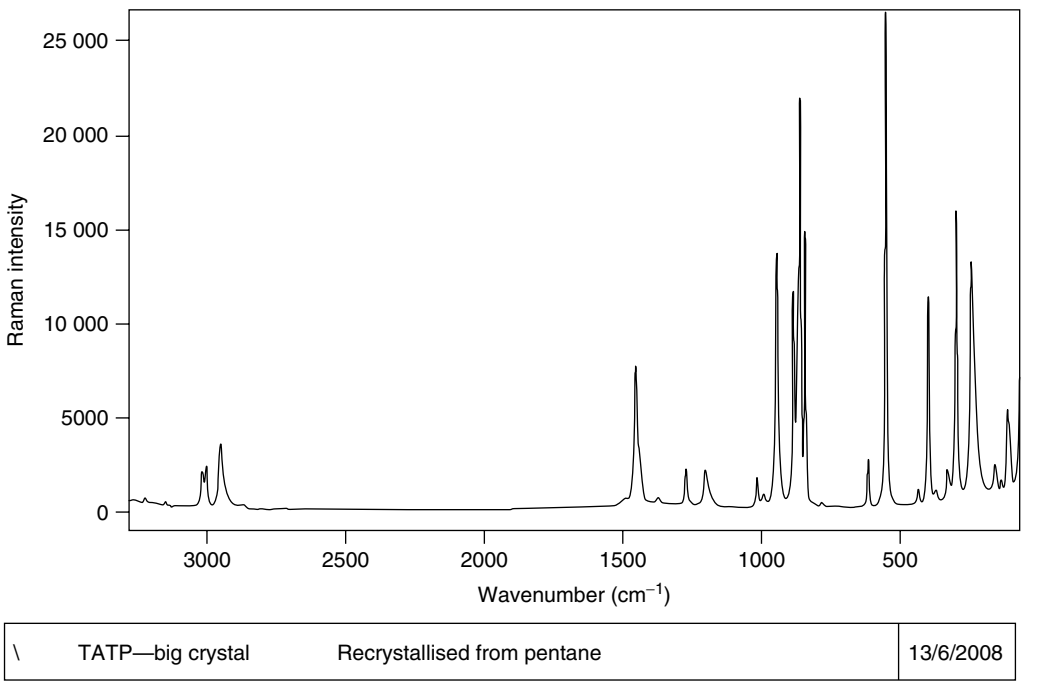


Figure A18 Raman spectrum of TATP

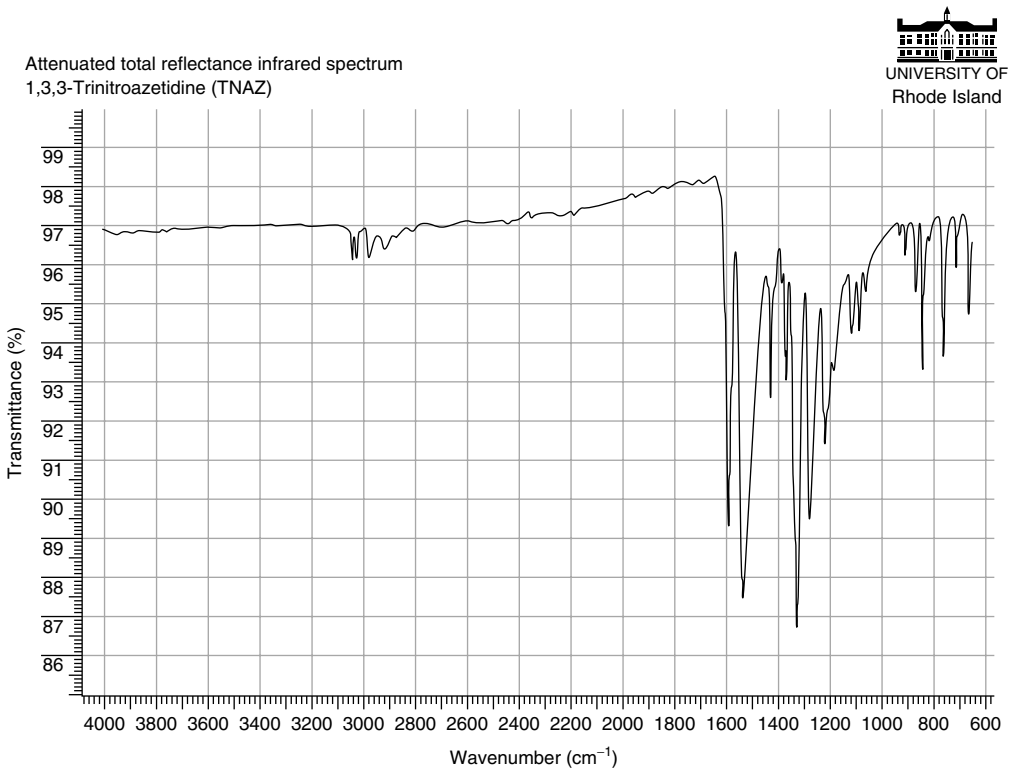


Figure A19 Infrared spectrum of 1,3,3-trinitroazetidine (TNAZ)

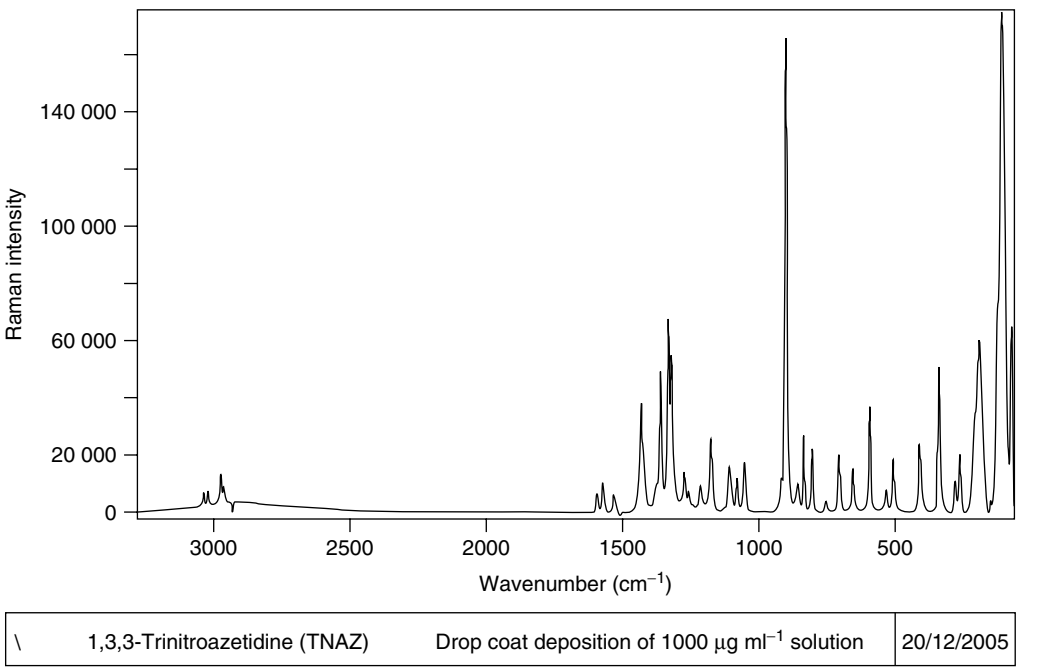


Figure A20 Raman spectrum of TNAZ

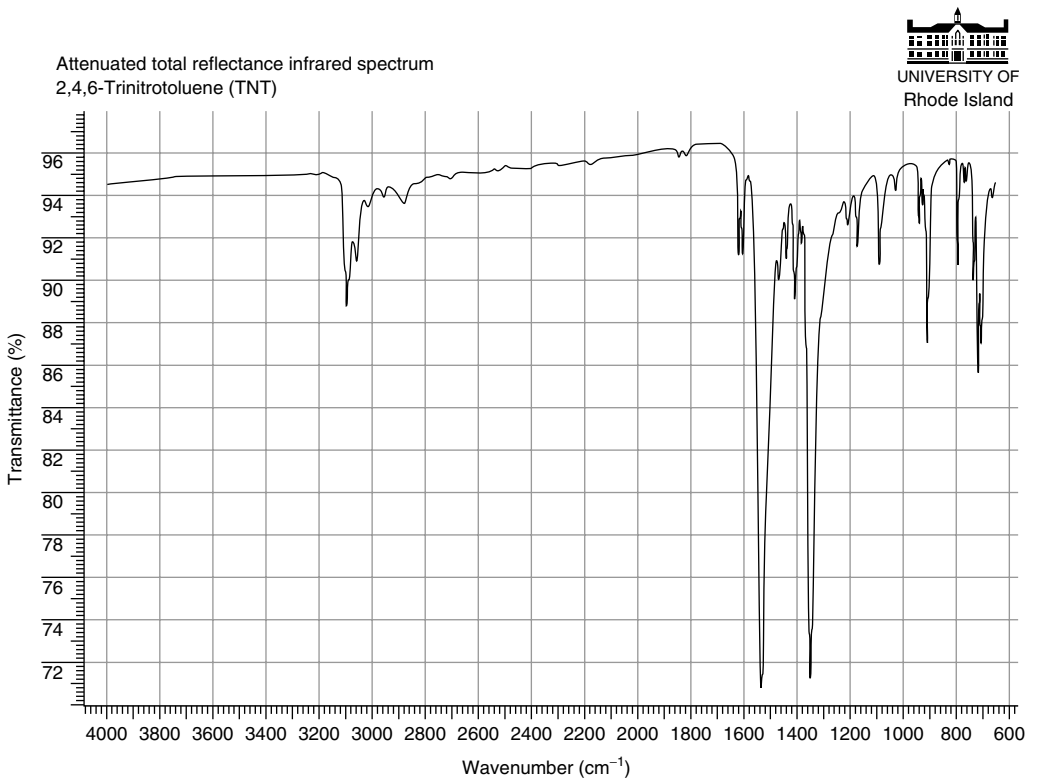


Figure A21 Infrared spectrum of 2,4,6-trinitrotoluene (TNT)

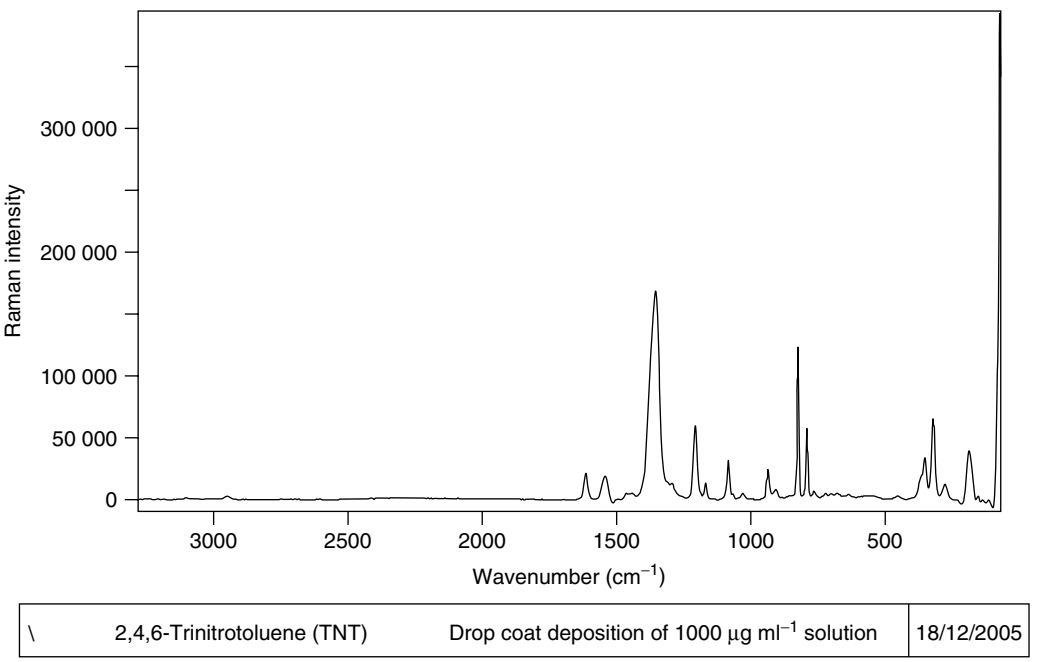


Figure A22 Raman spectrum of TNT

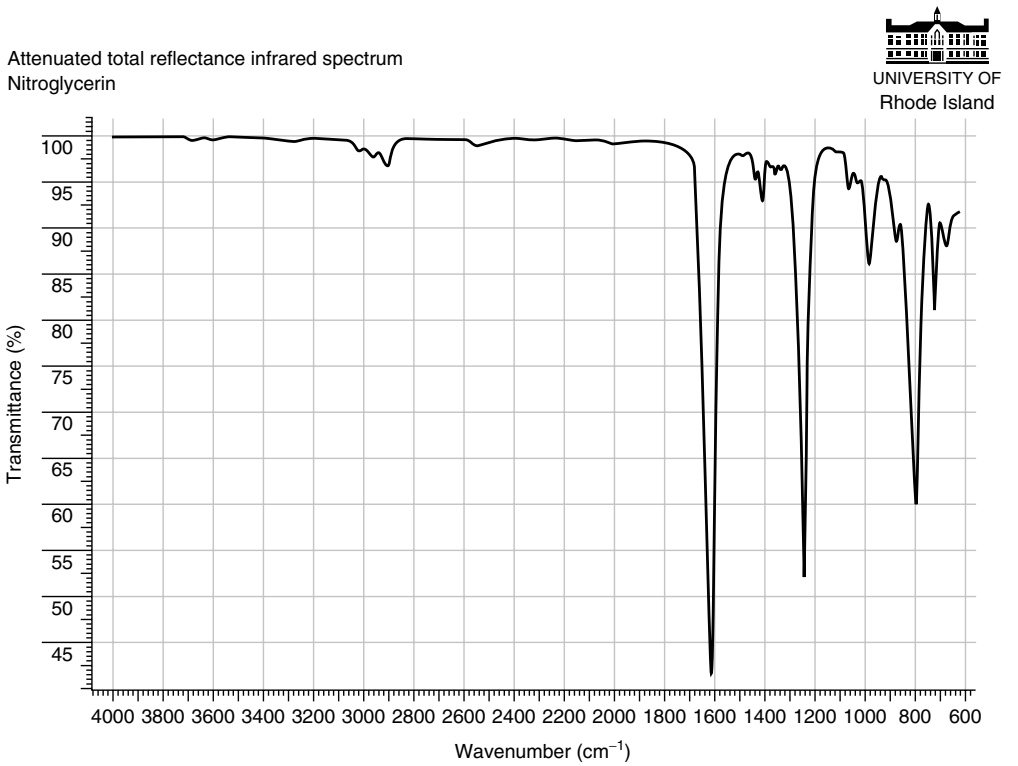


Figure A23 Infrared spectrum of nitroglycerine (NG)

References

[1] Reynolds, J., Nunes P., Whipple, R. & Alcaraz, A. (2006). *On-Site Analysis of Explosives in Various Matrices*, *Nato Security through Science Series*, Springer, Netherlands.

[2] Kelleher, J.D. (2002). Explosives residue: origin and distribution, *Forensic Science Communications* **4**(2).

[3] NIJ (2000). *A Guide for Explosion and Bombing Scene Investigations*, NCJ 181869 U.S. Department of Justice.

[4] National Center for Forensic Science *Instructors' Training Curriculum Guide to Explosion and Bombing Scene Investigations*, TWGFEX.

[5] Phillips, S.A., Lowe, A., Marshall, M., Hubbard, P., Burmeister, S.G. & Williams, D.R. (2000). Physical and chemical evidence remaining after the explosion of large improvised bombs. Part 1: firings of ammonium nitrate/sugar and urea nitrate, *Journal of Forensic Sciences* **45**(2), 324–332.

[6] Cullum, H., Lowe, A., Marshall, M. & Hubbard, P. (2000). Physical and chemical evidence remaining after the explosion of large improvised bombs. Part 2: firings of calcium ammonium nitrate/sugar mixtures, *Journal of Forensic Sciences* **45**(2), 333–348.

[7] Lowe, A.M., Marshall, M., Walker, C.L. & Hubbard, P. (2001). Physical and chemical evidence remaining after the explosion of large improvised bombs. Part 3: firings of calcium carbonate ammonium nitrate /sugar, *Journal of Forensic Sciences* **46**(3), 535–548.

[8] Oxley, J.C., Smith, J.L., Resende, E., Rogers, E., Strobel, R.A. & Bender, E.C. (2001). Improvised explosive devices: pipe bombs, *Journal of Forensic Sciences* **46**(3), 510–534.

[9] Gregory, O., Downey, M., Cumminskey, C., Platek, M., Oxley, J.C., Smith, J.L. & Bernier, E.T. (2008). Microstructural characterization of pipe bomb fragments, Accepted *Material Characterization*.

[10] Higgs, D.G., Jones, P.N., Markham, J.A. & Newton, E. (1978). A review of explosives sabotage and its investigation in civil aircraft, *Journal of the Forensic Science Society* **18**, 137.

[11] Walsh, G.A., Inal, O.T. & Romero, V.D. (2003). A potential metallographic technique for the investigation of pipe bombings, *Journal of Forensic Sciences* **48**(5), 484–500.

[12] Crowson, A., Cullum, H.E., Hiley, R.W. & Lowe, A.M. (1996). A survey of high explosives traces in public places, *Journal of Forensic Sciences* **41**(6), 980–989.

[13] Walker, C., Cullum, H. & Hiley, R. (2001). An environmental survey relating to improvised and emulsion/gel explosives, *Journal of Forensic Sciences* **46**(2), 254–267.

[14] George, V., Jenkins, T.F., Phelan, J.M., Leggett, D.C., Cragin, J.H., Webb, S.W., Oxley, J.C., Smith, J.L., Berry, T.E. & Miyares, P.H. (2000). Progress on determining the vapor signature of a buried landmine. *Proceedings of the 14th Annual International Symposium on Aerospace/Defense Sensing, Simulation, & Controls*, April 2000, Orlando, pp. 240–261, 258–269.

[15] Oxley, J.C., Smith, J.L., Foote, E. & Yue, J. Thermal decomposition of tnt and rdx in soil 1 & 2: batch studies at elevated temperature, *Journal of Environmental Management*, submitted.

[16] Beveridge, A.D. (1992). Development in the detection and identification of explosive residues, *Forensic Science Review* **4**, 17–47.

[17] Higgs, D.G. & Hayes, T.S. (1982). Post-detonation traces of nitroglycerine on polymeric materials: recovery and persistence, *Journal of Forensic Sciences* **22**, 343.

[18] Jenkins, R. & Yallop, H.J. (1970). The identification of explosives in trace quantities on objects near an explosion, *Explosivstoffe* **6**, 139–141.

[19] Fedoroff, B.T. (1960). *Encyclopedia of Explosives and Related Items*, PATR-2700, in ten volumes. US Army Armament Research and Development Command, Picatinny Arsenal, Dover.

[20] Dobratz, B.M. & Crawford, P.C. (1985). *LLNL Explosives Handbook*, UCRL-52997 Change 2. Lawrence livermore national laboratory, Livermore.

[21] Dionne, B.C., Rounbehler, D.P., Achter, E.K., Hobbs, J.R. & Fine, D.H. (1986). Vapour pressure of explosives, *Journal of Energetic Materials* **4**, 447–472.

[22] Oxley, J.C., Smith, J.L., Moran, J. & Shinde, K. (2005). Determination of the vapour density of triacetone triperoxide (tatp) using a gas chromatography headspace technique, *Propellants, Explosives, Pyrotechnics* **30**(2), 127–130.

[23] Oxley, J.C. (2003). The thermal stability of explosives, in *Handbook of Thermal Analysis and Calorimetry: Applications to Inorganic and Miscellaneous Materials Volume 2*, P.K. Gallagher & M.E. Brown, eds, Elsevier, Chapter 8, pp. 349–369.

[24] Persson, P. (1989). *Nitromethane Safety and Performance Issues*. Report for US Army AMCCOM contract no. DAAL 03-86-D-0001, New Mexico Institute of Mining and Technology, April 23.

[25] Meyer, R., Kohler, J., & Homburg, A. (2007). *Explosives*, 6th Edition, Wiley-VCH, Weinheim.

[26] Oxley, J.C., Smith, J.L. & Naik, S. Determination of urea nitrate and guanidine nitrate vapour pressures by isothermal thermogravimetry, Accepted to *Journal of Energetic Materials*.

[27] Esslinger, K.J., Siegel, J.A., Spillane, H. & Stallworth, S. (2004). Using STR analysis to detect DNA from exploded pipe devices, *Journal of Forensic Sciences* **49**(3), 481–484.

[28] White, P.C. (ed) (2004). *Crime Scene to Court: The Essentials of Forensic Science*, 2nd Edition, Royal Society of Chemistry, Cambridge.

[29] TWGFEX stands for Technical Working Group for Fire and Explosives; managed by the National Center for Forensic Science, University of Central Florida (http://ncfs.ucf.edu/twgfex/).

[30] Phillips, S. (2001). Pyrotechnic residue analysis – detection and analysis of characteristic particles by scanning electron microscopy/energy dispersive spectroscopy, *Science and Justice* **41**(2), 73–80.

[31] Kosanke, K., Dujay, R. & Kosanke, B. (2003). Characterization of pyrotechnic reaction residue particles by SEM/EDS, *Journal of Forensic Sciences* **48**(2), 531–537.

[32] Glattstein, B., Landau, E. & Zeichner, A. (1991). Identification of match head residues in post-explosion debris, *Journal of Forensic Sciences* **36**(5), 1360–1367.

[33] Yinon, J. & Zitrin, S. (1993). *Modern Methods & Applications in Analysis of Explosives*, Wiley, New York.

[34] Yinon, J. & Zitrin, S. (1981). *The Analysis of Explosives*, Pergamon Press, Oxford.

[35] Yinon, J. (ed) (2007). *Counterterrorist Detection Techniques of Explosives*, Elsevier.

[36] Beveridge, A. (ed) (1998). *Forensic Investigation of Explosions*, Taylor and Francis, pp. 315–342.

[37] Miller, F.A. & Wilkins, C.H. (1952). Infrared spectra and characteristic frequencies of inorganic ions, *Analytical Chemistry* **24**(8), 1253–1294.

[38] Pristera, F., Halik, M., Castelli, A. & Fredericks, W. (1960). Analysis of explosives using infrared spectroscopy, *Analytical Chemistry* **32**(4), 495–508.

[39] University of Rhode Island database see http://expdb.chm.uri.edu/.

[40] Smith, K.D., McCord, B.R., MacCrehan, W.A., Mount, K. & Rowe, W.F. (1999). Detection of smokeless powder residue on pipe bombs by micellar electrokinetic capillary electrophoresis, *Journal of Forensic Sciences* **44**(4), 789–794.

[41] Oxley J.C., Smith J.L., Higgins C., Bowden P., Moran J.S., Brady J., Aziz C.E. & Cox E. Efficiency of perchlorate consumption in road flares, propellants and explosives, *Journal of Environmental Management*, submitted.

[42] Oxley, J.C., Smith, J.L. & Valenzuela, B. (1995). Ammonium perchlorate decomposition: neat and solution, *Journal of Energetic Materials* **13**(1&2), 57–91.

[43] Baytos, J.F. (1991). *Field Spot-Test Kit for Explosives*, LA-12071-MS, Los Alamos National Laboratory, University of California, United States Department of Energy, contract W-7405-ENG-36, issued July.

[44] Parker, R.G., Stephenson, M.O. & McOwen, J.M. (1975). Analysis of explosives and explosive residues. Part 1: chemical tests, *Journal of Forensic Sciences* **20**(1), 133–139.

[45] http://www.sys-2.com/products_explosivetestkit.htm. (accessed Nov 2008).

[46] McCrone Research Institute and others provide information on this technique, e.g. see website at http://www.mcri.org/home/section/10-15-16-55/microscopy-of-explosives-(1722).

[47] Parker, R.G., McOwen, J.M. & Cherolis, J.A. (1975). Analysis of explosives and explosive residues, part 2: thin-layer chromatography, *Journal of Forensic Sciences* **20**(1), 254–256.

[48] Nam, S.-I. (1997). *On-Site analysis of explosives in soil; evaluation of thin-layer chromatography for confirmation of analyte identity*. CRREL Special Report, August.

[49] McKay, G.L. (2002). Forensic characterisation of organic peroxide explosives (TATP, HMTD, DADP), *Kayaku Gakkaishi* **63**(6), 323–329.

[50] Thompson, R.Q., Fetterolf, D.D., Miller, M.L. & Mothershead II, R.F. (1999). Aqueous recovery from cotton swabs of organic explosives residue followed by solid phase extraction, *Journal of Forensic Sciences* **44**(4), 795–804.

[51] www.epa.gov/testmethods/pdfs/8095.pdf EPA method 8095 explosives by gas chromatography. (accessed Nov 2008).

[52] Oxley, J.C., Smith, J.L., Resende, E., Pearce, E. & Chamberlain, T. (2003). Trends in explosive contamination, *Journal of Forensic Sciences* **48**(2), 1–9.

[53] Yinon, J. (2003). *Advances in Forensic Applications of Mass Spectroscopy*, CRC.

[54] Muller, D., Levy, A., Shelef, R., Abramovich-Bar, S., Sonenfeld, D. & Tamiri, T. (2004). Improved method for the detection of TATP after explosion, *Journal of Forensic Sciences* **49**, 935–938.

[55] Widmer, L., Watson S., Konrad Schlatter, K. & Crowson, A. (2002). Development of an LC/MS method for the trace analysis of triacetone triperoxide (TATP), *The Analyst* **127**, 1627–1623.

[56] Schulte-Ladbeck, R., Kolla, P. & Karst, U. (2003). Trace analysis of peroxide-based explosives, *Analytical Chemistry* **75**(4), 731–735.

[57] Oxley, J.C., Smith, J.L. & Chen, H. (2002). Decomposition of multi-peroxidic compound: triacetone triperoxides (TATP), *Propellants, Explosives, Pyrotechnics* **27**, 209–216.

[58] Oxley, J.C., Smith, J.L., Chen, H. & Cioffi, E. (2002). Decomposition of multi-peroxidic compounds: part II: hexamethylene triperoxide diamine (HMTD), *Thermochemica Acta* **388**(1–2), 215–225.

[59] Zhang, J., Oxley, J.C., Smith, J. & Cioffi, E. (2000). Mass spectra of unlabeled and isotopically labeled hexamethylene triperoxide diamine (HMTD), *Propellants, Explosives, Pyrotechnics* **25**, 1–4.

[60] Xu, X., Van De Craats, A.M., Kok, E. & De Bruyn, P.C.A.M. (2004). Trace analysis of peroxide explosives by high performance liquid chromatography – atmospheric pressure chemical ionization – tandem mass spectroscopy (HPLC-APCI-MS/MS) for forensic applications, *Journal of Forensic Sciences* **49**(6), 1230–1236.

[61] Tamiri, T. (2005). Characterization of the improvised explosive urea nitrate using electrospray ionization and atmospheric pressure chemical ionization, *Rapid Communications in Mass Spectroscopy* **19**, 2094–2098.

[62] Almog, J., Burda, G., Shloosh, Y., Abramovich-Bar, S., Wolf, E. & Tamiri, T. (2007). Recovery and detection of urea nitrate in traces, *Journal of Forensic Sciences* **52**(6), 1284–1290.

[63] Almog, J., Klein, A., Tamiri, T., Shloosh, Y. & Abramovich-Bar, S. (2005). A field diagnostic test for the improvised explosive urea nitrate, *Journal of Forensic Sciences* **50**(3), 582–586.

[64] Oxley, J.C., Smith, J.L. & Naik, S. (2008). Determination of urea nitrate and guanidine nitrate vapor pressures by isothermal thermogravimetry, Accepted *Propellants, Explosives, Pyrotechnics*.

[65] Oxley, J.C., Smith, J.L., Naik, S. & Moran, J.S. (2009). Decompositions of urea and guanidine nitrates, Accepted to *Journal of Energetic Materials*, Vol II.

[66] Oxley, J.C., Smith, J.L., Rogers, E. & Yu, M. (2002). Kinetic studies on ammonium nitrate formulations: the search for explosivity modifiers, *Thermochemica Acta* **384**(1–2), 23–45.

[67] Oxley, J.C., Smith, J.L. & Wang, W. (1994). Compatibility of ammonium nitrate with monomolecular explosives, part I, *Journal of Physical Chemistry* **98**, 3893–3900.

[68] Brower, K.R., Oxley, J.C. & Tewari, M.P. (1989). Homolytic decomposition of ammonium nitrate at high temperature, *Journal of Physical Chemistry* **93**, 4029–4033.

[69] Oxley, J.C., Smith, J.L. & Wang, W. (1994). Compatibility of ammonium nitrate with monomolecular explosives, Part II: nitroarenes, *Journal of Physical Chemistry* **98**, 3901–3907.

[70] Oxley, J.C., Smith, J.L., Ye, H., McKenney, R.L. & Bolduc, P.R. (1995). Thermal stability studies on homologous series of nitroarenes, *Journal of Physical Chemistry* **99**, 9593–9602.

[71] Oxley, J.C., Kooh, A., Szeckeres, R. & Zheng, W. (1994). Mechanisms of nitramines thermolysis, *Journal of Physical Chemistry* **98**, 7004–7008.

[72] Hiskey, M.A., Brower, K.R. & Oxley, J.C. (1991). Thermal decomposition of nitrate esters, *Journal of Physical Chemistry* **95**, 3955–3960.

[73] Fine, D.H., Yu, W., Goff, U.E., Bender, E.C. & Reutter, D.J. (1984). Picogram analysis of explosive residues using the thermal energy analyser (TEA), *Journal of Forensic Sciences* **29**, 732–746.

[74] Douse, J.M.F. (1985). Trace analysis of explosives at the low nanogram level in handswab extracts using columns of Amberlite XAD-7 porous polymer beads and silica capillary column gas chromatography with thermal energy analysis and electron capture detection, *Journal of Chromatography* **328**, 155–165.

[75] Gaurav, D., Malik, A.K. & Rai, P.K. (2007). High-performance liquid chromatographic methods for the

analysis of explosives, *Critical Reviews in Analytical Chemistry* **37**, 227–268.
[76] Sigman, M.E. & Ma, C.Y. (2001). Detection limits for GC/MS analysis of organic explosives, *Journal of Forensic Sciences* **46**(1), 6–11.
[77] Parker, R.G. (1975). Analysis of explosives and explosive residues. Part 3: monomethylamine nitrate, *Journal of Forensic Sciences* **20**(2), 257–260.
[78] Beardah, M.S., Doyle, S.P. & Hendey, C.E. (2007). Effectiveness of contamination prevention procedures in a trace explosives laboratory, *Science and Justice* **47**, 120–124.
[79] Oxley, J.C., Smith, J.L., Kirschenbaum, L., Shinde, K.P. & Marimganti, S. (2005). Accumulation of explosives in hair, *Journal of Forensic Sciences* **50**(4), 826–831.
[80] Oxley, J.C., Smith, J.L., Kirschenbaum, L. & Marimganti, S. (2007). Accumulation of explosives in hair: part ii: factors affecting sorption, *Journal of Forensic Sciences* **52**(6), 1291–1296.
[81] Hiley, R.W. (1998). Quality control in the detection and identification of traces of organic high explosives, in *Forensic Investigation of Explosions*, A. Beveridge, ed, Taylor & Francis, pp. 315–342.
[82] Crowson, A., Hiley, R.W. & Todd, C.C. (2001). Quality assurance testing of an explosive trace analysis laboratory, *Journal of Forensic Sciences* **46**(1), 53–56.
[83] Crowson, A., Doyle, S.P., Todd, C.C., Watson, S. & Zolnhofer, N. (2007). Quality assurance testing of an explosive trace analysis laboratory – further improvements, *Journal of Forensic Sciences* **52**(4), 830–837.

Further Reading

Oxley, J.C., Smith, J.L., Bernier, E.T., Sandstrom, F.W., Weiss, G.G., Recht, G.W. & Schatzer, D.S. (2008). Characterizing the performance of pipe bombs, *Journal of Forensic Sciences*, submitted.

Related Articles

Bomb Scene Management
Crime Scene Investigation
Crime Scene Management
Crime Scene Documentation
Crime Scene Photography: US Perspective
Evidence Interpretation: a Logical Approach
Explosions: Scene Investigation
Fire: Scene Investigation
Sampling and Estimation of Quantities
Sampling Trace Evidence
Trace Evidence: Transfer, Persistence, and Value
Training and Certification (in Criminalistics)

SARAH L. LANCASTER, MAURICE MARSHALL AND JIMMIE C. OXLEY

Explosions: Investigation of *see* Fire and Explosion Investigations: Overview

Extended Suicide *see* Suicide (Behavior)

Extraction

Introduction

One of the most crucial stages in the development of a DNA profile is the ability of the scientist to recover cellular material from a substrate and to extract DNA from it in a manner that is suitable for further processing steps such as polymerase chain reaction (PCR). Forensic samples are, however, not always easy to deal with. Samples can be deposited in the most inconvenient of places (*see* **DNA: Sources of**), exposed to the harshest of environments, and mixed with many potential inhibitors of DNA profiling. Many of these factors are outside the control of the forensic scientists.

An Ideal Method

A method suitable for the extraction of forensic samples requires the following:

- DNA is efficiently recovered from the biological sample;

- the extracted DNA is protected from further fragmentation or nuclease action;
- the extracted DNA is free of inhibitory substances that may have a detrimental effect on further processing; and
- the method minimizes the risks of cross-contamination from other samples or sources.

In recent years, as additional types of forensic evidence are developed and implemented, methods are also required that enable the following:

- the coextraction of DNA and RNA and
- the extraction of DNA from plant material of various kinds.

Typical steps in the process are as follows:

- visual examination and presumptive tests to identify the biological source of the sample;
- recovery of biological material;
- lysis of the cells and release of the DNA into solution;
- removal of inhibitory compounds;
- determination of the amount of DNA recovered; and
- if necessary, concentration of the DNA extract.

Any method developed for forensic use must be validated according to the accreditation or other quality system in place in the laboratory. Each of the methods described in this article can include numerous, small, laboratory-specific variations, such as incubation times and temperatures; however, the basic steps are common to all procedures.

Proteinase K/Organic Extraction

The earliest method described for the recovery of DNA from forensic samples involved the use of the enzyme proteinase K, followed by an organic extraction with a solution of phenol and chloroform [1]. Despite the fact that this is a time-consuming method, has multiple transfer steps, and uses hazardous chemicals, it remains a favorite in many forensic laboratories.

Cellular material is recovered from the substrate by scraping, soaking, or other means (*see* **DNA: Sources of**). A simple Tris/HCl buffer containing proteinase K and a detergent such as sodium dodecyl sulfate (SDS) is added to the cell pellet. The proteinase is an enzyme that breaks open the cell and nuclear membranes, releasing the cell contents including the DNA into solution. The phenol/chloroform step coupled with an ethanol precipitation or alternate concentration step successfully removes inhibitors and results in a DNA extract amenable for further analysis. It is particularly useful when small amounts of DNA are expected, such as swabs taken from car steering wheels or touched items.

The DNA that is recovered is double stranded DNA, relatively free of impurities, and can be used for many applications (*see* **Variable Number Tandem Repeats**).

Differential Extraction

Forensic laboratories typically employ a differential extraction method for the separation of spermatozoa from other cells such as vaginal epithelia [1]. Strong disulfide bonds prevent lysis (bursting or breaking up) of the spermatozoa nuclei in the presence of proteinase K and SDS, (a detergent), allowing the more fragile epithelial and other cells to lyse and release their DNA. Separation of the spermatozoa from the lysed cells is achieved by centrifugation and removal of the supernatant (the liquid portion sitting above the separated solids) containing the lysed cellular components. The spermatozoa can then be lysed by the addition of a reducing agent such as dithiothreitol (DTT) in the presence of the extraction buffer, proteinase K, and SDS. Each fraction, the "sperm" and "epithelial" fractions, can then be processed separately typically by organic extraction, phenol/chloroform extraction, ethanol precipitation, or other methods (see below).

Alternative approaches to the separation of sperm cells from epithelial cells include the use of laser microdissection [2] and cell sorting using a glass microdevice method [3]. These approaches have potential for integration into more automated extraction protocols.

Chelex®

With the advent of PCR, Chelex® extraction became widely used [4]. This method generates single stranded DNA suitable for PCR and became the method of choice in many laboratories because of its simplicity, low cost, and lack of harmful reagents.

Chelex®100 is a resin composed of styrene divinylbenzene copolymers containing paired iminodiacetate ions that act as chelating groups. These chelating groups have a high affinity for polyvalent (multiple charges) metal ions such as Mg^{2+} ions. Thus, degradation of DNA in the sample is minimized by chelation of the metal ions inactivating the nuclease enzymes that digest DNA. Chelex® beads in suspension are added to cells recovered from a substrate, and after incubation, the DNA is released from the cells by boiling. The alkalinity of the solution and the high temperature cause the cell membranes to burst, releasing the DNA.

Chelex® is inhibitory to the PCR reaction and it has been observed [5] that long-term storage of samples containing Chelex® beads results in a deterioration of the DNA profile quality.

While Chelex® is suitable for a wide range of sample types, additional clean up and concentration steps are often required for dirty or samples containing very small amounts of DNA. Nevertheless, Chelex® extraction remains a popular choice in forensic laboratories.

Solid Phase Extraction Technologies

In recent years, alternative technologies suitable for manual or automated applications have become more widely used. These methods can also be employed after the separation of spermatozoa and epithelial cells. One of the earliest such methods to be used was DNA IQ™ [6–9].

DNA IQ™

The DNA IQ™ system utilizes a two-step method. First, the cellular material is recovered from the substrate by incubation of the sample to 95 °C with a lysis buffer. Then DNA IQ™, a silica-coated, paramagnetic resin, is added to the sample. The resin binds to DNA with high affinity in the presence of chaotropic elements such as guanidinium, which lyse the cells, denature proteins, and inhibit DNases. A magnet is then applied, which causes the resin, with DNA attached, to congregate on the tube wall nearest to the magnet. DNA is then recovered from the resin using an elution buffer. DNA can be recovered from most forensic sample types, using proteinase K to assist in the extraction of DNA from hair, bone, and spermatozoa [9] and is reported to be particularly suitable for very small samples.

The amount of DNA captured by the resin is roughly proportional to the amount of resin used and can be varied according to the sample type. This has advantages when extracting DNA from DNA database samples where the amount of starting material is large, as a consistent amount of DNA can be extracted and further expensive quantitation steps may not be required [8]. The DNA IQ™ system has been successfully implemented onto robotic platforms [10] and has been used by many laboratories to extract DNA from casework samples, for example, from volume crime cases [7] and envelope flaps [11].

MagAttract®

The MagAttract® system (Qiagen Inc.) also utilizes paramagnetic silica beads to capture DNA after release from cells using chaotropic agents. A similar lyse, bind, wash, and elute protocol is followed [12, 13]. Again, the final elution volume can be altered depending on the concentration of DNA likely to be recovered. This technology has been optimized for use on several robotic platforms, enabling laboratories to customize specific protocols.

Charge Switch™ Technology

Charge Switch™ (DNA Research Innovations Ltd., UK) also uses magnetic bead technology. In this case, however, the magnetic beads are coated with a unique chemical compound that acts as an ionic switch that is dependent on pH. The overall process is the same as that employed by other magnetic technologies. First, the cells are recovered from the substrate and lysed. Then at pH below 6.5, the DNA from lysed cells binds to the particles that are subsequently captured using a magnet. The pH is raised to 7 to allow a washing step to take place before the pH is raised again to 8. At this pH, the charge is neutralized and the DNA is eluted. Again, modifications to the elution volume can be made to alter the concentration of the recovered DNA and the technology has been shown to be applicable to a wide variety of forensic samples [14].

QIAamp Technology

In 1998, Greenspoon *et al.* [5] described the use of QIAamp® spin columns (Qiagen Inc.) as an

alternative extraction method for the extraction of DNA from database samples resulting in DNA that survived long-term storage at −20 °C. In this process, cells are lysed and then placed into a spin column containing a silica-gel-based membrane to which the DNA binds. Following washing steps, aided by centrifugation or vacuum suction depending on the protocol being followed, the DNA is eluted into solution ready for quantitation and/or amplification.

A recent addition to the QIAamp® spin columns is the use of an additional column, known as a *QIAshredder® homogenizing column*, after cell lysis and prior to the binding of the sample to the silica membrane [15]. The use of such columns enhanced the effectiveness of the QIAamp® spin column process, increasing the performance of this extraction system relative to either a Chelex® extraction or an organic extraction.

Solutions for Specific Sample Types

FTA Card

Whatman FTA® cards are extensively used for the collection of reference samples for forensic analysis. The FTA® technology lyses cells on contact, releasing and immobilizing the DNA. An added advantage is that infectious pathogens and other microorganisms are deactivated, enabling long-term room temperature storage of samples [16].

Once the sample is immobilized onto the card, a small punch is taken typically 2 mm across. A simple series of washing steps follows, which removes any inhibitory substances that might be present. The washed punch is then placed directly into the amplification. This method is easily automated [17].

Feces

The extraction of DNA from fecal remains presents particular challenges. It has been shown that swabbing of the outer surface of the stool [18] or collection of tissue paper smeared with feces [19] offers the best chance of success. In the extraction process, it is important to remove as much inhibitory material as possible, particularly the polyphenolic compounds [20]. A number of specific solutions have been developed for the successful extraction of DNA from feces; these include the QIAamp® DNA stool Mini Kit (Qiagen Inc.) and the Ultraclean™ Fecal DNA Isolation Kit (MO BIO Laboratories Inc.).

Extraction from Bone

There are many methods described in the literature for the extraction of DNA from samples of bone (for examples see [21–23]). The most significant difference between the methods described is whether or not a demineralization step is used. There are common themes to most of the methods. First, most bones can be successfully cleaned of surface dirt and potential contaminating substances including extraneous DNA using one or more of the following methods; abrasion with sandpaper etc. The cleaned pieces of bone are then crushed into small fragments, often with the aid of a freezer mill and liquid nitrogen until a fine powder is obtained.

The powder can be demineralized using several changes of 0.5 M ethylene diamine tetra-acetic acid (EDTA), pH 7.5 at 4 °C with agitation over one to two days [21, 22] or as part of the lysis procedure [23]. For most of the methods described, subsequent organic extraction and concentration with a column-based method such as a Centricon™ 30 microconcentrator is suggested.

Teeth

Like bones, teeth are relatively easy to clean prior to extraction. Some complexity is introduced when it is necessary to preserve the external hard surfaces of the tooth, which is important if physical and biochemical tests are also to be carried out. Methods of extraction that do not crush or damage the physical structure of the tooth are required, although this may reduce the maximum recovery of DNA [24].

Hair

As with teeth and bones, hair can be cleaned prior to extraction. One suggestion is the use of sonication in 2% SDS, followed by rinsing in water prior to extraction as an effective method [25]. One method used to optimize the recovery of DNA from root and shafts of telogen hairs (*see* **Hair: Microscopic Analysis**) is to grind the hair after washing using glass grinders, followed by Chelex® extraction. Alternatively, organic extraction methods [26] or solid phase extraction methods such as DNA IQ™ can be used.

Formaldehyde-fixed and Paraffin-embedded Tissues

Formaldehyde-fixed and paraffin-embedded tissues provide a valuable resource for forensic analysis.

Removal of the paraffin can be accomplished by washing the tissue in xylene [27] or by microwave treatment of the paraffin blocks to "melt" the paraffin [28]. A variety of extraction methods can then be used, including organic extraction and Chelex® procedures.

Coextraction of DNA and RNA

Recently, there has been interest shown in the development of RNA-based methods for the identification of body fluids in forensic samples [29]. To maximize the usefulness of such techniques, it is important that both DNA and RNA can be obtained simultaneously from the same sample and methods that accomplish this have been published [30, 31]. In brief, these methods include steps to protect the fragile RNA from RNases and methods to selectively precipitate DNA in the presence of RNA. As the development of the RNA-based methods is in its infancy and has not been widely adopted by forensic laboratories to date, it is likely that refinements to these methods will be made.

Extraction of Plant Material

The use of plant material as a source of forensic DNA evidence has also been recognized. Many methods for the extraction of DNA from plant material are based on the use of hexadecyltrimethlyammonium bromide (CTAB) [32], although these have been more recently replaced by custom-designed commercial kits such as DNeasy Plant Kits from Qiagen Inc. [33].

Contamination

No account of the extraction of DNA from forensic samples would be complete without some mention of contamination. Lessons learned from the study of ancient specimens [34] describe the risks of contamination and precautions that can be taken to minimize the risks. As DNA is recovered from ever smaller amounts of cellular material, this is of utmost importance. Recent adoption of low copy number DNA (*see* **Low Copy Number DNA**) techniques in laboratories typically includes the monitoring of the laboratory environment for extraneous DNA, which has highlighted the importance of protective clothing, frequent changes of gloves, and the decontamination of equipment and surfaces.

References

[1] Gill, P., Jeffreys, A.J. & Werrett, D.J. (1985). Forensic application of DNA fingerprints, *Nature* **318**, 577–579.

[2] Elliot, K., Hill, D.S., Lambert, C., Burroughes, T.R. & Gill, P. (2003). Use of laser micro-dissection greatly improves the recovery of DNA from sperm on microscope slides, *Forensic Science International* **137**, 28–36.

[3] Horsman, K., Barker, S.L.R., Ferrance, J.P., Forrest, K.A., Koen, K.A. & Landers, J.P. (2005). Separation of sperm and epithelial cells on micro-fabricated devices: potential application to forensic analysis of sexual assault evidence, *Analytical Chemistry* **77**, 742–749.

[4] Walsh, P.S., Metzger, D.A. & Higuchi, R. (1991). Chelex®100 as a medium for the simple extraction of DNA for PCR-based typing of forensic material, *Biotechniques* **10**, 506–513.

[5] Greenspoon, S.A., Scarpetta, M.A., Drayton, M.L. & Turek, S.A. (1998). QIAamp spin columns as a method of DNA isolation for forensic casework, *Journal of Forensic Sciences* **43**, 1024–1030.

[6] Greenspoon, S.A., Ban, J.D., Sykes, K., Ballard, E.J., Edler, S.S., Baisden, M. & Covington, B.L. (2004). Application of the Biomek®2000 laboratory automation workstation and the DNA IQ™ system for use in the DNA extraction of high volume forensic casework, *Journal of the Canadian Society for Forensic Science* **49**, 29–39.

[7] Komonski, D.J., Marignani, A., Richard, M.L., Frappier, J.R.H. & Newman, J.C. (2004). Validation of the DNA IQ IQ™ system for use in the DNA extraction of high volume forensic casework, *Journal of the Canadian Society for Forensic Science* **37**, 103–109.

[8] Promega Corporation (2002). *DNA IQ™ System – Database Protocol*, Promega Corporation, Madison, WI.

[9] Promega Corporation (2002). *DNA IQ™ System – Small Sample Casework Protocol*, Promega Corporation, Madison, WI.

[10] McLaren, B. (2006). Automation in a forensic laboratory: an update, *Profiles in DNA* **9**, 19–20.

[11] Ng, L.-K., Ng, A., Cholette, F. & Davis, C. (2007). Optimization of recovery of human DNA from envelope flaps using DNA IQ™ system for STR genotyping, *Forensic Science International Genetics* **1**, 283–286.

[12] Kishore, R., Hardy, W.R., Anderson, V.J., Sanchez, N.A. & Buoncristiani, M.R. (2006). Optimisation of DNA extraction from low yield and degraded samples using the Bio Robot® EZ1 and Bio Robot® M48, *Journal of Forensic Sciences* **51**, 1055–1061.

[13] Montpetit, S.A., Fitch, I.T. & O'Donnell, P.T. (2005). A simple automated instrument for DNA extraction in forensic casework, *Journal of Forensic Sciences* **50**, 1–9.

[14] Taylor, M., Bridge, C. & Baker, M. (2005). Improved sensitivity for forensic DNA purification using Charge Switch® Technology, *DNA Research Innovations Ltd. Quest* **2**, 51–54.

[15] Castella, V., Dimo-Simonin, N., Brandt-Casadevall, C. & Mangin, P. (2006). Forensic evaluation of the QIAshredder/QIAamp DNA extraction procedure, *Forensic Science International* **156**, 70–73.
[16] Burgoyne, L., Kijas, J., Hallsworth, P. & Turner, J. (1994). *Proceedings from the Fifth International Symposium on Human Identification*, Promega Corporation, Madison, WI, p. 163.
[17] Belgrader, P., Del Rio, S.A., Turner, K.A., Marino, M.A., Weaver, K.R. & Williams, P.E. (1995). Automated DNA purification and amplification from blood-stained cards using a robotic workstation, *Biotechniques* **19**, 426–432.
[18] Johnson, D.J., Martin, L.R. & Roberts, K.A. (2005). STR-typing of human DNA from human fecal matter using the Qiagen QIAamp® Stool mini kit, *Journal of Forensic Sciences* **50**, 802–808.
[19] Roy, R. (2003). Analysis of human fecal material for autosomal and Y chromosome STRs, *Journal of Forensic Sciences* **48**, 1035–1040.
[20] Monteiro, M., Bonnemaison, D., Vekris, A., Petry, K.G. & Bonnet, J. (1997). Complex polysaccharides as PCR inhibitors in feces, *Journal of Clinical Microbiology*, **35**, 995–998.
[21] Hochmeister, M.N., Budowle, B., Borer, U.V., Comey, C.T. & Dirnhofer, R. (1991). Typing of deoxyribonucleic acid (DNA) extracted from compact bone from human remains, *Journal of Forensic Sciences* **36**, 1649–1661.
[22] Hagelberg, E., Sykes, B. & Hedges, R. (1989). Ancient bone DNA amplified, *Nature* **342**, 485.
[23] Loreille, O.M., Diegoli, T.M., Irwin, J.A., Coble, M.D. & Parson, T.J. (2007). High efficiency DNA extraction from bone by total demineralisation, *Forensic Science International Genetics* **1**, 191–195.
[24] Smith, B.C., Fisher, D.L., Weedn, V.W., Warnosk, G.R. & Holland, M.M. (1993). A systematic approach to the sampling of dental DNA, *Journal of Forensic Sciences* **38**, 1194–1209.
[25] Linch, C.A., Smith, S.L. & Prahlow, J.A. (1998). Evaluation of the human hair root for DNA tying subsequent to microscopic comparison, *Journal of Forensic Sciences* **43**, 305–314.
[26] Hellmann, A., Rohleder, U., Schmitter, H. & Wittig, M. (2001). STR typing of human telogen hairs – a new approach, *International Journal of Legal Medicine* **114**, 269–273.
[27] Banerjee, S.K., Makdisi, W.F., Weston, A.P., Mitchell, S.M. & Campbell, D.R. (1995). Microwave-based DNA extraction from paraffin-embedded tissue for PCR amplification, *Biotechniques* **18**, 768–773.
[28] Gill, P., Kimpton, C.P. & Sullivan, K. (1992). A rapid polymerase chain reaction method for identifying fixed specimens, *Electrophoresis* **13**, 173–175.
[29] Juusola, J. & Ballantyne, J. (2003). Messenger RNA profiling: a prototype method to supplant conventional methods for body fluid identification, *Forensic Science International* **135**, 85–96.
[30] Alvarez, M., Juusola, J. & Ballantyne, J. (2004). An mRNA and DNA co-isolation method for forensic casework samples, *Analytical Biochemistry* **335**, 289–298.
[31] Chromczynski, P. (1993). A reagent for the single step simultaneous extraction of RNA, DNA and proteins from cell and tissue samples, *Biotechniques* **15**, 532–534.
[32] Doyle, J.J. & Doyle, J.L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue, *Phytochemistry Bulletin* **19**, 11–15.
[33] Bless, C., Palmeter, H. & Wallace, M.M. (2006). Identification of *Acer rubrum* using amplified fragment length polymorphism, *Journal of Forensic Sciences* **51**, 31–38.
[34] Lindahl, T. (1997). Facts and artifacts of ancient DNA, *Cell* **90**, 1–3.s

SALLY-ANN HARBISON

Extraction: Differential *see*
Differential Extraction

Eyewitness: Suggestibility of

Eyewitnesses can be quite accurate in their memory for a witnessed event, but many factors related to eyewitness suggestibility can influence the accuracy of eyewitness accounts. Eyewitness suggestibility refers to the tendency for individuals to incorporate misleading information into their memories for a witnessed event, which may lead to inaccurate eyewitness accounts (*see also* **Hypnosis and Memory**; **Memory: Repressed**). Typically the misinformation occurs on a postevent basis. The process by which one may become an eyewitness includes several stages, all of which may lead to errors in memory for the witnessed event.

The order of the phases as a witness experiences them include: (i) witnessing the event; (ii) possible postevent discussion with other witnesses; (iii) exposure to postevent information, potentially including misinformation, during the delay and/or when

remembering the episode, such as during police questioning; and (iv) a time delay between witnessing the event and being asked to recollect the event. These sources of misleading information may come from external sources, such as postevent narratives, questionnaires, cowitnesses, and investigative interviews, as well as internal sources, such as autosuggestion. The purpose of this article is to review these sources of adult eyewitness suggestibility as they pertain to the process by which witnesses may encounter various intervening events that can skew their memory for the original event. Although many of these and other factors related to eyewitness testimony have been studied with children (*see* **Children: as Witnesses**) and older adults (*see* **Eyewitness Testimony**), we have limited our review to the research on healthy, young adults. However, we included a brief overview of research designed to assess individual differences in adult suggestibility as it pertains to eyewitness memory; in this overview, age differences are noted. Additionally, we have focused our review on six major areas of forensically relevant research in eyewitness memory that reflect both the classic study of and current trends in eyewitness memory research: misleading postevent information and the misinformation effect, cowitness contamination, time delay, suggestive interviewing, individual differences in suggestibility, and false memory formation.

History and Theories of Adult Eyewitness Memory

Alfred Binet (1900) was one of the first to scientifically study the role of suggestive questioning on memory for objects and events [1]. He formally distinguished between *suggestion*, or external, deliberate misinformation, and *autosuggestion*, or internal, self-generated memory distortion. He reported that both can lead to memory errors. In 1908, Hugo Münsterburg published his seminal book *On the Witness Stand*, in which he described errors in witness memory and suggestive interviewing that may bring about false confessions [2]. Münsterburg's writings set the stage for modern scientific investigation of faulty eyewitness accounts, the study of forensic interviewing, and the nature of obtaining confessions. Both Binet's and Münsterburg's works emphasized the fallibility of memory in light of misinformation and foreshadowed theoretical explanations of why eyewitness memory is often inaccurate. In 1932, Sir Frederic C. Bartlett published the results of numerous experiments on memory errors made when participants recalled meaningful material, including memory for (witnessed) faces [3]. Bartlett's work serves as a methodological prelude to much modern eyewitness memory research. Furthermore, he theorized that memory is a reconstructive process rather than verbatim reproductions of what was experienced and therefore prone to errors.

It was not until the early 1970s that psychologists began again to formally investigate eyewitness memory within the context of Bartlett's influential theory of reconstructive memory. Elizabeth Loftus and her colleagues developed a research methodology with which to study the role of suggestive misleading information on memory [4]. As illustrated in Figure 1, subjects in such experiments witness an event, are presented with misleading information, and are tested for their memory for the event. For example, in a classic study by Loftus *et al.* (1978), subjects in both control and misled groups viewed an event depicting an automobile-pedestrian accident, including the view of a stop sign at the intersection between two streets [5]. The subjects answered a series of questions, including one critical question. The control group was asked "Did another car pass the red Datsun while it was stopped at the stop sign?" and thus did not receive any misinformation. The misled group was asked the same question, but "yield sign" replaced "stop sign" in the question; thus, the misled group received misinformation about the sign at the intersection. Later, all participants were administered a two-alternative forced choice recognition test. In the key choice they were shown two slides, one of the car stopped at a stop sign (from the original event) and one of the car stopped at a yield sign. In this study and in numerous other studies, Loftus and her colleagues found that compared to control conditions, misled subjects were much more likely to erroneously remember misleading information as original event information, a phenomenon called the *misinformation effect* [5–8].

Loftus and her colleagues argued that the memory trace for the originally witnessed information was altered and replaced by misinformation via reconstructive processes. Others researchers soon criticized this trace alteration position on grounds that using the methodological procedure shown in Figure 1 to test the misinformation effect produces a bias

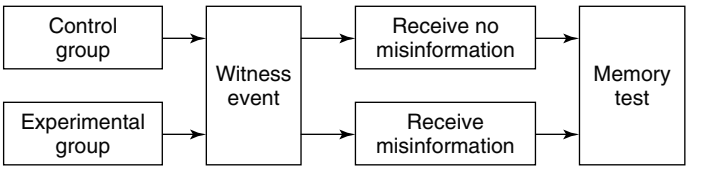


Figure 1 Research methodology used to study the misinformation effect

toward memory for misinformation, and further contended that original event memory traces are still intact but not retrieved, possibly because the misled information is more recent and interferes or blocks accessibility to original details [9, 10]. This notion captures the essence of the coexistence hypothesis and soon researchers began to use a more general term, namely, "*memory impairment*", which was introduced to generically refer to multiple forms of memory distortion while not implying any neuropsychological processes.

One theoretical explanation of memory impairment whether due to suggestion or autosuggestion focuses on inaccurate monitoring of the source of information, leading to confusion between original event information and misinformation [11, 12]. As a result, witnesses may misattribute the source of the misleading information to the original event, producing the misinformation effect. Another recent theory proposed to explain internally generated confusions in memory and resultant false memories for an event is the FTT, which posits that individuals create two independent memory traces for the original event, a verbatim trace and a gist trace [13]. The verbatim trace includes an exact memory representation for the event, while the gist trace includes a schema-based representation of the event. When an individual witnesses an event, both verbatim and gist traces are formed. Initially, access to strong verbatim traces allows adults to produce quite accurate recollections and to simultaneously suppress memory errors. However, verbatim traces fade quickly overtime [14] such that many details of the event are forgotten, but gist traces remain fairly intact. Thus, one's recollection for an original event typically may rely more on gist representations than on verbatim representations. Because gist does not fade overtime as much as the exact details, and as gist is schematic, memory for the original event may be more prone to error when one relies on gist more than verbatim traces.

Extensions of the basic misinformation paradigm shown in Figure 1 have led to current research on sources of adult suggestibility that includes investigation of misleading information in the form of cowitness suggestions and suggestive investigative interviewing. Likewise, changes in memory for the event may occur without deliberate misinformation and instead develop internally via autosuggestion.

Sources of Adult Suggestibility

As we have chronicled above, the psychological study of adults' susceptibility to accepting misleading information as accurate, original event information, includes the scientific investigation of cowitness suggestion, suggestive interviewing by police investigators, and autosuggestion. For instance, witnesses speaking about an event before reporting the event can alter one's memory for the original event and lead to cross-contamination of witness reports [15, 16]. Recently researchers have begun to uncover the extent to which discussing an event with others influences memory. For example, one approach has been to utilize a confederate (an experimenter's assistant who poses as a subject in the study) to provide accurate and inaccurate information after a witnessed event, usually during a memory test, in order to sway the real subject to conform to the confederate's answers. A few studies have used the latter approach and have found that in most situations, individuals conform to the confederate's reports and provide inaccurate information about the event that they witnessed [16–21]. Postevent information in the form of witness discussion may bias a witness to report the postevent information, which may be inaccurate, especially if they do not trust their own memory. Furthermore, sometimes witnesses may be unaware of the influence of others' recollections of the event. In this case, a witness may unknowingly misattribute what others said as a memory for what she/he originally saw or heard. In the course of a

forensic investigation, information supplied by witnesses to investigators may serve as the basis for further police interviews, construction of lineups, and formulation of theories about the witnessed event. Therefore, it is important that a witness not be contaminated by others' view of what happened. After witnessing an event and potentially talking with other witnesses, usually witnesses are interviewed by law enforcement personnel. It is during these interviews when witnesses may be prone to suggestion and changes in memory for the event may result. Furthermore, the eyewitness interview or interrogation can lead to extreme effects of eyewitness suggestibility in the form of false confessions (*see* **Interrogative Suggestibility**) and false eyewitness identification (*see* **Eyewitness Lineups: Identification from**), both of which are beyond the scope of this article.

The scientific study of suggestive interviewing includes investigation of factors that are related to faulty eyewitness memory. Such factors include, but are not limited to, the wording of investigative questions, the time delay between the witnessed event and recall or recognition of the event, the repetition of misinformation, and individual differences in eyewitness memory. In an early example, Loftus and Palmer (1974) presented subjects with a video of a car accident, then later asked them a series of questions about the event [6]. The key question involved a manipulation of the wording of the question in which different subjects were asked to estimate the speed of the moving car that *contacted, hit, bumped* or *smashed* the other car. Subjects who heard *smashed* estimated faster speeds than subjects who heard other verbs. One week later, all subjects returned and were asked additional questions, including whether there was broken glass at the scene of the accident. Those individuals who heard *smashed* were more likely to recall broken glass than those who heard other verbs; in fact, there was no broken glass. Loftus and Palmer's study illustrates how the mere wording of questions can alter one's memory representation for a witnessed event. Theoretically, misleading information may be mistaken for original event information and erroneously remembered as what was witnessed, indicating a source monitoring confusion. Thus, the misleading postevent information creates a bias in memory for recollection of misinformation. From the vantage point of FTT, the misleading event may make it difficult to retrieve an original verbatim trace and/or provides a more recent verbatim representation that is gist-consistent and may successfully compete with the original information at test.

The time delay, or retention interval, between witnessing an event and reporting about the event may lead to inaccuracies in eyewitness memory [22, 23]. For example, subjects in Odinot and Wolter's (2006) study viewed a video of an event and were tested either one, three, or five weeks later. They found that the longer retention intervals resulted in fewer items accurately recalled and more items inaccurately recalled. The decrease in accuracy is not surprising, as this result is consistent with a long history of research on the relationship between retention interval and forgetting [14, 24]. However, subjects also recalled more inaccurate items with a longer delay, in line with our earlier discussion that they may have relied more on the gist of the event. Once again FTT predicts just this sort of dissociative effect of memory; with time, verbatim traces fade (accurate recall decreases with a delay) but the gist traces may remain intact but are schema-based and more prone to supporting memory error (inaccurate recall increases with a delay). Similar results have been found when misinformation is presented to witnesses. The retention interval between witnessing an event and receiving misinformation and the delay between receiving misinformation and reporting from memory what was witnessed both can lead to memory impairment [5, 23, 25, 26].

The study of individual differences in eyewitness memory and testimony include examination of the effects of age on suggestibility and subsequent eyewitness errors. Generally speaking, children are more susceptible to suggestion than adults and a review of this topic is found in the Encyclopedia article Children as Witnesses referred to earlier. Results with older adults have been mixed. Many researchers have found that older adults are more susceptible to suggestive postevent information and as a consequence make more eyewitness memory errors than younger adults [27–30]. However, there are some exceptions in the research literature, in which like the aforementioned studies, older adults demonstrate poorer overall memory accuracy, but they are not more susceptible to the effects of misleading information [31, 32]. Loftus *et al.* (1992) examined numerous other individual differences variables, including gender, occupation, and educational level [29]. Aside from clear age effects, Loftus *et al.*'s results were mixed.

Another approach to studying individual differences in adult suggestibility to misleading information involves a formal test of suggestibility. Gudjonsson (1987, 1997) developed the Gudjonsson Suggestibility Scale II (GSS2) in which individuals are first exposed to a story, and then after a delay are asked a series of questions about the story [33, 34]. After being told that they made many mistakes on their first test, individuals are questioned again. The GSS2 provides two scores that indicate an individual's susceptibility to suggestion: "Yield", which is the tendency to accept misinformation as reality, and "Shift", which is the tendency to change one's answers in response to negative feedback and social pressure. The sum of Yield and Shift scores provides a measure of total suggestibility. Much of the research that utilizes the GSS2 is related to personality characteristics [35, 36] and the interrogation of witnesses [37–39]. For example, Baxter *et al.* demonstrated that subjects were more likely to "shift" their memory reports to accept more misinformation when the interviewer displayed a firm as opposed to a friendly demeanor [37]. In an examination of cross-cultural differences, Pollard *et al.* found that individuals in the United Kingdom were more likely to report misinformation from the onset than their US counterparts, and individuals in the United States were more likely to "shift" their answers compared to the UK sample [39]. Polczyk *et al.* found that older adults were more likely to report misinformation (Yield) than young adults, but older adults did not differ from young adults in the likelihood that they would "shift" their answers in response to negative feedback and social pressure [38]. The GSS2 is an assessment of memory suggestibility in general (*see also* **Interrogative Suggestibility**); however, another example of individual differences in suggestibility involves the variability of response to hypnotic suggestion that may or may not invoke retrieval of true memories (*see* **Deception: Truth Serum**).

False Memories

Inaccuracies in eyewitness memory, whether due to suggestion or self-generated, autosuggestion, all result in distortions that make the formation of false memories for the witnessed event quite likely. False memories due to autosuggestion are considered spontaneous false memories as opposed to those implanted via overt suggestion. The study of spontaneous false memories has proliferated in the past decade and has dominated the use of a methodological procedure known as the *Deese–Roediger–McDermott* (*DRM*) procedure in which subjects are asked to remember schema-based word lists designed to elicit illusory memories [40, 41]. For example, subjects may study *bed, rest, awake, tired, dream*, etc. then are asked to recall or recognize list words. Typical results are that subjects falsely remember a critical (lure) word (e.g., *sleep*) that was not presented but is strongly associated to the list words [42]. Furthermore, much like the research on eyewitness suggestion, subjects falsely remember lure words with high confidence [43] and report details about their (non)occurrence in the list [44, 45]. Also, spontaneous false memories persist overtime [46, 47], a fact consonant with the notion that with delays people are relying on gist traces.

The formation of false memories has been explained by two major theories: FTT and activation-monitoring. FTT posits that false memories occur because of retrieval of gist memory traces that were repeatedly strengthened during encoding as each presented item converges on the theme [13]. Activation-monitoring suggests that false memories occur as a result of a combination of relational (gist-based) processing at study and poor memory monitoring at retrieval [12, 48]. The study of false memories using word lists provides researchers with a methodology that produces high rates of false memories analogs to those studied in eyewitness situations. Furthermore, because word list studies are highly controlled from a procedural standpoint, compared to other stimulus materials they provide greater opportunities to test theories of false memory formation. Such theoretical tests allow researchers to better predict the circumstances under which false memories occur and also how false memories can be reduced. It is important to note that while we have focused on the DRM paradigm, highly similar false recollections occur with more ecologically valid materials including pictures, stories, narratives, and classroom lectures [47].

Summary

Psychologists have also examined factors that can reduce the effects of suggestibility in eyewitness memory. Warnings about misinformation have been shown to reduce the effects of misinformation, but such effects depend on when and how the warning is

administered [49–51]. Likewise, warnings of spontaneous false memories have been found to reduce false memories when the warning is administered after study [52] but such results have been mixed. Nonsuggestive interviewing techniques, such as in the cognitive interview, increase accuracy of memory for a witnessed event [53, 54].

Within the past 10 years, the courts and criminal justice personnel have become more open to considering psychological research on eyewitness memory, particularly with the advent of forensic DNA testing. Wells *et al.* (1998) and Scheck and Neufeld (2000) reported that more than 100 people convicted prior to the use of DNA testing have since been exonerated, and more than 75% of those falsely convicted were convicted on the basis of eyewitness testimony [55, 56]. Currently, the Innocence Project currently reports over 200 exonerations due to DNA evidence (www.innocenceproject.org). Wells *et al.* (2000) reported that, consequentially, there has been a rise in the use of eyewitness memory research in the criminal justice system [57]. Factors that affect and have the potential to reduce adult suggestibility continue to be explored to further our understanding and to inform others (particularly those in the criminal justice system) about the various factors that can lead to inaccuracies in eyewitness memory reports [58, 59].

References

[1] Binet, A. (1900). *La Suggestibilite*, Schleicher, Paris.

[2] Münsterburg, H. (1908). *On the Witness Stand*, Doubleday & Page, New York.

[3] Bartlett, F.C. (1932). *Remembering: A Study in Experimental and Social Psychology*, Cambridge University Press, Cambridge.

[4] Loftus, E.F. (1979). *Eyewitness Testimony*, Harvard University Press, Cambridge.

[5] Loftus, E.F., Miller, D.G. & Burns, H.J. (1978). Semantic integration of verbal information into visual memory, *Journal of Experimental Psychology: Human Learning and Memory* **4**, 19–31.

[6] Loftus, E.F. & Palmer, J.C. (1974). Reconstruction of automobile destruction: an example of the interaction between language and memory, *Journal of Verbal Learning and Verbal Behavior* **13**, 585–589.

[7] Ross, D.F., Read, J.D. & Toglia, M.P. (1994). *Adult Eyewitness Testimony: Current Trends and Developments*, Cambridge University Press, New York.

[8] Davis, D. & Loftus, E.F. (2007). Internal and external sources of misinformation in adult witness memory, in *Handbook of Eyewitness Psychology (Vol. 1): Memory for Events*, M.P. Toglia, J.D. Read, D.F. Ross & R.C.L. Lindsay, eds, Lawrence Erlbaum Associates, Mahwah, pp. 195–237.

[9] Berkerian, D.A. & Bowers, J.M. (1984). Eyewitness testimony: were we misled? *Journal of Experimental Psychology: Learning, Memory, and Cognition* **9**, 139–145.

[10] McCloskey, M. & Zaragoza, M.S. (1985). Misleading postevent information and memory for events: arguments and evidence against memory impairment hypotheses, *Journal of Experimental Psychology: General* **114**, 1–16.

[11] Johnson, M.K., Hashtroudi, S. & Lindsay, S.D. (1993). Source monitoring, *Psychological Bulletin* **114**, 3–28.

[12] Roediger III, H.L., Balota, D.A. & Watson, J.M. (2001). Spreading activation and the arousal of false memories, in *The Nature of Remembering: Essays in Honor of Robert G. Crowder*, J.S. Nairne, I. Neath & A.M. Suprenant, eds, American Psychological Association, Washington, DC, pp. 95–115.

[13] Brainerd, C.J. & Reyna, V.F. (2005). *The Science of False Memory*, Oxford University Press, New York.

[14] Read, J.D. & Connolly, D.A. (2007). The effects of delay on long-term memory for witnessed events, in *Handbook of Eyewitness Psychology (Vol. 1): Memory for Events*, M.P. Toglia, J.D. Read, D.F. Ross & R.C.L. Lindsay, eds, Lawrence Erlbaum Associates, Mahwah, pp. 117–155.

[15] Memon, A. & Wright, D.B. (1999). Eyewitness testimony and the Oklahoma bombing, *The Psychologist* **12**, 292–295.

[16] Wright, D.B., Self, G. & Justice, C. (2000). Memory conformity: exploring misinformation effects when presented by another person, *British Journal of Psychology* **91**, 189–202.

[17] Gabbert, F., Memon, A. & Allan, K. (2003). Memory conformity: can eyewitnesses influence each other's memories for an event? *Applied Cognitive Psychology* **17**, 533–543.

[18] Gabbert, F., Memon, A., Allan, K. & Wright, D.B. (2004). Say it to my face: examining the effects of socially encountered misinformation, *Legal and Criminological Psychology* **9**, 215–227.

[19] Meade, M.L. & Roediger III, H.L. (2002). Explorations in the social contagion of memory, *Memory & Cognition* **13**, 140–144.

[20] Mudd, K. & McGovern, J.M. (2004). Conformity to misinformation and time delay negatively affect eyewitness confidence and accuracy, *North American Journal of Psychology* **6**, 227–238.

[21] Wright, D.B., Mathews, S.A. & Skagerberg, E.M. (2005). Social recognition memory: the effect of other people's responses for previously seen and unseen items, *Journal of Experimental Psychology: Applied* **3**, 200–209.

[22] Odinot, G. & Wolters, G. (2006). Repeated recall, retention interval, and the accuracy-confidence relation in eyewitness memory, *Applied Cognitive Psychology* **20**, 973–985.

[23] Zaragoza, M.S. & Mitchell, K.J. (1996). Repeated exposure to suggestion and the creation of false memories, *Psychological Science* **7**, 294–300.
[24] Ebbinghaus, H. (1888/1913). *Memory: A Contribution to Experimental Psychology*, Teachers College, Columbia University, New York.
[25] Belli, R.F., Windschitl, P.D., McCarthy, T.T. & Winfrey, S.E. (1992). Detecting memory impairment with a modified test procedure: manipulating retention interval with centrally presented event items, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **18**, 356–367.
[26] Loftus, E.F. (1975). Leading questions and eyewitness report, *Cognitive Psychology* **7**, 560–572.
[27] Dodson, C.S. & Krueger, L.E. (2006). I misremember it well: why older adults are unreliable eyewitnesses, *Psychonomic Bulletin & Review* **13**, 770–775.
[28] Karpel, M.E., Hoyer, W.J. & Toglia, M.P. (2001). Accuracy and qualities of real and suggested memories: Nonspecific age differences, *Journal of Gerontology: Series B: Psychological Sciences and Social Sciences* **56B**, 103–110.
[29] Loftus, E.F., Levidow, B. & Duensing, S. (1992). Who remembers best? Individual differences in memory for events that occurred in a science museum, *Applied Cognitive Psychology* **6**, 93–107.
[30] Mitchell, K.J., Johnson, M.K. & Mather, M. (2003). Source monitoring and suggestibility to misinformation: adult age-related differences, *Applied Cognitive Psychology* **17**, 107–119.
[31] Coxon, P. & Valentine, T. (1997). The effects of the age of eyewitnesses on the accuracy and suggestibility of their testimony, *Applied Cognitive Psychology* **11**, 415–430.
[32] Neuschatz, J.S., Preston, E.L., Burkett, A.D., Toglia, M.P., Lampinen, J.M., Neuschatz, J.S., Fairless, A.H., Lawson, D.S., Powers, R.A. & Goodsell, C.A. (2005). The effects of post-identification feedback and age on retrospective eyewitness memory, *Applied Cognitive Psychology* **19**, 435–453.
[33] Gudjonsson, G.H. (1987). A parallel form of the Gudjonsson suggestibility scale, *British Journal of Clinical Psychology* **26**, 215–221.
[34] Gudjonsson, G.H. (1997). *The Gudjonsson Suggestibility Scales Manual*, Psychology Press, Hove.
[35] Bain, S.A., Baxter, J.S. & Ballantyne, K. (2007). Self-monitoring style and levels of interrogative suggestibility, *Personality and Individual Differences* **42**, 623–630.
[36] Forrest, K.D., Wadkins, T.A. & Larson, B.A. (2006). Suspect personality, police interrogations, and false confessions: maybe it is not just the situation, *Personality and Individual Differences* **40**, 621–628.
[37] Baxter, J.S., Boon, J.C.W. & Marley, C. (2006). Interrogative pressure and responses to minimally leading questions, *Personality and Individual Differences* **40**, 87–98.
[38] Polczyk, R., Wesolowska, B. & Gabarczyk, A. (2004). Age differences in interrogative suggestibility: a comparison between young and older adults, *Applied Cognitive Psychology* **18**, 1097–1107.
[39] Pollard, R., Trowbridge, B. & Slade, P.D. (2004). Interrogative suggestibility in a US context: some preliminary data on normal subjects, *Personality and Individual Differences* **37**, 1101–1108.
[40] Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall, *Journal of Experimental Psychology* **58**, 17–22.
[41] Roediger III, H.L. & McDermott, K.B. (1995). Creating false memories: remembering words not presented in lists, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **21**, 803–814.
[42] Gallo, D.A. (2006). *Associative Illusions of Memory: False Memory Research in DRM and Related Tasks*, Psychology Press, New York.
[43] Toglia, M.P., Neuschatz, J.S. & Goodwin, K.A. (1999). Recall accuracy and illusory memories: when more is less, *Memory* **7**, 233–256.
[44] Norman, K.A. & Schacter, D.L. (1997). False recognition in younger and older adults: exploring the characteristics of illusory memories, *Memory & Cognition* **25**, 838–848.
[45] Gallo, D.A., McDermott, K.B., Percer, J.M. & Roediger III, H.L. (2001). Modality effects in false recall and false recognition, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **27**, 339–353.
[46] Seamon, J.G., Luo, C.R., Kopecky, J.J., Price, C.A., Rothschild, L., Fung, N.S. & Schwartz, M.A. (2002). Are false memories more difficult to forget than accurate memories? The effect of retention interval on recall and recognition, *Memory & Cognition* **30**, 1054–1064.
[47] Neuschatz, J.S., Lampinen, J.M., Toglia, M.P., Payne, D.G. & Cisneros, E.P. (2007). False memory research: history, theory, and applied implications, in *Handbook of Eyewitness Psychology (Vol. 1): Memory for Events*, M.P. Toglia, J.D. Read, D.F. Ross, & R.C.L. Lindsay, eds, Lawrence Erlbaum Associates, Mahwah, pp. 239–260.
[48] Roediger III, H.L., Watson, J.M., McDermott, K.B. & Gallo, D.A. (2001). Factors that determine false recall: a multiple regression analysis, *Psychonomic Bulletin & Review* **8**, 385–407.
[49] Echterhoff, G., Groll, S. & Hirst, W. (2007). Tainted truth: overcorrection for misinformation influence on eyewitness memory, *Social Cognition* **25**, 367–409.
[50] Eakin, D.K., Schreiber, T.A. & Sergent-Marshall, S. (2003). Misinformation effects in eyewitness memory: the presence and absence of memory impairment as a function of warning and misinformation accessibility, *Memory & Cognition* **29**, 813–825.
[51] Greene, E., Flynn, M.S. & Loftus, E.F. (1982). Inducing resistance to misleading information, *Journal of Verbal Learning & Verbal Behavior* **21**, 207–219.

[52] McCabe, D.P. & Smith, A.D. (2002). The effect of warnings on false memories in young and older adults, *Memory & Cognition* **30**, 1065–1077.
[53] Memon, A. & Higham, P.A. (1999). A review of the cognitive interview, *Psychology, Crime, & Law* **5**, 177–196.
[54] Fisher, R.P., McCauley, M.R. & Geiselman, R.E. (1994). Improving eyewitness testimony with the cognitive interview, in *Adult Eyewitness Testimony: Current Trends and Developments*, D.F. Ross, J.D. Read & M.P. Toglia, eds, Cambridge University Press, New York, pp. 245–269.
[55] Wells, G.L., Small, M., Penrod, S.D., Malpass, R.S., Fulero, S.M. & Brimacombe, C.A.E. (1998). Eyewitness identification procedures: recommendations for lineups and photo spreads, *Law and Human Behavior* **22**, 603–607.
[56] Scheck, B., Neufeld, P. & Dwyer, J. (2000). *Actual Innocence*, Random House, New York.
[57] Wells, G.L., Malpass, R.S., Lindsay, R.C.L., Fisher, R.P., Turtle, J.W. & Fulero, S. (2000). From the lab to the police station: a successful application of eyewitness research, *American Psychologist* **55**, 581–598.
[58] Toglia, M.P., Read, J.D., Ross, D.F. & Lindsay, R.C.L. (2007). *Handbook of Eyewitness Psychology (Vol. 1): Memory for Events*, Lawrence Erlbaum Associates, Mahwah.
[59] Lindsay, R.C.L., Ross, D.F., Read, J.D. & Toglia, M.P. (2007). *Handbook of Eyewitness Psychology (Vol. 2): Memory for People*, Lawrence Erlbaum Associates, Mahwah.

MICHAEL P. TOGLIA, KERRI A. GOODWIN AND JEFFREY S. NEUSCHATZ

Eyewitness Lineups: Identification from

The police lineup is a common tool for trying to establish the identity of the perpetrator of a crime. A lineup is likely to be used when there was at least one eyewitness (*see* **Eyewitness Testimony**) who observed the culprit committing the crime (or observed someone just prior to or just after the crime was committed who was likely to have been the culprit). When police investigators have a suspect in the crime, that person or his/her photo is shown to the eyewitness to see if the eyewitness will identify that person as the culprit. Because showing only one person (or one photo) is considered a suggestive procedure, the suspect (or his photo) is embedded among other people (or photos of other people). These other people (or their photos) are not suspects and are called *fillers*. This array of people (or photos) is generally called a *lineup*. The eyewitness is then asked if s/he recognizes anyone in the lineup as having been the culprit. Photographic lineups are much more common than are live lineups and the number of lineup members typically ranges from six to eight. A positive identification of someone from a lineup can be a very powerful form of evidence, commonly resulting in arrest, indictment, and ultimate conviction.

The Development of a Science of Lineups

Behavioral scientists have conducted hundreds of experiments, starting in the mid- to late 1970s, in which unsuspecting people have viewed simulated crimes followed by lineups. On the basis of these experiments, scientists have published several hundred articles warning that the risk of eyewitnesses identifying innocent people is much higher than most people seem to believe. The rate of mistaken identification is not a constant number but instead varies as a function of dozens of factors. These include factors that are not under the control of the criminal justice system (such as witness viewing conditions, stress, whether the witness and culprit are of the same race or not), which are called *estimator variables*. However, the rate of mistaken identification also includes factors that are directly under the control of the criminal justice system (such as prelineup instructions to the eyewitness, the choice of fillers used in the lineup, and cues from the lineup administrator that can influence the eyewitness), which are called *system variables* [1]. Psychological scientists have taken a special interest in system variables because of the potential to use system variables to reduce the chances of mistaken identification.

Although the eyewitness identification literature drew the attention and respect of many psychological scientists, it was not until the mid- to late 1990s that the legal system began to take significant notice. This notice by the justice system was due largely to the advent of forensic DNA testing in which claims of innocence by a subset of convicted people could be tested. In 1996, a US Justice Department report on the first 28 exonerations on the basis of DNA

testing revealed that 24 were cases of mistaken eyewitness identification [2]. By 2007, the number of DNA-based exonerations of individuals who were convicted by juries stood at more than 200 and over 75% of those were cases of mistaken eyewitness identification (see www.innocenceproject.org for an up-to-date internet site that tracks these cases). These DNA exonerations have resulted in greater communication between the justice system and the psychological science on the problem of mistaken identification.

Causes of Misidentification in Lineups

Researchers have discovered a large number of factors that contribute to eyewitness misidentification. Some of these factors are relatively obvious in the sense that they represent conditions of witnessing that are not conducive to forming a good memory of the facial characteristics of the perpetrator, such as poor lighting, distant viewing, short exposure durations to the perpetrator's face, and the use of disguises by the perpetrator. Other factors might be less obvious, such as the stress and fear that sometimes accompany witnessing a crime. Surveys suggest that many people believe that stress and fear could make memory better by making the eyewitness more alert or that the emotion will somehow help the eyewitness form a clear and lasting memory [3].

But, research does not support this contention, indicating instead that stress impairs the formation of memories needed for accurate identification decisions from lineups [4]. Similarly, the use of weapons visible to the eyewitness draws attentional focus to the weapon and detracts from processing of the face of the perpetrator [5]. Another factor that impairs eyewitness identification performance is whether the eyewitness and the culprit are of the same or different race or ethnic background. Research has consistently shown that it is more difficult to recognize a stranger who was viewed on only one prior occasion if that person was of a different racial or ethnic appearance than the witness [6].

Relative Judgments

Having a weak memory *per se*, however, does not fully explain why eyewitnesses make mistaken identifications from lineups. In particular, having a weak memory does not explain why a witness would make a positive identification rather than simply saying "I don't know". A dominant conceptualization among eyewitness scientists is that eyewitnesses are motivated to make an identification (in the interests of justice) and sometimes rely on a "relative-judgment process" rather than absolute recognition to make the identification decision [7]. A relative judgment is one in which the eyewitness compares each lineup member to the other lineup members and decides which one more closely resembles his/her memory relative to the other lineup members.

The relative-judgment strategy can be effective if the actual perpetrator is in the lineup, but it will lead to a mistaken identification if the actual perpetrator is not in the lineup. Experiments have shown that removing the perpetrator from a lineup results in witnesses shifting their identification decision to another lineup member, even though they were warned that the actual perpetrator might not be in the lineup [8]. This phenomenon is especially pronounced when witness's memories for the perpetrator are weak.

Prelineup Instructions

Along these same lines, the failure to use prelineup instructions warning the eyewitness that the perpetrator might not be in the lineup (or suggesting that the perpetrator is in the lineup) leads to high rates of mistaken identification when the actual perpetrator is not in the lineup [9]. This prelineup warning has little effect on the eyewitness's ability to identify the perpetrator if the actual perpetrator is in the lineup. No empirically determined estimates exist of how often perpetrator-absent lineups are shown to witnesses in actual cases. But, the absence of the actual perpetrator in the lineup is not necessarily an unusual situation. It simply means that the person-of-interest (or suspect) is not the perpetrator and that the investigation has focused on the wrong person.

It is not likely that the rate of perpetrator-absent lineups in actual cases can be represented by a single base rate or percentage. Instead, the rate of perpetrator-absent lineups is likely to vary from one case to another and one jurisdiction to another depending on how much or little evidence the detectives feel that they need in order to conduct a lineup [10]. Jurisdictions that are quick to assemble a lineup on the basis of mere hunches would be expected to run a higher rate of perpetrator-absent lineups than would jurisdictions that require good evidence against the suspect before conducting a lineup.

Lineup Composition

Another factor that contributes to mistaken identification is the use of lineup fillers who fail to fit the eyewitness's description of the perpetrator, which leads the eyewitnesses to prefer the suspect (who does fit that description) even when the suspect is not the perpetrator [11]. Consider a case, for instance, in which the eyewitness described the perpetrator as a tall man in his 20s with short dark hair and no facial hair. Suppose that the suspect fits that description but the fillers were short, or had light colored hair, or had long hair, or were in their 40s, or had facial hair. This would make the suspect stand out in the lineup as being the person that the investigators obviously believe committed the crime.

A technique called the *mock witness procedure* has been developed by eyewitness scientists to assess such biases. Mock witnesses are people who have never previously seen the suspect and are simply given the eyewitness's verbal description of the perpetrator. The mock witnesses then view the lineup (or a photo of it) and select the person they believe is the suspect in the case. A fair six-person lineup should result in only 1/6 of the mock witnesses picking the suspect and the remaining 5/6 of the witnesses selecting fillers. The inverse of these ratios (e.g., inverse of $1/6 = 6$) is known as *functional lineup size* [12]. Hence, if 1/3 of the mock witnesses pick the suspect, the functional size would be 3, even if the nominal size (number of persons in the lineup) was 6, 8, or 10. The higher the functional size, the more it is presumed to be protective of mistaken identifications of an innocent suspect.

Influence of Lineup Administrator

Another type of factor that can cause mistaken identifications from lineups is the behavior of the person who administers the lineup. The common practice is for the case detective to administer the lineup. That means that the lineup administrator knows very well which lineup member is suspected of the crime and which lineup members are merely fillers. Most initial lineup procedures use photographs and there is no videotaping of the procedure. When photo-lineup procedures are used, there is no right to have defense counsel or other neutral parties present to observe the procedure. Eyewitness scientists have argued that this procedure creates considerable opportunity for the lineup administrator to influence the eyewitness inadvertently and have argued instead that the lineup procedure should be conducted using an administrator who does not know which person is the suspect and which are fillers [13].

This is known as the *double-blind lineup procedure* [8, 14]. Experimental simulations have shown that lineup administrators do influence eyewitness identifications in a manner consistent with what the administrator is led to believe about which person is the suspect [15]. The precise ways in which this influence occurs are not yet well established, but they appear to be very similar to the experimenter-expectancy effect in which the results of an experiment are influenced by the expectations of the experimenter who tests the research participants [16].

False Confidence

A lineup results in two primary outcomes: the identification itself and the confidence that the eyewitness expresses in the identification. The confidence expressed by the eyewitness is important because tentative or low confidence identifications are usually not considered strong evidence, often do not result in charges, and tend to be unpersuasive to judges and juries. A confident eyewitness, in contrast, is very persuasive [17]. Hence, eyewitness scientists have devoted a great deal of research to the question of how strong or weak the correlation is between eyewitness identification confidence and eyewitness identification accuracy.

As might be expected, the strength of the relation between confidence and accuracy varies considerably across studies as a function of numerous other variables. A meta-analysis of 30 different studies indicated that the average correlation might be as high as $r = 0.41$ [18]. That means that accuracy is accounting for only about 16% of the variance across eyewitnesses in their confidence.

The concept of false confidence refers to a highly confident eyewitness who has nevertheless made a mistaken identification. False confidence can occur for a variety of reasons, including the obvious situation in which an innocent person was identified who happens to have very high coincidental resemblance to the culprit. More interesting, however, is the phenomenon of "confidence malleability" in which a mistaken eyewitness who was initially low in confidence later becomes highly confident.

Evidence for confidence malleability has been shown dramatically with the postidentification feedback effect. After making mistaken identifications, eyewitnesses who are given feedback suggesting that they identified the right person (e.g., "Good, you identified the actual suspect") undergo a distortion in their memory about their initial uncertainty and come to believe that they were confident all along [19]. The postidentification feedback effect has been demonstrated across a variety of experiments [20]. The result of such feedback is the creation of a confident (yet mistaken) eyewitness.

References

[1] Wells, G.L. (1978). Applied eyewitness testimony research: system variables and estimator variables, *Journal of Personality and Social Psychology* **36**, 1546–557.

[2] Connors, E., Lundregan, T., Miller, N. & McEwan, T. (1996). *Convicted By Juries, Exonerated By Science: Case Studies in the Use of Dna Evidence to Establish Innocence After Trial*, National Institute of Justice, Alexandria, VA.

[3] Kassin, S.M., Tubb, V.A., Hosch, H.M. & Memon, A. (2001). On the "general acceptance" of eyewitness testimony research, *American Psychologist* **56**, 405–416.

[4] Morgan, C.A., Hazlett, G., Doran, A., Garrett, S., Hoyt, G., Thomas, P., Baranoski, M. & Southwick, S.M. (2004). Accuracy of eyewitness memory for persons encountered during exposure to highly intense stress, *International Journal of Psychiatry and the Law* **27**, 265–279.

[5] Steblay, N.M. (1992). A meta-analytic review of the weapon focus effect, *Law and Human Behavior* **16**, 413–424.

[6] Meissner, C.A. & Brigham, J.C. (2001). Twenty years of investigating the own-race bias in memory for faces: a meta-analytic review, *Psychology, Public Policy, and Law* **7**, 3–35.

[7] Wells, G.L. (1984). The psychology of lineup identifications, *Journal of Applied Social Psychology* **14**, 89–103.

[8] Wells, G.L. (1993). What do we know about eyewitness identification? *American Psychologist* **48**, 553–571.

[9] Steblay, N.M. (1997). Social influence in eyewitness recall: a meta-analytic review of lineup instruction effects, *Law and Human Behavior* **21**, 283–298.

[10] Wells, G.L. (2006). Eyewitness identification: Systemic reforms, *Wisconsin Law Review* **2006**, 615–643.

[11] Wells, G.L., Rydell, S.M. & Seelau, E.P. (1993). On the selection of distractors for eyewitness lineups, *Journal of Applied Psychology* **78**, 835–844.

[12] Wells, G.L., Leippe, M.R. & Ostrom, T.M. (1979). Guidelines for empirically assessing the fairness of a lineup, *Law and Human Behavior* **3**, 285–293.

[13] Wells, G.L. (1988). *Eyewitness Identification: A System Handbook*. Carswell Legal Publications, Toronto.

[14] Wells, G.L., Small, M., Penrod, S.J., Malpass, R.S., Fulero, S.M. & Brimacombe, C.A.E. (1998). Eyewitness identification procedures: recommendations for lineups and photospreads, *Law and Human Behavior* **22**, 603–647.

[15] Haw, R.M. & Fisher, R.P. (2004). Effects of administrator-witness contact on eyewitness identification accuracy, *Journal of Applied Psychology* **89**, 1106–1112.

[16] Rosenthal, R. (2002). Covert communication in classrooms, clinics, courtrooms, and cubicles, *American Psychologist* **57**, 838–849.

[17] Wells, G.L., Lindsay, R.C.L. & Ferguson, T.J. (1979). Accuracy, confidence, and juror perceptions in eyewitness identification, *Journal of Applied Psychology* **64**, 440–448.

[18] Sporer, S., Penrod, S., Read, D. & Cutler, B.L. (1995). Choosing, confidence, and accuracy: a meta-analysis of the confidence-accuracy relation in eyewitness identification studies, *Psychological Bulletin* **118**, 315–327.

[19] Wells, G.L. & Bradfield, A.L. (1998). "Good, you identified the suspect:" Feedback to eyewitnesses distorts their reports of the witnessing experience, *Journal of Applied Psychology* **83**, 360–376.

[20] Douglass, A.B. & Steblay, N. (2006). Memory distortion in eyewitnesses: a meta-analysis of the postidentification feedback effect, *Applied Cognitive Psychology* **20**, 859–869.

GARY L. WELLS AND LISA E. HASEL

Eyewitness Memory ***see*** Elderly in Court

Eyewitness Testimony

One of the most famous cases involving mistaken eyewitness memory arose out of events that occurred in 1984, when a man broke into two separate women's homes, raped both women, and stole their money and other belongings. One of the victims, 22-year-old Jennifer Thompson identified the same man in a photo lineup and in a subsequent in-person

lineup; his name was Ronald Cotton. Cotton was arrested and charged with two counts of rape and burglary. Even though Cotton had an airtight alibi, the confident eyewitness testimony of Jennifer Thompson was sufficient for the jury to convict him. Cotton was sentenced to life in prison. Nine years after Cotton's conviction, he learned about DNA testing while watching the O.J. Simpson trial. Cotton asked to be tested, and the results proved that Cotton was innocent. In June of 1995, Ronald Cotton was exonerated of the crimes he had once been convicted of and was given $5000 in compensation [1].

As Cotton's case reveals, eyewitness identification is neither always reliable nor accurate (*see* **Eyewitness Lineups: Identification from**). Because eyewitness testimony is a powerful component of criminal court cases, it is necessary to understand the memory processes and variables that influence identifications, and to learn what can be done to prevent these miscarriages of justice.

How Memory Works

After witnessing a crime and being asked to identify the culprit, eyewitnesses sometimes express great confidence in the accuracy of their identifications, saying, "I was so frightened. I'll never forget that face." However, research and real-life cases reveal that eyewitnesses may misidentify the perpetrator. Contrary to popular belief, memory does not function like a videotape recorder. Events are not recorded in the mind so that they can be played back later; rather, complex reconstruction takes place to form the "memory" for the particular event (*see* **Memory: Reconstructive**).

Acquisition Stage

Memory formation consists of three stages: acquisition, retention, and retrieval [2]. The acquisition stage occurs when the original event is perceived. The amount of information encoded into a person's memory system is actually only a small portion of the person's surrounding environment.

Retention Stage

After the information from the event is stored in memory, many factors can affect the memory. This period of time between the event and eventual recall of it is called the *retention stage*. We have known since the nineteenth century that memory fades during the retention interval [3].

The passage of time is not the only factor that can cause memory to falter during the retention stage. Eyewitnesses may overhear a conversation between other witnesses or read a newspaper article about the crime; these are both examples of "postevent information" that can alter the original memory. Sometimes postevent information can even cause a witness to introduce nonexistent objects into the memory. After a person witnesses a crime, he/she may be exposed to information that alters the original memory with the new information or that causes the eyewitness to compromise between the original memory and the new information. For example, in a study conducted by Loftus and her colleagues [4], participants viewed a series of 30 slides depicting an accident involving a red Datsun and a pedestrian. Half the participants watched the car approach an intersection in which there was a yield sign, and the other half watched the car approach the intersection with a stop sign.

After viewing the slideshow of the auto-pedestrian accident, the participants were asked a series of questions, one of which was critical. Half the participants were asked, "Did another car pass the red Datsun while it was stopped at the *stop sign*?" The other half were asked the same question with the words "*yield sign*" replacing the words "stop sign". For half the subjects, the question contained misleading information; that is, they saw a stop sign but were asked about a yield sign or they saw a yield sign and were asked about a stop sign. Twenty minutes later, participants were administered a recognition test.

When given consistent information from the critical question (i.e., saw a stop sign and asked about a stop sign), 75% of the participants accurately identified the correct slide (either depicting a red Datsun at an intersection with a stop sign or a red Datsun at an intersection with a yield sign). If the question they had previously answered included misleading information (i.e., saw a yield sign but asked about a stop sign), only 41% of the participants accurately identified the correct slide. Thus, the misleading question reduced accuracy for identifying the correct slide to below chance level. As demonstrated by this study, postevent information in the form of misleading questions can reduce the accuracy of memory.

Retrieval Stage

In the last stage, the retrieval stage, witnesses try to recall information about an event. They may simply provide a narrative about what happened or answer specific questions. Retrieval can occur many times, for example, during an investigation by law enforcement or during trial testimony. Some of the factors that influence accuracy and completeness of an eyewitness's account during the retrieval stage include the environment in which the retrieval process takes place, the type and wording of questions to obtain information, and who is asking the questions.

Estimator and System Variables

At any point during the three stages of memory, a number of other variables can influence the way in which the information is perceived, retained, and recalled. Researchers often refer to these factors as "estimator variables" and "system variables". Estimator variables include the characteristics of a witness (e.g., an eyewitness's age and race) and the characteristics of an event (e.g., amount of time the eyewitness attentively viewed the culprit, whether or not a weapon was present, and crime scene lighting conditions) [5].

System variables include factors that are under the control of law enforcement or others in the legal system. These include factors such as whether or not the culprit was in the lineup (culprit-present *vs.* culprit-absent lineups), whether the eyewitness was given prelineup instructions indicating that the culprit may or may not be in the lineup, and the composition of the lineup (e.g., the suspect looks significantly different from the other people included in the lineup). The lineup presentation method (simultaneous *vs.* sequential) and any influence the person administering the lineup may have on an eyewitness (e.g., feedback about the person an eyewitness chose as the culprit) are also considered system variables [5].

Important witness characteristics to consider are the age and race of the eyewitness. In particular, research shows that older adults and young children are less accurate than adults with respect to eyewitness memory. Research also suggests that witnesses are better able to identify strangers of the same race than strangers of a different race. This phenomenon is known as *cross-racial identification*. Other important estimator variables to consider are the characteristics of the event. For example, the amount of time that an eyewitness has to view the culprit may affect the reliability of memory. Critically, research indicates that eyewitnesses often overestimate the duration of crimes. Another factor that may affect eyewitness memory is the presence of a weapon; this phenomenon is known as *weapon focus*. Specifically, research has shown that the presence of a weapon during a crime can capture attention, and while the weapon may be well remembered, other details can be impaired in memory, including the identity of the culprit [6].

Sometimes investigators include a suspect in a lineup who is not actually the real perpetrator. Showing these culprit-absent lineups, although not intentional, can present problems. Eyewitnesses may expect the culprit to be present and result in the eyewitnesses choosing the person who looks most similar to the culprit they remember, and subsequently a misidentification occurs (*see* **Eyewitness Lineups: Identification from**).

The composition of a lineup is also crucial. A biased lineup may contain a suspect that is the only person of a certain race when the nonsuspects in the lineup (i.e., fillers) are of a different race or a suspect described as six-feet tall is placed in a lineup with fillers that are all much shorter. As recommended by the National Institute of Justice's (NIJ) *Eyewitness Evidence Guide for Law Enforcement* [7], the composition of a lineup would include fillers who fit the verbal description of the culprit as given by the eyewitness. While some lineup biases may seem subtle and unimportant, they may still greatly affect eyewitness identification. In addition to composition, the lineup presentation method is important. Some research suggests that sequential lineups, in which individuals or photos are presented sequentially rather than simultaneously, may reduce erroneous identifications.

Another system variable that is important to address is the influence that the lineup administrators may have on eyewitnesses. Research shows that confirming feedback (e.g., "Good, you identified the suspect") from a lineup administrator may inflate the eyewitness's confidence in his or her accuracy of the identification. Thus, researchers suggest the use of double-blind lineups in which neither the administrator of the lineup nor the eyewitness knows which lineup/photospread member is the suspect. By using

the double-blind technique, the investigator can neither provide any inadvertent cues nor can he or she provide confidence-inflating feedback.

Estimator and system variables may affect an eyewitness's testimony in several ways. It is important to remember that any combination of estimator and/or system variables could occur in a single case. Depending on which variables are present in a case, an eyewitness's memory may be greatly impaired. Because law enforcement has greater control over system variables than estimator variables, it is important that law enforcement makes every effort to prevent mistaken identifications. In particular, the NIJ recommends the implementation of specific lineup procedures to reduce erroneous identifications. In addition, it is important for the jurors in a trial to be aware of the factors that may influence an eyewitness account.

Juror Knowledge

Over recent decades, researchers have examined whether jurors correctly weigh the variables that affect the reliability of eyewitness identifications. Several studies have revealed that potential jurors show limited understanding of eyewitness memory. For example, Deffenbacher and Loftus [8] administered the Knowledge of Eyewitness Behavior Questionnaire (KEBQ) to college students and lay people. The KEBQ contained 14 multiple-choice questions that assessed jurors' understanding of several eyewitness behaviors. Overall, the results indicated that subjects performed moderately above chance on some questions, but subjects performed at a less than chance level on other questions.

In particular, subjects held misconceptions about the effects of violence and interviewing methods on an eyewitness' memory. Also, subjects misunderstood the slight, positive correlation between confidence and accuracy, the reliability over time of memory for faces, the lower accuracy of cross-race identifications and the fact that trained observers are no more accurate as eyewitnesses than the average citizen. The authors concluded that cautionary instructions and expert testimony are necessary to ensure that jurors correctly evaluate eyewitness evidence.

Likewise, Kassin and Barndollar [9] surveyed college students and lay people about 21 eyewitness phenomena. The results showed that subjects performed at or below chance for 11 of the items. While most subjects correctly understood that severe stress, the eyewitness' attitudes and expectations, and the wording of questions can influence an eyewitness' testimony, most subjects incorrectly believed that violence and exposure time do not affect an eyewitness memory. Subjects showed limited understanding of the forgetting curve, the accuracy–confidence relationship, the inferiority of cross-race identifications, and the lack of increased accuracy of trained observers. These findings indicate that jurors are insensitive to some of the factors that affect an eyewitness' testimony.

Other research suggests that jurors are less aware of estimator variables than system variables. For instance, Benton *et al.* [10] surveyed actual jurors and found that although jurors held misconceptions about both estimator variables and system variables, jurors were generally more insensitive to questions concerning estimator variables. Specifically, jurors performed below chance for 15 of the 22 questions on estimator variables including, but not limited to the following: cross-race identifications, the accuracy–confidence relationship, weapon focus, forgetting curve, exposure time, unconscious transference, and violence. Jurors also performed below chance for four of the eight questions on system variables including the following: lineup instructions, lineup presentation format, description matching, and lineup fairness. The authors concluded that in order to prevent jurors from erroneously assessing eyewitness evidence, expert testimony is crucial.

Additional research suggests that some jurors are too trusting of eyewitness evidence. In particular, Schmechel *et al.* [11] surveyed a large, random sample of potential jurors from Washington, DC and assessed whether the potential jurors correctly understood what makes an eyewitness more or less reliable. The findings indicated that 65% of subjects believed eyewitness identification evidence to be somewhat or very reliable. Furthermore, 46% of subjects incorrectly understood the reconstructive nature of human memory. A substantial proportion of subjects were insensitive to the factors that affect eyewitness testimony including: the accuracy–confidence relationship, trained observers, cross-race identifications, show-ups, weapon focus, violence, prelineup instructions, and lineup presentation format. These results suggest that a sizable proportion of jurors will

begin a trial with dangerous misconceptions about the reliability of eyewitness testimony.

Conclusion

So far, we know from hundreds of research studies that eyewitnesses sometimes inaccurately remember the details of a crime and, even worse, misidentify the perpetrator of a crime. We also know that eyewitness memory is susceptible to distortion during each of the three stages of memory. Whether an eyewitness views a crime in which a weapon is present or whether an eyewitness views an unfair lineup, both estimator and system variables affect the accuracy of eyewitness identifications. Although real-life cases lend support to the empirical finding that eyewitness testimony is fallible, not all participants in the legal system appreciate this fact. In particular, the juror knowledge studies reveal that jurors are generally insensitive to the factors affecting eyewitness testimony. Taken together, these studies suggest that jurors may draw erroneous conclusions from eyewitness evidence and as a result, wrongfully convict innocent individuals. To prevent wrongful convictions, it is important that law enforcement follow the recommended guidelines set forth by the NIJ, and that juries are given cautionary instructions and/or expert testimony on the reliability of eyewitness testimony.

References

[1] Scheck, B., Neufeld, P. & Dwyer, J. (2003). *Actual Innocence: When Justice Goes Wrong and How to Make It Right*, New American Library, New York.

[2] Loftus, E.F. (1979). *Eyewitness Testimony*, Harvard University Press, Cambridge.

[3] Ebbinghaus, H.E. (1964). *Memory: A Contribution to Experimental Psychology*, Dover, New York.

[4] Loftus, E.F., Miller, D.G. & Burns, H.J. (1978). Semantic integration of verbal information into a visual memory, *Journal of Experimental Psychology: Human Learning and Memory* **4**, 19–31.

[5] Wells, G.L. & Olson, E.A. (2003). Eyewitness testimony, *Annual Review of Psychology* **54**, 277–295.

[6] Steblay, N.M. (1992). A meta-analytic review of the weapon focus effect, *Law and Human Behavior* **16**, 413–424.

[7] Technical Working Group for Eyewitness Evidence (1999). *Eyewitness Evidence: A Guide for Law Enforcement [Booklet]*, United States Department of Justice, Office of Justice Programs, Washington, DC.

[8] Deffenbacher, K. & Loftus, E.F. (1982). Do jurors share a common understanding concerning eyewitness behavior? *Law and Human Behavior* **6**, 15–30.

[9] Kassin, S.M. & Barndollar, K.A. (1992). On the psychology of eyewitness testimony: a comparison of experts and prospective jurors, *Journal of Applied Social Psychology* **22**, 1241–1249.

[10] Benton, T.R., Ross, D.F., Bradshaw, E., Thomas, W.N. & Bradshaw, G.S. (2006). Eyewitness memory is still not common sense: comparing jurors, judges and law enforcement to eyewitness experts, *Applied Cognitive Psychology* **20**, 115–129.

[11] Schmechel, R.S., O'Toole, T.P., Easterly, C. & Loftus, E.F. (2006). Beyond the ken: testing jurors' understanding of eyewitness reliability evidence, *Jurimetrics Journal* **46**, 177–214.

KALLY J. NELSON, NICCI BOWMAN-FOWLER, SHARI R. BERKOWITZ AND ELIZABETH F. LOFTUS

Wiley Encyclopedia of
FORENSIC SCIENCE

Volume 3
F–L

Editors-in-Chief

Allan Jamieson
The Forensic Institute, Glasgow, UK

Andre Moenssens
Forensics and Law Center, Columbia City, IN, USA

WILEY



Facial Comparison

Introduction

In general, people are thought to be reasonably good at recognizing faces. Although research indicates that recognition of "familiar" faces can be extremely efficient, even with disguises and poor-quality images, this does not apply to recognition of "unfamiliar" faces [1, 2]:

> At an image level, the variability in the appearance of a given face under variations of lighting, pose and expression far exceeds the variation between two different faces under the same conditions [2].

This describes, in a single sentence, the central issue at hand with regard to the area of facial comparison in terms of both biometric (automated) and "human-based" facial comparison. This does not take away from the fact that facial comparison is regarded as a highly valuable tool for identification, and in many cases is indeed valuable. The number of crimes in which facial comparison plays a role will certainly increase, especially considering the growing number of security cameras.

Within the context of person identification, different processes can be defined. "Recall" is here defined as the process of retrieving descriptive information of a person from long-term memory in the absence of the person, his/her photograph, or other image. Recall requires observation, retention, and reproduction of a person's features. Recall is essential for the production of composite images, as produced by a police artist for investigational purposes. However, these images can only be used as investigative tools, and can never be used as proof of identity. "Recognition" is, in this article, defined as the process of identifying or matching a person, his/her photograph, or image with a mental image that one has previously stored in long-term memory. Recognition requires observation and retention of a person's features and the process of comparison of the retained information with an external image whether it be the life person, a photograph, or composite image. Recognition is important for investigation as well as witness statements. "Comparison" is defined as the process of identifying or matching a person or his/her image with another photograph or image without the use of retained information. For comparison, retention and reproduction do not play a role.

Summarizing, recall requires observation, retention, and reproduction; recognition requires observation and retention; and comparison requires observation. Obviously, for forensic purposes with respect to surveillance images we talk about recognition or comparison in case human operators are involved and about comparison in case (automated) biometrics is involved.

Human Observer

The main questions with respect to forensic image comparison are as follows: "Is the person in image A the same person as in image B?" or "is the person in image A a different person than in image B?" These images can vary greatly in terms of

source (wet passport photo, digital passport photo with and without compression, analog video, and digital video with low or high compression) and quality. In principle, we try to reduce the question to a comparison of two images. In doing so, we test the hypothesis that it is the same person in both images, as well as the hypothesis that two different people are concerned.

Are we actually capable of doing this? Can an unequivocal judgment be made with regard to identity? Can an estimate of a likelihood ratio be made with regard to the hypotheses "the images are of the same person" and "the images are not of the same person"?

Issues Involved in Facial Comparison

The obstacles encountered when carrying out facial comparison can be divided into issues related to the image material itself and issues related to the comparison process.

Image Material. *Image Quality*. The quality of the image material provided for facial comparison varies widely. Quality is poorly defined in this context, but is mainly used in the sense of "identifiability" or "degree of visibility of details". Image quality is influenced by a number of factors that should be described in the report if possible. Image quality is determined by sharpness, contrast, clarity, distance, distortion (teller machines!), resolution, and compression. Figure 1 gives some examples of surveillance and document images of varying quality. These examples vary from "useless" (a) due to image quality and pose, to "good" with minor compression artefacts (d). It would be advisable to summarize image quality in a single number/several numbers to indicate the expected evidential value of an image. Unfortunately, even an approximation of this is not possible at this time, and it is a tricky problem due to the many factors involved in "image quality".

Posture and Position in the Image. Besides image quality, the usefulness of an image is also influenced

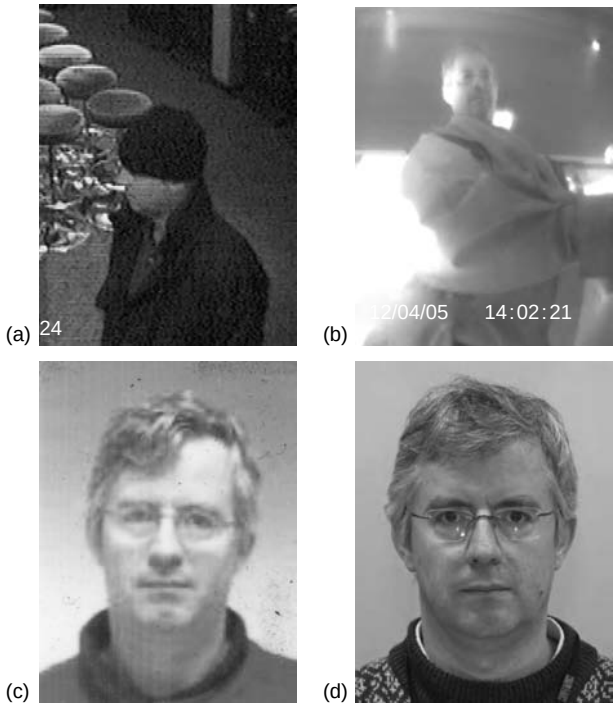


Figure 1 Typical examples of surveillance and document quality images. Quality ranges from "useless" due to image quality and pose (a), "nearly useless" due to camera position and distortion (b), "somewhat useful" due to poor sharpness (c), and "good" with minor information loss due to compression (d)

by the posture and position of the person depicted. Security cameras are often placed high up, and people do not look directly into the camera. The fact that the representation of shapes and distances within the face changes with changes in distance and pose relative to the camera is still poorly acknowledged, resulting in too much reliance on distance measurements in the pictures of the face. Problems can also arise with "controlled" photos: most passport photos are currently taken of the person looking straight at the camera, but sometimes the head is tilted or turned slightly, as a result of which the ears may or may not be fully visible. Mug shots are usually taken of the suspect facing straight ahead and from the left, as a result of which the identifying features on the right side of the face, and the right ear, are usually not visible. Of course, hats, glasses, and disguises also affect the possibilities of comparison: in the case of a camera placed high up, if the suspect is wearing a baseball cap with a bill, this is often sufficient to cover most of the face.

Reconstruction. To overcome problems with posture and position, reconstruction of recording conditions can be used. To reposition someone in the image, the camera system must still be intact and the suspects must cooperate. Foils are used for validation in this procedure, but the influence of the choice of the foils (according to the description of the perpetrator), the methods of processing of the data (one-to-one comparisons with perpetrator images or grouping of "similar" people), and conclusions that can be drawn from such investigation have not been validated yet.

A method that is still in the experimental phase is the use of 3D models for reconstruction [3]. The advantage of this method is that it does not require cooperation by a suspect at the scene of the crime, and that repositioning can take place in a virtual environment. However, making the 3D models is a laborious and time-consuming process. The reliability of the reconstructed models has not been tested yet. The models provide a less natural image, and the value of measurements based purely on 3D models is limited [4].

The use of models of foils for determining the identifying value when making comparisons by means of positioning of 3D models also has not been studied yet.

A benefit of model reconstruction is that all conditions, including lighting, can be modeled in accordance with the images to be investigated as much as possible.

Comparison Process. *Features*. Facial comparison should take place according to a list of facial features. These lists can be drawn up on the basis of historical lists used for facial comparison, feature lists obtained from anthropologists, and discussions with people from other agencies involved in person identification (an example list is provided in Table 1). A problem that remains is the difference in the interpretation of what some features "are". This interpretation depends on education and anatomical knowledge, and experience has shown that it is not always assessed consistently even by the same observers. It is also questionable whether the interpretation of the feature designations is in line with the knowledge of and interpretation by the client (police and judge).

Similarities and Differences. If there is a consensus on the interpretation of the features, the features can be compared. The comparison involves looking at the similarities and differences of a feature on two images. This involves two problems: (i) can people assess differences and similarities (human as a measuring instrument) and (ii) how are the observed similarities and differences interpreted?

Human as a Measuring Instrument. Research has shown that, as a measuring instrument, humans are not good at assessing facial features in which sizes and proportions play a role [5]. The evidential value of sizes and proportions in visual material is also limited because of the image distortion caused by unknown camera equipment, unknown zoom settings, and unknown pixel aspect ratios of digital cameras. Not every human observer assesses the (differences in) shapes of facial features in the same way either [6, 7]. In addition, people are not very consistent in the assessment of similarities between pairs of photos [8]. Although the comparison of familiar or famous people on videos and photographs is efficient, the comparison of unknown people on videos with photographs or in real life is not efficient [1]. However, another study showed that the use of moving pictures can improve the recognition of static, frontal photographs if the images are shown on the same medium (monitor), whereas multiple static images have little added value [9]. In addition

Table 1 Example list of features to be observed for facial comparison

Face	Shape Proportions Hairline/hair growth
Forehead	Shape "Bumps" Horizontal creases Eyebrows
Eyes	Distance Position Color Eye shape (shape of eye slit) Upper eyelid crease Bags Wrinkles at corners of eyes
Nose	Length of nose Width of nose Prominence Symmetry Creases at bridge of nose Profile of nose Shape of tip of nose Shape of nostrils Septum
Middle section of face	Cheekbones/cheek line Cheek-eye groove Cheek-nose groove
Ear	Size Position Shape of helix (outer edge of ear) Darwin's tubercle (bump on edge) Shape of antihelix (fold in ear) Earlobe
Mouth area	Shape of philtrum (groove between nose and upper lip) Moustache or moustache "shadow" Beard or beard "shadow"
Mouth	Mouth size Mouth shape Upper lip Lower lip
Chin	Shape Groove between mouth and chin Dimple Double chin
Lower jaw/throat	Shape Adam's apple
Characteristic features	Skin marks Scars Creases and wrinkles

The list is based on anatomical knowledge, stability over time, and visibility in pictures

to interpretation differences depending on the image material, we also have to deal with the "other race effect", the effect that recognition of people from another, rather than one's own, race may be less efficient [10–13]. These effects have mainly been shown to occur in the case of recognition (memory task: "have you seen this person before?"). One of the theories is that minor variations in facial shape are coded better when looking at images of people of one's own race [12, 13]. Other race effects also appear to occur in facial comparison. Research on humans as a measuring instrument in facial comparison is limited; most of the research concerns "recognition" or "rapid" comparison. We do not know of any studies concerning the quality of man as a measuring instrument according to a structured method.

Interpretation of Observations. Debate on the interpretation/assessment of features as "similar" or "different" often occur during discussions of results like "I see a difference, but it can be explained" and "this is similar to". Even the term "not observable" can cause a difference in interpretation. An example of this in the case of wrinkles: not observable because someone does not have wrinkles, or not observable owing to the photo quality? Does "no wrinkles visible" in both images constitute a similarity if the image quality is good? The question of quality can also be involved in this discussion; a suggestion is to rate the "quality" and "usefulness" of the individual images first, before proceeding with the comparison. Then a selection is made from the features to prevent "looking for" features in a poor-quality image. The drawback to this is that features that may be useful to a limited extent may be eliminated from a comparison too quickly. Another issue is the practical feasibility of this approach; the images should be assessed by two different people in two different rounds to prevent bias as much as possible. Along with the requirement that the comparison should be carried out by more than one person, this may lead to logistic problems.

Individualizing Value of Features. An important question in facial comparison is "How unique is a feature?" Although we have put much time and energy into answering this question, we have not yet been able to answer it, or find literature data that answer

this question. One of the problems is that (anthropological) investigations focus largely on measuring parts of the skeleton and the like, such as measuring the distance between points in the face that can be determined by palpation but are difficult to identify on images. Another problem is that most known measurements have been taken from a very limited, usually Caucasian, population. No answer can be given to the question "How unique is this nose?", and so we depend on the subjective interpretation of the observer. This subjective interpretation is known to depend on the appearance of people in the observer's own environment [12, 14]. Currently we – and others [15, 16] – only designate birthmarks and scars as "strongly identifying".

Final Assessment of Comparison. The combination of "similarities and differences" and "individualizing value" is weighed to achieve a final assessment on the images. Cognition psychology research has shown that people are good at recognizing "familiar" people, but not as good at recognizing "unfamiliar" people. "Familiar" people are people with whom social interaction has taken place, and providing moving images does not improve the recognition task [1, 17]. The extent to which a systematic approach of the comparison task influences the reliability of the final assessment is not known. However, we do know that there is little correlation between the instinctive "certainty" about a comparison judgment and the performance of human observers [1, 17].

Biometrics

Automated facial comparison and recognition still function far from optimal. The best systems have a verification equal error rate (EER) of about 1.5%, and a false reject rate (FRR) of about 10% at a false accept rate (FAR) of 0.1% if "document quality" images are used [18]. The facial recognition systems are still sensitive to aging of the subjects: the FRR increase to about 20% at an FAR of 1% if the picture is 3 years old.

Looking at the performance of facial biometrics systems for investigational purposes of criminals or terrorists on watch list (identification), the best systems find about 60% of the suspects on a list of 1000 people at a setting, resulting in 1% false alarms [19]. Increasing the number of people on the watch list results in about 2–3% points decrease in the number of correct hits with each doubling of the database size. Note that these results are based on experiments with standardized high to medium quality pictures from collaborating subjects. The performance of biometric facial recognition systems decreases quickly with changes in pose and position, lighting, and facial expression. Surveillance images taken under uncontrolled circumstances, using low-quality cameras, mostly placed in conspicuous places (high at the ceiling, low at automatic teller machines (ATMs), "hidden" at the side of entrances) and recording people who do not look into the camera, with changing expression and sometimes partly covered faces, are hardly or not at all recognized by biometric systems. In the past, biometrics also has been proposed as a solution for detecting document fraud by look-alikes: people using genuine documents from other people with similar appearance. Unfortunately, this problem cannot be solved by the current biometric systems, because these systems are based on relative common features and use a relative low resolution (300 dots per inch (DPI) passport images). Although these images have advantages because of the limited memory capacity of embedded memory chips in passports, this limits the usefulness of the facial image for identification purposes. Because of the low resolution, the most identifying features such as birthmarks, wrinkles, and scars are unrecognizable.

Summary

Facial comparison is a wide-ranging forensic technique: it involves disciplines such as image analysis, 3D technology, anthropology, photogrammetry, statistics, and cognitive psychology (human as a measuring instrument). Many steps in the comparison process are still mediocre, poorly or not supported by scientific data. Quality metrics for visual material and the comparison process are still lacking; important steps in the process are subjective. The issues "quantification of image quality", "human as a measuring instrument", and "individualizing value of features" are difficult problems because of the large number of factors involved, the difficulty of measuring man's functioning as a measuring instrument, and the scope of the investigation needed to test individualizing features. In most cases, psychological research concerns recognition of people; as far as we know,

research with regard to facial comparison is limited. It is often assumed that comparison works according to the same basic principles as recognition, but without memory load. Research, development, and elaboration of facial comparison methods still will require much more time, because little hard data is available. Performance of current biometric systems is still too low for forensic purposes, although biometric systems may be useful for automated investigation of standardized image databases. The quality issues of surveillance material limit the options of applicability of biometric systems for uncontrolled situations.

References

[1] Bruce, V., Henderson, Z. & Burton, A.M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images, *Journal of Experimental Psychology. Applied* **7**, 207–218.

[2] Hancock, P.J.B., Bruce, V. & Burton, A.M. (2000). Recognition of unfamiliar faces, *Trends in Cognitive Sciences* **4**, 330–337.

[3] Blanz, V. & Vetter, T. (2003). Face recognition based on fitting a 3D morphable model, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**, 1–12.

[4] Goos, M., Alberink, I.B. & Ruifrok, A.C.C. (2006). 2D/3D image (facial) comparison using camera matching, *Forensic Science International* **163**, 10–17.

[5] Vanezis, P., Lu, D., Cockburn, J., Gonzalez, A., McCombe, G., Trujillo, O. & Vanezis, M. (1996). Morphological classification of facial features in adult caucasian males based on an assessment of photographs of 50 subjects, *Journal of Forensic Sciences* **41**, 786–791.

[6] Kemp, R., Towell, N. & Pike, G. (1997). When seeing should not be believing: photographs, credit cards and fraud, *Applied Cognitive Psychology* **11**, 211–222.

[7] Henderson, Z., Bruce, V. & Burton, A.M. (2001). Matching faces of robbers captured on video, *Applied Cognitive Psychology* **15**, 445–464.

[8] Scheuchenpflug, R. (1999). Predicting face similarity judgements with a computational model of face space, *Acta Psychologica* **100**, 229–242.

[9] Pike, G.E., Kemp, R.I., Towell, N.A. & Phillips, K.C. (1997). Recognizing moving faces: the relative contribution of motion and perspective view information, *Visual Cognition* **4**, 409–437.

[10] Bothwell, R.K., Bringham, J.C. & Malpass, R.S. (1989). Cross-racial identification, *Personality and social psychology bulletin* **15**, 19–25.

[11] Lindsay, D.S., Jack, P.C. & Christian, M.A. (1991). Other-race face perception, *Journal of Applied Developmental Psychology* **76**, 587–589.

[12] Valentine, T. & Endo, M. (1992). Towards an exemplar model of face processing: the effects of race and distinctiveness, *The Quarterly Journal of Experimental Psychology* **44A**, 671–703.

[13] O'Tool, A., Deffenbacher, A., Valentin, D. & Abdi, H. (1994). Structural aspects of face recognition and the other-race effect, *Memory and Cognition* **22**, 208–224.

[14] Bruce, V. & Young, A. (1986). Understanding face recognition, *British Journal of Psychology* **77**, 305–327.

[15] Bromby, M. (2003). At face value? The use of facial mapping and CCTV image analysis for identification, *New Law Journal. Expert Witness Supplement* **153**, 302–304.

[16] Bowie, L., Plews, S. & Bromby, M. (2004). *When Evidence is a Question of Image*, Law Society Gazette.

[17] Bruce, V., Henderson, Z., Greenwood, K., Hancock, P.J.B., Burton, A.M. & Miller, P. (1999). Verification of face identities from images captured on video, *Journal of Experimental Psychology. Applied* **5**, 339–360.

[18] Phillips, P.J., Scruggs, W.T., O'Toole, A.J., Flynn, P.J., Bowyer, K.W., Schott, C.L. & Sharpe, M. (2007). FRVT2006 and ICE 2006 Large-Scale Results, http://www.frvt.org/FRVT2006/docs/FRVT2006andICE2006LargeScaleReport.pdf 2008.

[19] Phillips, J.P., Grother, P., Michaels, R.J., Blackburn, D.M., Tabassi, E. & Bone, M. (2003) Face recognition vendor test 2002, http://www.frvt.org/DLs/FRVT_2002_Evaluation_Report.pdf 2008.

ARNOUT C. C. RUIFROK

Facial Recognition *see* Facial Comparison, Visual Recognition Systems in Identification

Facial Reconstruction

Once one has built the so-called biological profile, the last thing one can do to help in identifying those remains, is to give the biological profile along with a face. Let us give a practical example: a human skeleton is found; anthropologists determine that it belongs to a 40- to 60-year-old Caucasoid male, about 1.70 m tall, with two dental crowns and

no other particular signs. Comparison with missing persons gives no positive results. All one can do then is to try and divulge the biological profile across television and newspapers, inviting people to say whether they recognize the description or if it suits someone they know. However, rather than divulging a "sterile" description, it may be more effective to show a face along with it. Facial reconstruction from the cranium therefore serves the purpose of triggering people's minds and reaching a suspicion of identity [1–17]. If someone says the biological profile, along with the face, reminds them of Mr Brown, then further analyses are brought forward to verify whether the skeleton indeed belongs to Mr Brown (*see* **Identification of Human Remains**). It is not however an identification method.

Facial reconstruction (preferably referred to by some authors as facial approximation) is practically based on Galen's theory concerning facial structure ("as walls to houses, so are bones to all living creatures, for other features take form from them and change with them"); the face is supposed to be influenced by the underlying bone tissue.

The first facial reconstruction, according to Pliny, was made by Lysistrathos in the fifth century BC using wax casts; this tradition was maintained by the most important Roman families, which kept masks reproducing faces of their forefathers (*majors*) in their houses. The first scientific reconstruction was carried out by His on J. S. Bach's cranium. For several years, this technique was considered as a reliable method in order to obtain the real appearance of dead people, until 1913 when Martin and von Eggeling independently carried out two different reconstructions on the same cranium.

The technique was then dropped for years, until 1924 when Gerasimov developed a method based on the anatomical reconstruction of muscles of the face (the so-called Russian method). This is in stark contrast to the American technique, which is based on the measurement of soft tissue depths overlying the bone. The latter represents the American school and its importance was reported by the Federal bureau of investigation (FBI) Law Enforcement Bulletin in 1946.

Because the relationship between bone and soft tissue is a basic starting point for facial reconstruction, tissue depths are the fundamentals upon which a facial reconstruction is built. For this reason, in literature soft tissue depths have been published for both sexes and several ethnic groups (and more recently for children) [6–10]. These had been already measured in the late nineteenth century in several ways: mainly from pins inserted into the faces of cadavers at specific cranial landmarks. Cadaver data, however, is slightly less reliable mainly because postmortem dehydration may cause some loss of depth. Data nowadays is obtained from ultrasound and computed tomography (CT) technology. Therefore, more modern databases of populations that are radiologically obtained are probably more reliable. Thanks to this type of technology, there is nowadays increasing data concerning the soft tissue facial depths of living individuals on the basis of sex, age, and ancestry. One should, however, always keep in mind that the tissue depth which is used by the operator is an average and depends on an enormous number of variables (weight, sex, age, etc. combined)–some of which are difficult to assess on the skeleton. Thus, the actual depth remains in reality as an unknown variable. The same is applicable to the approximation of other facial areas such as the nose and mouth.

Facial reconstruction can be performed either manually or via sophisticated softwares [18–26] (among other things, bidimensional drawings also can be created which are of a more simplistic nature). Both methods have their advantages: the first is cheaper but consumes more time; the second is quicker but more expensive. The error is comparable, because both techniques follow similar anatomical guidelines. The following is a description of one of the most common manual methods.

Facial reconstruction is carried out on the cranium, or, preferably, on a cast of the cranium, which is usually made in plaster. The first step involves designing the profile from a lateral cranial X-ray (Figure 1), which then guides the 3D final reconstruction. Thirteen cranial points are located on the cranial profile in the X-ray, which guide the reconstruction of soft tissue profile. This allows for a reconstruction specifically made to that particular profile. Tissue depths are given by anatomical tables present in the literature, and the profile is adapted according to variables such as nasal height, height of teeth, chin, etc.

Once the profile is defined, then the 3D reconstruction is performed. A cast of the cranium is usually created; therefore, the reconstruction can be made without working on the real cranium. This is done in order not to damage the original since it may have peculiar lesions which need to be re-examined and

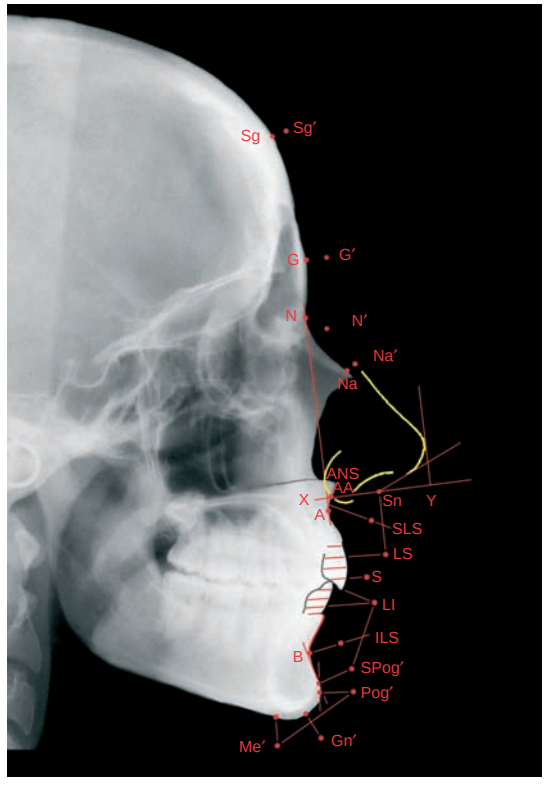


Figure 1 An example of the complex calculations involved in the reconstruction of the profile from the radiograph

are important for the cause of death. A separate cast is made for the mandible.

The 3D technique consists of applying so-called tissue depths onto 31 cranial points. These cranial depths usually consist of toothpicks or rubber depths, which demarcate the exact thickness of the soft tissue at that specific point, according to general anatomical tables taken from living populations and adapted for race, sex, and age. Then plasticine muscle models are put on the bone surface, complying with the anatomical insertions.

At first temporal and masseter muscles are applied onto the cranial cast, followed by buccinator and mouth orbicular muscles. Then mouth elevator and upper lip elevator muscles, zygomatic muscular groups and upper lip, and mouth corner depressor muscles are added. However, vascular and nervous structures, along with adipose tissues, are not identifiable, and this produces a substantial difference between the final reconstruction and the real face appearance. Parotid glands models are placed on the reconstruction as well. The advantage of an accurate muscular reconstruction is to create a more specific guide and to reduce risks of subjective interpretation. A plasticine stratum of skin is finally placed on the muscle groups according to anthropological data coming from the biological profile (Figure 2).

Eyes can be made of plasticine, plaster, resin, and ophthalmic prostheses (Figure 3). The authors use eye resin models that are set into the orbits according to anatomical methods, usually by centering the pupil with the junction of two straight lines between the middle point of lower and upper orbital edges and between outer and inner eye corner.

The nose is modeled mainly according to the results of the 2D profile reconstruction mentioned above. Width of the nose is usually calculated by adding one third of the bone nasal aperture on each side (Figure 4).

The *rima oris* (mouth aperture) is evaluated according to the distance between the two canine teeth (Figure 5).



Figure 2 The various phases of manual facial reconstruction: (a) application of tissue depths on specific landmarks; (b) application of the deep muscles; (c) completion of the muscular structure; (d, e) application of "skin"; and (f) details can then be added such as beard or hair although these may be misleading

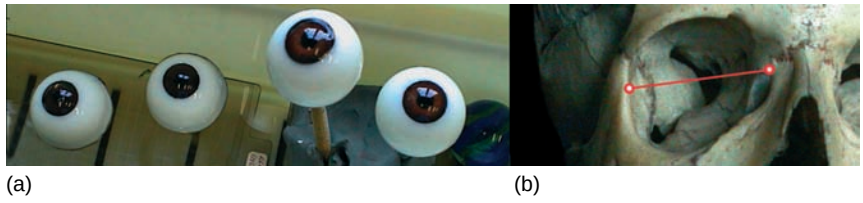


Figure 3 (a) Example of resin eye models and (b) anatomical area of eye alignment within the orbit



Figure 4 The red horizontal line represents the extreme limits of the soft tissue portion of the nose

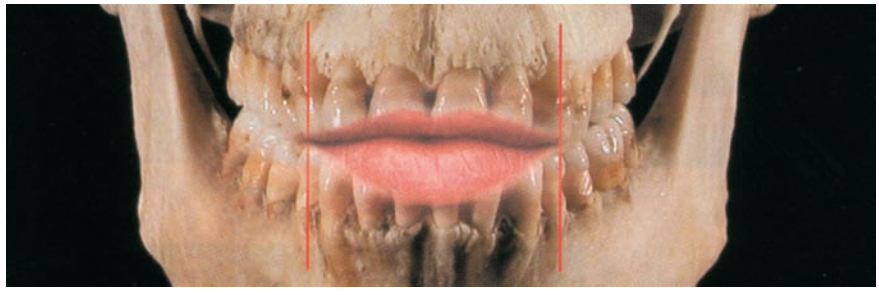


Figure 5 Superimposition of lips onto the dentition, which shows the generic correspondence between the corners of the mouth and the first premolar/canine

Ears, eye color, wrinkles, and eyebrows are obviously not identifiable from the cranium. Reconstruction is the object of a debate, which involves different opinions. There are some people who believe that the facial reconstruction must be as unevocative as possible – therefore gray, with no color. Others prefer color and interpretation. The reconstruction is a last attempt to provide some information

about the identity of subject. The aim of technique requires "facial interpretation", and, according to some authors, must even be similar to a "caricature", in order to stress the personal characteristics and physical peculiarities of that person. Ears, hair styles, eye color, wrinkles, and eyebrows can alternately be added in conformity with race, sex, age, etc., if one wants to give greater "character" to the reconstruction, though the risk may mislead observers.

Sometimes a smiling expression may be useful in order to show a particular odontological profile, such as missing teeth, a wide diasthema, or the presence of dental prosthesis.

It is clear that facial reconstruction is a very gross anatomical exercise – there are many variables which have to be taken into account and also exceptions. The technique shows its clear limits, where the relationship between bone and soft tissue is not easily predictable (for example, mouth and nose). Recent years have seen a considerable proliferation of witty and useful articles that try to make it more accurate particularly in relation to these difficult areas [11–17].

As far as testing the accuracy of approximation is concerned, several types of tests, among which blind tests, have been pursued [27–35]. One of the first tests was a comprehensive study of Helmer [35] who worked with two independent teams that performed double blind facial reconstructions on double castings of 12 skulls. Results proved the large variation in the quality of the reconstruction but the resemblance between the reconstruction and the actual face varied from "slight resemblance" (42%) to "close resemblance" (38%). Subsequent studies in this sense were performed, although it is almost intuitive that the evaluation of "resemblance" can never be free of extremely subjective parameters. Stephan and coworkers [27, 29–34] have given considerable thought (and articles) to this issue, with interesting and useful results. Accuracy of facial approximation has been assessed by resemblance ratings (i.e., the comparison of a facial approximation directly to the target individual) and recognition tests (i.e., the comparison of a facial approximation to a photo array of faces including foils and a target individual). It has been proven that recognition tests are more valid assessment parameters than resemblance ratings since a study in particular has shown that similar resemblance ratings can be given to different facial approximations of the same person, whereas recognition tests produce vastly different results (one is "identified", the other is not). Furthermore, there seems to be no statistically significant difference between resemblance ratings of facial reconstructions to target individuals and to individuals incorrectly identified as the target individuals. The same group of researchers tested 16 facial approximations on 37 assessors of varying age. Only one approximation ended in true positive identification rates above chance at statistically significant levels. It is therefore rare for facial approximations to be sufficiently accurate to allow identification of a target individual above chance. This also suggests that facial reconstruction is not useful in excluding individuals to whom skeletal remains may not belong. At times it may even be misleading. It should therefore be used when all else has failed.

Other authors are slightly more positive and suggest that better results can be expected when the facial reconstruction is performed by a trained scientist and with the more elaborate methods. Thus excellent, good to middle resemblances were reached in 9 out of 25 cases, but success increased with a more elaborate method and with an expert operator in one particular study [28]. Statistical comparisons, however, between measurements on the facial reconstruction and on the actual face showed that some anthropological distances were constantly underestimated or overestimated. Other authors [25] have recently found that in a test with computer-modeled facial reconstructions recognition tests yielded percentage hit rates of 50% above chance. These remain, however, sporadic studies that inevitably have a very strong subjective element (that of recognition) that makes the assessment of accuracy almost impossible. One can actually verify the error as far as anthropometrical measurements are concerned (e.g., correspondence of interpupillary distance and bigonial width between the reconstruction and the actual face), but it is a completely different matter to pinpoint the main factors in recognition.

As mentioned previously, several authors have focused in the past 20–25 years on the potential of computer-aided facial reconstructions, which imply the use of very sophisticated techniques that are used currently by very few teams [18–26]. Nonetheless computerized facial reconstruction in the future may become more user-friendly, more easily accepted, and its future assessment may promote further improvement.



Figure 6 Example of a real case reconstruction. To the left is the facial reconstruction made from a skeleton found in a river. Transmission to the public led to a suspicion of identity and subsequent identification of the victim via the comparison of odontological antemortem and postmortem traits. To the right is the frontal and lateral picture of the living victim

In conclusion, the field is quite active at the moment and much research is still required, although stronger connections with the cognitive sciences should be sought for the achievement of more accurate reconstructions since still much has to be learnt on what actually triggers recognition in the human brain: shape, reflectance, emotion, etc. [36].

It should also be mentioned that in cases in which remains are mummified or putrefied, some authors prefer not to clean the cranium from soft tissues but to rehydrate the tissues and work on reconstructing the original face.

Whichever method one chooses, the important issue is to divulge the final result (Figure 6), which must be released throughout the media, in order to reach the largest population possible and make it easier to achieve a suspicion of identity. It is not an identification method; therefore, once there is an idea of who the person may be, positive identification must be performed by other methods.

References

[1] Prag, J. & Neave, R. (1997). *Making Faces. Using Forensic and Archaeological Evidence*, British Museum Press, London.

[2] George, R.M. (1987). The lateral craniographic method of facial reconstruction, *Journal of Forensic Sciences* **32**(5), 1305–1330.

[3] Gerasimov, M.M. (1993). *Principles of Reconstruction of the Face on the Skull*, Nauka, Moscow. Quoted by Fedosyutkin, B.A. & Nainys, J.V. The relationship of skull morphology to facial features, in *Forensic Analysis of the Skull*, M.Y. Iscan & R.P. Helmer, eds, Wiley-Liss, New York, pp. 199–213.

[4] Claes, P., Vandermeulen, D., De Greef, S., Willems, G. & Suetens, P. (2006). Craniofacial reconstruction using a combined statistical model of face shape and soft tissue depths: methodology and validation, *Forensic Science International* **159**(Suppl 1), S147–S158.

[5] Greef, S.V. & Willems, G. (2005). Three dimensional cranio-facial reconstruction in Forensic identification: latest progress and new evidences in the 21st century, *Journal of Forensic Sciences* **50**(1), 12–17.

[6] Rhine, J.S. & Moore, C.E. (1984). Reproduction tables of facial tissue thicknesses of American Caucasoids, in *Forensic Anthropology*, *Maxwell Technical Series I*, Maxwell Museum of Anthropology, Albuquerque.

[7] Rhine, J.S. & Campbell, H.R. (1980). Thickness of facial tissues in American blacks, *Journal of Forensic Sciences* **25**, 847–858.

[8] De Greef, S., Claes, P., Vandermeulen, D., Mollemans, W., Suetens, P. & Willems, G. (2006). Large-scale in-vivo Caucasian facial soft tissue thickness database for craniofacial reconstruction, *Forensic Science International* **159**(Suppl 1), S126–S146.

[9] Utsuno, H., Kageyama, T., Deguchi, T., Umemura, Y., Yoshino, M., Nakamura, H., Miyazawa, H. & Inoue, K. (2007). Facial soft tissue thickness in skeletal type I Japanese children, *Forensic Science International* **172**(2–3), 137–143.

[10] El-Mehallawi, I.H. & Soliman, E.M. (2007). Ultrasonic assessment of facial soft tissue thicknesses in adult Egyptians, *Forensic Science International* **117**, 99–107.

[11] Stephan, C.N. & Henneberg, M. (2003). Predicting mouth width from inter-canine width – a 75% rule, *Journal of Forensic Sciences* **48**(4), 725–727.

[12] Stephan, C.N. & Davidson, P.L. (2008). The placement of the human eyeball and canthi in craniofacial identification, *Journal of Forensic Sciences* **53**(3), 612–619.

[13] Stephan, C.N. (2002). Position of superciliare in relation to the lateral iris: testing a suggested facial approximation guideline, *Forensic Science International* **130**(1), 29–33.

[14] Stewart, T.D. (2003). The points of attachment of the palpebral ligaments: their use in facial reconstruction on the skull, *Journal of Forensic Sciences* **28**(4), 858–863.
[15] Vermeulen, L. (2005). Manual facial reconstruction in forensic medicine, *Revue Belge de Médecine Dentaire* **60**(3), 227–236.
[16] Wilkinson, C.M., Motwani, M. & Chiang, E. (2003). The relationship between soft tissues and the skeletal detail of the mouth, *Journal of Forensic Sciences* **48**(4), 728–732.
[17] Rynn, C. & Wilkinson, C.M. (2006). Appraisal of traditional and recently proposed relationships between the hard and soft dimension of the nose in profile, *American Journal of Physical Anthropology* **130**(3), 364–373.
[18] Turner, W., Tu, P. & Kelliher, T. (2006). Brown R computer-aided forensics: facial reconstruction, *Studies in Health Technology and Informatics* **119**, 550–555.
[19] Quatrehomme, G., Cotin, S., Subsol, G., Delingette, H., Garidel, Y., Grévin, G., Fridrich, M., Bailet, P. & Ollier, A. (1997). A fully three-dimensional method for facial reconstruction based on deformable models, *Journal of Forensic Sciences* **42**, 647–650.
[20] Nelson, L.A. & Michael, S.D. (1998). The application of volume deformation to three-dimensional facial reconstruction: a comparison with previous techniques, *Forensic Science International* **94**, 167–181.
[21] Miyasaka, S., Yoshino, M., Imaizumi, K. & Seta, S. (1995). The computer aided facial reconstruction system, *Forensic Science International* **74**, 155–165.
[22] Vanezis, P., Vanezis, M., McCombe, T. & Niblett, T. (2000). Facial reconstruction using 3-D computer graphics, *Forensic Science International* **108**, 81–95.
[23] Clement, J.G. & Marks, M.K. (2005). *Computer Graphic Facial Reconstruction*, Elsevier Acadamic, Amsterdam.
[24] Vandermeulen, D., Claes, P., Loeckx, D., De Greef, S., Willems, G. & Suetens, P. (2006). Computerised craniofacial reconstruction using CT-derived implicit surface representations, *Forensic Science International* **159**(Suppl 1), 164–174.
[25] Wilkinson, C., Rynn, C., Peters, H., Taister, M., Kau, C.H. & Richmond, S. (2006). A blind accuracy assessment of computer-modeled forensic facial reconstruction using computed tomography data from live subjects, *Forensic Science, Medicine and Pathology* **2**(3), 179–187.
[26] Wilkinson, C. (2005). Computerized forensic facial reconstruction: a review of current systems, *Forensic Science, Medicine and Pathology* **1**(3), 173–177.
[27] Stephan, C.N. & Cicolini, J. (2008). Measuring the accuracy of facial approximations: a comparative study of resemblance rating and face array methods, *Journal of Forensic Sciences* **53**(1), 58–64.
[28] Quatrehomme, G., Balaguer, T., Staccini, P. & Alunni-Perret, V. (2007). assessment of the accuracy of three-dimensional manual craniofacial reconstruction: a series of 25 controlled cases, *International Journal of Legal Medicine* **121**, 469–475.
[29] Stephan, C.N. & Henneberg, M. (2006). Recognition of forensic facial approximation: case specific examples and empirical tests, *Forensic Science International* **156**(2–3), 182–191.
[30] Stephan, C.N. (2002). Do resemblance ratings measure the accuracy of facial approximations? *Journal of Forensic Sciences* **47**(2), 239–243.
[31] Stephan, C.N. & Arthur, R.S. (2006). Assessing facial approximation accuracy: how do resemblance ratings of disparate faces compare to recognition tests? *Forensic Science International* **15**(Suppl 1), S159–S163.
[32] Stephan, C.N. (2003). Anthropological facial "reconstruction" – recognizing the fallacies, unembracing th eerrors and realizing method limits, *Science and Justice* **43**(4), 193–200.
[33] Stephan, C.N. (2002). Do resemblance ratings measure the accuracy of facial approximations? *Journal of Forensic Sciences* **47**(2), 239–243.
[34] Stephan, C.N. & Henneberg, M. (2001). Building faces from dry skulls: are they recognized above chance rates? *Journal of Forensic Sciences* **46**(3), 432–440.
[35] Helmer, R.P., Rohricht, S., Petersen, D. & Mohr, F. (1993). Assessment of the reliability of facial reconstruction, in *Forensic Analysis of The Skull*, M.Y. Iscan & R.P. Helmer, eds, Wiley-Liss, New York, pp. 229–246.
[36] Russell, R. & Sinha, P. (2007). Real-world face recofnition: the importance of surface reflectance properties, *Perception* **36**(9), 1368–1374.

Related Articles

Anthropology

Anthropology: Ancestry and Stature Determination

Anthropology: Age Determination of Remains

Identification of Human Remains

Sex Determination of Remains

Species Determination of Osseous Remains

CRISTINA CATTANEO AND DAVIDE PORTA

Factitious Disorder *see*

Posttraumatic Stress Disorder

False Confessions ***see*** Confessions: Evidentiary Reliability of; Interrogative Suggestibility

False Memories ***see*** Memory: Repressed

Falsifiability Theory

Austrian-born British philosopher Karl R. Popper (1902–1994) is best known for his theory of falsifiability, which he argued depending on the "testability" of the premises underlying the relevant scientific discipline. His 1934 German-language book, *The Logic of Scientific Discovery* [1], was hailed by many other scientists as one of the most important works of the twentieth century.

In his 1963 publication, *Conjectures and Refutations* [2], Popper amplified that the demarcation between true science and pseudoscience is that scientific knowledge is based on falsifiable hypotheses that permit the inference (conclusion) to be drawn that, at least for the present, the underlying premise is probably true or probabilistically likely to be true. In contrast, pseudoscience, or in Popper's words "metaphysics", rests on unfalsifiable or untestable premises from which invalid inferences may well be drawn.

In Conjectures, Popper used his theory of falsifiability in an attempt to discredit psychoanalysis and Marxism, both of which had at some time been postulated as being based on scientific principles, by showing the premises underlying these fields were incapable of verification.

Although likely oversimplified, Popper advocated that the hallmark of science is that the premises upon which a scientific principle rests are capable of being refuted by empirical evidence derived from a physical experiment or through observation.

The term *falsifiability* is often misunderstood by liberal-arts-trained lawyers and judges. It does not mean that a premise is false, but that a hypothesis is capable of being shown to be false by experimentation, observation, or examination of data [3–6].

The ability to explain the meaning of concepts supporting expert opinions is especially crucial to forensic scientists who may be asked to discuss the principle of falsifiability on the witness stand (**Expert Opinion: United States**). Since 1993, the words "falsifiability" and "Karl Popper" have acquired special meaning in legal circles, at least in jurisdictions where courts follow the Daubert (***Daubert v. Merrell Dow Pharmaceuticals***) rules on admissibility of expert testimony.[a]

In the *Daubert* decision, the US Supreme Court delineated a number of different factors by which trial judges are asked to decide whether proffered expert testimony will qualify as "scientific knowledge" shown to be sufficiently reliable to be admitted at trial. One of these factors is its "falsifiability".[b] The court adopted the Popper model of distinguishing reliable scientific knowledge from junk science (unreliable expert opinions) by combining "testability" with "falsifiability". The court explained

> [A] key question to be answered in determining whether a theory or technique is scientific knowledge that will assist the trier of fact is whether it can be (and has been) tested. "Scientific methodology today is based on generating hypotheses and testing them to see if they can be falsified; indeed, this methodology is what distinguishes science from other fields of human endeavor [7].

For this proposition, the Court cited a number of sources, prominent among them being Popper's work [8].

Ever since 1993, courts attempting to apply *Daubert* to specific areas of technical expertise have sought to determine whether the underlying assumptions on which an expert's discipline rests have been tested or are subject to testing. In attempting to resolve this issue, varying results are obtained in court decisions, the outcome in each case depending upon the techniques, the testifying experts, and the tribunal. In doing so, some jurists have exhibited a lack of understanding of the concept of testability. This deplorable result may, in part, be due to the inability of testifying experts to articulate clearly the meaning of falsifiability.

Since the decision in *Kumho Tire* (**Judicial Notice of Scientific Principles and Facts**), however, which held that a *Daubert*-like inquiry requiring proof of

reliability extends to all forms of expert opinion testimony, including that based on technical or specialized knowledge from which opinions are derived, it has become less important to determine whether a particular field qualifies for the label "scientific".

While the US Supreme Court chose Popper's work as the basis for its "falsifiability" and "testability" concepts in the context of the rules of court evidence (**Federal Rule of Evidence 702**), a number of other scientific researchers are less than accepting of Popper's theories. Indeed, some scientists suggest that Popper's view was outmoded even when it was proposed, and that the *Daubert* decision is based, therefore, on a somewhat discredited view of science.[c] Scientist-lawyer Adina Schwartz argued that Popper's falsifiability (testability) concept was an inadequate logical criterion that was developed by philosopher Popper "without appealing to the views of those professing to do science". Dr Schwartz postulated that the Supreme Court's choice of Popper's definition of science over those postulated by other equally prominent philosophers of science was a poor exercise of reasoning [9, 10].

End Notes

[a.] *See* **Computer Animation and Simulation Evidence** – Daubert v. Merrell Dow Pharmaceuticals. Also, **Judicial Notice of Scientific Principles and Facts** – Kumho Tire v. Carmichael.

[b.] The factors are (i) whether the type of evidence can be and has been tested by a scientific methodology, which subsumes its "falsifiability"; (ii) whether the underlying technique has been subjected to peer review and publication in the profession (*see* **Peer Review as Affecting Opinion Evidence**); (iii) how reliable the results are in terms of a potential error rate (*see* **Error Rates in Forensic Methods**); and (iv) whether a technique has been generally accepted in the relevant profession (*see* **General Acceptance Test for Novel Expert Evidence**). *See also*, **Expert Opinion: United States**.

[c.] See, e.g., Crease, R.P., who shows that Popper's principle was proven to be wrong throughout the history of science, in "finding the flaw in falsifiability", at http://physicsworld.com/cws/article.print/16478. See also, Goldstein, Rebecca, a philosopher at Harvard University, who stated that "Popper's characterization of how science is practiced – as a cycle of conjecture and refutation – bears little relation to what goes on in the labs and journals", in "Falsifiability" at http://www.edge.org/q2008/q08_9.html.

References

[1] Popper, K.R. (1959). *The Logic of Scientific Discovery*, Basic Books, New York. The work was an expanded translation of his earlier *Logik der Forschung*, 1934.

[2] Popper, K.R. (1962). *Conjectures and Refutations: The Growth of Scientific Knowledge*, Basic Books, New York, hereinafter *Conjectures*. The book was reprinted in 1968 by Harper & Row.

[3] Feigi, H. & Brodbeck, M. (1953). *Readings in The Philosophy of Science*, Appleton-Century-Crofts, New York.

[4] Franklin, J. (2001). *The Science of Conjecture–Evidence and Probability Before Pascal*, The Johns Hopkins University Press.

[5] Cartwright, N. (1999). *The Dappled World: A Study of the Boundaries of Science*, Cambridge University Press.

[6] Popper, K. (1977). "falsification-testability" and scientific theory, in *The Structure of Scientific Theories*, 2nd Edition, C.G. Hempel, T.S. Kuhne & S. Suppe, and others, eds, University of Illinois Press.

[7] Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579, 593, 113 S.Ct. 2786, 2796 (1993).

[8] Popper, K. (1989). The criterion of the scientific status of a theory is its falsifiability, or refutability, or testability, in *Conjectures and Refutations: The Growth of Scientific Knowledge*, 5th Edition, p. 37.

[9] Schwartz, A. (1997). A 'Dogma of Empiricism' revisited: Daubert v. Merrell Dow Pharmaceuticals, Ind. and the need to resurrect the philosophical insight of Frye v. United States, *Harvard Journal of Law and Technology* **10**, 149.

[10] Moenssens, A.A., Henderson, C.E. & Portwood, S.G. (2007). *Scientific Evidence in Civil and Criminal Cases*, 5th Edition, Foundation Press, p. 23.

Related Articles

Error Rates in Forensic Methods

General Acceptance Test for Novel Expert Evidence

Peer Review as Affecting Opinion Evidence

ANDRE MOENSSENS

Family Law ***see*** Parental Rights and Prerogatives

FDR ***see*** Firearm Discharge Residue: Analysis of

Federal Rule of Evidence 702

All forensic experts are expected to be familiar in some measure with Federal Rule of Evidence 702. While technically applicable only in the United States in federal, civil, and criminal trials, discussion of the rule is featured often in the forensic literature, in symposia, and in seminars that purport to prepare experts for court appearances. Like the court decisions in *Frye v. United States*, *Daubert v. Merrell Dow Pharmaceuticals*, and *Kumho Tire v. Carmichael*, Federal Rule of Evidence 702 has been mentioned by courts worldwide, even in jurisdictions where it does not represent local law.

The rule is part of the Federal Rules of Evidence (FRE), which have been in effect in the United States since 1975. It was amended in 2000, and defines the use of expert testimony in civil and criminal litigation. FRE 702 is included in Article VII of the federal rules that is titled, "Opinions and Expert Testimony." Article VII deals not only with expert testimony in Rule 702 but also with the following:

- Opinion Testimony by Lay Witnesses (FRE 701).
- Bases of Opinion Testimony by Experts (FRE 703).
- Opinion on Ultimate Issue (FRE 704).
- Disclosure of Facts or Data Underlying Expert Opinion (FRE 705).
- Court Appointed Experts (FRE 706).

As it currently exists, Rule 702 provides the following:

> If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case.

By the 2000 Amendment, the drafters sought to conform FRE 702 to the requirements of the US Supreme Court decision in *Daubert v. Merrell Dow Pharmaceuticals*, which had interpreted Rule 702 in the context of the "general acceptance" principle of *Frye v. United States*.

Related Articles

Daubert v. Merrell Dow Pharmaceuticals

Expert Opinion: United States

Frye v. United States

General Acceptance Test for Novel Expert Evidence

Ultimate Issue Evidence by Experts

ANDRE MOENSSENS

Female Aggression ***see*** Aggression: Gender Differences in

Fibers

Introduction

The process of interpretation begins when the expert receives a case: relevant information will be collected, pre-assessment will take place and transfer probabilities will be assessed. We refer the reader to

see **Case Assessment and Interpretation** for preassessment and early interpretation and focus here on interpretation after analytical results. Let us consider a case where, after thorough laboratory examination and consideration of similarities and dissimilarities, two or more fiber samples cannot be distinguished, the question remains as to how best we can interpret this result? In some cases, especially in the United States, forensic scientists express this finding in terms of a "*could have come from*" statement. What is the value of this type of statement? This question probably constitutes the main challenge faced by the forensic fiber examiner.

Unlike glass or paint evidence, where a proof of common source can sometimes be achieved through a physical fit, fibers (excluding pieces of fabrics and some exceptional cases) are always considered as class evidence. In other words, there is probably always more than one item in the world that shares some common characteristics with a specimen of evidential fibers.

However, this does not mean that fibers have little or no value. The evidential value of fibers depends upon a multitude of parameters and is often stronger than a layperson would expect. Fibers can even constitute primary evidence under some circumstances. Their main strength resides in their ability to reconstruct an event or a series of events. It is the aim of this article to present and discuss the interpretation and value of fiber evidence.

Factors Influencing the Value of Fiber Evidence

Fiber evidence is difficult to interpret. While the fundamental parameters such as relevance of the trace, transfer/persistence, and frequency in the population of interest are similar to those related to other trace evidence such as glass or paint, the problems associated with fibers are more difficult. For example, it is not possible to build and maintain a database of clothing types in the same manner than one would do with a database of glass types. In most cases, the questions raised in relation to the significance of fibers are extremely complex.

Since the mid-1980s, different aspects of the evidential value of fibers have been studied, and there is now a clearer view on the significance of fiber evidence [1, 2]. It is generally accepted that the value of fiber evidence depends on many factors that are often complex and interacting. These factors can be classified as known (or usually determinable) or unknown (or difficult to determine) [3].

Factors that are known or that are usually determinable are as follows:

- **The circumstances of the case**

The discovery of fibers may be more or less significant, depending on the circumstances. The forensic scientist needs to consider whether or not fibers can be explained by an alternative hypothesis; i.e., by coincidence or by a real transfer, but one which has an explanation that is not incriminating, or the other, where the explanation may be incriminating. In order to attempt to select between these options, the scientist needs to know the relevant and often detailed information about the circumstances of the case and details relevant to the individuals allegedly involved.

- **The time that has elapsed before collection of the evidence**

The longer the time between the transfer event and the collection, the weaker the evidence. The risk that the relevant fibers are being lost and replaced by nonrelevant fibers *via* secondary or a higher degree transfer increases with time.

- **The suitability of the fiber types for recovery and comparison**

Some fibers are highly colored or luminescent, which means they are not only easier to find and collect but also have more features to examine and to compare with each other. This increases the probability to discriminate fibers that do not have a common origin, and therefore enhance the significance.

- **The extent of the comparative information derived from the samples**

This information may be limited by the laboratory resources or by the size and type of fibers. It is obvious that a "fiber match" is much more significant when complementary and highly discriminating techniques have been used in comparison to simple screening tests.

- **The number of types of matching fibers**

Garments often contain more than one fiber type. This means that a "multiple match" increases the significance of fibers. It should be noted that, in cases involving a blended fabric, the absence of one or more fiber types in the pool of recovered fibers

does not necessarily preclude the blended fabric from being the source of the questioned fibers. Different components of a blended fabric may have different shed potentials, and the number of fibers transferred of different types is not necessarily proportional to the stated composition of the garment. This phenomenon is well known and is called *differential shedding*.

- **Whether there has been an apparent cross transfer of fibers or not**

This situation arises where, for example, fibers found on the victim's garment are not differentiable from the fibers of the suspect's garment *and* fibers found on the suspect's garment are not differentiable from the fibers of the victim's garment. The demonstration of a cross transfer constitutes the ideal situation following the Locard exchange principle. In this case, the significance of fiber evidence is dramatically increased because the chance of finding "matching fibers" in both samples by pure coincidence is remote.

- **The number of matching fibers recovered**

In general, the larger the number of fibers, the smaller is the chance of finding these fibers by chance only. The discovery of a small or an unexpected number of fibers is difficult to interpret and requires good knowledge of transfer and persistence theories.

- **The location of the recovered fibers**

Some locations are more prone to secondary (nonrelevant) fiber transfer than others. In a break-in case, for example, fibers found at the edge of a smashed window are more significant than fibers found on the ground. The link between the perpetrator and the evidence in the latter case is quite unclear, which decreases the significance. Similarly, fibers found on undergarments in a sexual case may be more significant than fibers found on outer garments, for example.

- **The methods used to conduct the examinations**

More discriminating and complimentary methods bring more significant comparative features and decrease the risk of a coincidental match.

Factors that are unknown or that are difficult to determine are as follows:

- **The extent and force of contact**

These factors influence fiber transfer and persistence with respect to both the number of fiber types and the number of fibers.

- **The degree of certainty that specific items were definitely in contact**

In some cases, there is a higher degree of certainty that the items submitted are those involved in the alleged incident. For example, there may be good eyewitness accounts or the suspect may have been apprehended quickly. In other cases, it is quite possible that the clothing is not relevant to the incident. For example, there might have been an opportunity to dispose of the clothing, or eyewitness statements might be unclear.

- **Donor fiber shed potential**

Some fabrics shed more fibers than others, and this must be considered when assessing the significance of the number of fiber types and the number of fibers. This means that it is impossible to give a simple rule that would define a cutoff number of fibers beyond which the primary transfer is certain, as opposed to secondary or higher transfers.

- **Frequency of occurrence of the matching fiber types**

The significance is weighted by the frequency of occurrence of the matching fibers. It is obvious that a match involving common fibers is less significant than that involving rare fibers. It should be noted that, owing to the high degree of variability of fibers, this frequency is generally much smaller than one might expect in the first instance.

Information Available to Assist the Interpretation of Fiber Evidence

The interpretation of fiber evidence requires a systematic study of the different factors described above in light of the context of the case and the hypotheses alleged by the different parties (generally defense and prosecution). This process requires a great deal of experience and logical reasoning. Some valuable information exists through specialized literature or unpublished research projects, and can be used as an aid by the fiber expert during the interpretative process. However, in some cases, the lack of relevant information may prompt the setup of simulation experiments in order to confirm or deny some hypotheses that remain unresolved. Although interpretation problems have engendered great debates within the forensic community over the last 25 years,

significant advances have been made in interpretation modeling (see below).

Most of the information currently available in the area of the interpretation of fiber evidence is related to the following issues:

- fiber transfer and persistence [4–19];
- frequency of occurrence of fiber types or chance of a coincidental match [20–43]; and
- statistical interpretation [44–47].

Fiber Transfer and Persistence

Knowledge of fiber transfer and persistence assists in answering the question as to whether the number of fibers and the number of fiber types found in a given case are likely under the allegation of contact. In other words, knowledge on transfer and persistence assists in answering the competitive questions "what is the probability of finding the number of fibers and fiber types found in a given case if there was a contact?" and "what is the probability of finding the number of fibers and fiber types found in a given case if there was no contact?".

This area of investigation includes the following topics:

- transfer studies;
- studies on differential shedding;
- experiments on the alteration of fiber characteristics in a manner consistent with localized conditions specific to a case; and
- persistence studies.

Since 1975, numerous transfer and persistence studies involving fibers have been undertaken. The results give a general guidance when assessing the number of fibers expected to be found under a given set of circumstances. It is beyond the scope of this article to examine these results in fine detail, but the most important findings are given below.

The number of fibers to be found mostly depends on

- the area of contact;
- the kind of physical contact (number of contacts, pressure, friction, time, etc.);
- the construction of the fabrics involved (donor and recipient);
- the fiber types (generic class, density, diameter, etc.);
- the length of the fiber fragments; and
- whether or not, and how, the recipient is mobile after the transfer event has occurred.

Small fiber fragments on the outer surface of a garment are more likely to be transferred and are generally more persistent on the recipient. Transferred fibers are more or less rapidly lost and redistributed on other locations (on the same recipient or not). Experimental studies have shown that fibers transferred at the time of contact may range from only a few to many hundreds or thousands. When the recipient is mobile, it has been shown that there is a rapid loss of fibers (typically 80% of loss within the first 4 h and 5–10% remaining after 24 h). On the other hand, fibers can persist for periods of many days and even weeks when transferred to a recipient, which remains subsequently relatively undisturbed such as a car seat or a dead body.

As already stated, in blended fabrics, the number of fibers transferred of the different types is not necessarily proportional to the stated composition of the garment. This issue, called *differential shedding*, is important when interpreting findings involving a fabric of blended composition.

The presence of only a few matching fibers may mean that they have been deposited by means of a secondary or subsequent transfer as opposed to a primary one (secondary transfer is the first indirect transfer after primary transfer, taking place *via* an intermediary object; common in contacts involving seating). Other reasons may also account for this situation, including a long time gap between contact/transfer and examination, a redistribution of fibers (e.g., due to the washing of the garment), the use of an inefficient method for fiber recovery, and a coincidental transfer. In all cases, caution is necessary when interpreting the finding of a small number of fibers, especially to items such as underclothing.

In real life, the variables that contribute to the number of fibers that may be transferred are so numerous and unknown that the most reliable assessment of the number of fibers to be found is done by the simulation of the suspected contact with the actual fabrics in the case under investigation. However, pressure of casework in most laboratories would make this type of research work impossible in all but the most critical cases.

Frequency of Fibers

"Fibers are mass produced; they are not like fingerprints or DNA, are they?" This question is typically asked by the defense in court. It is therefore necessary to consider fiber frequencies. Knowledge of these frequencies helps in the assessment of the relative rarity of the fiber features observed during the examination. In other words, the knowledge on the frequency of fibers mainly assists to answer the question "what is the probability of finding the fiber features observed in a given case if there was no contact?". A large number of studies have been achieved over the last 15 years in this area. Crucial information comes from the following sources:

- population studies;
- target fiber studies;
- data collections; and
- trade enquiries.

The population studies have analyzed examples of the normal fiber population that may be expected on a given surface. This is vital information to have at hand when trying to answer the defense question "well surely you would expect this type of fibers to be present on any such surface?" Fiber population studies involve sampling the given surface and classifying the recovered fibers into fiber type and color categories.

Conclusions that can be drawn from the studies that have been carried out so far have shown the following:

- Synthetic fibers form only a low percentage (13–20%) of any population yet studied.
- Most of the color/morphology/generic type combinations (approximately 65%) are represented by a single fiber type only. Synthetic fibers exhibit a very high degree of polymorphism.
- The chance of one type representing >1% of the total population is remote. A collective of synthetic fibers (same type, same morphology, and same color) can therefore be said to be the result of a recent contact with a specific textile and can be considered as being highly significant.

Target fiber studies give some idea of what proportion of the population has fibers of a particular type on their clothing (or car seat, cinema seat, pub seat, etc.). This data incidentally automatically takes into account not only the rarity of a fiber type but also the tendency for this fiber to be transferred or to persist on the receiving fabric. Without being comprehensive, a list of target fiber studies that have been carried out to date include the following:

- a blue wool nylon pullover from Mark & Spencer with sales over 1 million;
- red wool from ladies' pullover and brown polyester of men's trousers on car seats;
- green cotton fibers from Mark & Spencer's leggings and red acrylics from a pullover on car and cinema seats;
- red acrylics from a scarf sold over a five-year period in quantities of 5–10,000 annually in nine European countries;
- blue wool fibers on seats in public houses throughout the United Kingdom; and
- blue wool, gray, and black polyester (all common) as well as blue acrylic (less common) on outer clothing from English households.

Findings of similar studies in different parts of the world regularly appear in the literature. So far, target fiber studies suggest that it is a rare event to find a fiber, even from a very common garment on clothing (or a seat) without the donor recently having been in close contact with the garment in question.

Data bases are generally built using garments received in forensic laboratories for examination. They yield to very detailed information such as the

- frequency of morphological characters within a fiber type;
- frequency of a certain polymer composition within a fiber type;
- frequency of uncommon fiber types in the general population;
- frequency of usage of fiber types in different textiles divided into different categories; and
- frequency of certain fiber-type combinations in different textiles.

One of the criticisms of data bases is that the collection of materials takes too long to accumulate and, hence, the data quickly becomes out of date. However, this criticism only really applies where the measurement of color data is involved. To overcome this problem, some authors have built up a comprehensive and easily updated database using data on

garments offered for sale in mail order catalogs. However, this approach is feasible only in countries where mail order is popular.

Trade enquiries involve tedious and time-consuming investigative work. However, such enquiries can bring extremely valuable information on the relative frequency of a given type of fiber, including sales figures and manufacture figures such as quantity of items dyed using the same dye batch, for example.

Statistical Treatment (Probabilistic Model)

The interpretative process is not entirely free from subjective opinion when dealing with fiber evidence because of the number of interactive factors that must be considered. As a result, a probabilistic model has been proposed as an aid to this process.

Taking note of the fact that the decision process at the time of the trial is based on inference, it appears that the logical rules for thinking about facts in legal cases are those of probability. Only probabilities are able to measure the uncertainty that exists and evolves during every criminal case. Therefore, only a model supported by the mathematical theories of probability is valid.

During the past 25 years, many researchers have shown increasing interest in a probabilistic model based on the Bayes Theorem. This model describes how the elements of proof combine with themselves and evolve during the decision process of the trial, avoiding distortion based on inappropriate use of intuition. The Bayesian framework, or more precisely on the likelihood ratio approach, can help in formulating and assessing the most relevant questions faced by the forensic scientist. The relevant questions are dependant upon the circumstances of the case, but they are all derived from two fundamental questions:

- What is the probability of the evidence if the alleged story is true?
- What is the probability of the evidence if the alleged story is not true (or the complimentary story is true)?

The ratio of these two probabilities is called a *likelihood ratio*. The latter measures the value of the evidence in terms of a pair of hypotheses, indicating if the given set of observations supports one hypothesis more than the other. It clearly shows that the concept of evidence, including fiber evidence, is relative.

The main challenge of a probabilistic model is in the fact that such a model uses data that, in the fibers context, are incomplete and difficult to quantify. Ideally, for each case, a target fiber study and transfer and persistence experiments should be carried out using relevant materials to the context of the case. Of course, such an approach is not practical and the assessment relies rather on the accumulated knowledge derived from experimental studies.

However, the Bayesian approach can still be used as a logical framework to provide the most accurate guide to interpretation currently possible. Conservative estimates that can be seen as "measures of belief" and simulations can also help in going through the process. In other words, this approach greatly minimizes the subjective component of the interpretation process. It is also worth noting that the Bayesian approach has been the initiator of numerous experimental studies, which brought crucial data for the interpretation of fiber evidence, regardless of whether the scientist applies the Bayesian framework or not.

Examples. Various examples of the application of the Bayesian approach to the assessment of the evidential value of fibers were developed in [44, 45] and were further discussed in [46]. Inferential guidelines can be extracted from these scenarios. They ultimately assist the forensic scientist in the evaluation of fiber evidence.

As noted in [45], the main points are described as follows:

- Evaluation of the evidential value of forensic evidence (including fiber evidence) is a matter of probability assessment.
- For the assessment of the strength of the evidence (E), it is necessary to consider the probability of the evidence under two given competing explanations for its occurrence, respectively presented by the prosecutor and by the defense (H_1 and H_2). The value of the evidence is estimated using a likelihood ratio $LR = P(E|H_1)/P(E|H_2)$. Hence, the likelihood ratio is defined not only by the evidence but also by the strategy chosen either by the prosecution or by the defense.
- Evidence (recovered trace and known material) may be to some extent incompatible with the offense. Hence, $P(E|H_1)$ is not always equal to 1

because of their number and position, and because of the phenomena of transfer, persistence, and recovery. There is a need to consider transfer probabilities.

- When traces (recovered evidence and known material) are supposed to be associated with the offense, we also have to allow for the probability of their absence before the commission of the offense. The corollary of this is that, if traces are supposed to be unrelated to the offense, the probability of their presence by chance also has to be considered. There is a need to consider background probabilities.
- The evaluation of trace evidence has to consider all potential recovered traces and not only the concordant evidence. This number of groups has to be considered.
- The number of declared offenders has a significant impact on the likelihood ratio.
- The definition of H_2 excludes the implication of the defendant. Thus, the scientist has to specify the relevant population in which adequate forensic surveys have to be made. There is a need to estimate frequencies correctly.
- In all investigation and evaluation of transfer traces, all possible exchanges have to be investigated including extreme situations in which expected evidence has either not been found or not been produced.

Conclusions

Background data available to the forensic fiber examiner to assist in interpreting the significance of fiber findings has increased greatly in the past 25 years. There is also a deeper understanding of the factors that need to be considered, and acceptance that it is incumbent on the practitioners to assist the court in assessing significance.

Given the circumstances facing forensic fiber examiners in different climatic and socioeconomic environments, local databases and research to put available data in a local context are essential. This type of data is now emerging.

The Bayesian approach has been studied and proposed in the area of fiber evidence. It is increasingly seen as a valid logical framework to provide the most accurate guide to interpretation currently possible. However, it is fair to say that the complexity of fiber evidence (when compared to other evidence categories like glass, for example) remains a challenge, which somewhat inhibits its systematic application in casework.

Ultimately, it is worth noting that the value of fibers is more in their ability to reconstruct an event or a series of events (i.e., "fibers can tell a story") rather than necessarily uniquely identify the source at the origin of a transfer. To make most of their use, they should therefore be used at the activity level and not the source level.

References

[1] Gaudette, B.G. (1988). The forensic aspects of textile fiber examination, in *Forensic Science Handbook*, R. Saferstein, ed, Prentice Hall, Englewood Cliffs, Vol. **2**, pp. 209–272.

[2] Robertson, J. & Grieve, M. (1999). *Forensic Examination of Fibres*, 2nd Edition, Taylor & Francis, London and Philadelphia.

[3] Grieve, M. (1999). Interpretation of fibres evidence – influential factors, quality assurance, report writing and case examples, in *Forensic Examination of Fibres*, 2nd Edition, J. Robertson & M. Grieve, eds, Taylor & Francis, London and Philadelphia, pp. 343–378.

[4] Pounds, C.A. & Smalldon, K.W. (1975). The transfer of fibers between clothing materials during simulated contacts and their persistence during wear. Part III-a preliminary investigation of the mechanisms involved, *Journal of the Forensic Science Society* **15**, 197–207.

[5] Pounds, C.A. & Smalldon, K.W. (1975). The transfer of fibers between clothing materials during simulated contacts and their persistence during wear. Part I-Fiber transference, *Journal of the Forensic Science Society* **15**, 17–27.

[6] Pounds, C.A. & Smalldon, K.W. (1975). The transfer of fibers between clothing materials during simulated contacts and their persistence during wear. Part II-Fiber persistence, *Journal of the Forensic Science Society* **15**, 29–37.

[7] Kidd, C. & Robertson, J. (1982). The transfer of textile fibers during simulated contacts, *Journal of the Forensic Science Society* **22**, 301–308.

[8] Robertson, J., Kidd, C.B.M. & Parkinson, H.M.P. (1982). The persistence of textile fibers transferred during simulated contacts, *Journal of the Forensic Science Society* **22**, 353–360.

[9] Cordiner, S.J., Stringer, P. & Wilson, P.D. (1985). Fiber diameter and the transfer of wool fibers, *Journal of the Forensic Science Society* **25**, 425–426.

[10] Parybyk, A.E. & Lokan, R.J. (1986). A study of the numerical distribution of fibers transferred from blended fabrics, *Journal of the Forensic Science Society* **26**, 61–68.

[11] Salter, M.T., Cook, R. & Jackson, A.R. (1987). Differential shedding from blended fabrics, *Forensic Science International* **33**, 155–164.
[12] Roux, C., Chable, J. & Margot, P. (1996). Fiber transfer experiments onto car seats, *Science & Justice* **36**, 143–151.
[13] Siegel, J.A. (1997). Evidential value of textile fiber-transfer and persistence of fibers, *Forensic Science Review* **9**, 81–96.
[14] Merciani, P., Monard, F.S., Buzzini, P., Massonnet, G. & Taroni, F. (2003). A study of the cross transfer of fibers, *Forensic Science International* **136**(1), 123.
[15] Marshall, L., Griffin, R.M.E. & Robinson, K. (2003). Recovery of transferred carpet fibres from shoes, their persistence on the outer surface and the value of fibres recovered from the inner surfaces of the shoes, *Forensic Science International* **136**(1), 123–124.
[16] Akulova, V., Vasiliauskiene, D. & Talaliene, D. (2002). Further insights into the persistence of transferred fibres on outdoor clothes, *Science & Justice* **42**(3), 165–171.
[17] Watt, R., Roux, C. & Robertson, J., Australian and New Zealand Forensic Science Society (2004). The influence of front loading and top loading washing machines on the persistence, redistribution and secondary transfer of textile fibres during laundering, *Proceedings of the 17th International Symposium on the Forensic Sciences*, Wellington.
[18] Palmer, R. & Banks, M. (2005). The secondary transfer of fibres from head hair, *Science & Justice* **45**(3), 123–128.
[19] Roux, C., Langdon, S., Waight, D. & Robertson, J. (1999). The transfer and persistence of automotive carpet fibres on shoe soles, *Science & Justice* **39**(4), 239–251.
[20] Biermann, T.W. & Grieve, M.C. (1996). A computerized data base of mail order garments: a contribution toward estimating the frequency of fiber types found on clothing. Part 1: the system and it's operation, *Forensic Science International* **77**, 65–73.
[21] Biermann, T.W. & Grieve, M.C. (1996). A computerized data base of mail order garments: a contribution toward estimating the frequency of fiber types found on clothing. Part 2: the content of the data bank and it's statistical evaluation, *Forensic Science International* **77**, 75–91.
[22] Biermann, T.W. & Grieve, M.C. (1998). A computerized data base of mail order garments: a contribution toward estimating the frequency of fiber types found on clothing. Part 3: the content of the data bank – is it representative? *Forensic Science International* **95**, 117–131.
[23] Home, J.M. & Dudley, R.J. (1980). A summary of data obtained from a collection of fibers from casework material, *Journal of the Forensic Science Society* **20**, 253–261.
[24] Cook, R. & Wilson, C. (1986). The significance of finding extraneous fibers in contact cases, *Forensic Science International* **32**, 267–273.
[25] Jackson, G. & Cook, R. (1986). The significance of fibers found on car seats, *Forensic Science International* **32**, 275–281.
[26] Palmer, R. & Chinherende, V. (1996). A target fiber study using cinema and car seats as recipient items, *Journal of the Forensic Science Society* **41**, 802–803.
[27] Roux, C. & Margot, P. (1997). The population of textile fibers on car seats, *Science & Justice* **37**, 25–30.
[28] Grieve, M.C. & Biermann, T. (1997). The population of coloured textile fibers on outdoor surfaces, *Science & Justice* **37**, 231–239.
[29] Bruschweiler, W. & Grieve, M.C. (1997). A study on the random distribution of a red acrylic target fiber, *Science & Justice* **37**, 85–89.
[30] Kelly, E. & Griffin, R.M.E. (1998). A target fiber study on seats in public houses, *Science & Justice* **38**, 39–44.
[31] Wiggins, K.G. & Allard, J.E. (1987). The evidential value of fabric car seats and car seat covers, *Journal of the Forensic Science Society* **27**, 93–101.
[32] Cantrell, S., Roux, C., Maynard, P. & Robertson, J. (2001). A textile fibre survey as an aid to the interpretation of fibre evidence in the Sydney region, *Forensic Science International* **123**, 48–53.
[33] Dignan, S.J. & Murphy, K.J. (2002). Fibre evidence from fingernail clippings, *Canadian Society of Forensic Science Journal* **35**(1), 17–21.
[34] Grieve, M.C., Biermann, T.W. & Davignon, M. (2003). The occurrence and individuality of orange and green cotton fibres, *Science & Justice* **43**(1), 5–22.
[35] Palmer, R. & Oliver, S. (2004). The population of coloured fibres in human head hair, *Science & Justice* **44**(2), 83–88.
[36] Watt, R., Roux, C. & Robertson, J. (2005). The population of coloured textile fibres in domestic washing machines, *Science & Justice* **45**(2), 75–83.
[37] Wiggins, K., Drummond, P. & Hicks Champod, T. (2004). A study in relation to the random distribution of four fibre types on clothing (incorporating a review of previous target fibres studies), *Science & Justice* **44**(3), 141–148.
[38] Houck, M.M. (2003). Inter-comparison of unrelated fiber evidence, *Forensic Science International* **135**, 146–149.
[39] Grieve, M.C. & Biermann, T.W. (2003). The individuality of blue polyester fibers used to provide forensic evidence, *Forensic Science International* **136**(1), 121–122.
[40] Cresswell, S.L., Cunningham, D. & NicDaeid, N. (2003). Textile survey of cinema seats in Glasgow, *Forensic Science International* **136**(1), 117.
[41] Was-Gubala, J. (2004). Comparative population studies of fibres secured in Poland, Czech Republic and Germany, *Zagadnien Nauk Sadowych* **60**, 58–77.
[42] Grieve, M., Biermann, T. & Schaub, K. (2005). The individuality of fibres used to provide forensic evidence – not all blue polyesters are the same, *Science & Justice* **45**(1), 13–28.
[43] Marname, R., Elliot, D. & Coulson, S. (2006). A pilot study to determine the background population of foreign

fibre groups on a cotton/ polyester T-shirt, *Science & Justice* **46**(4), 215–220.
[44] Champod, C. & Taroni, F. (1997). Bayesian framework for the evaluation of fiber transfer evidence, *Science & Justice* **37**, 75–83.
[45] Champod, C. & Taroni, F. (1999). Interpretation of fibres evidence – the Bayesian approach, in *Forensic Examination of Fibres*, 2nd Edition, J. Robertson & M. Grieve, eds, Taylor & Francis, London & Philadelphia, pp. 379–398.
[46] Aitken, C. & Taroni, F. (eds) (2004). Fibres, *Statistics and Evaluation of Evidence for Forensic Scientists*, 2nd Edition, John Wiley & Sons, Chichester, pp. 381–398.
[47] Causin, V., Schiavone, S., Marigo, A. & Carresi, P. (2004). Bayesian framework for the evaluation of fiber evidence, *Forensic Science International* **141**, 159–170.

Related Articles

Evidence Interpretation: a Logical Approach

Examination of Fibers and Textiles

Report Writing for Courts

Statistical Evidence in Court

Trace Evidence: Transfer, Persistence, and Value

CLAUDE ROUX AND JAMES ROBERTSON

Filaments: Light, Examination *see* Light Bulbs and Filaments: Examination of

Fingermarks: Detection and Recovery *see* Friction Ridge Skin: Fingerprint Detection and Recovery Techniques

Fingerprints *see* Friction Ridge Examination (Fingerprints): Interpretation of

Fingerprints: DNA *see* Friction Ridge Skin: Interaction between Fingerprint Detection and DNA/Biological Material

Fire: Accrediting and Certifying Fire Investigators *see* Fire Investigator: Standardization, Accreditation, and Certification

Fire: Chemistry of

Fire and Energy

Fire is a complex phenomenon, which evolves in the three dimensions of space and time. It is a chemical reaction that occurs in the vapor phase but involves solids and liquids as well, through an energy feedback loop. The energy of the fire drives the chemical processes of pyrolysis and vaporization, which must be maintained if the fire is to keep burning. Fire is an exothermic chemical reaction, i.e., one that gives off energy in the form of heat and light. The understanding of fire requires an understanding of the basic concepts of energy.

Much of what is known about energy involves the transformation of energy from one form to another. For example, gasoline contains chemical energy, and

when burned in an internal combustion engine, the chemical energy is converted to heat energy, which, in turn, is converted into mechanical energy. The power company burns coal to boil water. The steam produced from the boiling liquid moves a turbine, which spins a magnet inside a coil of wires to produce electrical energy. When that electrical energy is passed through a filament to make light or through a resistance element to heat water or air, it is transformed back into heat energy. When supplied to a motor, the electrical energy is converted into mechanical energy.

It is useful to think of energy as *the ability to do work*. Heat energy is actually a form of work. Heat can be used to boil water, and the increase in pressure above boiling water in a closed vessel can be used to move matter, as in a steam engine. Count Rumford's experiment, wherein he used the frictional heat from cannon boring to boil water, allowed for an understanding of the concept of energy transfer [1].

If a glass of ice water is placed in a room, heat will flow from the room into the glass until the ice melts. Eventually, the water will be the same temperature as the room, and heat transfer will cease. Energy transfer that takes place by virtue of a temperature difference exclusively, i.e., flowing from an area of high temperature to an area of lower temperature, is called *heat flow*. In the eighteenth century, Lavoisier and others thought that because heat was flowing, it must be a substance, and they named that substance *caloric*. Count Rumford and James P. Joule, an English physicist, came to understand that what was flowing was energy.

In honor of Joule's work, the basic unit of energy was named after him. The unit of work in any system of measurement is the unit of force multiplied by the unit of distance. In the metric system, the unit of force is the newton and the unit of distance is the meter. A newton is that force that gives 1 kg an acceleration of $1\,m\,s^{-2}$; 1 Nm equals 1 J.

But how does the movement of an object translate to the energy from a fire? As Rumford demonstrated, work can be measured through its ability to raise the temperature of a volume of water by a fixed number of degrees. A *calorie* is defined as the amount of energy required to raise the temperature of 1 g of water $1\,^{\circ}C$. (A corresponding unit, described in terms of degrees Fahrenheit and pounds of water, is a British thermal unit (Btu). One Btu is the quantity of heat required to raise the temperature of one pound of water from 63 to $64\,^{\circ}F$. One Btu equals 252 calories.) Different substances require different amounts of heat to increase their temperature by a specific amount. This parameter is called *the specific heat capacity* of the material and is expressed as joules per kilogram per degree Celsius. In a fire, depending on the duration, different materials can reach different temperatures despite being exposed to the same amount of energy. Just as lifting a weight from one floor of a house to another can take place rapidly or slowly, so can the process of increasing the temperature of a body of water.

The amount of work required to raise a given weight to a given height is the same whether it takes a second, a minute, or an hour. Thus, the *rate* at which work is done becomes important when considering energy or heat transfer. The amount of work done *per unit time* is the quantity of interest.

Power is defined as the work done divided by the time interval. Appliances that use energy are described by the power they produce, either in Btu per hour, or in watts. A *watt* is defined as $1\,J\,s^{-1}$, so, therefore, the time component is built in.

Just like a heating appliance, the size of fire can be described in terms of watts, or more commonly, kilowatts or megawatts. To understand fire, it is necessary to understand ignition sources and fire in terms of its power in watts. Consider a controlled fire, a gas-fired stove-top burner. A typical burner is rated at $12\,000\,Btu\,h^{-1}$. A total of $12\,000\,Btu\,h^{-1}$ equals $3.33\,Btu\,s^{-1}$. One Btu equals 1054.8 J, so that $12\,000\,Btu\,h^{-1}$ burner is putting out just over $3500\,J\,s^{-1}$, or 3500 W. A $40\,000\,Btu\,h^{-1}$ burner, the size typically found in a gas-fired water heater, puts out 11 720 W, and a $125\,000\,Btu\,h^{-1}$ gas furnace delivers 36.6 kW.

The power of a fire in kilowatts is known as its *heat release rate* (*HRR*). The HRR is the single most important property of fire, because it allows predictions of how that fire will behave [2]. The HRR affects the temperature of the fire, its ability to entrain air, and the identity of the chemical species produced in the fire.

The power of fire is an important property to know, but it is equally important to know how that energy is distributed. A total of 20 kW spread evenly throughout a structure by a furnace's circulation fan will keep that structure comfortable on a winter day. Confining or focusing the energy can result in dramatically different consequences. The concept of

heat flux is therefore an important consideration. Heat flux is a measure of the rate of energy falling on or flowing through a surface, and is measured in terms of watts or kilowatts per unit area. Fire can be described as having a heat flux in watts per square centimeter ($W\,cm^{-2}$) or in kilowatts per square meter ($kW\,m^{-2}$); $1\,W\,cm^{-2}$ equals $10\,kW\,m^{-2}$.

All combustible materials have a property known as *critical radiant heat flux*, which is defined as the amount of energy necessary to cause ignition. Typical critical radiant heat fluxes are on the order of $20\,kW\,m^{-2}$.

On a clear day, energy from the sun strikes the earth with a radiant heat flux of approximately $1\,kW\,m^{-2}$, an amount that varies depending on time and location. This is enough energy to cause sunburn in 30 min or less. We cannot increase the HRR of the sun, but we can increase the radiant heat flux by focusing the energy that falls on a large area onto a smaller area, using a magnifying glass. Figure 1 shows what happens if we use a magnifying glass (or a concave mirror) to decrease the surface area by 96%. The $1\,kW\,m^{-2}$ heat flux becomes $25\,kW\,m^{-2}$ at the focal point, which is sufficient to ignite most combustibles. Figure 2 shows burn lines scorched onto the underside of a wooden roof by concentrated sunlight striking a concave makeup mirror in a bathroom. Parallel lines of charring were produced over several days, eventually resulting in the ignition of the roof. Table 1 shows the observed effect of typical radiant heat fluxes.

Chemical Processes

When exposed to heat, materials can undergo physical or chemical changes. Physical changes are

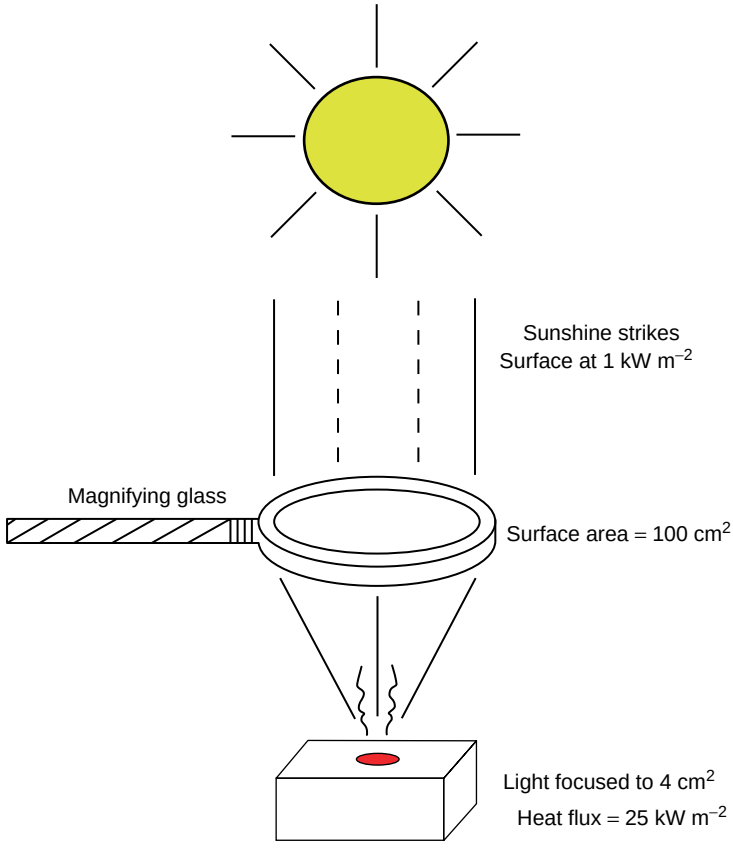


Figure 1 A magnifying glass increases the radiant heat flux from the sun by reducing the surface area that the light covers [Reproduced from Ref. 3. © Taylor and Francis Group, 2006.]



Figure 2 Burn lines scorched into the underside of a roof on successive days. A concave mirror located on a bathroom vanity focused the sun's energy out through the window [Reproduced from Ref. 3. © Taylor and Francis Group, 2006.]

Table 1 Typical radiant heat fluxes

Approximate radiant heat flux ($kW\,m^{-2}$)	Comment or observed effect
170	Maximum heat flux as currently measured in a postflashover fire compartment.
80	Heat flux for protective clothing thermal protective performance (TPP) test.[a]
52	Fiberboard ignites spontaneously after 5 s.[b]
29	Wood ignites spontaneously after prolonged exposure.[b]
20	Heat flux on a residential family room floor at the beginning of flashover.[c]
16	Human skin experiences sudden pain and blisters after 5-s exposure with second-degree burn injury.[a]
12.5	Wood volatiles ignite with intended exposure[d] and piloted ignition.
10.4	Human skin experiences pain with 3-s exposure and blisters in 9 s with second-degree burn injury.[a][b]
6.4	Human skin blisters in 18 s with second-degree burn injury.[a][b]
4.5	Human skin becomes blistered with a 30-s exposure, causing a second-degree burn injury.[a]
2.5	Common thermal radiation exposure while fire fighting.[f] This energy level may cause burn injuries with prolonged exposure.
1.4	Thermal radiation from the sun. Potential sunburn in 30 min or less.[g]

Source: NFPA 921, Guide for Fire and Explosion Investigations. With permission

[a] From NFPA 1971, Standard on Protective Ensemble for Structural Fire Fighting
[b] From Lawson, "Fire and the Atomic Bomb"
[c] From Fang and Breese, "Fire Development in Residential Basement Rooms"
[d] From Lawson and Simms, "The Ignition of Wood by Radiation," pp. 288–292
[e] From Tan, "Flare System Design Simplified," pp. 172–176
[f] From US Fire Administration, "Minimum Standards on Structural Fire Fighting Protective Clothing and Equipment"
[g] From Bennett and Myers, Momentum, Heat, and Mass Transfer

The unit kilowatt per square meter defines the amount of heat energy or flux that strikes a known surface area of an object. The unit (kW) represents 1000 W of energy and the unit (m^2) represents the surface area of a square measuring 1-m long and 1-m wide. For example, $1.4\,kW\,m^{-2}$ represents 1.4 multiplied by 1000 and equals 1400 W of energy. This surface area may be that of the human skin or any other material

changes in state, i.e., from solid to liquid or from liquid to gas. A change in state involves no change in the identity of the molecules or atoms involved. State changes are reversible, i.e., upon cooling, vapors can condense into liquids and liquids can solidify. Chemical processes brought about by an exposure to heat in the presence of oxygen typically include pyrolysis and oxidation. Pyrolysis is the irreversible breaking down of large molecules into small molecules as the result of heating. In a fire, those small molecules are usually oxidized by reaction with atmospheric oxygen. Pyrolysis is also known as *thermal decomposition*.

Gases

Most substances that we encounter in our everyday experience are not pure elements, but mixtures of compounds. The simplest substances are the gases. In a gas, the individual components, atoms, or more commonly, molecules, are not tightly bound to one another. A pure gas can be an elemental gas or it can be a compound. Elemental gases, with the exception of the inert or noble gases (helium, neon, argon, krypton, xenon, and radon), usually exist as diatomic gases, that is, there are two atoms of the same element bound to each other. Thus, hydrogen, oxygen, and nitrogen occur as H_2, O_2, and N_2, respectively. Gases of compounds have at least two different atoms bound together, such as carbon monoxide, CO, or methane, CH_4. The gas, with which we are most familiar, air, consists of approximately one part oxygen to four parts nitrogen.

Gases behave in a predictable fashion when exposed to changes in volume, pressure, or temperature. An important property of all gases is that regardless of which gas we are considering, equal volumes of gases contain equal numbers of molecules. Most gases come very close to obeying the ideal gas law, $PV = nRT$, where P is the pressure, V is the volume, n is the number of molecules, R is a number ($8.314472\,\mathrm{K^{-1}\,mol^{-1}}$) known as the *universal gas constant*, and T is the temperature *in degrees kelvin*. If the size of the container is reduced by half, the pressure doubles. If the number of molecules is doubled, the pressure doubles, or if the container is allowed to expand, the volume doubles. Raising the temperature from 25 to 324 °C (298–596 K) doubles the volume if the pressure stays constant or doubles the pressure if the volume stays constant.

Combustion reactions in gases were among the first chemical reactions studied and understood. The combustion of hydrogen in air to make water is perhaps the simplest. Two volumes of hydrogen plus one volume of oxygen yields two volumes of water vapor.

$$2H_2 + O_2 \rightarrow 2H_2O \qquad (1)$$

The simplest carbon-based combustion reaction involves methane. One volume of methane plus two volumes of oxygen yields two volumes of water vapor plus one volume of carbon dioxide.

$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O \qquad (2)$$

Note that if there is insufficient oxygen, carbon monoxide instead of carbon dioxide will be produced. It is worth noting that, while the above reactions have described volumes of oxygen, most combustion reactions take place in air, where oxygen comprises only about 20% of the available gas. Thus, a volume of methane actually requires 10 volumes of air in order to burn completely.

The chemical equations that describe combustion reactions list only the initial and final chemical species. An important by-product of the reaction is heat. The amount of heat produced (energy released) depends on the substance being oxidized. The *heat of combustion* is measured as energy per unit mass, usually reported as kilojoules per kilogram ($\mathrm{kJ\,kg^{-1}}$), but can also be reported as Btu per pound or calories per gram. In the case of gases, the heat of combustion is often expressed as kilojoules per cubic meter or Btu per cubic foot. The HRR is then the heat of combustion multiplied by the *mass loss rate*, the rate at which the fuel is consumed.

A cubic foot of natural gas has an energy content of approximately 1000 Btu, or more than a million joules. If all this energy is released in one second, the explosion will have a HRR of 1 MW. If burned over an hour, as would be typical for a pilot light, the HRR is less than 300 W. The energy released is the same in both cases. The difference is in how fast that energy is released. When one considers the energy available if a $40\,000\text{-Btu}\,\mathrm{h^{-1}}$ burner is left open without a pilot light for an hour, one begins to understand the necessity for safety devices on gas-burning appliances.

Natural gas is composed mainly of methane, CH_4. Liquefied petroleum (LP) gas is composed mainly of

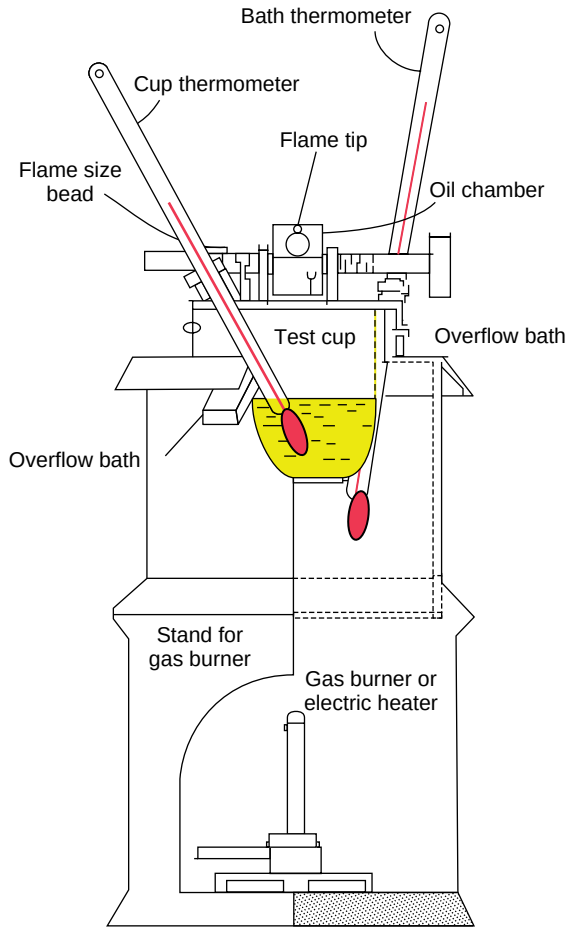


Figure 3 Tag closed cup flash point testing apparatus [Reproduced from Ref. 3. © Taylor and Francis Group, 2006.]

propane, C_3H_8. Where a given volume of natural gas requires two volumes of oxygen to burn, a volume of propane requires five volumes of oxygen. Propane, likewise, contains more energy per cubic foot than natural gas. A cubic foot of natural gas contains about 1000 Btu, where a cubic foot of propane contains about 2500 Btu. Interestingly, a cubic foot of natural gas requires 10 ft^3 of air, and a cubic foot of propane requires 25 ft^3 of air to burn efficiently. This comparison suggests a relationship between the amount of oxygen consumed and the amount of heat capable of being released by a given volume of gas. Hydrocarbon fuels are remarkably consistent in this regard. Completely burned hydrocarbon fuels release 13.1 $kJ\,g^{-1}$ of oxygen consumed (±5%). This relationship forms the basis of oxygen consumption calorimetry as a means for measuring the HRR of fire.

Liquid

The molecules in liquids and solids have an affinity for each other, which reduces the space between the molecules and results in condensation. Liquids typically occupy 1/100 to 1/300 of the volume of a similar number of gas molecules. While many people think of liquids as burning, the fact is that liquids only burn when their molecules are vaporized into the gas phase. Once in the gas phase, or, more correctly, vapor phase, the molecules behave as though they were gases. Combustion can only occur when there is a sufficient number of molecules of

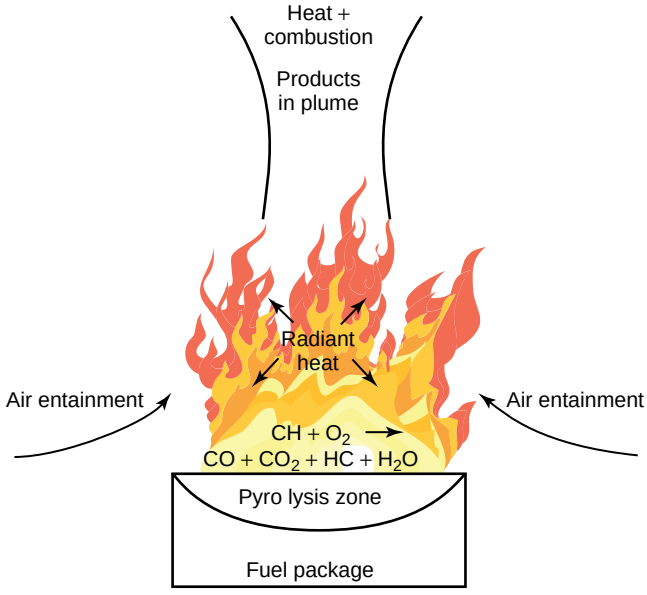


Figure 4 Schematic representation of some of the processes taking place in the combustion of a solid fuel. Combustion continues for as long as the flame radiates sufficient energy back to the surface to produce sufficient quantity of gases and vapors [Reproduced from Ref. 3. © Taylor and Francis Group, 2006.]

the liquid fuel vaporized in the space above a pool of liquid. The temperature at which the concentration of vapors above the liquid in a still pool enclosed in a specified apparatus will ignite when exposed to a flame is known as the *flash point*. A common flash point apparatus, the tag closed cup tester, is shown in Figure 3. The flash point is also referred to as the *piloted ignition temperature*. Typically, ignition cannot be achieved when a liquid is below its flash point, but it is possible to get a sufficient number of molecules of the liquid into the air space above it through atomization or through wick action.

Solids

The chemical processes involved in the combustion of solids are far more complex than those involving liquids and gases. If solids were simply liquids at a temperature below their freezing point (or melting point), describing the combustion behavior of solids would be relatively easy. A candle is an example of this kind of solid. The heat from the flame melts the solid wax to produce liquid wax, and the liquid is then vaporized so that it can combine with oxygen in the air. Most solids, however, and almost all combustible solids, do not melt or vaporize to form liquids or gases and can only be transformed into liquids and gases by breaking down large molecules into new and smaller molecules. There are no gaseous "wood molecules" or "wool molecules", nor are there gaseous molecules of many polymers. When wood burns, it is first necessary to break it down into a liquid or gas through a process known as *pyrolysis*. Whether an ignition source is capable of igniting a particular fuel (a *competent* ignition source) depends on whether it can cause the production of sufficient vapors through pyrolysis. Once these vapors are ignited, the fuel will continue to burn for as long as the resulting flame transfers sufficient energy to the surface of the solid so that it continues to produce vapors. Figure 4 shows the processes involved in the burning of a solid.

Almost all combustible substances are either natural or synthetic polymers. Wood is a mixture of three natural polymers: cellulose, hemicellulose, and lignin. Silk and wool are proteins, another kind of natural polymer. The six most common synthetic polymers are polyethylene, polypropylene, polyvinylchloride, polystyrene, polyester, and polyurethane [3]. When exposed to heat, polymers undergo an irreversible

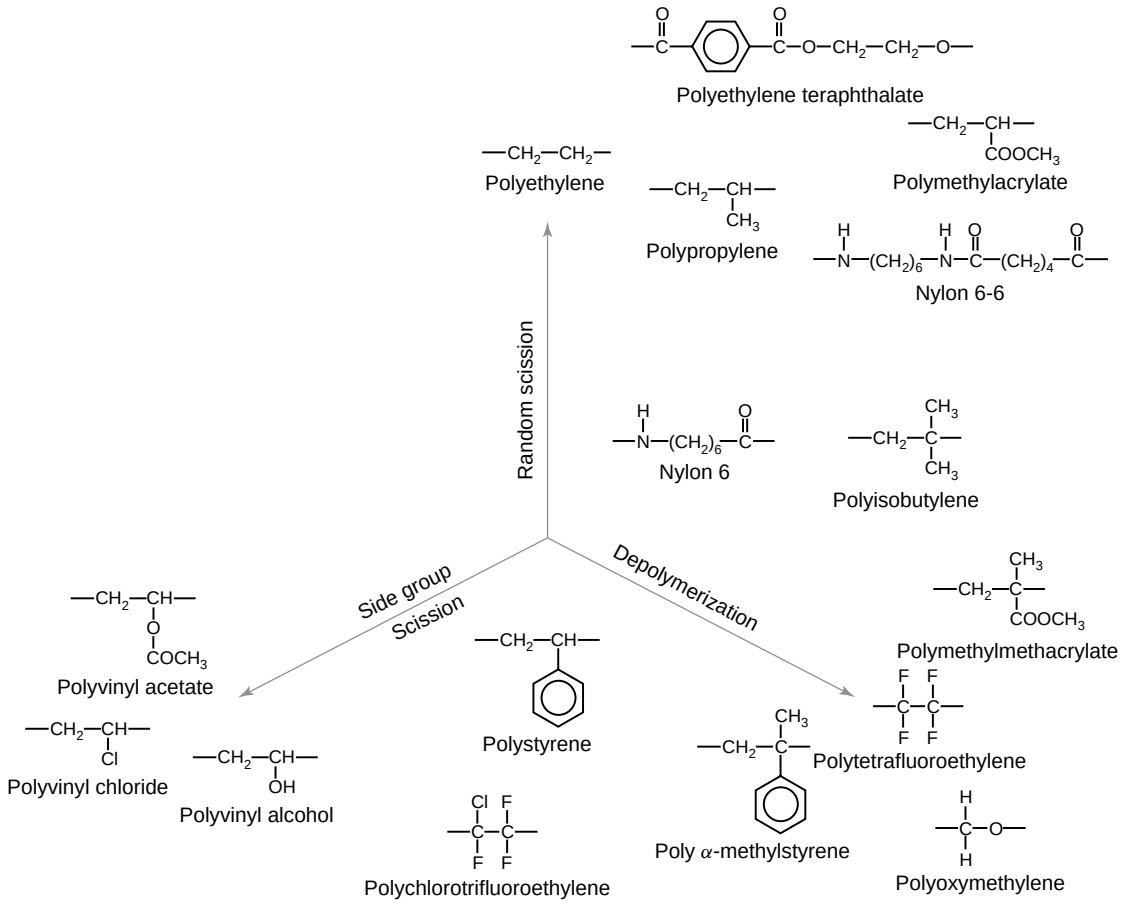


Figure 5 Some commonly encountered polymers and their approximate classification according to their pyrolysis mechanisms [Reproduced from Ref. 7. © Elsevier, 2003.]

decomposition, caused by the breaking of chemical bonds, or pyrolysis. The chemical bonds within the molecules undergoing pyrolysis break in the order of their bond strength. Weaker bonds break before stronger bonds. Bond strength is the amount of energy in kilojoules (kJ) per mole required to dissociate one atom from another. For example, the bond dissociation energy of the carbon–chlorine bond is 330 kJ mol^{-1}, whereas the bond dissociation energy of a carbon–carbon single bond is 350–380 kJ mol^{-1} [4]. When exposed to fire temperatures, therefore, polyvinyl chloride (PVC) loses its chlorine atoms, which rapidly combine with atmospheric water to form hydrogen chloride (HCl). The carbon atoms remain bonded to each other, and, in fact, form double bonds because of the loss of the chlorine.

The progressive dissociation of chemical bonds results in three basic mechanisms by which pyrolysis typically occurs: random scission, side group scission, and depolymerization. Depolymerization, also known as *monomer reversion*, is the simplest mechanism. The polymers simply dissociate into the original monomer. Styrene polymers, acrylics, and polytetrafluoroethylene (PTFE, frequently used as a nonstick surface coating) undergo monomer reversion. Monomer yield for these plastics is greater than 80%. For other plastics, like PVC, the monomer yield is close to zero [5].

Polymers containing many bonds of equal or nearly equal strength tend to undergo random scission. Polyethylene, for example, breaks down into normal alkenes (straight chain hydrocarbons with a double

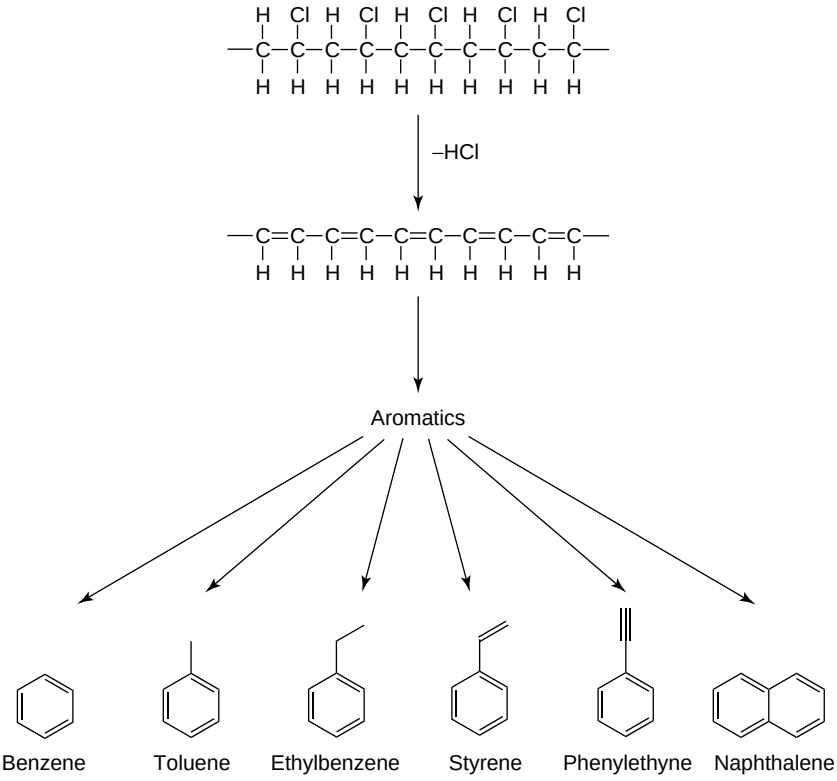


Figure 6 Products of the pyrolysis of polyvinyl chloride (PVC) [Reproduced from Ref. 7. © Elsevier, 2003.]

bond), alkanes (straight chain hydrocarbons) and dienes (straight chain hydrocarbons with two double bonds). Asphalt undergoes the same kind of thermal decomposition, but without the dienes [6]. It is therefore possible to distinguish between asphalt smoke residue and polyethylene smoke residue, even though they are quite similar. Both asphalt and polyethylene (polythene) are common fuels, and their chemical residues are commonly found in samples of fire debris.

Side chain scission involves the selective loss of side groups attached to a long chain backbone. The loss of chlorine from PVC is a typical side chain scission reaction. Many polymers will undergo a combination of mechanisms. Figure 5 is a diagram showing the typical pyrolysis mechanisms of several polymers.

Pyrolysis is, in many cases, a sequential process. After the chlorine atoms are stripped from the PVC polymer (or any polymer that undergoes side chain scission) the remaining backbone of the polymer is highly reactive. The formation of aromatic molecules is relatively straightforward. Aromatics make up over 90% of the detectable residue from pyrolyzed PVC [7]. Figure 6 shows the side group scission mechanism of PVC and the resulting products. Given that there are no aromatic functional groups present in the original PVC polymer chain, it is far from intuitively obvious that PVC should degrade into aromatic compounds. These compounds are some of the same compounds found in gasoline and other petroleum products, but the relative ratios of the aromatics (for example, the xylenes) are different than the relative ratios in petroleum products. For a further discussion of interpretation of aromatics in fire debris, *see* **Fire Debris: Laboratory Analysis of**.

Solids exposed to heat can be expected to do one of three things: melt, dehydrate, or char. Since they are mixtures, as opposed to pure compounds, most solids that melt do not have a sharp melting point, but typically exhibit what is called a *glass transition temperature*. Substances that dehydrate lose water

molecules. These are molecules that are chemically bound to the solid matrix but are only weakly bound. Substances such as gypsum wallboard and concrete contain bound water, and this gives them some of their fire resistance. When the water is released, calcination takes place in the gypsum wallboard. The water released from concrete is thought to play a role in a phenomenon known as *spalling*.

Charring occurs when volatiles (small molecules) are created through pyrolysis and then driven off by the external heat source, where they ignite and burn in air. The flaming combustion above the surface of the solid provides more heat for additional pyrolysis and volatilization. What is left behind is known as *char*. Char can insulate the material below it from the heat source because it is generally less conductive than the solid from which it originated. It is possible to determine an average charring rate for substances exposed to a calibrated energy source such as a furnace, but it is difficult to apply such charring rates to uncontrolled fires. When exposed to postflashover conditions, the charring rate of most common woods has been found to be in the range of $0.5-0.8\,\mathrm{mm\,min^{-1}}$ [8].

The various chemical and physical reactions that materials undergo as a result of exposure to heat are called *fire effects*. The visible or measurable manifestations of these effects are called *fire patterns* [9]. A more detailed discussion of the interpretation of fire patterns appears in **Fire: Dynamics and Pattern Production**.

References

[1] Benjamin, T. & Rumford, C. (1804). An enquiry concerning the nature of heat and the mode of its communication, *Philosophical Transactions of the Royal Society* **94** 77.

[2] Drysdale, D. (1985). *An Introduction to Fire Dynamics*, John Wiley & Sons, p. 13.

[3] Lentini, J. (2006). *Scientific Protocols for Fire Investigation*, CRC Press, p. 47.

[4] McMurry, J. & Fay, R.C. (1998). *Chemistry*, 2nd Edition, Upper Saddle River, Prentice Hall.

[5] Beyler, C. & Hirschler, M. (1995). Thermal decomposition of polymers, in *SFPE Handbook of Fire Protection Engineering*, National Fire Protection Association, Quincy, MA, pp. 99–119.

[6] Lentini, J. (1998). Differentiation of asphalt and smoke condensates from liquid petroleum distillates using GC/MS, *Journal of Forensic Sciences* **43**(1), 97–113.

[7] Stauffer, E. (2003). Concept of pyrolysis for fire debris analysts, *Science and Justice* **43**(1), 29–40.

[8] Babrauskas, V.(2005). Charring rate of wood as a tool for fire investigations, *Fire Safety Journal* **40**, 528–554.

[9] NFPA 921 (2008). *Guide for Fire and Explosion Investigations*, National Fire Protection Association, Quincy, MA, p. 12.

JOHN J. LENTINI

Fire: Dynamics and Pattern Production

Fire Dynamics and Fire Pattern Production

Although fire is a chemical reaction, the way in which a particular fuel burns is likely to depend more heavily on its physical state, the geometry of the fuel, and its environment than on its chemical nature. For example, a piece of dimensional timber may be exceedingly difficult to ignite, even with a blow torch. If that timber is sliced into kindling-sized sticks, it may be easily ignited with a match. A crumpled-up wad of newsprint will burn much more fiercely if the bottom of the wad is ignited than the top. A vehicle burning in an open field will be much more amenable to examination after a fire than a vehicle that has been consumed inside a garage. The term *fire dynamics* includes both the chemistry covered previously (*see* **Fire: Chemistry of**) and the engineering disciplines of heat transfer and fluid dynamics. It is this combination of chemistry and physics that makes the understanding of fires interesting and difficult. Because of the number of critical factors involved that may affect the behavior of fires, they are sometimes said to be subject to the "Mongolian butterfly" syndrome, wherein the small wind generated by a butterfly flapping its wings in Asia will affect the annual rainfall halfway around the world. The study of fire dynamics attempts to take some of the apparent chaos and make sense of it. In order for this to happen, a basic understanding of the concepts of ignition and flammability is required.

Ignition

All fires begin with ignition, and it is the goal of all fire investigations to determine the ignition source. It is not enough, however, to simply find "an" ignition source. The proposed ignition source must have sufficient energy to cause the proposed "first fuel ignited" to combine with atmospheric oxygen. For ignition to occur, the substance under consideration must first be capable of propagating self-sustained combustion. Ignition is defined as the process by which this propagation begins [1]. In the case of solids, ignition occurs when the heat generation rate in a given volume of material exceeds the heat dissipation rate, and the rate of reaction increases as temperatures rise further. The increasing heat causes chemical bonds to break, and the material decomposes into volatile substances. These volatile substances either ignite in the presence of a pilot, or they auto ignite. Piloted ignition temperatures for solids are generally in the range of $250-450^{\circ}$C, while auto ignition temperatures are generally in excess of 500°C. Either the pilot (a preexisting flame) or the hot environment provides the ignition energy for the gases and vapors evolved by heating.

How easily a substance ignites depends on its size and distribution. A layer of coal dust may burn relatively slowly, while that same dust dispersed into the air will ignite explosively. In general, gases, vapors, and dusts in air ignite more easily than solids.

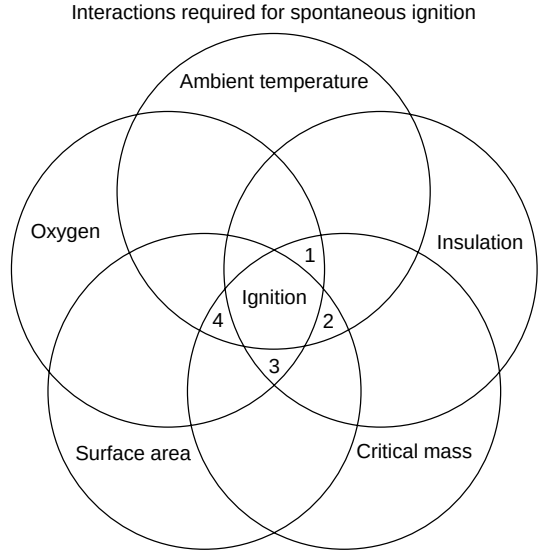


Figure 1 Interactions required for spontaneous ignition to occur

Solids may be either "thin" (<2 mm) or "thick." Generally, thin materials ignite more easily. For thick materials, the fuel's ability to conduct heat away from the surface (thermal conductivity), its heat capacity,



Figure 2 Effect of location on the burning of three identical fuel packages. Each photo was taken 17 ± 1 s after ignition. (a) A couch cushion burning in the center of the room. (b) An identical cushion burning against a wall. (c) An identical cushion burning in a corner [Photo courtesy of Jamie Novak, St. Paul, MN Fire Dept.]

and its density [2] are important. Poor heat conductors ignite at lower temperatures than good conductors. Thus, polyurethane foam will ignite sooner than a polyethylene trash can, which will ignite sooner than a wooden desk.

There are two special kinds of ignition that occur without the application of an external heat source. In chemical ignitions, simply combining a fuel and an oxidizer can result in a fire. A common example of a chemical ignition is the contamination of powdered swimming pool chlorine (calcium hypochlorite) with an organic material. The calcium hypochlorite provides the oxygen, and the organic material provides the fuel. The heat generated by the combustion of the organic contaminant can cause a decomposition reaction to occur in the pile of the remaining calcium hypochlorite. While not technically "combustion," an observer would be hard pressed to note many differences between this reaction evolving heat and light and "a fire."

Self-heating and spontaneous ignition occur as a result of exothermic (i.e., heat producing) reactions, whereby heat is liberated at a rate sufficient to raise the temperature of the material. The exothermic reactions in this case are faster than rusting but slower than fire. The typical spontaneous heating scenario involves the oxidation or polymerization of vegetable oils. Mineral oils such as motor oil, or even gasoline, even when saturated in a pile of rags, do not undergo spontaneous heating at ambient temperatures.

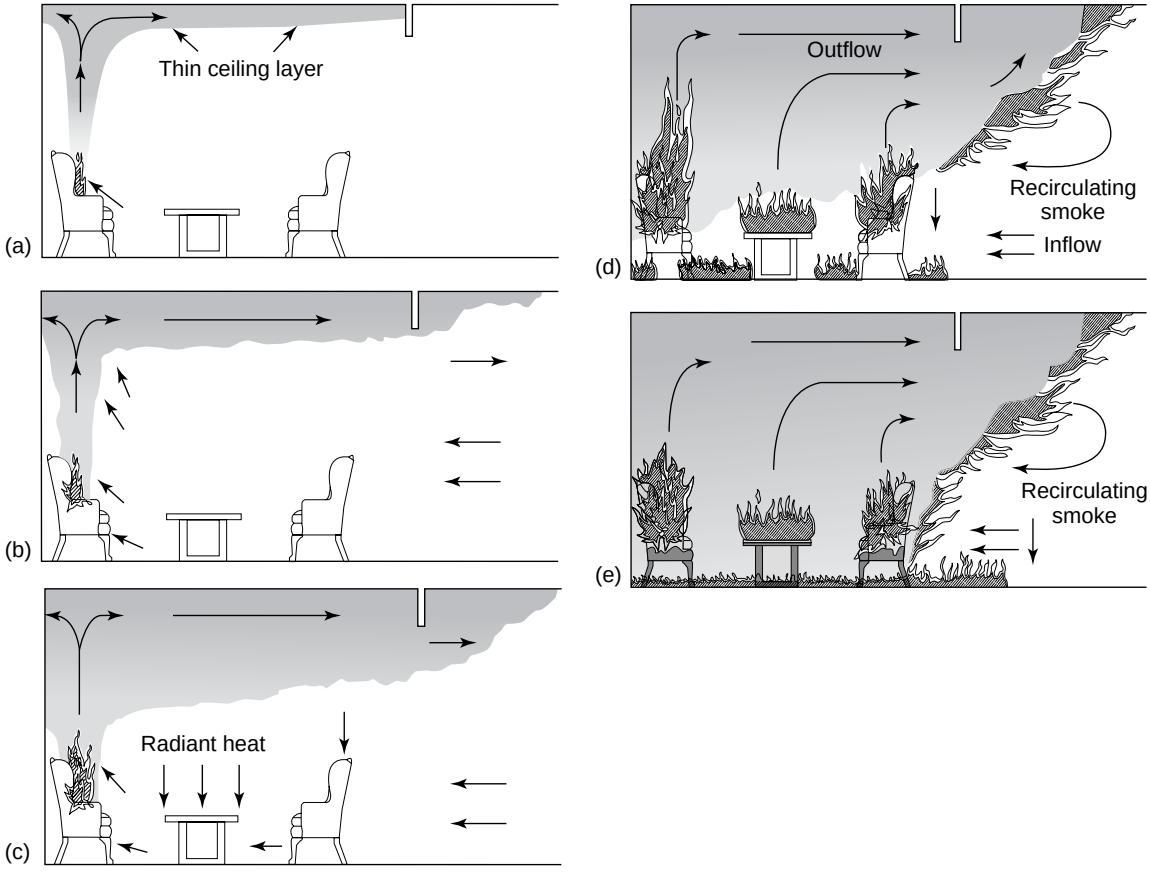


Figure 3 Progression of a typical compartment fire from the free burning stage through full room involvement. (a) Early compartment fire development. (b) Upper layer development in compartment fire. (c) Preflashover conditions in compartment fire. (d) Flashover conditions in compartment fire. (e) Postflashover or full room involvement in compartment fire. [Reproduced with permission from reference [1]. © NFPA, 2004.] This reprinted material is not the complete and official position of NFPA on the referenced subject, which is represented only by the standard in its entirety

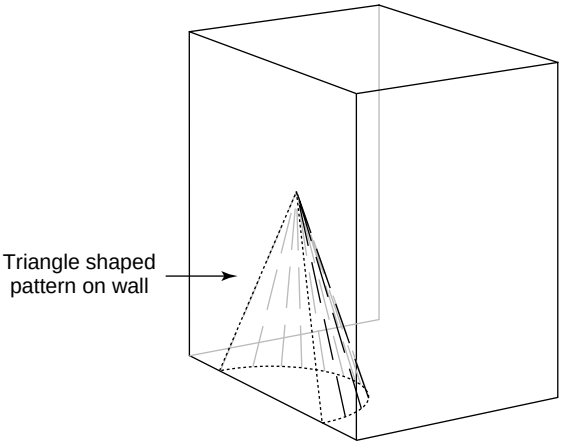


Figure 4 Schematic drawing showing the intersection of an early fire plume with a wall surface to produce an inverted cone pattern

Spontaneous ignition follows the same rules as piloted ignition or ignition by radiant heat, but the balance required to achieve ignition is much more delicate. Figure 1 shows the requirements for obtaining ignition without an external heat source. A balance must be maintained between providing the material under consideration with sufficient oxygen, but also sufficient insulation. The surface area of the material must be large enough to absorb large quantities of oxygen, but configured in such a way that the heat generated does not escape.

Although it is spontaneous, self-heating to ignition is usually a very slow process. This is because the heat is concentrated at the center of the mass of fuel, where there is usually insufficient oxygen for flaming combustion, so the first combustion is a smoldering combustion. Once the smoldering mass penetrates to the surface, flaming combustion can take place. Consequently, spontaneous ignition is frequently preceded by a long period wherein large quantities of smoke are produced.

The time frame in which spontaneous ignition occurs can be remarkably long. Solvents present in a liquid stain in a pile of cotton rags, for example, will use up much of the heat generated by the polymerization of drying oils within the stain in the process of evaporating. It may only be after all of the solvent has evaporated that the heat of the reaction can be absorbed by the cotton substrate. Spontaneous heating to ignition frequently takes place in a matter of hours, but may take as long as 4 days, or even longer.

Flames

A flame is a luminous zone of burning gases, vapors, and fine suspended matter where combustion is



Figure 5 An example of an inverted cone pattern

taking place. The kind of flame that fire investigators are generally concerned with is called a *diffusion flame*. In a diffusion flame, fuel gases or vapors and oxygen combine in a reaction zone because of differences in concentration. Substances move, or diffuse, from areas of higher concentration to areas of lower concentration. Thus, the wax vapors in a candle move away from the wick, where they are most concentrated, into the reaction zone where the vapors combine with oxygen from the air moving toward the flame. Michael Faraday's study of the physics and chemistry of a candle flame, presented in his Christmas Lectures to the Royal Academy, was one of the earliest studies of fire dynamics. Faraday [3] recognized the complexity of the simple candle flame when he wrote, "There is no more open door by which you can enter into the study of natural philosophy than by considering the physical phenomena of a candle. There is not a law under which any part of this universe is governed which does not come into play, and is not touched upon, in these phenomena."

Diffusion flames can be either laminar (orderly) or turbulent (chaotic). Any flame taller than 30 cm (about 1 ft) is likely to contain sufficient disorder to be considered turbulent [4].

Flame temperatures vary widely, depending on whether the flame is premixed or diffuse, laminar or turbulent, the amount of fuel involved and its net energy release. In everyday structure fires, flame temperatures range from about 600 to about 1200°C. The start of flashover, described later, is usually defined as a temperature in the upper layer of 500–600°C. Temperatures of 1300°C are not uncommon under postflashover conditions. At one time, high temperatures were believed to be the result of the use of accelerants. This is an appealing notion; but it is now known that the temperature of a well-ventilated gasoline fire is no higher than the temperature of a well-ventilated wood fire [5].

Flammability

The term *flammability* describes a number of different properties of a material, and many different tests have been designed to measure these properties. The properties of a fuel that may be of interest include the following:

- minimum ignition energy (measured in millijoules);
- flash point (measured in degrees centigrade);
- burning rate or mass loss rate (measured in grams per second);
- heat of combustion (measured in kilojoules per gram);
- heat release rate (measured in kilowatts or megawatts); and
- flame spread rating (measured in arbitrary units compared to the rate of flame spread on red oak or expressed as distance per unit time).

Depending on the fire in question, one or all of these measures of flammability might be of interest to the fire investigator.

In understanding the behavior of fires, the central question is whether a fuel proposed to have been ignited by a particular ignition source is capable of being ignited by that source, and capable of sustaining ignition once it does become ignited. Knowledge of the flammability properties of the

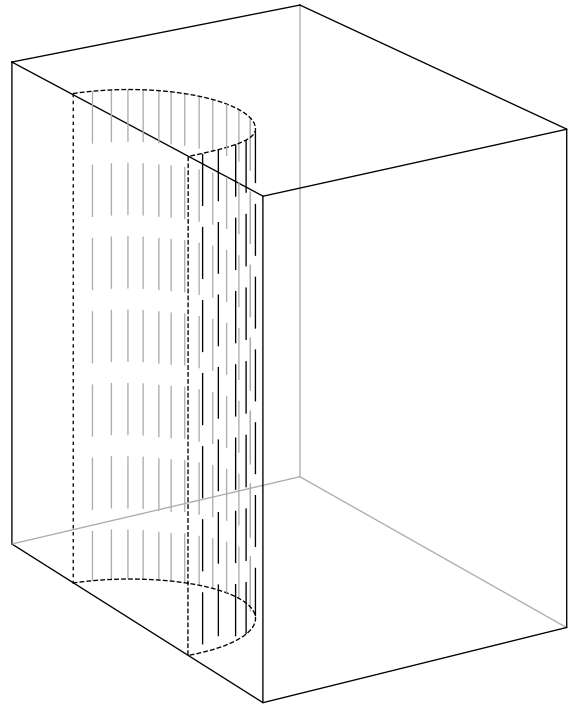


Figure 6 Schematic drawing showing the intersection of a fire plume with a wall surface to produce a columnar pattern

materials involved will often help to settle these questions, but in assessing the progress of a structure fire, the interaction between the burning fuel and the structure, itself, is often the most important interaction.

Heat Transfer

It is the transfer of energy in the form of heat that produces the most damage and leaves behind the artifacts that fire investigators interpret. Heat in fires is transferred by three mechanisms: conduction, convection, and radiation. Usually, heat transfer of all three types is taking place. "Conduction" is defined as heat transfer to another body or within a body by direct contact. Faster moving atoms or molecules in a body transfer heat by colliding with nearby molecules or atoms. An example would be the metal handle of a pot becoming hot because the pot is heated. "Convection" is defined as heat transfer by circulation within a fluid medium such as a gas or a liquid. Warming a fluid (be it liquid or gas) makes it less dense. An example is the warm air rising from just above a warm road surface. "Radiation" is defined as heat transfer by means of electromagnetic energy. An example is the heat from the sun. Just as the heat from the sun arrives at the earth after traveling through a vacuum, radiation in a fire requires no medium and travels at the



Figure 7 An example of a columnar pattern. This fire burned for less than 5 min. The fire did not have much time to interact with the ceiling, so only the beginnings of a V-shaped pattern are visible at the upper edges

speed of light. Infrared is an example of radiated heat.

With rare exceptions, the fire patterns observed in a structure fire are caused by convective or radiative heat transfer. At the beginning of a fire, almost all of the heat transfer is by means of conduction. As a hot gas layer develops (see compartment fires, below), radiation makes more of a contribution, and after flashover, radiation becomes the dominant means of heat transfer.

Compartment Fires

Most people have some familiarity with fire behavior because of their previous exposure to campfires, brush fires, and trash fires. In fire litigation cases, it is often necessary to help the trier of fact get beyond their "understanding" of the behavior of fire so that they understand the behavior of a fire "inside a structure." For decades, fire investigators, as a profession, failed to appreciate the significant differences between unconfined (free burning) fires and compartment fires, and made misinterpretations of what they observed. "Low burning" was thought to indicate that someone intervened to make the fire burn downward, and usually, the proposed mechanism for such downward burning was the addition of an ignitable liquid, or accelerant.

By looking at a simple example of a confined fire, we can begin to understand how interacting with the structure significantly affects the growth and spread of the fire. If we take a typical wastebasket fire with a heat release rate of 150 kW, and place it in the center of a large room, the flame height will be about 1.3 m (just over 4 ft). If the wastebasket is moved next to a wall, cool air can no longer enter the fire plume from the direction of the wall. The length of the mixing region, where flame and air come together, will be extended, and the flame will be taller. Even though we have removed half of the essential mixing air by moving the wastebasket next to the wall, it will not be twice as tall, but it will be taller. If we move the wastebasket into a corner, the flames will be taller still, because only 25% of the radius of the fire plume is open. Thus, it is possible for a fuel package in the corner of a room to appear much more heavily damaged than one might expect. In fact, experiments have been conducted wherein one end of a sofa was ignited, but the other end exhibited much more complete damage, because it was close to the corner. Figure 2 demonstrates the effect of the location of a fuel package with respect to walls. All three photos are of identical polyurethane chair cushions taken 17 s (± 1 s) after ignition.

While its location in a room can dramatically affect the flame height and burning rate of a fuel package, it is the interaction with the ceiling and the resulting production of a "hot gas layer" that produces a phenomenon that is foreign to most people's experience and central to the progress of most structure fires.

If an upholstered chair in a living room is ignited, say by a cigarette, and makes the transition to flaming combustion, the fire's behavior in the early stages will be no different than if the chair were outdoors. When the products of combustion reach the ceiling, however, things begin to change. At first, a thin ceiling layer develops. This layer consists of small particles, aerosols, and gases produced by the fire, i.e., the products of combustion. Some of these products are only partially burned and are subject to later ignition. These products rise because they are less dense than the surrounding air, and they begin to fill up the room from the top down. The particles and gases in the hot gas layer absorb energy from the

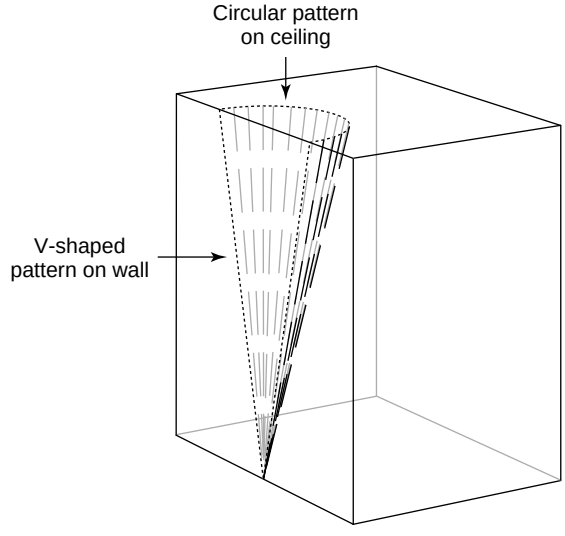


Figure 8 Schematic drawing showing the intersection of a fire plume with a wall surface to produce a V-pattern, and with the ceiling to produce a semicircular pattern

fire plume, and begin to reemit some of the energy. As the fire continues, the layer at the ceiling grows thicker as the burning chair produces more smoke and releases more energy. The hot gas layer then becomes a heat source in itself, emitting some of its energy in the form of radiation, which moves in all directions, including downward. Thus, even without any "help," this fire has the capacity to ignite objects below it. As the fire on the chair continues to grow, it acts as an "energy pump," moving energy into the hot gas layer. When the temperature of the hot gas layer reaches 500–600°C, "flashover" occurs. At this point, the radiant heat flux at the floor of the compartment rapidly rises to at least 20 kW m^{-2}. The transition from preflashover burning to postflashover burning has been described as changing from "a fire in a room" to "a room on fire." The room will continue to burn long after flashover occurs, but the fundamental nature of the fire changes. Prior to flashover, the fire grows as additional fuel is ignited. Once flashover occurs, however, every exposed combustible surface in the room is already on fire, so the only factor that can affect growth is the amount of ventilation available. Prior to flashover, a fire is said to be "fuel controlled." After flashover, the fire is said to be "ventilation controlled." The progression of a typical compartment fire is illustrated in Figure 3.



Figure 9 An example of a V-pattern. This fire was determined to have started at a trash can located at the bottom of the V

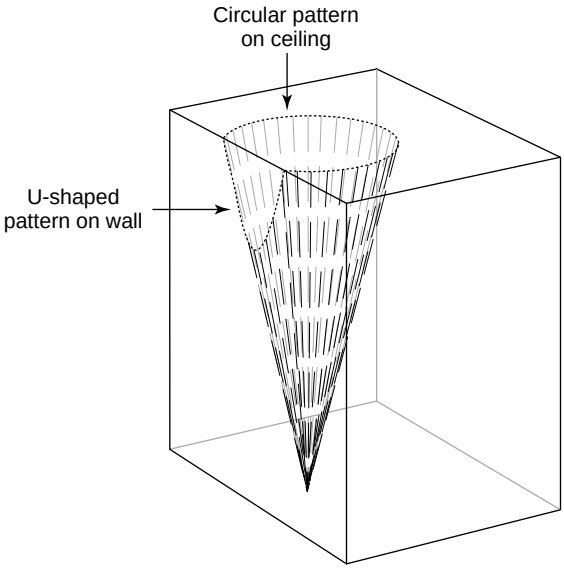


Figure 10 Schematic drawing showing the intersection of a fire plume with a wall surface to produce a U-pattern on the wall and a circular pattern on the ceiling

In the free burning stage, the primary means of fire spread is via convection, i.e., movement of the fluid (air and products of combustion) as a result of differences in buoyancy. After flashover, the primary means of fire growth and spread is radiation [6]. This radiation behaves exactly like visible light, and causes the production of fire patterns in straight lines. "Protection patterns" are formed when objects protect the surfaces underneath or behind them from exposure to radiant heat. Protection patterns are essentially shadows.

One thing that usually happens when flashover occurs is that the windows break. Because window glass often fails when radiant heat fluxes reach 20 kW m^{-2}, breakage of the windows is often considered the event that defines when flashover occurs [7]. The breakage is due to the increased thermal stresses caused by the vast difference in temperature between the exposed glass and unexposed glass under the edge of the window frame.

One of the first tasks of a fire investigator investigating a structure fire is to determine those compartments where flashover occurred and those compartments where it did not. Because postflashover burning can produce patterns that obscure or obliterate patterns that were created prior to flashover, careful interpretation is required in postflashover compartments.

Pattern Development

The observations of "fire patterns" on the surface of a fire scene are the major tools that a fire investigator uses in determining where a fire started and how it moved (*see* **Fire: Scene Investigation**). A "fire effect" is an observable or measurable change in or on a material as a result of a fire. A fire pattern is an identifiable shape formed by a fire effect or group of fire effects. Fire effects are the underlying data that are used by the investigator to identify fire patterns. The recognition, identification, and proper analysis of fire patterns depend on the understanding of the dynamics of fire development and heat and flame spread. This analysis includes an understanding of the way that the three modes of heat transfer (conduction, convection, and radiation) produce the fire effects, and the nature of flame, heat, and smoke movement within a structure.

The patterns that are of most interest to fire investigators are those recorded on two-dimensional surfaces such as walls, ceilings, and floors. Patterns may either be caused by fire intensity or by fire spread. The patterns seen by an investigator represent the entire history of the fire. Each time another fuel package is ignited or the ventilation to the fire changes, the rate of energy production and the heat distribution change. Any burning item can produce a plume and thus a fire pattern. Determining which pattern was produced at the point of origin by the first material ignited is the goal of the fire investigator, but success often becomes more elusive as the size and duration of the fire increases. Patterns can be generated by fire plumes, hot gas layers, and the movement of fires from one compartment to another or from one level of the building to another.

Mature fire plumes are cones, and where those cones intersect ceilings and walls, characteristic patterns are generated. Early in a fire's development, the cone is upright, as shown in Figures 4 and 5. In the midpoint of a fire's development, the plume may assume a columnar shape. Figures 6 and 7 illustrate the development of a columnar pattern. Such patterns are seldom seen and require that the fire be extinguished after the cone stage, but before the formation of an inverted cone.

If the cone of a fire plume has its point at the base of a wall, then a V-pattern will be produced on the wall. See Figures 8 and 9. If the base of the fire is some distance away from the wall, then the wall will intersect the fire plume higher up, and a U-shaped pattern will be produced, as shown in Figures 10 and 11. A semicircular pattern will be produced on the ceiling if the fire is close to the wall, as shown in Figure 12, and a circular pattern will be the result if a fire is in the middle of the room. In a typical fire scene, many or all of these patterns may be observed.

Recognizing fire patterns and how they were produced is a requisite skill if a fire investigator is to accurately determine the origin of the fire, which is the most important activity that must be conducted prior to determining the cause.



Figure 11 An example of a U-pattern. This fire was started at the wall behind the door in the foreground. There was a V-shaped pattern behind the door. The U-pattern was produced farther away from the origin



Figure 12 An example of a semicircular pattern on a ceiling. The semicircular pattern was helpful in determining that the fire started inside the wall, at the location of a three-way switch. One of the switch contacts was found to have melted, and there was arcing on the branch circuit conductor in the wall cavity

References

[1] NFPA 921 (2004). *Guide to Fire and Explosion Investigations*, National Fire Protection Association, p. 12.

[2] Quintiere, J. (1998). *Principles of Fire Behavior*, Delmar Publishers, pp. 69–81.

[3] Faraday, M. (2002). A course of six lectures delivered before a Juvenile Auditory the Royal Institution of Great Britain during the Christmas Holidays of 1860–61, in *The Chemical History of a Candle*, Dover Publications, Mineola.

[4] Quintiere, J. (1998). *Principles of Fire Behavior*, Delmar Publishers, p. 26.

[5] NFPA 921 (2004). *Guide to Fire and Explosion Investigations*, National Fire Protection Association, p. 37.

[6] NFPA 921 (2004). *Guide to Fire and Explosion Investigations*, National Fire Protection Association, p. 18.

[7] Babrauskas, V. (2003). *Ignition Handbook*, Fire Science Publishers, p. 533.

JOHN J. LENTINI

Fire: Scene Investigation

This article describes the basic steps to be followed in carrying out a fire investigation. Each fire investigation is different from every other, if for no other reason than that fires occur at different places and at different times. There are sufficient similarities among fires; however, that a standard approach to the process is possible. Not every step will be necessary or practical in every investigation, and there are certainly situations that will not be addressed here that will require some initiative and creativity on the part of the investigator. Many of the routine steps will be possible only if the investigator is the "first responder", or one who is on the scene before it has been significantly altered. Many investigations do not even start until other investigators have had a 6- or 12-month head start, and have reached a conclusion that, for example, a particular product failure caused the fire. An investigator retained by the product manufacturer will thus be relying on the documentation produced by the earlier responders.

Procedures will also vary according to the problem being addressed. In many cases, the cause of the fire is well known, but the investigator may be tasked with addressing a particular question about a particular system.

Recognize the Need

Before covering the subject of *how* a fire investigation is conducted, it is first necessary to understand *why* a fire investigation is conducted. This is the first step of the scientific approach, described in

National Fire Protection Association (NFPA) 921 as "Recognize the Need" [1]. What is it that the entity calling for the investigation needs? The main task of most fire investigators is to determine *where* the fire started and *how* it started, but the word "cause" often has many deeper meanings, depending on the context. In addition to the cause of a fire, one might want to understand the cause of a fire death. There might be a very simple cause determination, such as an obvious cooking fire, but the next question arises "Why did the victim fail to escape"? Was there something wrong with the victim? Was he or she incapacitated? Was there something wrong with the building? Were the exits blocked? Did the smoke alarm function properly, and if not, why not? Did the fire spread too quickly, and if so, why? Fire investigators are frequently asked to determine why a fire spread in the way that it did. This could be due to the nature of the first fuel ignited, the interior finish, the nature of the building ventilation system, the failure of a fire protection system, such as a sprinkler or vent damper or automatic door closure, or because someone propped open the door to the exit stairway.

In many fires, the cause of the fire (the circumstances that brought the fuel and the ignition source together) is reasonably well known because there are reliable eyewitnesses. An inspection of the scene may well verify that, but it is also necessary to understand *why* it happened. It is important to know whether there was a defect in the design or manufacture of an item or whether it was broken or a safety device had been defeated. Many times someone has defeated a safety device. Even after the cause of the ignition is known, it may be, however, necessary to understand why the fire suppression system failed to extinguish the fire.

Even if there is no criminal case, the person who wired around the safety device and caused the loss may be civilly liable to the owner of the building. The owner of the building may be civilly liable to the tenant next door, if he failed to maintain the fire suppression system.

In the public sector, the first goal is usually to determine whether the fire was accidental or deliberate (although this term is eschewed by many observers who say that the motive cannot be ascertained from the science, which is probably true in many cases). That is sometimes the limited assignment of the first investigator sent out by an insurance carrier, but usually an insurance carrier is looking for someone, anyone, who might be held responsible for the fire, other than its named insured.

Presumption of Accidental Cause

The most important thing a fire investigator can do is to keep an open mind. It must be remembered that fires are common, fires can be very difficult to investigate, and *more than 80% of all fires are accidental* [2]. It is this last fact that seems to elude many investigators. The fire scene should always be treated as a crime scene in the first instance. As long as the potential exists for the fire to be determined to be incendiary, there should be site security in place, and mandated procedures designed to protect the integrity of any criminal prosecution. In the United States, the controlling legal authority on a fire investigator's right of entry is the Supreme Court's interpretation of the Fourth Amendment in two Michigan cases, *Michigan v. Tyler* and *Michigan v. Clifford* [3, 4]. In *Tyler*, the court held that an "expectation of privacy" exists even after a building burns; hence, a warrant is required except when there are "exigent circumstances". In *Clifford*, the court held that the need for a warrant existed even in cases of "administrative" searches, i.e., those searches that start out being conducted for purposes other than the collection of evidence in a criminal investigation. Most entries are made either under "exigent circumstances" or with the permission of the owner. When the owner is unavailable to give such permission, it is easier to obtain a warrant than to later explain why no warrant was obtained.

Planning the Investigation

When the fire investigation assignment is received, getting sufficient information will allow the formulation of a reasonable plan. Learn the basics regarding when the fire occurred: which fire department responded; who owns and leases the building; what the building was used for; and what other parties may have been affected by the fire. Private sector investigators may need to coordinate their efforts with both public sector investigators and private fire investigators retained by other parties. Learn to what extent suppression and overhaul may have altered the scene. The investigator may be able to decide whether it is necessary to recruit additional talent or obtain heavy

equipment to assist with the investigation. If there is a credible suspicion of an electrical failure, it is a good practice to bring a forensic electrical engineer to the site. If there is a collapse of a steel truss roof over a 50 000 square foot retail store, it will probably be necessary to rent a crane and hire a cutting and welding contractor.

It is useful to learn as early as possible whether there were any eyewitnesses to the fire. This can save an immense amount of work. Witnesses may have even photographed or videotaped the fire in progress. If a building has a security system, there may be records of alarm activations or possibly video tapes from security cameras.

For most occupancies, the use of the structure is apparent, and the influence of activities of the occupants or the kind of fuels present on the spread of the fire will be simple enough to determine. For industrial occupancies, it is usually necessary for the investigator to become familiar with the peculiarities of the processes taking place by consulting with the plant manager or someone knowledgeable about the systems. The fire history of the facility is also something that should be examined.

Prior to visiting the scene, it is necessary to ensure that one has permission to do so. Authority to enter and investigate most often comes in the form of permission from the owner, but may also take the form of a search warrant or an order of the court. Having made as many preparations as practical, the site visit is the next step.

The Initial Survey

Fire scenes are dangerous places. Safety hazards include the obvious ones, such as structural instability, energized electrical circuits, and leaking fuel gases, but the investigator also needs to consider the less obvious hazards, such as fire gases (CO and HCN), toxic hazards such as asbestos, and biohazards. If conditions exist that prevent the inspection from being conducted safely, steps must be taken to alleviate those conditions before the investigation can proceed.

A brief “walk through” is considered useful by some investigators, and is indispensable in some scenes. For a single house, however, many start with the exterior and then move to the interior, starting with the area of least damage and then moving to the area of greatest damage. Except in the most obvious of cases, it is best not to use the walk through to formulate hypotheses. The best approach is to keep an open mind for as long as possible. Do not hasten the process. Although the investigator is obviously making observations at this point, one should avoid drawing inferences from those observations. Ideally, the walk through will help the investigator formulate a general plan for the conduct of the site inspection.

Documentation

The site condition should be documented before the investigator makes any changes (or any further changes). Documentation can take many forms, but the best form is still photography. If an observation is made, it should be documented. With the advent of digital photography, there is no longer any reason not to take a photograph of every pertinent artifact. Digital cameras, or film cameras with time legends, also serve to document the time spent at a scene, or at least the time between the first and the last photograph.

Other forms of documentation are essential as well. Written notes will document those observations not well suited to photography, and video documentation helps to orient photographs.

The documentation prior to disturbance of any evidence takes time, and it is during that time that the investigator can make the observations necessary to help decide how to reconstruct the fire scene. A preprinted sheet for documenting common building characteristics, such as the one shown in Figure 1, helps this investigator to avoid missing important details.

Another critical piece of documentation, required on all but the simplest investigations, is the sketch. For most buildings, the minimum sketch should be a floor plan. Sometimes, particularly in commercial or industrial buildings, an “as built” drawing of the building can be obtained from the building owner/operator (assuming that the fire did not destroy such documents). Most commercial occupancies have fire escape plans framed on the wall in several places. These are usually sufficient, at least as the starting point for a sketch. A floor plan is necessary because the people reading an investigator’s report (or listening to his testimony) are not going to be familiar with

FIRE SCENE FIELD NOTES

JOB #_____ Policy or Claim #_____ Date of Loss_____Time _____
Assignment Rcvd_____
Insured:_____ Date of Investigation_____
Address:_____ Building Last Occupied_____

CONSTRUCTION Approx. Dimensions_____Square Feet_____# Floors_____# Bedrooms___# Baths_____

Orientation Front door faces [ ]N [ ]E [ ]S [ ]W

Exterior Finish	Interior Finish			
	Floors	Subfloor	Ceilings	Walls
[ ]Frame	[ ]Carpet	[ ]Plywood	[ ]Sheetrock	[ ]Sheetrock
[ ]Metal/Plastic Siding	[ ]Tile/Linoleum	[ ]Planks	[ ]Plaster/Lath	[ ]Plaster/Lath
[ ]Brick Veneer	[ ]Hardwood	[ ]Particle Board [ ]Tile	[ ]Panel	[ ]Panel
[ ]Stone Veneer	[ ]Plywood	[ ]Slab	[ ]Tile	[ ]Ply Panel
[ ]Brick	[ ]Particle Board	[ ]_____	[ ]_____	[ ]_____
[ ]Mobile Home	[ ]_____			
[ ]Other _____				

Other Construction Data:_____

HVAC: Make of Heater:_____

[ ]Central	[ ]Natural Gas	[ ]LP Gas	[ ]Other
[ ]Electric	[ ]Forced Air	[ ]Forced Air	# Units_____
[ ]Forced Air	[ ]Floor Furnace	[ ]Floor Furnace	[ ]Wood Stove
[ ]Baseboard	[ ]Space Heaters	[ ]Space Heaters	[ ]Fireplace
[ ]Ceiling	[ ]Vented	[ ]Aboveground Tank	[ ]Coal
[ ]Wall Mounted	[ ]Unvented	[ ]Submerged Tank	[ ]Kerosene
[ ]Space Heaters	[ ]Open Line Flow_____	% Full___ ___ PSI	[ ]_____
[ ]Heat Pump		[ ]Open Line Flow_____	

Other Heating System Data:_____

Air Conditioning:[ ]Central [ ]Room Units Make:_____

DETECTION/SUPPRESSION Smoke Detectors:_____Fire Extinguishers:_____Sprinklers:_____

SECURITY
Days Since Fire_____
Doors Open_____
Windows Open_____
Insureds Home?_____
of Occupants_____
Smokers?_____
Melted Metals [ ]Fe [ ] Cu [ ]Al
_____
_____
_____

Appliances: (type, brand)
Refrigerator_____
Range_____
Oven_____
Dishwasher_____
Microwave_____
Freezer_____
Washer_____
Dryer_____
Water Heater_____
Compactor_____

Furniture (#)
_____Beds
_____Dressers
_____Couches
_____Chairs
_____Dining
_____
_____
Clothing:
[ ]Hangers
[ ]Cloth

INVENTORY

[ ]Pots & Pans	[ ]Food	[ ]Televisions _____	Other electronics_____
[ ]Dishes	[ ]Guns	[ ]Computer	_____
[ ]Silverware	[ ]Photographs	[ ]Pets	_____

Explosions: [ ]None in Evidence
[ ]Before Fire
[ ]During Fire
[ ]Reported [ ]Before [ ]During

Weather: _____
Lightning_____
Wind_____

Fire Suppression:[ ] effective [ ] ineffective [ ] not attempted Response time _____Control time_____

Department(s) _____ Dept. Contact:_____
_____

ATS 851A, 08/04

Figure 1 Checklist describing building conditions

the building. A floor plan is indispensable for getting other people oriented so that they can understand what the photographs show, and what they might mean.

The level of detail on the sketch is up to the investigator, and will be dictated by the needs of the investigation. In many cases, a simple plan view showing rooms and doorways will suffice. If ventilation or escape is an issue, then certainly the location of the windows should be included. If fire modeling is contemplated, then it is necessary to draw not only the doors and windows but also the height of each opening, the ceiling height, the location of each fuel package and the composition and thickness of all interior finishes must be recorded. The best time to gather this information is while the building is being sketched. Getting this information later may not be possible.

The sketch is as good a place as any to orient the building to the earth. A compass is an inexpensive device that should be in every investigator's camera case. "Left" and "right" can be very confusing descriptions, as it is then necessary to say where one is standing and which direction one is facing. Compass directions are much more consistent and easy to understand. An approximate scale is also useful. Such a scale might look like this: ~10 ft. Writing out "1 in. equals 10 ft" is only useful if the sketch is not enlarged or reduced.

Reconstruction

During the process of sketching, documenting, and becoming thoroughly familiar with the fire scene, the investigator has had a chance to make observations about the fire and to formulate a plan for reconstruction. This almost always involves debris removal and may require several hours or days of strenuous physical activity to remove the debris that covers the artifacts that will allow a determination of origin and cause. Although there is heavy equipment that can accomplish this debris removal, starting with heavy equipment will almost always destroy more evidence than it uncovers. Too many investigations have been compromised by bobcats and backhoes used by investigators looking for "pour patterns" or spalling. The causative agent frequently lands in a pile of debris in the parking lot, with no possible way to determine where it was located before it was run over by the heavy equipment. The problem with heavy equipment, efficient though it may be, is that instead of clearing out a doorway or a hallway, it takes down the walls on either side.

Sometime after the burned roofing material is carried outside, interesting artifacts will begin to be revealed. It is necessary to stop periodically and document this potential evidence as it is found. It is frequently necessary to move things out of the way so that debris clearing can continue, then move those things back to their original location to understand their role, if any, in the progress of the fire. In a well-executed debris removal operation, the debris is moved *once*.

Once the debris has been moved out of the way, items should be placed back into their original locations, if possible. Tables and chairs frequently leave "footprints", protected areas on floors that allow for the furniture to be exactly repositioned. Use caution with freestanding items such as tables, however, since it is possible to get them repositioned exactly backwards.

Once the reconstruction is as complete as it is going to get, any portions of the scene that have been changed should be rephotographed, paying particular attention to burn patterns that indicate movement or intensity of the fire, fuel packages, and ignition sources.

A common mistake in the documentation of the scene is the failure to take overall shots of the area where close-ups are going to be taken. Photographs of burn patterns or of small objects that might be ignition sources are just not as useful if there is no photograph that places the subject of the close-up into its context.

Some of the most important artifacts are those that allow the collection of *sequential data*. The fire scene "records" each movement of the fire from one place to another, but often these patterns do not help us learn *when* a particular event occurred, or in which order two or more events occurred. Patterns that give sequential data are the most instructive. Sometimes, it is the orientation of burn patterns on an object that allows inferences to be drawn regarding the object's position when the pattern was created. Fallen objects such as pieces of the ceiling may exhibit smoke or heat damage on the top or bottom surface. It is important to examine this "debris" as it is being removed. Sometimes, the task of moving debris is so large that focus on details such as this can be lost.

The evidence is now exposed and has been documented. The fire investigator is presented with what is likely to be all of the physical evidence there is to see. It is hoped that the work that has been done will allow for the development of a credible hypothesis as to where the fire started, and what may have caused it.

Inventory

In a residential or mercantile occupancy, the inventory is physical evidence that frequently requires documentation. The investigator should at the very least determine that the inventory is consistent with the reported type and level of occupancy. Are there sufficient beds, linens, toiletries, and food for the number of persons reported to be living in a residence? (Do not open the refrigerator to check the food supply until you are ready to leave, especially in the summer.) Private fire investigators are often asked to include an inventory section in their report in order to verify the accuracy of an insurance claim. Assuming that it is the kind of item not likely to have been consumed in the fire, the absence of some items may indicate a fraudulent claim. In fact, it may indicate that those items were removed prior to the fire on the basis of the homeowner's knowledge that the fire was going to occur. Prior removal of contents, or substitution of contents prior to the fire, can be used as evidence of prior knowledge and may overcome a presumption that a fire was accidental. An accident is by definition unexpected.

This author has investigated many fires where the homeowner could not part with his beloved guns or his television, so they were removed before the fire was set. Then, in attempt to "double dip" the missing items were claimed on the insurance inventory. The homeowner may have even been able to document the purchase of items removed prior to the fire. Sentimental or irreplaceable items, such as photographs, are another class of items that are frequently removed prior to an intentionally set fire. Since these items have little or no economic value, the claim that some burglar stole them and then set fire to the house would be curious.

Substituted inventory is an even better indicator of what happened and who is responsible. In one fire investigated by the author, the homeowner first put his valuables in storage, then went to the town dump and obtained what he hoped would be sufficient junk furniture and appliances to support a claim after those materials had been burned to powder. He made the mistake of using too much gasoline, however, and the fire went out for lack of air. There was a trail of broken picture tube glass leading from the driveway into the living room, where a television set with a broken picture tube sat, with not a scorch mark on it.

In one of the author's first fraud investigations, the house was reduced to ashes, but the homeowner had stashed his coin collection in the trunk of his car, where it was later spotted by an alert insurance adjuster. If there are outbuildings on the property, the investigator should document their contents, to make sure that undamaged property is not later claimed as having been destroyed in the fire.

In cases such as these, it is necessary for the arsonists to unwittingly cooperate in their own downfall, if this kind of approach is to work. There may be a legitimate excuse for the photos and other mementoes being found in the chicken house. In such a case, the homeowners must deny removing anything prior to the fire.

Avoiding Spoliation

It often happens that an investigator reaches a point in the scene investigation that it becomes apparent that a particular device malfunctioned and caused the fire. This may be a light switch or a ceiling fan or a cooking or heating appliance, a computer, or any one of the thousands of manufactured products. It may be that a contractor performing a service, such as refinishing the floor or reroofing a commercial building, was on site shortly before the fire, and the evidence seems to point to some careless act on the part of one of the contractor's employees. When this occurs, it is the investigator's job to stop any further activity that might prejudice the rights of an entity that may soon become a party to litigation. The investigation can resume at a later time. If the property is insured, the insurance carrier will look to whoever caused the fire for compensation. That party, in turn, will likely insist on the opportunity to view the evidence. Failure to accommodate potential defendants may result in sanctions against the plaintiff seeking damages, up to and including dismissal of the lawsuit. In noncriminal fire investigations, the rights of all parties to have access to the evidence has been embodied since

the 1970s in ASTM E 860, *Standard Practice for Examining and Preparing Items that are or may Become Involved in Criminal or Civil Litigation* [5].

Spoliation has so far not been an issue in criminal cases, but it is likely to come up in the future, so the concept should not be dismissed out of hand by arson investigators. Spoliation is defined as the "loss, destruction, or material alteration of an object or document that is evidence or potential evidence in a legal proceeding by one who has the responsibility for its preservation" [6]. To avoid spoliation, potential defendants must be put "on notice" and be told when the investigation will proceed, and allowed to send a representative to participate in the continuation of the investigation.

Spoliation is a relatively new concept for many investigators. In the past, an investigator would go to the site, determine the cause, and if that cause was an appliance, it would be collected and taken to an electrical engineer for further evaluation. If the engineer found evidence that the appliance was responsible for the fire, the insurance carrier would make a claim against the manufacturer, and the case would proceed. Such behavior by a fire investigator in the twenty first century, however, will not only prevent the investigator's client from collecting from the manufacturer the defective product, but it is also likely to get the investigator sued, even if he has correctly identified the cause of the fire and it is, in fact, defective.

Defendants no longer simply accept the first determination, and the courts have supported their position. Not only is it necessary to let the manufacturer see its defective product but it is also often necessary to let them see it *in place*, to let them have their own investigator confirm that the product was, in fact, at the origin of the fire, and to allow that investigator to independently rule out other potential sources of ignition. A manufacturer of clothes dryers, defending itself in a product liability action, will likely be successful if the washing machine was not at least preserved for inspection by the manufacturer's chosen investigator.

Claims of spoliation can be taken to apparently ridiculous extremes, and realizing this, NFPA's Technical Committee on Fire Investigations has undertaken to offer guidance not only on what activities constitute spoliation but also on those activities that should *not* be considered spoliation. It is usually necessary to conduct significant debris removal prior to uncovering the cause of the fire. Debris removal to uncover the area of origin should not be considered spoliation nor should removal of the offending device from the scene be considered spoliation, if such removal is necessary to either protect the device, or to identify its manufacturer. Further, it may be possible to eliminate an appliance by a simple nondestructive examination off-site. Note that any device once suspected but later eliminated needs to be preserved so that it can be examined by other parties.

It is scenarios such as this one that can influence criminal cases as well. It is no longer acceptable for a public sector investigator to "eliminate" a reasonable potential source of accidental ignition, and then allow it to be destroyed. A defendant may then claim that he has been prejudiced by the destruction of potentially exculpatory evidence. Such a claim will be especially effective if the only "evidence" of an incendiary cause is the investigator's inability to find any source of accidental ignition in the area identified as the origin.

Claims of spoliation against public agencies are not common, but public sector investigators should not dismiss the idea of preserving evidence for the civil investigation. Running over a fire scene with a bobcat or a backhoe to look for "pour patterns" or spalling, particularly if no such "evidence" is uncovered, makes it next to impossible for the real cause of the fire to be discovered.

Evidence Collection and Preservation

There are basically two types of evidence that fire investigators collect. The most common is samples of flooring, furniture, and other materials suspected of containing ignitable liquid residues (ILR). For a discussion of ILR analysis, *see* **Fire Debris: Laboratory Analysis of**. The other category of evidence is ignition sources that are suspected of having caused the fire, or are collected for the purpose of proving that they did *not* cause the fire. As with other facets of the investigation, evidence collection will be more successful if the fire investigator has a plan. If there is more than one investigator on the scene, all should agree to the plan for evidence collection. The standard that describes appropriate procedures for evidence packaging and labeling is ASTM E1459, *Standard Practice for Evidence Labeling and Related Documentation* [7]. Other standards applicable to fire investigation are described in **Fire Investigator: Standardization, Accreditation, and Certification**.

With respect to the samples collected for ignitable liquid residue analysis, if possible, the samples least likely to contain residues should be collected first. This will minimize the risk of cross-contamination. If comparison samples are to be collected (and it is strongly recommended that they be collected), they should be collected first. Select materials that are identical or nearly identical to the materials suspected of containing ignitable liquid residue. Label the evidence containers with the following information: a unique case number or file number, identifying the fire scene; the sample's serial number; a description of the substrate material; the location of the sample; the date collected; and the initials of the investigator collecting the sample. The labeled containers should be placed at the location where the sample will be collected, and photographed in place before the sample is placed in the can. A second photograph should be taken showing the sample in the can next to the former location of the sample. The location of the sample is the single most important attribute of the sample, so it is important that this information be thoroughly documented. When collecting samples for ignitable liquid residue analysis, the investigator should wear disposable latex gloves, and change gloves between each sample. The gloves should be left behind at the sample location and not placed in the sample container.

There are several choices available for sample containers. Containers should be free of contamination and vapor-tight. Glass has the advantage of transparency and resistance to corrosion, but it breaks. Metal cans do not break, but in the time between the site inspection and any trial, they are likely to corrode. Corrosion can be avoided or delayed by the use of polyester-lined paint cans. The coating is opaque gray or opaque tan, and is the kind used for water-based paints. Such a coating will not influence the laboratory analysis, but it is a good practice to save one can from each batch purchased, in case it becomes necessary to prove this.

Evidence bags made for the purpose are suitable containers, but their only real advantage is their light weight and small volume, which makes it possible for the investigator to carry a large number of them. Major disadvantages include difficulty in sealing, poor containment of wet samples and susceptibility to tearing. Many laboratories find it necessary to repackage evidence delivered in bags. When submitting evidence in bags, a negative control (empty bag) is required.

The location of any sample collected should be noted on the investigator's sketch. An evidence transmittal form (see Figure 2) should be filled out as soon as practical after the samples are collected. Samples should not be held longer than absolutely necessary prior to submission to the laboratory. If hand delivery is likely to create a significant delay, then the use of a common carrier is preferable.

In the context of a joint inspection, all parties should have the opportunity to request that certain items of evidence be preserved for later examination. The person responsible for collection of potential ignition sources (usually the investigator retained by the property owner or their insurance carrier) should photograph each item in place, and then label the item with a label or tag that contains all of the information listed above for ignitable liquid residue samples. Once all interested parties have had a chance to view the items in place, they should be carefully transported to a testing facility. Wrapping the items with shrink wrap, available at low cost at trailer rental facilities and other places, will prevent the loss of loose parts of appliances.

Origin Determination

The determination of the origin of a fire is a fire investigator's most important and difficult task. If one does not find the correct origin, it is practically impossible to find the correct cause. Conversely, if one is able to correctly determine the origin, it is frequently not difficult to find the cause. NFPA 921 provides a long list of factors that need to be "considered" when trying to locate the origin, but the guidance is necessarily incomplete. What do you do after you consider the effects of ventilation? Other texts offer advice on burn pattern interpretation, for example, to examine the beveled edge of a hole to determine which way a fire came through a wall or a floor. What those sketches show is the direction the fire was moving when it *last* penetrated the floor or wall, which is sometimes all one can hope for.

Some things to consider when looking for the origin are the artifacts produced by the confinement

APPLIED TECHNICAL SERVICES, INCORPORATED

1190 Atlanta Industrial Drive Marietta, GA 30066 (770) 423-1400 Fax (770) 424-6415 e-mail fire@atslab.com

EVIDENCE TRANSMITTAL

Date: _____ Verbal Report To: _____

Phone #: _____

Submitted By: _____ Written Report To: _____

_____ _____

_____ _____

_____ _____

File #: _____ Invoice To: _____

Insured: _____ _____

Claim Number: _____ _____

Date of Loss: _____ _____

ATS Reference # _____ _____

Description of Evidence: Container, Size Type of Material, Condition of Material (burned or unburned) — Location Collected

1. _____ _____
2. _____ _____
3. _____ _____
4. _____ _____
5. _____ _____
6. _____ _____

Special Instructions: _____

Chain of Evidence
(Signature Required)

From: _____ To: _____ Date: _____ Time: _____

From: _____ To: _____ Date: _____ Time: _____

From: _____ To: _____ Date: _____ Time: _____

From: _____ To: _____ Date: _____ Time: _____

ATS 800 (10/99)

Professional Engineers
Design • Consulting • Testing and Inspection
Members in AAFS, ACS, ASM, ASME, ASNT, ASQC, ASTM, AWS, FSCT, IAAI, NACE, NCSL, NFPA, SAFS
GEORGIA SOCIETY OF PROFESSIONAL ENGINEERS, NATIONAL SOCIETY OF PROFESSIONAL ENGINEERS

Figure 2 Evidence transmittal used for submission of fire debris samples to a private laboratory

of a hot gas layer. If one observes a smoke horizon or a heat horizon, one can conclude that, in all likelihood, that artifact was recorded on the wall *before* the ceiling in that room failed. This can be very useful, at least when it comes to placing the origin on the correct level of a structure. The absence of evidence of confinement can be almost as useful. Fires that originate in the attic frequently cause ventilation of the roof before the ceiling fails, so there is no confinement. Lightning fires frequently ignite roofs and attics, resulting in characteristic damage.

Structural framing also causes fire to penetrate walls and floors in predictable ways, and the artifacts of a surface penetration can help identify which side of the surface was involved first. If a fire attacks a floor from above, the fire is not going to be influenced by the joists under the floor. If it attacks the floor from below, then frequently, the holes in the floor will be rectangular, with the joists directing the flow of hot gases. The same is true for walls. Straight up and down holes in the drywall, with edges coincident with the wall studs, indicate that the fire came through from the other side. Figures 3 and 4 show such a pattern, and illustrate how it was created. Frequently, such determinations are "no-brainers", but it is easy to be misled by other factors, such as the "lowest burn" or "heaviest burn". That is when one needs to "consider the effects of ventilation".

At other times, the smoke and heat horizons may be the only evidence remaining. In the case where a structure collapses and burns in the basement, one does not usually expect to see a smoke or a heat horizon on the foundation walls unless the fire started in the basement. If the fire started upstairs and burned its way into the basement, there would have been a hole where the fire penetrated downward, thus preventing the formation of a hot gas layer there.

If there are humans, particularly dead humans, in the structure, their activities need to be considered. The individual in Figure 5 caught his shirt on fire in the kitchen, but did not stay there. His presence in the hallway in the area of deepest and heaviest burning, and in an area where there were no potential sources of accidental ignition could have triggered a major criminal investigation. Fortunately, an astute investigator discovered the pan of burned food in the kitchen.

Examination of damage to the electrical system is frequently useful in determining the origin, or in testing a hypothesis about the origin. An electrical



Figure 3 Rectangular burn pattern on a gypsum drywall surface

circuit has the capacity to act as a fire detector. When the insulation on an electrical circuit is compromised by a fire, several things can happen. The carbonized insulation becomes conductive, and arcing through char can occur. This may or may not sever the circuit. Frequently, the current will meet sufficient resistance to flow that it does not open the circuit protection device, and the circuit can be attacked in another area, either upstream (toward the service panel) or downstream (away from the service panel). When the arc causes the wire to sever, current flow downstream ceases, and the point of arc severing furthest downstream can be said to have preceded any other arc severing on that circuit. Usually, the fire department or power company disconnects power to the structure sometime during the fire, and any electrical activity observed necessarily occurred before that happened. Sometimes the fire itself cuts



Figure 4 Location of a "drop down" fire in the interior wall space behind the pattern shown in Figure 3. The cross member stopped the hot coals, and those produced the pattern



Figure 5 The only significant fire damage in this residence was in the immediate vicinity of the body. The victim's clothing was ignited by a small cooking fire that self extinguished [Reproduced with permission from Gardiner Associates.]

off the power, when it burns the service drop, or causes the service entrance cable to arc to the distribution panel. In a fire that exhibits electrical damage in only one part of the structure, the search for the origin should focus on that area. Electrical arc damage is sometimes the only kind of artifact that survives a fire.

There have been only limited studies of the error rate of origin determinations by fire investigators. The studies that have been done, however, suggest that the error rate is alarmingly high. The US Bureau of Alcohol, Tobacco, Firearms, and Explosives (ATF) conducts fire investigation training courses at the Federal Law Enforcement Training Center near Brunswick, Georgia. One of the opening exercises that the ATF employs is to invite course attendees to examine the scene of a test fire and determine the origin and cause. No statistics have been kept on these exercises, but it has been reported that the number of correct determinations is appallingly low. In one exercise that was recorded, ATF Special Agent Steve Carman burned two similarly furnished bedrooms at a Las Vegas, Nevada test site. Fifty-three fire investigators from both public and private sectors, with varying degrees of experience, were asked to determine which quadrant of each room held the origin of the fire. In each case, only three investigators were able to determine the correct quadrant, and none correctly identified the point of origin [8]. The odds of correctly identifying the quadrant of origin by chance alone are 25%. Agent Carman's results indicate that fire investigators who use the area of lowest or deepest charring to indicate the area of origin in fully involved compartments frequently identify the wrong location. It seems likely that the origin of far more fires should

be declared "undetermined", or, alternatively, much larger areas of origin should be determined.

Hypothesis Development and Testing

This is the critical stage of the investigation. Using his or her education, training and experience, and the facts as they have been developed thus far, the investigator may have sufficient information to formulate a fire scenario that accounts for the condition of the scene. Because this step is so dependant upon the individual investigator's experience, it is a subjective exercise, and because it is subjective, it requires further testing if it is to be believed. Investigators who are trying to follow the scientific method as outlined in NFPA 921 may object to this characterization, because they will be basing their hypothesis "solely on the empirical data that the investigator has collected". However, the development of the hypothesis is, nonetheless, a subjective exercise using inductive reasoning, which involves starting with a particular experience and making generalizations. Inductive reasoning leads to probabilities, not certainties. An investigator who has never seen a cigarette thrown into a wastebasket can cause a fire may believe it does not happen and so is unlikely to formulate (much less test) such a hypothesis. The editors of *Fire Findings* conducted an experiment to determine whether a cigarette thrown into a trashcan with paper would start a fire, and it required 132 trials before the first ignition [9]. Had they stopped after 100 trials, they could well have published a report stating that cigarette will not ignite paper in trashcans. They ran a total of 300 tests and got five ignitions.

The overall hypothesis about the cause of a fire may include numerous subhypotheses, the accuracy of which may or may not influence the overall determination. For example, it may be part of the investigator's hypothesis that a particular door was open during the fire. If the door was open, then there should be evidence of burning on the exposed surfaces of the doorjamb and the hinges. If, on the other hand, the door was closed, then there should be mirror images, much like ink blots, on the mating surfaces, as shown in Figure 6.

Note that the hypothesis "test", described above, involves only observation and logic. No laboratory test is required. For hypotheses that state, "That burn pattern was caused by a liquid accelerant," a laboratory test is required. Except in the most obvious cases (none of which occur in a fully involved compartment), it is not possible to credibly state that an accelerant caused a particular burn pattern unless there is confirmation by a laboratory that follows ASTM methodology.



Figure 6 Mirror patterns on the mating surfaces of a door hinge, showing that the door was closed at the time of the fire. Part of the initial investigator's hypothesis was that the exterior door was left open on purpose to provide ventilation. All three hinges were found in the closed position, with screws still in place

For a discussion of the mythology of fire investigation, *see* **Arson Investigation: Misconceptions and Mythology**.

Hypothesis testing requires that the investigator compare his or her hypothesis with *all* of the relevant credible data. It also requires the investigator to be willing to abandon a hypothesis if it is disproved, and to seriously question that hypothesis if the data are merely "consistent" with it.

The identification of the origin of the fire is a hypothesis. Beware of the trap of circular logic. Once an investigator has identified an origin, every potential ignition source outside of that origin is automatically "eliminated", because only ignition sources within the area of origin could have started the fire. This type of thinking can lead to serious errors, the most common being the declaration that

"Since I could find no sources of accidental ignition within the area of origin, this must have been an incendiary fire." Such declarations are sometimes possible, but only in rare cases. For example, if there is a small fire confined to a bedroom closet, and there are no lights or wires or other devices in the closet, it may be possible to state with certainty that the ignition source was an open flame, even if no match of lighter is found. NFPA 921 allows for such a determination in cases "where the area of origin is clearly defined and all other potential heat sources at the origin can be examined and credibly eliminated" [10]. The phrase "clearly defined" is not clearly defined. This ambiguity has been exploited by some investigators to allow them to state that "clearly defined" means whatever they want it to mean (*ipse dixit*). The intent of the passage was to allow a reasonable determination to be made when the ignition source was absent, but not to give license to simply declare a fire incendiary because no accidental cause could be found. The technical committee's discussion of the issue came about when it finally attempted to address the problem of the "negative corpus" fire determination.

Negative corpus is shorthand for "negative *corpus delicti"*. The committee chose not to even mention the phrase in the document to avoid giving the concept any credibility. The same section of the document gives extensive guidance on the concept of "elimination" of a potential cause and states,

> Elimination, which actually involves the developing, testing and rejection of alternate hypotheses, becomes more difficult as the degree of destruction in the compartment of origin increases, and it is not possible in many cases. Whenever an investigator proposes the elimination of a particular system or appliance as the ignition source on the basis of appearance or visual observation, the investigator should be able to explain how the appearance or condition of that system or appliance would be different if it were the cause of the fire.

"Clearly defined", in this investigator's view, means that anyone, even someone completely untrained in the investigation of fires, could look at the damage and unhesitatingly, and without fear of contradiction, point at the location and state, "that's where the fire started". Figure 7 shows what this author had in mind when signing off on the "clearly defined" language.

In those cases where the origin of the fire is less "clearly defined", the absence of any potential heat sources in the proposed area of origin is *data*. It is, in fact, data that is contradictory to the hypothesis that this is the origin, and should cause the investigator to carefully reexamine the hypothetical origin. Is there another area that could reasonably be considered the origin of the fire? Will an investigator



Figure 7 A "clearly defined" origin. The fire began in the wardrobe in the corner. There were no lights or other potential ignition sources in the wardrobe, so a determination that the ignition source was an open flame can be justified. This photo also shows the development of a heat horizon, a smoke horizon, and a well-formed V-pattern caused by the intersection of the fire plume with the wardrobe door [Reproduced with permission from Gardiner Associates.]

retained by another party agree with the hypothetical origin? Certainly, in a fully involved compartment, every potential heat source within that compartment requires careful examination. Heavier damage on one side of the compartment or the other is almost certainly a result of increased ventilation rather than the longer burning time. If there is more than one fully involved compartment, there must be credible data to allow the investigator to decide which one burned first.

The circumstantial evidence that fire investigators use to draw inferences and form hypotheses is frequently subject to more than one interpretation. It is seldom as clear as the "footprints in the snow" analogy used to explain the concept of circumstantial evidence to jurors.

Data that contradicts an investigator's hypothesis about either the origin or the cause needs to be taken seriously. It cannot be discarded merely because "it does not fit". A credible hypothesis should fit all the relevant data.

Once an investigator has developed a testable hypothesis, and then tested it by comparing it against all of the data, a most difficult step remains. That is, the examination of any other hypothesis that explains the data equally well. By this point, it is difficult to maintain an open mind, but be certain that anyone who has an interest in finding an alternate hypothesis will put one forward. The investigator should be able to state that such hypotheses were considered, and be able to articulate exactly why those alternate hypotheses were discarded. In many cases, where an investigator's findings are discredited, it is not because such findings are incorrect, it is because the investigator is unable to credibly articulate why his or her hypothesis is true, and the alternate is not.

There will be times when more than one hypothesis will fit the data. In such cases, the investigator has the obligation to report on all credible hypotheses.

Reporting Procedure

The report may well be the single most important document generated during the fire investigation. It will be the document that prosecutors, defense attorneys, adjusters, and other fire investigators look to in order to understand a fire investigator's procedures and opinions. In some jurisdictions, if the report does not address a particular issue, that issue will be out of bounds when the investigator testifies. If there is an error in the report, it will become the focus for cross-examination. If there is insufficient detail in the report, readers will not understand how an investigator reached his or her conclusions. The quality of the report frequently determines the fate of the case.

Reports can take many forms, ranging from a few comments at the end of a fire report to a very detailed description, running a hundred pages or more. The format of any given report is usually governed by the requirements of the fire investigator's client, whether that is an insurance company, a police agency, a defense attorney, or a prosecutor. Some report formats are dictated by court rules.

A routine fire investigation report should follow the format of a scientific report, containing a background section, a synopsis of important witness observations, a description of the investigator's observations with a sketch and photographs to illustrate those observations, and a conclusion.

ASTM E 620 is the *Standard Practice for Reporting Opinions of Technical Experts*, and generally applies to fire investigators. E 620 requires that the report contains all facts that are pertinent to the opinion rendered, and that there be an identification of those facts that are based on the investigator's observations, as opposed to those facts that the investigator has learned from other sources. It also requires that the report contain the logic and reasoning of the investigator, whereby each of the opinions and conclusions were reached.

There are times when it is appropriate to include a "Discussion" section in a report, particularly if there are alternate hypotheses that explain the condition of the scene. The "Discussion" section is the appropriate place to describe the relative merits of each of these alternate hypotheses.

It sometimes happens that a report deadline arrives before an investigator has been able to obtain or review all of the relevant data. Even in such cases, it is frequently possible to arrive at certain opinions. In most cases, even if an investigator believes that all of the data has been reviewed, it is useful to add a disclaimer to the effect that the investigator reserves the right to amend or supplement the report should additional information come to light.

Summary

To be in a position to conduct a credible investigation, each agency should have a written procedure for investigations, which refers to NFPA 921. Fire investigators need to understand the purpose of their investigation and should formulate a plan for each one. No investigation should be undertaken until the investigative team makes sure that it can be done safely.

Documentation of observations is critical, both to help investigators communicate their findings and to help them remember. Investigations frequently need to be suspended to allow interested parties to be given the opportunity to see a fire scene before it is disturbed. Reconstruction and additional documentation are required once the initial evaluation is completed.

Hypothesis formulation and testing is the most important function of the investigator. Origin determination is a hypothesis that is validated by finding the cause within the area of origin. Finding no cause tends to invalidate the hypothetical origin, and in such a case, the investigator needs to revaluate the origin determination.

Reports can take many forms. They are used to communicate results and memorialize findings. Most reports should follow the basic outline of a scientific report, so that the reader understands both the investigator's observations and his or her reasoning. Record keeping should be consistent, and destruction of documents should be avoided.

References

[1] NFPA 921 (2004). *Guide for Fire and Explosion Investigations*, National Fire Protection Association, Quincy, MA, p. 14.
[2] Karter, M. (2008). *Fire Loss in the United States During 2007*, National Fire Protection Association, Quincy, July Available from http://www.nfpa.org.
[3] Michigan v. Tyler, 436 U.S. 499 (1978).
[4] Michigan v. Clifford, 464 U.S. 287 (1984).
[5] ASTM E860-07 (2007). *Standard Practice for Examining And Preparing Items that are or may Become Involved in Criminal or Civil Litigation*, ASTM International, West Conshohocken.
[6] NFPA 921 (2004). *Guide for Fire and Explosion Investigations, National Fire Protection Association*, Quincy, MA, p. 13.
[7] ASTM E1459-92 (2005). *Standard Practice for Evidence Labeling and Related Documentation*, ASTM International, West Conshohocken.
[8] Carmen, S. (2007). Understanding post-flashover fires: recognizing the importance of ventilation. *Presentation to the Oregon Chapter of the International Association of Arson Investigators (IAAI)*, Bend, September 12 2007.
[9] Sanderson, J. (1998). Cigarette fires in paper trash, *Fire Findings*, **6**(1), 1.
[10] NFPA 921 (2004). *Guide for Fire and Explosion Investigations, National Fire Protection Association*, Quincy, MA, p. 137.

JOHN J. LENTINI

Fire and Explosion Investigations: Overview

In forensic science books, the discussions of fires and explosions are frequently grouped together, but there are striking differences between the two. Both fires and explosions are events that cause the destruction of their causative agents, and both require considerable skills, in the field and in the laboratory, to understand the events.

In both fires and explosions, there is sometimes a perceived conflict between emergency responders and investigators. There are times when the emergency response, such as fire scene overhaul, can further complicate the investigator's task. Usually, this is simply an occupational hazard. The extinguishment of the fire or the removal of fire or explosion victims for medical treatment must take priority over the investigation, and most investigators understand this.

Whilst most fires are accidents explosions are generally not accidents. The vapor phase explosions caused by leaks of fuel gas or highly volatile liquids are the exception. While the average residential fire must be considered accidental until it is proven otherwise, there is usually little point in considering the potential accidental causes for a car blowing up in a crowded market. The nature of explosion investigations tends to be different from the nature of fire investigations. Those conducting explosion investigations are typically affiliated with the national government (and perhaps health and safety regulators), while the responsibility for investigating fires usually rests with local authorities. Fires also carry with them the potential for civil litigation, as the

assignment of liability may be on a product incorrectly manufactured or a service negligently provided.

The goal of investigators of both fires and explosions is to scientifically establish what happened and who is responsible [1]. This frequently involves the coordination of several scientific and investigative disciplines, but always requires that there be a principal investigator, who puts it all together, and does the "reality check" at the end. This principal investigator need not be a formally trained scientist, but he is certainly required to think like a scientist, and to follow the scientific method [2].

Problems can arise when the fire investigator is also the person responsible for prosecuting the case if the fire is determined to be incendiary. The knowledge that comes with fulfilling the police responsibilities tends to make objectivity more difficult in fulfilling the scientific responsibilities. Division of responsibilities is perhaps the best way to overcome some of these difficulties.

The succeeding articles will cover the following topics:

- The chemistry of fire (*see* **Fire: Chemistry of**)
- Fire dynamics and fire pattern development (*see* **Fire: Dynamics and Pattern Production**)
- Investigation of fire scenes (*see* **Fire: Scene Investigation**)
- The mythology of arson investigation (*see* **Arson Investigation: Misconceptions and Mythology**)
- Using computers to model fire behavior (*see* **Fire Modeling and Its Application in Fire Investigation**)
- Laboratory analysis of fire debris (*see* **Fire Debris: Laboratory Analysis of**)
- Investigation of explosion scenes (*see* **Explosions: Scene Investigation**)
- Laboratory analysis of explosive residues (*see* **Explosion Debris: Laboratory Analysis of**)
- Standardization, certification and accreditation (*see* **Fire Investigator: Standardization, Accreditation, and Certification**)

References

[1] NFPA 921 (2004). *Guide to Fire and Explosion Investigations*, National Fire Protection Association, p. 6.
[2] Lentini, J. (2006). *Scientific Protocols for Fire Investigation*, CRC Press, p. 537.

JOHN J. LENTINI

Fire and Explosions: Mythology of Arson Investigations *see* Arson Investigation: Misconceptions and Mythology

Fire Debris: Laboratory Analysis of

Introduction

The laboratory analysis of fire debris is one of the most important hypothesis tests that can be performed in an investigation, especially when the investigator forms a hypothesis that the fire was set using ignitable liquids [1]. The vast majority of fire debris analyses are tests for ignitable liquid residues (ILR). Other tests of fire debris will be discussed once ILR analysis is covered.

As a result of what has been learned in the last 20 years about fire pattern production (*see* **Fire: Dynamics and Pattern Production**), laboratory analysis is the only valid way to conclusively determine that an ignitable liquid was used to start a fire, at least in a compartment that has become fully involved. Even when the compartment is not fully involved, and there is a burn pattern similar to the one shown in Figure 1, the investigator still needs to determine the identity of the ignitable liquid.

In the past, the laboratory analysis was regarded as "the icing on the cake," because by the time samples were collected, the fire investigator had already decided what had caused the fire, and the stated purpose of the laboratory analysis was to help determine the identity of the flammable or combustible liquid used to start the fire. Investigators were accustomed to receiving negative reports from their laboratory, even when they "knew" that a fire had been intentionally set with ignitable liquids.

Back in those days, the term *ignitable liquid* had not yet been coined. "Flammable or combustible liquids" were generally referred to, even by people in the laboratory, who had no clue how such liquids



Figure 1 "Obvious Pour Pattern." The only surface that was burned in this mobile home was the floor. There were no furnishings in the house. The carpet was tested and found to be positive for the presence of a medium petroleum distillate such as mineral spirits or charcoal lighter fluid. This is one of the unique cases where visual observation alone can lead to valid conclusions about what caused the pattern

may have been used, as "accelerants." Many of the findings were, in fact, false negatives, because the sensitivity of laboratory methods was just not very good. The sensitivity has improved dramatically since 1974, to the point where one can now see the petroleum products that are a natural part of our background, but many fire investigators still distrust negative reports from laboratories, based on this earlier experience.

The development of analytical procedures has paralleled the development of standards for fire debris analysis. The first vague outline of a standard was not published until 1982 [2], and prior to that, analysts would report that a sample contained "an oily liquid" that had "sufficient similarities" to a known sample of gasoline or kerosene or diesel fuel.

Isolating the Residue

Initial Sample Evaluation

Once the paperwork is done, the first critical step in the analysis of fire debris is the selection of a separation technique. Choosing an inappropriate technique could result in a false negative, a misidentification, the destruction of evidence, or any combination of these. The first step in this selection process (and the first step in any chemical analysis process) is to look at the sample. One purpose of looking at the sample is to ensure that it is what it purports to be. Sample characteristics will determine the most appropriate method for isolating any ILR that may be present. Once the visual examination has taken place, the next step is a "nasal appraisal." Occupational safety experts will no doubt frown on this recommendation, but it can be done carefully. There is no need to put one's nose in the can, even though the analyst can be reasonably certain that the fire investigator who collected the sample has already done exactly that. Unless the sample is a liquid sample for comparison purposes, it can be safely appraised by removing the lid, and waving the hand gently over the top of the sample to see if it exhibits any obvious odors. If there is an odor present, it becomes possible to do a rapid and accurate analysis by removing a small piece of the sample and extracting it with solvent. If the odor is very strong, it is advisable to remove a small piece of the sample and place it into a separate container for analysis.

ILR Isolation Method Selection

Solvent extraction according to American Society for Testing and Materials (ASTM) standard practice E 1386 is an appropriate method for rinsing out empty containers, for extracting small aliquots of samples with a high concentration of ILR, and for isolating residues from very small samples [3].

The vast majority of samples, however, are likely to be samples of burned building materials, floor coverings, and furnishings, which do not exhibit a strong odor, and are best analyzed by passive headspace concentration, as described in ASTM E

1412. This technique is essentially nondestructive [4]. If the analyst decides to use another technique later, running passive headspace concentration will not interfere with that. A schematic representation of the passive headspace concentration principle is shown in Figure 2.

Other methods of isolation have been studied thoroughly over the last 25 years, and while they have some utility, none match the advantages of passive headspace concentration using an activated carbon strip. Static headspace sampling (Figure 3), which involves warming the container and sampling the vapors in the headspace directly and then injecting those samples into the gas chromatograph, is a good screening technique, but it does not result in the production of an archiveable extract, nor does it detect compounds much heavier than C_{15}. Static headspace sampling is described in ASTM E 1388 [5]. Dynamic headspace concentration, described in ASTM E 1413 [6], was useful in demonstrating the effectiveness of adsorption/elution as a valid approach to ILR isolation, but it is destructive, requires far more attention than passive headspace concentration, and the apparatus can be cumbersome and finicky. It is no more sensitive than passive headspace concentration, and it is possible to achieve "breakthrough," which occurs when the carrier gas first moves the analytes onto the adsorbent, then blows them off again. The only advantage that it offers is speed. Solid-phase microextraction (SPME), described in ASTM E 2154 [7], is another option, but it is very labor intensive, and, like headspace sampling, does not have the potential to produce an archiveable sample that can be analyzed again. Using passive headspace concentration results in a solution that can be injected many times, and when the carbon strip is left in the solution, it will gradually readsorb ILR as the eluting solvent evaporates. Juhala reported the readsorption by small potions of activated charcoal that fell off Plexiglas beads in 1982 [8]. The solution can be reconstituted years later if a second look at the sample is desired. Because of the transient nature of many fire debris sample containers, the archived activated carbon strip is often the best evidence after a few years have passed.

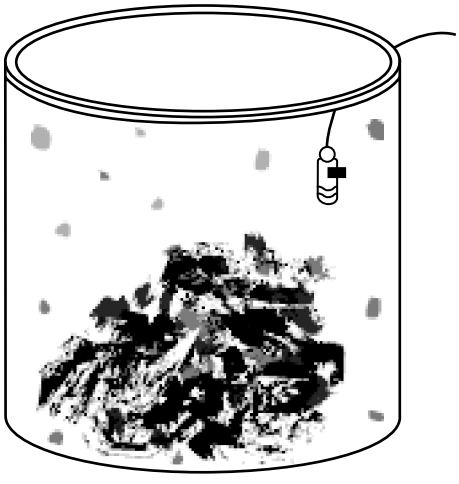


Figure 2 Schematic drawing of passive headspace concentration using an activated carbon strip. Vapors are produced by heating the container with debris to 80°C. The ACS adsorbs the vapors for 16 h, then is rinsed with diethyl ether spiked with 100 ppm perchloroethylene, and the resulting solution is analyzed by GC-MS



Figure 3 Using a gas-tight syringe to withdraw a sample of headspace from a fire debris sample. The headspace sample, about 500 µl, is injected directly into the GC-MS injection port

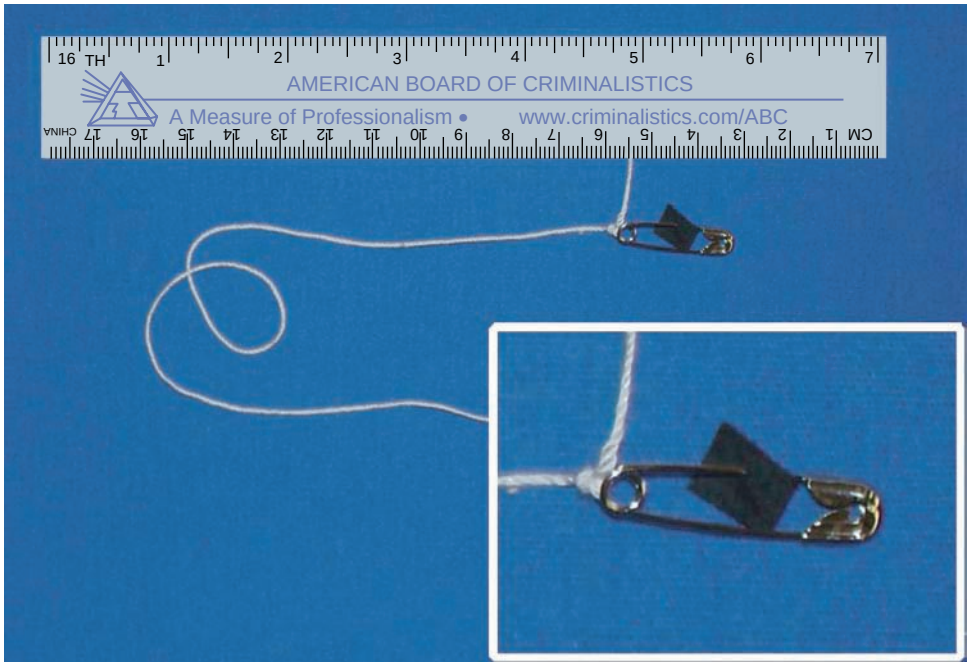


Figure 4 Close-up view of an activated carbon strip

The only equipment required for passive headspace concentration is a convection oven, vials, activated carbon strips (ACS) and a solvent dispenser. Caseload will determine the required oven size.

Every laboratory should optimize the parameters in its ACS procedure to make sure they are getting the best results possible. "Good" results from ACS are those where the chromatogram of the concentrated headspace vapors of a standard closely matches the chromatogram of that same standard in the eluting solvent.

A 10×10 mm carbon strip is the minimum size recommended. A typical carbon strip is shown in Figure 4. This 100-mm^2 strip can easily accommodate the headspace vapors from 10 µl of any ignitable liquid placed on a Kimwipe® in a quart can. It is possible to overload carbon strips, and this results in a preferential adsorption of heavier hydrocarbons with the loss of aromatics, but this effect is generally not large enough to affect the identification. Usually, samples that are capable of overloading the carbon strip will exhibit a strong odor, and the analyst can take an aliquot of the sample, or reduce the analysis time.

For samples with a high water content, there exists a danger that the vapor pressure in the sample container will cause the lid to pop off. This has the potential for contaminating the sample oven. A "pressure relief device" is easy to construct for such samples. Puncture a small hole in the lid, and cover with cellophane tape. Figure 5 shows a can so equipped.

The analysis time of a typical ACS procedure is 16 h. The analyst is encouraged to experiment with different adsorption times and temperatures, with the goal of finding a balance between the maximum recovery and the minimum time necessary. Sixteen hours is convenient because the samples can be put in the oven at 4 : 00 p.m. and taken out at 8 : 00 a.m. the next day. One of the major advantages of ACS is that it requires very little attention from the analyst. Once the strip is in the can and the can is in the oven, nothing is going to happen until it is time to take the strip out of the can, put it in the vial, and add the eluting solvent. At this point in the analysis, however, the analyst must be extremely attentive to the procedure. Once the carbon strips are taken out of the sample container, they have an identical appearance. It is at the point of placing the strips into



Figure 5 One-quart can be equipped with a "pressure relief device," a short strip of cellophane tape over a small hole pierced in the lid

their vials that an unrecoverable error can occur – a strip can be misidentified.

Many laboratories use a preconcentration step, wherein they add approximately 500 μl of solvent to the strip in a vial; then after the strip has had a chance to equilibrate, it is removed and the solvent is evaporated down to 100 μl or so. This will result in a fivefold increase in concentration, but that increase can usually be achieved electronically with very little loss of signal or increase in noise. The first analysis, in this analyst's view, does not require the preconcentration step. That can be accomplished at a later time, if necessary, but preconcentration runs the risk of skewing the results. The only appropriate way to evaporate the solvent is to blow a stream of dry nitrogen over it. Heating the solvent may result in the loss of lighter ILR components.

Solvent Selection

The solvent used for the elution is another critical choice. The most popular eluting solvent is carbon disulfide; a highly toxic, carcinogenic, teratogenic, smelly, nasty liquid that will ignite upon exposure to boiling water. It does work very well to elute aromatics and aliphatics approximately equally from activated carbon strips, but so does diethyl ether. There have been studies to indicate that carbon disulfide is superior to diethyl ether, or to pentane, the other solvent recommended by ASTM Committee E 30 [9]. Another study [10] found only marginal differences between diethyl ether and carbon disulfide when applied to carbon strips exposed to 10 μl samples of ILR. Carbon disulfide was originally selected as a fire debris solvent because of its relatively quiet signal when passing through a flame ionization detector (FID). When using a mass spectral detector, that advantage disappears because the detector is turned off while the solvent is passing through. Some concerns have been expressed about the capability of diethyl ether to form explosive peroxides, but that will not occur if the ether is kept in a refrigerator and used on a regular basis. Explosions of cans of ether have been reported only when those cans have been unused for years. Carbon disulfide, diethyl ether, and pentane are all highly flammable, but with respect to fire, carbon disulfide poses the greatest risk in that it has the lowest ignition temperature and the broadest flammable limits [11]. A comparison of the properties of the three solvents recommended by ASTM E 1412 is shown in Table 1.

Internal Standards

There are two places in the analysis of fire debris where the use of internal standards is appropriate. Addition of an internal standard to the sample itself allows the analyst to develop at least a qualitative feel for the "tenacity" of the sample, and for the effectiveness of the isolation procedure. This may be accomplished by the addition of 3-phenyltoluene (20 μl of a 2.5% solution of 3-phenyltoluene in ether). In the eluting solvent, a second internal standard consisting of 0.1% (100 ppm) perchloroethylene is added.

If the 3-phenyltoluene does not appear in the chromatogram, this means that the sample is exceedingly tenacious, and suggests that the tenacity of the sample might be the reason that the chromatogram appears so flat. If the perchloroethylene peak does not appear, that means something has gone wrong with the injection. Comparison of the signal from the sample to the perchloroethylene signal allows for a semiquantitative determination of the amount of ILR present.

Table 1 Elution solvent comparison [a]

	Carbon disulfide	Diethyl ether	Pentane
Flash point	−30 °C (−22 °F)	−45 °C(−49 °F)	−40 °C (−40 °F)
Lower explosive limit (LEL) (% vol in air)	1.0	1.9	1.5
Upper explosive limit (UEL) (% vol in air)	50.0	48.0	7.8
Specific gravity	1.3	0.7	0.9
Boiling point	40 °C (115 °F)	35 °C (95 °F)	36 °C (97 °F)
Autoignition Temp	90 °C (194 °F)	180 °C (356 °F)	260 °C (500 °F)
Exposure limit, TWA[b]	4 ppm	400 ppm	600 ppm
Exposure limit, STEL[c]	12 ppm	500 ppm	750 ppm
Carcinogenic	Yes	No	No
Teratogenic	Yes	No	No
IDLH	500 ppm	19 000 ppm[LEL]	15 000 ppm[LEL]
FID signal	Small	Large	Very large
Cost per liter[d]	$431.70	$71.05	$62.85

[a] Refer to your laboratory's hazard communication literature or MSDS for more complete information
[b] Time-weighted average for an 8-h exposure
[c] Short-term exposure limit (up to 15 min)
[d] J. T. Baker, online catalog 24 Nov 2008. All prices are for case quantities of Ultra Resi-analyzed grade

For 10 μl standards of known ignitable liquids isolated according to ASTM E 1412, the two internal standard peaks are roughly on the same order of magnitude as the sample peaks. Blanks are run using the 3-phenyltoluene applied to filter paper, and the blank strip is eluted with the spiked solvent. The advantages of using internal standards far outweigh any disadvantages. Some analysts may perceive a danger of being accused of "contaminating" a sample, but this is easily overcome by having a proper blank in the file.

Advantages and Disadvantages of Isolation Methods

Two drawbacks have been cited for use of the ACS method, the time required to perform the adsorption, and the relative lack of sensitivity compared to SPME. If laboratories were in the habit of providing same-day service, then the first argument might have merit. If laboratories "batch" their fire debris work and make their clients wait up to six weeks for results, a 16-h versus a 15-min analysis time is meaningless. With respect to sensitivity, the ACS method is capable of routinely detecting 0.1 μl of ignitable liquid from a nontenacious background.

The fire debris analyst's job is to help the fire investigator understand whether a foreign ignitable liquid was present at the fire scene. Fire debris laboratories now have the ability to detect the solvent in polyurethane finish five years or more after it has been applied to a hardwood floor. There is normally no need for more analytical sensitivity than that.

The "screening" of fire debris samples does not make much sense, unless there is a national emergency and the investigator needs results right away. For routine analysis, passive headspace concentration, according to ASTM E 1412, should be the norm. Solvent extraction, as described in ASTM E 1386, is appropriate for sampling aliquots of very strong samples or for extracting very small samples or empty containers.

Steam distillation, shown in Figure 6, may be selected in cases where it is desired to produce a neat liquid extract of the fire debris. The benefit of this is that a vial of the liquid can be brought into a courtroom, shown to a jury, lit on fire on a Q-tip, and passed around. Since steam distillation is only appropriate on very concentrated samples, however, it is preferable to make sure that the sample is preserved, and the sample itself can be passed around for the jury to smell. For the most part, steam distillation is a technique whose time has come and gone.

A comparison of the advantages and disadvantages of the various isolation techniques is shown in Table 2.

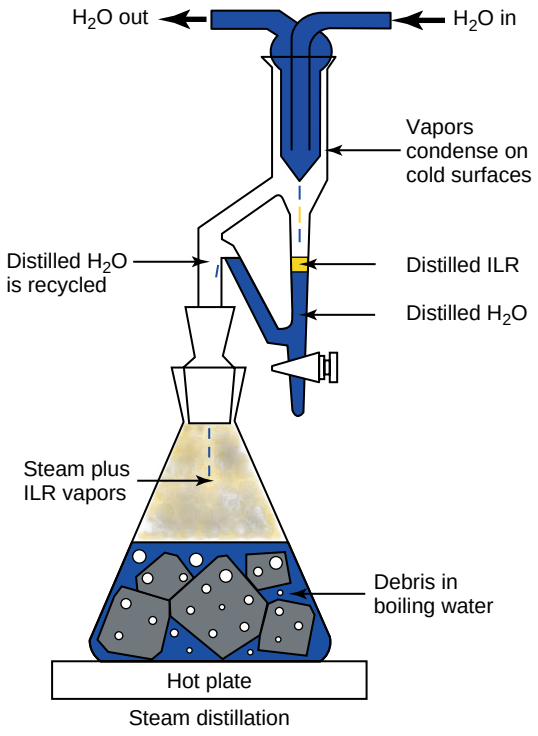


Figure 6 Steam distillation apparatus. The sample is boiled. Vapors condense on a "cold finger" and fall into the trap, which allows the water to recycle, while the immiscible oil layer builds up on top of the water column

Analyzing the Isolated ILR

Despite the improvements in separation and detection technology, the overall approach to identification of ILR is the same as it was in the early 1970s. A chromatogram from the sample is compared with chromatograms from known standards, and the analyst determines whether there are "sufficient similarities" to make an identification. What has changed is the quantity of information available because of the increased resolution provided by capillary columns and the ability to obtain a mass spectrum up to 10 times per second, as well as the reaching of a consensus on the meaning of "sufficient." So, while there is more information to compare, the technique is still one of pattern recognition and pattern matching.

An argument can be made that when one looks at a mass spectrum, one is looking at structural details, rather than simply matching patterns, but the patterns still have to match. The same argument has been made about structural elucidation in the use of FTIR for drug identification. While it would be nice to think that analysts routinely consider molecular structure, the day-to-day operation is one of pattern matching.

Fire debris analysis involves a different set of parameters than those employed by environmental scientists, who are typically trying to quantify the components of an oil spill or contaminants at a Superfund site. Environmental methods typically assume the presence of gasoline or other petroleum products, and then look for benzene, toluene, ethylbenzene, and xylene (BTEX) to quantify the amount present. Unless they are trying to identify the source of the spill, environmental analysts are usually not employing the same skill set used by fire debris analysts.

If one considers the nature of many petroleum products and the processes that are going on when these products are isolated from debris samples, one can begin to appreciate why chromatographic patterns look the way they do. Many petroleum products are straight-run distillates, particularly the medium and heavy petroleum distillates (HPD). The overall pattern of these products is a Gaussian (bell-shaped) distribution of peaks, dominated by the normal alkanes.

A fractionation process similar to distillation occurs when an ILR is isolated from a sample. This is caused by the very low vapor pressure of ILR compounds above C_{18}. If they do not get into the air in the headspace, they will not be adsorbed onto the carbon strip.

The ASTM standards give examples of many of the patterns that an analyst is likely to see in positive samples, and there is a detailed compilation of hundreds of patterns available in a standard text [12]. Both the ASTM standards and the gas chromatography mass spectrometry (GC-MS) Guide provide sufficient parametric information to allow the analyst to set up an instrument to provide patterns that look very much like the ones in the texts. While these texts are a great resource, it is imperative that every fire debris analysis laboratory has its own library of ILR. This provides for patterns with the exact same retention times, mass spectra with the exact same fragmentation patterns, and a quality assurance tool to let the analyst know when there has been some drift in the instrument, and when it is time to run a new set of standards. Whatever approach is taken, it will take time to develop the ability to recognize ILR patterns.

Table 2 Comparison of ILR isolation techniques

Method	Advantages	Disadvantages
E 1385 steam distillation	Produces a visible liquid, simple to explain	Labor intensive, destructive, not sensitive, requires expensive glassware
E 1386 solvent extraction	Useful for small samples and empty containers, does not cause significant fractionation, useful for distinguishing HPDs from each other	Labor intensive, expensive, co-extracts nonvolatile substances, increased risk of fire, solvent exposure, destructive
E 1388 headspace sampling	Rapid, more sensitive to lower alcohols, nondestructive	No archiveable sample, not sensitive to heavier compounds, poor reproducibility
E 1412 passive headspace	Requires little analyst attention, sensitive, nondestructive, produces archiveable sample, inexpensive	Requires overnight sampling time
E 1413 dynamic headspace, using activated charcoal	Rapid, sensitive, produces archiveable sample, inexpensive	Labor intensive, subject to breakthrough, destructive
E 1413 dynamic headspace, using Tenax	Rapid, sensitive	Labor intensive, requires thermal desorption, no archiveable sample, destructive
E 2154 solid-phase microextraction (SPME)	Rapid, highly sensitive, useful for field sampling with portable GC/MS	Labor intensive, expensive, requires special injection port, reuse of fibers, no archiveable sample

Table 3 Ions characteristic of compound classes

Ion	Compounds
57	Alkanes
83	Cycloalkanes
91	Toluene, xylenes
105	C_3 alkylbenzenes
117	Indan, methylindans
119	C_4 alkylbenzenes
131	Methyl, dimethylindans
142	Methylnaphthalenes
156	Dimethylnaphthalenes

The mass spectrometer provides the ability to simplify what can be very complex and confusing patterns. This is a result of the ability of the data analysis software to separate out only those peaks having particular ions present in the pattern. For instance, if one wants to look at the normal alkanes, one only needs to obtain an extracted ion chromatogram (EIC) for $m/z57$ (hereafter referred to as *ion 57*) and most of the balance of the components will disappear. It is thus possible to break down a total ion chromatogram (TIC) into its component parts. This approach to data analysis is known as *mass chromatography* and was first proposed by Smith in 1982 [13]. There are basically two ways to approach mass chromatography, the single ion approach, and the multiple ion approach. Dolan [14] has proposed referring to the single ion chromatograms as "extracted ion chromatograms" (EIC) and multiple ion chromatograms as "extracted ion profiles" (EIP). Whether the analyst prefers EICs or EIPs, the same ions, the ones shown in Table 3, are used to identify the compounds commonly found in petroleum products.

When working with software that scales the EIC or profile to the tallest peak, some caution is advised when using multiple ions. To the extent that the second, third, and fourth ions contribute to a pattern, they tend to make it more complicated, thus defeating some of the purpose of extracting the ions in the first place. To the extent that these additional ions do not contribute, they may convince the analyst that he or she is seeing more than what is actually present. Finally, to the extent that these additional ions are present at substantially lower concentrations, they would be better observed on their own, rather than in the profile. Figure 7 is a comparison of ion 57 from a kerosene standard versus a profile based on ions 57, 71, 85, and 99. The profile is slightly more complicated, but since ion 57 is the base peak for almost every component in the chromatogram, the ion 57 profile is the tallest.

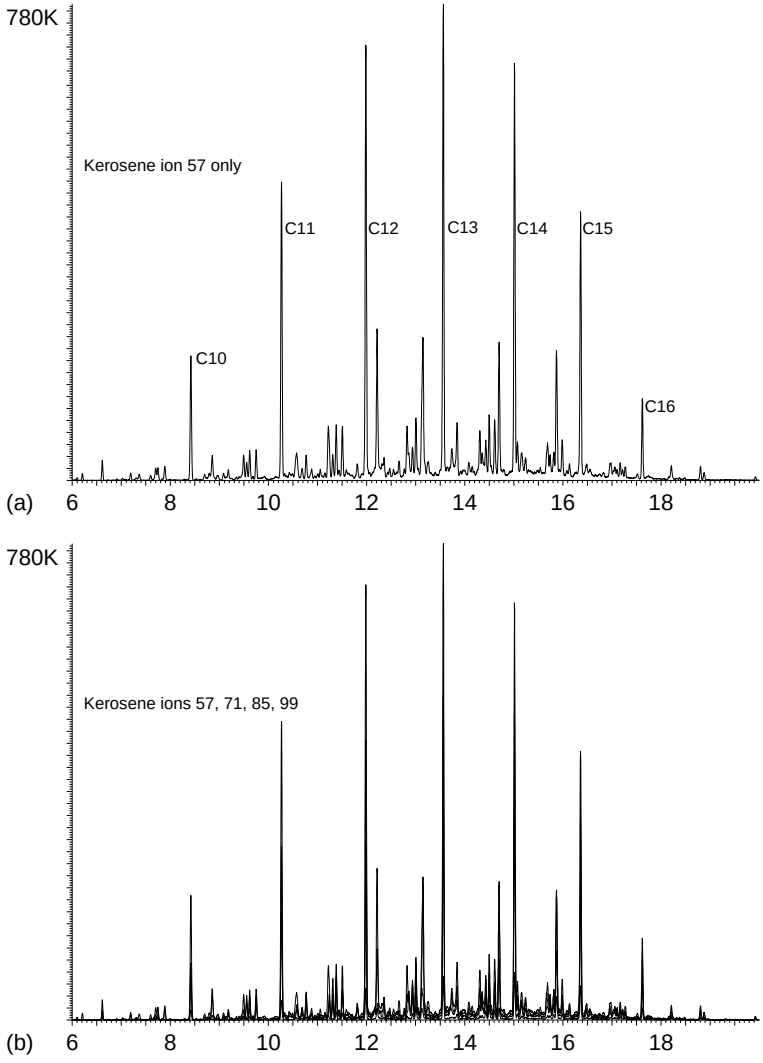


Figure 7 Ion profiling vs. ion chromatography. Comparison of the ion 57 chromatogram (a) with the ion profile combining ions 57, 71, 85, and 99 from a kerosene standard (b). The four ions in the profile are plotted in the merged format. Plotted individually, they are all very similar, with ion 57 presenting the tallest peaks

Another example is shown in Figure 8, which presents ions 128, 142, and 156, the naphthalenes, from a gasoline sample. In the top chart, the ions are combined into a profile, which, although it gives the analyst an idea about the relative abundance of the three ions, shows very little detail for the dimethylnaphthalenes represented by ion 156. When the extracted ions are presented separately, the analyst still gets the quantitative data by reading the abundance numbers next to the *Y* axis, but also gets to see the fine details in the shape of the peaks at the right side of the chart.

A similar example is shown in Figure 9(a). These charts show ions 91, 105, and 119 from a 75% evaporated gasoline "standard." (The term *standard*, when used herein, is intended to mean "known reference material," rather than a standard reference material. The gasoline was purchased from a gas station.) The effect is even more pronounced at 90% evaporation, shown in Figure 9(b), when the ion 91 peaks are

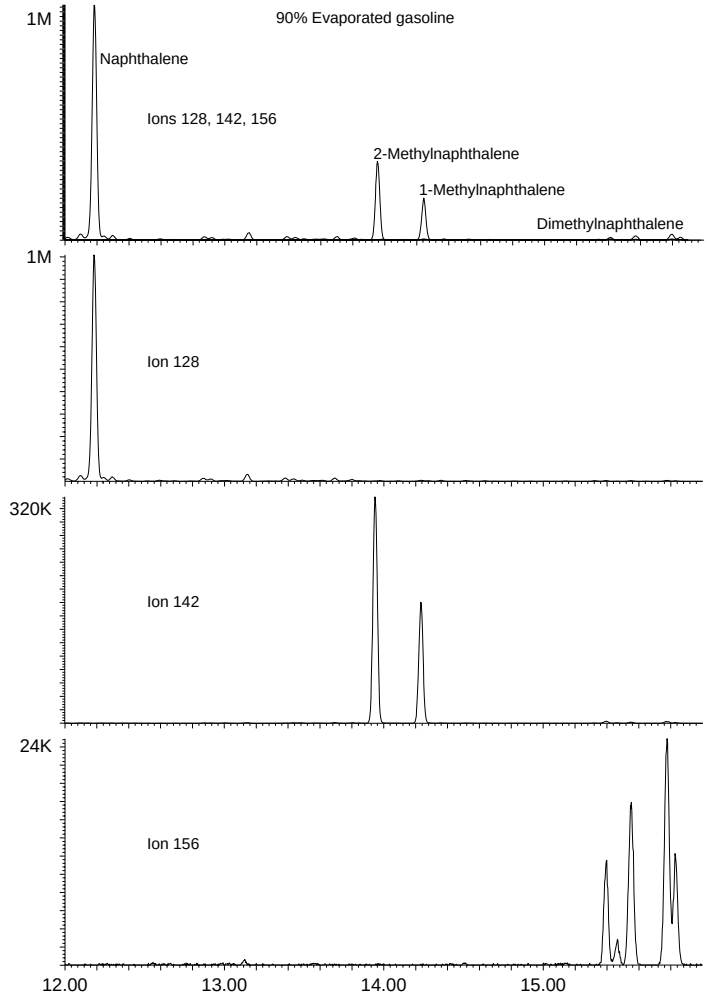


Figure 8 Ion profiling vs. ion chromatography. Comparison of the ion profile of the naphthalenes from a gasoline standard (top chart). Compare with the detail provided by presenting the three ions (128, 142, and 156) separately

smaller. Whether an analyst chooses extracted ion chromatography or extracted ion profiling is largely a matter of taste. The typical set of six ion chromatograms used in many laboratories to document the presence of gasoline is shown in Figure 10, and the set of three ion chromatograms used for distillates is shown in Figure 11. The details of each are discussed below.

Criteria for Identification

Most of the chromatograms that an analyst uses to make a positive identification of an ILR will match at the level of the TIC. Samples with low concentrations of ILR or high backgrounds, or a combination of both, can sometimes yield a positive identification for ILR if the analyst is very careful. Generally, but not always, if the TIC from the sample does not match the TIC from the standard, the sample is likely to be negative. Examples of cases where this is not true are presented following this general discussion of criteria for identification of an ILR in routine cases.

ASTM E 1618 identifies eight classes of ignitable liquids identifiable by GC-MS [15]. These classes are as follows:

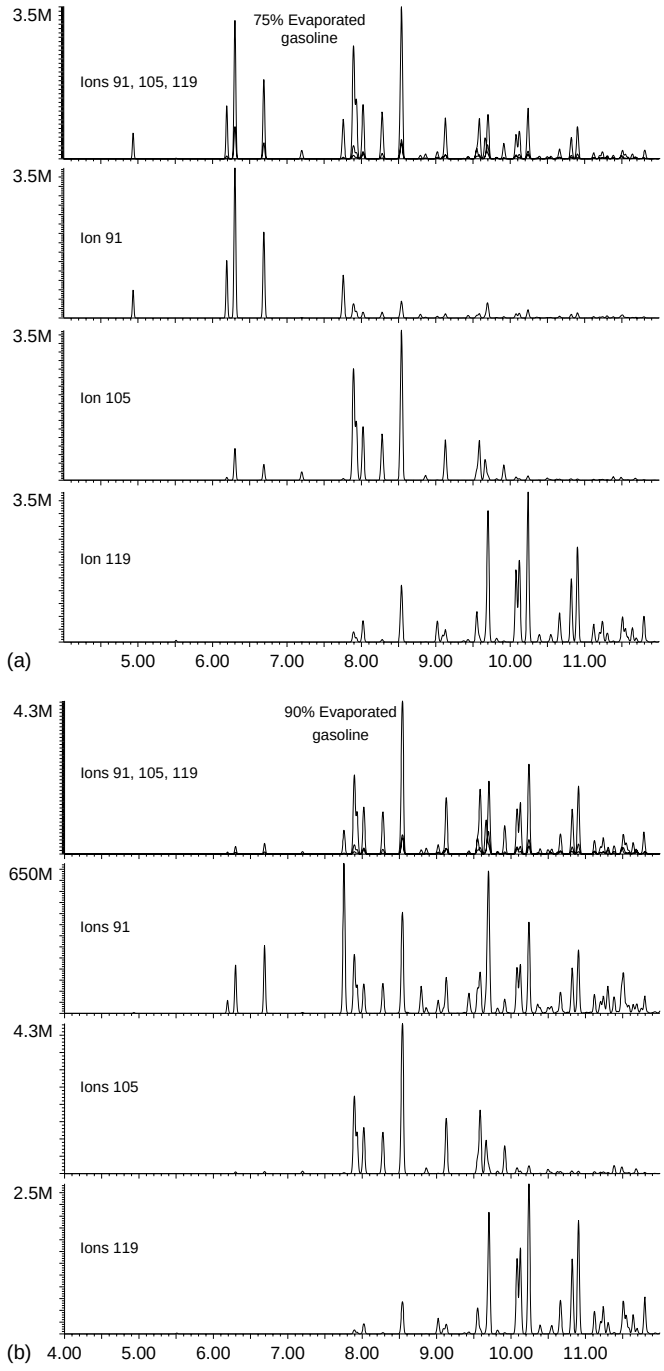


Figure 9 (a) Ion profiling vs. ion chromatography. Comparison of ion profiles from 75%-evaporated gasoline. The top chart shows a combination of ions 91, 105, and 119. The next three charts show those ion chromatograms plotted independently. (b) Ion profiling vs. ion chromatography. Comparison of ion profiles from 90%-evaporated gasoline. The top chart shows a combination of ions 91, 105, and 119. The next three charts show those ion chromatograms plotted independently. Note the improvement in the level of detail presented for the xylenes

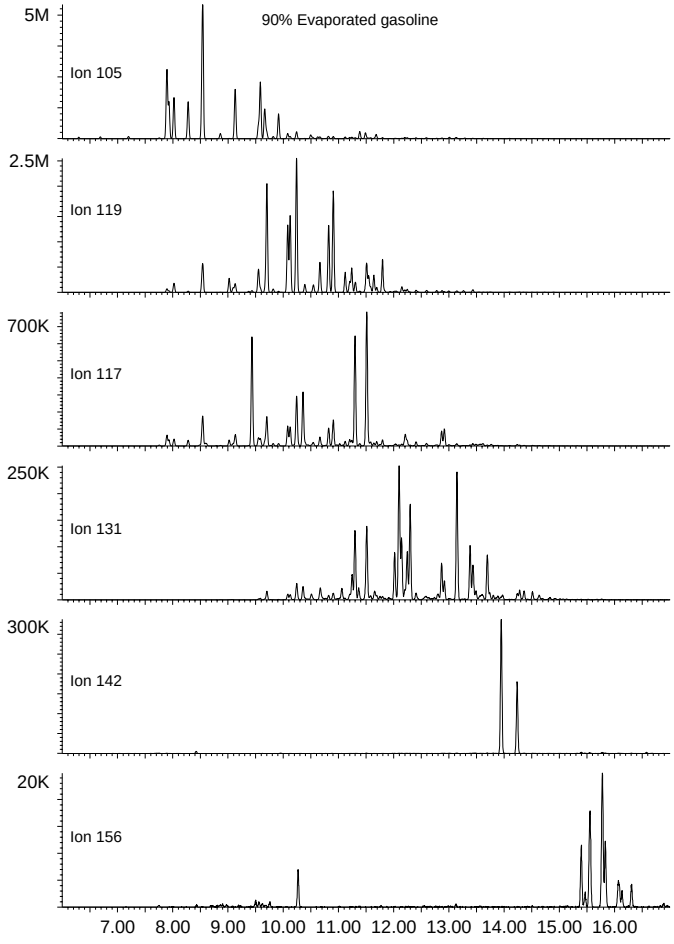


Figure 10 A typical set of six ion chromatograms used to document the presence of gasoline. Ions commonly used are 105 for C_3 alkylbenzenes, 119 for C_4 alkylbenzenes, 117 for indan and methyl indans, 131 for dimethyl indans, 142 for 2- and 1-methylnaphthalene, and 156 for dimethylnaphthalenes

- gasoline
- petroleum distillates
- isoparaffinic products
- aromatic products
- naphthenic–paraffinic products
- *n*-alkane products
- oxygenated solvents
- miscellaneous.

With the exception of gasoline, and the oxygenated solvents, each of the above classes can be divided into three further classes according to the average molecular weight: light (C_4–C_9), medium (C_8–C_{13}), and heavy (C_8–C_{20+}).

The standard presents criteria for the various kinds of compounds (alkanes, cycloalkanes, aromatics, and condensed ring aromatics) found in each one of these classes. The alkanes include both straight-chain and branched hydrocarbons, and can be extracted using ion 57. The cycloalkanes are mostly substituted cyclohexanes, which can be seen by extracting ion 83. Cycloalkanes also have a strong ion at 55, and as the length of the substituted alkane increases, ion 57 begins to dominate the mass spectra. “Aromatics” means alkyl-substituted benzenes with a single ring. These can be extracted using ions 91, 105, and 119. Ion 91 will show toluene and xylenes, ion 105 will show xylenes and C_3 alkylbenzenes, and ion 119 will

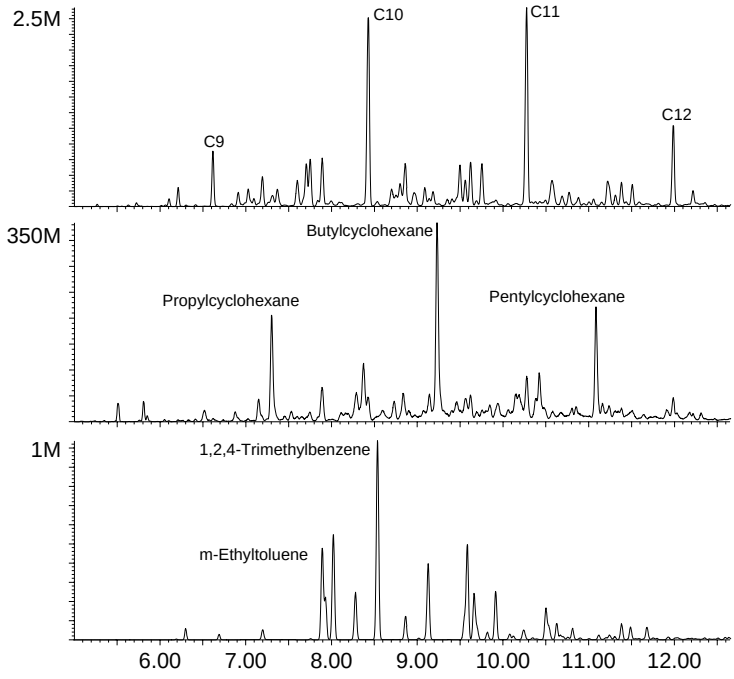


Figure 11 A typical set of three ion chromatograms used to document the presence of distillates. These are the ion 57, 83, and 105 chromatograms from smokey bear charcoal lighter, a medium petroleum distillate

show C_4 alkylbenzenes. There will be some overlap in the EIC, as shown previously in Figure 9. "Condensed ring aromatics" refers to indans and naphthalenes. Indans have a five-member ring attached to a benzene ring, and may be substituted. The naphthalenes that we usually see are naphthalene itself, 2- and 1-methyl naphthalene (written in that order because that is the order of elution from a nonpolar column) and the dimethyl naphthalenes. Naphthalene presents a single peak when ion 128 is extracted, there are two peaks for the dimethyl naphthalenes seen when ion 142 is extracted, and eight peaks in the dimethyl naphthalene chromatogram seen when ion 156 is extracted. The indans can be visualized by extracting ions 117 and 134.

Identification of Gasoline

The composition of petroleum products as found in fire debris is influenced by three factors: crude oil parentage, the effects of petroleum refining processes, and the effects of weathering. Gasoline is the way it is largely because of the second of these influences, petroleum refining. Although all crude stocks contain aromatics, aliphatic hydrocarbons comprise the bulk of most crude oils. Because the aliphatics cause knocking in gasoline engines, the value of the crude stock is enhanced by reformation through a process of dehydrogenation. Toluene and xylenes are the most abundant compounds produced in these processes. Gasoline does contain numerous light aliphatics, as light as butane, that are present when the gasoline is first pumped, but by the time it has been through a fire, most of the aliphatics lighter than toluene have dissipated. Consequently, toluene is usually one of the first tall peaks seen in a sample of gasoline, and once it has weathered to 50% or more, the toluene is much lower than the C_3 alkyl benzene or xylene peaks. Gasoline changes considerably as it evaporates, which is what makes it one of the more challenging classes to identify. When a Gaussian distribution of normal alkanes changes because of evaporation, it is still a Gaussian distribution of normal alkanes, just a heavier one. Most of the samples of gasoline that this author has seen in samples of fire debris have been more than 75% evaporated.

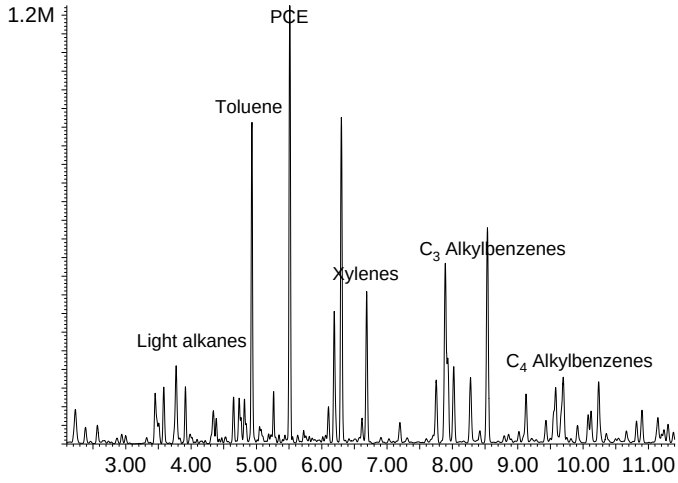


Figure 12 Total ion chromatogram of fresh gasoline. On a piece of filter paper 10 μl was spotted and the headspace was concentrated using ASTM E 1412. The eluting solvent was diethyl ether spiked with 100 ppm perchloroethylene, shown here eluting between toluene and the xylenes

Gasoline is probably the most frequently misidentified ILR. That is because many of the compounds present in gasoline as it comes from the pump are also produced when polymers degrade as a result of exposure to heat. The key to avoiding misidentifications is making sure that the ratios between groups of compounds and within groups of compounds are consistent with the standard. Toluene is a very common pyrolysis product. It is an unusual fire debris sample that does not contain toluene. Since it is one of the first compounds to evaporate, one does not expect to see a tall toluene peak in the absence of equally tall xylene peaks. Figure 12 shows the chromatogram of a 10 μl standard of gasoline adsorbed using an activated charcoal strip, and eluted with diethyl ether spiked with 100 ppm perchloroethylene. The toluene and xylene peaks are almost equally tall. If a fire debris sample contains toluene from gasoline, it will be accompanied by xylenes and the higher peak groupings of gasoline. Toluene that is not so accompanied comes from something other than gasoline.

Xylenes are also produced by the decomposition of plastics, but they are the first group of compounds that can be examined for correct intergroup ratios. Figure 13 shows ion 91, the base ion for xylenes, from gasoline in three different stages of evaporation, kerosene, and a medium petroleum distillate (MPD). The relative ratios for the three peaks are almost indistinguishable. Note that the ethylbenzene peak in the 50% evaporated gasoline is slightly lower. This trend continues as the degree of evaporation increases, but generally, we will find that if the xylenes in our sample are from a petroleum product, they will have this ratio. Note that there are three isomers of xylene, ortho-, meta-, and para-, but that the three peaks seen in these chromatograms actually represent four compounds, because meta- and paraxylene cannot be resolved except in the longest columns. If xylenes are found in ratios other than the one shown in Figure 13, particularly if ethyl benzene is the tallest peak in the group, one can safely conclude that neither gasoline nor any other petroleum product was the source of the xylene.

The next group to consider is the C_3 alkyl benzenes. This is by far the most important of the patterns in any sample of gasoline. Like the xylenes, it is also found in all petroleum products from which the aromatics have not been removed, and unless evaporation has decreased the concentration of the lighter compounds, the peak ratios will always be the same. This group seems to show up in many samples that contain no gasoline, and it is the failure to exhibit the proper ratio of peaks that is sometimes the analyst's only clue that he is looking at a 'difficult' sample. Figure 14 shows just one such sample, as well as four standards that exhibit the proper ratios

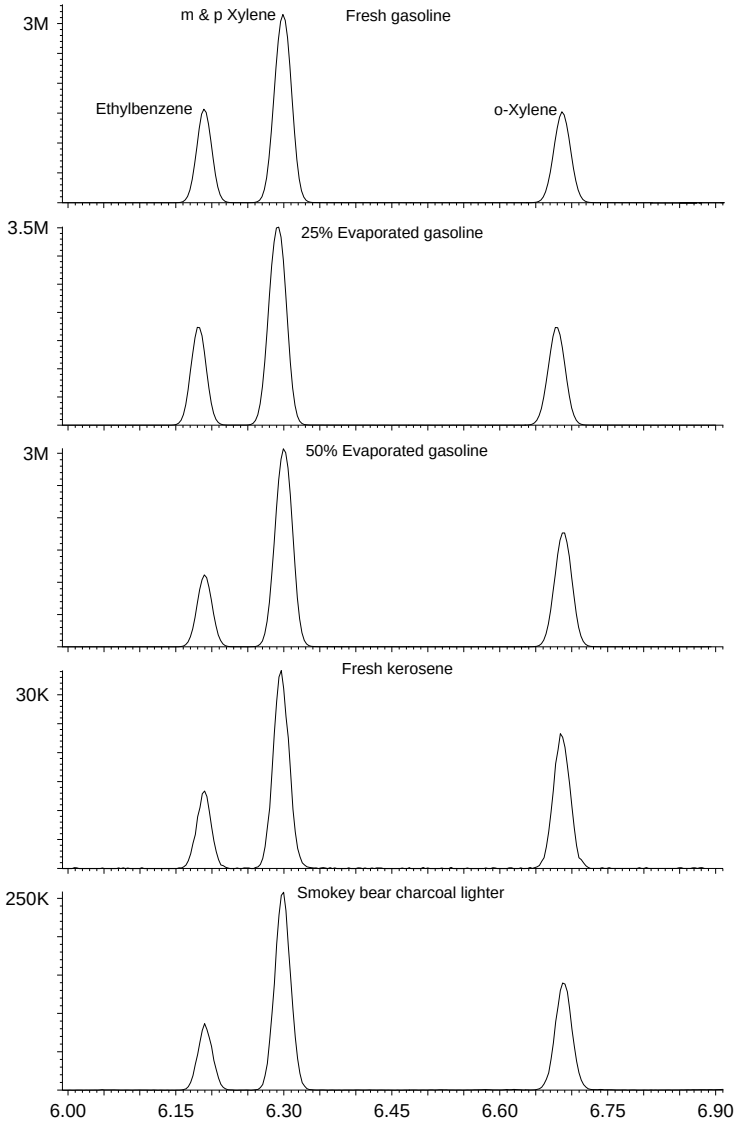


Figure 13 Ion chromatogram for ion 91 from gasoline in three different stages of evaporation, kerosene, and a medium petroleum distillate. Note that the relative ratios for the three peaks in the xylene group are almost indistinguishable. Evaporation causes the ethylbenzene peak in the 50%-evaporated gasoline to be slightly shorter

of components in the C_3 alkyl benzene group. As with the xylenes, meta- and paraethyltoluene cannot be resolved, although the resolution is somewhat better here than in the xylenes. In nearly all cases where the C_3 alkyl benzene group is, in fact, present as a component of a petroleum product, the pseudocumene peak will be the tallest. In fact, in almost all samples found positive for gasoline, because most of the residues encountered will be highly evaporated, the pseudocumene peak will be the tallest peak on the chart.

Note that in the questioned sample at the top of Figure 14(a), it is the second peak that is the tallest, not the pseudocumene peak. The first peak has the same retention time as *m*- and *p*-ethyltoluene, but the shape is wrong for petroleum, and the mass

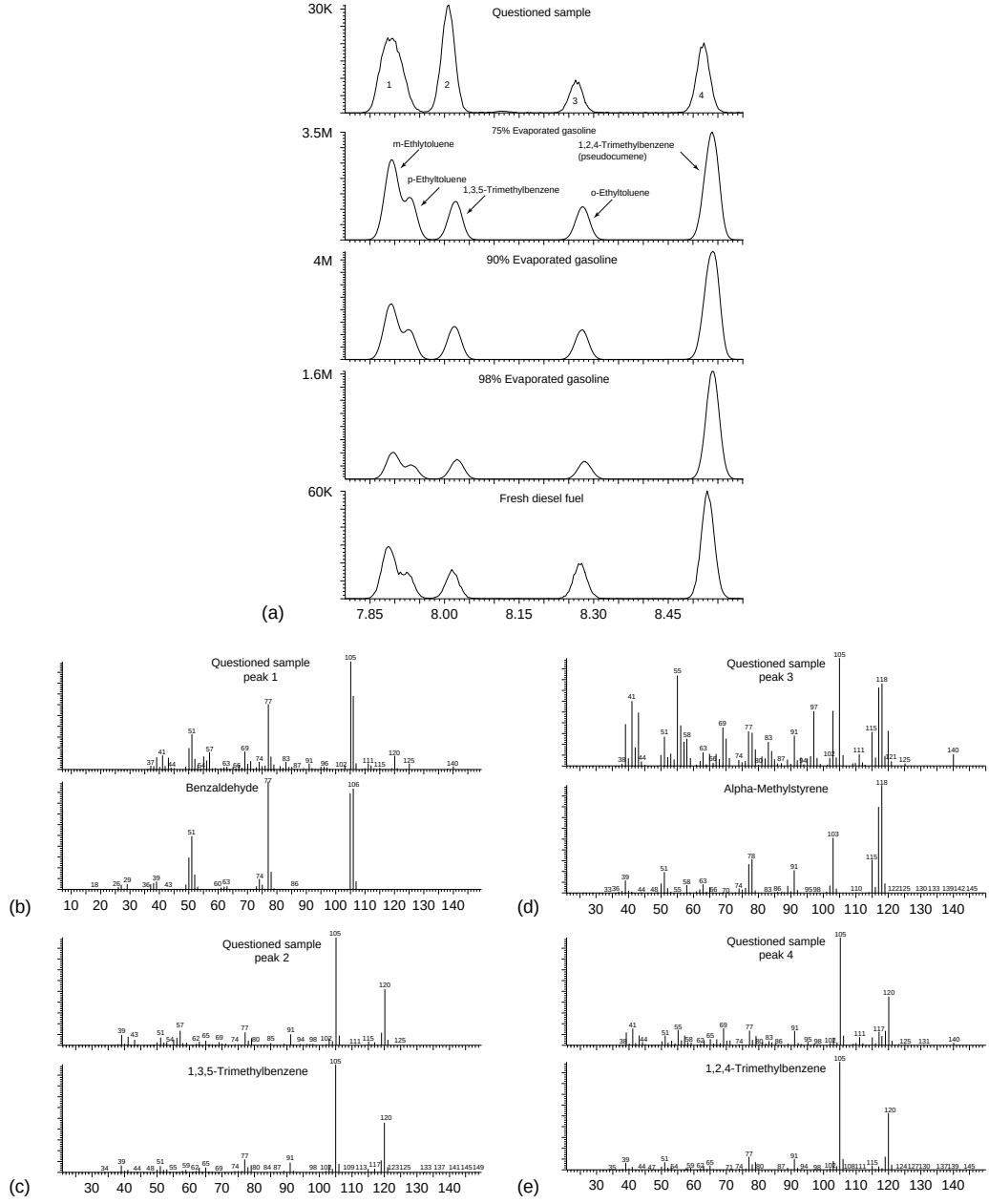


Figure 14 A study of the C_3 alkylbenzene group. Part (a) shows an unknown sample (top), gasoline in three stages of evaporation, and fresh diesel fuel. The TIC did not look like gasoline, but based on this and a few other ion chromatography comparisons, an analyst called the *sample positive for gasoline*. The mass spectra shown in (b) through (e) showed that two of the peaks, #1 and #3, were not gasoline components. (b) Mass spectrum and library match of the first peak in the top chart shown in Figure 14(a). This peak actually represents benzaldehyde, and not *m*-ethyltoluene. (c) Mass spectrum of the second fully resolved peak in the top chart shown in (a). As with gasoline, this peak is identified as 1,3,5-trimethylbenzene. (d) Mass spectrum of the third major peak in the chart shown at the top (a). This mass spectrum changed across the peak, but clearly contained α-methylstyrene. 1-decene, a common decomposition product, was also suggested by the spectrum. There may be some *o*-ethyltoluene co-eluting as well, but that cannot be demonstrated. (e) Comparison of the mass spectrum fourth peak in the chart shown in (a) with 1,2,4-trimethyl benzene (pseudocumene). This is a reasonably straightforward match

spectrum indicates that the peak actually represents benzaldehyde. The mass spectrometer unequivocally identified the second peak in the questioned sample as 1,3,5-trimethyl benzene, but the third peak, which may have represented *o*-ethyltoluene, also contained a significant 118 ion, indicating the presence of α-methylstyrene, a common pyrolysis product.

Simply matching the components when the ratios are not right can lead to misidentifications. The odds against three background components co-eluting in exactly the right concentrations to skew the peaks in this group are pretty high. Three other samples from the same fire scene were misidentified, and exhibited similar peak ratios and mass spectral characteristics. As with almost all misidentifications, there was not a good pattern match with the TIC. Extracted ion chromatography or extracted ion profiling can be very useful; however, the analyst should remember that it is a spectral as well as a chromatographic technique. The spectra should be examined, especially when the peak ratios are not the same as seen in the standard. A word about the use of the mass spectrometer beyond generating mass chromatograms is in order.

Most analysts use the mass selective detector or mass spectrometer as a tool for generating EIC and EIP. Obviously, if one is looking at very simple mixtures or single components, the mass spectrum is necessary in order to make an identification. A sometimes-overlooked function of the mass spectrometer is the evaluation of EIC and profiles. As was shown in Figure 14(a), an EIC with the intra-group ratios just slightly "off" has the potential to be misleading. Figures 14(b) through (e) shows the mass spectra of each of those five peaks, as well as the library's best match. The first peak, which is coincident with *m*-ethyltoluene, has a large peak at m/z 77, but no such ion is present in the spectrum of *m*-ethyltoluene. There may be some *m*-ethyltoluene hidden under this peak, but there is definitely some benzaldehyde as well.

One way to tell whether there are co-eluting compounds under a chromatographic peak is to examine the mass spectra at different points across the peak. If it is a pure compound, the mass spectrum will change little, if at all. In Figure 14(c), the second peak in the five-peak group is a near-perfect match for 1,3,5-trimethylbenzene, which is the third peak in the five-peak group required to identify gasoline. The third peak, whose mass spectrum is shown in Figure 14(d), clearly contains more than one substance, but the strong peak at 118, as well as the retention time, tell us that there is α-methyl styrene co-eluting. Like the second peak, the fourth peak is a pure compound, and a nearly perfect match for pseudocumene. Because the peak ratios are off, and especially because two of the four peaks in the ion 105 chromatogram represent compounds not found in gasoline, it must be concluded that the first identification was in error.

Background subtraction is a tool found in most data analysis software, and can frequently resolve questions as to whether a particular compound is actually present. A detailed evaluation of the quality of the spectra underlying a mass chromatogram, whether it is a single ion extraction or a multiple ion profile, should be carried out periodically, but it should routinely be carried out on any sample where either the peak ratios or retention times are "just a little off."

Note in Figure 14(a) that when a sample is sufficiently evaporated, the peaks at the left side of the chart begin to diminish. This is not an unexpected result. The initial guidelines for identifying gasoline stated that the *m*-ethyltoluene/pseudocumene five-peak group was still present in gasolines that had lost 90% of their fresh weight. Actually, this group does not disappear until the gasoline is more than 98% evaporated. If this group cannot be positively identified, an analyst is on very thin ice indeed when identifying a sample as containing gasoline. In such a case, there must be a peak-for-peak match of all of the higher peak groupings.

As with many of the other components of gasoline, this 5-peak group is present in almost all petroleum products from which the aromatics have not been removed. The bottom chart in Figure 14(a) is from unevaporated diesel fuel.

The C_4 alkyl benzenes are best viewed by extracting ion 119. Because there are more ways to build a C_4 alkyl benzene than a C_3, this is a more complex pattern, but it is present in almost all petroleum products. Figure 15 shows the patterns found in highly evaporated gasoline, as well as in kerosene and diesel fuel that are produced by the presence of the C_4 alkyl benzenes. Note that in kerosene and diesel fuel, the pattern is more complex than in gasoline.

The next group of compounds that need to be present in order to make a solid identification of gasoline is the indans. These can be extracted using ions 117 and 131. The doublet at 11.3 and 11.5 min in the

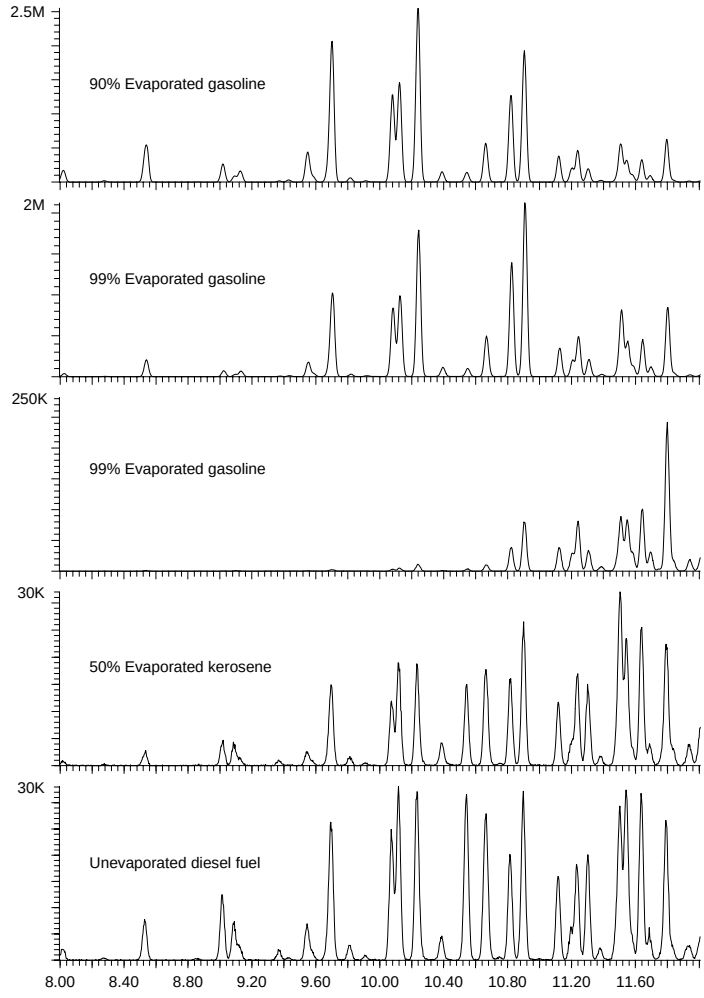


Figure 15 C_4 alkylbenzenes represented by ion 119 in gasoline in three different stages of evaporation and unevaporated diesel fuel. All samples were prepared by spotting 10 µl on a piece of filter paper, and processed using ACS adsorption/elution

ion 117 chart is, in this analyst's experience, always present. It has traditionally not been an absolute requirement of the ASTM standards, but it probably should be. The standard states, "Indan (dihydroindene) and methylindans are usually present." The first tall peak in the 117 chart is indan, appearing at 9.4 min. Figure 16 shows the peak groupings characteristic of gasoline for the indans, methylindans, and dimethyl indans. Like the C_3 alkyl benzenes, the indans and alkyl-substituted indans can be found in roughly the same proportions in gasoline that is 98% evaporated. Other petroleum products, particularly the distillates, contain indans and methylindans, but their mass chromatographic patterns are more complex than the ones for gasoline.

The next group of compounds that should be present in a sample identified as containing gasoline is the methyl- and dimethylnaphthalenes. Naphthalene is also present, but naphthalene is so common that its presence in a sample is all but meaningless. 2-Methylnaphthalene elutes before 1-methylnaphthalene, and is almost always more abundant than the 1-methylnaphthalene. In any case where this ratio is reversed, the extract should

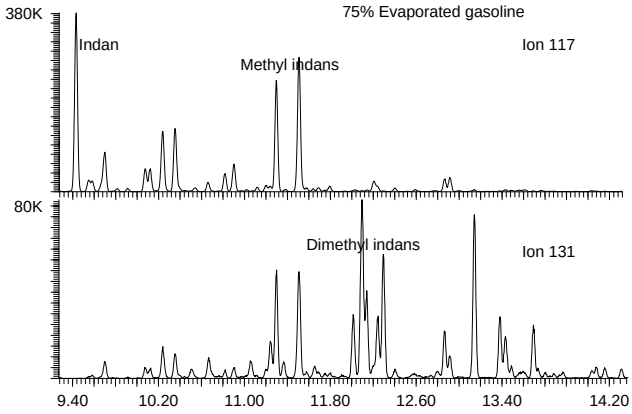


Figure 16 Ion chromatograms for indan, the methylindans, and dimethylindans found in 75%-evaporated gasoline

be considered suspect. The boiling point of the methylnaphthalenes is high enough that the ratio, unlike xylenes or C_3 alkyl benzenes, is unlikely to be affected by evaporation. All petroleum products are likely to contain both the naphthalenes and the dimethylnaphthalenes, and the ratios should be highly comparable.

There are some aromatic products in the marketplace that meet almost all of the ASTM criteria for the identification of gasoline. Such products are used as solvents for stains, insecticides, and industrial and commercial products. The way to differentiate these aromatic products from gasoline is to look for the presence of alkanes. All gasolines contain high percentages of alkanes when fresh, but even highly evaporated gasolines do contain some branched alkanes. Straight-chain hydrocarbons are an unusual finding in gasoline, because they are an undesirable component in that they cause knocking when burned in gasoline engines. Knocking is the premature detonation of a fuel, and occurs more readily with straight-chain hydrocarbons than with branched-chain hydrocarbons or aromatics. Resistance to knock is known as the *octane rating*. Octane rating is based on a scale of 0–100, where 0 is the resistance to knock of 100% *n*-heptane and 100 is the resistance to knock of 100% iso-octane (2,2,4-trimethylpentane). *N*-octane actually has an octane rating below zero.

Identification of Distillates

Most distillates that are likely to be encountered in fire debris are straight-run distillates from crude oil. They have not been subjected to cracking or reforming, so they lack the high aromatic content found in gasoline, but there are still aromatics present. The distillates are characterized by an abundance of normal alkanes, but there are also branched alkanes and cycloalkanes present.

Distillates are usually easy to find, except for light petroleum distillates (LPDs), which might be missed unless the analyst checks for them in every sample before calling it negative. Because of their high volatility, they tend to be present at low concentrations in fire debris samples, and are typically crowded against the left side of the chart, where they may be mistaken for decomposition products. This is a particular hazard when the higher boiling components are decomposition products. LPDs do not generally exhibit the Gaussian distribution so frequently seen in MPDs and HPDs. When the peaks that elute prior to eight min are examined, and the analyst sees a mixture of cycloalkanes, branched alkanes, and perhaps some normal alkanes in the C_7 through C_9 range, an LPD should be suspected. The analyst should have in the library as many LPDs as possible, because the patterns tend to vary from one to the other, unlike the heavier distillates. Figure 17(a) shows three different brands of cigarette lighter fluid, each exhibiting a different pattern of peaks. The recommended EIP for LPDs are 57, 55, 83, and 91, as can be seen in Figure 17(b). This particular LPD is unusual in that it contains xylenes, but no toluene is present. Because of their high volatility, LPDs are not persistent in the environment, and are not usually expected to be found as background material. There are some

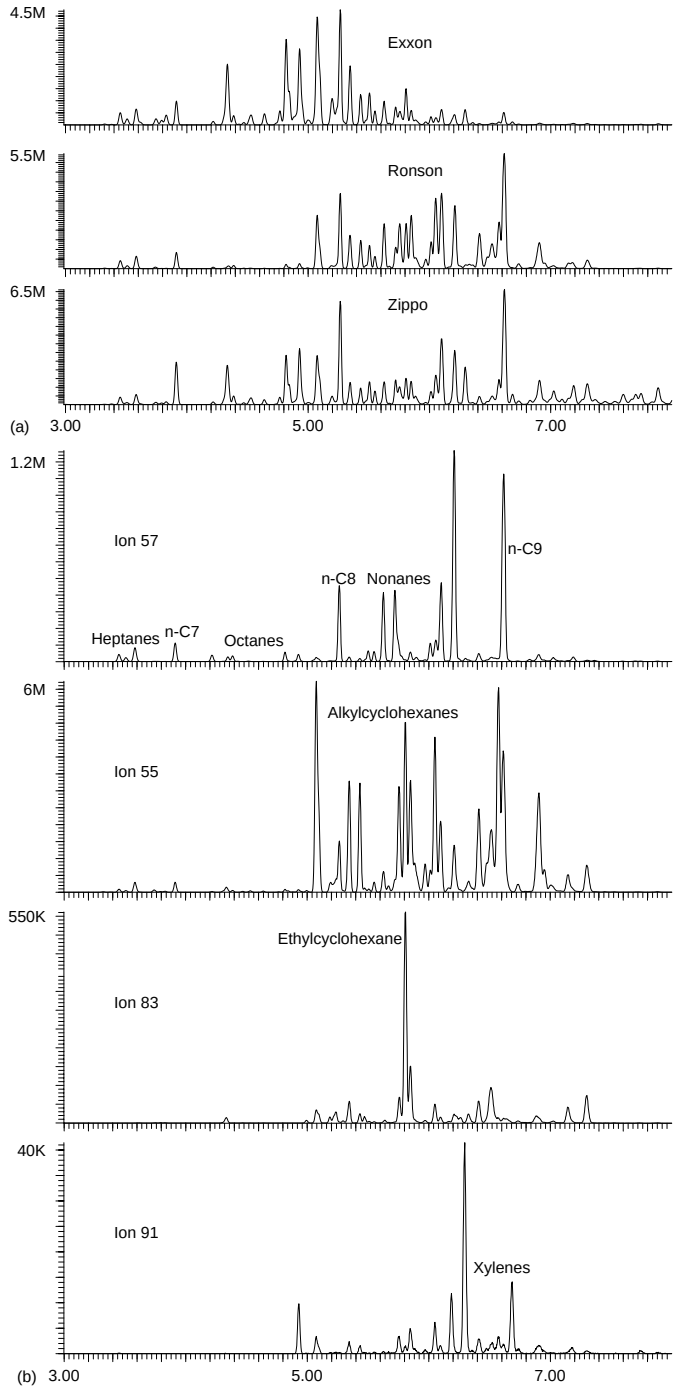


Figure 17 (a) Total ion chromatograms of three different brands of cigarette lighter fluid. (b) Ion chromatograms of ions 57, 55, 83, and 91 from a standard of Ronson Lighter Fluid

cleaning agents and automotive products that contain LPDs, but if the sample was collected from somewhere other than the garage, the workshop, or under the kitchen sink, a finding of LPD generally indicates the presence of a foreign ignitable liquid.

This is not the case with medium or HPD, which are far more common in our environment [16]. A typical medium petroleum distillate, Sparky charcoal lighter fluid, is shown in Figure 18(a). The ions necessary to make a good identification of a medium petroleum distillate are 57 and 83, shown in Figures 18(b) and (c). The analyst should also check for ion 105 or 91, in order to be certain that the aromatics have not been removed, in which case, a proper description of the residue would be a de-aromatized distillate. (De-aromatized distillates were once a separate ASTM class, but difficulty in defining the cutoff percentage caused the committee to discontinue the use of this class in 2006.) One does not notice much difference between the TIC and ion 57, because distillates are dominated by the normal and branched alkanes. The alkylcyclohexanes, which elute about halfway between the normal alkanes, present an overall appearance similar to that of the alkane chromatogram. If the normal alkanes are not present, or if they are present at approximately the same concentration as the branched alkanes, an isoparaffinic product should be suspected. When the cycloalkanes are present at a concentration greater than about 25%, a naphthenic–paraffinic source is indicated. In most distillates, the cycloalkanes will be present at about 5–10% of the concentration of the normal alkanes.

Medium petroleum distillates cover a wide range of products, and may be used as fuels, such as lamp oil or charcoal starter, or as solvents, such as mineral spirits or insecticide carriers. This wide range of products and formulations requires that the library of reference materials include numerous MPDs.

HPDs include kerosene and diesel fuel. Except for HPDs that are formulated for specific applications, such as jet fuel, the carbon number range of HPD can vary. Thus, unless one has a sample of the unevaporated liquid, it is difficult to determine the degree of evaporation of the residue isolated from a sample. In northern parts of the USA, diesel fuel sold in the winter may actually be kerosene.

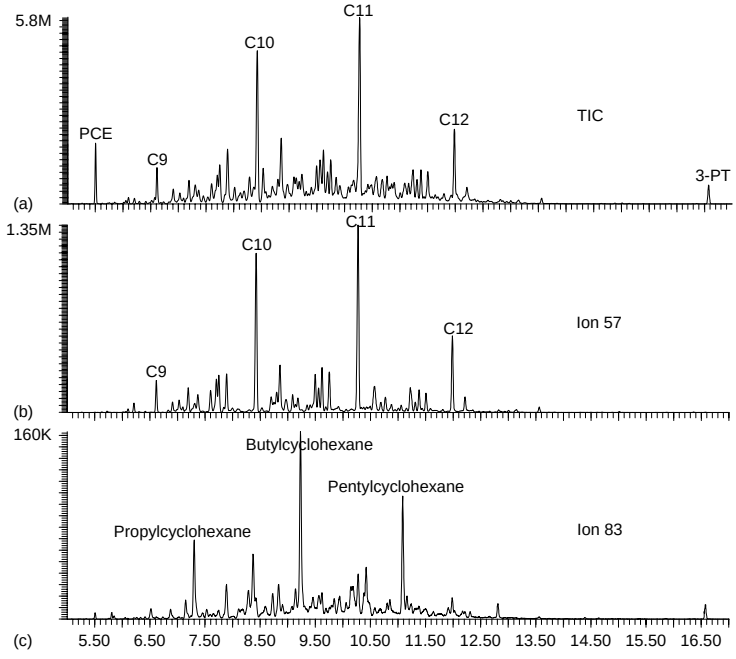


Figure 18 Total ion chromatogram of a typical medium petroleum distillate, Sparky Charcoal Lighter Fluid (a), and ion chromatograms of ion 57 (b) and ion 83 (c)

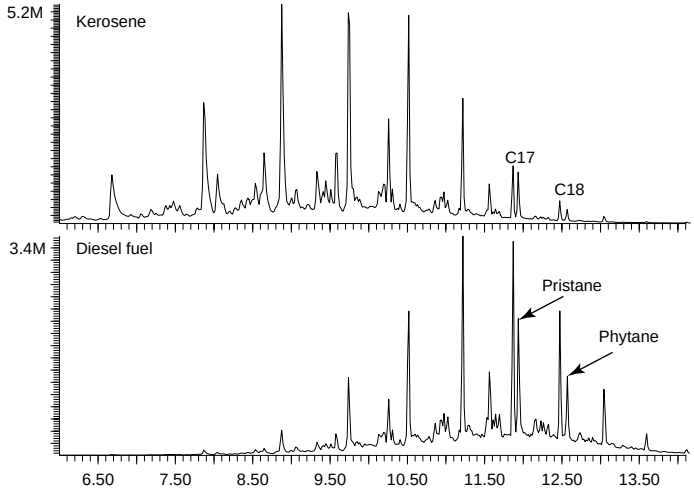


Figure 19 Total ion chromatograms showing a comparison of kerosene and diesel fuel

Figure 19 shows a comparison of kerosene and diesel fuel. These distillates come with built-in carbon number markers, in the form of pristane and phytane. Pristane is 2,6,10,14-tetramethyl pentadecane, $C_{19}H_{40}$, and it elutes immediately after normal heptadecane. Phytane is 2,6,10,14-tetramethyl hexadecane, $C_{20}H_{42}$, and it elutes immediately after normal octadecane. Thus, one looks for the two doublets on the high side of the bell-shaped curve and can count up and down from there. In Figure 19, these doublets occur at 12 and 12.5 min. One can see that the kerosene, which is a known standard that has been evaporated to 50% of its original volume, ranges from C_{11} to C_{19}, while the diesel fuel range is from C_{12} to C_{21}. If the kerosene is evaporated further, it will look more like the diesel fuel. Had the diesel fuel been evaporated less, it would look more like the kerosene.

As with the medium petroleum distillates, interpretation of a finding of HPDs should be approached with caution. There are numerous household products that contain HPDs, including many of the same kinds of products in which MPDs are found. The safer charcoal lighters are made from kerosene rather than mineral spirits. They are safer because of their higher flash point. Figure 20 shows a TIC of lemon oil furniture polish. This polish has a carbon number range from C_{12} to C_{22}, with C_{17} being the tallest peak on the chart. The limonene peak at the left side of the chart could easily be attributed to a pine substrate. Pinenes and limonene are very common in samples containing structural (coniferous) wood. This particular furniture polish, if found in a fire debris sample, could easily be reported as diesel fuel.

It is not even necessary for an actual liquid to be present in order for a distillate to be detected. Figure 21(a) shows the TIC of a piece of pine wood that was stained with Minwax® finish 10 months before it was subjected to headspace concentration. The naturally occurring terpenes, α- and β-pinene and d-limonene, are the dominant peaks on the chart, but the mineral spirits solvent is still clearly visible. When ions 57 and 83 are extracted, the terpenes disappear, to yield the charts shown in Figures 21(b) and (c). These results show the critical necessity of asking for comparison samples, particularly when samples of flooring are submitted for analysis. The flooring does not need to be recently painted in order to exhibit the distillate solvent used to apply the floor coating. This author has reported finding distillates 24 months after application [17] and has detected distillates in samples of finished flooring and furniture up to ten years old. There is no reason to believe that these solvents do not persist indefinitely, trapped in either the wood matrix or in a polymer coating matrix.

In addition to the distillates that are present as distillates, there are some distillate-like residues produced as the result of decomposition of other products. Asphalt is what is left at the bottom of the distillation pot after all of the volatiles have been

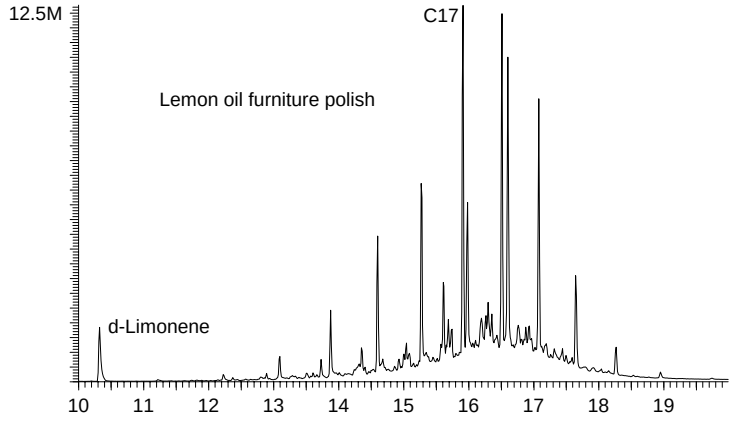


Figure 20 Total ion chromatogram of lemon oil furniture polish. This sample could easily be reported out as diesel fuel

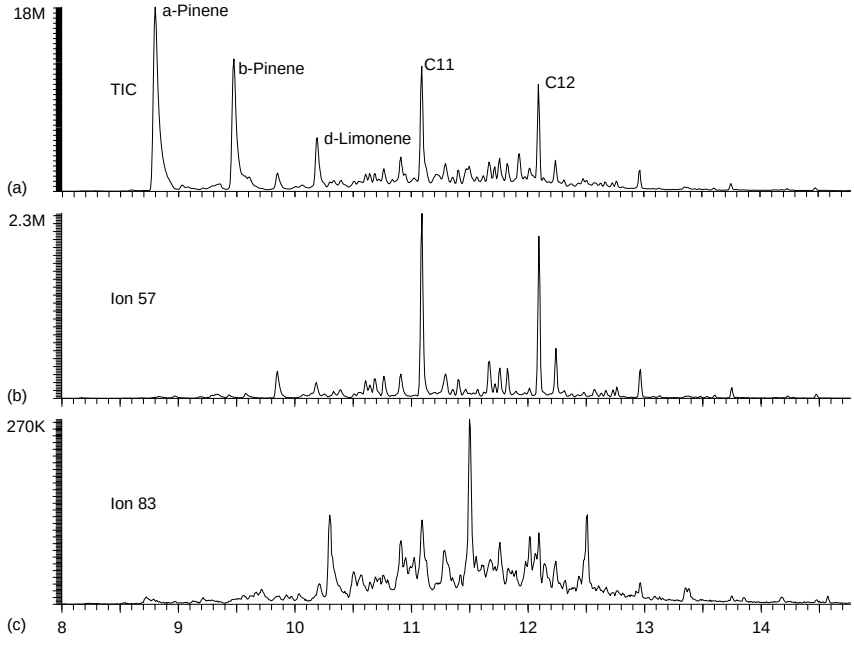


Figure 21 Total ion chromatogram of a piece of pine wood stained with Minwax® Finish 10 months prior to its analysis (a), and ions 57 and 83 from the same sample (b and c)

distilled from crude oil. It contains hydrocarbons ranging from C_{30} to C_{60}. When these long-chain hydrocarbons undergo pyrolysis, they do so in much the same way as the long-chain hydrocarbons in polyethylene, i.e., via random scission. This results in the presence of normal alkanes in the range of C_9-C_{18} [18]. When these pyrolysis products are present in fire debris that is subjected to headspace concentration, they produce a chromatogram that can be and has been mistaken for the chromatogram of a HPD. Such a chromatogram is shown in Figure 22(a). In 1982, it was reported that roof shingles could produce "accelerant-like residues," and at that time, there were no methods available for distinguishing HPDs

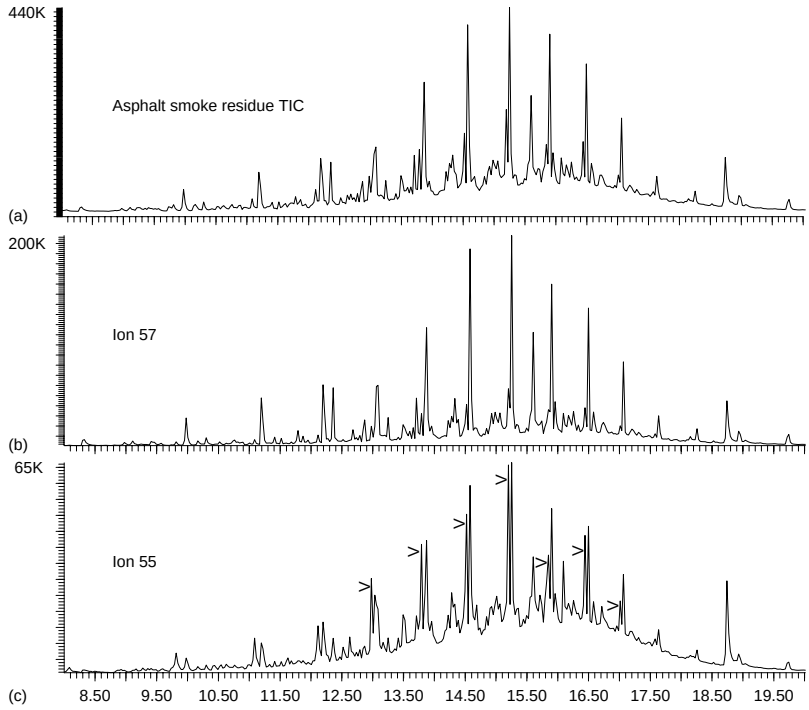


Figure 22 (a) Total ion chromatogram of asphalt smoke residue, of the type that can be mistaken for a petroleum distillate (b and c). Comparison of ions 57 and 55 from the asphalt smoke residue shown in the TIC. The growth of the peak in front of the normal alkane when comparing 55–57 demonstrates the presence of alkenes, which allows the identification of the residue as a decomposition product rather than a foreign petroleum distillate

from asphalt shingle residues [19]. The increased use of capillary columns allowed for the occasional visualization of a double-peak kerosene, sometimes called *pseudo kerosene*, but even with capillary columns, it was not uncommon for asphalt residue to be misidentified as a liquid petroleum distillate. This author learned how to make the differentiation in 1995 in connection with the investigation of an insurance claim that had been erroneously denied because of a finding of HPDs where none should have been present. Figures 22(b) and (c) show how the distinction is made. One compares ion 57 with ion 55. If a second peak appears in front of the *n*-alkane peak, or if a small peak grows larger, one can conclude (particularly after examining the mass spectrum) that the second peak is the 1-olefin. A sample such as this must be classified as asphalt smoke condensate or asphalt decomposition residue, and unless the possibility that asphalt residue got into the sample can be categorically eliminated, this material must be identified as being native to the scene.

The amount by which the olefin peak "grows" will vary depending on the sample. When one looks at kerosene and diesel fuel, the only difference between the appearance of the ion 55 chart and the ion 57 chart is that the abundance of ion 55 is lower. The relative abundances of the individual peaks with respect to each other do not change.

Another way to distinguish asphalt decomposition products from HPDs is to look for the cycloalkanes. They are not present. There will be a Gaussian pattern observed when ion 83 is extracted, but this is due entirely to the presence of the olefins. The ion 83 and ion 55 ion chromatograms show the same peaks. Asphalt smoke condensates also contain little, if any, pristane and phytane.

Polyethylene is another substance that produces a distillate-like appearance in the chart, but if the capillary column has any resolution at all, it will be obvious that one is looking at polyethylene residue, rather than at a distillate. Figure 23(a) shows the TIC of polyethylene smoke condensate. All of the peaks

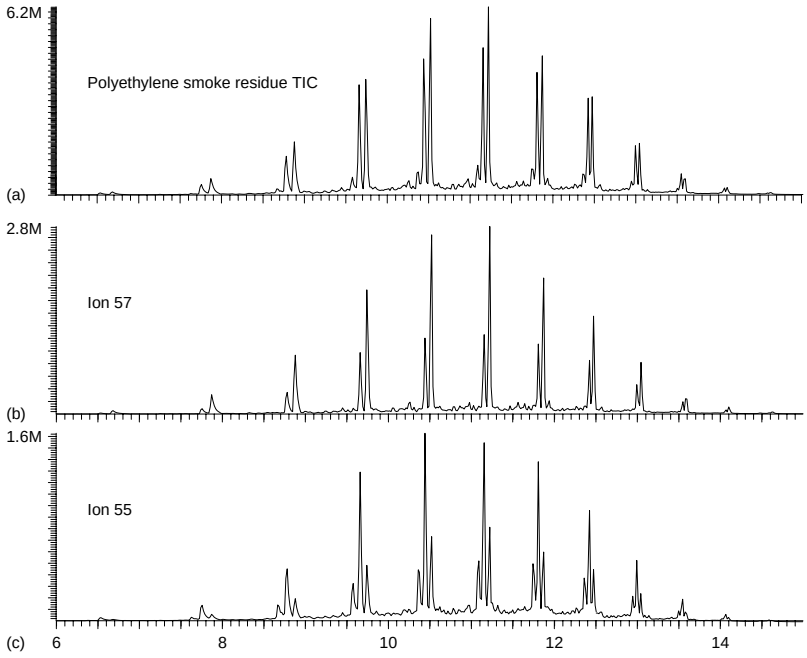


Figure 23 (a) Total ion chromatogram of polyethylene smoke residue. The first peak in each doublet is the alkene and the second is the alkane. Pristane, phytane, cycloalkanes, and aromatics are all absent. Ion 57 (b) and 55 (c) chromatograms of polyethylene smoke residue. In the ion 55 chromatogram, a third peak, representing the n, $(n-1)$-diene appears

are doublets and most are actually triplets, owing to the presence of the 1,$(n-1)$-diene in the mixture. When one looks at ion 57 and compares it with ion 55, as shown in Figure 23(b), the growth in the olefin peak is obvious, in that it becomes taller than the alkane peak. The diene peak also grows, because the dienes contain more ion 55 than ion 57. Lubricating oils are subject to the same decomposition processes as asphalt and polyethylene.

A careful examination of the chromatographic and mass spectral data will prevent the analyst from misidentifying decomposition products as distillates, but automation of pattern recognition presents some dangers if not handled properly. With all computer "answers," one should always do a "reality check." The same pattern recognition software that allows a mass spectrum from an unknown compound to be matched against the spectra of 100,000 compounds can be applied to chromatograms. The chromatogram is converted using Microsoft Excel into a bar graph that has the same general appearance as a mass spectrum. This graph can then be compared against a library of known ignitable liquids that have been similarly transformed. Like the mass spectral libraries, "extra" peaks do not necessarily keep the library from recognizing a "match." The polyethylene smoke condensate chart shown in Figure 23(a), matched up against a library that contains only ignitable liquids, will yield a match of exceptionally high quality for diesel fuel. This is why it is necessary to populate ignitable liquid libraries used for this purpose with known background chromatograms. The chart would make an even better match for polyethylene smoke condensate, but only if it is in the library.

Identifying Other Classes of Products

The remaining classes of ILR can frequently be identified by what is not present, as well as what is. A de-aromatized distillate, for example, will have a signal from the aromatics that is less than 1% of the signal from the aliphatics. Otherwise, it will look the same with respect to both the alkanes and the cycloalkanes. Because of environmental regulations governing the aromatic content of distillates there

has been, since 1973, an ASTM standard for this determination. ASTM D 3257-01 is entitled *Standard Test Methods for Aromatics in Mineral Spirits by Gas Chromatography* [20], and uses a specified test blend for calibration. If the calculation of aromatic content is an issue, forensic scientists might avoid a *Daubert* challenge by using this established method, rather than devising a new one.

The normal alkane products are very easy to recognize, as long as one makes sure that one is looking at a homologous series of normal alkanes as opposed to a homologous series of aldehydes or some other group of homologs that are pyrolysis or decomposition products. Groups of compounds that differ from each other only in that they have one additional CH_2 group look pretty much the same as a series of normal alkanes, and require some caution. Figure 24 shows a series of normal alkane products marketed by Exxon, as well as a sample of Lamplight Farms Ultra Pure Lamp Oil®, which also consists of normal alkanes. One common source of normal alkanes is carbonless forms (documents), and the bottom chromatogram in Figure 24 shows the chromatogram of concentrated headspace vapors from a 5 cm × 5 cm square of a carbonless form. These forms contain microspheres filled with the normal alkane solvent that, when broken, causes the color to develop in the ink. There are only a few microliters per square foot, but this concentration is easily detectable. Figure 25 is a scanning electron micrograph, showing the bottom surface of a carbonless form. Normal alkane products are also found in some brands of linoleum floor covering. Any time a sample is collected from a floor likely to have linoleum in its structure (kitchen, bathroom, laundry room), a finding of normal alkanes is probably not meaningful. A comparison sample is an absolute necessity in such cases.

Isoparaffinic hydrocarbons are made by removing the normal hydrocarbons with a molecular sieve. These liquids are becoming more common as petroleum refiners move to more environmentally friendly and odor-free replacements for the straight-run distillates such as mineral spirits. Similar to the isoparaffinic hydrocarbons are the naphthenic–paraffinic products, which are characterized by an abundance of cycloalkanes. Whereas one might expect to find cycloalkanes present at less than 5% in the isoparaffinic hydrocarbons, they may be present at up to 30% in naphthenic–paraffinic products. The distinction requires looking at the ion profiles for the cycloalkanes versus the alkanes, as well as looking at the abundance numbers on the left side of the chart. Figure 26 shows a comparison of the ion 57 and ion 83 chromatograms from Isopar H®. Contrast this with Figure 27, which compares those same ion chromatograms from a naphthenic–paraffinic solvent, Vista LPA 170®. Instead of having an ion 83 chromatogram that is 2% the height of the ion 57 chromatogram, in the naphthenic–paraffinic solvent, the ion 83 is more than 30% of the height of the ion 57 chromatogram. Also, note that the tall peaks in the ion 57 chromatogram from the naphthenic–paraffinic product are not normal alkanes, but are branched alkanes.

Another way to make the distinction between the isoparaffinic products and the naphthenic–paraffinic products is to take an average mass spectrum. Representative isoparaffinic hydrocarbons are shown in Figure 28. A mass spectrum of any of these looks very much like the mass spectrum of a single alkane. Ion 57 is the base peak, and the other fragments are spread out in a Gaussian distribution with a spacing of 14 mass units between them. A typical isoparaffin average mass spectrum, that of Isopar L®, taken over a 3-min portion of the chromatogram, is shown in Figure 29.

The same exercise can be performed on the naphthenic–paraffinics, several of which are shown in Figure 30. The average mass spectrum of a typical naphthenic–paraffinic product, shown in Figure 31, looks quite different from that of the isoparaffinic product. While it still exhibits a base peak of 57, it will be very rich in 55, 69, and 83.

The identification of other ILR requires an individual examination of the peaks in the chromatogram. It is possible to find just about any compound in just about any sample. Findings of alcohols, turpentines, aromatic solvents, and other “flammable” liquids need to be viewed with great caution. ASTM E 1618 recommends not making an identification of these single compounds unless they are present in such concentrations that the signal is at least 2 orders of magnitude greater than the background. Just about every fire debris sample will contain methanol, though most laboratories do not routinely check for methanol. Just about every fire debris sample will contain toluene [21]. Unless these substances are present at concentrations sufficiently high for the

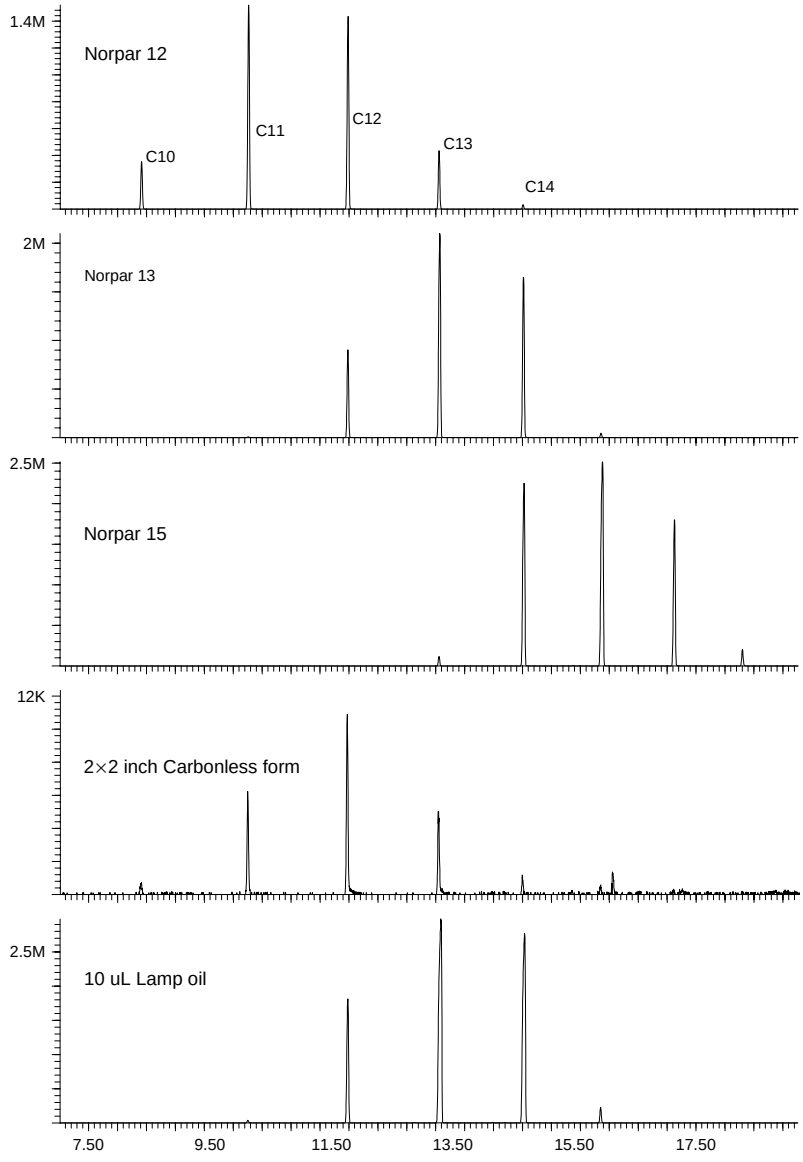


Figure 24 Ion 57 chromatogram of three Exxon Isopar® products, compared with an ACS extract from a carbonless form, and the ACS extract of a 10 μl sample of Lamplight Farms® Ultrapure Lamp and Candle Oil

analyst to feel comfortable saying they are not native to the background, they should not be reported.

Identity of Source

The ability to match a weathered ILR with a proposed source of unburned ignitable liquid has eluded fire debris analysts until just recently. The environmental forensics community has developed some tools to identify the sources of petroleum spills, but they have the advantage of having huge quantities of material available for characterization. Petroleum hydrocarbons contain trace quantities of biomarkers, substances that have changed little since they were first

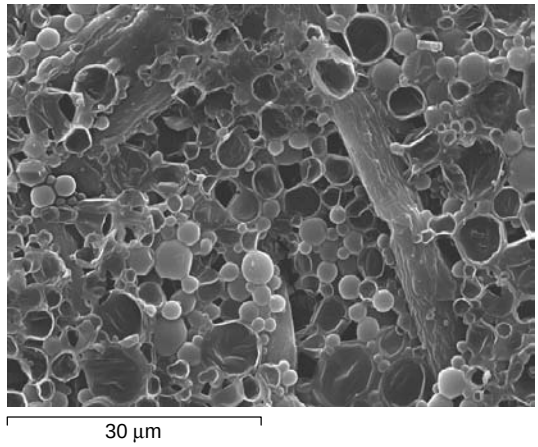


Figure 25 Scanning electron micrograph of the underside of a carbonless form. The microspheres range in size from 2 to 20 µm, and are filled with normal alkanes

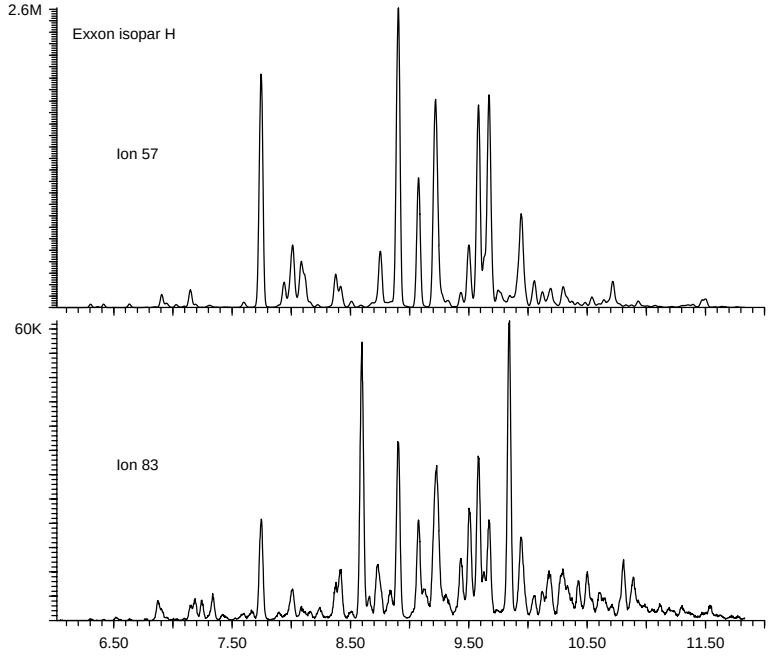


Figure 26 Comparison of the alkane (ion 57) and cycloalkane (ion 83) chromatograms from Exxon Isopar H®. The cycloalkane peaks are only about 2% of the height of the branched alkane peaks

synthesized inside a living organism. The relative amounts of these biomarkers frequently allows for the identification of the source of a major spill. Polycyclic aliphatic and aromatic hydrocarbons are also useful in "fingerprinting" hydrocarbons. Examination of additives, such as oxygenates, or in the case of old gasolines, the alkyl lead compounds, can also provide clues as to the identity of the source of a spill. There are three levels of influence on the composition of petroleum. The primary influence is the genesis of

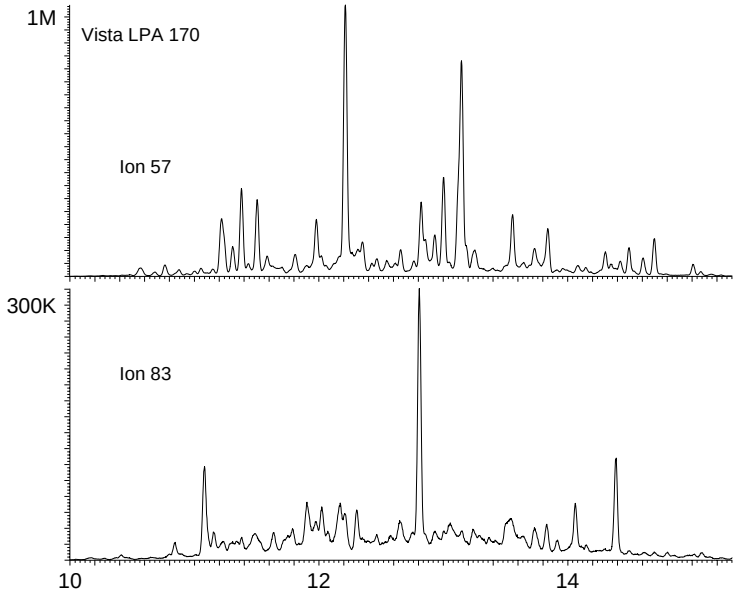


Figure 27 Comparison of the alkane (ion 57) and cycloalkane (ion 83) chromatograms from Vista LPA 170, a naphthenic–paraffinic product. The height of the cycloalkane chart is about 30% of the height of the branched alkane chart

the petroleum, i.e., including its geographic location, the original source of the oil (terrestrial or marine), and the conditions in the rock where the oil resided for millions of years.

The second level of influence affecting the hydrocarbon fingerprint is processes imposed on the hydrocarbon by the refiner. Activities such as distillation, hydrocarbon cracking, isomerization, and alkylation change the composition of the fuel.

The tertiary influences are those that occur after the petroleum product leaves the refiner. The most important tertiary influence that is of interest to fire debris analysts is, of course, weathering, or evaporation. Mixing in the service station tank is also critical. Because the tanks are seldom empty, each time a new delivery is made, the fuel service station tank assumes a new, temporary, identity. For a more extended discussion on chemical fingerprinting in the environmental forensics arena, see *Introduction to Environmental Forensics*, reference [22].

Some of the most comprehensive work to identify the source of ILR from fire debris was done by Dale Mann in 1987 [23, 24]. Using a 60-m column, Mann learned that it was possible to make comparisons on samples of fresh gasoline, and correctly identify the source by examining the relative ratios of the light hydrocarbons, ranging from *n*-pentane to *n*-octane. It is, unfortunately, exactly these hydrocarbons that are lost before a gasoline is even 25% evaporated. Mann's second paper describes the limitations of making comparisons between ILR and fresh samples, because of the contamination of the residues with pyrolysis products, changes introduced by the isolation method, and the loss of volatiles resulting from weathering. Comparison of residues with fresh sources was, therefore, not frequently studied after Mann's extensive work. Recently, however, Dolan and Ritacco reported that they had devised a way to measure the relative abundances of 20 peak pairs in gasolines, and were able to identify the source of 30 samples that were evaporated to 25 and 50%. This process uses the relative abundances of branched alkanes that occur generally in the center of the gasoline chromatogram. Because the peaks are sequential, eluting only a few seconds apart, the relative ratios are not significantly affected by evaporation. The peaks examined are all minor components. The major components of gasoline are similar enough to each other that they are not useful in discriminating between sources [25].

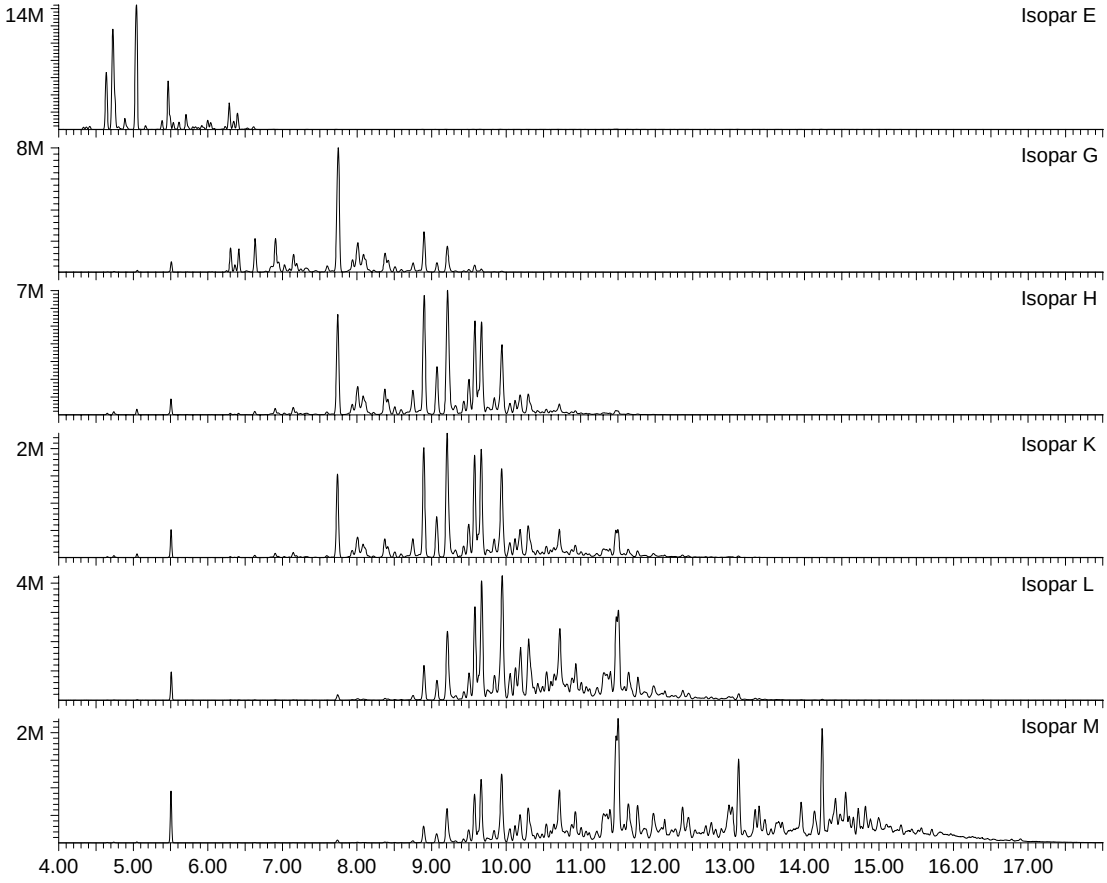


Figure 28 Total ion chromatograms of six isoparaffinic hydrocarbon products, Exxon Isopars ® E, G, H, K, L, and M

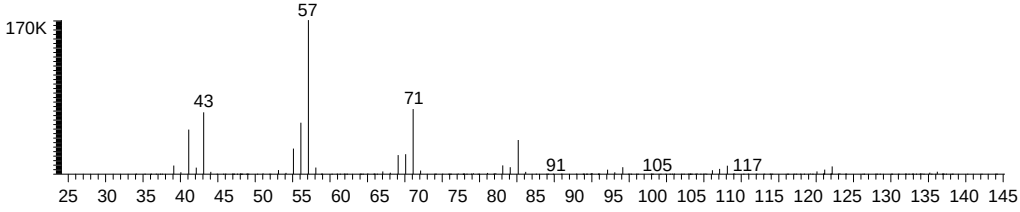


Figure 29 Average mass spectrum taken from Isopar L, over the range of 9–12 min. This spectrum has a very similar appearance to that of a branched alkane

Following up on this work, Wintz and Rankin applied principal components analysis (PCA), only to learn that while the ratios within the pairs show little, if any, change between unevaporated and 50%-evaporated gasoline, only a few of those pairs were actually useful in distinguishing between one gasoline source and another, and some of those pairs occurred in the LPD range, so they are unlikely to be found in gasolines evaporated to more than 50% [26].

Further work by Barnes and Dolan identified six ratios of sequentially eluting minor components in 50%-evaporated gasoline, and four ratios in 75%-evaporated gasoline. Using these ratios, they were able to successfully discriminate among 16 gasoline

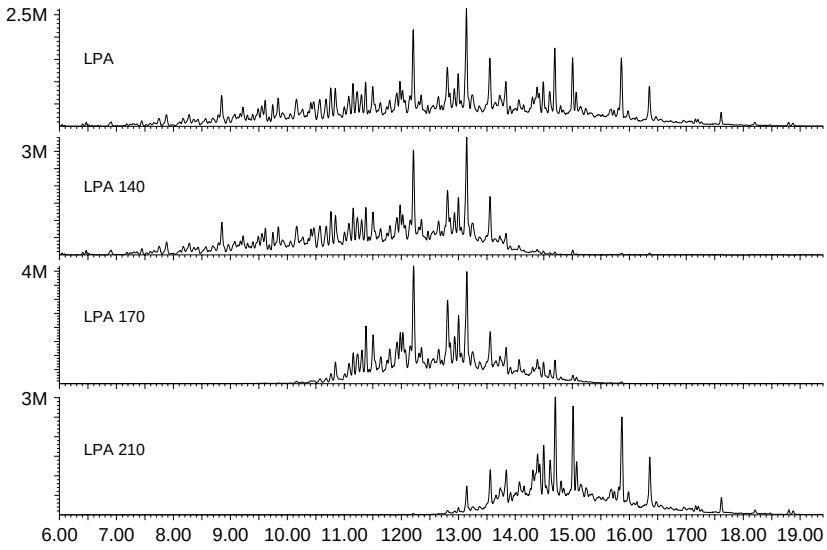


Figure 30 Total ion chromatograms of 4 naphthenic–paraffinic products, Vista LPA, LPA 140, LPA 170, and LPA 210. The numbers approximately correspond to flash points

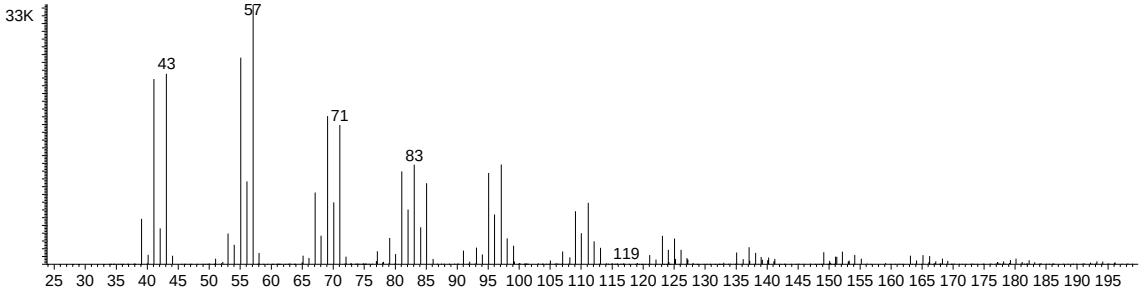


Figure 31 Average mass spectrum of Vista LPA®, taken from 8 to 17 min

samples evaporated 50%, and 10 gasoline samples evaporated 75%. They used a blind study to confirm the ability of the technique to correctly identify the source of three gasolines, matched against a library of 10 candidate sources. The techniques uses a 60-m column, and is automated using a target compound program that takes the abundance of the base peak from the mass spectrum of each compound of interest, and divides the base peak abundance of the later eluting compound by the base peak abundance of the earlier eluting compound. The authors state that this technique not only appears to work but also is most useful for eliminating a source, and caution that the analyst's statement of conclusions reflect that a common origin is indicated, but is not a certainty [27]. Sandercock and DuPasquier have reported being able to distinguish 32 out of 35 gasolines by collecting selected ion monitoring (SIM) data of the polynuclear aromatic hydrocarbons from a 200 μl sample, and then applying PCA to the data [28].

While doing comparisons of gasolines, or any ignitable liquids for that matter, the best approach would seem to be to first take the proposed source liquid, and evaporate some of it so that it matches the residue in terms of carbon number range. Once this has been accomplished, a detailed examination of the finer points in the chromatogram should be carried out.

A really "good match" between chromatograms not only matches the peaks but also matches the valleys. When one is trying to determine if an ILR could have come from a particular source, the valleys should match. Examining multiple EICs for both quality and quantity would also seem to be a necessary step when doing comparisons. This practice has been codified with respect to the analysis of distillates heavier than kerosene in ASTM D 5379-00 *Standard Practice for Oil Spill Source Identification by Gas Chromatography and Positive Ion Electron Impact Low Resolution Mass Spectrometry* [29]. The approach in this method is to compare 24 EIC of polynuclear aromatics and other biomarkers. This is similar to Sandercock and DuPasquier's work.

Pepler has reported making brand identifications of gasolines by using high-resolution mass spectrometry to detect high molecular weight additive packages in gasolines that have lost all of their volatile materials [30].

Given the state of the art, it is probably a good thing that comparisons are not frequently requested.

Other Fire Debris Analyses

Most public crime laboratories in the United States limit their practice of fire debris analysis to the detection of ignitable liquids. There are, however, other chemical analyses that are occasionally requested in the context of a fire investigation. If chemical incendiary materials are suspected of having been used, the laboratory may be asked to check for such substances as glycols, brake fluid, chlorine residues, and residues of road flares. Glycols and brake fluid are detectable using gas chromatography-mass spectrometry. Nonhydrocarbon chemical incendiary substances may be detected by a combination of elemental analysis and microscopic examination.

Laboratories are also sometimes asked to check for the presence of animal or vegetable oils in the samples of debris suspected of having undergone spontaneous heating. Such oils are not amenable to the direct analysis by GC-MS because of their high molecular weight, but it is possible to derivatize vegetable oils that produced fatty acid methyl esters (FAME), which can be analyzed by GC-MS. In a typical FAME analysis, the debris samples are extracted with *n*-heptane. The filtered extract is then derivatized using 2 N potassium hydroxide in methanol. After agitation and centrifugation, the heptane solution is analyzed directly. Depending on the extent of the fire damage, it may be possible to characterize the original oil, but much caution is required, as the fatty acids that are most susceptible to spontaneous heating are the ones likely to be diminished in concentration. Generally, samples are simply reported as containing vegetable oils or not. The technique described above is capable of detecting quantities of vegetable oil as small as 10 mg.

Other laboratory analyses frequently requested include the nonchemical variety, such as the examination of appliances or devices, which are believed to have been instrumental in the ignition of the fire. Such analyses are generally not conducted in crime laboratories, but are instead conducted in private facilities by electrical engineers or other experts [31].

Reporting Procedures

The analyst's report can be one of the most important documents generated during the investigation of a fire. The report, therefore, should be written carefully enough so that readers are not misled. The objective of fire debris analysis is to determine whether there is any *foreign* ILR present in a sample.

The report from a forensic science laboratory is supposed to be a *scientific* report. As such, it should include an introduction, a section detailing the test methods and results, and a discussion and conclusion section, if necessary. The laboratory report should also state what was done with the evidence. There is an unfortunate tendency in some agencies to provide checklists and merely state "positive for gasoline," or "negative for accelerants," rather than preparing a real laboratory report. Reading such a report, a reviewer cannot tell whether the analyst used a GC-MS or a Ouija board.

Given that it is now possible to store templates with all possible results in them, the excuse that it is "too time-consuming" to prepare an understandable narrative report should be regarded as unacceptable. Although laboratory reports are intended to assist investigators, they should also be understandable to individuals requested to review that work. As a minimum, a report should include the following:

- An identification of the fire in question.
- A description of how the sample was delivered to the laboratory, when, and by whom.

- A description of the samples, including container size, substrate material, and a reported location from where the sample was collected.
- A description of the isolation procedure used to separate the ILR from the sample substrate.
- A description of the analytical technique applied to the sample extract.
- The results of the analysis of the data.
- A discussion of the meaning of the results, if there is any chance of misinterpretation or misuse.

In the discussion section, the analyst can provide examples of potential sources for whatever ILR may have been identified. This is also an appropriate place to put in a disclaimer about the possibility that the ILR may not be foreign to the background.

- A conclusion or bottom line, understandable to even an attorney is a helpful thing.
- A sentence stating what has happened to the sample.

If a substance that is natural or incidental to the background is found, it is the analyst's job to say so. ASTM E 1387 and E 1618 both allow for the inclusion of disclaimers, on both positive and negative reports. With respect to negative reports, a disclaimer to the effect that negative results do not preclude the possibility that ignitable liquids were present at the fire scene, but may not have been detected for a variety of reasons including detection thresholds, is suggested as an aid to help readers avoid misunderstanding the report. Likewise, in the case of a positive report, the standard states, "It may be appropriate to add a disclaimer to the effect that the identification of an ILR in a fire scene does not necessarily lead to the conclusion that a fire was incendiary in nature. Further investigation may reveal a legitimate reason for the presence of ILR." Certainly, a finding of gasoline in the living room is noteworthy. It is not so noteworthy to find gasoline in the basement near the chain saw.

On occasion, it seems that the fire debris analyst is too eager to "help" with an arson investigation. Reporting a sample as being "positive" when all the analyst has identified is background compounds is neither helpful nor morally defensible. Karen and Paul Stanley of Akron, Ohio, were falsely accused of setting the fire that killed their infant son. The fire debris samples were submitted to a laboratory that reported finding "turpentine" in a sample consisting of charred Douglas fir. The laboratory analyst at the very least should have reported the finding of turpentine with a caution stating that the turpentine was indistinguishable from naturally occurring turpentine found in coniferous woods. The case was dismissed after the prosecutor saw the light [32].

A similar case of an analyst being too "helpful" occurred in a homicide case in Georgia, where the laboratory analyst reported finding "toluene, a flammable liquid" on a suspect's "clothing." Actually, the "clothing" included a pair of tennis shoes, and the examination of a new pair of tennis shoes directly from the shoe store revealed the presence of a high concentration of toluene. A closer examination of the first analysis revealed that in addition to toluene, diethylene glycol and butylated hydroxytoluene (BHT) were also present in the suspect's shoes and the exemplar shoes in identical relative proportions. Less than a year earlier, there had been a presentation on the analysis of suspects' shoes in arson cases at an American Academy of Forensic Sciences (AAFS) seminar [33]. Perhaps, if the analyst had attended that presentation (or read the proceedings), the toluene would have been recognized for what it was.

Record Keeping

Each case file that includes a positive identification of an ILR should contain not only the sample charts but also charts of a standard to which that sample can be compared. This means that both the sample and standard should be printed with the same axes scales, so that the data can be easily reviewed. Even if there is no criminal prosecution, the analyst should be aware that many fire cases involve civil litigation, so the case file should be kept for a reasonable period of time.

Electronic data files should also be protected and stored. There are only a few versions of GC-MS data analysis software in widespread use, so it is possible for one analyst to review another's raw data. Such reviews should be facilitated by the preservation of the data. Instrumental data such as tune reports or spectrum scans should be kept, if only to keep track of the performance of the instrument. Certainly, if a case file has matching standards and samples, there is no need to review the tune report for that particular week. No matter what happened, the instrument tune

parameters would not cause a false positive. Blanks, on the other hand, should be stored somewhere, particularly for those cases where the detection limit of the procedure is being pushed.

Conclusion

The isolation and identification of ILR from fire debris samples is an important part of the fire investigation process. The fire debris analyst should be familiar with the common techniques of fire investigation and understand the language used by investigators. If an ILR is identified, a hypothesis that a fire was incendiary may be supported, and a long litigation process may ensue, either in criminal or civil court. The laboratory results are often the deciding factor in a prosecutor's decision to indict, or an insurance company's decision to resist a claim.

The technology for fire debris analysis has improved dramatically over the last decades, to the point where our methods are now as sensitive as they need to be, and possibly more sensitive. The analytical procedure requires focus and creativity, but also adherence to the generally accepted criteria for making an identification. The analysis should be conducted in such a way that it is capable of being reviewed by another analyst, and such reviews should be not only expected but also welcomed. Science is based on "multiple witnessing." Fire debris analysts need to be acutely aware of the stakes involved in what they are doing and of the need to communicate effectively about the meaning of their findings.

References

[1] Lentini, J. (2006). *Scientific Protocols for Fire Investigation*, CRC Press, Boca Raton, p. 137.

[2] AANotes (1982). *Arson Analysis Newsletter*, SEA, Columbus, Vol. **6**, Chapter 3, p. 57.

[3] ASTM (2005). *Standard Practice for Separation and Concentration of Ignitable Liquid Residues from Fire Debris Samples by Solvent Extraction*, E1386-00 (2005), ASTM International, West Conshohocken.

[4] ASTM (2007). *Standard Practice for Separation of Ignitable Liquid Residues from Fire Debris Samples by Passive Headspace Concentration With Activated Charcoal*, E1412-07, ASTM International, West Conshohocken.

[5] ASTM (2005). *Standard Practice for Sampling of Headspace Vapors from Fire Debris Samples*, E1388-05, ASTM International, West Conshohocken.

[6] ASTM (2007). *Standard Practice for Separation of Ignitable Liquid Residues from Fire Debris Samples by Dynamic Headspace Concentration*, E1413-07, ASTM International, West Conshohocken.

[7] ASTM (2001). *Standard Practice for Separation and Concentration of Ignitable Liquid Residues from Fire Debris Samples by Passive Headspace Concentration with Solid Phase Microextraction (SPME)*, E2154-01, ASTM International, West Conshohocken.

[8] Juhala, J.A. (1982). A method for adsorption of flammable vapors by direct insertion of activated charcoal into the debris samples, *Arson Analysis Newsletter* **6**(2), 32.

[9] Newman, R. & Dolan, J. (2001). Solvent options for the desorption of activated charcoal in fire debris analysis Proceedings of the American Academy of Forensic Sciences Annual Meeting. Seattle.

[10] Armstrong, A. & Lentini, J. (1997). Comparison of the eluting efficiency of carbon disulfide with diethyl ether: the case for laboratory safety, *Journal of Forensic Sciences* **42**(2), 307.

[11] NFPA (1994). *Guide to Fire Hazard Properties of Flammable Liquids, Gases and Volatile Solids*, NFPA 325, NFPA.

[12] Newman, R., Gilbert, M. & Lothridge, K. (1998). *GC-MS Guide to Ignitable Liquids*, CRC Press.

[13] Smith, R.M. (1982). Arson analysis by mass chromatography, *Analytical Chemistry* **54**(13), 1399.

[14] Stauffer, E., Dolan, J. & Newman, R. (2008). *Fire Debris Analysis*, Academic Press, Burlington, p. 290.

[15] ASTM (2006). *Standard Test Method for Ignitable Liquid Residues in Extracts from Fire Debris Samples by Gas Chromatography-Mass Spectrometry*, E 1618-06, Annual Book of Standards, Volume 14.02, ASTM, W. Conshohocken.

[16] Lentini, J., Dolan, J. & Cherry, C. (2000). The petroleum-laced background, *Journal of Forensic Sciences* **45**(5), 968.

[17] Lentini, J. (2001). Persistence of floor coating solvents, *Journal of Forensic Sciences* **46**(6), 1470.

[18] Lentini, J. (1998). Differentiation of asphalt and smoke condensates from liquid petroleum distillates using GC-MS, *Journal of Forensic Sciences* **43**(1), 97.

[19] Lentini, J. & Waters, L. (1982). Isolation of accelerant-like residues from roof shingles using headspace concentration, *Arson Analysis Newsletter* **6**(3), 48.

[20] ASTM (2001). *Standard Test Methods for Aromatics in Mineral Spirits by Gas Chromatography*, D 3257-01, Annual Book of Standards, Volume 6.03, ASTM, W. Conshohocken.

[21] Stauffer, E., Dolan, J. & Newman, R. (2008). *Fire Debris Analysis*, Academic Press, Burlington, p. 464.

[22] Murphy, B.L. & Morrison, R.D. (eds) (2002). *Introduction to Environmental Forensics*, Academic Press.

[23] Mann, D.C. (1987). Comparison of automotive gasolines using capillary gas chromatography I: comparison methodology, *Journal of Forensic Sciences* **32**(3), 606.

[24] Mann, D.C. (1987). Comparison of automotive gasolines using capillary gas chromatography II: limitations of

automotive gasoline comparisons in casework, *Journal of Forensic Sciences* **32**(3), 616.
[25] Dolan, J. & Ritacco, C. (2002). Gasoline comparisons by gas chromatography-mass spectrometry utilizing an automated approach to data analysis, *Proceedings of the American Academy of Forensic Sciences Annual Meeting*, Atlanta, February 16, p. 62.
[26] Wintz, J. & Rankin, J. (2004). Application of principal components analysis in the individualization of gasolines by GC-MS, *Proceedings of the American Academy of Forensic Sciences Annual Meeting*, Dallas, February, p. 48.
[27] Barnes, A.T., Dolan, J.A., Kuk, R.J. & Siegel, J.A. (2004). Comparison of gasolines using gas chromatography-mass spectrometry and target ion response, *Journal of Forensic Sciences* **49**(5), 1018.
[28] Sandercock, P.M. & Du Pasquier, E. (2004). Chemical fingerprinting of gasoline 2. Comparison of unevaporated and evaporated automotive gasoline samples, *Forensic Science International* **140**, 43.
[29] ASTM (2001). *Standard Practice for Oil Spill Source Identification by Gas Chromatography and Positive Ion Electron Impact Low Resolution Mass Spectrometry*, D 5379-00, Annual Book of Standards, Volume 11.02, ASTM, W. Conshohocken.
[30] Pepler, R. (2005). Petrol Branding, *Gardiner Associates International Fire & Arson Investigation Conference*, Brunel University, Uxbridge Campus.
[31] Lentini, J. (2006). *Scientific Protocols for Fire Investigation*, CRC Press, Boca Raton, p. 203.
[32] (2002). Prosecution expert rejects short as cause, *Akron Beacon Journal*, Akron Ohio.
[33] Cherry, C. (1996). Arsonist's shoes: clue or confusion, *Proceedings of the American Academy of Forensic Sciences*, Nashville, February 1996. AAFS.

JOHN J. LENTINI

Fire Investigator: Standardization, Accreditation, and Certification

Introduction

The adoption by the fire investigation community and the larger forensic science community of standards began in the mid-1970s in the United States, but did not reach fruition for 10 years in the case of fire investigators, and 15 years in the case of fire debris analysts. Adoption of standard methodologies began in the early 1990s, although some primitive "standards" were available for laboratory analysts as early as 1982.

As with the introduction of any new standard, be it for individuals, institutions, or methodologies, there was a substantial rearguard action that needed to be overcome before the standards could be characterized as "generally accepted". All of the forensic disciplines have struggled with these concepts as they tried to define what was meant by the various standardization terms, and how standardization programs could be implemented.

The Quality Triangle in Forensic Sciences

The concept of the quality triangle in the forensic sciences was first presented in the early 1990s. Lawrence Pressley, a former Quality Assurance Director at the FBI Laboratory, is generally credited with the image shown in Figure 1, which ties together accreditation, certification, standardization, and proficiency testing. As interdependent and seamless as these concepts may now seem, they have not come together without some resistance.

Fire investigators embraced certification in the late 1980s, but many fiercely resisted the idea of standardization (some still do). Crime laboratory directors embraced the idea of accreditation in the early to

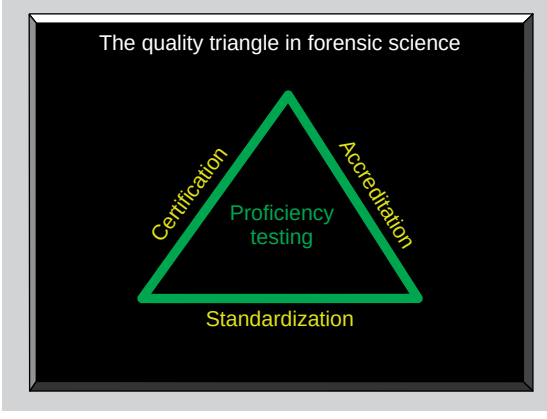


Figure 1 The Quality Triangle in Forensic Science represents the current paradigm for quality assurance in forensic science laboratories in the US

mid-1980s, but resisted the idea of individual certification, as did some individual practitioners (some still do). The Criminalistics Section of the American Academy of Forensic Sciences (AAFS) voted in the mid-1990s to join with the American Board of Criminalistics (ABC) in supporting certification of individuals, and the American Society of Crime Laboratory Directors (ASCLD) finally decided to support the ABC program in the late 1990s. Some conflicts still exist, such as the debate over who should pay for proficiency testing – the individual who needs it to maintain his or her certification or the laboratory, which needs proficiency testing to maintain its accreditation. Despite these peripheral disagreements, the forensic quality triangle is the paradigm of the current era.

On August 9 and 10, 2004, the American Bar Association House of Delegates, representing the ultimate consumers of the work product of all forensic scientists (including fire investigators), passed the following resolution:

> Resolved, That the American Bar Association urges federal, state, local and territorial governments to reduce the risk of convicting the innocent, while increasing the likelihood of convicting the guilty, by adopting the following principles:

1. Crime laboratories and medical examiner offices should be accredited, examiners should be certified, and procedures should be standardized and published to ensure the validity, reliability, and timely analysis of forensic evidence.
2. Crime laboratories and medical examiner offices should be adequately funded.
3. The appointment of defense experts for indigent defendants should be required whenever reasonably necessary to the defense.
4. Training in forensic science for attorneys should be made available at minimal cost to ensure adequate representation for both the public and defendants.
5. Counsel should have competence in the relevant area or consult with those who do where forensic evidence is essential in a case.

The consumers are now demanding what the leaders of the forensic science and fire investigation professions have been advocating for more than a decade. Not all of the parts in the quality triangle can be applied to fire investigation, due to the unavailability of accreditation for field investigators and the lack of proficiency testing. The entire triangle can, however, be applied to the laboratory analysis of fire debris. The individual parts of the forensic quality triangle as it applies to fire and explosion investigations are addressed.

Method Standardization

For over a century, the American Society for Testing and Materials (ASTM) (now known as *ASTM International*) served as a standards development organization and followed rules set forward by oversight bodies such as the American National Standards Institute (ANSI) and the International Organization for Standardization (ISO). The National Fire Protection Association, also following ANSI rules, promulgated standards for fire equipment and fire departments. These standards development organizations have always responded to a perceived need for standards coming from within a particular community.

ASTM Committee E30 on Forensic Sciences got its start in 1974, largely based on the perception of the forensic engineering community that standards were required in order to conduct valid investigations of incidents resulting in product liability litigation. The standards promulgated were general in nature, and by the early 1980s, the engineers had written all of the standards that they thought were necessary. The engineering standards, dealing with the collection of information [1], reporting of opinions [2], evaluation of technical data [3], and the prevention of spoliation [4] have since been generalized to apply to all forensic investigations, both civil and criminal. These engineering standards are frequently cited in fire investigations.

Committee E30 became dormant throughout most of the 1980s, but in 1989, ASTM approached the American Academy of Forensic Sciences and E30 was revitalized. The first new standards adopted by E30 were fire debris analysis standards. These were adapted from a document promulgated by the Forensic Science Committee of the International Association of Arson Investigators (IAAI) entitled "Guidelines for Laboratories Performing Chemical and Instrumental Analysis of Fire Debris Samples" [5]. The first edition of ASTM E 1387, Standard Test Method For Flammable or Combustible Liquid Residues In Fire Debris Samples by Gas Chromatography, was adopted in 1990 [6]. This standard test

method, as well as other fire debris analysis standard practices, has been well accepted in the forensic science community (*see* **Fire Debris: Laboratory Analysis of**).

Beginning in 1985, the National Fire Protection Association (NFPA) Technical Committee on Fire Investigations worked on developing a guide for fire and explosion investigations. The resulting document, known as *NFPA 921*, was first published in 1992 [7]. The initial reaction to NFPA 921 was mixed. Partly because the document described numerous "misconceptions" (*see* **Arson Investigation: Misconceptions and Mythology**) held by many fire investigators, the holders of those misconceptions fiercely resisted acknowledging NFPA 921 as a standard of care. By 2000, however, most responsible fire investigators accepted the necessity of a standard to bring some measure of quality to the profession. The IAAI president publicly endorsed the adoption of NFPA 921, and the US Department of Justice issued a research report entitled "Fire and Arson Scene Evidence: A Guide For Public Safety Personnel" [8]. In the last decade, numerous courts, ruling on reliability challenges have accepted NFPA 921 as the *de facto* standard of care for fire investigation. Investigators who follow NFPA 921 generally have their opinions admitted, and those who do not generally have their opinions excluded, at least with respect to civil cases [9]. (US courts seem reluctant to exclude the opinion testimony of law enforcement officers, even if their methodology is well outside generally accepted norms.)

Because of the relative rarity of explosions and the near monopoly of government investigative agencies in explosion investigations, standards development organizations have not yet prepared any standard documents. The US Justice Department did publish *Guide for Explosion and Bombing Scene Investigation* in 2000 [10].

Accreditation

As used in this context, the term *accreditation* is intended to apply to agencies and institutions, while the term *certification* applies to individuals. Currently, accreditation is available only to forensic science laboratories conducting analysis of fire debris and explosive residues. Accreditation is not available to agencies that conduct fire or explosion investigations in the field.

Accreditation is an increasingly important way by which agencies and companies demonstrate that their work meets certain minimum standards. Both public and private forensic science laboratories may seek accreditation from the American Society of Crime Laboratory Directors/Laboratory Accreditation Board (ASCLD/LAB) or from other "registered" companies (registrars) that audit the accredited organization to a particular standard. Testing and calibration laboratories are generally accredited to ISO 17025, *General Requirements for the Competence of Testing and Calibration Laboratories* [11] A modified version of ISO 17025 is used by ASCLD/LAB in its accreditations [12]. Accreditation is becoming more common for police and fire departments, as well. Whichever registrar and agency or company chooses, the agency is required to have in place a written quality assurance program, and be able to demonstrate to auditors sent by the registrar that the quality assurance program actually assures quality, and that individuals in the organization actually follow the program.

Certification

Certification is a voluntary process of peer review by which a practitioner is recognized as having attained the professional qualifications necessary to practice in one or more disciplines [13]. Some employers "certify" their employees, but real certification requires that the certificate be provided by someone who does not necessarily have an interest in the applicant's ability to be recognized as an expert. This article focuses only on certification offered by third parties.

In a properly designed certification program, the initial certification is only a gateway to continued professional development. Getting through the application and test process certainly requires the acquisition and demonstration of a set of knowledge, skills, and abilities that members of the profession believe are necessary, but it is the maintenance of certification that makes a program valuable. Valid certification programs have four necessary attributes:

1. a credentials check;
2. a validated process, usually a written examination, wherein the certification body tests the knowledge, skills, and abilities of the applicant;

3. a means of revoking the certification of individuals who violate a code of professional practice; and
4. a recertification program, which requires the certified individual to document continuing professional development.

Some private organizations offered certification programs, but they were really nothing more than diploma mills, which took the applicant's money in exchange for a certificate, but such certification had no real meaning. As a result, various boards have been set up by professional organizations to accredit certification programs. The Forensic Specialties Accreditation Board (FSAB) is a corporation created by the AAFS to accredit certification bodies. FSAB generally requires that certification bodies applying for FSAB accreditation demonstrate that their program embodies the four requirements listed above.

The ABC offers certification for fire debris analysts. In order to become certified, applicants must demonstrate appropriate education, training, and experience prior to being allowed to challenge an examination, which contains a large component of general forensic science knowledge in addition to knowledge of fire debris analysis. Documentation of annual participation in a proficiency testing program is required to maintain certification. Every five years, the certified individual is required to document a satisfactory level of continuing professional development. Recertification points can be obtained by attending meetings, serving on technical committees or as an officer of a forensic science organization, publishing articles or books, and giving presentations. Only a very small percentage of practicing fire debris analysts holds ABC certification. As is the case with many certification programs, the first individuals to become certified are the leaders of the discipline.

Fire scene investigators can obtain certification from the IAAI or from the National Association of Fire Investigators (NAFI). Both of these programs require an application process, passing an examination, and documenting continuing education. Because of the nature of the work, proficiency testing is not available for fire scene investigators.

As with fire debris analysts, a significant percentage of the practicing fire investigators in the United States have certification from neither the IAAI nor the NAFI.

References

[1] ASTM E1188-05 (2005). *Standard Practice for Collection and Preservation of Information and Physical Items by a Technical Investigator*, ASTM International, West Conshohocken.

[2] ASTM E 620-04 (2004). *Standard Practice for Reporting Opinions of Technical Experts*, ASTM International, West Conshohocken.

[3] ASTM E 678-07 (2007). *Standard Practice For Evaluation of Scientific or Technical Data*, ASTM International, West Conshohocken.

[4] ASTME 860, 07 (2007). *Standard Practice for Examining and Preparing Items That are or May Become Involved in Criminal or Civil Litigation*, ASTM International, West Conshohocken.

[5] IAAI Forensic Science Committee (1988). Guidelines for laboratories performing chemical and instrumental analyses of fire debris samples, *Fire and Arson Investigator*, **38**(4), 45.

[6] ASTM E 1387-90 (1990). *Standard Test Method for Flammable or Combustible Liquid Residues from Fire Debris Samples by Gas Chromatography*, ASTM International, West Conshohocken.

[7] National Fire Protection Association (2004). *NFPA 921: Guide for Fire and Explosion Investigation Investigations*, National Fire Protection Association, Massachusetts.

[8] National Institute of Justice (2000). *Fire and Arson Scene Evidence: A Guide for Public Safety Personnel*, U.S. Department of Justice, National Institute of Justice, Washington, DC.

[9] Lentini, J.J. (2007). The standard of care in fire investigation, *Canadian Association of Fire Investigators Journal*.

[10] National Institute of Justice (2000). *A Guide for Explosion and Bombing Scene Investigation*, U.S. Department of Justice, National Institute of Justice, Washington, DC.

[11] International Organization for Standardization (1999). ISO 17025, *General Requirements for the Competence of Testing and Calibration Laboratories*, International Organization for Standardization, Geneva.

[12] American Society of Crime Laboratory Directors, Laboratory Accreditation Board (2001). *Program Manual*, ASCLD/LAB, Garner.

[13] American Board of Criminalistics (2007). http://www.criminalistics.com viewed on 31 December 2007.

JOHN J. LENTINI

Fire Modeling and Its Application in Fire Investigation

Introduction

Why do we investigate fire and explosion incidents?

The answer is that fire and explosion incidents must be investigated for the purpose of understanding failure. A failure can be defined as an "... unacceptable difference between expected and observed performance" [1]. Individuals involved in the forensic investigation and analysis of fire and explosion incidents are usually tasked with rendering opinions as to the origin, cause, responsibility, and/or prevention of such incidents. From a technical perspective, the investigation will provide a physical description of fire events, identify failure modes and causative factors, provide feedback on design, identify noncompliance with appropriate standards and practices, or evaluate the validity of analytical tools. From a forensic perspective, investigators identify "more probable than not" scenarios to a "reasonable degree of scientific certainty" that help the client, as well as the court, to understand the technical issues and perhaps to determine fault or assign responsibility.

The investigation process contributes to a compilation of fire statistics for use in the development of educational programs, fire prevention codes, standards, and training. Furthermore, such investigations produce a compilation of requisite understanding of specific fire scenarios by applying fundamental knowledge to the incident under study. Thus, it is essential that valid and accurate determinations of the origin and cause of a fire be obtained so that the aggregate product of the investigation process provides essential information for protection of lives and property. NFPA 921 *Guide for Fire and Explosion Investigations* was developed to assist in improving the fire and explosion investigation process and the quality of information resulting from the investigation process based on reliable scientific principles [2].

The Essence of the Investigation Process

The essence of the investigation of fire and explosion incidents is embodied in the "... use of a systematic approach with attention to all relevant data". NFPA 921 recommends the use of the scientific method, which has traditionally been used in the physical sciences and forms the basis for scientific and engineering investigations, including fire and explosion incidents [2].

The scientific method as applied to fire and explosion incidents involves the following steps: recognize the need, define the problem, collect data, analyze the data, develop a hypothesis, and test the hypothesis. As discussed in NFPA 921, a typical investigation may include the documentation of an incident scene during a scene inspection or a review of any previous scene documentation; the collection of evidence; interviews with eyewitnesses; review and analysis of previous investigations; and the collection of data from other appropriate and relevant sources. This data is then analyzed to produce a hypothesis or group of hypotheses to explain the origin and cause of the fire or explosion event [2].

The most important step of the scientific method as it applies to fire investigation is the testing of hypotheses using "... the principle of deductive reasoning where the investigator compares his or her hypothesis to all of the known facts as well as the body of scientific knowledge associated with the phenomena relevant to the specific incident". The use of the scientific method will frequently uncover new data for analysis, which may result in the need for previous determinations to be reanalyzed. This iterative process is highly desirable and quite necessary for a reliable investigation [2]. Valid implementation of the method in the investigation requires the use of fundamental knowledge associated with the combustion and fire sciences.

Fire Is a Complex Phenomenon

Above all, "combustion" is a chemical process that involves the reactive combination of a fuel and an oxidizer. Generally, combustion phenomena can be categorized into two groups: "desired combustion" (e.g., a building furnace or a campfire) and "undesired combustion" (e.g., a house fire). Although sometimes treated by different professional communities as distinct processes, undesired combustion involves the

same fundamentals as desired combustion. The study of desired combustion (i.e., the combustion sciences) has been somewhat limited to gaining fundamental scientific knowledge. On the other hand, undesired combustion (i.e., the fire sciences) has included a wider diversity of the unwanted fire problem and has tended to be more empirical in its study and practical application. The body of knowledge with respect to the combustion and fire sciences includes many textbooks, most notably Glassman [3], Cox [4], Friedman [5], Quintiere [6], Karlsson and Quintiere [7], Drysdale [8], and the (Society of Fire Protection Engineers) *SFPE Handbook* [9]. In addition, there is a relatively large body of peer-reviewed research in both the combustion and fire sciences literature.

The study of unwanted fires forces together many otherwise discrete specialties [4] such as thermodynamics, fluid transport, heat transfer, and combustion chemistry. The reason for this integration of fields is the highly coupled nature of the phenomena occurring simultaneously. Even with the wealth of information and knowledge in the combustion and fire sciences, because of the complexity, the engineer has been forced to exploit solutions that are empirically based. However, the advancement and evolution of the combustion and fire sciences have helped to build a more secure technical foundation for the discipline of Fire Protection Engineering [4]. The challenge for the fire investigator is the practical application of fundamental knowledge to the investigation of fires that have occurred in the field. Such fundamental and empirical knowledge can be exploited through the development of models to simplify the analysis of the combustion phenomena that limit practical application.

The Use of Applied Knowledge in the Investigation of Fires

In order to render accurate and reliable determinations associated with particular fire incidents, the investigator must "understand the fire." This understanding of specific fire events can be derived both qualitatively and quantitatively from more discrete aspects of basic knowledge, from empirical data or formulations, and/or from a compilation of quantitative tools that embody a collective form of basic knowledge. A significant increase in the wealth of fundamental knowledge and the exponential advancements in computer technology have allowed the development of computer fire models. Such tools can be used as part of a fire dynamics analysis.

Fire dynamics analysis consists of the use of mathematical equations or models derived from fundamental principles or empirical data. These range from simple algebraic equations to computer algorithms incorporating many individual equations. Fire dynamics analysis can be used to predict fire phenomena such as time to flashover, upper layer gas temperatures and concentrations, and smoke concentrations and flow rates. Characteristics of the environment, such as unburned fuel, surface temperatures, and actuation time of smoke detectors, heat detectors, and sprinklers are also amenable to analysis. In addition, fire dynamics analysis can be used to predict the effects of changes to the environment, such as the opening or closing of doors, window breakage, or other physical events [2].

The authority of fire dynamics analysis relies on its use of a robust scientific approach to evaluate hypotheses regarding fire origin, cause, and the growth or decay of the resulting fire. These types of analyses generally use building data and fire dynamic principles to predict the environment produced by a fire under a proposed hypothesis. An analysis of the spread of the fire and resulting damage, as well as the time line of the fire, are potential products of such an analysis. The results can be compared to physical and eyewitness evidence to support or refute the hypothesis [2]. Thus, mathematical modeling provides the investigator with tools for the testing of hypotheses associated with the origin and cause of a particular fire. Even when the origin and cause of the fire are not an issue, these modeling tools can also be used to evaluate the factors associated with producing the resulting damage to property or injury to people (i.e., "The origin and cause of the loss").

Brief History of the Use of Fire Models

While the most identifiable fire models were developed after the introduction of the personal computer in the early 1980s, the first book devoted to the use of modeling in fire research was published in 1959 [10]. Philip Thomas and Peter Hinkley [11] developed the first zone fire model in response to two major fires. The first occurred in 1953 at the General Motors hydraulic transmission plant at Livonia, Michigan. The second fire occurred a few years later at the Jaguar Automobile Production Plant in the United Kingdom. Both fires grew rapidly, due to the production of flammable vapors from the heating of

the built-up steel deck roof, and exceeded the ability of fire suppression resources. One result of these fires was an interest in determining if roof venting could have made a difference and could have provided protection in a manner that would eliminate the need to erect fire walls in these spaces [12].

Thomas and Hinkley [11] conducted extensive experiments and formulated a means of determining roof vent requirements based on equilibrium flow rates between the mass flow into the upper layer and the mass flow out of the roof vents. Since this work was done prior to the common availability of personal computers to practicing engineers, Thomas and Hinkley incorporated their results in the form of nomographs that provided engineers with a reasonable approximation of the venting requirements under a range of building dimensions and fire scenarios.

The initiation of preflashover zone fire modeling can be traced to the mid-1970s with the publication of a description of the fundamental equations by Quintiere [13]. On the basis of these equations, the first generation of zone fire models were published by Pape and Waterman [14] and by Emmons and Mitler [15]. Following the introduction of these two models, a number of zone fire models for mainframe computers were subsequently developed. In 1985, the first zone model, ASET-B, was written specifically for the newly available personal computers [16].

Since the 1980s and the introduction of the first generation of compartment fire models, many models of similar construct have been developed around the world. Many of these models provided a user-friendly Graphical User Interface (GUI) that helped to initiate a large expansion in their use and application. As early as 1985, it had become broadly accepted that future advancements in Fire Protection Engineering would be through the use of computer modeling [17].

Friedman [18], and more recently, Olenick and Carpenter [19], have compiled surveys of available fire models, which include some 170 associated models (http://www.firemodelsurvey.com). The most identifiable fire models in use today include FPEtool [20], Consolidated model of Fire Growth and Smoke Transport (CFAST) [21], and Fire Dynamics Simulator (FDS) [22] models that are all products of the Building Fire and Research Laboratory (BFRL) at the National Institute of Standards and Technology (NIST). Today, fire models are being used by engineers and architects, building officials, the fire service, investigators, building and fire code developers, materials and systems manufacturers, fire researchers, and educators.

Overview of Fire Models

What Is a Model?

Generically speaking, a model can be thought of as a tool for simulating and predicting the behavior of a complex entity or system. All models require input, the processing of the input, and the output of the predicted results. Usually, due to the complexity of the natural world as viewed as a system, a practical model requires that simplifying assumptions be made, but such simplifications tend to limit the range of applicability of the model. Models can evolve from empirical data or as a result of existing knowledge from Physics or Chemistry, or a combination of both.

Although a somewhat common perception, a model does not need to be mathematical in nature. There are two broad classes of models, namely, physical models and mathematical models. Physical models endeavor to reproduce phenomena in a simplified physical situation. Reduced-scale physical models are a common form of modeling because full-scale experiments can be prohibitively expensive, intractable, or altogether infeasible. Reduced-scale modeling does not simply mean conducting experiments at a reduced physical scale. In addition to reducing the linear dimensions of a physical system, it is also necessary to maintain mechanical, thermal, and chemical similarity in the reduced-scale model through dimensional analysis or from the fundamental equations describing the phenomena [10].

Mathematical models are sets of equations that describe the behavior of a physical system. Mathematical models can be further divided into two classes: probabilistic models and deterministic models. Probabilistic models attempt to simulate the range of possible outcomes of the random behavior of a system, while deterministic models assume that, given a well-defined physical system, the system's behavior can be entirely and accurately predicted. These classes have some crossover, in that probabilistic models include within their construct some deterministic modeling algorithms. In general, the aim in the use of physical models is to discover fundamental laws governing the behavior of systems. A mathematical model can be constructed on the basis of derived laws that can then be used to predict the behavior

of novel systems. Thus, physical and mathematical models can be interrelated and complementary [10]. In the context of this discussion, the focus is on deterministic fire models.

What is a Deterministic Fire Model?

A deterministic fire model is a mathematical representation of the processes encountered in "fire" by interrelated expressions based on known physics and chemistry. Although the implication from the term *fire model* is that the combustion process is modeled directly, due to the complexity of the phenomenon, "fire models" provide engineering approximations of the time-varying characteristics and impact of a user-defined mathematical representation of a fire. That is, the model usually requires that the user input a predefined growth rate of the fire. This requirement is the result of limitations associated with the modeling of the spread of flames on combustible surfaces, although as models improve in their predictive capabilities, this current situation will change [10]. In order to discuss the types of available fire models, it is helpful to understand the phenomena being studied.

What Are We Trying to Model?

All fires must behave according to fundamental principles, but can vary depending on the conditions in which they occur. It is the purpose of modeling to describe how a particular fire behaves based on fundamentals as well as the specific conditions.

There are two broad classes of the physical circumstances of a fire, that is, those fires that occur in the "open" or "unconfined" fires, and those fires that occur in an enclosed space, commonly referred to as *structure fires*, *compartment fires*, or *room fires*. More specific variables that characterize the environment include the fuel(s) involved, the configuration and arrangement of the fuel(s), the physical characteristics of the building, the location of the fire origin, the location and the capabilities of building occupants to mitigate the fire event, and any passive or active fire protection systems.

A good example of a fire burning in the "open" is a campfire. In the simplest terms, the wood of the campfire provides the gaseous fuel required for combustion and producing the volume of luminescent flames. This flaming region generates products of combustion (e.g., soot, carbon dioxide, carbon monoxide, and water vapor) that are transported vertically to the surrounding environment due to buoyancy through the formation of a thermal plume above the flaming region. (While buoyancy is a well-defined phenomenon, for the purposes of this discussion, the term *hot air rises* is sufficient to understand the concept.) The hot gases rise in the thermal plume and mix with the surrounding air until the temperature difference between the products of combustion and the surrounding air is zero. This heat transfer process is called *convective heat transfer*. The fire also transfers heat through thermal radiant energy to adjacent objects (e.g., the ground, human clothing, and skin) that will increase the temperature of those objects. In general, the energy transferred to the surrounding environment from a fire burning in the "open" is lost with no feedback process from the environment to the fire that can modify the fire's behavior.

In contrast, a fire burning in an enclosure will experience feedback from the local environment that can modify its burning characteristics. A good example of an enclosure fire is shown in Figure 1(a), which illustrates the cross section of a room with a floor, walls, a ceiling, and an open doorway. The room contains a wastepaper basket, a sofa, and a table. For this example, the wastepaper basket is the first item ignited and the sofa and table are potential secondary and tertiary fuels. In the initial stages of the fire growth, the thermal plume will rise and impinge on the ceiling surface above and will broadly redirect the flow from a predominately vertical direction to a predominately horizontal direction. As the hot gases flow horizontally in all directions from the ceiling impingement region, a thin layer is formed that is referred to as a *ceiling jet*. Initially, the ceiling jet will be relatively thin in depth, but as the ceiling jet reaches adjacent vertical walls, the flow will be generally redirected back toward the thermal plume and create a quiescent hot upper layer that is greater in depth than the original ceiling jet as shown in Figure 1(b). As the fire grows and as more products of combustion are produced over time from the fire, the hot upper layer will grow in a more uniform depth, and the interface between the hot upper layer and the cooler lower layer will descend.

As the hot upper layer descends and reaches the top of the open doorway, as illustrated in Figure 2(a), the difference in pressure on the inside of the doorway relative to the outside of the doorway causes the products of combustion from the hot upper layer

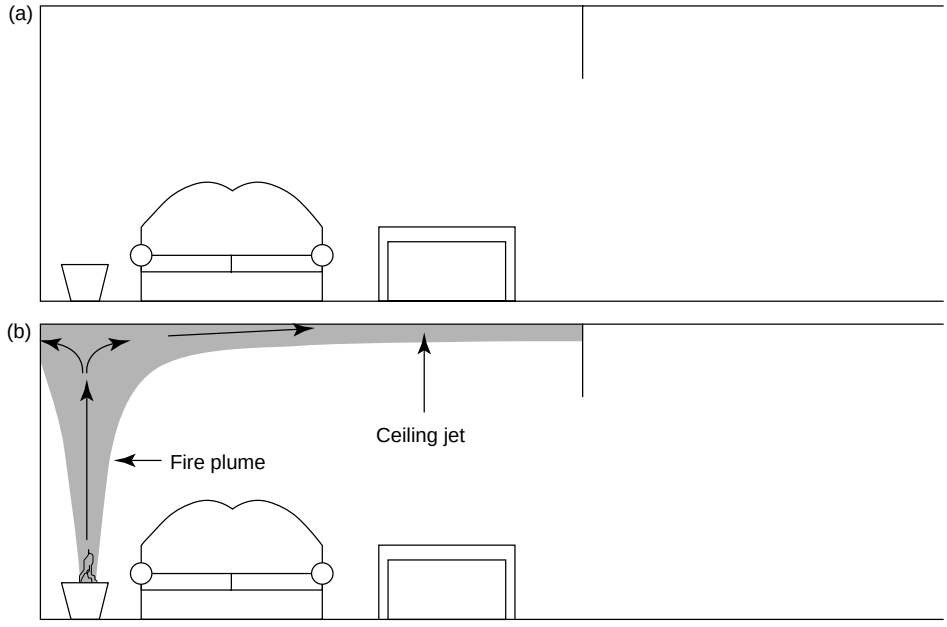


Figure 1 (a) An illustration of the cross section of a room with an open door and containing a wastepaper basket, a sofa, and a table. (b) An illustration of a flaming fire in the wastepaper basket forming a thermal plume above and producing a ceiling jet under a confined ceiling

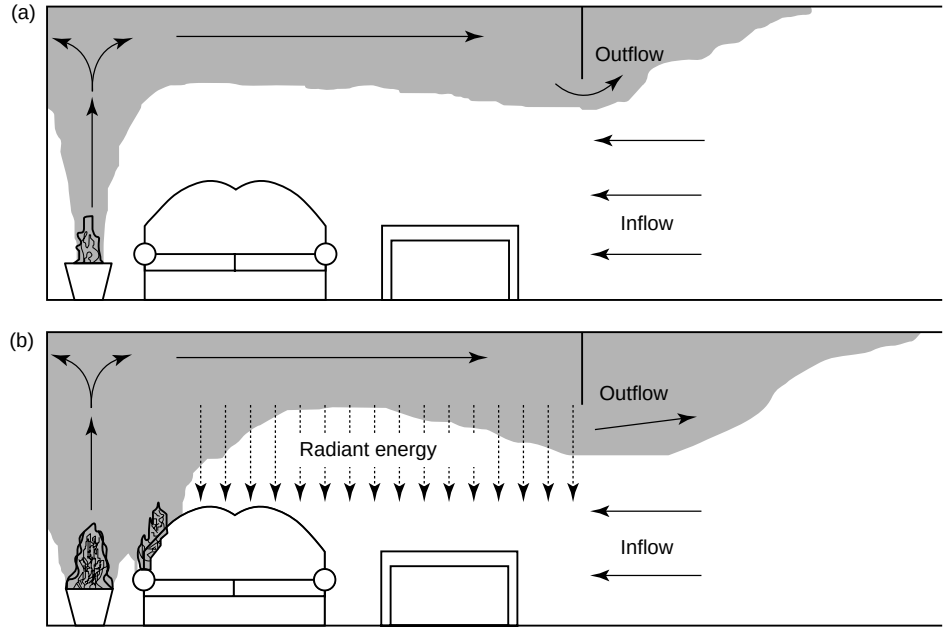


Figure 2 (a) An illustration of the increasing depth of the upper layer and the flow of hot products of combustion out of the top of the doorway. (b) An illustration of the growing fire and the descending interface between the hot upper layer and the cooler lower layer relative to the position of the ceiling

to flow out of the doorway. Owing to the law of conservation of mass, if the generation rate of the products of combustion from the fire does not exceed the exhaust rate of the products of combustion flowing out of the doorway, the hot upper layer will not descend any further. If the fire grows in size and the generation rate exceeds the exhaust rate, the bottom of the hot upper layer will continue to descend below the height of the doorway, as shown in Figure 2(b). The fire is analogous to a "pump" that produces and transports mass and energy to the hot upper layer from the fire source and as the fire grows, the size, and temperature of the hot upper layer will increase. As the upper layer increases in temperature, radiant energy becomes the dominant mode of heat transfer and will begin to transfer heat to the fuel below, as shown in Figure 3(a). Also, a well-defined flow pattern will be established at the doorway, with the hotter products of combustion flowing out the upper portion of the doorway and cooler ambient air flowing into the compartment in the lower portion of the doorway.

If the fire continues to grow and the upper layer temperatures approach approximately 500°C (900°F), there will be a continued increase in the intensity of the thermal radiant energy transferred to the exposed combustible contents in the room, and the surface temperature of these fuels will increase. When the upper layer temperature reaches approximately 600°C (1100°F), there is sufficient thermal radiant energy exposure (i.e., $20\,\mathrm{kW\,m^{-2}}$) to instantaneously ignite most combustibles in the room, as shown in Figure 3(b) [2]. This phenomenon, known as *flashover*, is illustrated in Figure 5. Flashover is not a discrete event in time, but an interval of time that is commonly characterized by (i) the transition from a localized fire to a general conflagration within the room when all fuel surfaces that can burn are burning; (ii) the transition from a fuel controlled fire, where there is plenty of air available and the fire is controlled by how much surface area is involved, to a ventilation controlled fire, where the size of the fire is determined by how much available air can reach the fire; (iii) the sudden propagation of flame through unburned hydrocarbons in the upper layer [8]. Once flashover conditions have been reached and the fire continues to grow in size, postflashover conditions or

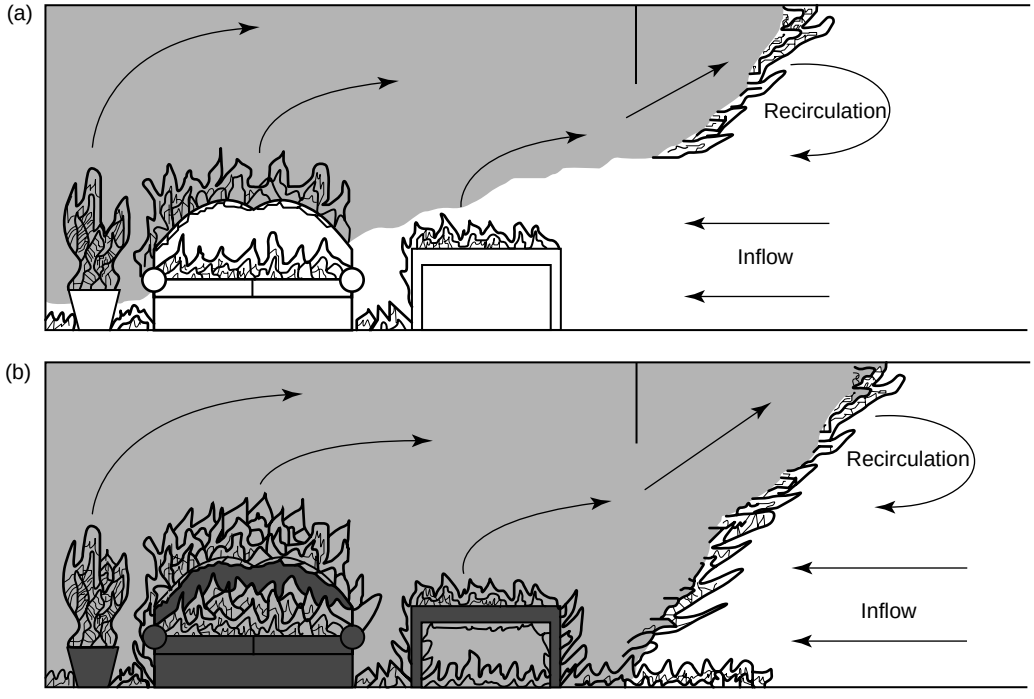


Figure 3 (a) An illustration of the onset of flashover due to thermal radiant energy from the upper layer. (b) An illustration of postflashover conditions or a fully developed compartment fire

full room involvement will follow in the majority of compartment fires unless the fuel is consumed, the fire is oxygen limited, or the fire is suppressed [2].

There are, of course, more phenomena associated with wanted and unwanted combustion that can be modeled and that are not addressed in this overview of fire modeling. Such additional phenomena can be used to test hypotheses associated with the investigation of fire and explosion incidents using the scientific method. The additional phenomena include patterns of thermal damage, activation of fire-detection devices, occupant egress from buildings, fire endurance of structures, generation of toxic gases, human respiratory uptake of toxic products, and fire suppression. The discussion of fire models for this article is necessarily limited to the commonly used models, with additional references for fire modeling included in the section "Further Reading" at the end of this article.

Types of Deterministic Fire Models

Deterministic fire models can range from simple correlations of empirical data to higher complexity models requiring weeks of computing time using multiple computer processors. The following sections provide an overview of the most commonly used types of fire models. Additional information on these and other simplistic correlations can be found in NFPA's *Fire Protection Handbook* [23] and SFPE's *Handbook of Fire Protection Engineering* [9].

Simple Correlations. A common example of a simple mathematical model is an equation describing the relationship between the heat release rate of the fire and the diameter of the surface area of the fire [24]:

$$\frac{L}{D} = 0.2\left(\frac{Q^{2/5}}{D}\right)$$

where, L is the average flame height in meters, D is the diameter of the fire in meters, and Q is the heat release rate in kilowatts. The correlation was developed on the basis of experimentally measured flame heights for different size fires (i.e., heat release rate and diameter).

FPEtool is the most well-known collection of simplistic computer simulation procedures that have been developed by multiple researchers [20]. Some of these correlations are for unconfined fires, including buoyant gas head [25], plume temperatures [26], rate of lateral flame spread [27], and radiant ignition of an adjacent fuel [28]. The remaining correlations are associated with confined fires and include average upper layer temperature [29]; ceiling jet temperatures [26]; mass flow through a vent [30]; smoke flow through an opening [25]; thermal detector and sprinkler response [31]; the occurrence of flashover [32]; and the limit on the size of a fire due to nonforced ventilation flows [33]. At the other end of the range of complexity are models that describe the behavior of fire in one or more rooms.

Deterministic mathematical fire models are further classified into two classes: (i) zone models and (ii) field models.

Zone Models. The zone model is the more common type of fire model available today. The dominant characteristic of the zone-type fire model is that it divides the individual compartment(s) into a relatively small number of discrete and relatively large zones that are implemented as control volumes. The most recognizable implementation is the two-zone model, where there is a hot upper layer and a cooler lower layer, as shown in Figures 4 and 5., but one-zone models have also been developed for both pre and postflashover zone models [34]. The size of and the conditions within each zone are calculated using the conservation equations (i.e., conservation of mass and energy) based on the fire source and thermal plume acting as a pump and transport mechanism for mass and energy to the upper layer by the thermal plume and ceiling jet. The model assumes that each zone is uniform with respect to the key variables such as temperature, soot, and gas concentrations (e.g., oxygen, carbon monoxide, and unburned hydrocarbons). In addition, the zone model assumes that when the mass and energy are generated from the fire source, they are instantaneously transported to the upper layer with no lag time associated with the transport. Thus, the transport time details are lost with the simplistic implementation of the assumptions associated with zone models [16]. The fundamental principles behind the zone fire model formulation are described by Quintiere [13].

While there are fundamental characteristics associated with the zone-type fire models, not all zone models have the same implementations with respect to the mathematical treatment of the relevant phenomena associated with compartment fires. For example, there are models that are able to handle only

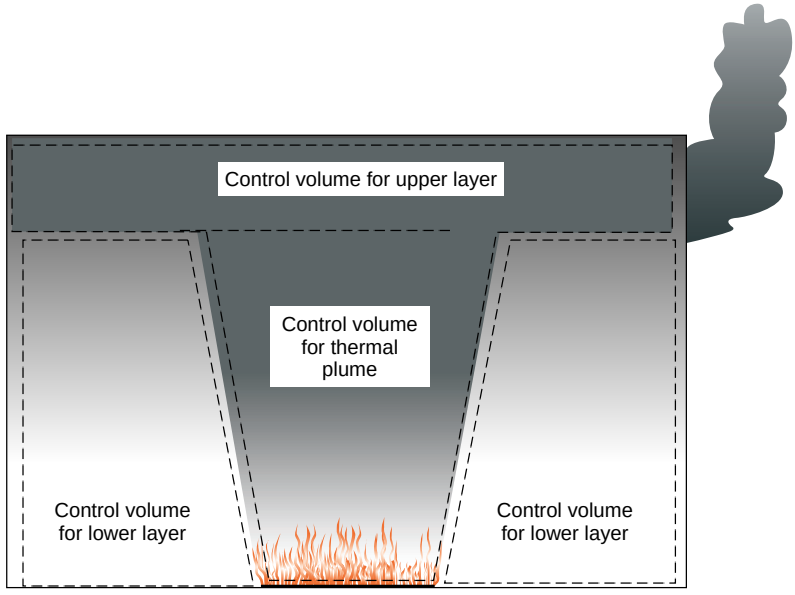


Figure 4 An illustration of a fire in a compartment with an opening on one end. The dotted lines show the zones or control volumes for the upper layer, the thermal plume, and the lower layer

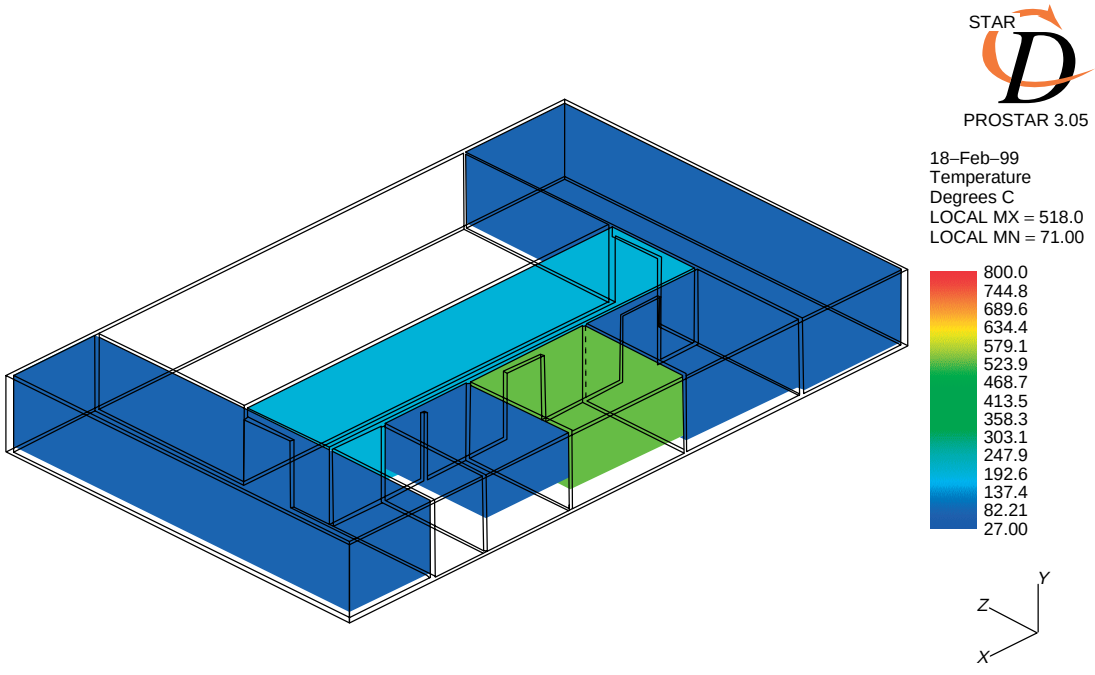


Figure 5 An illustration of a fire in a building in a middle room with an open doorway to the center hallway. There are three open doorways that connect other rooms to the center hallway and one partially closed doorway to an adjacent room. The depth of the upper layer interface is graphically illustrated for each room and the color gradient displays the average upper layer temperature

the compartment of fire origin, while others have multiroom capabilities. Other significant differences include the treatment of the combustion process and availability of oxygen, heat transfer between zones and to compartment surfaces, and pressure driven flows between zones and through vents such as doors, windows, and stairwells.

CFAST is one of the more well-known zone models in existence today. It is a multiroom model that predicts conditions within a structure resulting from a user-specified fire. CFAST Version 6 can accommodate up to 30 compartments with multiple openings between compartments and to the outside. The program inputs are a basic description of the building geometry (e.g., width, depth, and height); information describing common vents between compartments; the thermophysical properties of the compartment bounding surfaces; the heat release rate of the fire; and species yields of the products of combustion. The program calculates the average temperature, depth, and species concentrations in the hot upper layer and the cooler lower layer in each compartment. CFAST also includes very limited mechanical ventilation capabilities (i.e., HVAC systems), a ceiling jet algorithm, capability of multiple fires, heat transfer to other fuel targets, activation of thermal detection and suppression systems, and a flame spread model [21]. The CFAST model and associated technical documentation can be downloaded from the Internet (http://cfast.nist.gov/) at no cost.

Zone models, through their capacity to provide results relatively quickly on personal computers, allow for relatively inexpensive parametric studies to be conducted, and have proved to be a practical method for providing estimates of fire processes in enclosures.

Field Models. A field model differs from a zone model in that instead of a single point of averaged data for a finite area or volume, a "field" of multiple data points is resolved. The most identifiable field model for use in fire modeling incorporates the computational fluid dynamics (CFD) technique. Thus, the most important characteristic of a CFD fire model is the density of spatial resolution of the fire environment in terms of temperature and concentrations throughout the compartment relative to a zone model, which uses a single averaged data point. The CFD technique involves dividing up the volume of interest into relatively small and discrete control volumes, also known as *cells*, as shown in Figure 6. Like a zone model, the CFD technique utilizes the conservation equations (i.e., conservation of mass, momentum, and energy) to solve the flow variables at multiple points in the domain of interest, as shown in Figure 7. As the detailed structure is calculated on the basis of a more detailed implementation of the fluid mechanics equations, the CFD solution is more fundamentally based than zone models.

While many general-purpose CFD models have been developed and used in the modeling of fire, the most widely distributed and commonly used CFD model in the fire protection engineering community is FDS. FDS consists of two computer programs and

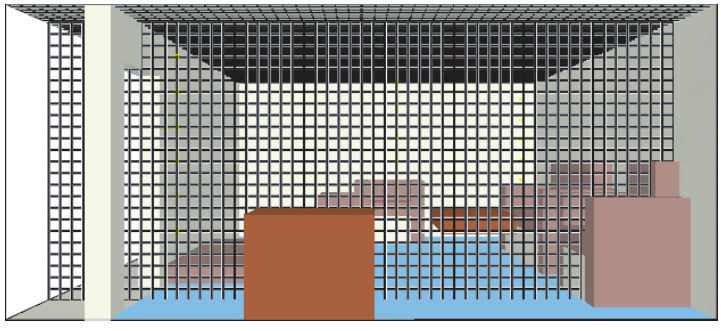


Figure 6 An example of the graphical output from the FDS model for a residential compartment fire with an open door. The size and number of cells in two dimensions are superimposed graphically in the model and are typical of the practical cell resolution associated with the modeling of a compartment fire

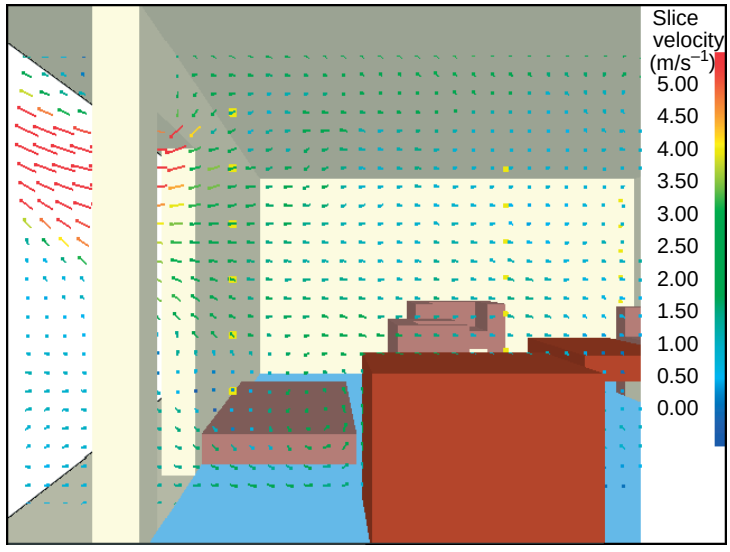


Figure 7 An example of the graphical data that can be superimposed onto the FDS model. Each graphic "arrow" is a vector within a plane or "slice" of the compartment that represents the predicted result for each of the selected cells. A vector provides the magnitude of the predicted value (in this case, the velocity of the flow) through its length and the direction through the orientation of the "arrow". The vectors within the doorway region in this case show the flow of hot gas out of the top of the doorway and the flow of relatively cooler gas (i.e., air) in at the lower portions of the doorway. The color (or gray-scale) gradient of the vectors is an additional means to visually characterize the velocity data

is designed to model fire-driven flows for low-speed, thermally driven flow with an emphasis on smoke and heat transport from fires. The first program is the actual FDS solver for the system of equations as described in McGrattan [22]. An input file is generated by the user that describes the geometry and sets up the simulation. This file is read by FDS and then run for a specified simulation time. The result is an output data file that can be viewed by the second computer program called *SmokeView* [35]. These two programs and associated technical documentation can be downloaded from the Internet (http://www.fire.nist.gov/fds) at no cost.

Current versions of FDS incorporate a more complex combustion model that releases fuel from the fire source and tracks the amount of fuel and oxygen in each cell. When the proportions of fuel and air are mixed in the right proportions, the fuel is burned (i.e., "mixed is burned"). This models assumes that the time to mix the fuel and air is much longer than the time for the fuel and air to react and is called *mixed is burned* model. FDS also handles thermal radiation, as well as the activation of thermal detectors and water spray from sprinklers.

While a CFD fire model provides greater resolution, to support more sophisticated applications of a model, there are trade-offs with respect to the required computer resources, the time required to run the models, the skill level of the user, and the associated costs of each. Thus, the use of CFD for all fire modeling applications is not always appropriate.

Overview on the Selection of Fire Models

Postfire reconstruction lends itself readily to the use of fire models. The appropriate selection and application of fire models to the investigation of fire and explosion incidents require an understanding of the fundamentals of fire dynamics, mathematical fire models, and their associated limitations with respect to the specifics of individual fire events. Several case studies of fire incidents using computer fire models have been published [36–41]. Such case studies generally demonstrate how a computer model used as part of a more general engineering analysis is a powerful tool. Most fire modeling applications that occur today, and that have occurred in the recent past, are performed in the context of the fire

litigation environment. Unfortunately, most of these applications are never published in the fire science literature.

Advantages and Limitations

The advantage of using fire models in fire investigation has been previously discussed with respect to their use within the context of the scientific method. The value of fire models is in their practical ability to address more complex phenomena, issues, and fire scenarios than other types of analytical tools. In short, fire models can help to quantify the limits of what can and cannot happen, based on scientific principles and analysis.

There are also limitations associated with the use of fire models. The current state of the art requires that the user provide the growth rate of the fire through a prescribed heat release rate history. Data sources for the thermophysical properties of commonly used building materials are somewhat limited. Similarly, the environmental conditions (geometry, fuel, and ventilation) for a specific fire scenario may not always have been documented at the fire scene or may not be available from other means. In addition, the uncertainty in the predictions is usually not obtainable for specific fire scenarios. The use of approximations in constructing a fire model also yields limitations with respect to the application to a specific scenario. The full limitations associated with the application to specific fire scenarios are, as yet, unknown. These issues involved in the evaluation of fire models have begun to receive attention in the technical fire community [10].

Any endeavor to model the behavior of fire also requires knowledge of the underlying physics and chemistry [5]. The previously discussed limitations tend to be overshadowed by the limitations associated with inappropriate use of fire models. Inappropriate use is usually the result of ignorance of fire dynamics and fire modeling. Fire modeling, as well as fire investigation, should only be conducted by individuals who have a sound knowledge of the principles of science and engineering and how these may be applied to understanding fire dynamics. Many fundamentals are intuitive, but misleading determinations can be drawn from postfire evidence and the results of fire modeling unless the investigator understands the science and dynamics of fire [42]. Thus, the onus of responsibility for appropriate application lies with the user of the model.

Model Selection

When solving technical problems, it can sometimes make good sense to limit the number of variables under study. It takes a high level of scientific knowledge of fire to ascertain which phenomena are important and which are not [17]. First and foremost, the appropriate selection of a fire model must include an assessment of the physics and chemistry required for the specific application and a determination of whether the model incorporates sufficient details for appropriate application.

Owing to the number of classes of models, a decision must first be made with respect to which class of models to use (e.g., zone or field model). For example, with respect to the prediction of the time for smoke detector activation, there are algorithms incorporated into both zone and field models. The selection of the most appropriate model needs to be based on the variables required to provide sufficient accuracy to predict the activation time. Accurate prediction requires that the local conditions (e.g., soot concentration and velocity) at the location of the smoke detector be resolved by the model to support the activation algorithm. A zone model would not be the most appropriate choice for this application because the upper layer (where the smoke detector would be located) is described by an averaged value. This limitation is not true for the field model. In addition, the zone model makes the assumption that the mass and energy generated by the fire source are immediately transported to the hot upper layer; thus, any transport time by the thermal plume and ceiling jet is ignored. As the transport time can be an important component to the time of smoke detector activation, the zone model would not be the most appropriate choice.

Once the selection of a class of models has been made, there can be a number of models within that class from which to choose. Again, this selection process must include an assessment of the Physics and Chemistry required for the specific application and a determination of whether a specific model incorporates sufficient details for appropriate application. Validation studies can provide some insight into this selection process.

Validation

The most important attribute of computer fire models is their ability to predict the relevant fire phenomena within their stated limitations. The predictive capability of a model depends on both the current state of fundamental knowledge for the phenomenon being modeled and the translation of that knowledge into a mathematical algorithm [10]. The predictive capability of fire models can be limited since these models rely on approximations for naturally occurring phenomena. There can also be additional uncertainty introduced depending on the knowledge and skill of the user. The level of uncertainty for a particular application can range from insignificant to rendering the model invalid and scientifically unreliable.

Validation is the process of measuring the level of uncertainty of the model's predictions. The general limitation of the validation process is that complete validation of any fire model for all fire scenarios is impractical. The American Society of Testing and Materials (ASTM) has developed guides to documenting and evaluating fire models [43, 44], and the SFPE has begun to take on the task of providing guidance documents on the use of fire models [45].

Numerous validation studies of computer fire models have been undertaken. The US Nuclear Regulatory Commission has begun the process of evaluating and validating models within the room of fire origin [46]. The model surveys by Friedman [18] and Olenick and Carpenter [19] also include reference to *validation studies*.

Summary

A failure can be defined as an "... unacceptable difference between expected and observed performance." Fire investigators and fire analysts are usually tasked with rendering opinions as to the origin, cause, responsibility, or prevention of such incidents. The essence of the investigation of fire and explosion incidents is embodied in the "... use of a systematic approach with attention to all relevant data". NFPA 921 recommends the use of the scientific method, which has traditionally been implemented in the physical sciences and forms the basis for scientific and engineering investigations, including fire and explosion incidents.

The challenge for the fire investigator is the practical application of fundamental knowledge to the investigation of fires that have occurred in the field. Such fundamental and empirical knowledge can be exploited through the development of models to simplify some of the complexities of the combustion phenomena that limit practical application.

In order to render accurate and reliable determinations associated with particular fire incidents, the investigator must "understand the fire." Fire models can be used as part of a fire dynamics analysis. The results can be compared to physical and eyewitness evidence to test an investigator's hypothesis. Modeling tools can also be used to evaluate the factors causing damage to property or injury to people (i.e., "The origin and cause of the loss").

A deterministic fire model is a mathematical representation of the processes encountered in "fire" by interrelated expressions based on known principles. Deterministic fire models can be simple correlations of empirical data or highly complex models requiring considerable computer processing power. The appropriate selection of fire modeling to the investigation of fire and explosion incidents involves an understanding of the fundamentals of fire dynamics, mathematical fire models, and their limitations with respect to the specifics of individual fire events.

Acknowledgments

Special thanks to technical writer, Nancy Milarcik Leyko of Combustion Science & Engineering, Inc. for her detailed review and comments that were helpful in providing a more technically comprehensible manuscript for a complex subject.

References

[1] Leonards, G.A. (1982). Investigation of failures, *Journal of the Geotechnical Engineering Division, ASCE* **108**(GT2), 187–246.

[2] NFPA 921 (2008). *Guide for Fire and Explosion Investigations*, National Fire Protection Association (NFPA), Quincy.

[3] Glassman, I. (1996). *Combustion*, 3rd Edition, Academic Press, New York.

[4] Cox, G. (1995). *Combustion Fundamentals of Fire*, Academic Press, New York.

[5] Friedman, R. (1998). *Principles of Fire Protection Chemistry and Physics*, 3rd Edition, National Fire Protection Association (NFPA), Quincy.

[6] Quintiere, J.G. (2006). *Fundamentals of Fire Phenomenon*, John Wiley & Sons, West Sussex.

[7] Karlsson, B. & Quintiere, J.G. (2000). *Enclosure Fire Dynamics*, CRC Press, New York.

[8] Drysdale, D. (1999). *An Introduction to Fire Dynamics*, John Wiley & Sons, West Sussex.

[9] SFPE (2008). *The SFPE Handbook of Fire Protection Engineering*, 4th Edition, Society of Fire Protection Engineers (SFPE), Bethesda, National Fire Protection Association (NFPA), Quincy.

[10] Beyler, C.L., DiNenno, P.J., Carpenter, D.J. & Watts Jr, J.M. (2008). Introduction to fire modeling, in *NFPA Fire Protection Handbook*, 20th Edition, National Fire Protection Association, Quincy, Section 3, Chapter 5.

[11] Hinkley, P.L. & Thomas, P.H. (1964). *Design of Roof-Venting Systems for Single Story Buildings, Technical Paper No. 10*, Fire Research Station, Borehamwood, p. 51.

[12] Richardson, J.K. (2003). *History of Fire Protection Engineering*, National Fire Protection Association (NFPA), Quincy, Society of Fire Protection Engineers (SFPE), Bethesda.

[13] Quintiere, J. (1977). ASTM STP 614, *Growth of Fires in Building Compartments*, American Society for Testing and Materials, Philadelphia.

[14] Pape, R., Waterman, T.E. & Eichler, T.V. (1981). *Development of a Fire in a Room from Ignition to Full Room Involvement – RFIRES*, NBS-GCR-81-301, National Bureau of Standards, Washington, DC.

[15] Mitler, H.E. & Emmons, H.W. (1981). *Documentation for CFC V: The Fifth Harvard Computer Fire Code*, NBS GCR 81–344, Home Fire Project Technical Report 45. National Bureau of Standards, Gaithersburg, p. 187, October.

[16] Walton, W.D. (2002). "Zone Computer Fire Models for Enclosures", *The SFPE Handbook of Fire Protection Engineering*, third edition, Society of Fire Protection Engineers (SFPE), Bethesda, National Fire Protection Association (NFPA), Quincy.

[17] Emmons, H.W. (1985). The needed fire science, *International Association for Fire Safety Science, Fire Safety Science, Proceedings, 1st International Symposium*, October 7–11, Gaithersburg, C.E. Grant & P.J. Pagni, eds, Hemisphere Publishing, New York, pp. 33–54.

[18] Friedman, R. (1992). International survey of computer models for fire and smoke, *Journal of Fire Protection Engineering* **4**(3), 81–92.

[19] Olenick, S.M. & Carpenter, D.J. (2003). Updated international survey of computer models for fire and smoke, *Journal of Fire Protection Engineering* **13**(2), 87–110.

[20] Deal, S. (1995). *Technical Reference Guide for FPEtool Version 3.2*, National Institute of Standards and Technology, Gaithersburg, NISTIR 5486-1, p. 137 April.

[21] Jones, W.W. Peacock, R.D., Forney, G.P. & Reneke, P.A. (2005). *CFAST: Consolidated Model of Fire Growth and Smoke Transport (Version 6). Technical Reference Guide*, NIST SP 1026, National Institute of Standards and Technology, Gaithersburg, p. 146. December.

[22] McGrattan, K.B. (2007). *Fire Dynamics Simulator (Version 5) Technical Reference Guide*, NIST Special Publication 1018-5, National Institute of Standards and Technology, Gaithersburg, p. 100, October.

[23] NFPA (2008). *Fire Protection Handbook*, 20th Edition, National Fire Protection Association, Quincy.

[24] Heskestad, G. (1983). Luminous heights of turbulent diffusion flames, *Fire Safety Journal* **5**, 103–108.

[25] Klote, J.H. & Milke, J.A. (1992). *Design of Smoke Control Management Systems*, ASHRAE, SFPE, Atlanta, pp. 21–32.

[26] Alpert, R.L. & Ward, E.J. (1983). *Evaluating Unsprinklered Fire Hazards*. SFPE Technology Report 83-2. Society of Fire Protection Engineers, Boston.

[27] Quintiere, J.G. & Harkleroad, M.F. (1984). New concepts for measuring flame spread properties, in *Symposium on Application of Fire Science to Fire Engineering: American Society for Testing and Materials (ASTM)*, Society of Fire Protection Engineers (SFPE), Denver.

[28] Babrauskas, V. (1982). *Will the Second Item Ignite?* NBSIR 81–2271, National Bureau of Standards, Gaithersburg.

[29] McCaffrey, B.J., Quintiere, J.G. & Harkelroad, M.H. (1981). Estimating room temperatures and the likelihood of flashover using fire test data correlations, *Fire Technology* **7**(2), 98–119.

[30] Lawson, J.R. & Quintiere, J.G. (1985). *Slide-rule Estimates of Fire Growth*, NBSIR 85–3196, National Bureau of Standards, Gaithersburg, pp. 56.

[31] Evans, D.D. & Stroup, D.W. (1985). *Methods to Calculate the Response of Heat and Smoke Detectors Installed Below Large Unobstructed Ceilings*, National Bureau of Standards (NBSIR) 85–3167, Gaithersburgm .

[32] Thomas, P.H. (1981). Testing products and materials for their contribution to flashover in rooms, *Fire Research, Borehamwood, England Fire and Materials* **5**(3), 103–111.

[33] Tewarson, A. (1988). Generation of heat and chemical compounds in fires, *The SFPE Handbook of Fire Protection Engineering*, 1st Edition, Society of Fire Protection Engineers (SFPE), Bethesda, National Fire Protection Association (NFPA), Quincy.

[34] Walton, W.D. & Thomas, P.H. (2002). Estimating temperatures in compartment fires, *The SFPE Handbook of Fire Protection Engineering*, 3rd Edition, Society of Fire Protection Engineers (SFPE), Bethesda, National Fire Protection Association (NFPA), Quincy.

[35] Forney, G.P. (2007). *User's Guide for Smokeview Version 5: A Tool for Visualizing Fire Dynamics Simulation Data*, National Institute of Standards and Technology,

Gaithersburg, NIST Special Publication 1017-1, p. 134, August.

[36] Bukowski, R.W. & Spetzler, R.C. (1992). Analysis of the happyland social club fire with HAZARD I, *Journal of Fire Protection Engineering* **4**(4), 117–131.

[37] Nelson, H.E. (1989). An engineering analysis of fire development in the hospice of Southern Michigan, December 15, 1985, in *Fire Safety Science – Proceedings of the 2nd International Symposium*, Hemisphere Publishing Corporation, New York.

[38] Levine, R.S. & Nelson, H.E. (1990). *Full-Scale Simulation of a Fatal Fire and Comparison of Results with Two Multiroom Models*, NISTIR 90–4268, National Institute of Standards and Technology, Gaithersburg.

[39] Alvares, N. (1995). Defining fire and smoke spread dynamics in the DuPont Plaza fire of 31 December 1986, in *Proceedings of the International Conference on Fire Research and Engineering*, Society of Fire Protection Engineers, Boston, MA, September 10–15.

[40] Sutula, J.A. (2002a). Applications of the fire dynamics simulator in fire protection engineering consulting, in *Fire Protection Engineering*, Society of Fire Protection Engineers, Bethesda. No. 14, Spring.

[41] Sutula, J.A. (2002b). Practical applications of the fire dynamics simulator in fire protection engineering consulting, in *Fire Protection Engineering*, Society of Fire Protection Engineers, Bethesda. No. 15, Summer.

[42] Drysdale, D. (2004). Fire dynamics and fire investigation, in *Proceedings of 1st International Symposium on Fire Investigation*, 27th–30th June, Fire Service College, Moreton-in-Marsh, Gloucestershirc.

[43] ASTM (1995). *Standard Guide for Documenting Computer Software for Fire Models*, ASTM E 1472, American Society for Testing and Materials, Philadelphia.

[44] ASTM (1998). *Standard Guide for Evaluating the Predictive Capabilities of Fire Models*, ASTM E 1355, American Society for Testing and Materials, Philadelphia.

[45] SFPE (2002). *Engineering Guide to the Evaluation of the Computer Model DETACT-QS, Society of Fire Protection Engineers*, Society of Fire Protection Engineers, Bethesda.

[46] NRC (2006). *Verification and Validation of Selected Fire Models for Nuclear Power Plant Applications*, Vol. 1 Main Report, Nuclear Regulatory Commission, Washington, DC, Electric Power Research Institute, Palo Alto, CA.

Further Reading

Anderson Jr, J.D. (1995). *Computational Fluids Dynamics: The Basics with Applications*, McGraw-Hill, New York.

Beard, A.N. (1995/1996). Limitations of fire models, *Journal of Applied Fire Science* **5**(3), 233–243.

Bukowski, R. (1991). Fire models: the future is now, *National Fire Protection Association (NFPA) Journal* **85**(5), 60–62, 64, 66–69.

Cox, G. (1995). Compartment Fire Modelling, in *Combustion Fundamentals of Fire*, G. Cox, ed, Academic Press, New York, Chapter 6, pp. 329–404.

Evans, D.D. (2000). *Use of Fire Simulation in Fire Safety Engineering and Fire Investigation*, National Institute of Standards and Technology, Gaithersburg, NISTIR 6588, November 2000.

Janssens, M.L. (2000). *An Introduction to Mathematical Fire Modeling*, 2nd Edition, Technomic Publishing Company, Lancaster, PA.

Joglar, F., Gautier, B., Gay, L. & Texeraud, J. (2007). *Verification and Validation of Selected Fire Models for Nuclear Power Plant Applications. Volume 6. MAGIC*, Electric Power Research Institute, Palo Alto, CA, Nuclear Regulatory Commission, Washington, DC, NUREG-1824, EPRI 1011999, p. 206, May, 2007.

Jones, W.W. (2001). *State of the Art in Zone Modeling of Fires*, National Institute of Standards and Technology, Gaithersburg, Vereinigung zur Forderung des Deutschen Brandschutzes e.V. (VFDB), International Fire Protection Seminar, 9th, Engineering Methods for Fire Safety, Proceedings, May 25–26, 2001, Munich, Germany, pp. A.4/89–126.

Mitler, H.E. (1985). The Harvard fire model, *Fire Safety Journal* **9**, pp. 7–16.

National Academy of Sciences (1959). Use of models in fire research. *Conference Proceedings*, Washington, DC, November 8–10, 1959.

Nelson, H.E. & Deal, S. (1991). Comparing compartment fires with compartment fire models, in *International Association for Fire Safety Science, Fire Safety Science Proceedings of the 3rd International Symposium*, July 8–12, 1991, Edinburgh, Scotland, G. Cox & B. Langford, eds, Elsevier Applied Science, New York, pp. 719–728.

Nelson, H.E. (2002). *From phlogiston to computational fluid dynamics*, in *Fire Protection Engineering* No. 13, Winter.

Peacock, R.D. & Reneke, P.A. (2007). *Verification and Validation of Selected Fire Models for Nuclear Power Plant Applications. Volume 5. Consolidated Fire Growth and Smoke Transport Model (CFAST)*, National Institute of Standards and Technology, Gaithersburg, Nuclear Regulatory Commission, Washington, DC, NUREG-1824, EPRI 1011999, p. 206, May (2007).

Quintiere, J.G. (1989). Fundamentals of enclosure fire 'Zone' models, *Journal of Fire Protection Engineering* **1**(3), 99–119.

Quintiere, J.G. (2002). Compartment fire modeling, 3rd Edition, P.J. DiNenno, eds, National Fire Protection Association, Quincy.

Spearpoint, M.J., Mowrer, F.W. & McGrattan, K.B. (1999). Simulation of a Compartment Flashover Fire Using Hand Calculations, Zone Models and a Field Model, *International Conference on Fire Research and Engineering (ICFRE3), Third (3rd). Proceedings, Society of Fire Protection Engineers (SFPE), National Institute of Standards and Technology (NIST) and International Association of Fire Safety Science (IAFSS). October 4–8, 1999, Chicago, IL*, Society of Fire Protection Engineers, Boston, MA, pp. 3–14.

Walton, W.D. (1985). *ASET-B A Room Fire Program for Personal Computers*, NBSIR 85–3144, National Bureau of Standards, Washington, DC.
Zukoski, E.E. & Kubota, T. (1980). Two-layer modeling of smoke movement in building fires, *Fire and Materials* **4**(1), 17.

Related Articles

Fire and Explosion Investigations: Overview
Fire: Chemistry of
Fire: Dynamics and Pattern Production
Fire: Scene Investigation

DOUGLAS J. CARPENTER

Firearm Discharge Residue: Analysis of

Introduction

Numerous serious criminal offenses as well as many terror activities involve firearms. During discharge of a firearm, a variety of materials (accompanying the bullet) are emitted from the muzzle and other possible openings in the firearm. Throughout this text, the term *firearm discharge residues* (*FDR*) is used for all these materials. FDR may originate from primer, gunpowder (propellant), lubricant and metals of the bullet, cartridge case, and the gun barrel. The term *gunshot residues* (*GSR*) is used in this text for primer residues only, although in the relevant literature GSR or cartridge discharge residues (CDR) are sometimes used instead of FDR, whereas the term *inorganic GSR* refers to *primer residues* and the term *organic GSR* refers to *propellant residues*. Analysis of FDR on relevant exhibits can assist in reconstruction of a crime and might provide evidence against suspects involved. For instance, detection of FDR on a suspect or his clothing may link him to a firearm-related crime.

The following topics are reviewed in this treatment:

1. Detection and analysis of primer residues (GSR);
2. detection and analysis of gunpowder (propellant) residues;
3. the techniques of associating firearms and ammunition with gunshot entries; and 4. estimation of the time since shooting.

Several publications provide a comprehensive background on some of these topics [1–7]. Detection of firearm imprints on the hands of a suspect may assist in linking him to a firearm-related event. Estimation of shooting distance may be useful in reconstruction of the crime. Both topics are not in the scope of this treatment and appear as separate contributions in the encyclopedia (*see* **Firearms: Identification of Handling of Firearms/Trace Metal Detection**; **Shooting Distance: Estimation of**).

The Chemistry of Firearms Ammunition

Cartridge Cases

The cartridge case is designed to house the primer and propellant and to securely retain the bullet (projectile) in the neck. The vast majority of cartridge cases are made of brass (approximately 70% copper and 30% zinc) but other materials such as steel, coated with zinc, brass, gilding metal (approximately 90% copper and 10% zinc), or copper; nickel-plated brass; cupronickel (approximately 80% copper and 20% nickel); gilding metal; aluminum; Teflon-coated aluminum; and plastic are also encountered.

Projectiles

Conventional projectiles for firearms are bullets. Pellets and slugs are used for shotguns. A wide range of bullet types is available based on differences in composition and the geometrical design of the bullet core and bullet jacket. Bullets are unjacketed, jacketed, or partially jacketed (semijacketed). When firearms had relatively low muzzle velocities, bullets were made of almost pure lead. Commercial lead-alloy bullets are hardened by alloying with antimony or, less commonly, with tin. The concentration of antimony may vary considerably, from 0.5 up to 12%. Unjacketed bullets may have their surfaces coated with a very thin layer of copper or brass. This is referred to as a *wash* or *coat* and is not a bullet jacket

in the conventional sense of the word. Unjacketed bullets are frequently lubricated with some form of wax or grease to prevent or reduce fouling in the barrel of the firearm. Jacketed bullets consist of a lead core surrounded by a jacket of harder material. Usually the core is left exposed at the base of the bullet; however, the bases of total metal jacket (TMJ) bullets are completely covered by the jacket. Jackets are commonly made of copper, brass, copper–nickel alloy, or mild steel.

Primers

Primers for center fire ammunition are housed in small metal cups that fit into a recess (primer pocket) in the center of the base of the cartridge case. In rimfire ammunition, the priming composition is housed inside the cartridge case in the hollow perimeter of the base. Priming compositions are mixtures which when subjected to percussion provide sudden burst of flame that serves to ignite the propellant within the cartridge case.

In general, small arms primers consist of the following components: an initiating explosive (initiator), an oxidizer, a fuel, a sensitizer, a friction material, and a binder. Examples of those types of compounds are the following:

Initiators:
lead or silver azide, mercury fulminate, lead styphnate, diazodinitrophenol.

Oxidizers:
barium nitrate, barium peroxide, potassium chlorate, lead dioxide.

Fuels:
antimony sulfide, gum arabic, calcium silicide, aluminum powder.

Sensitizers:
tetracene, trinitrotoluene (TNT), pentaerythritol tetranitrate (PETN), tetryl.

Friction materials:
ground glass, aluminum powder.

Binders:
gum arabic, dextrin, glue.

Sometimes, in the priming composition, more than one compound from any of the above classes may be used. Modern primers are mostly noncorrosive and nonmercuric (NCNM). A typical centerfire primer cap produced today might contain lead styphnate, antimony sulfide, barium nitrate, and tetracene. Not all primers, however, contain compounds of antimony and barium. Some. 22 caliber rimfire cartridges contain either lead and barium or only lead compounds.

In 1980, Dynamit Nobel AG commercialized a new ammunition, called *Sintox*^{TM}, developed to minimize airborne lead levels and possibly other metallic residues such as barium and antimony [5]. In this ammunition, lead styphnate is replaced by diazodinitrophenol and barium nitrate and antimony sulfide are replaced by a mixture of zinc peroxide and titanium metal powder respectively [7]. This primer also contains tetracene and cellulose nitrate. The use of this primer coupled with a TMJ bullet eliminates the health hazard problem. Since then several other ammunition manufacturers have started introducing lead-free ammunition.

Propellants

Modern smokeless gunpowder for small arms ammunition almost exclusively contains cellulose nitrate (nitrocellulose, NC) as the main explosive component (single based). Other explosive ingredients may also be present, for example, glycerol trinitrate (nitroglycerine, NG) (double based) and nitroguanidine (triple based, unlikely to be encountered in small arms ammunition). Propellants also contain stabilizers such as diphenylamine (DPA) or ethyl centralite (EC), flash inhibitors such as 2,4-dinitrotoluene (2,4-DNT), and 2,6-dinitrotoluene (2,6-DNT) and plasticizers such as diethyl phthalate and dibutyl phthalate. Some inorganic materials such as chalk, graphite, potassium sulfate, potassium nitrate, and barium nitrate are added to improve ignitability, facilitate handling, and minimize muzzle flash.

The Mechanism of FDR Formation

When the firing pin strikes the primer of a cartridge, it provides a burst of flame without the development of a detonating wave [7]. The flame causes the propellant to ignite. It burns fast (deflagration) but does not detonate. When the bullet leaves the muzzle of a firearm, the pressure is vented to the atmosphere. This typically occurs before all of the propellant is consumed, causing fragments of unburned gunpowder to be emitted from the firearm. Other particulates

originate from the decomposition of the primer and from the metals of the cartridge and the firearm. The process of their formation may be described in the following manner [8]: the expanding gases from the propellant deflagration initially compress the bullet axially, so that it expands radially. The enlarged bullet is then "swaged" through the barrel by the rapidly increasing gas pressure. This results in strong frictional heating. Also, the rifling rips small fragments from the bullet, and these particles are in part melted and in part vaporized. The metal vapors from the projectile mix with the vapors of the decomposition products of the primer and are ejected from all possible openings in the firearm and may be deposited on surfaces in close proximity to the firer. In flight, the vapors condense into liquid phase and then solidify to particles. Unless modified in the liquid state by impact, they appear as spheroids in the size ranging from less than 1 up to 10 μm [2, 9]. Thus, the composition of the formed particles may have contributions from the primer, the bullet, the cartridge case, the gun barrel, and the propellant, although the term *primer residue* is occasionally used.

Detection and Analysis of GSR (Primer Residues)

This section discusses various methods of detection and characterization of primer residues mainly on suspects, their clothing, and their belongings. The focus of the discussion is on GSR particle analysis by scanning electron microscopy combined with energy dispersive X-ray analysis (SEM/EDX), since nowadays this is the method of choice in advanced crime laboratories.

Color Tests

The earliest method to detect FDR on shooters' hands was introduced in 1933 by Teodoro Gonzales [5]. It is known as a *dermal nitrate test* or *paraffin test*. Molten paraffin was poured on the hands of the subject and allowed to solidify to form a cast on the hands. When the paraffin froze, it was peeled off and sprayed with a 0.2% solution of DPA in concentrated sulfuric acid. The reagent gives a deep blue coloration with nitrates and nitrites that may originate from gunpowder and primer residues. However, even though nitrates and nitrites may come from other sources (e.g., fertilizers, urine), this test is not specific to nitrates and nitrites only. Other oxidizers like chlorates, iodates, permanganates etc. may give a positive reaction. On the other hand, also false negative results were common. In spite of these drawbacks, the test has been used for many years. Finally, it was abandoned by most of the forensic laboratories. However, the test is still employed in number of places in the world [2]. Recently, the US Department of Energy's Sandia National Laboratories, working with Law Enforcement Technologies Inc., announced the development of the ISID™-1 (Instant shooter ID Kit). It appears that it is an updated version of the "dermal nitrate" or "paraffin" test. As in the case of the "dermal nitrate" test it was argued that "The concerns of non-specificity of the ISID™-1 kit need to be critically evaluated in light of its potential usefulness in the field and any legal issues associated with its documented 'false positives"' [10].

The Griess reagent used to detect nitrites is discussed in a separate contribution of the encyclopedia (*see* **Shooting Distance: Estimation of**).

In 1959, a test was developed for detection of lead, barium, and antimony, which may be found in modern ammunition primer residues [11]. The method did not gain wide acceptance because of lack of specificity of the color reactions, insufficient sensitivity, interference of the color reactions among the three elements, and the instability of the colors developed.

Bulk Instrumental Analytical Methods

Lead, antimony, and barium may be found on the hands of persons who were not involved with firearms but generally in smaller quantities compared to firers. Thus, developing a quantitative sensitive method for those elements could assist in differentiating between "innocent" people and those involved in shooting.

In this paragraph, bulk instrumental analytical methods are discussed. The term applies to those methods that determine bulk concentration of the constituents in the analyzed sample. With regard to GSR analysis, the analyzed sample is a collection of many particles. Contrary to the bulk method, a method for particle analysis, like SEM/EDX, can analyze discrete particles in the sample.

In the early 1960s, neutron activation analysis (NAA) was applied for detection of trace elements with high sensitivity [5, 12]. Although this method

was not applicable for lead, it was used to determine minute quantities of antimony and barium. In 1971, atomic absorption spectrophotometry (AAS) was developed for GSR analysis. Conventional flame AAS had sufficient sensitivity for detection of the lead quantities on firers' hands, but it was not sensitive enough for detection of antimony and barium. Development of graphite furnace atomic absorption spectrophotometry (GFAAS) solved this problem. GFAAS is relatively time consuming and is subject to some interferences due to the presence of high concentrations of nonanalyte constituents. Many forensic science laboratories have used NAA or GFAAS for GSR analysis in casework.

Some other bulk analytical methods have been studied recently for GSR analysis in hand samples [5]. Among these, two methods have been implemented in casework: inductively coupled plasma atomic emission spectroscopy (ICP/AES) [13] and inductively coupled plasma mass spectrometry (ICP/MS) [14, 15]. ICP/AES is a rapid, multielement technique that is relatively free of matrix interferences, but commercial instruments lack the sensitivity required for accurate determination of antimony at the relevant levels [14]. ICP/MS is now a widely accepted method for trace and ultratrace analysis of a variety of liquid, gaseous, and solid samples in many applications. Analytical benefits of the method are low detection limits, wide linear dynamic range, relative freedom from interferences, and good precision and accuracy [14]. Nonetheless, the latest survey [16] showed that the most popular bulk analytical method for GSR analysis was GFAAS.

GSR Particle Analysis by SEM/EDX

Up to the mid-1970s, the accepted methods for GSR analysis in casework were bulk analytical techniques described above. In the second half of 1970s, several studies were reported indicating that much higher evidential value could be obtained by examining the composition and morphology of discrete particles formed during a firearm discharge with the aid of SEM/EDX analysis. Continuous improvements in the SEM and EDX technologies are also reflected in the GSR analysis by SEM/EDX. Nowadays, most manufacturers of the SEM/EDX equipment offer automated search systems for GSR particles, which makes the analysis much easier when compared to manual search, which is very tedious. The principle of the automated search is similar to the manual search, namely, detection of a suspected particle by the backscatter electron (BSE) signal and its analysis by EDX. Nevertheless, even with the automated search systems the analysis may be quite long, amounting to several hours per square centimeter of a sample, depending on the surface density of particles having high average atomic number.

Sampling of GSR and Their Persistence on Various Surfaces. To collect GSR samples for SEM/EDX, most of the crime laboratories use 25- or 13-mm-diameter SEM aluminum stubs coated with adhesive tape (tape-lift method) [16]. The coated stubs are kept in GSR collection kits, which are commercially available or may be prepared in the laboratory. To collect particles from a surface (e.g., a hand), the stub is pressed repeatedly against the surface (see subsequent discussion).

The tape-lift method was shown to be the most efficient method of collection among three examined methods (tape-lift, glue-lift, and concentration methods) [17]. It was reported that in practice there might be an intrinsic difficulty in concentrating GSR on a small area due to pile-up and clogging of the filters used [18].

In the United Kingdom, swabs have been used to recover GSR and gunpowder residues from skin surfaces of a suspect. The propellant residues are extracted from the swabs by organic solvents and filtered through membrane filters on which the GSR particles are collected and examined by SEM/EDX [19].

A comprehensive study was conducted on collection efficiency of GSR particles from hair and hands using double-sided adhesive tape [20]. Tape-lift was found to be a suitable method for collecting GSR from hair, contrary to the previous claim [21]. No significant difference was found between collection efficiency of this technique and the more complicated method of swabbing the hair by use of a comb with a solvent-dampened cloth and filtration, proposed elsewhere [21]. It was found that 200–300 dabbings are necessary to achieve maximum collection efficiency from hair with the double-sided adhesive.

In several studies, the recommended collection procedure from hands by tape-lift method is to press the stub repeatedly against the hand until it has lost its stickiness per subjective assessment [1] The concern regarding continual dabbing was that skin debris

might conceal GSR particles from view. Nonetheless, it was found [20] that 50–100 dabbings are necessary to achieve maximum collection efficiency from hands, while stickiness appeared lost after about 20–30 dabbings. In a subsequent study [22], it was demonstrated that there is no substantial danger of concealing GSR particles by continuous dabbings of hands up to 50 times.

Clothing and other surfaces may be sampled for GSR by tape-lift method or by vacuum filtration [19, 23].

Once GSR particles are deposited on the hands, they are continuously lost as a result of the normal activities of a living person [12]. It is difficult to generalize as to how long GSR would be retained since it depends on the type of activity [12, 24]. While laboratory tests using bulk analytical methods have indicated that GSR remain on the hands for only a short time (perhaps 1 or 2 h) in actual cases, detectable amounts have been found several hours later, perhaps due to higher initial amounts of GSR [12]. Using particle analysis, it was reported that the probability of finding GSR on hands more than 3 h after firing is very low [21]. A subsequent study [25] has confirmed that there is a rapid drop in GSR particle count as a function of postfiring collection time, with the largest loss occurring in the first 3–4 h. Laboratory experiments showed that 30 min after firing no significant difference was observed between samples collected from the right and left hands with regard to the number of GSR particles found, although the shooter used one hand for firing [21]. The accumulated casework experience of about 12 years in Israel has shown that in only a very few cases a large number of particles was found in hand samples and in significantly lower number of these cases considerably more particles were found on one hand as compared to the other one, so that some inference could be made concerning the firing hand [26].

A much longer persistence of GSR was found on hair and clothing [21]. Persistence of 24 h was observed in laboratory experiments when the hair had not been washed [20]. So far few crime laboratories have introduced sampling hair for GSR, and this topic was not discussed in the recent meeting on GSR [27].

Compositions, Classification, and Interpretation. As was pointed out earlier, the composition of GSR particles may have contributions from several sources. It was shown that when using lead bullets in ammunition 70–100% of GSR particles are lead [8]. If the bullet is coated with copper or brass, the result is the same except that a substantial portion of all the particles contains some copper. If jacketed or partially jacketed bullets are used, the fraction of lead particles in the total collection is greatly reduced. Only a small portion of the particles contains copper and the total number of particles is smaller as well. It could be concluded from these observations that most of the lead in GSR comes from the projectile rather than from the primer. In another study [28] using X-ray diffraction (XRD), it was found that lead in the metallic form is found to be the main constituent of all GSR tested. The significant difference found between the GSR compounds' composition formed by ordinary shots (involving bullets) and primer shots can be explained by the far higher temperature and pressure of the former, resulting in almost total combustion of primer components as well as involvement of the projectile [28].

In their pioneering extensive study, Wolten *et al.* [1, 29, 30] proposed a classification scheme for GSR. They divided GSR into two categories: (i) unique or characteristic and (ii) consistent. The compositions they considered to be unique are as follows:

1. Pb, Sb, Ba
2. Ba, Ca, Si with traces of S
3. Ba, Ca, Si with traces of Pb if Cu and Zn are absent
4. Sb, Ba.

The compositions considered as consistent:

1. Pb, Sb
2. Pb, Ba
3. Pb
4. Ba if S is absent or present only as a trace
5. Sb (rare).

In both categories, one or several of the following and only the following elements could also be present: Si, Ca, Al, Cu, Fe, S, P (rare), Zn (only if Cu is also present), Ni (rare and only with Cu and Zn), K, and Cl.

It should be emphasized that the concentrations of the elements in the GSR particles are highly variable due to the very fast process of their formation.

The modifications proposed by Wallace and McQuillan [31] resulted only in two alleged "unique"

compositions for GSR: (a1) Pb, Sb, and Ba and (a2) Sb and Ba.

The definition of "unique" for these compositions was based on the experimental fact that they had thus far (at the time of definition) been observed only in GSR (*see also* **Identification and Individualization**).

The problem of classification of various compositions of GSR particles has continued to be a source of concern in casework and of interest for research, in particular the alleged "unique" compositions to GSR. Nowadays, most of the experts prefer to use the term *characteristic* instead of *unique* [27]. These compositions were found among particles produced by fireworks [32] and particles collected from various parts of cars and individuals involved in various automobile-related jobs [33]. It was observed, in particular, that particles of composition (a2) might be found, although in very small amounts, in residues from fireworks and among particles collected from automobiles. Very few particles of composition (a1) were found in the residues of one of the types of the tested fireworks. However, an additional element of magnesium (unusual to GSR) was present in those particles. In another study, Torre *et al.* [34] reported that some types of brake linings contain lead, antimony, and barium and they can represent a source of particles showing GSR-like elemental profiles. Most of these particles could be easily discriminated from GSR by means of the high levels of iron or the presence of "prohibited" elements in the EDX spectrum. However, particles with iron at minor or trace levels and lacking "prohibited" elements were also found but did not have spherical morphology. In view of these findings, the morphological criterion becomes more important for identification of GSR than was suggested by the ASTM committee [35].

It was proposed [36] to include the characteristic particles' composition: Pb, Ba, Ca, Si, and Sn, formed in discharge of Sellier Bellot, Prague (SBP) ammunition in the category of "unique" or characteristic compositions. This proposition was based on the extensive computerized casework experience of the authors that could lead to a statement that such a composition has thus far been observed only in GSR associated with the SBP ammunition.

Stoney pointed out [37] that it is not possible to reach uniqueness through statistics. Nevertheless, with the advent of autosearch systems for GSR in numerous crime laboratories, it may be much easier than before (using a manual search) to experimentally assess the rarity of various GSR compositions in a similar manner as the contribution of automated systems for fingerprint and firearms identification.

In any case, it is very important to compare the GSR compositions found on a suspect with the GSR compositions of the spent cartridge cases (if found) at the scene of crime. Sometimes, the evidential value of such comparisons may be much higher than the extent of "uniqueness" of the GSR particles found on the suspect, for instance, if the GSR particles found on a suspect are consistent with the GSR in the spent cartridge cases of a rare ammunition.

Several studies [38–40] reported the formation of "mixed compositions" GSR particles when firing different ammunitions from the same firearm. These findings are consistent with previous studies [8, 31], which claimed that conventional thorough cleaning of a handgun does not remove all GSR particles.

Until now, there has been no consensus among the forensic science laboratories regarding the number of particles needed to confirm the presence of GSR. There is much more agreement regarding the wording of reports when a laboratory decided that there is a positive result. In such a case, the statement: "The sample is consistent with the suspect having discharged a firearm, having been in the vicinity of a firearm when it was discharged, or having handled an item with GSR on it" resembles the phrasing of most of the laboratories [16, 27]. Regarding the vicinity of a discharged firearm, two recent studies [41, 42] showed that the maximum number of GSR could be found in the right front quadrant at a distance of 2–4 m with respect to the shooting firearm position and shooting direction. It was observed that there is an imminent danger of contamination of people who appear at the scene several minutes following shooting.

The possibility of GSR contamination of the law enforcement personnel or of the law enforcement premises should be taken into account when interpreting results in any particular case [27]. It was recommended that GSR sampling should be done at the scene, where permissible, and as expeditiously as possible and if the subject's hands cannot be sampled before placing the subject in a police vehicle, the subject's hands should be bagged in order to prevent possible contamination. It was suggested that in order to better distinguish between GSR from contact with law enforcement and a civilian shooting event,

a policy such as that set forth in some EU countries (e.g., Germany) should be set, which mandates that all domestic law enforcement use ammunition with taggants detectable by SEM/EDX [27].

Proficiency Tests for GSR Analysis by SEM/EDX. Recently, proficiency tests for GSR particles examination have been designed. For example, the tests prepared by Collaborative Testing Services, Inc. (CTS) were based on GSR suspensions in organic solvents. Owing to the large variation among the samples, it was impossible to assess and compare precisely the proficiency of the various laboratories participating in the test. Lately, a major advance was achieved in the European Network of Forensic Science Institutes (ENFSI) proficiency tests on identification of GSR by SEM/EDX. In these tests developed by Niewoehner *et al.* [43], for the first time "synthetic GSR particles" (pseudoparticles) were prepared by microelectronics technology on the stubs. All the participating laboratories received the same samples, namely, stubs having a flat surface with the same number of pseudo (two-dimensional) particles, in the same locations on the stub, and with the same size distribution. In the GSR2003 test, the pseudoparticles were of Pb, Sb, and Ba composition and having sizes of 2.4, 1.2, 0.8, and 0.5 μm. Although these samples differ from real-life samples of GSR, it seems that at present they are the best to evaluate the proficiency of the procedures used and the SEM/EDX equipment for GSR detection and analysis.

Other Instrumental Methods

Several other instrumental methods have been studied for GSR analysis. Most of them are still in the research phase and cannot be used in casework. Hellmiss *et al.* [44] used SEM equipped with AES analyzer instead of EDX. This method permitted the analysis of elements with low atomic number and elements whose peaks overlap in EDX. Using simultaneous Auger electron spectroscopy and ion sputtering etching, a profile of the in-depth composition variations of GSR was studied.

Micro-X-ray fluorescence (XRF) has been used recently in growing number of crime laboratories for elemental analysis of trace evidence, sometimes instead of SEM/EDX and sometimes in combination with SEM/EDX. Micro-XRF has also been studied for GSR analysis [45, 46]. However, it cannot be used for single particle examinations owing to the problems of spatial resolution. Nonetheless, it has been implemented operationally for GSR deposit analysis around the bullet holes.

Ikeda *et al.* [47]) demonstrated the application of scanning electron microscopy/wavelength dispersive X-ray spectroscopy (SEM/WDX) for detection and analysis of GSR particles from hands, hair, and clothing of a shooter. The authors reported that the method has been implemented in casework.

Time-of-flight secondary ion mass spectrometry (TOF-SIMS) was studied recently for the analysis of GSR [48]. The method may be used as a complimentary technique to SEM/EDX, since it has the capability of depth and trace elements profiling.

Additional techniques that have been studied for GSR analysis are focused ion beam (FIB), Raman microscopy, and laser-induced breakdown spectroscopy (LIBS).

Detection and Analysis of Gunpowder Residues

As was discussed in the previous section, SEM/EDX is a well-established method for detection and analysis of GSR (primer residues) on double-sided adhesive-coated aluminum stubs used to sample a suspect, his clothing, or his belongings, and is used by most of the forensic laboratories in the world. The method however, has the following drawbacks:

1. It is quite slow even when using an automated search system.
2. It has a relatively low "success" rate of detection (about 10%) [19, 26].
3. Not all primer residues are considered "unique" to discharge of firearms.
4. There is a low variability in compositions of primers.

Detection of gunpowder residues on suspects may have an additional evidential value for linking them to a firearm-related offense.

Analytical methods for the detection and identification of propellants or their residues are similar to those employed for the analysis of explosives or explosive residues in postblast samples, e.g., gas chromatography/mass spectrometry (GC/MS), gas chromatography/thermal energy analyzer (GC/TEA), or liquid chromatography/mass spectrometry (LC/MS).

The amounts of propellant residues that may be left on the shooter's hands several hours after shooting may be very small (normally in the nanogram levels) [19].

To the best of the author's knowledge, only two methods have been used operationally for detection and identification of propellant residues on shooters' hands: high performance liquid chromatography with a pendant mercury drop electrode detector (HPLC/PMDE) and GC/TEA due to their high sensitivities i.e., between tens to hundreds of picograms for NG, 2,4 DNT, and 2,6 DNT. The Forensic Science Service in the United Kingdom employs the same method for sampling and analysis of traces of explosives and gunpowder residues on suspects and their clothing [19]. These residues are recovered from the skin by a swabbing kit (nonwoven cotton cloth prewetted with a mixed solvent of 80% isopropanol and 20% water). The swabs are used to sample hands, face, and neck of a suspect. In addition, there is a nail scrapper and a swab threaded through a comb to sample the hair. Clothing and other items are sampled by vacuum filtration on 1-μm membrane filters. Sample extracts are filtered for GSR (primer) particles and cleaned up by solid phase extraction (SPE) in a microcolumn procedure using Chromosorb 104 as adsorbent. The samples are then analyzed by HPLC/PMDE and by GC/TEA.

In general, the question of which and how many techniques are required for confirmation or identification of an explosives trace is complicated, and there is no single answer [49]. Although comprehensive identification by MS would be a method of choice, it is not sensitive enough in many real-life cases.

The swab method to collect primer and gunpowder residues has not been adopted worldwide, in spite of its potentially higher evidential value. It is reasonable to assume that development of an operational method to analyze propellant residues in combination with primer residues on the stubs would encourage at least some of the laboratories to adopt analysis of both types of residues.

Micellar electrokinetic capillary electrophoresis (MECE) with a diode array UV detector was studied to analyze propellant residues on adhesives. Despite the high discrimination power of the technique, it appears that it is not sensitive enough to be implemented in casework [50].

In recent years, ion mobility spectrometry (IMS) technology has gained widespread use in the detection of trace explosive evidence owing to its portability, good sensitivity (comparable to GC/TEA), reasonable selectivity, and high speed of analysis [51]. A method in which the clothing is first sampled by double-sided adhesive for GSR analysis by SEM/EDX, followed by vacuum collection for propellant residue examination by GC/TEA, IMS, and GC/MS was introduced recently into casework in the Israel Police [52]. A novel method was reported for extraction and analysis (IMS and GC/TEA) of gunpowder residues on double-sided adhesive-coated stubs after they were examined for GSR by SEM/EDX [53].

Chemical Analysis for Associating Firearms and Ammunition with Gunshot Entries

The most conclusive way to link a fired projectile to a firearm is by identification of the unique markings of the barrel on the projectile (*see also* **Firearms: Bullet and Cartridge Case Identification**). Quite often in casework, however, the recovered projectiles are so damaged or disintegrated that the quality of marks (if still identifiable) is not sufficient for comparison. In such circumstances, it is impossible to achieve a definite linkage. Some level of association might be obtained by comparison of class characteristics such as the caliber or the width of lands and grooves if this can be assessed from the damaged projectile or its fragments.

There is no method for conclusively associating a firearm or ammunition with a gunshot entry. A class characteristic linkage may be obtained by measuring the diameter of the gunshot entry and by comparing chemical compositions of FDR in the bullet hole perimeter and around it, FDR in the barrel of the weapon, FDR in the spent cartridge case, and the composition of the fired projectile. Also, unfired ammunition, if apprehended with the suspect, may be analyzed for the above purpose.

Analysis of all the components of the FDR (GSR, propellant residues, and metal particles from a projectile) may serve this purpose.

The level of association will depend on the rarity of the examined compositions and on the extent of contamination that each component of FDR leaves in the firearm. The higher the contamination, the lower the extent of association obtained for a particular shot.

GSR Analysis. There is little variability in primer compositions used nowadays for manufacturing various types of ammunitions [3, 5, 7]. Samples may be collected from gunshot entries by double-sided adhesive tapes on stubs (that are used to collect samples from suspects or their clothing), examined by SEM/EDX, and compared with samples collected from the barrel of the suspected firearm and the spent cartridge cases found at the crime scene. Alternatively, the gunshot entry may be cut if it is in clothing or skin mounted on a stub and examined after coating with carbon in the high-vacuum SEM/EDX or without any coating in the variable pressure SEM/EDX. Micro-XRF may also be used for the analysis [45, 46], but with less accuracy because its spatial resolution is not sufficient for single particle analysis.

If spent cartridge cases are not found at the scene and no suspected firearm is apprehended, evidence may be obtained from the shooting tests of unfired ammunition if found in the possession of the suspect.

However, the problem of "mixed primer compositions" should be taken into account, since each round leaves "primer memory" in the barrel [8, 31, 38–40].

Gunpowder Residue Analysis. Several studies were conducted for analysis and comparison of unfired and fired propellants. Andrasko [54] used a GC equipped with a flame ionization detector (FID) and an HPLC equipped with a variable wavelength detector. He reported that the amount of propellant flakes recoverable from clothing (shot from distances shorter than about 50 cm) around the bullet hole was sufficient for the analytical techniques used. A similarity in composition was observed between the propellant particles recovered from the spent cartridges or those recovered from the clothing and the unfired propellant from the same ammunition box. Andrasko showed that the composition of subsequent shootings with the same weapon using ammunition from various manufacturers did not indicate any contamination from the propellant flakes from previous shootings.

MacCrehan *et al.* [55, 56] used MECE for analysis and comparison of fired and unfired propellants. Similar to Andrasko, MacCrehan *et al.* found that compositional analysis of the propellant provides information that can associate residue samples with unfired gunpowder. In shooting experiments using consecutively different ammunition brands (several rounds for each brand), they found only a trace of the previously fired powder in the first shot and none in the next shots.

It is apparent that the behavior of gunpowder residues is different from that of GSR (primer residues). As mentioned previously, GSR of "mixed compositions" are formed when firing different ammunitions from the same firearm. Examined propellant residues might, therefore, be more useful than GSR for associating firearms and ammunition with gunshot entries.

Analysis of Projectiles. Elemental composition is a well-established point of comparison for source discrimination for many types of materials that cannot be distinguished on the basis of their physical properties [57]. One application where elemental analysis has been used is the compositional analysis of bullet lead (CABL). In the past, NAA enabled good discrimination between various lots of bullets [58]. Recently, ICP/MS has been successfully used for this purpose as it is one of the most sensitive methods for elemental analysis [59, 60].

It was shown that considerable variability exists among lead sources despite manufacturers' efforts to control the concentrations of selected elements [57]. On the other hand, in a considerable number of cases, very different elemental signatures may be found even within a single box of bullets, in particular, regarding trace elements [60]. Randich *et al.* [61] have shown that data for lead alloys supplied to two major ammunition manufacturers confirm that multiple indistinguishable shipments of lead alloys from secondary lead refiners to the ammunition manufacturers are made each year and over a period of many years. They have also shown that distinguishable compositions can come from the same melt or "source" of lead alloy. Therefore, they argued "that there is no scientific validity to any conclusions more positive than attributing the possible association as to molten source among bullets from different samples". A similar view was expressed in the recent report on the subject by the US National Academy of Sciences [62].

In examination of very small bullet fragments found in victim's clothing, the risk of contamination from various sources increases, making the trace element analysis less suitable. In such circumstances, a more suitable method for comparison of small quantities of lead is the determination of lead isotope composition, measured as lead isotope ratio [63].

Lead occurs in nature as a mixture of four stable isotopes. Three of these isotopes, having atomic masses of 206, 207, and 208, are the end products of three natural radioactive decay series. The fourth isotope of atomic mass 204 occurs naturally as a nonradiogenic isotope. There is a relatively large variation in the lead isotopic composition in nature. This fact may be useful for discrimination of various sources of lead in ammunitions. Several technologies were used to measure lead isotope ratio in research and casework for discrimination between ammunitions: thermal ionization mass spectrometry (TIMS), sector field-inductively coupled plasma mass spectrometry (SF-ICP/MS), and secondary ionization mass spectrometry (SIMS) [59, 60, 63, 64]. Recently, a new technology of multicollector inductively coupled plasma mass spectrometry (MC-ICP/MS) has been introduced in the field, resulting in better performance, sample throughput, precision, and accuracy with respect to the isotope ratio as compared to the previous technologies mentioned above [65].

It was shown [66] that various mechanical or chemical methods of cleaning do not completely remove lead deposits ("lead memory") from barrels of firearms as was found for GSR. The implication of this effect is that the lead isotope composition of an FDR deposit collected from a barrel at a certain time will be a combination of the lead deposits produced by various firings prior to the sampling. Consequently, this phenomenon lowers the level of association that may be obtained by comparative analysis of lead between a firearm, the fired ammunition, and the gunshot entry in a particular shooting.

Estimation of Time Since Discharge

In a series of papers, Andrasko *et al.* [67–70] presented a novel method for estimation of time since discharge for some types of weapons and spent cartridge cases. Using solid phase microextraction (SPME), samples are taken from the atmosphere inside the barrel of the weapon or the cartridge case and analyzed by GC/TEA, GC/FID or GC/MS, all of which can detect a variety of combustion products. Estimation of time since last discharge is based on the escape rate of the volatile discharge residues from the barrel or the cartridge case as a function of time. The method was studied for shotguns, rifles, pistols, revolvers, and spent cartridges cases. The type of ammunition used, the temperature of storage, cleaning of the barrel after shooting, and the number of shots fired can all influence the results.

References

[1] Wolten, G.M., Nesbitt, R.S., Calloway, A.R., Loper, G.L. & Jones, P.F. (1977). *Final Report on Particle Analysis for Gunshot Residue Detection ATR-77(7915)-3*, The Aerospace Corporation, El Segundo.

[2] Thornton, J.I. (1994). The chemistry of death by gunshot, *Analitica Chimica Acta* **288**, 71–81.

[3] Meng, H.H. & Caddy, B. (1997). Gunshot residue analysis-review, *Journal of Forensic Sciences* **42**, 553–570.

[4] Schwoeble, A.J. & Exline, D. (2000). *Current Methods in Forensic Gunshot Residue Analysis*, CRC Press.

[5] Romolo, F.S. & Margot, P. (2001). Identification of gunshot residue: a critical review, *Forensic Science International* **119**, 195–211.

[6] Zeichner, A. (2003). Recent developments in methods of chemical analysis in investigations of firearm-related events, *Analytical and Bioanalytical Chemistry* **376**, 1178–1191.

[7] Wallace, J.S. (1990). Chemical aspects of firearms ammunition, *AFTE Journal* **22**, 364–389.

[8] Wolten, G.M. & Nesbitt, R.S. (1980). On the mechanism of gunshot residue particle formation, *Journal of Forensic Sciences* **25**, 533–545.

[9] Basu, S. (1982). Formation of gunshot residues, *Journal of Forensic Sciences* **27**, 72–91.

[10] Martini, L. & DeMaria, A. (2002). Evaluation of law enforcement technologies inc. ISID01 "Instant shooter ID kit", *AFTE Journal* **34**, 404–406.

[11] Harrison, H.C. & Gilroy, R. (1959). Firearms discharge residues, *Journal of Forensic Sciences* **4**, 184–199.

[12] Krishnan, S.S. (1982). Detection of gunshot residues: Present status, in *Forensic Science Handbook*, R. Saferstein, ed, Prentice-Hall.

[13] Koons, R.D., Havekost, D.G. & Peters, C.A. (1988). Determination of barium in gunshot residue collection swabs using inductively coupled plasma-atomic emission spectrometry, *Journal of Forensic Sciences* **33**, 35–41.

[14] Koons, R.D. (1998). Analysis of gunshot primer residue collection swabs by inductively coupled plasma-mass spectrometry, *Journal of Forensic Sciences* **43**, 748–754.

[15] Reis, E.L.T., Souza Sarkis, J.E., Neto, O.N., Rodrigues, C., Kakazu, M.H. & Viebig, S. (2003). A new method for collection and identification of gunshot residues from the hands of shooters, *Journal of Forensic Sciences* **48**, 1269–1274.

[16] Singer, R.L., Davis, D. & Houck, M.M. (1996). A survey of gunshot residue analysis methods, *Journal of Forensic Sciences* **41**, 195–198.

[17] De Gaetano, D., Siegel, J.A. & Klomparens, K.L. (1992). A comparison of three techniques developed for

sampling and analysis of gunshot residues by scanning electron microscopy/energy dispersive x-ray analysis (SEM-EDX), *Journal of Forensic Sciences* **37**, 281–300.

[18] Zeichner, A., Foner, H.A., Dvorachek, M., Bergman, P. & Levin, N. (1989). Concentration techniques for the detection of gunshot residues by scanning electron microscopy/energy dispersive x-ray analysis (SEM/EDX), *Journal of Forensic Sciences* **34**, 312–320.

[19] King, R.M. (1993). The work of explosives and gunshot residues unit of the forensic science service (UK), in *Advances in the analysis and detection of explosives*, J. Yinon, ed, Proceedings of the 4th international symposium on the analysis and detection of explosives, 1992, Jerusalem Israel. Kluver Academic Publishers, 91–100.

[20] Zeichner, A. & Levin, N. (1993). Collection efficiency of gunshot residue (GSR) particles from hair and hands using double-side adhesive tape, *Journal of Forensic Sciences* **38**, 571–584.

[21] (1980). *SEM/MPA Firearms Discharge Residues*, Metropolitan Police Forensic Science Laboratory, London, Vols. 1 and 2.

[22] Zeichner, A. (2001). Is there a real danger of concealing gunshot residue (GSR) particles by skin debris using the tape-lift method for sampling GSR from hands? *Journal of Forensic Sciences* **46**, 1447–1455.

[23] Andrasko, J. & Petterson, S. (1991). A simple method for collection of gunshot residues from clothing, *Journal of Forensic Science Society* **31**, 321–330.

[24] Kilty, J.W. (1975). Activity after shooting and its effect on the retention of primer residues, *Journal of Forensic Sciences* **20**, 219–230.

[25] Jalanti, T., Henchoz, P., Galluser, A. & Bonfanti, M.S. (1999). The Persistence of gunshot residue on shooters' hands, *Science & Justice* **39**, 48–52.

[26] Zeichner, A. & Levin, N. (1995). Casework experience of GSR detection in Israel, on samples from hands, hair and clothing, using an autosearch SEM/EDX system, *Journal of Forensic Sciences* **40**, 1082–1085.

[27] Wright, D.M., Trimpe, M.A. (2006). Summary of the FBI Laboratory's Gunshot Residue Symposium, May 31–June 3, 2005, *Forensic Science Communications* **8**.

[28] Tassa, M., Leist, Y. & Steinberg, M. (1982). Characterization of gunshot residues by x-ray diffraction, *Journal of Forensic Sciences* **27**, 677–683.

[29] Wolten, G.M., Nesbitt, R.S., Calloway, A.R., Loper, G.L. & Jones, P.F. (1979). Particle analysis for the detection of gunshot residue. I: scanning electron microscopy/energy dispersive x-ray characterization of hand deposits from firing, *Journal of Forensic Sciences* **24**, 409–422.

[30] Wolten, G.M., Nesbitt, R.S., Calloway, A.R. & Loper, G.L. (1979). Particle analysis for the detection of gunshot residue. II: occupational and environmental particles, *Journal of Forensic Sciences* **24**, 423–430.

[31] Wallace, J.S. & McQuilLan, J. (1984). Discharge residues from cartridge-operated industrial tools, *Journal of Forensic Science Society* **24**, 495–508.

[32] Mosher, P.V., McVicar, M.J., Randall, E.D. & Sild, E.H. (1998). Gunshot similar particles produced by fireworks, *Canadian Society of Forensic Science Journal* **31**, 157–168.

[33] Garofano, L., Capra, M., Ferrari, F., Bizzaro, G.P., DiTullio, D., Dell'Olio, M. & Ghitti, A. (1999). Gunshot residue – further studies on particles of environmental and occupational origin, *Forensic Science International* **103**, 1–21.

[34] Torre, C., Mattutino, G., Vasino, V. & Robino, C. (2002). Brake linings: a source of non-GSR particles containing lead, barium and antimony, *Journal of Forensic Sciences* **47**, 494–504.

[35] ASTM Committee E-30 on Forensic Sciences (2002). *ASTM E 1588–95, (Reapproved 2001) standard Guide for Gunshot Residue Analysis by Scanning Electron Microscopy/Energy-Dispersive Spectroscopy*, Vol. 14.02, Annual Book of ASTM Standards.

[36] Zeichner, A. & Levin, N. (1997). More on the uniqueness of gunshot residue (GSR) particles, *Journal of Forensic Sciences* **42**, 1027–1028.

[37] Stoney, D.A. (1991). What made us ever think we could individualize using statistics, *Journal of Forensic Science Society* **31**, 197–199.

[38] Zeichner, A., Levin, N. & Springer, E. (1991). GSR particles formed by using different types of ammunition in the same firearm, *Journal of Forensic Sciences* **36**, 1020–1026.

[39] Gunaratnam, L. & Himberg, K. (1994). The identification of gunshot residues particles from lead-free Sintox ammunition, *Journal of Forensic Sciences* **39**, 532–536.

[40] Levin, N., Tsach, T., Bergman, P. & Springer, E. (2000). *A Survey of Titanium and Zinc Particles in Samples Collected from Suspects*, Proceedings of the 2nd EAFS meeting, Cracow, Poland.

[41] Fojtasek, L., Vacinova, J., Kolar, P. & Kotrly, M. (2003). Distribution of GSR particles in the surroundings of shooting pistol, *Forensic Science International* **132**, 99–105.

[42] Fojtasek, L. & Kmjec, T. (2005). Time periods of GSR particles deposition after discharge-final results, *Forensic Science International* **153**, 132–135.

[43] Niewoehner, L., Andrasko, J., Biegstraaten, J., Gunaratnam, L., Steffen, S. & Uhlig, S. (2005). Maintenance of the ENFSI proficiency test program on identification of GSR by SEM/EDX (GSR2003), *Journal of Forensic Sciences* **50**, 877–882.

[44] Hellmiss, G., Lichtenberg, W. & Weiss, M. (1987). Investigation of gunshot residues by means of Auger Electron Spectroscopy, *Journal of Forensic Sciences* **32**, 747–760.

[45] Flynn, J., Stoilovic, M., Lennard, C., Prior, I. & Kobus, H. (1998). Evaluation of x-ray microfluorescence spectrometry for the elemental analysis of firearm discharge residues, *Forensic Science International* **97**, 21–36.

[46] Charpentier, B. & Desrochers, C. (2000). Analysis of primer residue from lead free ammunition by x-ray microfluorescence, *Journal of Forensic Sciences* **45**, 447–452.
[47] Kage, S., Kudo, K., Kaizoji, A., Ryumoto, J., Ikeda, H. & Ikeda, N. (2001). A simple method for detection of gunshot residue particles from hands, hair, face and clothing using scanning electron microscopy/wavelength dispersive x-ray (SEM/WDX), *Journal of Forensic Sciences* **46**, 830–834.
[48] Collins, P., Coumbaros, J., Horsley, G., Lynch, B., Kirkbride, K.P., Skinner, W. & Klass, G. (2003). Glass-containing gunshot residue particles: a new type of highly characteristic particle? *Journal of Forensic Sciences* **48**, 538–553.
[49] Zitrin, S. (1999). Recommendations for combining techniques to confirm identifications, in *Proceedings of the workshop on explosives trace analysis methods*, S.A. Phillips & R. Hiley, eds, DERA, UK; *Science and Justice* **34**, 261–268.
[50] Northrop, D.M. (2001). Gunshot residue analysis by micellar electrokinetic capillary electrophoresis: assessment for application in casework. Parts I and II, *Journal of Forensic Sciences* **46**, 549–572.
[51] Fetterolf, D.D. & Clark, T.D. (1993). Detection of trace explosive evidence by ion mobility spectrometry, *Journal of Forensic Sciences.* **38**, 28–39.
[52] Zeichner, A., Eldar, B., Glattstein, B., Koffman, A., Tamiri, T. & Muller, D. (2003). Vacuum collection of gunpowder residues from clothing worn by shooting suspects, and their analysis by GC/TEA, IMS and GC/MS, *Journal of Forensic Sciences* **48**, 961–972.
[53] Zeichner, A. & Eldar, B. (2004). A novel method for extraction and analysis of gunpowder residues on double-side adhesive coated stubs, *Journal of Forensic Sciences* **49**, 1194–1206.
[54] Andrasko, J. (1992). Characterization of smokeless powder flakes from fired cartridge cases and from discharge patterns on clothing, *Journal of Forensic Sciences* **37**, 1030–1047.
[55] MacCrehan, W.A., Patierno, E.R., Duewer, D.L. & Reardon, M.R. (2001). Investigating the effect of changing ammunition on the composition of organic additives in gunshot residue (OGSR), *Journal of Forensic Sciences* **46**, 57–62.
[56] MacCrehan, W.A., Patierno, E.R. & Duewer, D.L. (2002). Associating gunpowder and residues from commercial ammunition using compositional analysis, *Journal of Forensic Sciences* **47**, 260–266.
[57] Koons, R.D. & Grant, D.M. (2002). Compositional variation in bullet lead manufacture, *Journal of Forensic Sciences* **47**, 950–958.
[58] Lukens, H.R., Schlesinger, H.L., Guinn, V.P. & Hackelman, R.P. (1970). *Forensic Neutron Activation Analysis Of Bullet-Lead Specimens*, USAEC Report GA-10141, Gulf General Atomic Incorporated, San Diego.
[59] Dufosse, T. & Touron, P. (1998). Comparison of bullet alloys by chemical analysis: use of ICP-MS method, *Forensic Science International* **91**, 197–206.
[60] Keto, R.O. (1999). Analysis and comparison of bullet leads by inductively-coupled plasma mass spectrometry, *Journal of Forensic Sciences* **44**, 1020–1026.
[61] Randich, E., Duerfeldt, W., McLendon, W. & Tobin, W. (2002). A metallurgical review of the interpretation of bullet lead compositional analysis, *Forensic Science International* **127**, 174–191.
[62] The U.S. National Academy of Sciences Report (2004). Forensic Analysis: Weighing Bullet Lead Evidence.
[63] Andrasko, J., Kopp, I., Abrink, A. & Skiold, T. (1993). Measurement of lead isotope ratio in lead smears and bullet fragments and its application to firearm investigations, *Journal of Forensic Sciences* **38**, 1161–1117.
[64] Stupian, G.W., Ives, N.A., Marquez, N. & Morgan, B.A. (2001). The application of lead isotope analysis to bullet individualization in two homicides, *Journal of Forensic Sciences* **46**, 1342–1351.
[65] Plazne, I., Ehrlich, S. & Halicz, L. (2001). Isotope-ratio measurements of lead in NIST standard reference materials by multiple-collector inductively coupled plasma mass spectrometry, *Fresenius Journal of Analytical Chemistry* **370**, 624–628.
[66] Zeichner, A., Ehrlich, S., Shoshani, E. & Halich, L. (2006). Application of lead isotope analysis in shooting incident investigations, *Forensic Science International* **158**, 52–64.
[67] Andrasko, J., Norberg, T. & Stahling, S. (1998). Time since discharge of shotguns, *Journal of Forensic Sciences* **43**, 1005–1015.
[68] Andrasko, J. & Stahling, S. (1999). Time since discharge of spent cartridges, *Journal of Forensic Sciences* **44**, 487–495.
[69] Andrasko, J. & Stahling, S. (2000). Time since discharge of rifles, *Journal of Forensic Sciences* **45**, 1250–1255.
[70] Andrasko, J. & Stahling, S. (2003). Time since discharge of pistols and revolvers, *Journal of Forensic Sciences* **48**, 307–311.

ARIE ZEICHNER

Firearm Examination: Ballistics

Internal Ballistics

The Association of Firearm and Tool Mark Examiners (AFTE) Glossary defines internal ballistics as "the study of the motion of the projectile(s) within the firearm from the moment of ignition until it leaves the barrel" [1].

When the firing pin strikes the primer[a] of a cartridge,[b] the priming compound inside the primer is crushed between the anvil (which may form part of the cartridge case or part of the primer itself) and the firing pin. This causes the priming compound to detonate, producing a flash which in turn ignites the propellant. The propellant inside the cartridge then begins burning. It should be noted that the propellant burns very rapidly but does not explode. As the propellant burns an enormous amount of gas is formed, which in turn causes pressure to be built up within the cartridge. The pressure continues to increase inside the cartridge. Eventually the projectile begins to move out of the mouth of the cartridge case and enters the lead; at this point the projectile makes contact with the rifling,[c] which retards the movement of the projectile due to the sudden resistance increase between the projectile and the rifling. The pressure exerted on the base of the projectile overcomes this resistance and the projectile moves forward, rotating with the rifling. The pressure caused by the build up of gases will accelerate the projectile's movement down the barrel until the projectile leaves the muzzle [2–4].

The following scenario serves as an example in which the study of internal ballistics may play a role in the examination of an internal ballistic–related case. Mr X, a well-known hunter and reloader, is out one day testing his latest "load" for his next hunting trip. The first cartridge that is fired causes the firearm to "blowup", severely wounding Mr X. The firearm together with the reloaded ammunition is submitted to the forensic firearms laboratory. In the first instance, the forensic firearm investigator would examine the fired cartridge case and any cartridges for signs of pressure and overloading. Additionally, through the aid of different types of testing equipment, such as pressure barrels fitted with piezoelectric transducers,[d] the amount of pressure generated inside the submitted reloaded cartridges can be measured and recorded. This will aid the forensic firearm investigators in determining whether or not the pressure generated by the reloaded cartridges meets the set specifications, such as those developed by Sporting Arms and Ammunition Manufacturers' Institute (SAAMI) or Commission Internationale Permanente (CIP), with respect to the amount of pressure that can be safely generated by a specific cartridge. If it is found that the generated pressure exceeds the recommended maximum pressure, the reason for the firearm blowing-up could possibly then be attributed to possible incorrect reloading procedures having been followed [3, 5, 6].

Specialized equipment used to measure chamber pressure, such as pressure barrels fitted with piezoelectric transducers, is not necessarily found in all the forensic ballistic laboratories. In some cases, it may be necessary to seek the assistance of a commercial ammunition manufacturing facility, an ammunition/firearm "proof"[e] facility, or another government agency to assist in a specific instance where chamber pressures need to be established.

The internal ballistic properties of a firearm and cartridge will determine a number of factors, such as projectile speed, rate of spin, stability in flight, and recoil.

Intermediate Ballistics

Intermediate ballistics is defined by Moss *et al.* " . . . as the study of the transition from internal to external ballistics, which occurs in the vicinity of the gun muzzle" [5, p. 53].

Intermediate ballistics is a field of study, which is positioned between the fields of internal and external ballistics. Following the firing process, there are also a number of other components, besides the projectile, which exit the muzzle. These are known as *gunshot residues* (GSRs) and consist of primer residues, gases produced by the burning propellant, soot, partially burnt and unburnt propellant particles, and particles arising, from the projectile. These components affect the intermediate ballistics.

The intermediate ballistics of a firearm influences a number of factors associated within the field of forensic ballistics, such as the accuracy of the projectile, level of noise created during the firing process, muzzle flash,[f] and the dispersion of primer residues, soot and burnt and partially burnt propellant particles. Thompson [6, p. 159] is of the opinion that the region of intermediate ballistics extends down range from the muzzle of the firearm to the point where the projectile finally passes out of the influence of the propellant gases. He states that the region of intermediate ballistics is typically " . . . half a dozen or so calibres long . . .". However, depending on the ammunition and firearm type, this may extend to a greater distance from the muzzle [2, 4, 6, 7].

Besides the theoretical aspects surrounding intermediate ballistics, the study of intermediate ballistics

from a forensic firearms perspective is more concerned with how it relates to the field of range determination. This is discussed in section "Range Determination".

External Ballistics

External ballistics can be defined as a study of the projectile's trajectory between the muzzle of the firearm and the target, and the influence that various factors have on the trajectory of the fired projectile. The question of "where did that shot come from?" is typically answered through the study of external ballistics. As has already been alluded to, there are a number of factors that influence the trajectory of a projectile. These include the projectile's mass, caliber, shape, sectional density, spin, etc. There are also external factors that influence the trajectory of a projectile, and these include atmospheric density, temperature, pressure, viscosity, and gravity [1, 6].

The forensic firearms investigator will mainly use external ballistics to determine the position from where a projectile could have been fired through trajectory[g] determination. This is often important in reconstructing a shooting incident.

The remaining speed, energy, time of flight, and drop for a projectile can be determined for specific distances based on tabulated information. A trajectory of a projectile can be calculated and determined through the use of various mathematical formulas and calculations, which is a very complex and time-consuming process. Trajectories of projectiles are now solved through the use of external ballistics computerized programs. A number of these programs are available on the commercial market. The more accurate programs are those that will allow the input of the angle of incidence (entry into the target) and projectile details such as caliber, ballistic coefficient, muzzle velocity, etc., and will be able to essentially work backward to calculate the possible point of discharge of the firearm. One should always keep in mind that although it might be possible to determine an area where a firearm was discharged, the *exact* position of the shooter is often difficult to pinpoint, particularly over long ranges [6].

Terminal Ballistics

Terminal ballistics can be divided into two sections, namely, terminal ballistics with respect to the study of nonbiological matter and terminal ballistics with respect to the study of biological matter (which is commonly referred to as *wound ballistics*).

Terminal Ballistics (Nonbiological Matter)

Terminal ballistics is defined in the AFTE Glossary as "the scientific study of the impact or penetration of projectiles in any substance heavier than air and their effects on the object struck" [1].

Terminal ballistics looks at the effect that the projectile has on the object (target) and the effect that the object (target) has on the projectile. When a forensic firearm investigator is called upon to reconstruct a shooting incident, he or she is faced with the terminal effects that the fired projectile had on its target. With this information, the forensic firearm investigator can assist in the reconstruction of the scene of a shooting incident. It is important therefore to be able to interpret the terminal effects of the fired projectile. A fired projectile will invariably do one of the three things, or a combination thereof, when striking an object. It will penetrate an object, or perforate an object, or ricochet from an object.

Terminal Ballistics (Biological Matter)

Wound ballistics is defined in the AFTE Glossary as "the scientific study of the effects of penetrating projectiles on the human or animal body. It is a branch of terminal ballistics" [1, 2].

No study of wound ballistics would be complete without a thorough understanding of the human anatomy. Besides the obvious importance of being able to better understand the interaction between a projectile and various parts of the human body that may be struck by a projectile, an understanding of human anatomy will place the forensic firearms investigator in a better position to interpret the results of a medicolegal examination report.

An important aspect of wound ballistics involves the mechanism by which a wound is created. In some instances, these "aspects" are visible and in other "aspects" they may not be so visible [4].

The identification of wounds plays an important role from an investigative perspective. This includes the correct identification of a wound, be it an entrance or exit wound, as well as correctly identifying the general characteristics of the wound that may indicate the range at which the shot was fired [2]. In some

jurisdictions, notably in the United States, courts have required that, beyond the investigator's visually perceived description, the study of wounds requires the testimony of a forensic pathologist or medical examiner's assistance with training in gunshot wounds.

One of the most evidential aspects of terminal ballistics, from a forensic perspective, is the ability of the forensic firearms investigator to determine the range at which a shot was fired. This can play a significant role in either supporting or refuting an allegation of the distance between the muzzle of the firearm and the target (victim) at the time the firearm was discharged.

Range Determination (General Aspects)

Range (distance) determination is defined by the AFTE Glossary as "The process of determining the distance from the firearm, usually the muzzle, to the target based upon patterns of gunpowder or gunshot residues deposited upon that target. Where multiple projectiles, such as shot, have been fired the spread of those projectiles is also indicative of the distance." The aforesaid approaches consider that standard ammunition is used for both rifled and smooth-bored barrels (firearms) [1].

The value of accurately determining the distance between the muzzle of the firearm and the target lies in either supporting or refuting an allegation relating to the position of the shooter and the target (normally the victim). For this very reason, it is essential that the forensic firearms investigator correctly identifies and evaluates the evidence that he or she has before him or her in determining the correct range.

A factor that must always be kept in mind with cases involving the determination of the range between the muzzle of the firearm and the target is that an exact distance can never (and should never) be given. The result obtained from conducting a range determination examination will provide the forensic firearms investigator with a minimum and maximum range between the muzzle of the firearm and the target or a median value with uncertainties.

Range determination can be divided into two sections, namely the range at which a shot was fired with a cartridge loaded with a single projectile, such as those used in rifles and handguns (where the examination of GSR is used to determine the range) and the range at which a shot was fired with a cartridge loaded with multiple projectiles, such as pellets loaded in shotgun cartridges (where the examination of the spread of shot is used to determine the range) [2, 4, 7].

Generally, rifle-barreled firearms are used to shoot single projectiles and smooth-bore firearms (shotguns) are used to fire cartridges loaded with multiple projectiles (pellets). However, shotguns are also commonly used to fire single projectiles such as rifled slugs.

Range Determination (Single Projectile Cartridges). With respect to range determination involving a rifled barrel, the forensic firearms investigator relies on the pattern of GSRs found on the target to determine the range between the muzzle of the firearm and the target.

When a cartridge with a single projectile is fired, not only does the projectile exit the barrel but so does soot, gaseous primer residues, partially burnt/unburnt propellant and projectile particles (i.e., GSRs). These gaseous residues and particles disperse from the muzzle of the firearm expanding in diameter with distance from the muzzle of the firearm. As these particles are very light they do not travel far from the muzzle and in many cases are unlikely to strike a target further than 1 m away. The further these particles move from the muzzle of the firearm the larger diameter and less dense this circular pattern becomes [2, 4, 7].

On the basis of the examination of the GSR pattern, the forensic firearms investigator can conduct test firings using the same firearm and ammunition onto similar target surfaces to determine the approximate distance between the muzzle and the target. It is common practice to enhance the lead and propellant present in a GSR pattern using colorimetric tests [7].

Range Determination (Multiple Projectile Cartridges). The fundamental principles involved in determining the distance at which a shot was fired with using single or multiple projectile cartridges are closely related. With cartridges loaded with multiple projectiles, the method used to determine the range between the muzzle of the firearm and the target primarily involves the examination of the spread of the pellets. Therefore, it is possible to estimate the firing range over a relatively long distance. As with cartridges loaded with single projectiles, the presence

of GSRs on a target does have significance when it comes to a shot at a close range [7].

Barrel length, whether a cartridge is loaded with a single or multiple projectiles, is relevant to the amount of GSR emitted from the muzzle.

The main factor that is considered when determining the range between the muzzle of the firearm and the target when using a cartridge loaded with multiple projectiles is the spread of the pellets. This is due to the fact that when a cartridge loaded with pellets is fired using a smooth-bored firearm, the pellets in a pattern will spread in a reasonably predicable manner. As the distance between the muzzle of the firearm and the target increases, the diameter of the pellet pattern increases and the density of the pellet pattern decreases. The pattern can be reproduced by conducting tests with the same firearm and ammunition so as to determine an approximate range between the muzzle and the target [2, 7].

End Notes

[a.] A primer contains a stable but shock-sensitive explosive mixture, which when struck (typically by the firing pin) explodes and ignites the propellant held within a cartridge.

[b.] A cartridge is a single unit of ammunition consisting of a primer, cartridge case, propellant, and projectile(s).

[c.] A combination of helical grooves, which are either cut or formed inside the barrel of a firearm. The purpose of rifling is to impart spin on a projectile, which aids in the stabilization of the projectile in flight.

[d.] A piezo-electric transducer is a device that can be used to measure pressure through the use of a quartz or tourmaline crystal.

[e.] A "proof facility"' is a facility which tests the safety of a firearm or ammunition by measuring the pressure generated in the firearm (using overloaded cartridges) or ammunition during test firing.

[f.] Muzzle flash occurs when the hot gasses and burning propellant particles exit the muzzle of the firearm along with the fired bullet and mix with the oxygen in the surrounding air.

[g.] A trajectory of a fired bullet can be described as the curved path, which the fired bullet follows between the muzzle of the firearm and the target.

References

[1] Association of Firearm and Tool Mark Examiners Standardisation Committee (1994). *Glossary of the Association of Firearm and Tool Mark Examiners (Appendix F, Firearm Ignition Systems)*, 3rd Edition, Available Business Printing, Chicago.

[2] Di Maio, V.J.M. (1985). *Gunshot Wounds*, Elsevier, New York.

[3] National Institute of Forensic Science (2007). *National Training Curriculum for Forensic Firearm Examiners*, National Institute of Forensic Science, Australia.

[4] Sellier, K. (1994). *Wound Ballistics, and the Scientific Background*, Elsevier, Netherlands.

[5] Moss, G.M., Farrar, C.L. & Leeming, D.W. (1995). *Military Ballistics*, Brassey's, United Kingdom.

[6] Thompson, I. (2001). *Ballistic Allsorts*, Hyde Park Press Pty, SA.

[7] Sellier, K. (1991). *Forensic Science Progress: Volume 6, Shot Range Determination*, Springer-Verlag, Berlin.

Further Reading

Basmajain, J.V. (1971). *Grant's Method of Anatomy*, Williams and Wilkins, Baltimore.

Fatteh, A. (1976). *Medicolegal Investigation of Gunshot Wounds*, J. B. Lippincott Company, USA.

Hatcher, J.S. (1962). *Hatcher's Notebook*, Stackpole Books, Harrisburg.

Lowry, E.D. (1968). *Interior Ballistics: How a Gun Converts Chemical Energy into Projectile Motion*, Doubleday & Company, Garden City.

Rinker, R. (2002). *Understanding Ballistics*, Morris House Publishing.

MICHAEL JACKSON

Firearms: Bullet and Cartridge Case Identification

Overview

The identification of fired cartridge cases and fired bullets to the firearm responsible for their discharge is correctly termed *firearms identification*. This is a subdiscipline of the field of forensic firearms

investigation and is wholly laboratory based due to the need for specialized equipment including a comparison microscope, camera(s), sensitive scales, and other measuring apparatus.

Firearms identification is by far the most specialized form of toolmark examination. In firearms identification, the firearm is the "tool", while the ammunition components receive toolmarks as they come into contact with various components of the firearm during the loading, chambering, firing, extraction, and ejection cycles. For example, the passage of the bullet through a barrel will cause a striated toolmark along the long axis of the bullet where it is in intimate contact with the bore, while the firing pin will cause an impressed toolmark where it strikes the primer of the cartridge upon discharge.

Whereas traditional toolmark examinations can sometimes be very difficult to successfully carry out due to the enormous amount of variation possible during their formation, firearms identification is simplified in that the cartridge will be marked more or less in the same predictable manner every time the firearm is discharged. Despite this, however, firearms identification is often challenging. Competent experts require a broad knowledge of firearms and their operation, ammunition types, the principles of toolmark identification, manufacturing techniques, materials science, and various optical instruments to make correct determinations. There are several main texts dealing with the technicalities of firearms identification [1–4].

Toolmarks

In order to understand firearms identification, a brief discussion on toolmarks and a description of some important definitions is necessary. A toolmark is formed when a harder object comes into contact with a softer object with sufficient force to cause deformation to the softer item. If the two items are of similar hardness, and if sufficient force is applied, both items will be deformed and damaged to some degree. There are two main types of toolmarks: striations and impressions, although a toolmark can consist of a combination of these two marks. The quality of any toolmark produced is dependent on variables such as the amount of force used in its formation, the relative angles between the two objects when they make contact, and the relative hardness of the two objects. Also relevant is the amount of use the tool has had since making the original mark as that part of the tool responsible for making any mark may change with use, abuse, corrosion, or by other means. If a tool surface changes over time for whatever reason, test marks made with the actual tool may not be able to be identified to a toolmark in question owing to the changes to its microscopic features.

Striations

Striations are formed when force is applied by a harder object along the surface of a softer object, leaving a series of parallel scratches, referred to as *striae*. The depth and spatial relationship of the striae to each other are made by the features existing on that part of the tool in contact with the softer item. The features responsible for making the marks may be quite gross and visible to the naked eye, but usually a microscope and correct lighting is needed to visualize them properly.

Impressions

Impressions are formed when a harder object contacts a softer object without either surface sliding against each other. The features existing on the surface of the harder object are then recorded on the softer object, creating a negative pattern of the detail where contact was made. Again, the features responsible for making the mark may be coarse and visible to the naked eye, but a microscope with correct lighting is normally needed to visualize them properly.

Class Characteristics

Class characteristics refer to the design features of any particular tool, and are determined prior to manufacture. They are usually discernable to the naked eye. An example of class characteristics in a firearm include the bore diameter, the number, width and depth of the lands and grooves cut or formed into the bore, and the direction and rate of twist of the rifling. Class characteristics are used to determine the type of firearm used in a shooting incident, or at least narrow the field of search considerably. Manufacturers of firearms and tools are free to employ whatever class characteristics they wish when making their products, hence many firearms from

different makers have similar class characteristics, resulting in a large group of firearms that leave similar marks. Conversely, manufacturers sometimes choose such unusual class characteristics that it can narrow the search considerably to one or only several possibilities, to the exclusion of all others.

Some simple examples of using class characteristics in a practical sense follow:

- A .30 caliber bullet can be excluded as having been discharged from a firearm with a .22 caliber bore.
- A .32 caliber bullet with rifling engravings of six lands and grooves with a left twist can be excluded as having been fired in a second .32 caliber firearm with six lands and grooves with a left twist, if the widths of the lands and grooves differ.
- A .22 caliber fired cartridge case with a round hemispherical firing pin impression could only have been discharged from other .22 caliber firearms with firing pins that leave the same size and shape firing pin impressions, assuming there are no other discernable class characteristics present on the cartridge case to narrow the field of search further. All .22 caliber firearms that do not leave the same shape and size of firing pin impression can be excluded as having discharged that cartridge.

Subclass Characteristics

Subclass characteristics are a smaller group within any set of class characteristics and are created during manufacture. They are reproducible features resulting from manufacturing processes that can be mistaken for individual characteristics. The highest potential for the production of subclass characteristics occurs in components that have been consecutively manufactured – the toolmarks left behind by the various machining processes using the same tool may not change appreciably over time on successively made items. Therefore, it is important to understand the manufacturing processes that are applicable to various firearm components and establish what changes can be distinguished over time in the microscopic imperfections of consecutively manufactured components. If subclass characteristics are present on any firearm part, these may or may not be confused with any individual marks left on ammunition components, depending on the orientation of the subclass features to the surface being marked.

Individual Characteristics

Individual characteristics are irregularities or features that are randomly acquired or generated and are crucial in the identification process. They are normally microscopic in nature and require correct illumination under a microscope to be seen. They are formed during manufacture and with use of the item, although subsequent abuse, corrosion, and damage can create further individual characteristics over time. Individual characteristics allow the expert to definitively identify, or exclude, a firearm as having discharged a bullet or cartridge case.

The Comparison Microscope

The comparison microscope is the most important instrument used in the forensic firearms laboratory and its basic features have not changed much since its introduction to the field of forensic firearms in the mid-1920s. The comparison microscope is essentially two microscopes joined together with two sets of stages movable in the x, y, and z axis, various adjustable light sources, and two sets of objective lens. By virtue of an optical arrangement in the bridge, both fields of view are connected so that two separate items can be simultaneously compared under the same magnification. Features vary slightly between makes and models; however, all have the ability to move the stages either independently or in concert. An important and necessary feature is a prism within the bridge, which creates a central vertical dividing line between both fields of view. When the central divider is moved from left to right, it enables the relative positions of various class and individual features to be easily viewed and compared.

The light sources can include incandescent bulbs and optic fiber types. However, the type of lighting utilized is not as crucial as providing even illumination from the same relative angle to the items being compared. As toolmarks are three dimensional, both striated and impressed marks are visualized by shining light obliquely across their surface topography. This causes shadows and reflections, which are seen as light and dark areas within the field of view. If the intensity of the light sources differs or the angle of

the light source differs substantially between the two sides, the resultant images formed will be dissimilar, making the comparison difficult, if not impossible.

The range of magnification used on the comparison microscope for most work varies between approximately 8× and 40×, although some circumstances may require lower or higher magnifications from 2× to 100×. It would be considered very unusual to observe toolmarks at magnifications over 100× using an optical microscope. Some European agencies have utilized scanning electron microscopes (SEMs) fitted with dual stages for comparisons of toolmarks; however, this is not the norm and this equipment has not been adopted by most forensic firearms laboratories. The advantage of using an SEM for comparisons is that the limited depth of field experienced at high magnification when using an optical system is negated.

Modern comparison microscopes can be fitted with cameras for both still photographs and video recording. Digital cameras are becoming more commonplace owing to their ease of use and the ability to print immediately, allowing instant assessment of the quality of the image. Measurements can also be taken using the microscope by a variety of means including graduated eyepieces and micrometers, and digitally.

The stages are usually multiadjustable to allow bullets and cartridge cases to be manipulated in three dimensions. This allows for every surface of the items to be examined minutely and optimizes the correct positioning to best visualize important or identifying features.

Bullet Identification

The forensic firearms expert faces three main types of problems when carrying out bullet comparisons. The same three problems are also applicable to fired cartridge cases.

1. Given a fired bullet, determine what type of firearm could have discharged it.
2. Given two or more bullets, determine if one or more firearms were used and how many bullets were discharged from each. There is still a requirement to also determine, if possible, what type of firearm was used.
3. Given one or more bullets and one or more firearms, determine which firearms discharged which bullets.

Finding answers to the problems above is achieved by using the comparison microscope and associated equipment to examine the class and individual characteristics of the fired bullets, imparted by the firearm responsible.

When bullets are fired through firearms with rifled barrels such as pistols, revolvers, and rifles, the circumference of the bullet in contact with the bore is engraved with the rifling existing in that barrel. The rifling is a series of helical grooves cut or formed into the inner surface of the barrel to impart a spinning motion to the projectile around its long axis. This spinning motion provides gyroscopic stability to the bullet and therefore, accuracy after it leaves the barrel. The number, width, and depth of the grooves, the direction of twist (clockwise or counterclockwise), and the rate of twist are determined by the manufacturer. These class characteristics vary from one firearm to another, but some forms of rifling are more popular and common than others.

The grooves in the rifling are unsurprisingly referred to as *grooves*, while the proud areas between the grooves are referred to as *lands*. The rifling engravings left on a fired bullet are therefore referred to as *groove impressions* and *land impressions* and they will slant to either the right or left side from the long axis of the bullet depending on whether the rifling rotated clockwise or counterclockwise, respectively.

In order to determine which type of firearm may have discharged a fired bullet (assuming no firearm was recovered), the rifling characteristics engraved by the responsible firearm are measured on the bullet. A comprehensive database of firearm manufacturer's specifications can then be referred to, which will eliminate most firearms, leaving only those with the same class characteristics. The Federal Bureau of Investigation compiles and regularly updates the most comprehensive reference data available for this task, referred to as the *general rifling characteristics* (*GRC*) file. This is made available free of charge to government forensic firearms laboratories worldwide upon written request.

It is evident that the absence of a suspect firearm does not preclude microscopic comparisons if two or more bullets are submitted to the laboratory. But if one or more suspect firearms are also submitted, test-fired samples have to be obtained to allow direct comparison. If test firing is to be carried out, all class characteristics of the suspect firearm(s) must first be

undifferentiable from the class characteristics seen on the crime bullets. Firearms with different class characteristics from those seen on the bullet can be eliminated as being responsible; there would be no need to conduct a microscopic comparison as they could not have been discharged from a weapon with incompatible class features.

Bullets recovered from crime scenes are rarely in good condition as they have usually impacted with various objects, including people, causing varying degrees of deformation. Therefore, the key consideration when test firing suspect firearms is to recover the bullets in a pristine condition, unmarked except as a result of its passage through the bore. Ideally, the bearing surface – the portion of the bullet in contact with the bore – should not be distorted or shortened from excessive mushrooming. This ensures that the entire striated surface of the bullet is available for examination, maximizing any information present.

Testing is usually carried out into a large water tank, although a number of other mediums, such as cotton waste and sawdust, have been used to trap fired bullets. If bullets are of hollow point design or of high velocity (greater than approximately $610\,\mathrm{m\,sec^{-1}}$), they will often fragment upon striking the water. Recovery in waste cotton can often address this issue. Such cartridges can also be "downloaded" by removing the bullet and reducing the amount of propellant in the cartridge case. This will normally reduce the velocity sufficiently to ensure that the bullet does not break up or overexpand when striking the water.

At least three test shots are initially discharged when preparing comparison samples but more are often required if the test results are poor. Ammunition that is the same as used in the crime should be chosen for the tests, or as close as possible to it if unavailable. For this reason, forensic firearms laboratories usually have large ammunition reference libraries containing as many different calibers and loads of ammunition possible, to ensure a suitable type is available for most situations.

Before the crime bullets are compared, comparisons should always be made between the test-fired bullets to assess the quality and quantity of individual characteristics present. The quality of the test results may be variable if the bore is rusty, pitted, or contains other debris that is expelled on each subsequent shot. If one test bullet cannot be matched with another, there is little hope of matching a test to a crime bullet. Another advantage to carrying out a test-to-test comparison is to check the functioning of the comparison microscope and that the magnification on both sides is even. Once the expert has established which features are consistently reproducing on the test-fired samples, those features can then be searched for on the crime bullet(s), firstly to correctly orientate the test to the crime bullet, then to either identify or exclude it as having been fired from the suspect firearm.

There are three main conclusions that can be reached upon completion of the microscopic comparison of fired bullets:

1. The bullet can be identified as being discharged from the firearm serial no. *xxxx.*
2. The bullet can be excluded as being discharged from the firearm serial no. *xxxx.*
3. The comparison is inconclusive; the bullet cannot be identified or excluded as having been fired in the firearm serial no. *xxxx.*

Inconclusive results are quite common. Often excessive damage has been occasioned to the bullet, or for various reasons there has been a lack of reproducibility of individual characteristics. Some laboratories may be more definitive in the event an inconclusive result is reached and divide an inconclusive result into further categories, which suggest a greater possibility, or not, as having been discharged from the firearm in question. All that the expert can honestly state, however, is that the bullet *may* have been discharged from the suspect firearm, *or any other firearm with those same class characteristics.*

The opinion formed by the expert must be based solely on his examination. It should not be colored by any other considerations such as being privy to other information in an investigation, which may suggest the guilt, or otherwise, of any person. As an example, eye witnesses may identify a person as having carried out a shooting. If this information is stated in the accompanying report when the examination is requested, this may create the belief in the mind of the expert that he will definitely be able to identify ammunition components to the firearm submitted for examination. All experts must be aware of the effect of "confirmation bias" when carrying out examinations.

In order for the expert to competently carry out bullet and cartridge case comparisons, there is a requirement to understand what constitutes an

identification and what does not. In other words, how much agreement in the individual characteristics is required to definitively state that a firearm, to the exclusion of all others, was responsible for discharging a bullet. This skill is developed over time underpinned by training, knowledge, and experience. But the only way to gain this understanding is to conduct thousands of "known matches" and "known nonmatches" using the comparison microscope.

When comparing bullets that have been test fired from the same firearm, the expert is comparing "known matches". There is no question that the bullets were fired from different guns, so the amount of agreement of striae on the bullet's bearing surface is usually substantial. That is not to say the majority of striae agree in known matches, for each comparison is different and the number of matching striae will vary. There is always a small amount of striae that do not match, but as long as the expert can reconcile the reasons for this, this is not an issue.

However, if the expert compares bullets he *knows* have been fired from two different firearms (with the same class characteristics), he is comparing "known nonmatches". The amount of matching striae in these situations is minimal. There is always some random matching of striae, but it is far less than what is seen in a known match. The expert therefore develops a "feel" for what constitutes an identification and what does not. Since the 1990s, however, this subjective approach has been criticized more and more by defense counsel, particularly in the United States. Their argument has been that this method is unreliable. To address these criticisms, conservative criteria for the identification of striated toolmarks was further developed and exhaustively tested. This approach to tabulating what is seen when comparing striated toolmarks is commonly referred to as *applying the principles of consecutive matching striae* (*CMS*).

Consecutively Matching Striae

CMS are sets or series of scratches that successively match without interruption, within a striated toolmark. Applying CMS to a striated toolmark simply allows the expert to *quantify* what is seen through the comparison microscope. After comprehensive research, two sets of conservative criteria were reached, one for two-dimensional marks and the other for three-dimensional marks, to allow the expert to confidently state whether an identification existed or not. CMS was not devised to replace the traditional subjective approach to matching striae, but to supplement it. Arguably, its greatest utility is application to toolmarks that are borderline as to whether an identification can be determined or not. If the criteria was met or exceeded, an identification could be reliably reported.

The conservative criteria for identification is as follows [5]:

In a two-dimensional toolmark, that is, a toolmark that has no appreciable surface contour, two groups of at least five matching striae appear in the same relative position or one group of eight.

In a three-dimensional toolmark, a toolmark that has discernable surface contour, two groups of at least three matching striae appear in the same relative position or one group of six.

In both cases, the possibility of any subclass characteristics influence must be ruled out.

Experience has shown that applying CMS could result in an identification being classed as inconclusive; this is due to the conservative nature of the CMS criteria. However, to date, it has never been reported in the literature that an identification has been erroneously arrived at by applying the CMS regime. Not all experts within the field of firearms and toolmarks agree on the necessity to apply CMS in conjunction with the traditional "pattern matching" approach, but it is accepted by many practitioners as being a valid additional tool to apply during the comparison of striated marks.

Cartridge Case Identification

Whereas bullets are normally engraved only with striated marks, cartridge cases are usually marked with a combination of striated and impressed marks. There is currently no criteria for the identification of impressed toolmarks; however, studies have been conducted into this issue[6, 7]. As the cartridge undergoes the loading, chambering, firing, extraction, and ejection processes, different components of the firearm mechanism contact various portions of the cartridge case, leaving a mark. As the cartridge case is made of a softer material than the action of the firearm, it will be marked accordingly. The main components that leave marks are those made by the firing pin, breech face, extractor, ejector, chamber,

the anvil, (in rimfire firearms) and the magazine. An identification of any or all of these marks can be made but only some (the firing pin, the breech face, and the anvil) are indicative of a discharge. Others indicate that a cartridge may only have been chambered in the firearm or cycled through the action.

Firing Pin Impressions

The firing pin (or hammer nose or striker, depending on the firearm) is responsible for striking the primer, thereby causing a discharge. An impressed mark is the normal result and its size, position, and spatial relationship to the other marks are important class characteristics. The tip of the firing pin usually contains sufficient individual characteristics to identify the firearm responsible, without the need to rely on additional marks. It should be kept in mind however that the firing pin, like many of the components relied upon for the comparison of fired cartridge cases, can be removed and changed so *all* the marks on a fired cartridge case should be taken into consideration before a conclusion is reached.

Breech Face Impressions

Breech face impressions are caused when the head of the cartridge is forced rearwards against the breech face upon discharge. The gas pressure, which drives the bullet from the barrel, also pushes in all other directions from within the cartridge case. In particular, the rearward action of the case opposite to the direction of the bullet will often leave an imprint of the features of the breech face on the case head. The primers of centerfire cartridges are especially prone to being marked by the breech face as they are made of a softer, thinner metal than the cartridge case.

The marks caused by the breech face are indicative of the machining processes used to form that part of the firearm. For example, if a milling machine was used to cut a circular recess for the case head, it could be expected that microscopic circular machining marks were left behind as a result of that process. A filed breech face might leave parallel marks on a case head – these may appear to be striations; however, the mark is still an *impression* of those striations. Other features of the breech face may also leave identifiable marks such as the aperture cut for the firing pin and various recesses cut for other components such as the extractor and ejector. If these shapes are present, they are often excellent class characteristics to assist determining the type of firearm used in their discharge.

Extractor Marks

The extractor is a small component on the side of a breech bolt that assists in removing a fired cartridge case or unfired cartridge from the breech of the weapon. As the action is opened, the extractor moves rearwards, whether attached to the breech bolt (as in a bolt action rifle) or not (as in a break action shotgun). The extractor may leave an impressed or striated mark, or both, depending on its particular arrangement in the firearm. The extractor on a self-loading firearm is more likely to leave a mark than, for instance, the multiple extractor of a revolver. This is primarily due to its greater relative force and speed of its operation.

The position of the extractor mark in relation to other marks is a useful class characteristic. If the mark contains sufficient individual characteristics, it may be used to identify a particular firearm.

Ejector Marks

The ejector is used to expel a fired case or cartridge from the weapon. Ejector marks are caused by the tip of the ejector as they make impact with the case or cartridge. A part of the mechanism, usually opposite the extractor, pushes on the cartridge rim. The rotating force produced while the extractor still grasps the rim on the other side, throws the cartridge or case from the action in a spinning fashion. The speed and force applied to the action when operated will often affect how far the case is thrown from the firearm and may also affect the quality of the ejector mark.

The position, shape, and size of the ejector mark are useful class characteristics. If the mark contains sufficient individual characteristics, it may be used to identify a particular firearm.

Chamber Marks

Chamber marks are produced by irregularities within the chamber, whether formed by the manufacturer such as an extractor cutaway, or by damage or corrosion. Chamber marks can consist of impressed

or striated marks. Impressions are produced when the case wall expands into unsupported areas of the case wall, while striations are caused when irregularities within the chamber scrape against the case wall upon loading, or more likely, upon extraction. They may be used to provide proof of discharge but as they may also be indicative of chambering only, some caution must be exercised when examining these marks to ensure the correct cause is determined.

Anvil Marks

Anvil marks are applicable to rimfire cartridge cases. When a rimfire cartridge is placed in the chamber, the rim of the cartridge overlaps the chamber mouth. The anvil could then be said to be an area extending outward in a ring, several millimeters from the chamber mouth. (In centerfire cartridges, the anvil is part of the cartridge itself and is located within the primer.) The anvil will be the location directly opposite the firing pin – this could be located at any point within 360° dependent on where the manufacturer has designed the position of the firing pin. Upon discharge, the cartridge rim is sandwiched between the tip of the firing pin and the anvil. This process will impart an impressed mark into the underside of the rim and any individual features existing at that location can be used to identify the firearm responsible.

Anvil marks have a much greater potential to be useful for identification if the firearm has been repeatedly dry-fired. Dry firing is the firing of a firearm when no cartridge is present. In a rimfire firearm, the firing pin will strike the anvil. Deformation to the chamber mouth and possibly the firing pin itself will occur over time, giving both the anvil and firing pin further individuality.

Miscellaneous Marks

Other parts of the firearm sometimes leave marks on cartridge cases, but these are usually due to particular peculiarities of a firearm and/or ammunition. These components can include magazine lips, ejection ports, feed ramps, and other parts dependent on how the action works.

Marks can also be accidental in nature and unrelated to the firearm. Cartridges that have been placed in a pocket with keys and coins, or left in a toolbox with tools may be marked at random by any number of hard surfaces. Noticeable marks can also be left by the cartridge manufacturer on ammunition components, which survive the firing process. Sometimes it is impossible to determine if a mark was firearm related or not, without the responsible firearm available for examination and testing.

References

[1] Hatcher, J., Jury, F. & Weller, J. (1957). *Firearms Investigation, Identification and Evidence*, 2nd Edition, The Stackpole Company.

[2] Gunther, J. & Gunther, C. (1935). *The Identification of Firearms – From Ammunition Fired therein with an Analysis of Legal Authorities*, John Wiley & Sons.

[3] Burrard, G. (1934). *The Identification of Firearms and Forensic Ballistics*, Herbert Jenkins Ltd.

[4] Mathews, J. (1962). *Firearms Identification*, The University of Wisconsin Press, Vol. 1.

[5] Biasotti, A. & Murdock, J. (1997). *Firearms and Toolmark Identification, Modern Scientific Evidence: The Law and Expert Testimony*, West Publishing, St. Paul, Vol. 2.

[6] Stone, R. (2003). How unique are impressed toolmarks? *AFTE Journal* **35**(4), 376–383.

[7] Collins, E. (2005). How "unique" are impressed toolmarks? – An empirical study of 20 worn hammer faces", *AFTE Journal* **37**(4), 252–295.

Related Articles

Firearms: Overview

Toolmarks

GERARD DUTTON

Firearms: Identification of Handling of Firearms/Trace Metal Detection

In the investigation of crimes involving the use of firearms, the forensic scientist is often asked to demonstrate a link between a suspect and a weapon. In homicide and suicide–homicide investigations in

which death is caused by shooting, a crucial question often is, who held the gun? The development of the scanning electron microscope (SEM)-based method for the unequivocal identification of gunshot residue (GSR) on the hands of a person suspected of having fired a weapon did not resolve the problem (*see also* **Firearm Discharge Residue: Analysis of Microscopy: Scanning Electron Microscopy**). The microscopic particles associated with the firing action are difficult to find, and their presence on the particular person does not necessarily prove that he fired the weapon [1, 2].

In its 1970 publication *trace metal detection technique* (*TMDT*), The National Institute of Law Enforcement and Criminal Justice in the United States has suggested applying the chelating agent oxine (8-hydroxyquinoline) to subjects' hands in order to detect metal traces that have been transferred from metallic objects, such as weapons, onto the holders' hands. The process involved spraying the hands, or dipping them in isopropanol solution of the reagent and, after air drying, observing them under short wavelength ultraviolet (UV) illumination [3]. The latent impressions resulting from handling a metallic object such as (but not only) firearms, show up as stains. There are stains that are colored while others fluoresce, depending on the particular metal involved. The test was thus named *TMDT*. Shortly afterward, the *TMDT* was thoroughly investigated by Stevens and Messler [4]. They ascribed the formation of the stains to the reaction between oxine and traces of metals that have migrated to the hand through handling the metallic objects (Figure 1(a)). In the case of firearms, the metal is mostly trivalent iron (Fe^{+3}). Stevens and Messler studied several factors that might affect the performance of the oxine reagent and suggested ways for optimal photography of the results [4]. They observed a correlation between the weapon's condition and the quality of the marks: stronger stains were obtained with low-quality, rusty handguns. The authors explained this observation by the presence of larger amounts of iron oxide (rust) on old or poorly maintained weapons, which reacts with the chelating reagent more readily than metallic iron. They also noticed a connection between the amount of perspiration on the hands and the quality of the marks: better marks were obtained on more perspiring hands.

In 1976, Goldman and Thornton reported a new and improved trace metal detection reagent: 3-(2-pyridyl)-5,6-diphenyl-1,2,4-triazine-p, p'-disulfonic acid, disodium salt trihydrate known also as *PDT* or *ferrozin* [5]. In contrast to oxine, *PDT* chelates only divalent iron, Fe^{+2} (Figure 1b). Its sensitivity in solution was found to be similar to that of oxine, but it had one striking advantage over the original TMDT test: the color developed is readily visible and may be easily photographed at room light. As the complex has a deep magenta color, no particular photography film was required. The authors also noticed that a positive reaction is dependent, to a considerable extent, on the amount of perspiration on the hands of the subject. The suggested procedure was quite simple: spraying the hands with methanolic solution of *PDT*. Traces of divalent iron show up as magenta stains within 1 min. Exposure to long wavelength UV light accelerates the color formation. Most heavy metal ions do not interfere with this reaction, but some anions, particularly cyanide, do by competing with *PDT* for the iron. In a following article [6], Thornton and Stoney recommended a modified *PDT* reagent. The improved reagent, 3-(2-pyridyl)-5,6-diphenyl-1,2,4-triazine (Figure 1c), has the same core structure as *PDT*, but lacks the two sulfonic acid groups. Consequently, it produces with divalent iron an analogous chelate (devoid of the sulfonic groups), whose color is very similar to that of the former. The major advantage of the modified reagent is its solubility in nonpolar solvents such as acetone, whereas *PDT* is soluble only in relatively polar solvents, which can also dissolve the stains of the magenta complex. The suggested technique [6] involves spraying the subject's hand with 0.5% reagent solution in acetone, and if divalent iron is present, a magenta stain should appear instantly. The authors reported the molar absorptivity of pyridyldiphenyltriazine–iron chelate to be 24 000 at 555 nm, slightly lower than 27 900 at 562 nm for the *PDT* chelate, but still sufficiently high to ensure a visible reaction with microgram amounts of iron. Later on, this modified reagent received the name PDT, since pyridyldiphenyltriazine is the parent compound, and the former reagent (Figure 1b) is but a chemical derivative [6]. Thus, in the rest of the article, reagent (Figure 1c) is referred to as PDT.

Shortly afterward, it was noticed that although the new PDT was advantageous over the previously tried reagents, it suffered from a significant drawback; in many cases, colored stains could not be

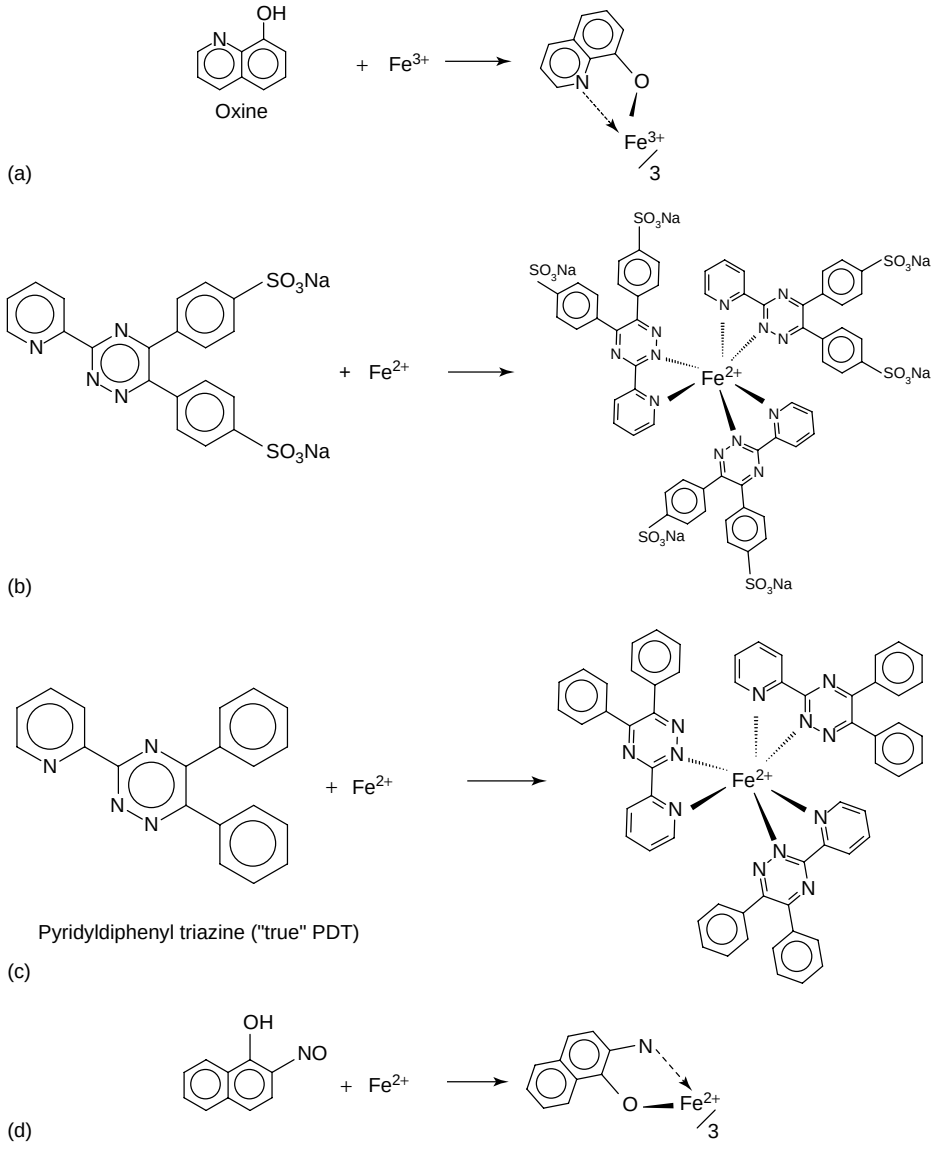
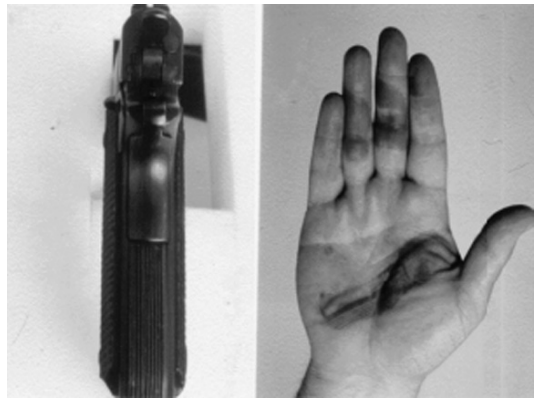


Figure 1 (a) The oxine reaction with trivalent iron; (b) the reaction between *PDT* with divalent iron; (c) the reaction between pyridyldiphenyltriazine ("true" PDT) and divalent iron; and (d) 2-nitrozo-1-naphthol reaction with divalent iron

developed even after a confirmed recent contact with firearms. This happened particularly when the tests were carried out more than 30 min after contact. The rationale to this phenomenon was suggested independently by Glattstein and Kraus in Israel [7] and Lee in Hong Kong [8]. Both groups assumed that the amounts of iron in its divalent state (Fe^{+2}) on the hands were insufficient for a positive PDT reaction. This could arise from either fast oxidation by air oxygen of the divalent iron, which is originally present [7], or from an initially small amount of Fe^{+2} on the hand due to rust (basically Fe^{+3}), covering the object [8], or a combination of both. Oxidation by air oxygen also explains why relatively old impressions could not be developed with PDT, since divalent iron is known to quickly oxidize by air to the trivalent state, particularly in a moist environment. Both groups suggested the



(a)



(b)

Figure 2 (a) Holding a Colt 45 pistol; and (b) the marks that are developed by spraying the hand with PDT solution [Reproduced with permission from Ref. 11. © IAI, 1998.]

use of the mild reducing agent, ascorbic acid, in conjunction with the PDT reagent. Ascorbic acid was found to considerably enhance the sensitivity of the PDT technique, allowing the detection of much older firearms impressions. Since 1980, the improved technique has been in operational use by crime-scene officers in the Israeli Police, and became a part of the "field test concept" there [9]. Spray cans containing PDT and ascorbic acid became commercially available under the trade marks Ferroprint™ and Ferrotrace™. It could visualize latent iron impressions on the hands hours after contact [10]. Quite often, the shape of the PDT impressions could be associated with certain firearms. Profiles obtained from commonly used handguns were reported by Glattstein *et al.* [11] (Figure 2).

In his 1986 article, "The Detection of Iron Traces on Hands by Ferrozine Sprays" [8], C-W Lee reported the results of a comprehensive study, which was aimed at optimizing the PDT technique. He tried several reducing agents (ascorbic acid was the most efficient), examined potential interference by traces of other metals (copper, zinc, nickel, aluminum, tin, lead, and chromium) and suggested ways to remove them, or interpret the results in their presence. Lee also outlined a procedure for transferring iron traces from hands onto filter paper for persons who refuse to have their hands sprayed by the reagent. Of the metal ions tested, only aluminum formed the familiar magenta color with PDT, but the color intensity of the product was 350 times lower than that of the iron complex. Ni^{+2} on the other hand, had an inhibitory effect on the formation of the iron magenta complex.

Despite the significant improvement of the PDT technique, there were still too many occasions in which the modified reagent showed insufficient sensitivity. The next series of comprehensive studies started with an attempt to design chelating agents

that would provide a deeper color (higher absorptivity) with iron. More than a dozen analogs of PDT have been synthesized and tried. In solution, eight of them exhibited higher sensitivity than the parent compound to Fe^{+2}. The gain, however, was insufficient (not more than 50% increase in color intensity) to justify the use of these noncommercial compounds as practical reagents [12]. Comment *et al.*, from the University of Lausanne compared the performance of four potential reagents for the detection of firearms impressions on the hands. Besides the three previously mentioned chelating agents, their study also involved a fourth one, 2-nitroso-1-naphthol (Figure 1d). PDT showed the highest sensitivity of all four. Furthermore, the marks developed by PDT were the most stable and showed the highest resolution [13]. A similar study on TMDT, which was concluded without specific preference, was published a few years earlier by Katzung [14].

In the last decade, more qualitative and quantitative characteristics of the PDT reaction have been reported. Studies were carried out not only for sensitivity enhancement but also for court argumentation. Leifer *et al.* reported that about 70% of a large number of volunteers developed visible stains after holding firearms. They did not notice significant differences between males and females in the effectiveness of the technique or in stain quality [15]. Avissar *et al.* presented, for the first time, quantitative data on the amounts of iron that migrate from a firearm to the palm. Iron levels between 21 and 315 ng cm^{-2} have been measured on the palm after a single grip that lasted 1 min. These authors found that the amounts of iron transferred to the palm depended primarily on the levels of palmar moisture and, to a lesser extent, on gripping duration. Thus, only a few seconds of gripping were required for developing good marks (corresponding to 80 ng cm^{-2} of iron) on highly moistured hands, whereas much longer gripping periods were necessary to develop marks of similar intensity on relatively dry hands [16]. Experiments by the same group that aimed at studying the effect of sweat components on metallic iron dissolution were carried out in aqueous solutions. The authors found that chloride ions in physiological concentrations remarkably enhanced iron dissolution, while serine, the major amino acid in palmar sweat, had a detrimental effect on this reaction. Urea, another sweat component, had only a minor effect on the dissolution rate. The authors proved that most of the iron is transferred from the weapon to the skin by a chemical process (dissolution) and not by mechanical friction. Almog *et al.* also quantified the correlation between moisture levels on the hand and the quality of the developed stains. They divided the population into three categories based on their iron uptake. Similarly to *fingerprint donor* for the person who leaves the latent finger marks, they used the term *iron acceptor*, or simply *acceptor*, for the person who adsorbs iron traces upon gripping an iron-made object (*see also* **Friction Ridge Examination (Fingerprints): Interpretation of**). Accordingly, the three categories mentioned above can be termed *good*, *average*, and *weak acceptors*, being determined by the intensity of the PDT mark, which, in turn, depends on the amount of perspiration on their hands [16, 17].

Four death investigations, in which the use of the PDT technique helped the investigators answer the critical question, "who held the lethal weapon", have been reported and discussed by Leifer *et al.* In one of those cases, a PDT impression developed not on the palm but on the hip. In another case, a suicide–homicide investigation, PDT results reversed the initial, "obvious" conclusion of the investigators [18] (*see also* **Suicide (Behavior)**).

References

[1] Zeichner, A. & Levin, N. (1993). Collection efficiency of gunshot residue (GSR) particles from hair and hands, *Journal of Forensic Sciences* **38**(3), 571–584.

[2] Zeichner, A. & Levin, N. (1995). Casework experience of GSR detection in Israel, on samples from hands, hair and clothing using an Autosearch SEM/EDX system, *Journal of Forensic Sciences* **40**(6), 1082–1085.

[3] (1970). *Trace Metal Detection Technique in Law Enforcement*, Pamphlet No. 71-1, National Institute of Law Enforcement and Criminal Justice, LEAA, Washington, DC.

[4] Stevens, J.M. & Messler, H. (1974). The trace metal detection technique (TMDT): a report outlining a procedure for photographing results in color, and some factors influencing the results in controlled laboratory tests, *Journal of Forensic Sciences* **19**(1), 496–503.

[5] Goldman, G.L. & Thornton, J.I. (1976). A new trace ferrous metal detection reagent, *Journal of Forensic Sciences* **21**(3), 625–628.

[6] Thornton, J.I. & Stoney, D.A. (1977). Improved ferrous metal detection reagent, *Journal of Forensic Sciences* **22**(4), 739–741.

[7] Glattstein, B. & Kraus, S. (1979). Metal traces detection on hands of suspects, *Israel Police Toolmarks and Materials Lab Research Report* (1), Sept (in Hebrew) 1–4.

[8] Lee, C.-W. (1986). The detection of iron traces on hands by ferrozine sprays: a report on the sensitivity and interference of the method and recommended procedure in forensic science investigation, *Journal of Forensic Sciences* **31**(3), 920–930.
[9] Kaplan, M. & Almog, J. (1983). Field diagnostic examinations for forensic purposes, *Police Chief* 30–33.
[10] Almog, J. & Glattstein, B. (1997). Detection of firearms imprints on hands of suspects: study of the PDT-based field test, *Journal of Forensic Sciences* **42**(6), 993–996.
[11] Glattstein, B., Nedivi, L. & Almog, J. (1998). Detection of firearms imprints on hands by the Ferrotrace spray: profiles of some common weapons, *Journal of Forensic Identification* **48**(3), 257–272.
[12] Almog, J., Hirshfeld, A., Glattstein, B., Sterling, J. & Goren, Z. (1996). Chromogenic reagents for iron(II): Studies in the 1,2,4-triazine series, *Analytica Chimica Acta* **322**(3), 203–208.
[13] Comment, S., Bonfanti, M. & Gallusser, A. (1998). Détermination de la main qui a tenu une arme sans forcément avoir tiré, *Revue Canadienne des Sciences Judiciaires* **31**(2), 79–94.
[14] Katzung, W. (1985). Möglichkeiten der sichtbarmachung von metallspuren durch anwendung der trace metal detection technique, *Kriminalistik und Forensische Wissenschaften* **57, 58**, 74–81.
[15] Leifer, A., Avissar, Y., Berger, S., Wax, H., Donchin, Y. & Almog, J. (2001). Detection of firearm imprints on the hands of suspects: effectiveness of PDT reaction, *Journal of Forensic Sciences* **46**(6), 1442–1446.
[16] Avissar, Y., Sagiv, A., Mandler, D. & Almog, J. (2004). Identification of firearms holders by the $[Fe(PDT)_3]^{+2}$ complex. Quantitative determination of iron transfer to the hand and its dependence on palmar moisture levels, *Journal of Forensic Sciences* **49**(6), 1215–1219.
[17] Avissar, Y.Y., Sagiv, A.E., Mandler, D. & Almog, J. (2005). Identification of firearms handling by the $[Fe(PDT)_3]^{+2}$ complex: chemical and time-dependent factors, *Talanta* **67**(2), 328–333.
[18] Leifer, A., Wax, H. & Almog, J. (2001). Who held the gun, decipherment of suicide-homicide cases using the PDT reagent, *Journal of Forensic Identification* **51**(4), 346–360.

JOSEPH ALMOG

Firearms: Number Restoration

see Serial Number Restoration: Firearm

Firearms: Overview

Overview

Forensic firearms investigation is the discipline concerned with the technical analysis of crime where firearms have been used. Firearm crime includes unlawful possession of firearms and related items, where shots have been discharged into the air or other objects such as cars or buildings, to serious crimes where people have been wounded or killed by the discharge of firearms. Firearms investigation could be considered as a specialized form of crime scene examination and such is the depth of knowledge and expertise required to competently investigate shooting crimes from a technical viewpoint; it is a complete field of study in itself.

Firearms investigation is often referred to as *forensic ballistics*; however, this is a technically incorrect term, and ballistics refers purely to the study of projectiles in motion. "Forensic firearms examination" is sometimes used but this can infer that the discipline is purely laboratory based. "Forensic firearms investigation" is the most appropriate term to describe the science as it encompasses the diverse range of examinations carried out at the shooting scene and in the laboratory.

The Laboratory

Most forensic firearms laboratories are owned and financed by their respective governments, usually a state or federal authority due to the costs involved in buying the necessary equipment and hiring/training staff. Modern police jurisdictions have police officers or civilian experts trained in the investigation of shooting incidents, usually located in their own separate section of the police forensic services division. Where there is no internal forensic firearms investigation section within the police department, this role may be contracted out to a larger law enforcement agency. There are some private individuals or companies who do have their own laboratories; however, these would not normally be responsible for the investigation of a shooting incident immediately after it had occurred. They would become involved at a later date, usually at the request of the lawyer

defending the person/s charged, to test the thoroughness of the work carried out by the government laboratory.

There is quite a wide variation around the world between different forensic firearms laboratories and the services they provide. There are a number of reasons for this including how the section may have developed historically, financial considerations, and the particular types of firearm crime most commonly encountered in that jurisdiction. For instance, it would be uncommon for a small jurisdiction to have a scanning electron microscope (SEM) for the analysis of gunshot residue from primer components if the authorities only encountered this evidence infrequently. If this expertise was required, there are usually protocols in place to have that evidence examined by larger laboratories equipped for such analysis. This may or may not be on a cost recovery basis.

Staffing of laboratories is commensurate with the amount of work encountered in that jurisdiction involving the criminal use of firearms. Some forensic firearms sections also have other responsibilities and roles such as servicing and repairing police firearms. The extra workload involved in providing an armoring service may then require staffing levels above that needed to carry out the forensic firearms investigation role. In such situations, personnel may either maintain their separate and distinct roles or carry out all the functions in that section.

The depth and breadth of any examination undertaken at the shooting scene or in the laboratory is normally commensurate with the seriousness of the crime. Consequently, minor incidents are usually not investigated to the same extensive level as a wounding or death. Examinations are very dependant on the circumstances of the incident in question and what may be of crucial importance in one investigation may be irrelevant in another. For example, whereas determining the distance between the victim and the offender may be critical in one incident, the condition of the safety mechanism of the firearm may be the most important factor to consider in another. What needs to be explored will be due to the relevant aspects of each particular case and after determining what is and what is not in dispute.

Broadly speaking, firearms investigation can be separated into several main roles: the investigation and reconstruction of the shooting incident, gunshot wound interpretation, laboratory examinations, report preparation, and expert testimony.

1. The reconstruction of the shooting incident

There are many considerations for the firearms investigator at the scene of a shooting crime but of prime importance are issues such as the following:

- Where was the victim located at the time of the shooting?
- How was the victim's body positioned at the time of impact?
- Where was the shooter located at the time of the shooting?
- How was the firearm held by the shooter?
- How many shots were fired?
- What was the sequence of the shots?
- Can the bullet paths and angles be determined?
- What was the physical condition of the responsible firearm when located in regards to the chamber, magazine, safety catch and so on?

The competent and thorough examination of a shooting scene is vital to the overall investigation mainly due to the potential at that juncture to lose critical evidence if it is not properly carried out. Staff in some jurisdictions are laboratory based only and do not attend shootings scenes; this role being undertaken by crime scene officers. However, it may be preferable that firearms investigators conduct shooting scene examinations as part of their role to place the incident in a more complete context; this may be especially relevant when conducting other examinations later in the laboratory. In the event that a death has occurred, again depending on the jurisdiction in which the shooting took place and the duties and responsibilities of the staff from the forensic laboratory, shooting investigation can also include assisting the forensic pathologist during autopsies upon gunshot victims.

2. Gunshot wound interpretation

Although the forensic pathologist (i.e., a medical examiner) is responsible for actually conducting the examination upon a shooting victim, it is not unusual for a firearms investigator to be present during the autopsy to offer advice and assistance in areas such as the following:

- interpretation of the wound/s such as which are entries and exits;
- the approximate distance at which the shot was fired in close range discharges;
- determination of the bullet path through the body;
- the possible orientation of the victim when struck;
- the possible sequence of shots in incidents where more than one impact is present; and
- recovery of projectiles and other evidence for later examination.

Once all the evidence has been recovered from the scene and during the autopsy (if applicable), a number of examinations are then carried out in the laboratory. These may include, but are not limited to, the examination of recovered projectiles and other types of evidence and/or tests to estimate the firing distance, to determine the bullet trajectory or path, etc.

3. Laboratory examinations

The expert conducts a wide range of laboratory examinations upon firearms, fired and unfired ammunition, bullet damaged clothing, and related items. Often these examinations cannot be completed until other forms of analysis are conducted first, such as testing for fingerprints and swabbing for DNA. The sequence of examination regime is not rigid but is dependent on the context of the case and on the priorities at the time, with due consideration to maximizing all the available evidence. Laboratory examinations can include the following:

- examination of suspect firearms for features, operability, function, safety, trigger pressure, etc.;
- microscopic comparisons between fired bullets and cartridge cases test fired in suspect firearms, to those recovered at the shooting scene;
- serial number restorations upon defaced/obliterated serial numbers on firearms;
- gunshot residue tests to indicate the manipulation or the use of a firearm and/or establish the approximate distance between the shooter and the target (close range discharges only);
- shotgun pellet pattern tests to establish the approximate distance between the shooter and the target;
- ejection pattern tests to assist determining where the shooter may have been positioned at the time of discharge; and
- various other tests as deemed necessary by the circumstances.

4. Preparation of reports and expert testimony

The final part of the investigation is the preparation of technical reports. Reports should ideally articulate what was done during each stage of the investigation, what exhibits were collected, and the results of all the examinations conducted. Reports should not be overly technical and contain jargon, making it difficult to understand by the layperson but the specifics referred to must still be technically correct. If requested by either the prosecution or defense legal representative, the expert would also attend court or other tribunal to orally give evidence of their involvement in the investigation.

There is only one forensic association worldwide that is primarily involved with firearm investigation, the Association of Firearm and Toolmark Examiners (AFTE). It has a membership of approximately 1000 individuals from 40 countries. A scientific journal is produced quarterly and an annual training seminar is held, usually within the United States. More information about AFTE and firearms investigation can be obtained from their website: www.afte.org. Another excellent website for reliable information on firearms identification is www.firearmsid.com.

Further Reading

DiMaio, V.J.M. (1999). *Gunshot Wounds, 2nd Edition*, CRC Press.

Haag, L. (2006). *Shooting Incident Reconstruction*, Academic Press.

Hueske, E.E. (2006). *Practical Ballistics and Reconstruction of Shooting Incidents*, Taylor & Francis.

Heard, B.J. (1997). *Handbook of Firearms and Ballistics*, John Wiley & Sons.

Warlow, T.A. (1996). *Firearms, the Law and Forensic Ballistics*, Taylor & Francis.

Related Articles

Firearm Examination: Ballistics

Firearms: Bullet and Cartridge Case Identification

Firearm Discharge Residue: Analysis of

Firearms: Identification of Handling of Firearms/Trace Metal Detection

Firearms: Scene Investigation

Shooting Distance: Estimation of

Serial Number Restoration: Firearm

Toolmarks

GERARD DUTTON

Firearms: Scene Investigation

Introduction

Shooting scene investigation is a specialized form of crime scene examination, but the techniques applicable to general crime scene investigation are still relevant. It relates to the recording and collection of specific evidence that is unique to a shooting scene, to enable accurate interpretation and reconstruction of the events that have occurred. Forensic firearms examiners provide a service and pool of expertise in matters relating to the investigation of shooting incidents. Their main functions are as follows:

- interpretation and reconstruction of shooting scenes;
- identification of firearms through the examination of cartridge cases and bullets;
- safety testing of firearms;
- general identification of firearms and firearm components;
- identification of ammunition;
- performance of ammunition;
- distance determination of shotgun patterns;
- distance determination from propellant patterns;
- restoration of erased serial numbers on firearms;
- checking firearms against an unsolved shootings database; and
- validating or negating victim/offender/witness statements.

The following provides an overview of the types of evidence that can be found at shooting scenes, as well as from victims and offenders.

The Shooting Scene

Documenting the position and condition of all relevant exhibits at a shooting scene is critical to ensuring that the combined evidence can be accurately evaluated and interpreted. There is only one opportunity to 'get it right' and failure to do so results in the loss of evidence, which cannot be recovered at a later stage.

Firearms

The condition of a firearm at a crime scene should be consistent with the circumstances of the alleged offense. If it is not, then it must be explained. The following details should be checked and recorded.

- condition of the breech i.e., unloaded or loaded with a live round or spent cartridge case;
- action cocked or uncocked, open or closed;
- position of hammer or cocking piece i.e., cocked or uncocked;
- magazine type and contents;
- position of any safety catch – on or off;
- relationship of the firearm to other exhibits and features at the scene; and
- any visible blood or biological tissue present on the firearm or within the muzzle of the barrel.

Before removing a firearm from the scene, consideration should be given to any requirements for DNA or fingerprint examination. It is essential that any such exhibits are not exposed to contamination.

Live Ammunition, Fired Cartridge Cases or Fired Bullets

Live ammunition can be found with the firearm or lying on the ground at the scene. Depending on the size of the cartridge case, fingerprint examination and DNA analysis should be considered.

Fired cartridge cases are unlikely to retain fingerprints or DNA evidence owing to the heat and pressure generated within the chamber of the firearm at the time of firing or because of other environmental and circumstantial factors. The only exception to this is if there is a suggestion the offender has handled

the cartridge cases postfiring, in an effort to remove evidence from the scene.

When searching for spent cartridge cases, consideration should be given to the use of the following to aid in the search:

- metal detectors
- explosive sniffer dogs
- line searches.

Fired bullets can retain trace evidence such as biological material and therefore can be a rich source of DNA evidence as well.

Projectile Impact Damage

This occurs when a bullet impacts an object during flight. Impact damage is commonly seen in building exterior and interior walls, windows, vehicles, and clothing at shooting scenes. There are two main considerations with this type of evidence. The first is to confirm the damage has been caused by a bullet. If the cause of the damage is uncertain or unknown, there are two presumptive tests that can be used to determine if the damage was caused by a bullet. Sodium rhodizonate can be used to determine if lead residues are present. Lead is the most common element used in the manufacture of most bullets inner core. Dithiooxamide (DTO) is used to test for copper residues. Copper is the most common metal used in the manufacture of the outer jacket material on many bullets. If both the lead and copper tests are to be conducted, it is imperative that the DTO test is conducted before the lead test. This is because the mild acidic solution used in the lead test also transfers copper residues.

The second consideration is to determine if the impact damage can be used to determine the path of the bullet and the approximate position of the shooter at the time of firing. A minimum of two separate reference points are required for this purpose, for example an entry and exit hole through an inner gyprock wall of a house. A colored plastic or wooden probe can be placed through both holes allowing an angle of incidence (entry) to be determined. Other measurements such as the height of the impact above the floor and its relationship to other features of the scene are also important. This combined information together with trigonometric calculations, can be used to determine the position of the shooter.

An alternative method of achieving a similar result, while demonstrating the approximate trajectory or path of the bullet, is the use of colored stringlines. Again, two reference points are required. By following the stringline/s back to a common point or to a maximum/minimum height, the approximate position of the shooter can be determined. Colored stringlines and probes can be readily photographed, which makes them useful for court presentation purposes as well. However, stringlines suffer from sag over extended distances, which can lead to inaccurate findings. Therefore, they should only be used over shorter distances.

Laser devices/pointers can also be used to identify possible trajectories if there are two or more linked features. They can be particularly useful over longer distances, where stringlines become impractical. However, lasers point in straight lines and if using such a device over an extended distance, allowance must be made for trajectory variations owing to factors such as bullet drop. Test firings over the distance in question may be necessary to determine the actual amount of bullet drop, as bullets do not travel in a straight line relative to the vertical plane, owing to the factors such as gravity. Unlike stringlines, the path indicated by a laser cannot be photographed in daylight conditions.

Bullet impacts into some materials can leave features, which assist in determining their direction of travel. The ‘pinch point’ is one example frequently seen on the sheet metal of motor vehicles. This is a small circular area of surviving paint at one end of a bullet entry hole or ricochet mark. It represents the first point of contact of the bullet and therefore provides directionality to the impact mark. Another feature of bullet impact marks in motor vehicles is called the ‘bow effect.’ These are fracture lines in painted sheet metal, which often must be enhanced with the use of a fingerprint powder that contrasts against the painted surface. The fracture lines have the appearance of a bow wave or a shock wave and show the direction of travel of the bullet.

Holes in Glass

There are a number of pitfalls interpreting holes in glass. Bullet holes are generally round and larger than the caliber of the bullet. Radial and concentric fracture lines emanate from the center of the hole. This applies to safety glass and window glass. The

bullet usually penetrates the glass and so a second strike inside the room is likely to be found. The presence of lead or copper close to the impact point may serve as an indicator of an impact being made by a bullet. This may be confirmed by sampling the area around the questioned hole using the DTO and/or sodium rhodizonate presumptive tests.

Where multiple bullet impacts occur to a glass window, it is possible to determine the 'sequence of shots' by reference to the radial fracture lines from one impact terminating against radial fracture lines from an earlier impact.

Stone damage may result in an irregularly shaped hole and the radial fracture lines may emanate from one edge of the hole, if the stone has not penetrated. The presence of the projectile is the best indicator of the cause of the damage, and a careful search should always be made for any relevant object.

A ball bearing fired from a slingshot often leaves a small round hole, which exhibits crushing around the outside with a cone of glass knocked out from inside. Damage by a ball bearing that has not penetrated the glass may leave a small round symmetrical hole with fine crushing around the outside edges. Again, a careful search for the projectile may confirm the cause of the damage.

A bullet produces a hole that has the shape of a truncated cone with the larger base at the point of exit and the smaller at the point of entrance of the bullet.

Bullet Ricochet

Ricochet is the deflection of a projectile(s) after impact. The amount of deflection depends on the medium involved. As a general rule, studies have shown that a bullet is deflected to some degree after passing through any material that offers even minor resistance, i.e., color-bond fencing, household glass windows, or windscreen glass. Bullets also ricochet off solid mediums such as roads or brick walls if the angle of incidence is other than 90° to it. As a general rule, bullets fired into yielding materials such as sand, have a higher angle of departure than the angle of incidence. Bullets fired into nonyielding materials such as concrete, have lower angle of departure than the angle of incidence.

Ricochet is an important consideration because it is a common defense raised by barristers and solicitors to account for the wounds sustained by a victim following the actions of the accused.

Shotgun Patterns

Shot shell cartridges contain multiple pellets. The dispersion pattern left by shot shell pellets can be used to express an opinion as to the range of the shot. Elongated patterns are indicative of angled shots. The dimensions of the pattern must be recorded accurately so that subsequent test firings using the actual firearm and similar ammunition can be made in an effort to replicate the pattern. The number of pellet impacts and their relative size assist in determining the type of shot shell load used. Shotgun wads can travel out to a distance of approximately 30 m, depending on the length of the shotgun barrel. They can also assist to identify the type of shot shell load.

Shooting Suicides

Shooting suicides can represent a particularly difficult form of shooting interpretation because it is possible that an apparent suicide is in fact a concealed murder. To resolve this question, the following are signs which generally indicate suicide:

- **History**

The victim may have a history of trouble or depression.

- **Firearm**

Should be identified as being one available to the victim.

- **Suicide note**

May be left at the scene. (Is it the victim's handwriting?)

- **Wound site**

Victims will usually shoot themselves in one of the preferred areas:

under the chin
inside the mouth
between the eyes
in the temple
in the center of the chest.

All these wounds exhibit evidence of having been fired from close to, or in contact with the body, i.e., powder tattooing (burnt and partially burnt

propellant grains embedded around the wound site), recoil injury, or splitting of the entry when it is over hard bone. A wound to any other part of the body should be viewed with suspicion unless accompanied by other indications of suicide.

- **Victim's reach**

With the muzzle of the weapon in contact with the entry wound, the victim must be able to reach the trigger by hand or by using some other improvised means to extend his/her reach.

- **Condition of the firearm**

Must be consistent with the circumstances.

- **DNA/fingerprints**

Should be checked on the firearm.

- **Multiple wounds**

Are unusual although not entirely uncommon. They may occur when the first shot was not crippling, such as a shot to the body or where the weapon was semiautomatic and the second shot has resulted from the victim's finger being retained inside the trigger guard.

- **Other injuries**

The victim may have injuries to the hand holding the muzzle caused by the foresight or to the hand pulling the trigger, caused by the trigger guard.

- **Signs of struggle**

There should be no signs of a struggle.

- **Position of firearm**

Should be consistent with having been held by the victim at the time of the shooting. However, there are occasions when recoil can cause the firearm to land further away than expected.

DNA versus Fingerprints

DNA, like fingerprints, is now accepted by the courts as powerful evidence in criminal proceedings. However, examiners are often faced with a difficult decision as to which of these is more important relative to a particular exhibit. This situation often arises with live or spent cartridge cases.

Smaller cartridge cases such as caliber. 22 long rifle, do not have the surface area to hold a substantial fingerprint and so DNA swabbing is the preferred option. However, larger cartridge cases such as caliber 45 ACP do have sufficient surface area and so fingerprint examination & DNA swabbing must both be considered viable options depending on the circumstances.

With regards to firearms, DNA swabbing should be undertaken first to avoid further contamination. However, DNA swabs should only be taken from areas on the firearm that are unlikely to yield fingerprints, such as the trigger, cocking handle, and chequered or rough areas of the grips.

The issue of how best it is possible to collect both fingerprints and DNA from a given exhibit is thoroughly discussed in **Friction Ridge Skin: Interaction between Fingerprint Detection and DNA/Biological Material**.

Scene Reconstruction

There are occasions when reconstructing a crime scene, or aspects of it, are necessary to identify possibilities or test a given scenario. It is usually based on information provided by persons involved in the shooting incident. Scene reconstructions are particularly common in police involved shootings or major homicides. It is important to keep an open mind with respect to any reconstruction because witnesses' accounts may not be entirely accurate.

Interpretation

Accurate interpretation of a scene is usually only possible once all the available evidence has been found and examined. Interpretation is basically reconstructing the chain of events related to the shooting from the available evidence. It allows the examiner to form conclusions on the basis of fact and therefore confirm or negate scenarios that may be posed by investigators, victims, or offenders. Courts and coronial inquests often rely on the opinions of forensic firearms examiners when determining the weight of the evidence before them. Accurate interpretation generally comes with experience, but also depends on proper education and training.

Other Considerations

Ambient lighting has a considerable effect on the ability of a shooter to take an aimed shot. Open sights on firearms require both the firearm and the target to be illuminated, while a telescopic sight requires only

the target to be illuminated. Other considerations are as follows:

- features that may affect line of sight;
- features that could result in an unintentional discharge such as a loss of footing on uneven ground;
- background features; and
- intermediate targets.

The Shooting Victim

A vast amount of evidence can be gathered from an examination of the victim of a shooting, both in a suicide and a homicide or nonfatal shooting.

Wound Position and Trajectory

The wound position and direction through the body, if both entry and exit are visible, may assist in determining the attitude of the victim at the time of the shooting. Wounds to the hands and arms may be defense wounds, where the victim has tried to ward off the shot.

Powder Tattooing

This term refers to burnt and partially burnt propellant particles that are expelled from the muzzle of a firearm and embed themselves in the skin surrounding a wound. Tattooing is generally seen on uncovered skin out to a maximum of 2 m, but more often at distances far less than this. The pattern depends on the firearm ammunition and environmental conditions (eg., wind in outdoor situations). The distribution of this propellant residue can allow forensic firearms examiners to express an opinion as to the range of the shot, otherwise known as '*distance determination*'. This is usually achieved by subsequent test firings using the actual firearm and ammunition of the same type. Therefore, it is important the pattern is well documented.

Residues on Clothing

If the victim is shot through clothing, at close range, the cloth retains the propellant and primer residues. Depending on the color of the clothing, these residues may be visible or invisible. Infrared imaging and other filtered light source examination may help to enhance the contrast of the residues against the background. Chemical analysis is the best method of determining the extent of the pattern produced, so that a forensic firearms examiner can conduct subsequent test firings to determine the distance from which the shot was fired.

Preservation of the pattern produced on the clothing is paramount. To ensure this, the clothing item should be removed from the body, dried, and laid flat so that the portion of cloth with the bullet hole is sandwiched between two sheets of paper or cardboard. This minimizes abrasion with other parts of the garment, which may remove loose material.

Postmortems

Forensic firearms examiners attend postmortem examinations for two main reasons:

- to collect firearms evidence, which includes clothing; and
- to assist the pathologist in the interpretation of gunshot wounds.

The Shooting Suspect

Suspect Weapon

Any relevant firearm in the possession of the suspect should be seized and forwarded for examination by a forensic firearm examiner. It is possible to macroscopically compare cartridge cases and bullets recovered at a shooting scene to the suspect weapon/s using a comparison microscope and comparative analysis techniques. Safety tests can also be conducted to confirm or negate allegations of accidental discharge.

Ammunition

If a suspect has ammunition in his/her possession, this too should be collected. Examination of the brand and type may provide a link to any ammunition at the shooting scene or it may be needed in tests to be conducted by the forensic firearm examiner. Macroscopic comparison of the headstamps on live ammunition or fired cartridge cases at the scene can be compared with those of live ammunition found in the possession of the suspect, to determine if the headstamps were formed by the same bunter tool

at the time of manufacture. This evidence can be particularly useful if the type of ammunition used is uncommon.

Medical Examination

The medical examination of the victim is an integral part of any investigation and may reveal crucial information with respect to the wound, firing range, and other circumstantial information. A medical examination of the suspect can reveal injuries that may have occurred in a struggle over a weapon, or bruising that may indicate that a poorly shouldered shotgun/rifle was used.

Gunshot Residue Analysis

If a suspect is located within approximately four hours of a shooting and he/she has not washed their hands, swabs may be taken for analysis of gunshot residue (GSR). When a person fires a firearm, gaseous residues are created, which escape from it. These may deposit distinctive combinations of elements upon the hands of the shooter and sometimes other areas such as in their hair or on their clothing. Some firearms such as revolvers or semiautomatic rifles or pistols, emit substantially more gaseous residues than a bolt action or lever action rifle because of their operating system. A 'GSR kit' can be used to collect these residues. GSR is microscopic and mainly consists of primer residues often containing the principle elements of lead, barium, and antimony. They are enhanced using chemical techniques and examined with the aid of a 'scanning electron microscope'.

It should be noted that the results of GSR analysis are not a definitive answer to whether a person has fired a firearm or not, as there may be other explanations. GSR is at best corroborative evidence. See **Firearm Discharge Residue: Analysis of** for a thorough discussion of this topic.

Further Reading

General Crime Scene Recording

Garrison, D.H. (1996). Recording bullet defects at crime scenes, *AFTE Journal* **28**(3), 168.

Goddard, C. (1980). Firearms evidence, *AFTE Journal* **12**(4), 93.

Haag, L.C. (1998). Cartridge case ejection patterns, *AFTE Journal* **30**(2), Spring, 300.

Hart, R.P. (1975). Evidentiary value of loaded firearms, *AFTE Journal* **7**(3), 35.

Rathman, G.A. (1988). The effect of material hardness on the appearance of bullet impact damage, *AFTE Journal* **20**(3), 300.

Silliman, J. (1977). Shotgun related evidence, *AFTE Journal* **9**(2), 111.

Shooting Scene Reconstruction

Courtney, M. (1994). The use of hand held laser pointers in the reconstruction of events at crime scenes, *AFTE Journal* **26**(3), 170.

French, M.L. (1997). Scene reconstruction using ballistic alignment laser, *AFTE Journal* **29**(3), Summer, 372.

French, M.L. (1997). Impact angle determination through plastic window, *AFTE Journal* **29**(1), Winter, 73.

Garrison, D.H. (1993). Shooting reconstruction Vs shooting re-enactment, *AFTE Journal* **25**(2), 125.

Garrison, D.H. (1993). Intent in shooting scene reconstructions, *AFTE Journal* **25**(4), 294.

Garrison, D.H. (1995). Reconstructing bullet paths with fixed & unfixed intermediate targets, *AFTE Journal* **27**(1), 45.

Garrison, D.H. (1995). Examining auto body penetration in the reconstruction of vehicle shootings, *AFTE Journal* **27**(3), 209.

Garrison, D.H. (1996). The effective use of bullet hole probes in crime scene reconstruction, *AFTE Journal* **28**(1), 57.

Haag, L.C. (2006). *Shooting Incident Reconstruction*, Butterworth Heineman.

Moran, B. (2001). Reconstruction of a double homicide, *AFTE Journal* **33**(2), Spring, 135.

Roberts, J. (1985). Reconstruction of a shooting to prove or disprove trajectory, *AFTE Journal* **17**(2), 53.

Salziger, B. (1999). Shots fired at a motor vehicle in motion, *AFTE Journal* **31**(3), Summer, 324.

Bullet Trajectory

Bunch, S.S. (1998). Some proposals for standardizing trajectory analysis & reporting, *AFTE Journal* **30**(3), Summer, 482.

Harrison, G.H. (1998). Road structures as it affects bullet path angles in vehicle shooting, *AFTE Journal* **30**(1), Winter, 89.
Kley, E.P. & Rowe, W.F. (1988). Trajectories of 00 buckshot, *AFTE Journal* **20**(4), 404.
Lattig, K.N. (1991). The determination on the point of origin of shots fired into a moving vehicle, *AFTE Journal* **23**(1), 524.
Nennstiel, R. (1985). Accuracy in determining long range firing position of gunman, *AFTE Journal* **17**(1), 47.
Nennstiel, R. (1991). Determination of the line of sight angle through firing experiments, *AFTE Journal* **23**(4), 919.
Trahin, J.L. (1987). Bullet trajectory analysis, *AFTE Journal* **19**(2), 124.
Van Arsdale, M. (1998). Determining bullet trajectory from a ricochet off windscreen glass, *AFTE Journal* **30**(2), Spring, 309.
Warren, G. (1991). Simple measurements of angles of elevation, *AFTE Journal* **23**(3), 869.
Zeldes, L., Lindberg, J. (1981). Laser use, *AFTE Journal* **13**(4), 21.

Bullet Ricochet/Deflection

Haag, L.C. (1987). The measurement of bullet deflection by intervening objects & the study of bullet behavior after impact, *AFTE Journal* **19**(4), 382.
Haag, L.C. (1989). Bullet ricochet: an empirical study & a device for measuring ricochet angle, *AFTE Journal* **21**(2), 182.
Nennstiel, R. (1984). Study of bullet ricochet on a water surface, *AFTE Journal* **16**(3), 88.
Rathman, G.A. (1987). Bullet ricochet & associated phenomena, *AFTE Journal* **19**(4), 374.

Related Articles

Crime Scene Investigation

DNA

Firearm Discharge Residue: Analysis of

Firearm Examination: Ballistics

Firearms: Bullet and Cartridge Case Identification

Firearms: Identification of Handling of Firearms/Trace Metal Detection

Firearms: Overview

Friction Ridge Skin: Fingerprint Detection and Recovery Techniques

Friction Ridge Skin: Interaction between Fingerprint Detection and DNA/Biological Material

Glass

Gunshot Wounds

NICHOLAS R. MAIDEN

Firearms: Shooting Distance *see* Shooting Distance: Estimation of

Firearms: Wounds Inflicted by *see* Gunshot Wounds

Firesetting

Introduction

> *If all that changes slowly may be explained by life, all that changes quickly is explained by fire. Fire is the ultra-living element. It is intimate and it is universal. It lives in our heart. It lives in the sky. It rises from the depths of the substance and offers itself with the warmth of love. Or it can go back down into the substance and hide there, latent and pent-up, like hate and vengeance* [1] p. 7.

In the setting of legal proceedings, firesetting is included under the crime of arson and may be present in civil suits regarding fire damage (*see* **Fire: Scene Investigation**; **Arson Investigation: Misconceptions and Mythology**; **Fire Modeling and Its Application in Fire Investigation**; **Fire Debris: Laboratory Analysis of**) and self-incineration [2] as well. From a psychological perspective, however, fire captures a wider variety of meanings. A source of warmth and light, fire may provide satisfaction and comfort, a sense of community, and awaken feelings of desire or passion, even as it contains destructive power. Fire is thus a dialectical image [1], conveying opposites, contrasts, and contradictions.

The myth of Prometheus conveys the notion that human mastery of fire was a transgression. Freud interpreted the myth to indicate that "the acquisition of fire was a crime" [3] p. 188, because, through mastery of the instinctual impulse to quench the fire, humankind became more powerful than the gods, entering a process of civilization through the renunciation of instinct. Anthropologists who have attempted to study the acquisition of fire from an empirical perspective have been unable to determine the precise timing of this development [4], nor have they reached consensus about the nature of the initial uses of fire [5]. The transition from a passive use of fire available in nature to the active control of fire is likely to have required mental as well as technological development, involving "self-restraint and reduction of spontaneity" [4] p. 521. Some anthropologists have observed that this restraint resulted in effects as consequential as social organization and the development of rules as well as language. Although the acquisition of mastery over fire was an historical attainment for society, each individual must go through his own process of developing knowledge about the uses and dangers of fire [6] p. 194–203.

The setting of fires for personal gain has been with us for millennia. In contemporary society, firesetting for profit, arson, is frequent and underprosecuted for want of identifiable perpetrators, and it can be deadly in result without any such intent. Arson fires can be set for insurance fraud, welfare fraud, bankruptcy scams, burglary, building stripping, property improvement, business modifications, as well as for employment [7]. That said, we turn our attention to the more perplexing domain of firesetting behavior in which financial gain is not the primary motive.

Childhood Discovery of Fire

In contemporary Western society, many children play with fire at least once in the process of the acquisition of knowledge about societal rules [8]. Risk factors for firesetting in childhood include covert antisocial behavior [9], parental stress, conduct problems, and hyperactivity, and in girls, anxiety or depression [10]. In adolescents, firesetting has been associated with shyness, aggressiveness [11], severe antisocial behavior, severe drug use, and risk-taking behavior [12]. Across age groups, heightened curiosity about fire has been associated with an increased risk for firesetting, accounting for the severity and persistence of firesetting behavior above and beyond the presence of antisocial conduct [13].

Younger children may find fire fascinating and may not have the cognitive capacity to appreciate its dangers. This susceptibility to experiment with fire has been a major focus of prevention, due to the potentially serious consequences. Public education programs have been implemented for the prevention of childhood fire play. In the United States, a common intervention following an initial firesetting incident is a home visit by a firefighter, who educates the child and his family about fire safety. Although this intervention has been shown to be effective in reducing firesetting behavior, more intensive treatments such as fire safety education programs in a clinical setting and psychotherapy have shown more promising results [14].

Studies of psychotherapy for firesetting behavior have principally involved the use of cognitive–behavioral techniques. However, these treatments have been brief, and it is as yet unknown what components of firesetting are most strongly modified by the cognitive–behavioral approach and whether the treatment effect would be augmented by a longer duration of therapy [14]. Owing to the observed importance of the intensity of fire interest in guiding children's firesetting behavior, further development of treatment modalities and techniques is needed. Intense fire interest and repeated firesetting can be associated with an underlying psychiatric disorder that requires treatment.

A history of firesetting in childhood has been associated with future violence, including cases of serial homicide (*see* **Serial Homicide**, [15]). Treatment and prevention of future firesetting behavior in childhood and adolescence may be concomitant with the resolution of psychopathology that would otherwise evolve with greater severity into adulthood. This can be stated no more than as a hypothesis currently, as there is virtually no data on the relationship between childhood firesetting and adults who become repeat firesetters.

Pyromania

The diagnosis of pyromania originated in the nineteenth century to explain some acts of repeated arson [7]. In current classifications, notably the Diagnostic and Statistical Manual, Fourth Edition, Text Revision (DSM IV-TR) (*see* **Psychopathology:**

Terms and Trends), pyromania is defined according to the following criteria:

(A) Deliberate and purposeful fire setting on more than one occasion.
(B) Tension or affective arousal before the act.
(C) Fascination with, interest in, curiosity about, or attraction to fire and its situational contexts (e.g. paraphernalia, uses, consequences).
(D) Pleasure, gratification, or relief when setting fires, or when witnessing or participating in their aftermath.
(E) The fire setting is not done for monetary gain, as an expression of socio-political ideology, to conceal criminal activity, to express anger or vengeance, to improve one's living circumstances, in response to a delusion or hallucination, or as a result of impaired judgment (e.g. in dementia, Mental Retardation, Substance Intoxication).
(F) The fire setting is not better accounted for by Conduct Disorder, a Manic Episode, or Antisocial Personality Disorder [16].

These stringent criteria define pyromania very narrowly as a specific disorder of impulse control and thought, which alone cannot account for the wide variety of mental conditions that are found in association with arson or other firesetting behaviors.

Although rare, pyromania may cause severe impairment and mental suffering [7]. In a small study of 21 patients recruited from research treatment settings for impulse control disorders, individuals who met criteria for pyromania at some point in their lives had high rates of mood disorders, anxiety disorders, substance-use disorders, and other impulse-control disorders [17]. Of note, the authors found that only two of the 21 subjects had been arrested for firesetting, and that 12 of 21 reported that they set "controlled" fires that did not rise to the level of criminal charges. On this basis, the authors state that the prevalence of pyromania may possibly be greater than is stated in the current literature.

Studies generally indicate that the prevalence of pyromania is low in offenders convicted of arson. A review of medical records and forensic mental health examinations in 90 arson recidivists over a 20-year period in Finland revealed only three defendants meeting DSM-IV-TR criteria for pyromania [18]. Another study of a similar population of defendants referred for psychiatric evaluation found 23% of arsonists with prior nonviolent offenses had a diagnosis of pyromania, while 12% of arsonists with a history of violent offenses had the diagnosis. However, this sample included approximately 10% of arson offenders, those referred for psychiatric evaluation [19]. A study using a variety of sources, from the FBI Behavioral Sciences Unit, from a forensic hospital, and from military disciplinary barracks found the rate of pyromania to be 1.3% [20].

While heuristically useful as a diagnostic category that explains a small subset of firesetting behavior, pyromania should be understood as a highly disabling disorder that is nonetheless uncommon and does not account for the vast majority of cases of firesetting.

Psychopathology

The exclusion criteria noted above in the DSM-IV-TR definition of pyromania indicate numerous other mental pathologies and psychological motivations with which firesetting behavior may be associated. Firesetting occurs in the context of a variety of other disorders. In this sense, the choice of fire is opportunistic in that fire is not the focus of the individual's mental illness, as in pyromania, but is rather the means of choice for self-expression under particular circumstances, or when the person has worsening symptoms of another psychiatric disorder.

Many of the research surveys on arsonists have involved small samples, or samples that are selected from specialized populations, such as defendants who have been referred for psychiatric evaluation in the context of criminal proceedings, and have been retrospective reviews of data collected for other purposes. These studies have consistently demonstrated that over half of the arsonists evaluated were intoxicated with alcohol at the time of the crime [18, 20, 21] and that a substantial number of arsonists suffered from an affective [20] or psychotic disorder [18, 20, 21]. Personality disorders were also common [18, 20, 22], as were other substance-use disorders [20, 23]. Some studies have suggested a high level of mental retardation in populations of arsonists [18, 22].

Sociodemographic and situational features have been associated with firesetting behavior. A majority of arson offenders were living alone or unmarried at the time of the crime [20, 24], many were unemployed, and among women, nearly half had a history of sexual abuse [24]. These features cannot be considered elements of psychopathology *per se*, but are rather factors that may contribute to the impact of a mental disorder.

Communication

The concept of *communicative arson* was introduced by Geller in 1992 [25], and arose from the observation that, in the context of psychiatric deinstitutionalization, the firesetting behavior of many mentally ill patients could not be explained by the symptoms of their disorders nor by pyromania [25, 26]. Communicative arson refers to acts of firesetting that occur in the context of the individual's decreased ability to convey his wishes and needs to others, due to a lack of adequate social skills. Treatment of individuals who set fires to convey anger, dissatisfaction with housing, or a need for care has focused on their social-skills deficits and on assertion training [27, 28]. In their discussion of the existing theoretical models for adolescent firesetting behavior, Glancy *et al.* remark that

> Libidinal explanations of psychodynamic theorists have given way to alternative paradigms of firesetting behaviour. Social learning models hold that juvenile firesetting is the manifestation of interpersonal failures that lead to the deviant expression of aggression and control ... it is theorized that firesetting can be linked to modelling of aggression and inadequate social skills [8] p. 53.

Observations of mentally ill patients who set fires indicate that the same applies to adult firesetters [27].

Although arson is classified as a crime against property, the potential destructiveness of uncontrolled fire as a threat to individuals as well as to communities would suggest that it is a form of violence that may arise from interpersonal conflicts, perceived threat, or aggressive impulses. The most common motive for arson is revenge [20, 22], which is aggressively targeted to a specified victim. Although revenge can be classified as a motive unrelated to mental disorder [7], in one study of defendants interviewed at an outpatient forensic evaluation clinic, "The common finding among all diagnostic groups was the high frequency of vindictiveness as the principal motivation behind the fire" [21]. On the other hand, when psychiatric patients are studied, arson correlates with histories of suicide attempts and not with past homicidal behaviors [29, 30].

As noted throughout this article, the choice to set a fire can be opportunistic or impulsive, based on emotional states that would otherwise manifest under different circumstances as other forms of destructive behavior, including interpersonal violence. On the other hand, the choice of fire can also be perceived by a person with poor social and verbal skills as an effective way to communicate. Firesetting requires no face-to-face encounter with another person, thus avoiding all eye contact; the rate of success in conveying a message is higher than most past efforts to influence outcome may have been for the individual; and far more often than not, someone must respond and intervene as a result of the fire – thus the person with poor social skills who is often ignored is not ignored this time, when he communicates by way of fire.

Conclusion

Firesetting is a destructive act with high potential lethality and a wide range of motivations. The controlled use of fire is essential for human society to exist, and individual learning of the safe and appropriate use of fire is an attainment during childhood that requires attention and intervention. Adult firesetting is often associated with a variety of psychopathological conditions and is not limited to persons with a diagnosis of pyromania. There is as yet little consensus on a comprehensive classification of firesetting behaviors nor is there consensus on what the categories or variables should be in order to develop a valid and reliable nosology for these behaviors. In contrast with the field of interpersonal violence, systematic, prospectively designed studies of risk factors for future firesetting in adults are lacking, and existing research on treatment modalities for juvenile and adult firesetting has yielded limited results at this time. The challenges posed by firesetting behavior have yet to be met.

References

[1] Bachelard, G. (1964). *The Psychoanalysis of Fire*, Routledge & Kegan Paul, London.

[2] Geller, J.L. (1997). Self-incineration. A review of the psychopathology of setting oneself afire, *International Journal of Law and Psychiatry* **20**, 355–372.

[3] Freud, S. (1953). *The standard edition of the complete psychological works of Sigmund Freud*, J. Strachey, ed, The Acquisition and Control of Fire (1932), in vol. 22, *New Introductory Lectures on Psycho-Analysis and Other Works*, Hogarth Press and the Institute of Psycho-Analysis, London.

[4] Goudsblom, J. (1986). The human monopoly on the use of fire: its origins and conditions, *Human Evolution* **1**, 517–523.

[5] Wrangham, R.W., Jones, J.H., Laden, G., Pilbeam, D. & Conklin-Brittain, N.L. (1999). The raw and the stolen: cooking and the ecology of human origins, *Current Anthropology* **40**, 567–594.
[6] Goudsblom, J. (1992). *Fire and Civilization*, Penguin Books, New York.
[7] Geller, J.L. (1992). Arson in review. From profit to pathology, *The Psychiatric Clinics of North America* **15**, 623–645.
[8] Glancy, G.D., Spiers, E.M., Pitt, S.E. & Dvoskin, J.A. (2003). Commentary: models and correlates of firesetting behavior, *The Journal of the American Academy of Psychiatry and the Law* **31**, 53–57.
[9] Kolko, D.J., Day, B.T., Bridge, J.A. & Kazdin, A.E. (2001). Two-year prediction of children's firesetting in clinically referred and nonreferred samples, *Journal of Child Psychology and Psychiatry, and Allied Disciplines* **42**, 371–380.
[10] Dadds, M.R. & Fraser, J.A. (2006). Fire interest, fire setting and psychopathology in Australian children: a normative study, *The Australian and New Zealand Journal of Psychiatry* **40**, 581–586.
[11] Chen, Y.H., Arria, A.M. & Anthony, J.C. (2003). Firesetting in adolescence and being aggressive, shy, and rejected by peers: new epidemiologic evidence from a national sample survey, *The Journal of the American Academy of Psychiatry and the Law* **31**, 44–52.
[12] Martin, G., Bergen, H.A., Richardson, A.S., Roeger, L. & Allison, S. (2004). Correlates of firesetting in a community sample of young adolescents, *The Australian and New Zealand Journal of Psychiatry* **38**, 148–154.
[13] MacKay, S., Henderson, J., Del Bove, G., Marton, P., Warling, D. & Root, C. (2006). Fire interest and antisociality as risk factors in the severity and persistence of juvenile firesetting, *Journal of the American Academy of Child and Adolescent Psychiatry* **45**, 1077–1084.
[14] Kolko, D.J. (2001). Efficacy of cognitive-behavioral treatment and fire safety education for children who set fires: initial and follow-up outcomes, *Journal of Child Psychology and Psychiatry, and Allied Disciplines* **42**, 359–369.
[15] Singer, S.D. & Hensley, C. (2004). Applying social learning theory to childhood and adolescent firesetting: can it lead to serial murder? *International Journal of Offender Therapy and Comparative Criminology* **48**, 461–476.
[16] American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition, Text Revision, American Psychiatric Association, Washington, DC.
[17] Grant, J.E. & Won Kim, S. (2007). Clinical characteristics and psychiatric comorbidity of pyromania, *Journal of Clinical Psychiatry* **68**, 1717–1722.
[18] Lindberg, N., Holi, M.M., Tani, P. & Virkkunen, M. (2005). Looking for pyromania: characteristics of a consecutive sample of Finnish male criminals with histories of recidivist fire-setting between 1973 and 1993, *BMC Psychiatry* **5**, 47.
[19] Repo, E., Virkkunen, M., Rawlings, R. & Linnoila, M. (1997). Criminal and psychiatric histories of Finnish arsonists, *Acta Psychiatrica Scandinavica* **95**, 318–323.
[20] Ritchie, E.C. & Huff, T.G. (1999). Psychiatric aspects of arsonists, *Journal of Forensic Sciences* **44**, 733–740.
[21] Leong, G.B. & Silva, J.A. (1999). Revisiting arson from an outpatient forensic perspective, *Journal of Forensic Sciences* **44**, 558–563.
[22] Rix, K.J. (1994). A psychiatric study of adult arsonists, *Medicine, Science, and the Law* **34**, 21–34.
[23] Jayaraman, A. & Frazer, J. (2006). Arson: a growing inferno, *Medicine, Science, and the Law* **46**, 295–300.
[24] Puri, B.K., Baxter, R. & Cordess, C.C. (1995). Characteristics of fire-setters. A study and proposed multiaxial psychiatric classification, *The British Journal of Psychiatry* **166**, 393–396.
[25] Geller, J.L. (1992). Communicative arson, *Hospital & Community Psychiatry* **43**, 76–77.
[26] Geller, J.L., Erlen, J. & Pinkus, R.L. (1986). A historical appraisal of America's experience with "pyromania"–a diagnosis in search of a disorder, *International Journal of Law and Psychiatry* **9**, 201–229.
[27] Geller, J.L. (1987). Firesetting in the adult psychiatric population, *Hospital & Community Psychiatry* **38**, 501–506.
[28] Harris, G.T. & Rice, M.E. (1984). Mentally disordered firesetters: psychodynamic versus empirical approaches, *International Journal of Law and Psychiatry* **7**, 19–34.
[29] Geller, J.L. & Bertsch, G. (1985). Firesetting behavior in the histories of a state hospital population, *Hospital & Community Psychiatry* **36**, 1056–1062.
[30] Geller, J.L., Fisher, W.H. & Moynihan, K. (1992). Adult lifetime prevalence of firesetting behaviors in a state hospital population, *The Psychiatric Quarterly* **63**, 129–142.

Further Reading

Barnett, W. & Spitzer, M. (1994). Pathological fire-setting 1951–1991: a review, *Medicine, Science and the Law* **34**(1), 4–20.
Kolko, D. (ed) (2002). *Handbook on Firesetting in Children and Youth*, Academic Press/Elsevier, San Diego.
Lewis, N.D.C. & Yarnell, H. (1951). *Pathological Firesetting (Pyromania)*, Nervous and Mental Disease Monographs, New York .

SUZANNE YANG AND JEFFREY L. GELLER

Flies *see* Entomology

Folie a Deux *see* Temporary Insanity

Folkrights ***see*** Behavioral Science Evidence

Footwear and Foot Impressions: Comparison and Identification

Introduction

This article can be used as guidelines for comparing shoemarks and drawing conclusions. It is a major update of the "Guideline for drawing conclusions regarding shoeprint examinations" as presented in Stockholm, Sweden 1999.

The new guidelines are intended for shoemark and tire track experts at the Netherlands Forensic Institute (NFI), scene of crime officers, and for anyone working in the field of comparative shoemark examination.

Throughout the development of these guidelines, the most important goal was thoroughness. Considering the variety and complexity of shoemark examination, there will always be a need for amplification and detailing.

Fifty years of knowledge and experience of the NFI was drawn on for the composition of this article, and long experience in presenting cases to the judiciary was also a factor in its composition.

Objective of the Guidelines

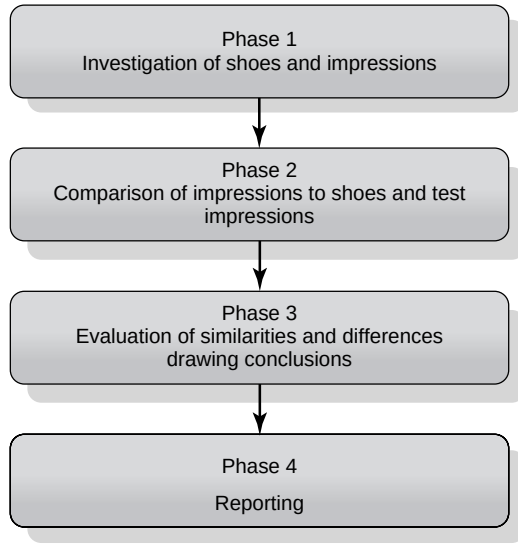
The guidelines exclusively relate to drawing conclusions in comparative shoemark examination. They lay out a method for objective examination, which takes into account not only the similarities between a shoemark and a shoe but also any differences. Whether there are similarities or discrepancies in a comparison, the document gives universal guidelines on how to reach a conclusion. When similarities are found, some steps must be taken before reaching a conclusion, and these steps are diligently covered in the guideline.

Observation. If there is reason to deviate from these guidelines, the reader should keep in mind that this article only indicates the direction in drawing conclusions. However, any deviation from the guidelines should clearly describe why the deviation was made, so that another examiner can verify this point of view.

For the essence of the guideline, we refer to section: "Evaluation of similarities/differences, and drawing conclusions."

Short Description of the Examination Phases

A comparative shoemark examination consists of the following phases:



Phase 1. In phase 1, the shoes and impressions are examined for class characteristics (dimensions, form, and pattern) and accidental characteristics, and described.

Phase 2. Examination is done to determine whether the class characteristics, presented by the impressions, correspond or differ from those of the shoes:

- if there are no correspondences, but some differences, then a negative conclusion must be drawn.

In case there are differences in dimensions and shape, it is investigated if this could be the result of how the shoemark was produced, for example by running, jumping, or slipping.

If the class characteristics correspond, the next step is an investigation into acquired characteristics; i.e., whether

- the degree and pattern of wear and tear is similar or different;
- any damage is visible in the sole of the shoe that may have caused the irregularities in the shoemark.

Phase 3. In phase 3, the similarities encountered are analyzed.

Ascertained is

- the characteristic value of the similarities;
- the extent of similarity in shape between the damage of the shoe and the irregularities in the impression.

If any characteristics in the shoes are encountered that are not visible in the impression or characteristics in the impression are encountered that are not visible in the shoe, an explanation is sought.

A conclusion is drawn after the analysis of all similarities and/or differences. That conclusion depends on the extent of similarity or difference in

- class characteristics;
- wear and tear phenomena; and
- accidental characteristics.

If the type of shoe is not very common, an investigation can be performed to quantity the value of these findings. The findings of this investigation can be included in the examination as additional information.

If insufficient similarities or differences are encountered to state with certainty whether the impression *was caused* or *not caused* by the shoe, the conclusion must be expressed less forcefully.

In the Netherlands, the terms used to express a conclusion are

Affirmative:	*was caused;* *was very probably;* *probably;* *possibly;*
The question remains open:	*inconclusive;*
Negative:	*not plausible;* *probably not;* *not.*

Phase 4. In phase 4, a report about the comparative shoemark examination is produced in accordance to the laboratory procedures and national standards when (available).

Basic Assumptions of the Guideline

The following basic assumptions apply to the use of the guideline:

1. The similarities should, in principle, be visible to the naked eye. An enhancement tool, such as a magnifying glass, is generally only used to assess if the characteristics in the shoe were created during manufacturing or through usage of the shoe (for example, features caused by air trapped in the production process or damage due to use).
   If case magnification is used, attention must be paid to the fact that similarities smaller than 1.5 mm must be clearly distinguishable in size from the background noise of the impression. Should the dimensions of an irregularity be of the same magnitude as the noise, the characteristics that seem to be similar will not be considered as corresponding.
2. Background noise can be created by
   (a) the structure of the background on which the impression is placed (carrier);
   (b) the size of the particles of the material with which or in which the impression is made;
   (c) the surface structure of the shoe;
   (d) disturbance of the impression; and
   (e) retrieving or storing the footwear impression evidence incorrectly.
3. The guidelines apply to generally occurring wear and tear of shoes (for example, worn patches in the middle on the front of the sole and on the back of the heel).
   Similarities that are related to the following are not included in the guideline:
   (a) patterns of wear and tear that point to an unusual gait or to an abnormality in the feet of the person wearing the shoes and
   (b) damage to the shoes created in the manufacturing process.
   When such similarities are encountered, the values of these are ascertained separately and included in the final conclusion. For evaluation of this, more detailed investigation is

often necessary, focused on the statistic podiatric aspects and the manufacturing process in question, respectively.

Evaluating Similarities/Differences and Drawing Conclusions

In comparative shoemark examination, three situations can be distinguished:

- corresponding class *and* accidental characteristics;
- corresponding class characteristics, *but no* corresponding accidental characteristics; and
- class characteristics that are different.

It is possible for a shoemark to differ from a test print. This can occur in the following ways:

- through the way the impression was made – due to the height, weight, and the actions of the person while making the impression;
- through disturbance of the impression;
- because of damage caused to the shoe in the time between the shoemark being made and the shoe being seized;
- because of wear and tear obliterating earlier damage to the shoe in the time between the shoemark being made and the shoe being seized and
- the degree of wear or tear to the pattern may have increased, in the time between the shoemark being made and the shoe being seized.

If the differences between a shoemark and a test print cannot be explained by the above, they are indications that the shoemark was not made by the shoe.

Corresponding Class and Accidental Characteristics

If the class characteristics of the shoe correspond with the class characteristics of the impression, it is investigated if there is damage in the shoe that corresponds to irregularities in the impression. If there are damages, the following is ascertained about these similarities:

- **The number**

The number of accidental characteristics corresponding with irregularities in the impression (or, expressed less strongly, which shows similarities to irregularities in the impression).

The similarities encountered must be indicated and labeled in work forms.

- **The characteristic value**

After the indicated similarities have been given a label, the characteristic value of each similarity is determined by subdividing these into one of the six defined classes. These classes are listed in Supplement 1. The classification of a similarity depends on the size and complexity of the shape. This classification and the associated value are based on experience at the NFI.

- **The extent of similarity in shape**

The extent of similarity in shape of each labeled corresponding irregularity with the particular accidental characteristic is ascertained. There are two categories into which the extent of similarity in shape can be subdivided:

(a) location, size, and shape correspond and
(b) location and size correspond and the shape corresponds roughly.

If the location or the size is different, it is investigated whether this can be explained by the way the impression was made (standing still, walking, running, jumping, or sliding). If this is not the case, this irregularity and damage may not be considered to correspond.

Results of Characteristic Similarities. The results of the characteristic value determination are entered in the result table (Supplement 2). The characteristic value and the extent of similarity in shape are indicated for every similarity.

The results indicated in the result table are plotted in two conclusion graphs (I and II), which are based on experience. The graphs are presented in Supplement 2.

Graph I is for irregularities that correspond with accidental characteristics in location, size, and shape.

Graph II is for irregularities that correspond with accidental characteristics in location, size, and *roughly* in shape.

In the graphs, the number of similarities with the same characteristic value is plotted horizontally. The corresponding logarithmic value of the score "S" can be read on the vertical axis for the characteristic

value line in question. This is repeated for all similarities with an equal characteristic value. When the number of similarities with the same characteristic value increases, the "S" score will thus increase logarithmically.

Results of Corresponding Wear and Tear. The value of a possibly corresponding pattern of wear and tear is then added to the log(S) value obtained.

The following assumptions are applicable in this regard:

1. If the (general) pattern of wear and tear to the shoe is not present or minor, and this corresponds with what the impression shows, *nothing* is added to the log(S) value.
2. If the profile components of the shoe demonstrate clear occurrences of wear and tear (partially worn down), the log(S) value is increased *by 1, at most*, depending on the quantity and the extent of similarity with the impression.
3. If the wear and tear has advanced to such an extent that various profile components have taken on characteristic shapes and/or areas between the profile components that demonstrate wear and tear with clear characteristic shapes, the log(S) value can be increased by *3, at most*, depending on the extent of the similarity.

If the wear and tear indicates an abnormality of the foot of the person wearing the shoes, the value of this can be determined by a more detailed podiatric examination.

Conclusion. The separately obtained log(S) values are added together on the conclusion line, next to graphs I and II. The corresponding verbal conclusion is stated next to this line.

Depending on the number total of log(S) (Σ log(S), which represents the number of similarities, their characteristic value, the extent of similarity in shape and the possible corresponding wear and tear, a conclusion can be drawn that

- the impression was *possibly* made by the shoe ($0.75 \leq \Sigma \log(S) < 3.5$);
- the impression was *probably* made by the shoe ($3.5 \leq \Sigma \log(S) < 6.5$);
- the impression was *very probably* made by the shoe ($6.5 \leq \Sigma \log(S) < 8.0$); and
- the impression was *made* by the shoe ($\Sigma \log(S) \geq 8.0$).

Supplement 3 shows an example with sample data for filling in the result table and the graphs.

Note: Not all laboratories use this system, and such a system should not be seen as the panacea. However, the experience in our laboratory shows that it combines pragmatism and scientific rigor. More information about the value of shoemarks can be found in the Statistics section.

Corresponding Class Characteristics, but no Corresponding Accidental Characteristics

When corresponding class characteristics are encountered and accidental characteristics that correspond or differ are absent, the question remains open whether the impression was made or not made by the shoe.

Conclusion. Considering the corresponding class characteristics, the shoe qualifies for having made the impression.

Owing to the absence of corresponding accidental characteristics, it *remains inconclusive* whether the impression was made by *this* shoe or not.

Class Characteristics that Differ

When class characteristics that differ are encountered (most importantly differences in dimensions), it is investigated if these can be explained, for example by how the impression is made. This is backed up by making test impressions. If the differences in dimensions cannot be explained by the way the impression is made, if there are differences in shape or pattern, or if the impression shows more wear and tear than the shoe does, a conclusion of exclusion must be drawn.

Conclusions. In these cases, depending on how clear the differences are, one of the following conclusions is drawn:

- it is *not considered plausible* that the impression was made by the shoe;
- the impression was *probably not* made by the shoe; and
- the impression was *not made* by the shoe.

Different Degree of Wear and Tear

If the degree of wear and tear of the shoe does not correspond with the degree of wear and tear presented in the impression, it must be possible to explain why; for example by considering the period between the time the impression was made at the crime scene and the time the shoe was impounded. Should this not be the case, a negative conclusion must be drawn.

Conclusions. In this case, depending on how clear the difference is, one of the following conclusions is drawn:

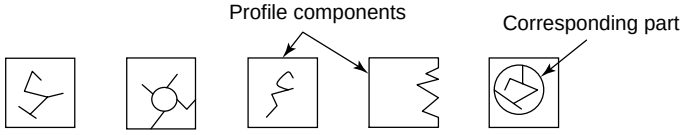
- it is *not considered plausible* that the impression was made by the shoe;
- the impression was *probably not* made by the shoe; and
- the impression was *not made* by the shoe.

SUPPLEMENT 1 Evaluation of similarities

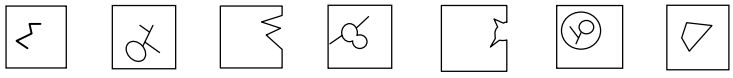
The value of a similarity between a particular accidental characteristic in the shoe and an irregularity in the impression depends on the number of corresponding components and the overall size of the similarity. The minimum acceptable size of a component is 1.5 mm. See also the explanatory notes on the next page. The similarity is subdivided into one of the following classes:

Similarity with "High Characteristic Value"

H1. Similarity consisting of six components, at least. One such similarity is necessary to ascertain that the impression was made by the shoe.

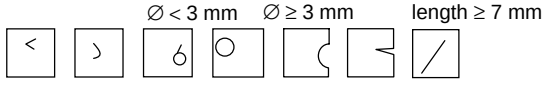


H2. Similarity consisting of four components, at least. Two such similarities are necessary to ascertain that the impression was made by the shoe.

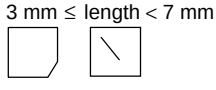


Similarity with "Medium Characteristic Value"

M1. Similarity consisting of two components, at least, or a line with a minimum length of 7 mm.
Note: a circle with a minimum diameter of 3 mm is considered as consisting of two components.
Three such similarities are necessary to ascertain that the impression was made by the shoe.



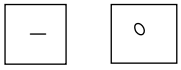
M2. Similarity consisting of one component with a size of 3–7 mm.
Six such similarities are necessary to ascertain that the impression was made by the shoe.



Similarity with "Low Characteristic Value"

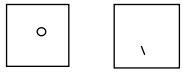
L1. Similarity consisting of one component, a line or a round shape with a size up to 3 mm.
Eight such similarities are necessary to ascertain that the impression was made by the shoe.

1.5 mm ≤ size < 3 mm



L2. Similarity consisting of a line or a round shape with a size less than 1.5 mm.
Ten such similarities are necessary to ascertain that the impression was made by the shoe.
(the lower limit "x" depends on the magnitude of the background noise in the impression. It must be possible to distinguish the irregularity in the impression, as being a result of the accidental characteristic in the shoe, from the background noise).

x < size < 1.5 mm



Explanatory Notes

- **Minimum size of a component**

For the classes L1 and higher, components must have a minimum size of 1.5 mm.

- **Two- and three-dimensional impressions**

A three-dimensional footwear impression (such as a cast) *can* be compared with the shoe or with a test impression made with the shoe. The components of the similarity can extend in three dimensions.

A two-dimensional impression (e.g., a shoemark on a sheet of paper) *must* be compared with the shoe *and* the two-dimensional test impressions made with it. For an explanation, please refer to the examples in Supplement 4.

- **Too few components**

If the number of components in a similarity is too low to comply with the requirements of a class, a lower class must be selected. For example, a similarity with five components is classified in class H2.

- **Partial similarity in shape**

If an irregularity in the impression shows fewer components than the accidental characteristic in the shoe, or the other way around, attention must be paid to the fact that only the similarity is evaluated.

For example, if an accidental characteristic consists of six components and the irregularity in the impression shows four components of these, the similarity does not comply with the criterion of an H1. Only the corresponding components are counted, and the similarity is classified as an H2 (and noted as corresponding, in Graph I).

If parts of components in a similarity do not correspond entirely as far as shape and size are concerned or if they are presented vaguely, it will be classified as roughly corresponding (noted in Graph II).

- **Schallamach**

If an impression is made with or in a fine medium, very small details (smaller than 1.5 mm) in the shoe sole may be presented. Detailing of the impression may be so large that the fine details of a Schallamach pattern can be distinguished in comparison to the transfer medium.

A corresponding Schallamach pattern, with sufficient corresponding details (more than 10) may lead to identification. Should there be fewer than 10 corresponding details, an L2 value will be taken separately for each clearly corresponding detail.

SUPPLEMENT 2 Result table and conclusion graphs for shoemark examination

Result table and conclusion graphs

Characteristic value of similarities *Enter a "1" for "corresponds", or a "2" for "roughly corresponds".*

	Similarity no.:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
High	Class H1: 6 components, at least															
	Class H2: 4 components, at least															
Medium	Class M1: 2 components, at least, or line ≥ 7 mm															
	Class M2: 1 component, 3 ≤ size < 7 mm															
Low	Class L1: 1 component, 1.5 ≤ size < 3 mm															
	Class L2: 1 round or line-shaped similarity, size < 1,5 mm															

Total corresponding		Total roughly corresponding	
H1		*H1*	
H2		*H2*	
M1		*M1*	
M2		*M2*	
L1		*L1*	
L2		*L2*	

Case number:

Number of shoe:

Number of impression:

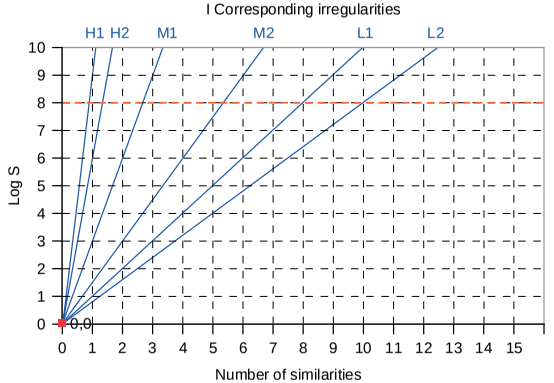
Numbers

Observation:

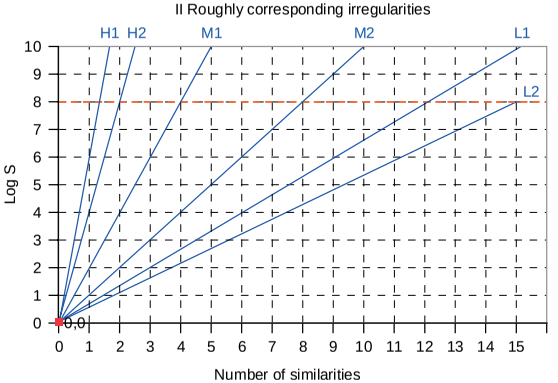
Wear and tear

No or minor wear and tear (0)	0
Corresponding characteristic wear and tear of a few profile components (0-1)	
Clearly corresponding, sharp outlined wear and tear of profile components or in-between areas (0-3)	
Corresponding wear and tear due to abnormal foot or gait (...)	
Total:	

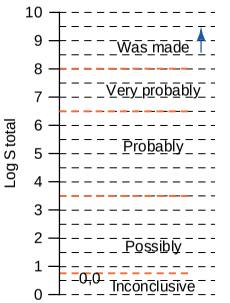
I Corresponding irregularities



II Roughly corresponding irregularities



Conclusion according to guideline



SUPPLEMENT 3 Example for the use of the result table and conclusion graphs

Result table and conclusion graphs

Enter a "1" for "corresponds", or a "2" for "roughly corresponds".

Characteristic value of similarities	Similarity no.:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
High	Class H1: 6 components, at least															
	Class H2: 4 components, at least															
Medium	Class M1: 2 components, at least, or line ≥ 7 mm															
	Class M2: 1 component, 3 ≤ size < 7 mm	1		2	1	2										
Low	Class L1: 1 component, 1.5 ≤ size < 3 mm		2													
	Class L2: 1 round or line-shaped similarity, size < 1,5 mm						1									

Total corresponding		Total roughly corresponding	
*H*1		*H*1	
*H*2		*H*2	
*M*1		*M*1	
*M*2	2	*M*2	2
*L*1		*L*1	1
*L*2	1	*L*2	

Case number:

Number of shoe:

Number of impression:

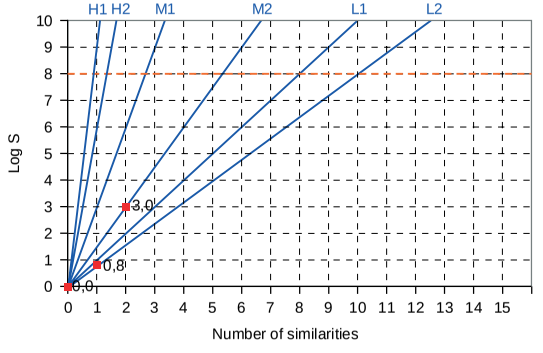
Numbers

Observation:

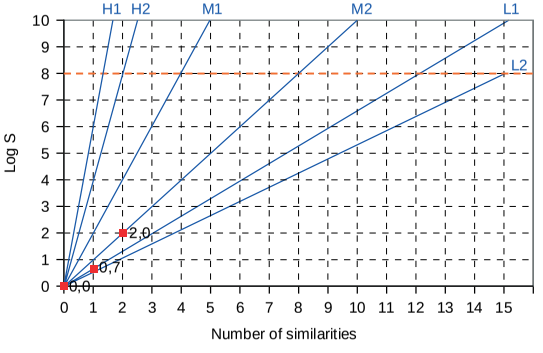
Wear and tear

No or minor wear and tear (0)	0
Corresponding characteristic wear and tear of a few profile components (0-1)	0,3
Clearly corresponding, sharp outlined wear and tear of profile components or in-between areas (0-3)	
Corresponding wear and tear due to abnormal foot or gait (...)	
Total:	0,3

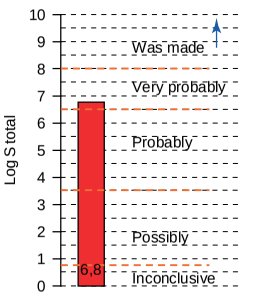
I Corresponding irregularities



II Roughly corresponding irregularities



Conclusion according to guideline



SUPPLEMENT 4 Characteristic value of similarities in two-dimensional impressions

In comparative examination of two-dimensional impressions, determine the characteristic value of the similarities by the representations of the accidental characteristics in test impressions.

Figure 1 shows two studs from a shoe with one or two line-shaped damages (accidental characteristics), respectively, each with a length of 2 mm.

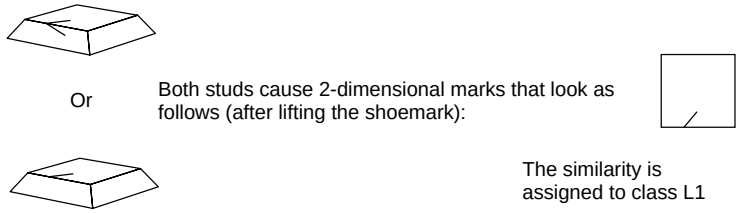


Figure 1 Two studs with one or two line-shaped damages

Figure 2 shows two studs from a shoe with, respectively, a circular damage of an approximate size of 3.5 mm, within which there are two other irregularities, each a length of at least 1.5 mm, and of a large circular damage with no further characteristics.

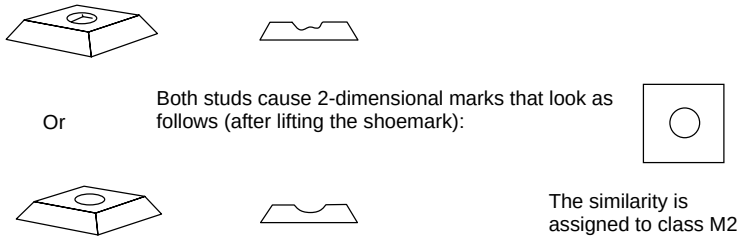


Figure 2 Two studs with two different circular damages

SUPPLEMENT 5 Definitions of terms used in this document

Accidental characteristics	See characteristics
Background noise	Disturbance of the footwear impression caused by the structure of the medium with which or in/on which the impression is made
Characteristics	Two types of characteristics are defined: Class characteristics These are features in footwear regarding the shape, pattern, and dimensions of the shoe as manufactured Accidental characteristics

	These are distinctive features of particular footwear caused primarily by general usage: wear and tear, and damage caused during use and secondarily by damage or imperfections caused during the manufacturing process
Characteristic value	A numerical value assigned to a similarity (of an irregularity in the impression with an accidental characteristic of the footwear), depending on its complexity
Class characteristics	See characteristics
Correspondence	This refers to similarities between a specific footwear impression and a specific shoe. Correspondence takes into account class and accidental characteristics.
Dimensions	The dimensions of shoe and footwear impressions are measured in millimeters The shoe sizes, used by manufacturers, are not considered accurate enough for investigative purposes Moreover, footwear impressions do not often show the size of the shoe that has made the impression
Footwear impression	The mark made by a shoe. The different types of footwear impression are Imprints in soft mediums (subsoil: sand, clay, snow, etc.) Marks standing out against hard or elastic surfaces, with or in a transferring medium. Mediums can include liquids, such as blood, mud, paint, and dust and any powdery material Note: when the term "footwear impression" is used, this refers to the original impression and also to any other acceptable reproduction of this, such as a photographic reproduction of a footwear impression on a negative or on paper, a digital photograph, a lifted footwear impression, a cast of a three-dimensional footwear impression
Form	The form is the basic outline of the sole of a shoe or a footwear impression
Impression	An impression is the mark made by a shoe as defined in "footwear impression", above
Mark	Another term for footwear impression
Pattern	The pattern refers to the ridges, lines, studs, and blocks of the tread of the sole of a shoe
Profile component	A distinctive part of a pattern (a ridge, line, stud, or block)
Shoeprint	Another term for footwear impression
Test impression	A test impression is made by the investigator for comparison with a footwear impression. A test impression is not always necessary, because when a cast of an imprint is available, the shoe can be directly compared with the cast. Test impressions are made when there is a lifted shoemark or a shoemark on a hard or elastic surface

Footwear and Foot Impressions: Databases

Introduction

Footwear impressions are recovered from all types of crime and over the past 20 years, considerable time and effort has been spent on the development of footwear databases. Their development has primarily been aimed at assisting an investigation by providing information about what possible footwear could have made the mark at the scene and provided background data to aid the evaluation of findings from a comparison [1–3].

Footwear sole databases have been developed by various forensic institutions and more recently, various commercial manufacturers have produced systems. It is the aim of this article to focus on the general principles and applications of footwear sole databases and the relationship with footwear impressions recovered from the scenes of crime.

The potential to gain useful intelligence through the use of footwear impressions is covered in article – **Footwear and Foot Impressions: Intelligence**.

The topics reviewed in this section are given as follows:

1. reference database of sole patterns from footwear
   (a) structure
   (b) coding
   (c) duplicate sole patterns
2. use of database to aid an investigation
3. use of database to aid evaluation of findings.

Reference Database of Sole Patterns from Footwear

There are several factories worldwide producing an enormous range of footwear including training shoes, casual footwear, traditional smart footwear, etc. The range of sole tread patterns as one would expect is even greater ranging from the plain leather sole on a smart dress shoe to tread patterns that are comprised of a single pattern component to footwear with tread pattern comprised of various complex geometrically shaped components.

The tread pattern on a specific brand and style of footwear may essentially be the same in appearance, but the size and spacing of the tread pattern components can vary depending on a number of factors. These include the size of the footwear, the method of manufacture, the number of moulds used by the manufacturer for a specific design, and the country of origin. Other footwear, especially training shoes, can have tread pattern components that change in appearance with wearing, an example being components that appears circular when new, can appear square when worn.

The primary function of the footwear sole database has been to provide a source of information that a footwear examiner can use, primarily to deal with two aspects:

1. To aid the investigative process by providing details of the possible footwear responsible for a mark at a crime scene.
2. To aid the examiners assessment of the evidential significance that the tread pattern makes to their final conclusion.

The sole reference database consists of a collection of sole tread patterns from items of footwear. The aim of the collection is to try and cover as many examples of footwear sole tread patterns as possible, although it is extremely difficult to deal with every available type and style of manufactured footwear. Therefore, most systems are based primarily on the most commonly encountered footwear. The collection of tread patterns means that it is possible to record features such as the manufacturer, the type and style, the size, and country of origin and a photographic record can be made about the appearance of the uppers. The database is made up of an image of the complete sole pattern. Every part of the tread pattern has to be present; otherwise, components may be missed when it is subsequently coded and may hamper any subsequent searches.

To actually record every pattern that could potentially be of interest during an investigation is extremely difficult and is generally unnecessary. This is due to the fact that the vast majority of crimes involve people wearing a limited number of footwear styles and in most instances of burglary, for example, in the United Kingdom, the footwear of choice is some form of training shoe.

The three main ways used to obtain sole tread pattern information are listed in the following section.

- **Direct from manufacturers.**

It is possible to obtain examples of new footwear patterns and relevant information by developing a direct relationship and agreement with the main manufacturers. It is difficult to have this relationship with all manufacturers, but any that can be established requires a lot of time and resources should be applied to make sure it is maintained. If it is not possible to deal directly, then another source of information can be to attend events organized by manufacturers where they present and market their goods to the trade buyers. It is normally possible on these occasions to record all required information including taking test prints and images of the uppers from the models of interest.

- **Footwear submitted to a laboratory or institute during casework.**

This refers to items of footwear that are taken from a suspect and submitted to a forensic laboratory/institute as part of the normal casework comparison process. Once it is submitted, it is possible to take test prints of the sole tread pattern and to take images of the uppers, plus details regarding the footwear such as size, country of origin, style number, and color. The footwear that will be encountered will provide examples of how certain patterns will look like during the various stages of wearing.

- **Footwear sole tread patterns taken by law enforcement agency in custody.**

Increasingly, an excellent source of information is obtained from the law enforcement agency. The police in the United Kingdom are authorized under the Police and Evidence Act (PACE) to take the footwear from a person arrested on suspicion of an offense. Various methods are available to make a record of the tread pattern on the sole of the shoe. The relevance of this source of tread patterns is that it is a record of what is actually being worn by people at that time and is more representative than the database made up entirely of tread patterns from submitted casework.

Coding of Tread Patterns

The main problem of how an examiner makes use of this potentially huge amount of information has been to develop a coding scheme of some form or another that allows different sole tread patterns to be differentiated from each other and allows small partial areas of the pattern to be identified.

The two main methods that have been developed and have obtained favor amongst examiners are provided in the following list:

1. use of generic pattern descriptors or icons and
2. use of a generic code name.

It is possible to describe a sole pattern fairly easily if the number of different components is small and their shapes fit a known descriptor. One of the problems associated with coding any footwear mark is trying to ensure consistency between the various people involved. This is important to ensure that all relevant impressions must be recovered from a search and everyone should talk in the same language. Consequently, there are normally rules associated with how the coding scheme is applied.

An example of a simple tread pattern, which can be coded easily and consistently, is the simple straight bar pattern illustrated in Figure 1.

It has therefore been possible to describe this generic pattern verbally as “a series of straight bars” or simply to use “straight”. This method has been used effectively in fairly simple databases based on Microsoft excel or a similar type software.

More sophisticated software and search algorithms have enabled the development of more complex and sophisticated software systems. By coding using icons, it is possible to code a mark from



Figure 1 This footwear has a simple pattern of evenly spaced straight bars

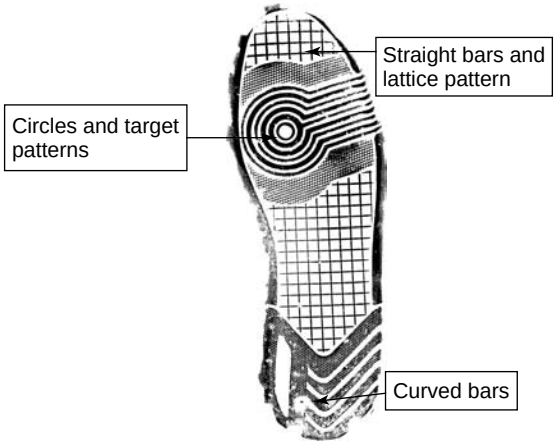


Figure 2 Complex pattern consisting of various tread components

the crime scene and search the reference collection to find any matches, search large numbers of footwear tread patterns for example, or to use icons to visually describe the pattern. By examining the various pattern components, it has been possible to build up a library of pattern descriptors and icons shapes that can be used to describe the various sole tread patterns and brand logos present on footwear.

Figure 2 shows an example of a complex tread pattern from an Adidas training shoe.

In this instance, it is possible to identify several pattern components than the straight bars in the example shown in Figure 1. It is possible to use random pattern, linked or lattice, straight, curved or wavy, circular, semicircular, oval, targets, lattice or linked pattern, complex, irregular, square, and logo. Of course, not all systems have the same descriptors or icons and it will depend on which patterns and shapes are used.

The following example shows a fairly complex pattern consisting of various components. A number can be identified in this example (Figure 2).

The use of icons allows for the use and development of computers to do all the hard work and the examiner identifies the most likely footwear contained in the database. There are rules and/or guidance to follow and to ensure consistency when more than one person is involved.

Use of a Generic Code Name

An alternative to the coding of footwear by using individual components as descriptors is to allocate a name to the tread pattern. This has an advantage in that everyone using the same coding system will know what the footwear looks like from the alpha numeric reference that has been allocated.

This is usually done in two ways. The first is to allocate the manufacturers name and a number. Usually, the number is based on something as simple as when it was first seen in the laboratory. An example would be Reebok 7; the seventh different tread pattern seen on the sole of Reebok footwear. The second alternative is to use the manufacturers name and the style name of the footwear as a reference; an example would be Reebok Classic.

In most commercially available systems, both elements will be utilized and the reference collection will have named tread patterns, with the ability to search using the various pattern components.

The reference database is therefore made up of complete sole tread patterns from footwear. It will be coded in some way and may contain more than one example of a specific pattern to show how the pattern changes in appearance with wearing. In some organizations, the reference collection is only the tip of the iceberg. The collection will additionally consist of test prints taken from all the footwear submitted. Of course, depending on the time the collection has been going and how often a particular pattern is encountered, there may be hundreds of examples of a pattern showing a whole range of sizes, moulds, and degrees of wear.

Use of Database to Aid an Investigation

At a crime scene, the perpetrator of the crime may leave footwear impressions. In some instances, these tend to be the more serious types of offense, and the law enforcement agency investigating this crime may need to address the following questions:-

- What does the complete tread pattern look like on the footwear responsible for the mark recovered at the scene?

- How manufacturer makes footwear with the identified tread pattern?
- What do the uppers look like in a footwear with the tread pattern?
- What is the size of the footwear responsible?

The information provided may give the investigators a lead or at least help when they have a suspect and are searching for clothing that they should submit for examination.

The process in this type of inquiry is for the footwear mark to be submitted to the footwear examiner who will examine and then code the tread pattern that is visible. Enhancement techniques may be used to ensure that the mark is clearly visible and that no other impressions are present. The mark will then be coded using the coding scheme most appropriate for the system. The database will be searched and "hits" will be found. These "hits" need to be considered against the submitted mark, but in many instances it will be possible to identify what type of footwear has the corresponding sole pattern, which manufacturer made them, and how the uppers will look like. With some partial impressions at a crime scene, the outcome will not be one single 'hit' but you are likely to get a list of potential "hits". This result is obtained due to the discriminating power of a coding scheme. If you make it simple, anyone will be able to use it after some training but it is likely that you will get lots of "hits", which will need further investigation. If you make a very complex coding scheme, then there is a need for more training, less people will be able to use it, it can introduce variation as people will have to make more decisions, but it is likely to reduce the number of "hits". Some common tread patterns are present on footwear produced by more than one manufacturer and the only visible difference between complete tread patterns may be the presence of a logo. This may mean that any feedback to the investigating officer is given with a "health warning".

The sizing of footwear impressions at crime scenes is normally an approximation at best and examiners will usually report a range of sizes. The underlying collection of the database can assist, as it may have a large number of different sizes and moulds of the appropriate sole pattern which can be used to estimate the size of the footwear responsible for the mark at a scene.

Use of Database to Aid Evaluation of Findings

This section does not elaborate the processes involved in doing a footwear mark comparison but provides general information regarding how an examiner uses the database to aid their interpretation [4–6]. Footwear mark comparisons are described more in detail in the article (*see* **Footwear and Foot Impressions: Comparison and Identification**).

In a footwear mark comparison to footwear submitted to an examiner for comparison, the examiner takes into consideration several aspects: pattern, size, mould, wear, and damage. In some forensic institutes/laboratories, examiners use the footwear collections that they hold to aid their evaluation of the evidence. They can estimate the commonality/frequency of the pattern being considered. It is also possible to gain useful information regarding the size, mould, and wear from the collection.

Ultimately, this information can assist the expert to express his or her opinion when a mark and a known sole cannot be differentiated.

References

[1] Birkett, J. (1989). Scientific scene linking, *Journal Forensic Science Society* **V29**(4), P271–P284.

[2] Mikkonen, S. & Astikainen, T. (1994). Database classification system for shoe sole patterns – identification of partial footwear impressions found at scenes of crime, *Journal Forensic Sciences* **V39**(5), P1227–P1236.

[3] Napier, T.-J. (2002). Scene linking using footwear mark databases, *Science and Justice* **V42**(1), P39–P43.

[4] Evett, I.-W., Lambert, J.-A. & Buckleton, J.-S. (1998). A Bayesian approach to interpreting footwear marks in forensic casework, *Science and Justice* **V38**(4), P241–P247.

[5] Taroni, F. & Buckleton, J. (2002). Likelihood ratio as a relevant and logical approach to assess the value of shoeprint evidence, *Information Bulletin Shoe Print and Tool Mark Examiners* **V8**(2), P15–P25.

[6] Champod, C., Evett, I.-W. & Jackson, G. (2004). Establishing the most appropriate database for addressing source level propositions, *Science and Justice* **V44**(3), P153–P164.

DAVID BALDWIN

Footwear and Foot Impressions: Foot Impressions and Linking Foot to Shoe

The forensic application of barefoot morphology involves the study of the physical features of feet and the impressions they make. This type of examination is used in cases when either an impression has been made at the scene of a crime or when the wearer of a shoe needs to be determined. However rare the application, in instances where it is needed to answer a question in a criminal case, barefoot impression evidence can become extremely important.

Barefoot Impressions as Physical Evidence

Barefoot impressions are a type of impression evidence that may be found at the scene of a crime. For example, if an individual involved in a homicide steps into a pool of blood with bare feet, they may track impression detail of the skin of the feet (friction skin and creases) that may be compared and identified in the same manner as fingerprints [1] (*see* **Friction Ridge Skin: Comparison and Identification**).

The shape and size features of the foot, defined as the morphology of the foot, may be examined and compared in barefoot impression evidence cases where the detail is insufficient to make a friction skin identification [2]. This type of examination is relevant when impressions have been made by a sock-clad foot, or when detail of the impression is limited due to a variety of deposition manner or surface type factors. It is also applied to the examination of the impressions found on the inner surfaces of footwear.

Barefoot impressions found on the insoles of footwear are analyzed to determine possible association or lack of association between a suspect and a particular shoe or boot. Scenarios in which this may provide important investigative information include cases in which discarded footwear has been associated to an impression from a crime scene, or in which footwear is found at a scene and may belong to either a suspect or a victim [3].

Accepted techniques for the comparison of footwear impression evidence include a side-by-side visual comparison, and the creation of an overlay tracing of the known item to be placed over the item in question to determine correspondence of features [4] (*see* **Footwear and Foot Impressions: Comparison and Identification**). The same techniques, side-by-side visual examination and overlay, are employed in the examination of barefoot impression evidence [5].

The Structure of the Foot and the Foot Inside a Shoe

A normal foot has 26 bones. Five metatarsal bones, which reach from the arch area to the metatarsal head area, are distinctly separated from one another. Phalanx bones are anterior of the metatarsal bones, forming five separate toes [6]. When a foot is placed inside a shoe, the forefoot and toes are most commonly constricted inward.

Barefoot impressions on the insoles of shoes are the result of pressure, heat, and sweat that occurs during the wearing of a shoe [3]. These impressions characteristically do not include friction ridge skin or crease detail, as socks are normally worn and the impressions result from repeated wearing of the shoe, with the foot constantly rubbing, rather than from a single touch of the foot inside the shoe. These impressions are often a clear recording of toes and the forefoot leading edge of the ball portion of the foot, referred to as the *metatarsal ridge* [2, 3, 5, 7]. In some footwear, impressions of the ball, arch, and heel areas may also be present (Figure 1).

In a barefoot comparison of shoe insole impression evidence, the recording of the features of the foot visible on the inner surface of the shoe are compared with the known exemplars taken from a person in question to include or exclude them as a source of the impressions inside the shoe. Comparisons also take place between the inside uppers of the shoes and the wear patterns as associated to the three-dimensional features of the foot which come into contact with the upper inside surface of shoes, such as toe nails, protruding toes, or toe joints [5].

As can be seen in Figure 2, an inked barefoot impression taken of a foot outside a shoe is wider in the forefoot and toe areas than is the impression of the same foot inside a shoe. A significant constriction



Figure 1 Foot impressions on shoe insoles

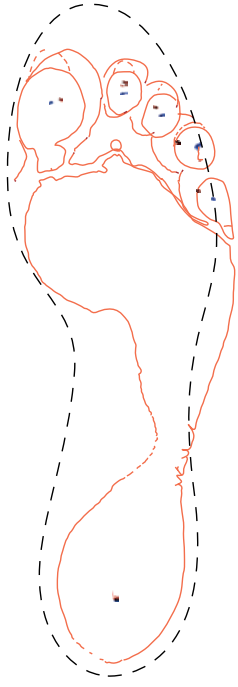


Figure 2 Tracing of a barefoot impression overlaid onto a tracing of the impression of the same foot inside a shoe

of the toes and metatarsal area is often observed when comparing a barefoot impression inside a shoe to one taken outside of a shoe. It is extremely important to take this constriction into consideration during an analysis of a barefoot impression inside a shoe.

Known Exemplars Used in Barefoot Comparison Cases

For the comparison of barefoot impression evidence, inked impressions are taken both walking and standing, both of naked and socked feet (Figure 3) [7]. Walking impressions involve the inking of the person's feet, and then having the person walk a length of paper approximately 15-feet long, down one side and back the other. This provides a series of naturally walking impressions for use in the comparison process. These successive impressions have been observed to "exhibit little or no change" from one another as they are made along the paper [8].

Known shoes of the person in question may also be seized for comparison to barefoot impressions found inside the shoes in question involved with a crime [8]. However, in practice, occasionally, only



Figure 3 Three barefoot impressions of the same right foot

inked impressions are available for comparison to the impressions of feet inside of shoes [9]. Thus, although it is generally recommended in forensic examinations to compare "like to like" exemplars, such as a shoe insole impression to another shoe insole impression, occasionally it is necessary to compare an inked bare foot impression to an impression inside a shoe.

Databases and Studies of Barefoot Impression

Some of the research in the area of foot morphology and forensic application has taken the form of the collection of large numbers of barefoot impressions from individuals and categorization of them to determine how variable the feet are.

The 1980 study of this type conducted in India collected footprints from 725 individuals. From the measurements taken, a series of indices were calculated to represent variation among the feet. These indices were combined to calculate probabilities. This appears to be one of the first applied databases using foot measurements and calculated statistics to estimate the variability of barefoot morphology in the human population [10]. Previous and subsequent similar studies have derived similar indices to one another, providing support for both the method employed and statistics demonstrating the high level of individuality of feet [11]. The five foot impressions in Figure 4 show the variability in shape of the features of the feet and toe positions between different individuals.

In 1986 and 1987, the FBI collected barefoot impressions from 500 individuals. Measurements, as described previously, were entered into a computer database and intercompared. It was noted that the size and shape features varied considerably and concluded that no left or right foot of any individual in the sample set was found to be identical. Further, results showed that only a few combined size and shape characteristics were required to discriminate a foot from this population [3].

The Royal Canadian Mounted Police (RCMP) has collected approximately 12 000 barefoot impressions for the study [12]. Two statistical analyses have been carried out on portions of these samples that have been entered into a database. The population used in the first preliminary study was 960, and for the second study it was increased to 5755 [13, 14].



Figure 4 Impressions of right feet from five different individuals

These studies concluded that measurements taken of the feet show a great degree of variability between the barefoot impressions of individuals, and a great degree of similarity between multiple impressions of feet taken of the same individual. The second study reports the statistical odds of a chance match of barefoot impressions, based solely on the positions and measurements of the characteristics of the feet recorded, with no ridges or creases considered, to be one in 1.27 billion. This number is not used in criminal cases, it is applied as an indication of the high degree of variability of barefoot morphological features for background and research purposes.

Both the footwear industry and military researchers have studied the population of human feet for the purpose of achieving comfort in footwear. Although not the goal of the studies, the results include comments which reflect upon the wide variety and differences in the structure of feet [3, 15, 16]. Although these are not to be considered forensic validation studies, they do support the concept of high degree of individuality of the morphology of the foot.

In addition to the construction and statistical analysis of databases, forensic texts that include sections on the examination and comparison of barefoot impression evidence, as well as articles specific to this discipline, have been published [2, 3, 17–19]. Through discussion, practice, peer review, and study, the development of forensic barefoot morphology impression comparison has been shaped over the last few decades. It is this process that has led to the current procedures and methodologies for the examination of this type of evidence. Both empirical and statistical studies validate the identifying nature of specific physical features of the foot, and that these features may be used to exclude an individual as the source of a barefoot impression, include them as a possible source of the impression or may lead to an inconclusive conclusion.

DNA Analysis Applied to the Insides of Shoes

It should be noted that DNA technology may be applied on samples taken from inside the shoes (*see* **DNA: Sources of**). This analysis could assist to assess if a contact occurred between an individual's feet and the shoes. However, in many circumstances, the chance of obtaining a profile may be small. Additionally, the DNA from more than one person is often found on the insides of shoes, not necessarily including the owner of the shoes [20].

Case Examples in the Comparison of Feet to Shoes

There have been many interesting cases in the past decade involving the examination and comparison of the insoles of shoes in determination of the wearer of the shoe, examples are mentioned later.

In 1948 in Canada, a shoe to foot comparison was conducted between a pair of shoes and two brothers who were suspects in a series of break and enters. One brother, Donald Kett, was convicted. The other brother, William Kett, claimed he was innocent and that all the shoes that were matched to the crime scene belonged to Donald. However, through comparison of the impressions inside the shoes, it was determined that William had worn the shoes, and he was also convicted [21].

In 1955, a burglar left his shoes at the scene of a crime in Surrey, England. Shoes were obtained from a suspect, and casts of the insides of the shoes were made and compared against one another. The shoes were found to have been worn by the same individual. This evidence was presented in court and a conviction was the result [22].

William Bodziak, a retired FBI Special Agent, describes a case in which a fleeing suspect lost his right shoe at the scene of a homicide. By comparing the sweat stains and depressions of the wearer's foot on the insole of the crime scene shoe to those found in the right shoe the suspect was wearing when apprehended, it was determined that there was a high probability that both the shoes were worn by the same person [3].

Robert Kennedy, a retired RCMP Officer, has conducted many cases involving the comparison of shoe insoles. One of his cases involved the homicide of a prostitute in Israel. A young man was the suspect, and when shoes were seized from his house, it was determined that his brothers and father also wore the same style of shoe. All the shoes were examined, and the victim's blood was found on the outside of one of the pairs of shoes. After the insole impressions of the shoes were examined, it was determined that the impression in one shoe did correspond to the suspect, but the impression in the other shoe did not. After

further examination of the insoles, it was determined that the insole that was different from the suspect's foot was not the original insole. Through a physical match of the glue residues on the bottoms of the insoles of all the shoes, the original insole was found in one of the other shoes. The impression on this insole was also found to correspond to the suspect, which linked the suspect to the pair of shoes on which the victim's blood had been found [23].

Another example of the use of barefoot morphology to link a suspect to a shoe associated with a crime in a series of break and enters in 1999 and 2000 that occurred in Manitoba, Canada. Similar shoe impressions were found at three crime scenes. When a suspect was located, and discarded shoes were found corresponding to the shoe impressions from the crime scenes, he denied ownership of the shoes. After examination and comparison it was determined that the suspect was likely the wearer of the shoes, and he was charged with the three offenses [24].

References

[1] Massey, S. (2006). Persistence of creases of the foot and their value for forensic identification purposes, *Identification Canada* **29**(4), 124–137.

[2] Kennedy, R. (1996). Uniqueness of bare feet and its use as a possible means of identification, *Forensic Science International* **82**, 81–87.

[3] Bodziak, W. (2000). *Footwear Impression Evidence*, 2nd Edition, CRC Press, Florida, pp. 381–411.

[4] FBI Scientific Working Group on Footwear and Tire Track Evidence (2007). *Guide for the Examination of Footwear Impression Evidence*, at www.theiai.org/guideline/swigtread/index.php (accessed August 11).

[5] Kennedy, R.B Instructor Sgt. RCMP (2004). *Barefoot Examination Course*, personally attended October 10–16, Belleville Police Department, Belleville.

[6] Rich, J., Dean, D. & Powers, R. (eds) (2005). *Forensic Medicine of the Lower Extremity: Human Identification and Trauma Analysis of the Thigh, Leg, and Foot*, Humana Press, New Jersey, pp. 380–399.

[7] Kennedy, R. (1997). Forensic barefoot morphology, *Identification Canada* **20**(3), 14.

[8] Kennedy, R. & Yamashita, B. (2006). Barefoot morphology comparisons: a summary, *Journal of Forensic Identification* **57**(3), 383–413.

[9] Massey, S. personal communication July (2007).

[10] Quamra, S. (1980). Naked foot marks – a preliminary study of identification factors, *Forensic Science International* **16**, 145–152.

[11] Laskowski, G. & Kyle, V. (1988). Barefoot impressions – a preliminary study of identification characteristics and population frequency of their morphological features, *Journal of Forensic Sciences* **33**(2), 378–388.

[12] Personal communication, Brian Yamashita PhD, RCMP research scientist. August (2007).

[13] Kennedy, R. (2003). Statistical analysis of barefoot impressions, *Journal of Forensic Science* **48**(1), 1–9.

[14] Kennedy, R. (2005). A large-scale statistical analysis of barefoot impressions, *Journal of Forensic Science* **50**(5), 1–10.

[15] Luximon, A., Goonetilleke, R. & Tsui, K. (2003). Foot landmarking for footwear customization, *Ergonomics* **46**(4), 364–383.

[16] Rossi, W. (1983). The high incidence of mismated feet in the population, *Foot and Ankle* **4**(2), 105–112.

[17] Cassidy, M.J. (1980). *Footwear Identification*, RCMP publication, Ottawa, pp. 137–146.

[18] Smericki, C. & Lovejoy, C. (1985). Identification via pedal morphology, *Identification News* **35**(12), 5–12.

[19] Belkin, R. & Korukhov, Y. (1986). *Fundamentals of Crimnalistics*, Progress Publishers, Moscow, pp. 65–70.

[20] Hiller, E., Dixon, P., Stewart P. & Yamashita B. (2005). Recovery of DNA from Shoes, *Journal Canadian Society of Forensic Science* **38**(3), 143–150.

[21] (1949). *RCMP Gazette* **11**(11), 17–20.

[22] McCaffery J.D. (1955). The shoe fits, *The Police Journal: a Quarterly Review for the Police Forces of the British Empire* **28**(2), 135–139.

[23] Shor, Y., Kennedy, R., Tsach, T., Volkov, N.M.D., Novolselsky, Y. & Vinokurov, A. (2003). Physical match: insole and shoe, *Journal of Forensic Science* **48**(4), 808–810.

[24] Kennedy, R., Gehl, S., Massey, S. & Saunders, G. (2002). Use of barefoot morphology in criminal investigations, *Identification Canada* December issue, **25**(4), 4–6.

LESLEY HAMMER AND ROBERT KENNEDY

Footwear and Foot Impressions: Intelligence

Introduction

Over the past few years, there has been an increasing awareness among law enforcement agencies worldwide that footwear impressions have the potential to make a significant contribution to police intelligence.

The evidence from crime scene statistics suggests that footwear impressions are present at a large number of crime scenes, and its percentage is similar to DNA and fingerprints. Footwear impressions are routinely recovered from the scenes of crime by officers using a range of techniques, including photography, adhesive lifts, and gelatin lifts, which are the three most common methods. Clearly, the majority of footwear impressions are recovered from the scenes of more commonly occurring crimes, such as burglary, theft, and criminal damage. It is this type of offense that has been the focus for the establishment of intelligence systems. The previously established systems were fundamentally based on human input and effort. Today the databases are much more sophisticated and provide increasing opportunities to maximize the contribution that footwear intelligence can make. [1–8]

The following topics are reviewed in this article:

1. why can footwear impressions be used as intelligence?
2. structure of footwear databases
3. use of databases to provide footwear intelligence.

Why can Footwear Impressions be Used as Intelligence?

The majority of impressions are recovered from the scenes of more common offenses such as burglary, theft, and criminal damage. There is a general acceptance amongst police investigators that a small number of people is responsible for the majority of crimes. This means that by gathering forensic intelligence and using other information available to the police, such as type of offense, location, modus operandi (MO), it is possible to gain invaluable information about linking crime scenes and to potentially identify the person who is responsible for the crime. Fingerprints have been used to provide intelligence to police for many years and more recently, the National DNA Database (NDNAD) in the United Kingdom has shown how useful DNA can be to the police in providing thousands of "hits", either to link different scenes or to show that a specific person may be linked to a scene or scenes.

The value of footwear impressions has been recognized by the UK government when they passed changes to the Police and Evidence Act (PACE), which enabled sole prints to be taken from the footwear that were worn by the persons arrested by the police. Footwear has the potential to play a significant role in providing intelligence to law enforcement agencies.

There are limitations to the use of footwear and these are the same factors that are also applicable to any comparison work. The simple fact is that as an item of footwear is worn over a period of time, the appearance of the sole will change and the detail that it produces in an impression will also change. Some features such as the pattern gradually changes, whereas fine characteristic detail can change within hours or days.

In using footwear, there is always a need to consider the time between when an impression was made and when the footwear of interest was recovered as the examiner may have to consider the differences that they find and make a judgment as to the meaning and significance of any difference found.

Structure of Footwear Databases

A footwear intelligence system is primarily made up of the following essential elements:

- reference database and
- scene and suspect databases.

Reference Database

The reference database/collection provides a source of information that can be used by the people involved in providing the intelligence. The database normally consists of prints taken from the sole of known footwear, where it is possible to gather information with regard to the manufacturer, size, country of origin, etc. (*See* **Footwear and Foot Impressions: Databases**.)

The sole reference database consists of a collection of sole tread patterns from items of footwear. The database is made up of an image of the complete sole pattern and also normally contains images of the uppers that are associated with each pattern. Depending on the system, these tread patterns will be coded, either by the use of pattern component descriptors such as straight bars, wavy bars, circles, target, square blocks, logo etc. These descriptors codes can be in different formats ranging from a simple letter or a word code to represent a particular component to the use of sophisticated icons. The other alternative

is to have an alpha-numeric descriptor for a particular pattern, which is usually based on the name of the manufacturer and a name or number, e.g., Smith 100, Smith Classic. The importance of maintaining this reference database is to provide a universal descriptor of a specific pattern that enables examiners to assign a code or name to each footwear impression that is recovered from a scene and to each item of footwear recovered from a suspect who is arrested and comes into custody.

There are various systems available but almost all are computer based, which makes the coding and searching much easier. One feature that helps to improve the effectiveness of a system includes, regularly updating of the reference collection to ensure that new tread patterns are captured and included in the reference. It is also useful to have some indication of how often a particular sole pattern is seen, its commonality.

Scene and Suspect Databases

The other two parts of an intelligence system are the databases/platforms for the footwear impressions recovered from crime scenes and the footwear recovered from the suspects who are arrested for an offense by the police/law enforcement agency.

The law enforcement agency will have a lot of information and available data, and some of this will be fairly sensitive and may have restricted access. The general information, that is, stored for a scene mark will include such things as, date of offense, date of scene impression recovered, division, address of scene, other forensic evidence recovered, offense code, and of course the code of the footwear impression. A similar set of data will be recorded for a suspect's footwear but will be more specific about the individual and may include name and related reference numbers.

Use of Databases to Provide Footwear Intelligence

- coding of footwear impressions and suspect footwear;
- assessment of the evidential quality of a scene mark;
- scene-to-scene linking;
- scene-to-suspect linking and screening of footwear.

Coding of Footwear Impressions and Suspect Footwear

The recovered footwear impression from the crime scene will be examined and a code allocated for the tread pattern will be apparent in the impression. A pattern code will also be assigned to the footwear taken from a suspect. It is often the case that at a particular period of time, there will be specific footwear that occur most commonly and make up the majority of the marks recovered and the footwear taken from the suspects. Using the allocated codes, it is possible to search the suspect and scene databases to identify if there are any "hits" obtained which would indicate other scenes or suspects that may be linked and would warrant further investigation. There is of course the potential to use other parameters that can reduce the scope of the search.

Assessment of Evidential Quality of a Scene Mark

When examining the scene impressions, it is possible and useful to assess the quality of the impression and its intelligence and evidential potential. In undertaking footwear comparison, there are a number of factors that are taken into consideration; pattern, size, mould, wear, and unique characteristic damage features. It is therefore possible to assess both the intelligence potential of the footwear impression and the evidential potential of the footwear impression and allocate a point on a scale. It is accepted that any impression where it is possible to discern a possible pattern has some intelligence use, as it can be coded and used in a search to show potential links to other scenes. It is also true that a footwear impression consisting of a single pattern component such as a single square block may show sufficient characteristic detail to enable an item of footwear to be conclusively identified as the boot/shoe responsible. It could be said, however, to have limited intelligence value as it would be very difficult to assign the correct pattern code. A square block is found on a very large number of tread patterns.

An example of a possible quality scale is shown in Table 1.

A footwear impression is of "good quality" if the features of the mark, such as the pattern components and any other detail, are clearly apparent and well defined. Footwear impressions that are poorly defined and do not even show any pattern components

Table 1 Example of Intelligence/Evidential Assessment Scale

Point on scale	Intelligence/evidential value	Description
1	No intelligence or evidential use	Very poor quality impression with no clear pattern or detail
2	Potential to provide Intelligence only	Generally poor quality impression but has pattern components apparent but the mark does not show any other detail.
3	Potential to provide limited evidential support	Generally poor quality mark but with limited details present which may include, pattern, mould, and wear.
4	Potential to provide significant level of support	Fair to good quality mark with pattern, mould, wear and possible damage features
5	Potential to provide strong support or conclusive evidential result	Clear well-defined good quality footwear impression with characteristic feature present

are unlikely to be of any significant evidential or intelligence value. A footwear impression is of poor evidential quality if it is only possible to observe the most basic aspects of the tread pattern and there is no other detail present. Impressions that only show the basic tread pattern may be poor evidentially but do have an intelligence value. It should be noted that this description is valid only if no unexplainable/meaningful differences exist between the mark and the actual sole.

Scene-to-Scene and Scene-to-Suspect Linking

Law enforcement agencies have made the use of intelligence as one of the core activities. For many years, they have used fingerprints and DNA and combined the information supplied by these forensic disciplines together with all the other information that they use to aid investigation. By identifying potential suspects, they can also analyze the information generated to examine the areas such as crime trends and also help the police to target specific locations. Footwear is an additional forensic area that can contribute toward the picturethat is generated. It does not have the same potential to produce identification as fingerprints and DNA do but it can be very effective.

Scene-to-scene linking can be undertaken at the very simple level by using the pattern code to find potential links and then using other information to investigate the link further. The way in which this is undertaken will depend on the computer system used and the nature of any search engine. The integration of other case data such as modus operandi and time of the offense, for example, into the system may provide increased resources for intelligence and more generally crime analysis.

It is also possible to undertake a more detailed comparison of the "linked" scenes to determine if there is any other corresponding detail or features within the marks which would provide stronger evidence that the same footwear was responsible for both marks.

Scene-to-suspect linking is basically the same as that for scene-to-scene linking. The links can be

identified just by using the pattern code, and that is generated when a scene mark is first put on the database and when the suspects footwear is coded and then searched. It is also an option to do a degree of comparison between "hits" that may be generated from a search. The scene mark and a print of the sole tread pattern from the suspect footwear can be compared, and similarities and differences assessed will enable an examiner to determine what evidential potential a full one-to-one comparison may have.

References

[1] Birkett, J. (1989). Scientific scene linking, *Journal Forensic Science Society* **V29**(4), P271–P284.
[2] Mikkonen, S. & Astikainen, T. (1994). Database classification system for shoe sole patterns – identification of partial footwear impressions found at scenes of crime, *Journal Forensic Sciences* **V39**(5), P1227–P1236.
[3] Girod, A. (1996). Computerized classification of the shoeprints of Burglars' soles, *Forensic Science International* **V82**(1), P59–P65.
[4] Geradts, Z. & Keijzer, J. (1996). The image-database REBEZO for shoeprints with development on automatic classification of shoe outsole designs, *Forensic Science International* **V82**(1), P21–P31.
[5] Girod, A. (2000). Efficiency of a computerized database of Burglarsoloes' standards, *Information Bulletin Shoe Print and Tool Mark Examiners* **V6**(1), P125–P132.
[6] Milne, R. (2001). Operation bigfoot – a volume crime database project, *Science and Justice* **V41**(3), P215–P217.
[7] Napier, T.-J. (2002). Scene linking using footwear mark databases, *Science and Justice* **V42**(1), P39–P43.

DAVID BALDWIN

Footwear and Foot Impressions: Overview

Comparison and Identification

Comparison is defined as "the process of ascertaining whether two or more objects have a common origin" [1]. In footwear and foot impression evidence, this occurs through the analysis of the impression in question from the scene of a crime and an evaluation of similarities and differences to a source in question. The source item and exemplar impressions made by the item are compared to the impression from the scene.

In the area of footwear comparison, impressions are made by the outsoles of the shoes or boots. A footwear comparison is conducted using both the actual shoe and test impressions made using the shoe in question (Figure 1). The characteristics that are compared range from class or manufactured characteristics, which are attributes of the design and manufacturing processes, to accidental or acquired characteristics, which are due to random occurrences. Class characteristics may include the item of footwear as a possible source of the scene impression or they may eliminate the item of footwear as a source of the impression. The presence of sufficient individualizing characteristics, as well as the absence of unexplainable/meaningful differences, in both a scene impression and an item of footwear may result in identifying a particular shoe as the source of an impression.

There are quantitative and qualitative elements involved in the comparison of shoe impression evidence. Physical size, shape, and relative position of both manufactured and acquired characteristics may be compared quantitatively through processes such as overlaying a test impression of a shoe on top of an actual size photograph of a crime scene impression or examining the exemplars side by side and using calipers. Differences, similarities, and value of associations and disassociations are interpreted qualitatively based on training and experience. Influences such as movement during impression deposition, effects of substrate material, and collection methods are studied during training and are continually evaluated through experience. Training and certification programs establish standards and expectations for footwear examiner experts. The International Association for Identification has established a recommended course of study for shoe and tire examiners and has a certification program in the area of footwear examination [2]. The Canadian Identification Society also has a footwear examination certification program [3].

It is noted that there are differences in the methods of reporting of comparison conclusions between agencies and between countries. However, comparison results commonly range from general correspondence of outsole design characteristics to an agreement of unique characteristics. More information about the interpretation of footwear impression



Figure 1 A shoe outsole and a test impression of the same shoe

can be found in the Statistics(Interpretation) section of the present Encyclopedia.

Foot impression comparisons are conducted by analyzing an impression of a bare or sock-clad foot from a crime scene and an evaluation of the similarities and differences to the foot of a person to determine if they may or may not have made the impression. If a barefoot impression includes the recording of skin detail, known as friction skin, then the impression is analyzed in the same manner as fingerprints and palm prints. However, if a conclusion based on the unique features of the skin is not possible, a comparison of the shape features of the feet may be conducted. This type of comparison may result in the inclusion or exclusion of a person as the possible source of the crime scene impression.

Foot impressions and linking foot to shoe

Foot impressions also may be present in the insides of shoes. This type of impression may be of value to determine who wore a particular shoe. A shoe may be compared to exemplar impressions of the foot of a person and/or to other items of their footwear. By comparing the size and shape of the features of the foot, with consideration for the possible effects of these features when inside of a shoe, a person may be included or excluded as the possible source of the impressions in an item of footwear (*see* **Footwear and Foot Impressions: Foot Impressions and Linking Foot to Shoe** for more information).

Footwear impressions

Footwear impressions may occur as a result of deposition of a material, removal of a material, or imprint into a surface or substrate (Figure 2). Impressions may be categorized as two- or three-dimensional based on the resulting depth of the impression. Deposition and removal of material usually results in a two-dimensional impression, and impressing into a substrate results in a three-dimensional impression.

An item of footwear may step into a material such as dirt, dust, oil, or blood. This material can transfer in subsequent steps to a relatively clean surface, leaving an impression of the outsole of the footwear. A clean footwear outsole may come into contact with a surface covered with material such as dirt, fine dust, or powder. This material may stick to the clean outsole as travel occurs across the surface.

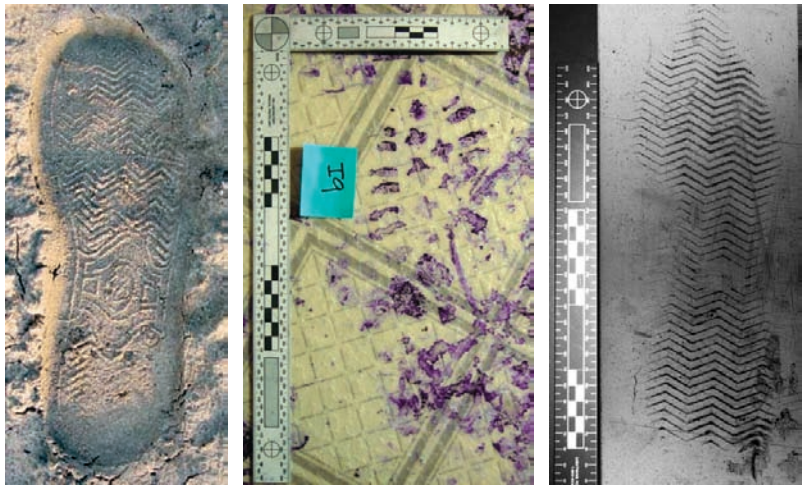


Figure 2 Examples of different types of crime scene footwear impressions (a) 3D impression in soil; (b) enhanced blood impression; (c) gel lift of a dust impression (color inverted)

This may leave an outline, or negative impression, of the outsole design behind.

A soft material, such as mud or snow, may hold an impression of a footwear outsole after a step into it has occurred. The condition of the material and the environment affects the quality of the impression over time. For example, the snow may melt, or wet mud may flow back into the impression, covering detail. If, however, the nature of the material is conducive to maintain the shape of the footwear outsole, a three-dimensional recording of the outsole will be recorded.

Detection and recovery techniques

The methods utilized for detection and recovery are dependant upon the condition and nature of the impression. Deposition or removal of material usually results in a two-dimensional impression and is usually photographed and lifted. A three-dimensional impression is usually photographed and cast.

Detection of footwear impressions may be accomplished by careful observation and through the use of lighting techniques, forensic light sources, application of chemicals, or lifting materials and equipment (Figure 3). The composition material of the impression will determine the method used for detection, possible enhancement and recovery. Dust and blood, for example, require a different series of processing techniques for two-dimensional impressions. There are also different casting materials and techniques utilized for three-dimensional impressions in dirt, mud, sand, and snow. Impressions on surfaces that may be removed from the scene may be collected, still attached to the original surface, such as a shoe impression on a piece of paper, or on flooring that may be cut out and removed.

Footwear outsole databases

Databases of footwear outsole impressions are utilized to search impressions from scenes of crime to determine what type of footwear may have left the impression. Once a possible source is determined, it may be possible through manufacturer information to determine a size or size range of shoes that made the impression. Outsole databases are created through the collection of footwear designs and entry into an electronic system, which allows searching and retrieval based on the patterns and designs of an impression.

Crime scene impressions of footwear may also be entered into database search and retrieval systems. Unidentified impressions may be compared to one another, potentially linking impressions from different crime scenes to one another (*see* **Footwear and Foot Impressions: Databases** for more information).

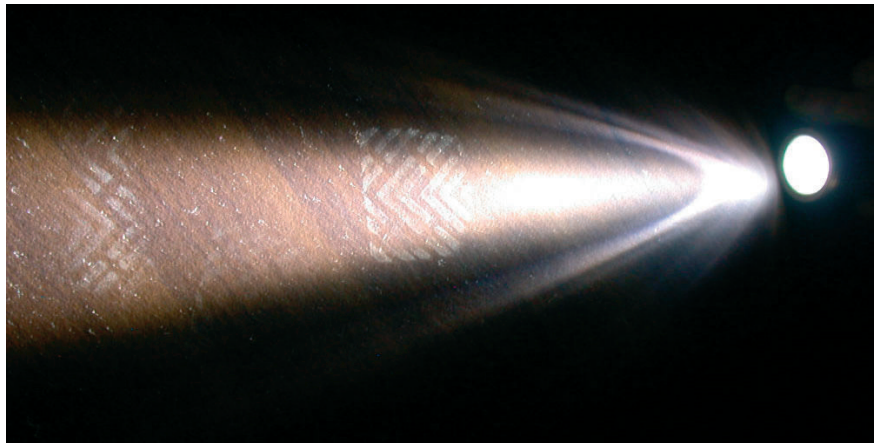


Figure 3 Using oblique light to search for shoe impressions in dust

References

[1] Saferstein, R. (2001). *Criminalistics: an Introduction to Forensic Science*, 7th Edition, Prentice Hall, New Jersey, pp. 61–74.
[2] Web page of the International Association for Identification: www.theiai.org (accessed on, 2008).
[3] Web page of the Canadian Identification Society: www.cis-sci.ca (accessed on, 2008).

Further Reading

Cassidy, M.J. (1980). *Footwear Identification*, RCMP publication, Ottawa.
Bodziak, W. (2000). *Footwear Impression Evidence*, 2nd Edition, CRC Press, Florida.
Abbott, J.R. (1964). *Footwear Evidence*, Charles Thomas Publisher, Springfield.

LESLEY HAMMER

Force: Use of by Police ***see*** Police Use of Force

Forensic Anthropology ***see*** Anthropology

Forensic DNA: an Overview ***see*** DNA: an Overview

Forensic Histology ***see*** Histology

Forensic Resources on the Web ***see*** Web Resources

Forged and Counterfeit Documents

Introduction

While the terms *counterfeit* and *forgery* are often used synonymously, there exists a distinct difference between counterfeit and forged documents. A counterfeit document can be defined as one wherein

a genuine specimen is used as a template for the creation of a document that infringes on the property rights of the issuer and misleads the acceptor as to its true identity, validity, and/or authenticity [1]. Examples of counterfeiting include constructing a fraudulent identity card, generating spurious travel documents, or the unauthorized printing of money. Forgery, on the other hand, involves the illegal alteration of an existing genuine document and includes actions such as modifying the amount for which a check is written or altering the date of inception of a contract. Alternatively, rather than referencing components of a document that have been contrived, forgery can also relate specifically to an intent to deceive [2].

Counterfeit documents serve many purposes for those attempting to defraud, allowing them to assume identities, make purchases, obtain credit, and orchestrate illicit activities that would not otherwise be possible. The types of items that are counterfeited or forged are broadly encompassing and include travel documents, such as drivers' licenses, passports, and visas; business and personal checks; travelers checks; and resident alien or immigration cards. The negative impact of counterfeiting and fraud to society can be extensive. Consider the following excerpt taken from the 9/11 Commission Report:

> Fraud in identification documents is no longer just a problem of theft. At many entry points to vulnerable facilities, including gates for boarding aircraft, sources of identification are the last opportunity to ensure that people are who they say they are and to check whether they are terrorists [3].

The costs incurred as a result of counterfeiting and forgery are realized by private citizens, consumers, businesses, the financial services industry, and the justice system. It is estimated that credit card fraud costs to both cardholders and card issuers approximates hundreds of millions of dollars per year [4]. In the year 2000, Visa and MasterCard reported losses in the United States of an estimated $1.0 billion under the traditional definition of identity theft, a 45% increase over these losses reported in 1996 [5]. The American Bankers Association reported that 80% of banks cited losses due to check fraud in 2006 with a total of $12.2 billion attempted to be drawn against accounts [6]. The US Social Security Administration estimated in 2000 that reports of misuse of Social Security Numbers increased more than fivefold over the number in 1998, with 81% of these directly attributable to identity theft [5].

Virtually any type of document can be simulated or altered, and examinations are often conducted on business and personal contracts; employment, banking and medical records; vehicle titles and license plates; letters of credit or reference; insurance documentation; certificates of stocks or bonds; and even sports memorabilia. Certain documents are classified as breeder documents, those that enable an individual to obtain other records that can be used for identification purposes [7]. These documents are often targeted for fraudulent reproduction by counterfeiters with the final objective being the acquisition of a more generally accepted, secure document (e.g., passport). Breeder documents may include birth and marriage certificates, social security cards, utility bills, military identification cards, and library cards.

Instruction in constructing these fraudulent materials can be found in books, magazines, and on the internet; lack of standardization of documents within and among countries has served to facilitate the perpetuation of counterfeits and forgeries. The equipment and materials used in counterfeiting are typically easily obtained, and a functional system may be as basic as a computer with desktop publishing software, camera, scanner, printer, and laminator. Although the cost of these software and hardware components is relatively low, often counterfeiters use fraudulent means for acquiring goods; therefore, the true cost of the items to them may be negligible. A more elaborate equipment set is typically not necessary in that while these methods of producing counterfeit documents will differ from those used in the genuine standards, the finalized document need not be a perfect simulation; it need only be sufficient in quality to pass the scrutiny of the initial reviewer. These individuals who constitute the first line of defense against counterfeit documents may have diverse backgrounds and skill sets and represent a variety of professions, including cashiers, tellers, clerks, immigration officers, and aviation employees. They may have varied levels of training in document comparison and feature recognition, and these verification examinations may be conducted in environments in which both the time and equipment available for these inspections is limited. The increasing prevalence of counterfeited and forged documents in circulation equates to an increased necessity for the

capability of differentiating those that are spurious from those that are genuine.

The Role of the Forensic Document Examiner

One of the primary roles of the forensic document examiner (FDE) is to analyze documents in order to determine genuineness. If a fraudulent document is detected, the FDE will then gather information that can elucidate how the forgery or counterfeit has been achieved. This may allow the examiner to identify and compare materials and methods of manufacture with those discovered in relation to one or several suspects, as well as to link different cases.

While one might consider documents to be specimens comprised strictly of paper substrates, a document can be any object containing marks or symbols that convey meaning to someone [2]. These marks can include characters that are handwritten, electronically generated, etched, or transferred from other surfaces and can be found on substrates such as plastics, metals, rubber, wood, and textiles. The role of the FDE is not to establish intent, but rather to objectively examine evidence and impartially report findings. In federal courts in the United States, determining intent of a defendant is a responsibility that lies solely with a judge or jury; therefore, in this context, inclusion of the word "forgery" in the reports and testimony of the FDE is highly discouraged [2].

Cited in the American Society for Testing and Materials (ASTM) standard E444-07, entitled "Standard Guide for Scope of Work of Forensic Document Examiners", a primary responsibility entailed in the comparison, analyses, and examination of documents is to "establish genuineness or nongenuineness, or expose forgery, or reveal alterations, additions, or deletions" [8]. The standard outlines the definition of an FDE, typical work encountered in the field, types of equipment utilized, the range of expertise required to satisfactorily examine evidence, and the reporting of results. The scope of work of the FDE includes the identification of equipment utilized to print, copy, or fax a document; identification or exclusion of substrates, inks, and writing instruments; determination of age, sequence, source, and alterations to documents; isolation of relationships between documents; recovery of eradicated or obliterated components of documents; and examination of content for evidence of spuriousness or to ascertain the identity of a person. In order to accomplish these objectives, the FDE must employ a multitude of techniques and utilize a host of instrumentation in the process of deriving accurate conclusions. These types of examinations can be challenging, but, nonetheless, standardized methods and instrumentation exist that can be used to determine whether a document is genuine or counterfeit, or whether the integrity of a once genuine document has been compromised. At minimum, the equipment to which an examiner should have access includes a microscope, an electrostatic detection device (EDD), and an instrument equipped with filtered lighting in order to make these determinations.

Microscope

Compound and stereo microscopes are primary tools of the FDE for providing magnification necessary to visualize and examine minute details of a document. Sufficient magnification and lighting are often essential in determining the printing process(es) employed, identifying individual and class characteristics, and gaining an understanding of how the document was constructed. Comparison microscopes allow for the side-by-side comparison of a genuine and a questioned document. More sophisticated methods, such as polarizing light, fluorescent, confocal, and atomic force microscopy, may also be used to evaluate physical and optical properties of documents. Scanning electron microscopy (SEM), in which a beam of electrons is bombarded across a surface resulting in the generation of high resolution images capable of resolving fine detail, is often also useful in document analysis. The stage of these instruments can be adjusted along and rotated about the x, y, and z axes, allowing for the viewing of virtually any aspect of the sample. Using this method, micrographs of ink particles and substrate fibers can be obtained that allow for detailed comparisons between questioned and genuine specimens. Figure 1 demonstrates differences in the printing quality of a genuine and counterfeit document that are easily visualized using approximately $50\times$ magnification.

Electrostatic Detection Device

An EDD is an invaluable tool that provides FDEs with a method to visualize indentations present on



Figure 1 A split photomicrograph showing differences in the printing features of the same area on a counterfeit (left) and genuine (right) document using approximately 50× magnification

paper documents. When entries are made on documents, indentations may be created on underlying pages. First introduced to the questioned document field in 1978 by Foster and Freeman, Ltd. as the Electrostatic Detection Apparatus (ESDA), electrostatic detection methods are considered minimally destructive, having a slight potential for effects to paper substrates resulting from a brief humidifying process, and the potential removal of small quantities of pencil lead from a document [9, 10]. Foster and Freeman™ currently sells the ESDA^2 and ESDA-*Lite* systems, while Projectina™ offers the Docustat. The electrostatic detection process, in general, is based upon the premise that physical contact with a substrate, such as from a writing instrument, causes a perturbation in paper fibers on underlying sheets of paper. This paper fiber disturbance results in a difference in electrostatic potential from undisturbed regions of the substrate, which can then be realized when a document is covered with a Mylar film, and a highly negative voltage is applied across it. Subsequent exposure of the document to toner particles, which adhere to indented regions, then allows for the visualization of these perturbations [11–13]. In addition to indented writing, an EDD can be used to visualize alterations that have been made to documents. In Figure 2, the date on a document was abrasively removed and then modified. Although in this case the previous number was not able to be determined, significant evidence of abrasion exists.

Video Spectral Analysis or Filtered Light Examination

Video spectral analysis (VSA) and filtered light examinations (FLEs) are nondestructive methods that can provide valuable insight regarding the optical properties of inks and other materials found on a document. FLE allows the user to view specimens at different wavelengths of light using various barrier and excitation filters; however, VSA can also be used to examine items using forms of radiant energy other than visible light, such as ultraviolet (UV) and infrared (IR). Evaluation of near infrared reflectance

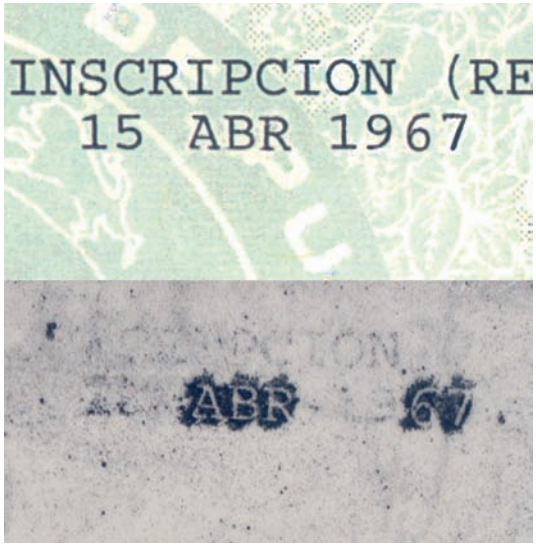


Figure 2 The examination of a suspected date change using an electrostatic detection device (EDD), with the original document (top) as compared to the EDD lift (bottom). Although an EDD can sometimes be used to visualize previous written entries, it was not possible to decipher the original date on this document; however, there is significant evidence of abrasion

(IRR) and infrared luminescence (IRL) properties can provide significant assistance when evaluating inks present on a document (Figure 3). VSA systems available on the market include the Video Spectral Comparator (VSC) from Foster & Freeman™ and the Docucenter and Docubox offered by Projectina™, all of which employ a combination of filters and photography to elucidate features of a document that may not otherwise be detected. Modern commercial equipment, including these instruments, typically has incorporated a charge coupled device (CCD) into the design, thereby providing users with the opportunity to examine and record images electronically.

Although the aforementioned instrumentation can be considered a minimum requirement for the FDE, other analytical methodologies can be used to conduct detailed chemical and physical analyses of materials used in the construction of counterfeit documents. Fourier transform infrared spectroscopy (FT-IR) is a technique that uses infrared radiation to excite bonds within molecules in order to isolate chemical functional groups. Using this instrumentation, components of documents can be analyzed to determine if they share the same chemical composition. Chromatography is an analytical technique used to separate individual components within a substance based on characteristics such as differences in solubility in a solvent. Depending on the material being analyzed, thin layer chromatography (TLC), gas chromatography (GC), or high-performance liquid chromatography (HPLC) can also potentially be used. In addition, mass spectrometry (MS) can be coupled with GC or HPLC to provide specific chemical identification of substances found in documents (e.g., colorants). Advanced methods of presentation of specimens to a mass spectrometer exist and include thermal desorption and direct analysis in real time (DART), which allows for noncontact sampling under ambient conditions of solid, liquid, and gas-phase specimens.

Other instrumental analysis methods can be used to assist in identifying inorganic compounds (e.g., metal complexes in pigments, minerals in paper, and plastics), such as X-ray fluorescence (XRF), SEM coupled with energy dispersive X-ray analysis (EDXA), and inductively coupled plasma mass spectrometry (ICP-MS). These are highly specialized techniques that can be used to detect trace components and/or identify unknown materials either found on or used in the construction of a document and can be invaluable tools in determining associations among specimens.

Irrespective of the request, the FDE must have a fundamental understanding of how documents are constructed. In order to determine if a specimen is counterfeit or forged, one should have an in-depth understanding of how the genuine was created, as well as a specimen file of known standards. This knowledge is critical for comparative examinations of suspect documents to be conducted objectively. Also, inherent limitations may exist that must be realized in order to derive accurate conclusions (e.g., manufacturers of components of documents may employ limited quality control measures in the production process, thereby resulting in variability in the genuine product). Following are discussed general examination procedures for questioned documents, along with some of the production methods and security features that are commonly utilized in the construction of genuine specimens. However, since there are numerous types of documents that can be counterfeited, this summarization should not be construed as an unabridged discussion of all possible



Figure 3 A split image showing differences in the infrared luminescence properties of a counterfeit (top) and genuine (bottom) document both printed using offset lithography

types of features that can be found in genuine documents.

Examination Procedures

The successful forensic examination of suspected counterfeit documents is contingent upon the use of a nonbiased approach to scientifically demonstrate that the questioned document is either genuine or counterfeit. Ultimately, the soundness of the findings must be irrefutable in a legal proceeding. In order to accomplish this objective, it is considered best practice that the FDE reference a genuine standard either obtained from or approved by the issuer of the document. In the event this authentic standard is unavailable, the examination may be conducted using a listing of features contained within the genuine that has been obtained from the issuing authority. It should be kept in mind that some documents produced by counterfeiters, termed fictitious instruments, are completely fabricated and have no genuine counterpart. In the case of forgeries, the document itself is typically genuine, but has been altered in some capacity; therefore, those features present in the genuine standard will typically also be evidenced in the forged document.

An examination is then conducted in which the questioned document is compared to either the genuine standard or reference list of features, and the similarities and differences recorded. This is best conducted by evaluating four categories of attributes to include (i) physical characteristics, (ii) printing processes, (iii) security features, and (iv) electronic media.

Physical Characteristics

Observation of physical characteristics, such as the dimensions, color, gloss, and physical composition of a document, is typically the first step in the

examination process. Although examiners should not rely solely on manufacturers' specifications, because there can be inherent oversight in the quality control process, many secure documents have strict regulations and manufacturing tolerances with regard to these attributes. For example, the sizes of identification cards are specified by the International Standards Organization (ISO) in the standard ISO 7810, "Identification Cards–Physical Characteristics", in which is detailed four official formats, ID-000, ID-1, ID-2, and ID-3. The most common size used for financial and identity cards is referred to as *ID-1* or *CR80*, which is defined as being 85.60 mm (3.370 in.) in width and 53.98 mm (2.125 in.) in height, and having a nominal thickness of 0.76 mm (0.030 in.) [14]. The currently issued US currency is 155.956 mm (6.14 in.) in length, 66.294 mm (2.61 in.) high, and 0.109 mm (0.0043 in.) thick. The size of a document may be subject to extremely narrow production tolerances, on the order of tenths of a millimeter. Therefore, counterfeit specimens may be readily identifiable when deviations are found in measurements, even as little as 0.5 mm from the genuine standard [15].

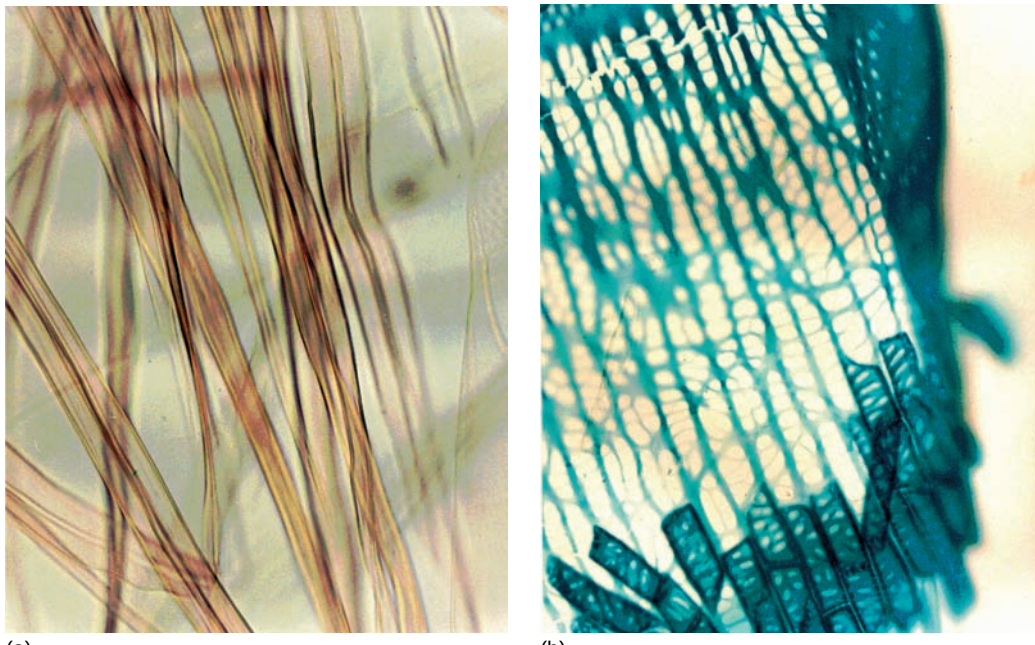
Color and gloss differences can be key in distinguishing among counterfeit, forged, and genuine specimens. Differences in colorimetric values between the suspect document and a genuine standard may be measured using a spectrocolorimeter, a spectrophotometer designed for the quantification of minute color differences, as well as for the evaluation of properties such as metamerism in documents. Using a spectrocolorimeter or microspectrophotometer, values for various colorimetric indices may be calculated, such as whiteness index, paper brightness, and colorant strength, which can be of additional value in comparisons between and among documents. Gloss characteristics of a counterfeit document may differ from those of an authentic specimen and may be quantified using a gloss meter (typically measured at various angles, such as 25, 60, and 85, depending upon the level of reflectivity of the surface).

Examination of the various components of a document *via* physical, optical, and chemical means can allow one to determine source or establish associations with other documents. This may include the analysis of paper fibers present in a cellulosic document (Figure 4a, b), the identification of polycarbonate, or the structural composition of a multilayered polyvinyl chloride (PVC) document. Other physical attributes of a document can be examined and include embossing characteristics, omissions of symbols or text, misspellings, and the formatting of designs, logos, and text with respect to both their design and their relative positions on the document. Regional conventions not indigenous to an area may be commonly encountered, such as improper denotations of monetary amounts, dates, or telephone numbers. All of these attributes can serve as class characteristics to help associate multiple documents to a common source. The document in Figure 5 contains a misspelling that can assist an examiner in establishing that a document is a counterfeit and then potentially associating it with others.

Printing Processes

Genuine documents are often designed such that fraudulent reproduction would be inherently difficult. One method for achieving this is to select printing processes that are not commonly available to the counterfeiter, and thus, would require time or expense that he or she would potentially be unable or unwilling to expend. The three most commonly utilized printing processes in creating genuine specimens, and most difficult to simulate by the counterfeiter, include offset lithography, intaglio, and relief printing (e.g., letterpress and flexography). When these processes are used in combination with one another, additional layers of security are then present, with the resultant composite product being a more robust document.

Offset Lithography Offset lithography is a planographic printing process wherein hydrophobic inks are applied to a plate containing hydrophilic nonimage areas that have been moistened using a dampening solution. As a result, the ink is repelled by the nonimage areas and adheres to only the hydrophobic image areas of the plate. The image is then offset to an intermediate blanket cylinder, and as the substrate travels between the blanket cylinder and an impression cylinder, the image is subsequently transferred. In security printing, a line offset method is primarily utilized rather than half-tone dots to create an image. The printed areas are typically comprised of fine lines that may form geometric patterns or other intricate designs. Spot colors are applied and may form a gradation of shades, gradually transforming from one color to another across the design, a technique referred to as *rainbow*, *split fountain*, or *iris*



(a) (b)

Figure 4 Photomicrographs of fibers used in the formation of paper: (a) eastern cottonwood fibers and (b) tupelo fibers

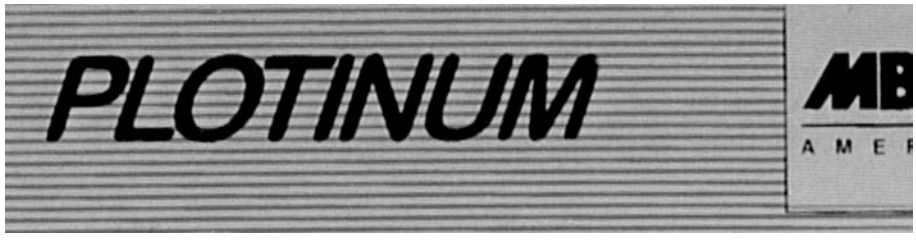
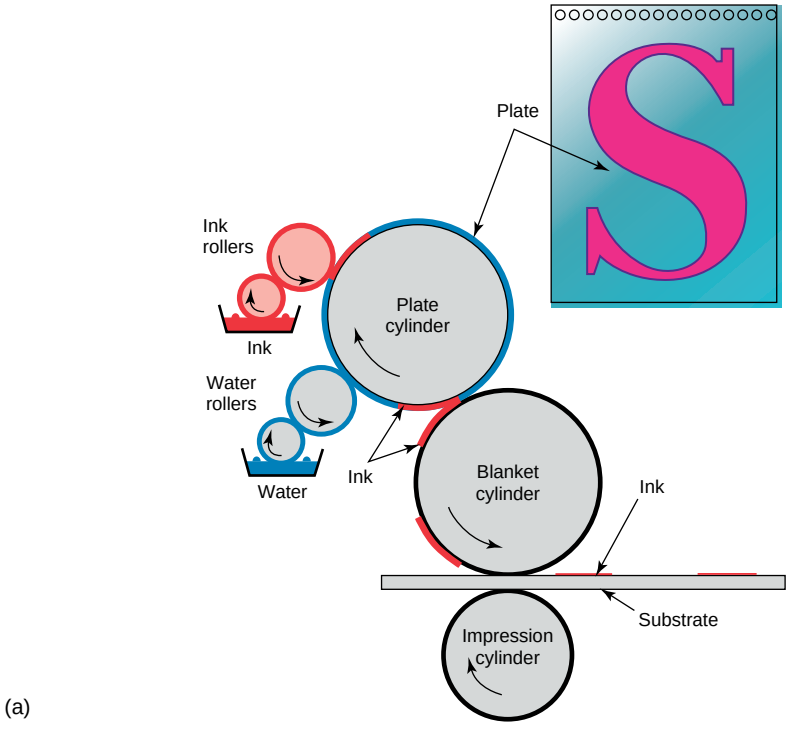


Figure 5 The word "PLATINUM" on a counterfeit credit card is misspelled and reads "PLOTINUM"

printing. The combination of this gradient coloration, resolution of the composite lines, and registration of components equates to an advanced degree of difficulty in counterfeiting documents prepared in this manner [15]. Figure 6 consists of an illustration of the offset lithography process, as well as a photomicrograph of the resultant print.

Intaglio Intaglio printing is a process wherein the image area of the printing plate or cylinder is etched or engraved. Once ink is applied, it subsequently fills the engraved areas, with the excess being removed by a doctor blade that is wiped across the surface. The plate then contacts the substrate under extremely high pressure, exceeding tens of tons per linear inch. Through capillary action, in combination with this high pressure, the ink from the etched image areas is transferred to the substrate, resulting in an embossed image on the document. A characteristic tactile quality is thus imparted to intaglio printed documents, which can serve as an additional security feature. Aged documents can lose some of this tactile character, and therefore, counterfeiters may attempt to artificially age documents to account for the lack of this property. Additionally with intaglio printing, ink is not entirely contained within the structure of the image upon removal of the plate, thereby resulting in a feathering effect that can be visualized

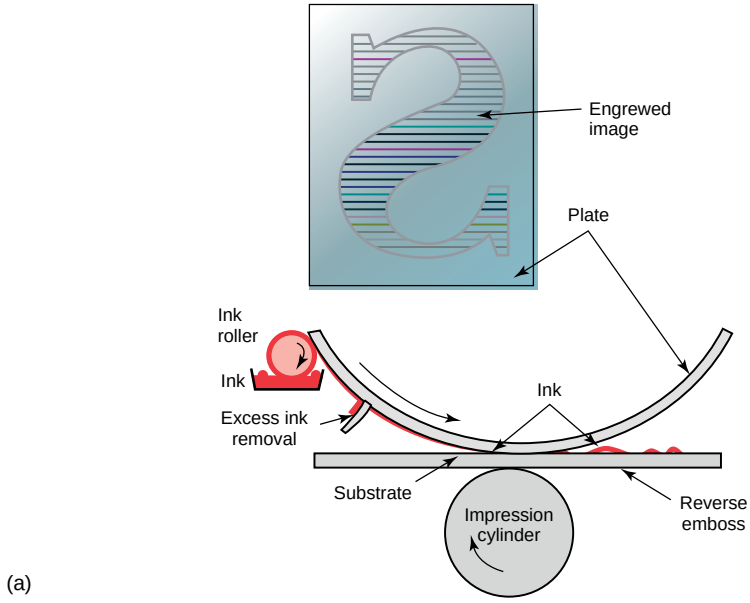


(a)



(b)

Figure 6 (a) A schematic representation of offset lithography printing; (b) a photomicrograph of an image printed using offset lithography



(a)



(b)

Figure 7 (a) A schematic representation of intaglio printing; (b) a photomicrograph of an image printed using intaglio

microscopically and is difficult to duplicate. Many intaglio presses afford an additional advantage of allowing the application of multiple colors, an arduous feat for the counterfeiter to attempt to duplicate with multiple passes of each individual color in that the registration will invariably be poor [15]. Figure 7 consists of an illustration of the intaglio process, as well as a photomicrograph of the resultant print.

Relief Printing

Letterpress and flexography are processes wherein a plate containing a raised image area is inked and, *via* impact, the impression is transferred directly to the substrate. The differences in the processes are that flexography utilizes a flexible rubber or photopolymer plate as opposed to the rigid plates that are commonly incorporated in letterpress printing. Letterpress is primarily used to print unique identifiers on a document, such as a serial number, that align uniformly with inherent design elements. Use of other means of creating these often results in the digits being misaligned with respect to the rest of the document. The counterfeiter is forced to either attempt to simulate this process or to utilize the same number or limited number set repeatedly, which may then be counterproductive to the primary goal of escaping detection [15]. Flexography can be used on a variety of surfaces and is typically used in high speed printing, as the low viscosity solvent or water-based inks utilized in this process are formulated to dry quickly. Figure 8 consists of an illustration of the letterpress process, as well as a photomicrograph of the resultant print.

A point of commonality of all of these printing processes is that the effects imparted by them are virtually impossible to reproduce *via* digital means, the primary methods utilized by many counterfeiters. The printing processes most commonly available to the general public typically include electrophotography (e.g., laser printers, photocopy machines), inkjet, dye diffusion thermal transfer (D2T2), and thermal mass transfer, all having unique properties that differentiate them from those typically utilized in security printing. Other types of processes can be utilized in genuine documents, but may be more readily reproducible by the counterfeiter. These include screen printing, flexography, and thermography. Some forgers construct miniaturized versions of production presses; however, large quantities of work cannot be processed with this equipment, and the same properties imparted through use of a commercial press are typically not evidenced. Counterfeiters may even have access to production presses themselves, gained through either fraudulent means or provided by a government, organization, or individual wishing to defraud. This possibility must always be considered, and close attention to detail must be therefore applied in each comparison made.

Security Features

A vast array of physical, optical, and chemical security features can be incorporated into documents to ensure their integrity, the objective being to convert a printed document into an end product for which the validity and authenticity can be verified. It is, therefore, critical that the printer(s), supplier(s), and issuer(s) adhere to secure operating procedures during all manufacturing and distribution protocols. Without strict security regulations, a breach of information and/or materials security may result that could serve to compromise the integrity of the document.

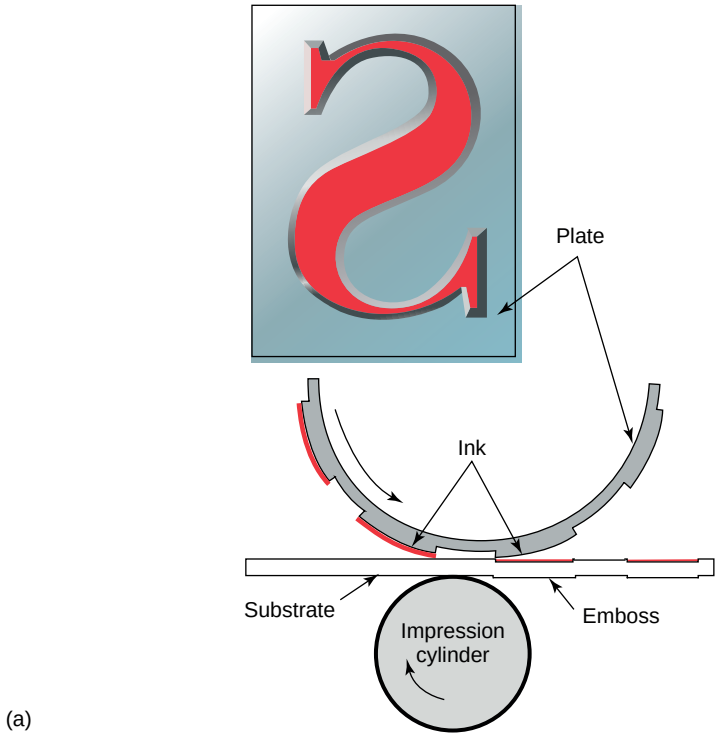
Security features can be considered overt, covert, or forensic. Overt features are those designed to be either detectable by the unaided eye under normal viewing conditions or discernible by touch (e.g., an identification number or watermark). These features are often key in inspection processes in which the expeditious evaluation of documents for genuineness is vital, such as at congested security checkpoints. Covert features are those not visible to the unaided eye or not tactile in nature. They may require specialized light sources or equipment in order to be detected (e.g., ultraviolet source or barcode reader). For forensic purposes, sometimes considered an additional category, these covert features are disclosed on a need-to-know basis and typically require specialized expertise and/or instrumentation for identification, as with the detection of a unique ingredient incorporated into an ink. It should be emphasized that it is often not necessary to be privy to the forensic features that are incorporated into a document if authenticity can be definitively determined on the basis of overt and/or covert features.

While it is well beyond the scope of this article to provide detailed descriptions of all the various security features available for use in constructing a document, following are described those commonly encountered. These features can be generally categorized into those incorporated within the substrate, those involving the printing process, and additive features.

Security Features Incorporated into the Substrate.

Watermarks

An image or pattern found in paper that appears less dense than the paper itself when viewed with transmitted light. Watermarks can be created using two different processes: fourdrinier (dandy roll) and



(a)



(b)

Figure 8 (a) A schematic representation of letterpress printing; (b) a photomicrograph of an image printed using letterpress

cylinder mould. During the papermaking process, an image imprinted onto a wire is attached to a dandy roll and impressed at varying degrees of pressure into the wet paper. The cylinder mould process uses areas of relief on the roll, rather than a wire with an image, resulting in a three-dimensional image in the paper, having various shades of gray. The watermark from the cylinder mould process usually contains far more detail than a dandy roll watermark. Other legitimate artificial or chemical watermarks can be used; however, the end product of these processes does not result in a true watermark. Typically, counterfeiters will use simulated watermarks that can be easily created using faint printing, stamps, and specialty inks (e.g., UV fluorescent) (Figure 9a). These are usually detectable to a trained examiner due to the lack of detail, ability to visualize without the aid of transmitted light, and/or the presence of UV fluorescence properties. Figure 9b illustrates a genuine watermark utilized in a secure document.

Planchets

Thin pieces of paper or synthetic material integrated into the substrate of a document during manufacturing. Planchets typically are found randomly distributed within bands on a document and may be iridescent, fluorescent, chemically treated, thermochromic, or contain microtext. Simulations may include drawing or printing colored disks onto paper; however, a genuine planchet will be removable from the substrate using a sharp object and will leave a void in its absence. It should be noted that authentic documents that have been frequently handled may be devoid of planchets; however, the voids left by their presence will remain. The images of planchets on a counterfeit document, created *via* a reprographic process, can be used to associate multiple documents if they are shown to be in identical positions. Figure 10 demonstrates the nonrandom distribution of planchets found in counterfeit documents having a common source.

Window thread

A ribbon that is laced through the substrate in an alternating pattern of visibility (windowed) and nonvisibility when viewed in reflectance. One iteration of the pattern is typically 1 cm in length, and when viewed in transmission, the entirety of the reflective ribbon can be visualized. The threads may be holographic in nature or contain microtext printing as additional security measures. Counterfeiters may draw an approximation of a window thread or attempt simulation by utilizing foil printing to create a ribbon having alternating areas of opaqueness; however, in neither case will the ribbon appear continuous under transmitted light. Figure 11(a)–(e) illustrates some of the different types of window threads that are found in secure documents, along with a counterfeit simulation.

Security fibers

Embedded into the substrate, these randomly distributed, colored fibers typically must be evaluated microscopically in order to ascertain whether they are actually individual fibers, or whether they have been drawn or printed onto or adhered to the surface of a document. Genuine security fibers can often be distinguished from the latter category of counterfeits in that they will present as being partially below and partially above the surface of the document (Figure 12). Additionally, these fibers may be fluorescent, iridescent, or invisible except when viewed under certain conditions. Fibers may also exhibit unique chemical signatures, for which identification using FT-IR may be warranted.

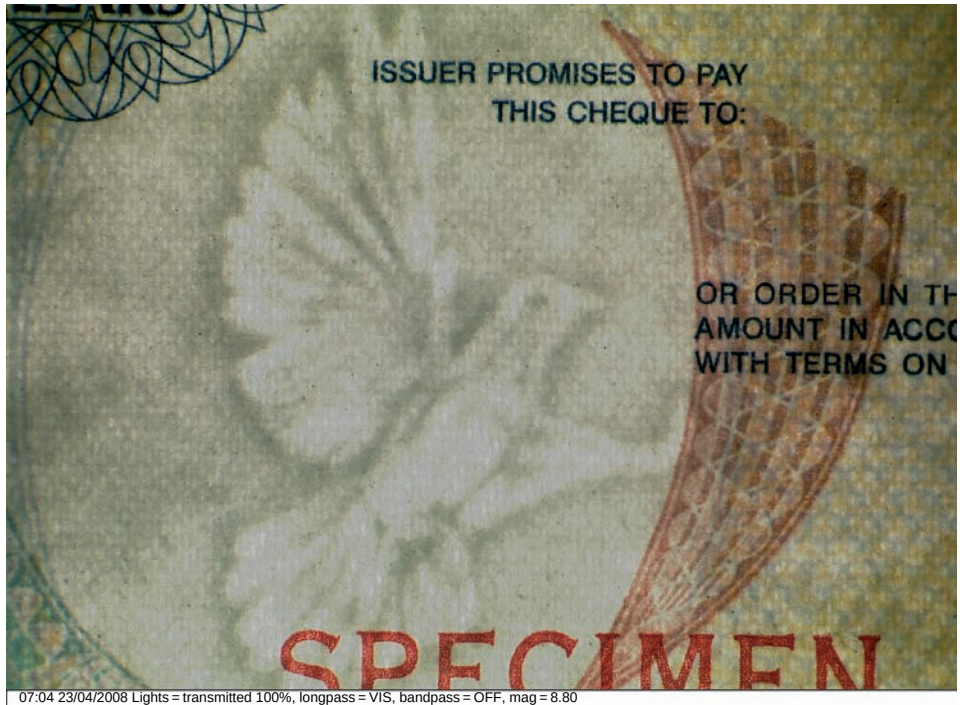
Security Measures Incorporated into the Printing.

Intricate printing

Features that are generally difficult to reproduce and require sophisticated technology to create. These can include complex patterns created with fine line detail and variable line thickness, custom geometric patterns, guilloche patterns (consisting of two or more bands of continuous, overlapping sinusoidal lines), microtext (generally 15–90 micrometers in size, see Figure 13a, b), two-sided registration (splitting of a graphic image such that one-half is displayed on the front of the document and one-half is displayed on the back of the document), and rainbow printing (i.e., split fountain printing, where two or more ink colors in a single unit merge together as a result of oscillating ink rollers). Figure 14 depicts a document containing both a guilloche pattern and rainbow printing. Depending upon the sophistication of the counterfeiter and the type of equipment being used, simulated intricate printing patterns are often readily detectable using examination under an appropriate level of magnification. Other printing features serve specifically as antiphotocopying mechanisms, such as void pantographs, wherein images such as the text “copy” or “void” are produced at a higher



(a) 07:06 23/04/2008 Lights = 365nm ultra violet, longpass = VIS, bandpass = OFF, mag = 8.80
integration = 1/3, iris = 73%, gain = 12dB, brightness = 10, contrast = 30, imaged width = 41.82 mm



(b) 07:04 23/04/2008 Lights = transmitted 100%, longpass = VIS, bandpass = OFF, mag = 8.80
integration = 1/8, iris = 50%, gain = 6dB, brightness = 28, contrast = 30, imaged width = 41.82 mm

Figure 9 (a) A simulated watermark visualized with 365-nm ultraviolet source is compared with (b) a genuine watermark visualized using transmitted light. The simulated watermark is not visible when viewed with transmitted light and the genuine watermark is not visible when examined using an ultraviolet source

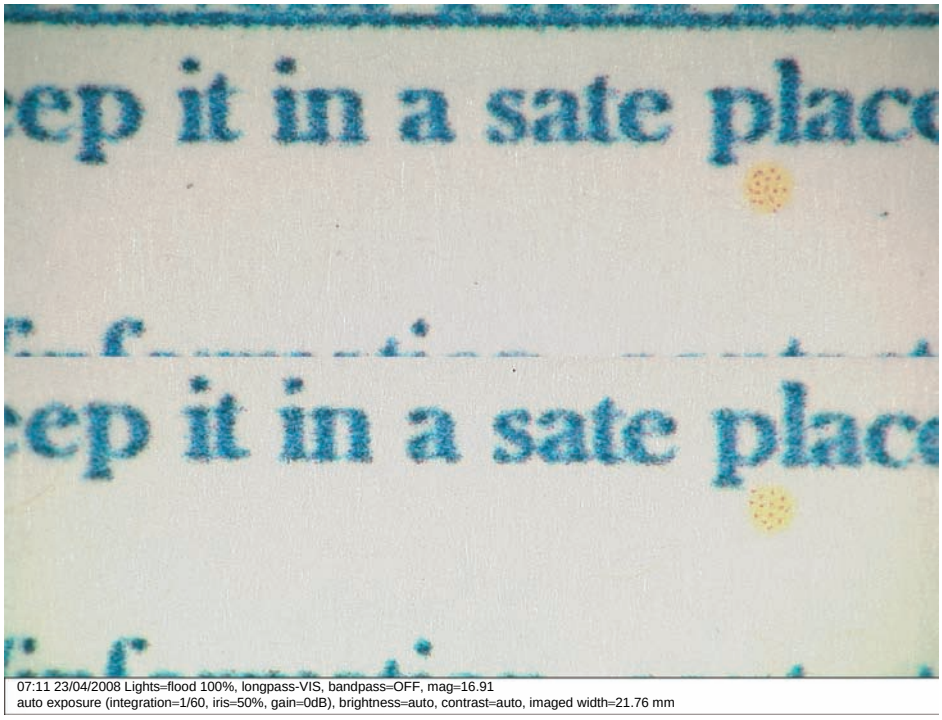


Figure 10 A split photomicrograph showing simulated planchets with the same relative placement and position. This would not be expected in genuine documents in that the planchets are randomly distributed in the substrate during the papermaking process. The top and bottom images are two counterfeit documents bearing different names that have been shown to be associated with a common production source. Also note that the word "sate" is misspelled and should read "safe"

screen ruling than that of the surroundings. Therefore, the pantograph is camouflaged in the original document; however, upon reproduction, it becomes readily visible.

Security inks

Numerous types of security inks are available, which exhibit various characteristics that can lend uniqueness to genuine documents. Iridescent or color-shifting inks exhibit different hues when viewed at varying angles and can be easily simulated for presentation to the untrained observer, but are very difficult to reproduce exactly. For example, counterfeiters often use metallic material sprinkled into the ink to give the perception that the ink is changing colors when tilted at different angles under a light source. Noniridescent inks include thermochromic (change in color due to a change in temperature), photochromic (change in color due to exposure to different wavelengths of light), fluorescent, infrared absorbent, and those exhibiting anti-Stokes shifts (fluorescence that is excited using a longer wavelength energy than that of emission). It is important to note that the changes that thermochromic and photochromic inks undergo when exposed to different temperatures or radiant energies are reversible. Metameric inks exhibit the property of appearing to be equivalent in color under one viewing condition, but perceptibly different under another. This phenomenon stems from the fact that the inks actually have different spectral properties. Despite this, the inks are matched under one illuminant or light source, such that the colorimetric difference derived from their spectra is below the threshold of detection by the human eye. An examination using some form of filtered light or VSA can be extremely helpful when characterizing both security and counterfeit inks. Also, GC/MS and FT-IR may be used to authenticate genuine security inks or link counterfeit inks in order to ascertain if they originated from a common source.

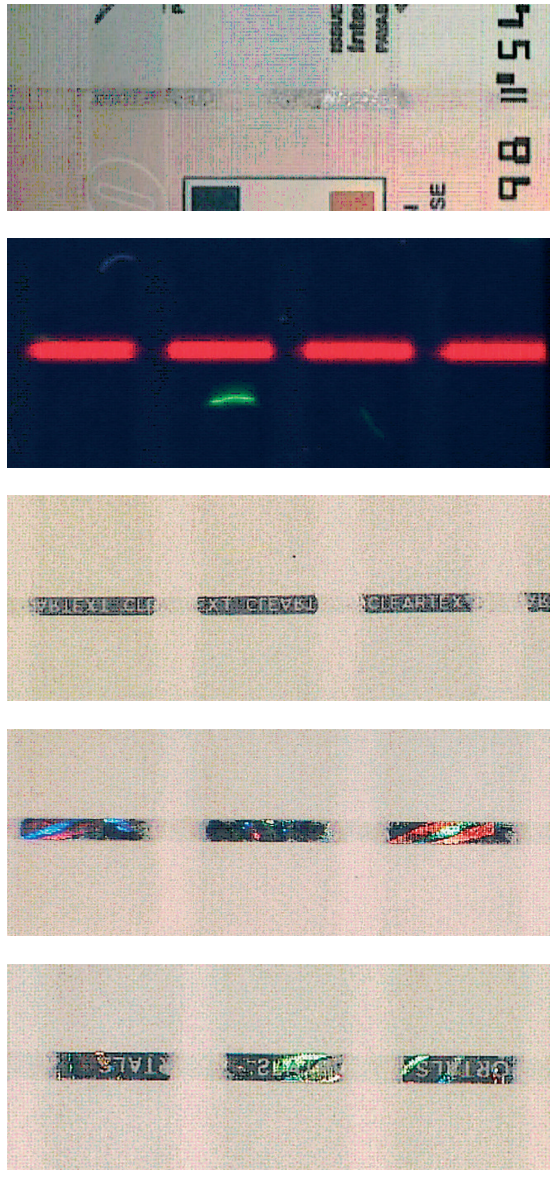


Figure 11 Examples of the different types of security threads from left to right: (a) demetallized, (b) holographic, (c) microtext, (d) fluorescent, and (e) a counterfeit

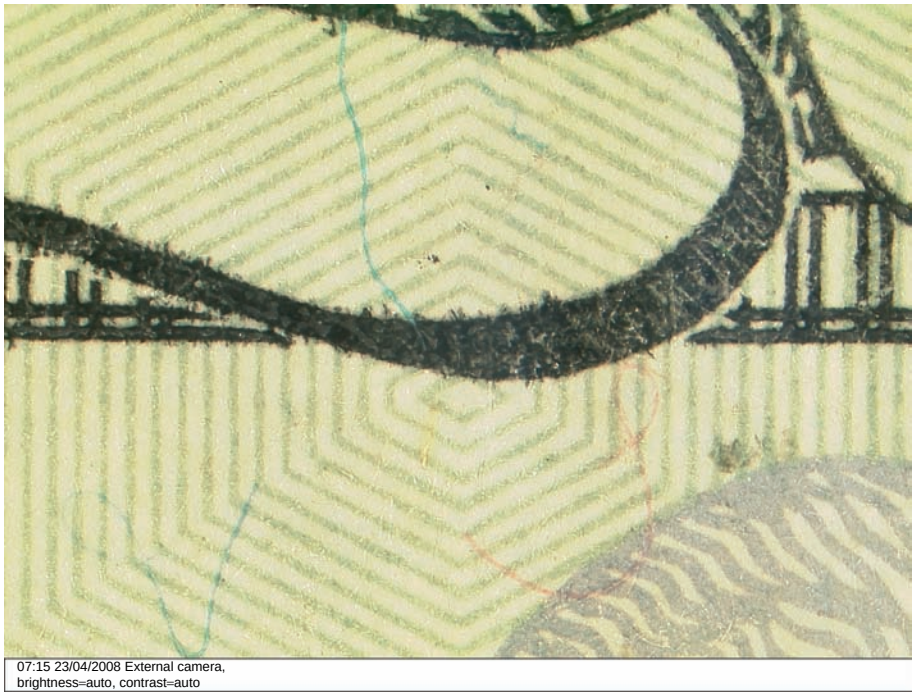


Figure 12 Red and blue security fibers found in a Federal Reserve Note

Security Features Added to a Document.

Optical variable device (OVD)
Optical variable devices (OVDs) can be either iridescent or noniridiscent in nature and include, among other structures, holograms and kinegrams. A hologram is a compilation of images, typically etched by laser into a photographic plate that forms a three-dimensional rendering of a photograph of an actual object, maintaining its original dimensions (Figure 15a, b). Color and conformational effects can be noted as the image is rotated through various angles of viewing. A kinegram is a two-dimensional vector image that appears to change in color and/or conformation when rotated and has the advantage of being visible at almost any viewing angle, even in low lighting conditions. OVDs are very difficult for the counterfeiter to successfully simulate and cannot be duplicated using electrophotographic methods.

Security laminate
– A coating, typically a foil or film, applied as a physical protective barrier for a document, as well as a means to ensure its authenticity. These coatings may involve the use of glass microbeads and thin interference layers of films that form a combination of varying refractive indices in order to produce a retroreflective effect. Retroreflective materials reflect light directly back at the source rather than diffusely, such that the angle of illumination must be equivalent to the angle of viewing for the feature to be visible. Therefore, retroreflective features that are integrated into a security design are not visible to the unaided eye and are best observed using a retroviewer, which allows the eye to be in alignment with the apparent angle of illumination, typically through the use of a semitransparent mirror [15].

Security threads
Thin ribbons, typically polyester, that may be metallized/demetallized or coated with pigments or other materials, resulting in their being rendered opaque, transparent, or fluorescent. Security threads may be designed such that they are only viewable using transmitted light or may contain microprinting. Inclusion of a security thread will result in a discernible increase in thickness of the document at the point of insertion. While optical properties may potentially be duplicated, simulated threads often are formed using

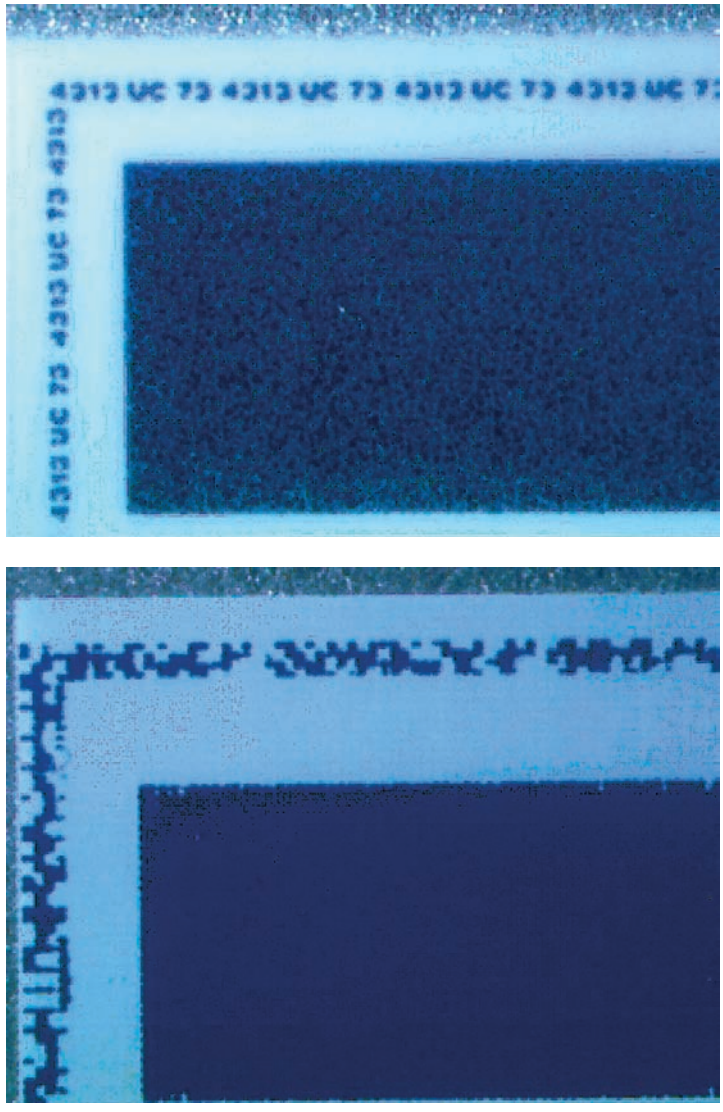


Figure 13 Microtext printing of the same area on (a) a genuine and (b) a counterfeit document

a different material, such as paper, or may be printed onto the substrate of a counterfeit document, which will not result in the perception of a difference in texture, key in the identification of a genuine security thread.

Perforations

Incorporating perforated printing, such as with serial numbers or images, in security documents can be especially useful in identifying page substitutions in a multiple page document if the holes are misaligned in one or more pages. The process involves physically creating holes through multiple pages in a single step and can be created *via* machine or laser.

Grommets or eyelets

Rings, eyelets, or rivets, typically metallic, used to secure a photograph to a document and may be engraved with information. Although not very common, grommets are used in some travel documents, such as passports, and are a very useful means of preventing photographic substitution. There may be



Figure 14 An intricate printing pattern, including guilloche and rainbow printing processes

evidence of tampering that can be visualized using microscopy. If the grommet has been substituted with a fraudulent replacement, an elemental analysis of the constituent metal may be warranted.

Security binding threads
Threads used to bind multiple page documents (e.g., passports) using stitching procedures that can be, in themselves, security features. In constructing these booklets, stitching methods more intricate than the conventional single, continuous stitch may be employed (e.g., locking stitches), which may be difficult for the counterfeiter to readily reproduce. When the integrity of the binding is in question, physical, optical, and chemical examinations can be conducted in order to determine the stitching method utilized, or if a thread has potentially been replaced or unbound.

Embossing and debossing
Processes used to create three-dimensional images in a substrate, typically formed using a die (and its counter die) with the application of heat and/or pressure. *Embossing* is a term that can be used to describe the formation of an image above the surface of a substrate (raised), while debossing results in an image inset below the surface (impressed). Embossed images may be coated in foil (e.g., tipping), as typically found with the raised numbers on a credit card. In some cases, the FDE may be requested to conduct an examination on embossing equipment as part of the analysis.

Foil stamping
Foil stampings may be colored, iridescent, or holographic and therefore serve as effective anticopying features. When observed with a microscope and sufficient oblique lighting, the foil should appear to be impressed below the surface of the document. The examination of counterfeit documents may reveal evidence of tampering, and, in many cases, foils are fraudulently constructed and attached to documents by means of an adhesive. In these cases, the debossing characteristic of this process will not be present.

Laser engraving
The beam emitted from a laser can be used to inscribe, or engrave, text or images into a document. Although laser engraving is not an etching process, which requires chemicals or machinery, it employs the same basic concept and can be used on a variety of substrates, such as plastics and metals. The effect is generally achieved using a computer-controlled operation, capable of precise engraving in three dimensions, a process termed *vector engraving*. This type of laser engraving results in a tactile quality being imparted to the document that can serve to be as distinct a component of the feature as are the contents inscribed. Pixel mode engraving can also be utilized, which allows for very high resolution characters or images to be imbedded in a document, without imparting the three-dimensional quality.

Chemical reactions
Involve the impregnation of a substrate with chemical substances that yield a reaction that can serve to prove either that a document is genuine or that it has been altered. A common method employed in altering documents involves the use of a strong oxidizing agent, such as sodium hypochlorite, which when combined with certain chemically treated security papers, results in the irreversible formation of a brown stain. Use of other chemical agents to remove inks can result in the development of other colors that are solvent-specific. Optical examinations, especially using a filtered light source, generally allow the FDE to detect such alterations.



(a)

(b)

Figure 15 Images of (a) a genuine and (b) a counterfeit hologram. Note the misspelling of MasterCard as "MasteCard" in the fraudulent version

Electronic Media

Electronic media is an all-encompassing term used to describe information on a document that can be authenticated using a machine-readable source. The more common electronic media include magnetic stripes, barcodes (one-dimensional and two-dimensional) optical memory stripes, optical character recognition, machine-readable zones, smart chips, and radio frequency identification devices (RFIDs). These are becoming more common in various types of security documents (e.g., passports, drivers' licenses) and will likely become far more ubiquitous in the future. The forensic examination of

(c) 2011 John Wiley & Sons, Inc. All Rights Reserved.

these features does require the appropriate reading device. For example, electronic information that has been fraudulently encoded into one of the tracks on a magnetic stripe may not correspond to the information embossed on the card. If this is the case, the encoded information should be reported to investigators to determine if the fraudulent information can be linked to an originating source.

Digital watermarks are a form of electronic media and may be categorized as invisible or fragile; these security features serve to protect the integrity of a digital image. Invisible watermarks are designed to be robust, such that information may be embedded in an image without detection. While virtually any electronic data may be contained within the watermark, this is typically restricted to less than 120 bits in size and used for applications such as copyright protection or to identify specifics of origin. Fragile watermarks are also invisible and reside in every pixel of an image. However, these serve to identify regions with which have been tampered, and altering the image results in the localized breakdown of the watermark, such that upon its extraction, areas that have been modified are readily apparent [16].

Biometrics, as they relate to security, pertain to electronic records of physiological or behavioral attributes that can be used to verify the identity of an individual. Biometric data can be placed on or encoded into a document and compared with information in a database in order to authenticate one's identity. Fingerprints and handwritten signatures are two of the more commonly recognized biometrics that can be encountered on documents; however, DNA, speaker identification, hand geometry, as well as iris and facial recognition technologies have all received significant attention, of late. While these features are not necessarily incorporated into modern documents, this does not preclude the potential for their future use. Biometric verification typically occurs in inspections conducted at the point of presentation of the document (e.g., by airport security personnel) and does not normally fall under the guise of a questioned document examination. However, the forensic examiner should be aware of these developing technologies in that biometrics are security features that may serve to establish the genuineness of a document or the identity of an individual. Fundamentally, all biometric authentication systems operate by using a software-based algorithm to match relative points of commonality between known standards and a questioned specimen. The overall patterns of the relative positions of the points within a biometric are then statistically evaluated on the basis of their uniqueness. If needed, these biometrics can be evaluated by a trained expert, and the FDE may have to determine if a signature is an original or digital reproduction, and if so, how the reproduction is printed. Oftentimes, multiple individuals who obtain a counterfeit document from the same source will have the same fingerprint pattern printed on their identity documents. This can occur if an original fingerprint image was scanned, copied, and incorporated into the document as an electronic file. For an extensive review of biometrics, the authors recommend Wayman *et al.* [17].

Conclusion

Fraudulent documents pose some of the greatest threats to modern society in terms of physical security, the integrity of our financial system, and our personal identities, themselves. With the escalating value of personal and financial data to a counterfeiter, along with the rapid evolution of technology and the increased amount of information available *via* the internet, corrupt individuals are becoming increasingly motivated to perpetrate these crimes for the purpose of financial gain. Security, in terms of both the physical document and the electronic information it contains, should be a major concern for governments and private entities alike, in that the impact of document fraud to society is extensive. Although the submission of counterfeit documents for forensic analyses usually takes place after the documents have been used and the crime has been committed, this function serves a particularly important role in combating fraud. Forensic examinations can allow for the detection of fraudulent documents, as well as provide investigative clues regarding the perpetrator(s), how the documents were constructed, if they are associated with other counterfeits, and if they were produced using materials seized from a suspect(s). Also, an FDE is sometimes called upon to provide expert testimony in a court of law to convey the findings of an examination. Indeed, this is a critical stage in the legal process and, therefore, it is incumbent upon the FDE to have a solid foundation in the examination of counterfeit materials so that he or she can provide unbiased and accurate information to the courts.

Acknowledgment

The authors would like to thank Jeffrey A. Payne, Forensic Document Examiner, of the US Secret Service for his review of and contribution to this article.

References

[1] Warner, R.D. & Adams, R.N. (2005). *Introduction to Security Printing*, PIA/GATF Press, Pittsburgh.
[2] Kelly, J.S. & Lindblom, B.S. (2006). *Scientific Examination of Questioned Documents*, 2nd Edition, CRC Taylor & Francis, New York.
[3] National Commission on Terrorist Attacks Upon the United States, (2004). *The 9/11 Commission Report: Final Report of the National Commission on Terrorist Attacks Upon the United States*.
[4] Federal Trade Commission (1997). *FTC Facts For Consumers: Avoiding Credit Card And Charge Fraud*, http://www.ftc.gov/bcp/conline/pubs/credit/cards.pdf.
[5] United States General Accounting Office, (2002). *Identity Theft: Prevalence and Cost Appear to Be Growing*, Mar. Report No.: GAO-02-363.
[6] American Bankers Association (2007). *Deposit Account Fraud Survey Report*.
[7] Social Security Administration (1997). *Report to Congress on Options for Enhancing The Social Security Card*, http://www.ssa.gov/history/reports/ssnreport.html.
[8] ASTM (2007). *ASTM E 444 Standard Descriptions of Scope of Work Relating to Forensic Document Examiners*.
[9] Foster, D.J. & Morantz, D.J. (1979). An electrostatic imaging technique for the detection of indented impressions in documents, *Forensic Science International* **13**, 51–54.
[10] Tolliver, D.K. (1990). The electrostatic detection apparatus (ESDA): is it really nondestructive to documents? *Forensic Science International* **44**, 7–21.
[11] Seward, G.H. (1998). Model for electrostatic imaging of forensic evidence via discharge through Mylar-paper path, *Journal of Applied Physics* **83**(3), 1450–1456.
[12] Seward, G.H. (1999). Practical implications of charge transport model for electrostatic detection apparatus (ESDA), *Journal of Forensic Sciences* **44**(4), 832–836.
[13] Licht, G.L. & Murano, E. (2004). ESDA effects in light of current discussions, *The American Society of Questioned Document Examiners* **7**(1), 7–21.
[14] ISO (2003). *ISO 7810 Identification Cards – Physical Characteristics*.
[15] Van Renesse, R.L. (2005). *Optical Document Security*, 3rd Edition, Artech House, Boston.
[16] Rabbani, M. (2007). *Digital Image Processing*, Rochester Institute of Technology, Rochester,16 November.
[17] Wayman, J.L., Jain, A.K., Maltoni, D. & Maio, D. (eds) (2005). *Biometric Systems: Technology, Design and Performance Evaluation*, Springer-Verlag London, New York.

Related Articles

Alterations: Erasures and Obliterations of Documents
Biometric Devices
Dating: Document
Ink Analysis
Ink Comparison and Interpretation
Interpretation: Document Evidence
Microscopy: High Power
Microscopy: Low Power
Microscopy: Scanning Electron Microscopy
Paper Analysis
Photography: Marks, Impressions, and Documents

DANNA E. BICKNELL AND GERALD M. LAPORTE

Foundation Testimony

In systems of justice where the admission of evidence is highly regulated, as is the case in the United States of America and some other adversary systems of jurisprudence, there often exists a requirement that, prior to the admission of a category of proof, testimony be adduced to show that the evidential prerequisites for admission of that type of proof have been satisfied. This is referred to as *laying the foundation* for admissibility, or authentication of the evidence. The necessary foundational testimony differs according to the type of evidence that the witness is expected to give. The admission of forensic evidence is often subject to such foundational prerequisites.

Thus, before a witness will be permitted to give opinion testimony as an expert, the court must be satisfied that the witness has the qualifications needed for giving the opinion testimony. This is part of laying the foundation for opinion testimony by an

expert witness (*see* **Chain of Possession of Tangible Evidence**).

When it comes to documentary evidence, or proof based on physical objects that have been examined by an expert, there are also special rules that specify the type of foundation testimony that the law requires (*see* **Documents: Authentication of**).

Similarly, any evidence to which objections on hearsay grounds may be raised, but which the proponent asserts comes within an exception to the rule against hearsay, must be shown to have been obtained under circumstances that justify its admission (*see* **Hearsay Evidence**; **Best Evidence Rule**; **Learned Treatises as Evidence**).

The required foundation testimony is explained in the separate articles referred to above.

ANDRE MOENSSENS

Fourth Amendment *see* Police Use of Force

Free Will *see* Compulsion

Friction Ridge Examination (Fingerprints): Interpretation of

Introduction

This article concentrates on the interpretational issues associated with fingerprint evidence or, more generally, evidence obtained from friction ridge skin comparison. The basic premises of the field, namely, the fact that [1] friction ridge skin arrangements are *extremely discriminating* and reflect the variability between donors, and that [2] friction ridge skin arrangements are *persistent*, have been covered elsewhere (*see* **Friction Ridge Skin: Comparison and Identification**). The same apply to the protocol analysis, comparison, evaluation and verification (ACE-V) used as a comparison methodology by most examiners. Only the main decisions of the ACE process will be discussed here.

Main Decisions Sought after During the Comparison Process (ACE)

Analysis allows an assessment of the quality of the mark and factors such a distortion, pressure, amount of residue, and the nature of the substrate and of the detection technique used. These factors can help set the *tolerances* that the examiner will have to allow when considering a potential correspondence. The analysis stage also leads to a decision regarding the capacity of the mark to be *compared* against known prints. In some instances, this conclusion is limited only to whether or not the mark is expected to be identified should the corresponding source be available. This decision is not within the scope of this article.

The comparison stage is essentially factual and leads to a documentation of the features that have been found in correspondence and the features that have been described as dissimilar.

Evaluation is the fundamental inferential step of the process, and is the topic of the present discussion. It will lead to the formulation of a conclusion, which may take three forms in the discipline: *individualization* (generally also referred to as *identification*), *exclusion*, and *inconclusive* [1].

Exclusion means that the mark and the print(s) do not have the same source. Observed differences in ridge flow (for example, the mark being a loop and the print showing a whorl pattern) will enable an examiner to exclude the print from being from the same friction ridge skin area as the mark. In fact, as soon as a difference is observed between the mark and the print that cannot be reconciled in the light of the tolerances defined during the analysis stage, an exclusion conclusion will be reached. The examiner would logically come to the deductive conclusion of exclusion upon assessing that the probability of a

match is zero, if the mark under consideration truly had been left by a designated friction ridge skin area. How an examiner assigns a probability of zero to the observation at hand is sometimes without debate (e.g., when the general pattern between the mark and the print cannot match). Sometimes, owing to, for example, a lack of clarity, the decision is more difficult to substantiate. Hence, a one-unexplained-discrepancy rule is not a standard. In fact, this rule is ill defined by a circular argument [2].

The absence of any irreconcilable discrepancy is a prerequisite for identification. In that case, the examiner will weigh the corresponding features in reference to the standards for identification (see below). An individualization conclusion is commonly understood to mean that a mark has been attributed to a particular individual to the exclusion of all others [3]. "Others" refers often to any human in the world living or dead. When that was the examiner's express or implied assertion, he obviously operates by tradition within the *earth population* paradigm. The relevance of such a paradigmatic posture has been discussed elsewhere (*see* **Identification and Individualization**). Another way to describe this approach, *borrowed from the practice of toolmarks examination*, is by saying that individualization will be reached when the examiner observes a level of agreement (across all legible features) that exceeds the highest level of correspondence he observed through his/her training and experience in comparisons involving *nonmatching* entities. An identification is then concluded when the mark shows sufficient quality (clarity) of friction ridges in agreement with the print so that the probability for such a match to happen if a print from another source is submitted is deemed to be impossible [4].

For all other cases, the comparison will be declared as "inconclusive" – indicating that none of the previous conclusions (individualization or exclusion) has been reached – even if there are significant and perhaps highly probative correspondences between the mark and print. Corroborative evidence of this type (i.e., less than "certain") based on friction ridge skin impressions is rarely brought to the attention of the Judiciary despite its potential to address the issue of identity. A policy (and essentially unscientific) resolution of the International Association for Identification (IAI) has enforced that view [5]. Unfortunately, such a cautious approach precludes the fact-finder from potentially very strong corroborative evidence [6, 7]. This situation arises from the misconception that fingerprint evidence must be categorical and from the unwillingness of the majority of examiners to accept the relevance of probabilistic evaluation.

Standards for Identification

There is variability regarding the identification standards applied by law-enforcement agencies. In a nutshell, we can distinguish between the empirical and the holistic approach (according to the distinction made in [8]).

The Empirical Approach

This practice calls upon the examiner to require a minimum number of minutiae (or points) in the agreement (without discrepancies) between a mark and a print before concluding on an individualization. The practice is still widespread, but with the number of points required varying from one country to another. Table 1 (updated from [9]) lists a few options.

None of these standards have a meaningful scientific or statistical underpinning (see [10] for an analysis of the 16-point standard). They represent an adopted, often dogmatic, practice used by fingerprint examiners. The 12 (or 8–12)-point rule can be traced to the pragmatic guidance given by Locard in 1914 [11]. The 16-point rule is associated with the very early and premature statistical work by Balthazard [12] and the misleading interpretation of a paper by Bertillon [13, 14].

Table 1 Numerical standards adopted in a selection of countries

Numerical standard	Country
16–17	Italy
8–12	Germany
10–12	Holland
12	Belgium, France, Israel, Greece, Poland, Portugal, Romania, Slovenia, Spain, Turkey, and South American countries
7	Russia

From a logical perspective, there is no argument to recommend any predetermined minimum number of features for the following main reasons (see [15] for additional statistical data):

- The relative frequency of general flow varies greatly from class to class. The empirical approach makes no distinction between general patterns.
- Minutiae frequency varies greatly as a function of their type and their position. Hence any system suggesting a fixed addition of points cannot be supported from a statistical perspective.
- When quality allows, features (frequently referred as level three features) such as pore positions and shapes (including edges) add to the identification process. No numerical standard accounts for them.

Such numerical standards are not an objective standard for identification, but rather an ill-defined and locally adopted quality control measure. In fact, while facing the same set of images (marks and prints from the same source), examiners differ extensively in the total number of annotated corresponding features [16, 17]. As stated by a group of European experts working under the auspices of Interpol: "The decision whether the total volume is sufficient for individualization may differ according to the approach. It is not the calculation of a number of points that automatically leads to identification. It is the total constellation of all the information, as a coherent complex of which the relations are the same, and the details, as far as present, fit within tolerances, which constellation is weighed and referenced with individual knowledge with or without an empirical standard" [18], p.28. That leads to the holistic approach.

The Holistic Approach. Following a three-year review, the IAI took a resolution that set the scene for the holistic approach since 1973. The resolution, slightly modified in Ne'Urim, states: "No scientific basis exists for requiring that a predetermined minimum number of friction ridge features must be present in two impressions in order to establish a positive identification." [19]. The examiners from the United States of America, Canada, Switzerland, the United Kingdom, and Nordic countries (Norway, Finland, Sweden, and Denmark) follow this approach.

The approach calls for an assessment by the examiner of each comparison based on its own merit (in terms of quality and quantity as revealed by the mark and the print). The expert concludes on an individualization when he/she is satisfied that there is "sufficient" correspondence (or sufficient discordance in the case of an exclusion) between the compared images. This informed judgment is based on training, experience, and expert knowledge. Allowing the whole range of features to be accounted for in the identification process is referred to as *ridgeology* [20].

The inferential steps taken by fingerprint examiners to reach their conclusions remain unclear and there is an important risk in this "opinionization" of the fingerprint field [21]. The main one is the lack of transparency in the inferential process.

It must be stressed that the inferential process is probabilistic in nature [6]. Currently there is no scientific way to substantiate probabilities of an adventitious match on an order of magnitude that would prevent a discussion about the probability of a coincidental match. Hence categorical conclusions of individualization (within this *earth population* paradigm) cannot be substantiated (*see* **Identification and Individualization**). The fingerprint profession has to recognize that conclusions of identification to the exclusion of all others are essentially outside the realm of the expert witness. His/her testimony should limit itself to expressing the contribution of the findings in favor of one proposition (identity of sources) versus another (nonidentity of sources). Here lies the first element of a "paradigm shift," to take the expression of Saks and Koehler [22].

From Identification to Activity Issues

The examination of marks allows the association of an individual with a particular surface or item. A contact between the donor and the surface or item is a prerequisite. At present, there is no scientific method that allows a reliable estimate of the timeframe in which the mark was left on the surface. Also, there is currently no reliable means of assessing the significance of the absence of the detection of marks on a given surface to indicate whether or not the object has been touched with bare hands. However, consideration of the position of the detected marks on the objects may help a court in assessing how

an object was handled, thereby providing potential information on the nature of the activities involved.

Bias, Errors, and Challenges

An analysis of the issues of bias associated with image comparisons is given by Thompson (*see* **Interpretation: Observer Effects**). Empirical studies have confirmed that observer effects can influence the interpretation and conclusions associated with fingerprint examination (see [23] and associated references). The misattribution of a mark recovered in the conduct of the Madrid bombing has been attributed in part to confirmation bias [24].

The fingerprint profession has claimed for many years that a misidentification was not a possible outcome (with the exception of intentional or fraudulent evidence). A recent publication provides accounts of 22 cases of misidentifications [25, 26]. Although Cole [25] suggests that these cases represent the tip of the iceberg, the accuracy of this statement remains nonvalidated. Error rates estimated through proficiency testing suggest that the case of a wrong individualization is a rather rare (although existent) event [26–29].

US court challenges (following the impact of the *Daubert* decision and its progeny) (*see* **Expert Opinion: United States**) have been well covered by Langenburg (*see* **Friction Ridge Skin: Comparison and Identification**). Associated with the court cases is an important body of scholarly publications questioning the reliability of fingerprint identification and its underpinning process, ACE-V (see [30] with discussions [31–34]). A recent paper by Cole can serve an excellent source for the relevant references [35]. Most of the arguments by these scholars are have merits and will be resolved partly when fingerprint evidence will adopt a reporting practice based on a probabilistic approach substantiated through adequate statistical research.

Statistical Models

Although probabilistic research has been sporadically present in the forensic literature associated with fingerprints (see [36, 37] for critical reviews), the fact that the practitioners ruled out probabilistic statements from their practice has led to the more and more untenable situation of a field claiming individualizing power without any firm statistical backing. However, systematic research on the selectivity of fingerprint features has always pointed toward an extreme selectivity of these features, even when partial and limited information is considered (e.g., [38–40]). Such models can serve readily to assess the weight of evidence (in terms of likelihood ratio) associated with fingerprint evidence. This is the second element of a paradigm shift: promoting systematic research on the selectivity of the fingerprint features, and hence moving from the sole examiner's informed judgment to the actual statistical selectivity that fingermarks may provide. As Saks and Koehler aptly put it: "Forensic identification scientists can help themselves immediately by forswearing exaggerated, definitive conclusions in favor of humbler, scientifically justifiable, and probabilistic conclusions" [41, p. 219.]

References

[1] Scientific Working Group on Friction Ridge Analysis Study and Technology (SWGFAST) (2004). Standards for conclusions, *Journal of Forensic Identification* **54**(3), 358–359.

[2] Thornton, J.I. (1977). The one-dissimilarity doctrine in fingerprint identification, *International Criminal Police Review* **32**(306), 89–95.

[3] Scientific Working Group on Friction Ridge Analysis Study and Technology (SWGFAST) (2002). *Friction Ridge Examination Methodology for Latent Print Examiners, ver.* 1.01 [updated 2002; cited], http://www.swgfast.org/Friction_Ridge_Examination_Methodology_for_Latent_Print_Examiners_1.01.pdf.

[4] Stoney, D.A. (1991). What made us ever think we could individualize using statistics, *Journal of Forensic Science Society* **31**(2), 197–199.

[5] *Identification News* (1980). Resolution VII amended, **30**, 3.

[6] Champod, C. & Evett, I.W. (2001). A probabilistic approach to fingerprint evidence, *Journal of Forensic Identification* **51**(2), 101–122.

[7] Champod, C. (1995). Locard, numerical standards and "Probable" identification, *Journal of Forensic Identification* **45**, 132–159.

[8] Groupe d'experts européens d'Interpol sur l'identification par les empreintes digitales-IEEGFI, Interpol (2000). Recherche d'une methode d'identification par les empreintes digitales, *29th European Regional Conference*, Reykjavik, 17–19th May 2000.

[9] Champod, C. (2000). Fingerprints (Dactyloscopy): standard of proof, in *Encyclopedia of Forensic Sciences*,

J. Siegel, P. Saukko & G. Knupfer, eds, Academic Press, London, pp. 884–890.

[10] Evett, I.W. & Williams, R.L. (1995). A Review of the sixteen points fingerprint standard in England and Wales, *Fingerprint Whorld* **21**(82), 125–143.

[11] Locard, E. (1914). La preuve judiciaire par les empreintes digitales, *Archives d'anthropologie Criminelle, de Médecine Légale et de Psychologie Normale et Pathologique* **29**(245), 321–348.

[12] Balthazard, V. (1911). De l'identification par les empreintes digitales, *Comptes Rendus des Séances de l'Académie des Sciences* **152**, 1862–1864.

[13] Bertillon, A. (1912). Notes et observations médico-légales – Les empreintes digitales, *Archives d'Anthropologie Criminelle, de Médecine Légale et de Psychologie Normale et Pathologique* **27**, 36–52.

[14] Champod, C., Lennard, C. & Margot, P.A. (1993). Alphonse Bertillon and dactyloscopy, *Journal of Forensic Identification* **43**, 604–625.

[15] Champod, C., Lennard, C.J., Margot, P.A. & Stoilovic, M. (2004). *Fingerprints and other Ridge Skin Impressions*, CRC Press, Boca Raton.

[16] Evett, I.W. & Williams, R. (1996). A review of the sixteen points fingerprint standard in England and Wales, *Journal of Forensic Identification* **46**(1), 49–73.

[17] Langenburg, G. (2004). Pilot study: a statistical analysis of the ACE-V methodology – analysis stage, *Journal of Forensic Identification* **54**(1), 64–79.

[18] Interpol (2004). *Interpol European Expert Group on Fingerprint Identification II – IEEGFI II. Part 2: Detailing the Method Using Common Terminology and Through the Definition and Application of Shared Principles*, Interpol, Lyon.

[19] Margot, P. & German, E. (1996). Fingerprint identification breakout meeting, in *Proceedings of the International Symposium on Fingerprint Detection and Identification*, J. Almog & E. Springer, eds, Israel National Police, Ne'urim, p. 21.

[20] Ashbaugh, D.R. (1999). *Qualitative-Quantitative Friction Ridge Analysis – An Introduction to Basic and Advanced Ridgeology*, V.J. Geberth, ed, CRC Press, Boca Raton.

[21] Cole, S.A. (2008). The 'Opinionization' of fingerprint evidence, *BioSocieties* **3**, 115–113.

[22] Saks, M.J. & Koehler, J.J. (2005). The coming paradigm shift in forensic identification science, *Science* **309**, 892–895.

[23] Dror, I. & Rosenthal, R. (2008). Meta-analytically quantifying the reliability and biasability of forensic experts, *Journal of Forensic Sciences* **53**(4), 900–903.

[24] United States Department of Justice, Office of the Inspector General – Oversight and Review Division (2006). *A Review of the FBI's Handling of the Brandon Mayfield Case (unclassified and redacted)*, United States Department of Justice, Office of the Inspector General, Washington, DC.

[25] Cole, S.A. (2005). More than zero: accounting for error in latent fingerprint identification, *The Journal of Criminal Law and Criminology* **95**(3), 985–1078.

[26] Cole, S.A. (2006). The prevalence and potential causes of wrongful conviction by fingerprint evidence, *Golden Gate University Law Review* **37**, 39–105.

[27] Haber, L. & Haber, R.N. (2004). Error rates for human latent fingerprint examiners, in *Automatic Fingerprint Recognition Systems*, N.K. Ratha & R., Bolle, eds, Springer Verlag, New York, pp. 339–360.

[28] Wertheim, K., Langenburg, G. & Moenssens, A. (2006). A report of latent print examiner accuracy during comparison training exercises, *Journal of Forensic Identification* **56**(1), 55–93.

[29] Koehler, J.J. (2008). Fingerprint error rates and proficiency tests: what they are and why they matter, *Hasting Law Journal* **59**, 1077–1098.

[30] Haber, L. & Haber, R.N. (2008). Scientific validation of fingerprint evidence under Daubert, *Law Probability and Risk* **7**(2), 87–109.

[31] Mnookin, J.L. (2008). The validity of latent fingerprint identification: confessions of a fingerprinting moderate, *Law Probability and Risk* **7**(2), 127–141.

[32] Haber, L. & Haber, R.N. (2008). Experiential or scientific expertise, *Law Probability and Risk* **7**(2), 143–150.

[33] Cole, S.A. (2008). Comment on 'Scientific validation of fingerprint evidence under Daubert', *Law Probability and Risk* **7**(2), 119–126.

[34] Champod, C. (2008). Fingerprint examination: towards more transparency, *Law Probability and Risk* **7**(2), 111–118.

[35] Cole, S. (2008). Out of the Daubert fire and into the Fryeing Pan? Self-acceptance versus meta-expertise and the admissibility of latent print evidence in frye jurisdictions, *Minnesota Journal of Law, Science & Technology* **9**(2), 453–541.

[36] Kaye, D.H. (2003). Questioning a courtroom proof of the uniqueness of fingerprints, *International Statistical Review* **71**(3), 521–533.

[37] Stoney, D.A. (2006). Fingerprint identification – scientific status S 34-2, in *Modern Scientific Evidence: The Law and Science of Expert Testimony*, D.L. Faigman, D.H. Kaye, M.J. Saks, J. Sanders & E.K. Cheng, eds, West Publishing, St. Paul, pp. 339–371.

[38] Neumann, C., Champod, C., Puch-Solis, R., Meuwly, D., Egli, N., Anthonioz, A. & Bromage-Griffiths, A. (2006). Computation of likelihood ratios in fingerprint identification for configurations of three minutiae, *Journal of Forensic Sciences* **51**(6), 1255–1266.

[39] Egli, N.M., Champod, C. & Margot, P. (2007). Evidence evaluation in fingerprint comparison and automated fingerprint identification systems – modelling within finger variability, *Forensic Science International* **167**(2–3), 189–195.

[40] Neumann, C., Champod, C., Puch-Solis, R., Egli, N., Anthonioz, A. & Bromage-Griffiths, A. (2007). Computation of likelihood ratios in fingerprint identification

for configurations of any number of minutiae, *Journal of Forensic Sciences* **52**(1), 54–64.
[41] Saks, M.J. & Koehler, J.J. (2008). The individualization fallacy in forensic science evidence, *Vanderbilt Law Review* **61**, 199–219.

CHRISTOPHE CHAMPOD

Friction Ridge Skin: Comparison and Identification

Introduction

The comparison of fingerprints, and more specifically, the friction ridges that compose a fingerprint, has been a mainstay of forensic science for 100 years. It is perhaps one of the quickest, cheapest, and easiest ways to reliably identify an individual.

The basic premise of fingerprint comparison is rather straightforward. Humans (and simians [1]) bear a series of ridges and furrows on the palmar and plantar surfaces of their epidermal layer of skin. The arrangement of the ridges and the features of the ridges have significant variation, which allows for identification. Furthermore, the fact that the ridge arrangements remain constant throughout our lifetime, makes fingerprint identification a useful technique for personal identification.

With respect to forensic application at scenes of crime, the principles remain the same, but the application can be more difficult. Sweat, secreted from the pores, and other contaminants, such as oils, blood, or dirt, form a residue on the ridges. When we touch a surface, this residue is transferred to a surface, leaving an impression of the friction ridges. This residue is typically not visible to the naked eye and must be developed using a chemical or physical (e.g., fingerprint powder) process. These "invisible" impressions are known as *latent-prints*, or *fingermarks* (or sometimes just "*marks*"). The unknown fingermarks obtained at scenes of crime are compared to fingerprints taken from a known source. Known standards are obtained by applying black ink to friction ridges and recording the ridge detail in a controlled fashion. In more recent times, this has been done electronically by scanning the ridges and creating a digital image.

While there are some differences between North American and European terminology, for the sake of consistency in this entry, the term "fingermarks" refer generally to unknown impressions developed at scenes of crime and "fingerprints" refer to the known standards. The same principles that are discussed apply equally to the ridges of the palms, toes, and feet.

History

There is evidence that fingerprints have been used as a means of personal identification in ancient cultures [2, 3]. Most notably the Chinese and Japanese have used finger and palm prints as a means of "signing" contracts and Chinese artists used their fingerprints as a personal mark of authenticity. Cultures in the Middle East practiced similar behavior. It is unknown what depth of understanding these cultures possessed regarding the nature of friction ridge skin.

Early pioneers of friction ridge skin dotted the sixteenth through eighteenth centuries [4]. However, it was not until the nineteenth century that two individuals, Sir William Herschel and Dr Henry Faulds, first explored the use of friction ridge skin as a means of personal identification and thus recognized its value [5]. Sir William Herschel was an employee of the civil service of India in 1858. He instituted the use of inked finger and palm prints as a means of signing contracts with local natives. He promoted the use of friction ridge identification throughout his civil service career. Courts, jails, and civil offices under Herschel were using friction ridges as a means of personal identification. Simultaneously, thousands of miles away in Japan, Dr Henry Faulds was systematically studying friction ridges. He too realized their value as a means of identification, but also studied (as did Herschel) their persistency. Faulds also recognized that the patterns of fingerprints would allow for classification, and fingermarks left by criminals at crime scenes could identify the perpetrator of the crime. Years later, Sir Edward Henry developed a classification system for the storage and retrieval of fingerprints.

It was Sir Francis Galton, however, who after writing the seminal friction ridge text "Finger Prints"

in 1892, earned the distinction of the father of fingerprints [6]. Galton expounded, through systematic study of friction ridge detail, upon the works of Herschel and Faulds. Galton is credited with the first (albeit quite crude) probability model for fingerprint individuality. Galton continued to study and publish on the topic of friction ridges for many years.

Further study throughout the twentieth century led to an increased understanding of friction ridge skin formation and structure. It is the mechanism of *formation* (morphogenesis) that explains the unique arrangements of friction ridge skin and it is the *structure* of friction ridge skin that explains the persistence of friction ridge skin. Noted scientists such as Drs Bert Wentworth, Harris Wilder, Inez Whipple, Harold Cummins, Charles Midlo, Alfred Hale, Michio Okajima, Sara Holt, and William Babler [7–13] all contributed significantly to the modern understanding of friction ridge skin formation and structure.

Features of Friction Ridges

The features of friction ridges can be categorized into Level 1 detail, Level 2 detail, and Level 3 detail [14, 15]. These levels refer, respectively, to ridge flow, ridge path, and ridge shape.

Level 1 detail is the flow of the ridges. In the fingers, the flow of the ridges can be classified into a pattern type. There are three major pattern types: loops, whorls, and arches (see Figure 1). There are many subclassifications for further pattern refinement as well. The flow of the ridges (and pattern type) are useful class characteristics, but lack strong discriminating power. Therefore, Level 1 detail is useful for exclusion, but not for statements of identity. Furthermore, because of the flexibility of skin, Level 1 detail can be grossly distorted and appear quite different in a fingermark than in the comparable fingerprint. For this reason, one must use caution when utilizing Level 1 detail, even when excluding (see Figure 2).

Level 2 detail is the path of the ridges. If one examines the fingerprint carefully in Figure 3, it can be observed that the ridges are not continuous. They split into two ridges, end abruptly, vary in length, or sometimes appear as a single dot. The occurrence of these events are known as *minutiae*. Minutiae are sometimes referred to as *points of identification*, *Galton points*, *dactyloscopic points*, or just colloquially as *points*.

Traditionally, identification of a source of a fingermark has sprung from a focus on the minutiae. Essentially the more minutiae one found in agreement between a fingermark and fingerprint, the more certain one could be that the impressions shared a common source. Today, examiners recognize that each minutia contains additional information useful in the comparison process, such as the rarity and clarity of the minutiae [16].

Each minutia has four components that must be considered during a comparison. When minutiae in a fingermark and fingerprint are compared, they are said to be in correspondence if all four of the components agree. The first component is the type of minutiae. There are three basic types of minutiae: bifurcations, ending ridges, and dots (Figure 3). There are combinations of these minutiae that are useful as well, such as short ridges and enclosures. The frequencies of minutiae types have been recorded and are useful [17–19]. When a minutia type is rare, it possesses more discriminating power. The second and third components of a minutia are angle and orientation. Minutiae in correspondence must be



Figure 1 Three major classes of fingerprint patterns. (a) Whorl pattern, (b) loop pattern and (c) arch pattern

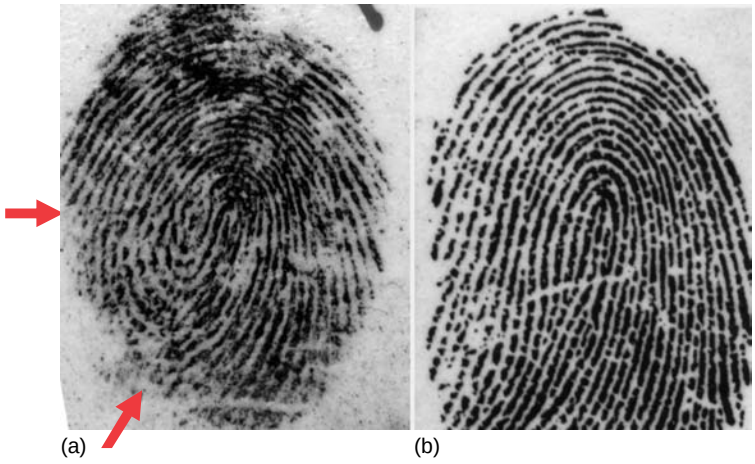


Figure 2 Distortion of pattern type. The fingermark in (a) was developed with black powder. The area between the red arrows is a distortion. The resulting effect gives the appearance of a whorl pattern, when in fact, it can be identified to the loop pattern in (b). [Image courtesy of Michella Snyder, Bureau of criminal Investigation, Richfield, Ohio USA]

oriented in the same direction along a ridge, with the same angle to be in agreement (Figure 4). The final component is ridge count. Minutiae in correspondence must have a similar spatial arrangement. Spatial distances can be distorted due to the flexibility of skin. However, the number of intervening ridges between minutiae is quite robust and cannot typically be altered, even under extreme distortional pressures. Thus the number of ridges between minutiae is counted. It is important that the path of each ridge is followed during a comparison and that the totality of the arrangement of the ridges is considered.

Level 3 detail can be represented by the shapes of the ridges and the relative location of the pores within the ridges (Figure 5). The use of pores for purposes of identification was first proposed by Edmond Locard [20, 21]. He noted that the relative relationships and positions of the pores varied significantly from person to person, and thus could be used to identify the source of a mark. The size and shapes of pores, however, did not vary significantly between persons, nor was this feature robust, since the pores constantly dilate and constrict. Locard did not recommend the use of pore size and shape for identification. Salil Chatterjee first suggested that edge shapes could be used for comparison purposes [22]. Chatterjee categorized edge shapes into eight classifications. These shapes can be compared when the clarity is sufficient and they are reproduced in the



Figure 3 Three major types of minutiae. In this inked fingerprint impression, the ridges are black and the furrows are white. The ridges end and split in various places. These are minutiae. The basic types of minutiae are dots (1), ending ridges (2), and bifurcations (3). Combinations of these three basic types are possible. For example, a spur (4) is an ending ridge in combination with a bifurcation. [Photograph courtesy of Ashley Boone, Bureau of Criminal Apprehension, Saint Paul, Minnesota, USA]

mark and the print. However, due to the flexibility of the skin and the three-dimensional nature of the

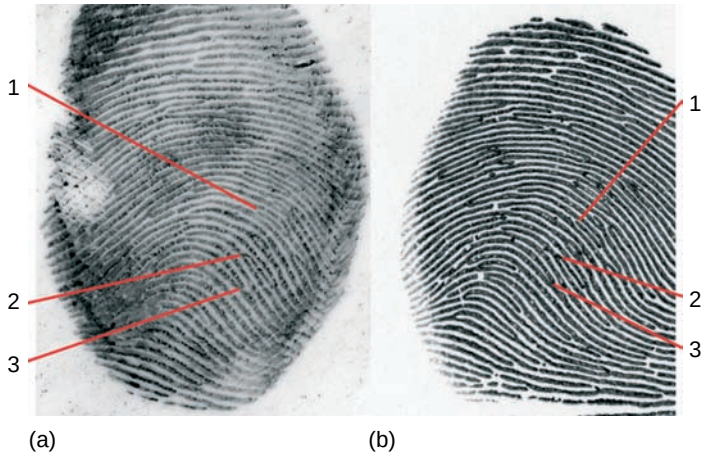


Figure 4 An example comparison of a developed latent fingermark and an exemplar. The image in (a) is a fingermark developed with black magnetic powder. It can be compared to the inked exemplar in (b). Friction ridge features present in the fingermark are sought in the exemplar. When comparing minutiae, they should correspond in location, type, orientation, and angle. For example, compare the minutiae that are marked as "1", "2", and "3" in the fingermark with their respective mates in the exemplar. Note the ridge count between them. [Images courtesy of Jennifer Kostroski, Bureau of Criminal Apprehension, Saint Paul, Minnesota, USA]

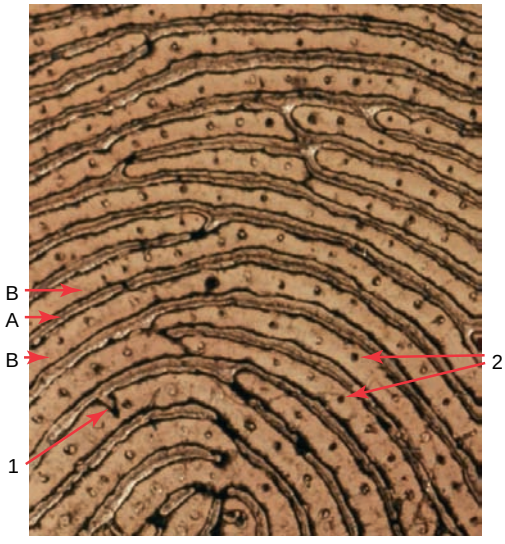


Figure 5 The presence of Level 3 detail: sweat pores and ridge shapes. The above image was produced by applying ink to the friction ridge skin and then pressing the finger against a clear glass plate. The furrows (A) appear slightly darker and thinner than the ridges (B). The ridges are not smooth along the edges, are not necessarily uniform in thickness, and occasionally bear distinct shapes (1). Sweat pores (2) can be seen dotting the ridges as well. [Photograph courtesy of Joshua Bergeron, Bureau of Criminal Apprehension, Saint Paul, Minnesota, USA]

friction ridge edges, it can be problematic to use edge shapes, as their presence can be intermittent or the shapes appear differently due to distortion of the skin from pressure and movement. Similar problems can occur when comparing pores as well. Level 3 detail can be powerful to use when present, but only a small percentage of actual case work contains such fingermarks bearing the highest clarity possible.

There is a class of additional features, sometimes referred to as *occasional features* or *accidental features*, in the friction ridge skin that are useful for identification purposes if present. These include incipient ridges (nascent ridges halted during development), creases and wrinkles, scars, and other skin damage (Figures 6–8).

The levels of detail present will be determined by the clarity of the impressions. The more clarity found in a mark, the more levels that can be observed. Thus, when a mark possesses minimal clarity, all that might be available is Level 1 detail, and perhaps only statements of exclusion can be made. When a mark possesses exceptionally high clarity, the amount of discriminating information contained in the ridges is also very high, and statements of identity can be made with very small areas of friction ridge skin (see Figure 9).



Figure 6 The presence of incipient (nascent) ridges in a fingerprint. Incipient ridges are undeveloped ridges that can be seen in the furrows. They generally lack sweat pores and are thinner than a fully developed ridge

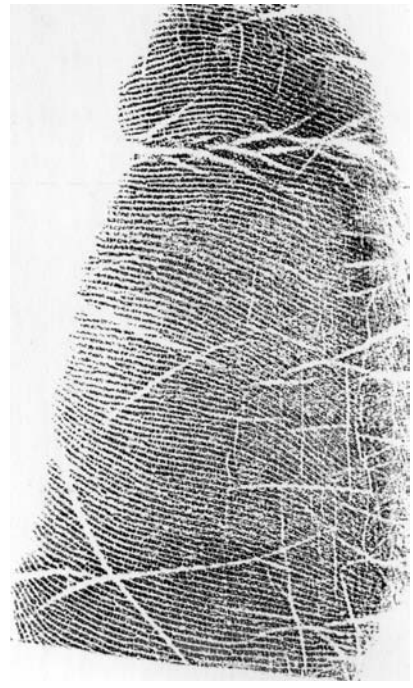


Figure 7 The presence of creases in a palm print. The above image is an inked palm print of the author's right palm (from the hypothenar region, commonly referred to as the "writer's palm" area). Palms typically bear a large number of creases which can be used for identification purposes

ACE-V

Analysis, comparison, evaluation, and verification (ACE-V) are the four stages that represent the overall examination of two friction ridge images (for example, a fingermark from the crime scene and a fingerprint from a suspect). During an examination, a friction ridge specialist first performs an analysis of the fingermark, followed by an analysis of the known exemplar. The purpose of the analysis is to assess the quantity, quality, and specificity of the features present in the fingermark. During the analysis, the specialist typically tries to assess the evidentiary value of the mark (i.e., its intrinsic value for discriminating a unique source). If the mark is determined to possess sufficient value, and the known exemplar is also sufficient for comparison purposes, then the specialist proceeds to the comparison phase. The comparison is the side-by-side comparison of the images. Features present in one image should be found in the other image, in the same relative position, orientation, and number of intervening ridges. Level 1, Level 2, and Level 3 details should all be in agreement given the quality and tolerance of the images. After the side-by-side comparison is complete, the specialist must formulate an opinion about the source of the images. This decision making process is the evaluation stage. While terminology may differ between agencies or countries, fingerprint specialists currently report one of the three general categorical conclusions: individualization (identification), exclusion, and inconclusive opinion [23]. A specialist reports an individualization if he or she has declared a match and believes the images originated from the same source. An exclusion is reported when the images could not share a common source of friction ridge skin. An inconclusive opinion is reported when the examiner could not determine an individualization or an exclusion. Finally, a specialist has another specialist (and in some agencies, multiple specialists) verify his or her conclusion. This check is a quality assurance mechanism and is the final phase of ACE-V known as *verification*. It is recommended by the standards of Scientific Working Group for Friction Ridge Analysis, Study, and Technology (SWGFAST) that at a minimum, all individualizations must be verified by at least one other specialist trained to competency [15]. However, some agencies will have



Figure 8 The presence of permanent scarring in a fingerprint. Fingerprint with a scar. Notice how the ridges pucker inwards, toward the scar. [Image courtesy of Kasey Wertheim, USA]

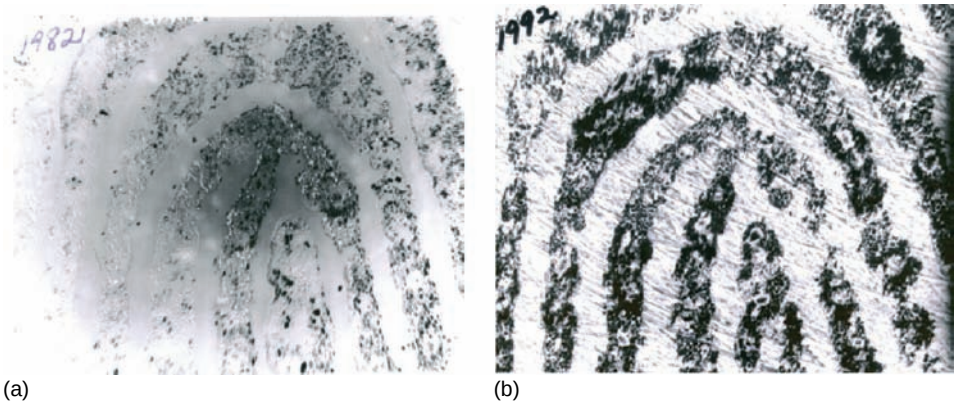
all conclusions, including exclusion and inconclusive opinions, verified.

The ACE-V approach was first suggested by Roy Huber, of the Royal Canadian Mounted Police (RCMP) in 1959 as a protocol for forensic examinations [25]. Michael Cassidy and Harold Tuthill, both of RCMP, adopted ACE-V for footwear examinations [26–27]. David Ashbaugh of RCMP suggested the use of ACE-V for friction ridges soon after [14]. However, wide spread knowledge and articulation of this approach really did not gain notice until the 1990s. In 1999, David Ashbaugh and Stephen Meagher of the Federal Bureau of Investigation (FBI), both testified in the first *Daubert* evidentiary hearing on the admissibility of fingerprint evidence [29]. In this hearing, they stated that friction ridge identification specialists utilize the ACE-V method as a means of comparing friction ridges and that this is the generally accepted method for doing so.

Probability Models

As previously noted, current practice by fingerprint specialists is to report a categorical opinion as a result of ACE-V examination of two images. It should be observed that there are two common practices to reach this opinion: minutiae thresholds (also called *the empirical* or *numerical standard*) and the holistic approach [30].

Many countries employ a minutiae threshold. This is a static number of minutiae that must be observed in correspondence to declare a positive match. The application of this approach can vary from country to country. For example, until recently, sixteen corresponding minutiae were needed to present an identification in court in England and Wales, but a match could be declared for “investigative purposes”,



(a) (b)

Figure 9 A comparison with a low number of minutiae, but a high level of clarity. The image in (a) is a fingermark deposited 10 years before the exemplar in (b) was taken. A high degree of clarity in the images reveals ridge shapes, edge shapes, and sweat pores. [The images above were presented by Stephen Meagher, FBI Latent Print Unit, in the first Daubert hearing for fingerprint evidence (*U.S. v. Byron Mitchell, 1999*)]

if a specialist found less than sixteen, depending on the circumstances of the evidence and the experience of the specialist. Those employing minutiae thresholds may only count minutiae towards the final opinion and these features must be true dactyloscopic minutiae (i.e., not produced by scars, include incipient ridges, etc.) [16].

With respect to the holistic approach, the specialist declares an individualization (or exclusion) opinion when he or she is personally satisfied that there is "sufficient" correspondence (or sufficient discordance in the case of an exclusion) between the compared images. This threshold of sufficiency is said to be based on training, experience, and expert knowledge. The expert evaluates the quantity, quality, and specificity of the ridge details present, utilizing all three levels of detail and any occasional features that may be present (e.g., scars, warts, creases, etc.). The evaluation of all of the features present determines whether "sufficiency" is reached to declare a definitive opinion. Ultimately, the specialist must be convinced that the chance of finding the examined features in someone else, *other than* the person against which the mark has been compared and identified, is so small that it may be discounted [32]. Whilst one cannot refute the importance of expertise, this opinion is subjective and not based *per se* on a statistical calculation for the corresponding features in the compared images.

In fact, the probabilistic approach as a means of reaching a conclusion has been rejected wholesale by profession, while simultaneously applauding the use of probabilities to support the individuality of friction ridge arrangements. The International Association for Identification (IAI), which is the professional organization representing the majority of fingerprint specialists in the world, declared the following resolution in 1980:

> Whereas the delegates of the IAI ... state unanimously that friction ridge identifications are positive, and officially oppose any testimony or reporting of possible, probable or likely friction ridge identifications found on hands and feet. Now therefore be it resolved that any member, officer or certified latent-print examiner who initiates or volunteers oral or written reports or testimony of possible, probable or likely friction ridge identification, or who, when required in a judicial proceeding to provide such reports or testimony, does not qualify it with a statement the print in question could be that of someone else, shall be deemed to be engaged in conduct unbecoming such member ... and charges may be brought [33].

Historically, probability models were explored as a means of providing statistics for a static minutiae threshold. This threshold was a "safe region" above which, once x number of corresponding minutiae were found, an individualization could safely be declared without possibility of duplication within the earth's population. Early models by Galton [6], Henry [35], and Balthazard [36] explored the probabilities of duplication of an entire fingerprint [6, 35,36]. In fact, the work of Balthazard led to the adoption of a 16–17 minutiae threshold, which was adopted by the French and Italian courts, and remains still to this day in France and Italy. More sophisticated models were proposed by Roxburgh [37], Amy [38, 39], Santamaria Beltran [40], Kingston [41], Gupta [42], Osterburg *et al.* [17], Sclove [44] and Stoney and Thornton [45–47]. However, all these models dealt with relatively small sets of fingerprints – the largest was Stoney and Thornton at 412 fingerprints (distal thumb tips). Champod and Margot [19, 48] reported statistical analyses for minutiae from nearly 1000 fingerprints (loop and whorl patterns from fore and middle fingers) using a computerized database [48]. Pankanti *et al.* [49] explored false match probabilities for minutiae in automatic fingerprint identification systems (AFIS databases and their matcher algorithms) [49]. Most recently, Neumann *et al.* [50, 51] reported on the use of likelihood ratios as a means of objectively assessing the evidential weight of corresponding minutiae between a mark and a fingerprint [50–51].

Current work on the validation of this model by Neumann *et al.* has led to the development of software that would allow a fingerprint specialist to annotate corresponding minutiae between two images [52]. The software calculates a likelihood ratio assessing the strength of the evidence, based on the minutiae annotated by the specialist. This statistic represents the probability of the corresponding minutiae under two competing hypotheses: the mark was made by the individual or the mark was made by another individual. Following recent challenges in US courts, government funding agencies in the United States such as Technical Support Working Group (TSWG) and National Institute of Justice (NIJ) supported the development of similar models. A nonexhaustive list of individuals currently developing models includes Champod and Egli at the University

of Lausanne (UNIL) in Switzerland, Meuwly at the Netherlands Forensic Science Institute, and Srihari at the State University of New York (SUNY) in Buffalo, New York. However, even as these models become available, current attitudes from fingerprint specialists have ranged from mild resistance to abject hostility towards the use of probabilities in fingerprint comparison [53].

It is very possible, in parallel developments in the underlying theories and computing power, that such models will be increasingly used by research scientists and practitioners in the near future. In fact, the courts may eventually require it. Note the dissenting opinion of a US appellate judge in *U.S v. Crisp* [54]:

> Fingerprint identification may also be seen as reliable because the examination community prevents its experts from testifying to a match unless they are certain of the match. Fingerprint experts, in other words, refuse to hedge their testimony in terms of probability .... This practice seems to have hastened the technique's acceptance by courts, which have been attracted to its seeming infallibility .... Professions of absolute certainty by an expert witness, however, seem out of place in today's courtroom. Even a DNA match has a small chance of being an error [54].

Legal Challenges

Fingerprint evidence has enjoyed widespread acceptance and use in the courts worldwide for approximately 100 years. However, the 1993 US legal decision in *Daubert et al. v. Merrill Dow Pharmaceuticals* raised the bar for admissibility of scientific evidence [55]. The *Daubert* decision required trial judges to assess the reliability of a scientific principle or method based on five factors. These five factors are:

1. General acceptance: Is the method or principle generally accepted within the relevant community?
2. Testing and validation: Has the method or principle been subjected to testing or validated through testing?
3. Known standards: Does the application of the method adhere to published and accepted standards?
4. Peer review: Has the method or principle been published and subjected to peer review?
5. Rate of error: What is the known or potential rate of error for the method?

Trials judges were asked to assess the proffered evidence under these factors. They were given a wide berth in their application and these are general recommendations for admitting the evidence and not a strict checklist.

In 1999, fingerprint evidence was put to the *Daubert* test in *U.S. v. Byron Mitchell* [56]. After a five day evidentiary hearing, Justice Curtis Joyner ruled to allow the presentation of fingerprint evidence. Upon appeal in 2004, the appellate court upheld the decision [57]. Since the decision, fingerprint evidence has been challenged hundreds of times in the United States. Stricter scrutiny and concerns have been raised by some courts and dissenting opinions, but fingerprint evidence has been mostly accepted.

There are a few exceptions to the admissibility of fingerprints at the time of this writing. In *Massachusetts v. Patterson* [58], the court recognized the reliability of ACE-V methodology as a standard for fingerprint comparison, but did not allow the evidence in the instant case because it was used for the individualization of a simultaneous impression [58]. In this instance, a cluster of fingermarks were observed that appeared to have been deposited simultaneously by the defendant. None of the marks stood alone and could be considered "identifiable" on their own; however, as an aggregate, the features were combined to effect an individualization. The Patterson Court rejected this notion stating that such an application of ACE-V had not been sufficiently tested. In *New Hampshire v. Langill* [59], the trial judge did not allow the fingerprint evidence because the judge found the examiner's documentation to be insufficient to demonstrate the accurate application of the method and because the conclusions of the initial examiner were not verified as a blind test by the verifying examiner (i.e., the second examiner was aware of the conclusions of the first examiner) [59]. The judge rejected the evidence until the analysts could demonstrate that ACE-V had been reliably conducted. The most recent rejection occurred in *Maryland v. Rose* [60], a death penalty case resulting from the shooting of a man during a carjacking [60]. The judge rejected the prosecution's assertion that the reliability of the ACE-V process has been demonstrated during 100 years of operational practice. The judge

also expressed concern about experts testifying to "an infallible method" that when applied correctly has an error rate of zero. This was particularly troubling because no evidence of systematic study of error rates or the possibility of chance duplication of fingermark fragments was presented by the prosecution. The judge felt that since it was a death penalty case, the assertion of 100% certainty, with a zero chance of error simply could not be demonstrated given the lack of research and data.

Future Directions

It is rather remarkable that two of the most important, direction-changing events have occurred nearly simultaneously (relatively speaking) in the fingerprint discipline's century of existence. These two events are the *Daubert* challenges in the United States, followed by the FBI's very public erroneous identification of Brandon Mayfield in the Madrid Train Bombing case [61]. These events have brought new scrutiny from commentators, courts, and academics upon fingerprint specialists and other forensic identification specialists as well.

In turn, forensic professionals have responded with an increased focus on conducting research, improving standards, and moving towards transparency. The Interpol European Expert Group on Fingerprint Identification (IEEGFI) in Europe and the SWGFAST are recognized for developing and promoting international standards. Furthermore, research solicitations and funding by US research groups such as the TSWG and the NIJ have led to important research that is advancing the field.

While the advancements in the field are welcome, they are, in our opinion, severely overdue. They have followed a long period of stagnation and it has taken *Daubert* and the *Mayfield Incident* to stir the pot. There will likely be a period of rapid change and progression in this discipline. These changes may result from the accurence of a series of events that are similar to what occurred within the forensic DNA community in the 1990s. Forensic DNA practices were subjected to significant academic debate and legal scrutiny that ultimately led to the formation of a blue ribbon panel of experts to evaluate the state of the science of forensic DNA. The result led to two reports by the National Research Counsel (NRC) that was sponsored by the National Academy of Sciences in the United States [62]. The recommendations of the NRC led to rapid implementation of standards and international cooperation and consensus to improve forensic DNA practices. Observing the successes of the forensic DNA model, the National Academy of Sciences has formed a new committee, called the *Committee for the Evaluation of the Needs of the Forensic Science Community*, to assess the state of the science in other forensic disciplines, including fingerprints. It is sincerely hoped that the recommendations of this committee will have the same positive effects experienced by the forensic DNA community and open the door for improvements in the fingerprint discipline.

References

[1] Cummins, H. & Midlo, C. (1943). *Finger Prints, Palms and Soles*, The Blakiston Company, Philadelphia, PA, pp. 176–177.

[2] Berry, J. & Stoney, D. (2001). History and Development of Fingerprinting, in *Advances in Fingerprint Technology*, 2nd Edition, H. Lee & R. Gaensslen, eds, CRC Press, Boca Raton, FL.

[3] Ashbaugh, D. (1999). *Quantitative-Qualitative Friction Ridge Analysis*, CRC Press, Boca Raton, FL, pp. 11–19.

[4] Ashbaugh, D.D. (1999). *Quantitative-Qualitative Friction Ridge Analysis*, CRC Press, Boca Raton, FL, pp. 33–41.

[5] Berry, J. & Stoney, D. (2001). History and Development of Fingerprinting, in *Advances in Fingerprint Technology*, 2nd Edition, H. Lee & R. Gaensslen, eds, CRC Press, Boca Raton, FL, pp. 25–31.

[6] Galton, F. (1892). *Finger Prints*, MacMillan & Company, London, pp. 100–113.

[7] Wentworth, B. & Wilder, H. (1918). *Personal Identification*, The Gorham Press, Boston, MA.

[8] Whipple, I. (1904). The ventral surface of the mammalian chiridium, *Zeitschrift für Morphologie und Anthropologie* **7**, 261–368.

[9] Cummins, H. & Midlo, C. (1943). *Finger Prints, Palms and Soles*, The Blakiston Company, Philadelphia, PA.

[10] Hale, A. (1952). Morphogenesis of volar skin in the human fetus, *The American Journal of Anatomy* **91**(1), 147–173.

[11] Okajima, M. (1970). Development of dermal ridges in the fetus, *Journal of Medical Genetics* **12**, 243–250.

[12] Holt, S. (1968). *The Genetics of Dermal Ridges*. Charles C. Thomas, Springfield, IL.

[13] Babler, W. (1979). Quantitative differences in morphogenesis of human epidermal ridges, *Birth Defects Original Article Series* **15**(6), 199–208.

[14] Ashbaugh, D. (1991). Ridgeology, *Journal of Forensic Identification* **41**(1), 16–64.

[15] Scientific Working Group for Friction Ridge Analysis, Study, and Technology (SWGFAST) (2002). *Friction*

Ridge Examination Methodology for Latent Print Examiners, 08-22-2002, v. 1.01.

[16] IEEGFI-II (2004). *Method for Fingerprint Identification; Interpol European Expert Group on Fingerprint Identification II*, Lyon, France.

[17] Osterburg, J., Parthasarathy, T., Raghaven, T. & Sclove, S. (1977). Development of a mathematical formula for the calculation of fingerprint probabilities based on individual characteristics, *Journal of American Statistical Association* **72**(360), 772–778.

[18] Stoney, D. & Thornton, J. (1987). A systematic study of epidermal ridge minutiae, *Journal of Forensic Science* **32**(5), 1182–1203.

[19] Champod, C. & Margot, P. (1995). Computer Assisted Analysis of Minutiae Occurrences on Fingerprints, in *Proceedings of the International Symposium on Fingerprint Detection and Identification*, Israel National Police, Jerusalem, Israel. J. Almog & E. Springer, pp. 305–318.

[20] Locard, E. (1912). Les pores et l'identification des criminels, *Biologica* **2**(24), 357–365.

[21] Locard, E. (1913). La poroscopie, *Archives d'Anthropologie Criminelle, de Médicine Légale et de Psychologie Normale et Pathologique* **28**, 528–546.

[22] Chatterjee, S. (1967). *Finger, Palm, and Sole Prints*, KOSA Publications, Kolkata, India.

[23] Scientific Working Group for Friction Ridge Analysis, Study, and Technology (SWGFAST) (2003). *Standards for Conclusion*, 09-11-2003, v. 1.0.

[25] Huber, R. (1959). Expert witness, *Criminal Law Quarterly* **2**, 276–295.

[26] Cassidy, M. (1980). *Footwear Identification*, RCMP Publication, Ottawa, Canada.

[27] Tuthill, H. (1994). *Individualisation: Principles and Procedures in Criminalistics*, Lightning Powder Company, Salem, OR.

[29] United States v. Byron Mitchell, CA-No. 96-407, Daubert Hearing Transcript, July 7–9 (1999).

[30] IEEGFI (2007). Method for Fingerprint Identification, *29th European Regional Conference, Interpol European Expert Group on Fingerprint Identification II: Reykjavik*, May 17–19, 2000. The document can be found at http://www.interpol.int/public/Forensic/fingerprints/Working Parties/IEEGFI/ieegfi.asp accessed on December 14.

[32] Champod, C., Lennard, C., Margot, P. & Stoilovic, M. (2004). *Fingerprints and Other Ridge Skin Impressions*, CRC Press, Boca Raton, FL, pp. 27–36.

[33] Identification News (1980). Resolution VII amended, *Identification News* **30**(8), 3.

[35] Henry, E. (1900). *Classification and Uses of Fingerprints*, 1st Edition, Routledge & Sons, London, pp. 56–58.

[36] Balthazard, V. (1911). De l'identification par les empreintes digitales, *Comptes Rendus, des Academies des Sciences* **152**, 1862–1864.

[37] Roxburgh, T. (1933). On the evidential value of finger prints, *Sankhya: Indian Journal of Statistics* **1**(50), 189–214.

[38] Amy, L. (1946). Valeur de la prevue en dactyloscopie, *Journal de la Société de Statistique de Paris* **88**, 80–87.

[39] Amy, L. (1946). Valeur de la prevue en dactyloscopie II, *Journal de la Société de Statistique de Paris* **88**, 188–195.

[40] Santamaria Beltran, F. (1955). A new method for evaluating ridge characteristics, *Fingerprint and Identification Magazine* **36**, 3–15.

[41] Kingston, C. (1964). *Probabilistic Analysis of Partial Fingerprint Patterns*, Ph.D. Dissertation, Department of Criminology, University of California, Berkeley.

[42] Gupta, S. (1968). Statistical survey of ridge characteristics, *International Criminal Police Reviews* **5**(218), 130–134.

[44] Sclove, S. (1979). The occurrence of fingerprint characteristics as a two-dimensional process, *Journal of American Statistical Association* **74**(367), 588–595.

[45] Stoney, D. & Thornton, J. (1986). A critical analysis of quantitative fingerprint individuality models, *Journal of Forensic Science* **31**(4), 1187–1216.

[46] Stoney, D. & Thornton, J. (1986). A method for the description of minutia pairs in epidermal ridge patterns, *Journal of Forensic Science* **31**(4), 1217–1234.

[47] Stoney, D. & Thornton, J. (1987). A measurement of fingerprint individuality, *Journal of Forensic Science* **32**(5), 1182–1203.

[48] Champod, C. (1996). *Reconnaissance Automatique et Analyse Statistique des Minuties sur les Empreintes Digitales*, Ph.D. thesis, Institut de Police Scientifique et de Criminologie de l'Université de Lausanne.

[49] Pankanti, S., Prabhakar, S. & Jain, A. (2001). On the Individuality of Fingerprints, *Proceedings of Hawaii International Conference on System Sciences (IEEE)*, Maui, Hawaii, December, pp. 805–812.

[50] Neumann, C., Champod, C., Puch-Solis, R., Egli, N., Anthonioz, A., Meuwly, D. & Bromage-Griffiths, A. (2006). Computation of likelihood ratios in fingerprint identification for configurations of three minutiae, *Journal of Forensic Science* **51**(6), 1255–1266.

[51] Neumann, C., Champod, C., Puch-Solis, R., Egli, N., Anthonioz, A. & Bromage-Griffiths, A. (2007). Computation of likelihood ratios in fingerprint identification for configurations of three minutiae, *Journal of Forensic Science* **52**(1), 54–64.

[52] Champod, C., Neumann, C., Langenburg, G. & Chamberlain, P. (2007). From Fingerprint Examination to Fingerprint Statistics, *Workshops at the 92nd Educational Conference of the International Association for Identification*, San Diego, CA, July 25–26.

[53] McKasson, S. (2001). I think therefore I probably am, *Journal of Forensic Identification* **51**(3), 217–221.

[54] United States v. Patrick Crisp, 324 F.3d 261 (4th Cir., 2003);. No. 01–4953 appeal of CR-01-236, March 31, 2003.

[55] Daubert v. Merrel Dow Pharmaceuticals, Inc., 509 U.S. 579; No. 92–102, (1993).

[56] United States v. Byron Mitchell, CA-No. 96–407, Daubert hearing transcript, July 7–9, and July 12–13, (1999).
[57] United States v. Byron Mitchell. No. 02-2859, On appeal for CA-No. 96–407. April 29, (2004).
[58] Commonwealth of Massachusetts v. Terry Patterson, SJC-09478, appeal of 432 Mass. 767 (2000), December 27, (2005).
[59] State of New Hampshire v. Richard Langill, No. 05-S-1129, January 19, (2007).
[60] State of Maryland v. Bryan Rose, Case No. K06-0545, October 19, (2007).
[61] Stacey, R. (2004). Report on the erroneous fingerprint individualization in the Madrid train bombing case, *Journal of Forensic Identification* **54**(6), 706–718.
[62] National Research Council – Committee on DNA Forensic Science (1996). *The Evaluation of Forensic DNA Evidence*, National Academy Press, Washington, DC.

GLENN LANGENBURG

Friction Ridge Skin: Fingerprint Detection and Recovery Techniques

Introduction

This entry presents an overview of the most common methods used to detect and enhance fingermarks in relation to forensic investigations. The list of techniques presented here is not exhaustive. For a comprehensive review of the subject, including further information on the properties of light and fingerprint chemistry, the reader should consult the publication by Champod and coworkers [1].

Detection of Visible Fingermarks

Visible fingermarks are those in which the papillary ridges are distinct from the substrate when observed with the naked eye. A positive image of the papillary ridges can be produced when the finger has been contaminated with a substance of contrasting color (e.g., blood, wet paint) that is then transferred to the substrate, and a negative image can be produced when the papillary ridges remove material (such as dust or soot) from the surface. Indented fingermarks can be produced when a finger comes into contact with a malleable substance that retains a three-dimensional image of the fingerprint ridges (e.g., window putty, candle wax).

Visible fingermarks can generally be detected without any treatment as long as there is sufficient contrast between the fingermark and its substrate. Weakly visible fingermarks may be enhanced using colored, oblique, or episcopic light illumination. Indented fingermarks can be difficult to visualize or photograph; therefore, good lighting conditions, powdering, or the use of a casting material (such as silicone) can be used to enhance the contrast.

Detection and Enhancement of Latent Fingermarks

Latent fingermarks are those that cannot be readily seen with the naked eye. They are created from the natural secretions and contaminants on the finger that are deposited onto the substrate in the shape of the papillary ridges – in much the same way as a rubber stamp deposits ink in the shape of its raised profile.

In order for a latent fingermark to be visualized, it must be given properties that differentiate it from its substrate. Usually, this is done by giving color or luminescence properties to the fingermark through either optical, physical, or chemical processing, or a combination of these. The most effective approach is to use a sequence of techniques that is appropriate for the type of substrate and the environmental conditions that the fingermarks have been exposed to.

Substrates are generally classified as being porous, nonporous, or semiporous. Porous substrates such as paper and cardboard absorb fluids, including natural secretions that have been deposited on the surface. Nonporous substrates such as glass and most plastics do not absorb such secretions, and thus the fingerprint deposit remains on the surface of the substrate. Semiporous substrates such as expanded polystyrene, latex gloves, or polymer banknotes possess both porous and nonporous characteristics. Substrates can be further divided into those that have been exposed to water, and those that have not. This factor determines the types of secretions that have potentially

washed or diffused away from the fingermark, hence defining the development techniques that can still be effective. In each particular scenario, the most appropriate sequence of reagents and techniques may be different.

Detection Techniques

Optical Techniques

All types of evidence must first be examined using optical techniques because they are nondestructive and can be performed relatively easily either in the laboratory or at the crime scene. In order to understand how optical techniques work, it is necessary to understand some of the basic properties of light.

Light Properties. The perception of color and luminescence is due to the interaction of light with matter. Light is a form of electromagnetic energy that comprises the ultraviolet (UV), visible, and infrared (IR) regions of the electromagnetic spectrum (Figure 1).

The human eye can only see the visible region of the spectrum, which is within the wavelength range of approximately 400–700 nm. Color perception is dependent on the wavelength of radiation; for example, radiation at 450 nm is observed as blue light, 550 nm is green, and 650 nm is red. When all colors of the visible spectrum are present with the same relative intensity, the result is white light (Figure 2).

Light in the visible region can be polychromatic (many colored), or monochromatic (single colored). When a single wavelength is present, there is pure monochromatic light. When a narrow range of visible wavelengths is present, there is a monochromatic band of light.

When a light beam is directed onto a surface, certain wavelengths are reflected from the surface, absorbed by it, or transmitted through it. The reflected wavelengths produce the color that an object appears to have. For example, an object that appears blue under white light reflects light around the wavelength 450 nm, while absorbing most other wavelengths. The same object under monochromatic green light (around 550 nm) appears black if it absorbs all the radiation at 550 nm.

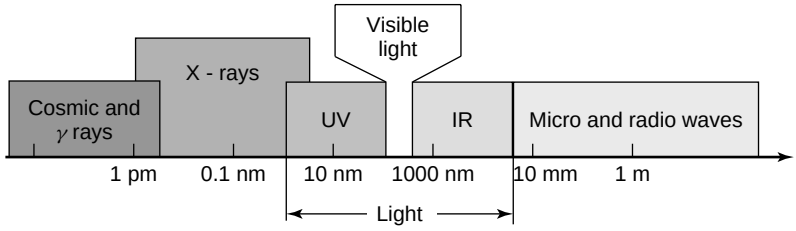


Figure 1 The electromagnetic spectrum. Light consists of ultraviolet (UV), visible, and infrared (IR) radiation [Reproduced from Ref. 1. © Taylor & Francis Group, 2004.]

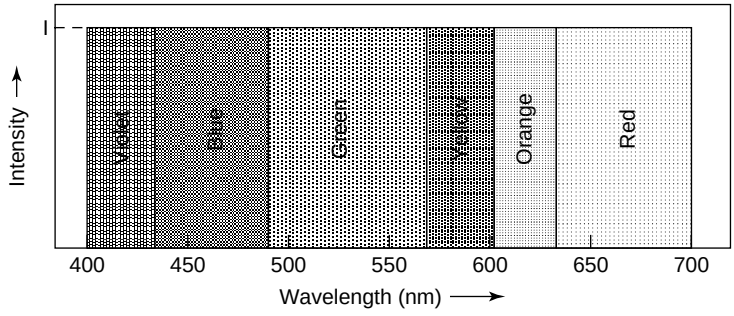


Figure 2 When each color of the visible spectrum is present with the same relative intensity, white light results [Reproduced from Ref. 1. © Taylor & Francis Group, 2004.]

Light Properties and Fingermark Detection. The absorption properties of light can be exploited when enhancing fingermarks. A forensic light source (FLS) that generates strong white light can be trained on the object, and filters can be used in the light path to block certain wavelengths in order to differentiate a contaminated or treated fingermark from its substrate. Laser light sources or alternate light sources (ALS) such as the Polilight® lamp are examples of FLSs used for fingermark detection. The application of UV light can be used in the same way if the substrate and fingermark possess different absorption properties in the UV region; however, a UV imaging system is required to capture the results.

Episcopic coaxial illumination can be effective for the detection of latent fingermarks on smooth, shiny surfaces. This technique relies on the diffused reflection from the fingermark ridges as opposed to the specular reflection from the substrate.

Some substances also possess photoluminescence properties that can assist in fingermark enhancement. This phenomenon occurs when matter absorbs a photon of energy in the visible or UV region and then emits the energy as a photon of light at a longer wavelength. During this complex process, some energy is lost in the form of heat and movement (molecular vibrations), and thus the light emitted is at a lower energy (longer) wavelength. Photoluminescence is a reproducible property, and, with FLSs, it is possible to use filters at particular excitation and observation wavelengths to increase contrast between a photoluminescent fingermark and its nonluminescent substrate. Images of the fingermark can be easily captured using standard photographic equipment and the correct filters. Where the substrate and the fingermark both possess luminescence properties and cannot be readily differentiated using traditional photographic techniques, more advanced imaging techniques can be employed such as fluorescence lifetime imaging (FLIM) [2–6] or chemical imaging [7–9].

In general, latent (or invisible) fingermarks – by their very nature – require additional processing rather than optical techniques alone. It is generally necessary to treat the evidence with a physical or chemical process that imparts color or luminescence to differentiate the fingermark from its substrate.

Preliminary optical examinations can assist in observing contaminants in a fingermark and determine whether a substrate possesses inherent properties that preclude treatment with a particular reagent. For example, if a certain reagent produces fingermarks that fluoresce at a particular wavelength, then the technique is redundant if the substrate strongly fluoresces at the same wavelength (this results in a lack of contrast and the fingermark remains invisible). After treatment with an appropriate reagent, optical techniques can be used for further enhancement.

Further information on light, chemistry, and fingerprint detection principles can be obtained from the publication by Champod and coworkers [1].

Detection Techniques at the Crime Scene

Fingerprint Powders. At the crime scene, fingerprint powders are traditionally used to develop fingermarks on smooth, nonporous substrates. The process of fingerprint development is a physical one, with the powder adhering to the moist, sticky, or greasy substances present in the latent fingermark deposit. Developed marks are photographed and can then be collected using adhesive tape or gel lifters.

While the application of powders is relatively easy and inexpensive, the technique generally lacks sensitivity. Only relatively fresh fingermarks are usually developed because, over time, latent fingermarks dry out and lose their stickiness, making it more difficult for the powder particles to adhere. The technique is also not suitable for developing fingermarks on porous surfaces as the latent deposit is quickly absorbed, making it unavailable to the powder. Items that can be transported back to the laboratory are not usually powdered as more sensitive techniques are available that should be applied where possible.

There are a multitude of different powders and brushes available for fingerprint detection, and the choice is often made according to experience or personal preference. Brushes are usually classified by the types of fibers used to make them (e.g., synthetic fibers, natural fibers, glass fibers, etc.). Traditional fingerprint powders consist of a resinous polymer for adhesion and a colorant (black, white, or gray) for contrast. Many different types of colored, metallic, or luminescent powders are also available commercially [10], and various pigments and dyes have been evaluated for use as fingerprint powders [11–16]. The selection of a powder must take into account the color and luminescence of the substrate. Aluminum flake powder is generally considered to be the most effective powder for general fingerprint development on smooth, nonporous surfaces [17].

Care must be taken when powdering, as excessive contact between the brush and the fingerprint destroys fragile fingermarks, and may result in the transfer of DNA to subsequent articles being powdered [18]. Magnetic powder, applied with a magnetic wand, can minimize such problems [19, 20]. The powder itself forms a "brush" in the magnetic field and the fingermark is developed when the finer powder particles adhere to it.

The Home Office Scientific Development Branch has conducted an extensive study on the most effective combinations of powders and brushes, and have prepared a flow chart to assist crime scene officers to select the most appropriate combination, depending on the type of surface encountered [21].

Small-Particle Reagent (SPR). Small-Particle Reagent (SPR) is a wet powdering technique that can be used to develop latent fingermarks on a wide variety of nonporous items. The process involves immersing or spraying the item with an aqueous suspension of an insoluble powder and then rinsing it with water to remove the excess reagent. The powder suspension, which also contains a low concentration of the detergent, is sensitive to the water-insoluble portion of the latent fingermark and is therefore effective on surfaces that are wet. Conventional SPR is a suspension of molybdenum disulfide particles [22], which gives a dark gray image of the treated fingermark. Particle size is a critical factor in achieving optimum results.

A modified SPR containing iron oxide instead of molybdenum disulfide has been proposed [23]. While it can improve contrast in some cases, comparative tests have shown that the modified formula was not as sensitive or selective as the original formulation [24]. A white SPR formulation based on zinc carbonate powder has also been developed [25]. As with molybdenum disulfide, the particle dimensions of the zinc carbonate affects fingerprint development. A fluorescent SPR powder using basic yellow 40 (BY40) in the original formulation is also available [26]. In this case, fluorescence visualization can best be achieved with excitation at 450 nm and observation using a barrier filter at 550 nm.

Powder suspensions similar to those used for the detection of latent fingermarks on adhesive surfaces (discussed later) have also been investigated for their potential to be applied to other nonporous surfaces [27]. Preliminary results suggest that these types of suspensions work differently from the conventional SPR technique.

Detection Techniques for Porous Surfaces

Ninhydrin. Ninhydrin (2, 2-dihydroxy-1, 3-indandione) reacts with primary and secondary amines to produce a dark purple compound known as *Ruhemann's purple* (Figure 3) [28]. Latent fingermarks generally contain amino acids as they are present in natural skin secretions. When ninhydrin is applied to a porous substrate carrying a latent fingermark, the compound reacts with the amino acids. If the fingermark contains a sufficient amount of amino acids, it becomes dark purple and can be readily viewed under white light. The mechanism for the reaction of ninhydrin has been well studied [29–31], and is given in Figure 4 [1].

Ninhydrin is usually made up in a solution that is applied by dipping or spraying the evidence. The solution contains a small amount of polar solvent (such as methanol or ethanol) to dissolve the ninhydrin, acetic acid to slightly acidify the reaction, and a carrier solvent. Chlorofluorocarbon (CFC) based carrier solvents are ideal for the ninhydrin development of latent fingermarks; however, the CFC ban worldwide has led to various replacement solvents being tested. Formulations based on hexane, heptane, or

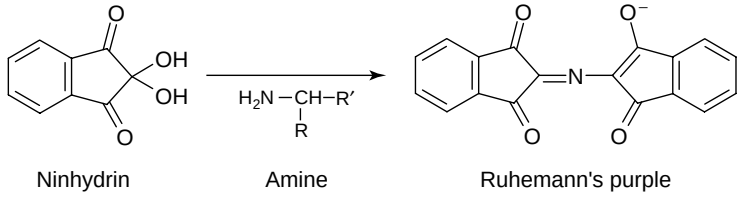


Figure 3 The chemical reaction of ninhydrin and a primary or secondary amine produces Ruhemann' s purple [Reproduced from Ref. 28. © RSC, 1910.]

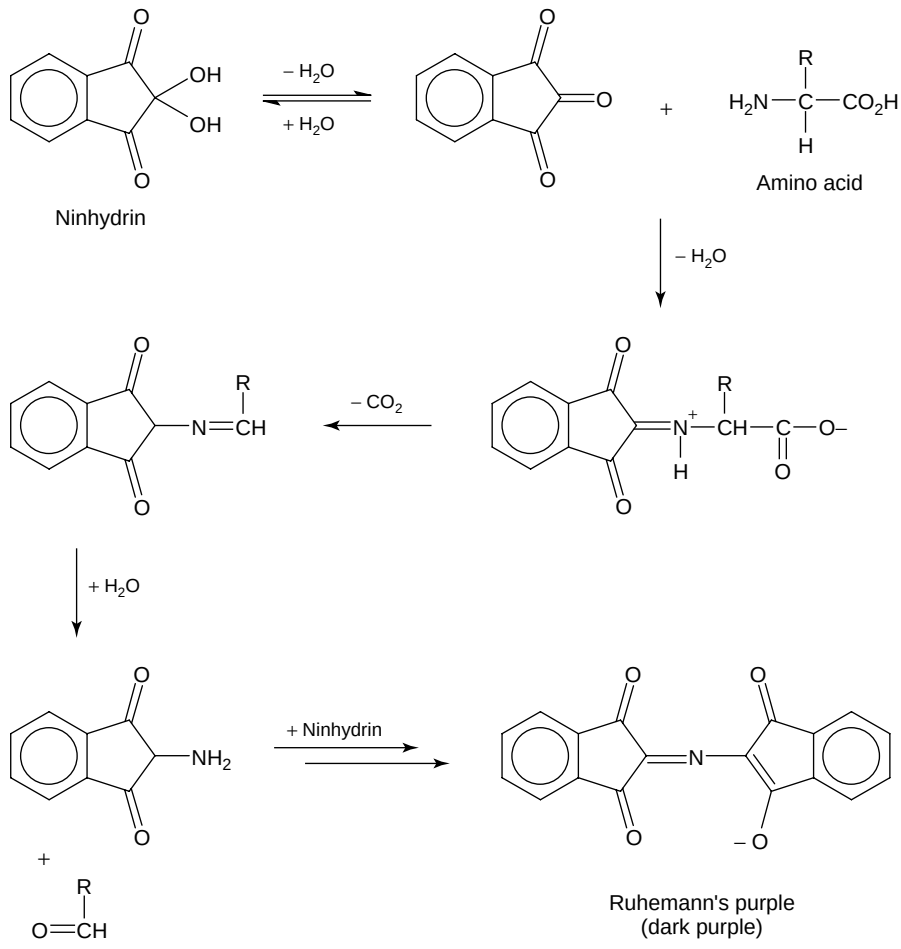


Figure 4 The mechanism of reaction between ninhydrin and amino acids [Reproduced from Ref. 1. © Taylor & Francis Group, 2004.]

petroleum ether have been proposed and have been successfully employed despite the high flammability of these formulations. Currently, the two most used formulations are based on hydrofluoroether (HFE) and hydrofluorocarbon (HFC) solvents.

Treated fingermarks are usually stored for 24 h at room temperature to allow optimal development to be achieved. Visualization is then undertaken in the absorption mode with white light illumination and a green–yellow bandpass barrier filter in front of the camera (e.g., 540–580 nm).

Secondary Metal Salt Treatment. If a fingerprint that has been developed with ninhydrin shows poor contrast, it can be further enhanced by treatment with a zinc or cadmium metal salt solution. The metal salt forms a fluorescent 1 : 1 coordination complex with the Ruhemann's purple (Figure 5) [30, 32]. Zinc nitrate treatment changes the color of the fingerprint to orange, while cadmium nitrate treatment changes the color to red. If the relative humidity is low, exposure to steam may be necessary to complete the reaction.

Visualization in the absorption mode is best achieved for zinc with a 490-nm bandpass filter in front of the camera (or the equivalent illumination from a FLS), and for cadmium with a 505-nm bandpass filter in front of the camera (or the equivalent illumination from a FLS). The fluorescence emissions for both complexes is very weak at room temperature,

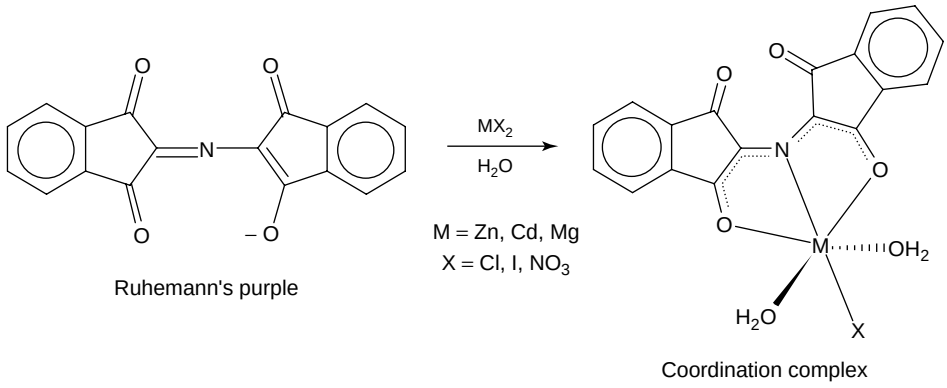


Figure 5 Treatment of Ruhemann's purple with certain metal salts produces a coordination complex [Reproduced from Ref. 32. © Blackwell Publishing Ltd, 1987.]

but this can be significantly augmented by cooling the fingermarks with liquid nitrogen (−196 °C, 77 K). The sample is placed in an insulated container and covered with a thin layer of liquid nitrogen. The luminescence can be seen by illumination in the blue–green (490 nm) region for the zinc complex and illumination in the green (505 nm) region for the cadmium complex, combined with a green–yellow (550-nm) bandpass barrier filter for the zinc complex and an orange (590-nm) bandpass barrier filter for the cadmium complex.

Diazafluorenone (DFO). The compound 1,8-diazafluoren-9-one (DFO) reacts with amino acids to give a product that is pale purple in color (Figure 6). While the color obtained is lighter than that of Ruhemann's purple, the advantage is that the product shows strong room-temperature luminescence without secondary treatment [33]. The reaction product [33] and mechanism [34] for the reaction between DFO and the amino acids present in the latent fingermark deposit appears to be similar to that of ninhydrin with amino acids (Figure 7).

DFO is generally applied by dipping the evidence into a solution of the compound and then heating at 100 °C for 20 min in an oven. Alternatively, it is possible to heat at 160 °C for 20–30 s or at 180 °C for 10 s in an ironing press [35]. Since the CFC ban, DFO formulations based on petroleum ether [36] and HFE7100 [37] have been evaluated. In each case, methanol was found to be an important component of the formulation because it contributes to the formation of a hemiketal that is the reactive species. Alternatively, DFO can be applied using a "DFO-dry" procedure, which involves placing DFO-soaked filter papers onto the evidence and heating it at 100 °C with a steam iron filled with a 5% acetic acid solution. The advantage of this technique is that ink running, damage to the documents, and background staining of the evidence is minimized.

The luminescence of DFO-treated fingermarks may start to diminish around 30 min after heating if they are kept in humid environments due to moisture absorption. However, the luminescence intensity can be restored by reheating (in an oven at 100 °C or in a press at 180 °C). Visualization can be achieved

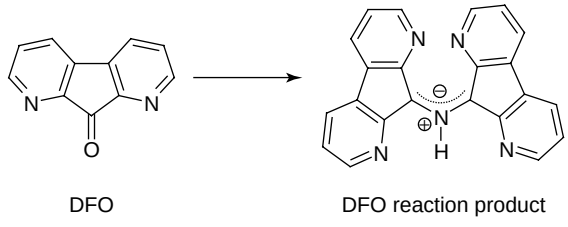


Figure 6 Suggested reaction product between DFO and amino acids [Reproduced from Ref. 35. © Elsevier, 1993.]

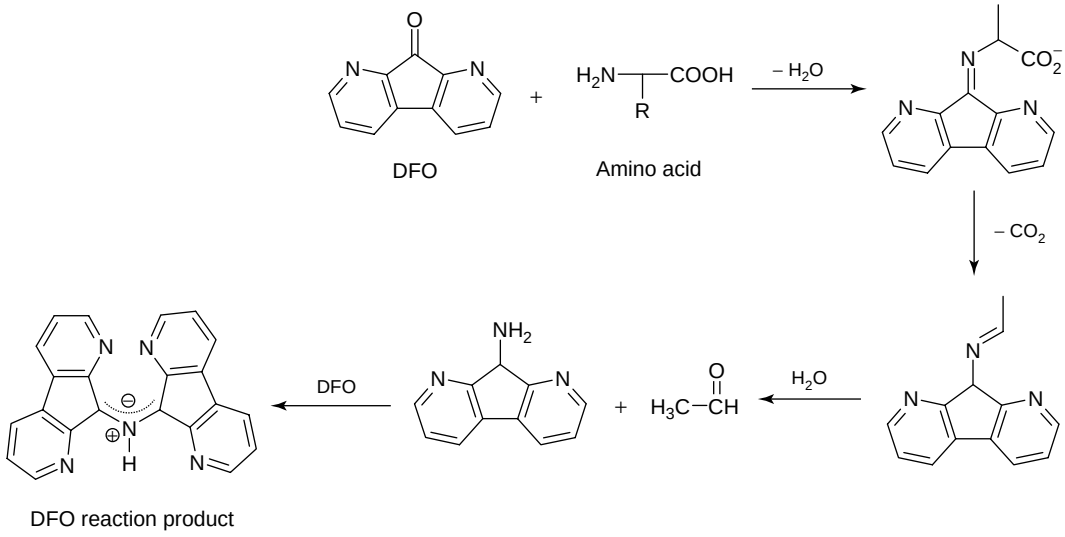


Figure 7 Suggested mechanism for the reaction of DFO and amino acids [Reproduced from Ref. 34. © Elsevier, 2000.]

over a broad excitation range ($\lambda_{ex.} = 430-580$ nm) with an observation filter selected, depending on the excitation wavelength ($\lambda_{em.} = 560-620$ nm). The best results are generally obtained with excitation between 530 and 570 nm.

1,2-Indanedione (IND) and 1,2-Indanedione-Zinc (IND-Zn). 1,2-Indanedione (IND, Figure 8) is a relatively new reagent that also reacts with the amino acids present in the latent fingerprint deposit. The resulting image is a bright pink color with strong luminescence at room temperature. While the color development with IND is inferior to that obtained using ninhydrin, it is an improvement on the color achieved using DFO. The luminescence produced by IND is also far stronger than that produced by DFO, giving it a distinct advantage for the development of fingermarks on porous substrates [38–41]. Other researchers, however, have reported that IND does not perform as well as DFO [42, 43] (Wilkinson, D. *Personal Communication,* 2007). This may be due to the use of different reagent formulations, the impact of different paper substrates, or the inherent sensitivity of IND to certain environmental conditions encountered around the world [40]. Recent studies have highlighted the fact that IND development is affected by both humidity and the nature of the substrate [41].

Excellent results have been reported for an IND formulation that contains zinc chloride [44]. This

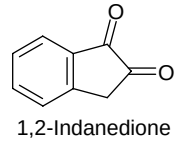


Figure 8 Chemical structure of 1,2-indanedione

combined indanedione-zinc formulation (IND-Zn) does not change the color of the development, as is the case with ninhydrin fingerprints that have been posttreated with metal salts, but allows the treatment to work more reliably, and, in some cases, can significantly improve development. Owing to the similarity in structures between ninhydrin and IND, it was originally hypothesized that the role of zinc was to form a coordination complex with the reaction product. However, recent studies suggest that the zinc is more likely to be a catalyst for the reaction between IND and amino acids [45] (Spindler, X. and Shimmon, R. *Personal Communication,* 2007).

Several formulations for IND [40] and IND-Zn [44] have been evaluated. In contrast to DFO, alcohols are detrimental to the reaction [46]. IND or IND-Zn is applied by dipping or spraying the item, then heating it using a dry heat press at approximately 165 °C for 10 s. Fluorescence visualization can be achieved by using various filter combinations, for example: (i) excitation at 500–520 nm

and observation using a 550-nm long-pass filter; or (ii) excitation at 520–540 nm and observation using a 590-nm bandpass filter.

Genipin. Genipin is a natural product obtained from the Gardinia fruit that reacts with the latent fingermark deposit to produce a dark blue image and room-temperature fluorescence (using excitation at ~590 nm and observation at ~620 nm) [47]. Development of genipin into a viable reagent for latent fingermark detection on paper is being investigated [48]. The potential advantages that genipin may have over ninhydrin and DFO include the following: the combination of color and fluorescence in a single reaction; the excitation and emission properties of genipin, which may give a better signal-to-noise ratio than DFO on paper surfaces that have an inherent fluorescence emission in the 500–600 nm range; and, the very low toxicity of the genipin compound, which is currently used as a food colorant and natural medication.

Physical Developer (PD). Physical developer (PD) targets the water-insoluble portion of the latent fingermark, which makes it suitable for use on porous items that have been wet or previously treated with amino acid–sensitive reagents such as those presented above. The PD reagent is an aqueous solution of silver ions containing a ferrous/ferric redox (reduction/oxidation) system, a citric acid buffer, and a cationic surfactant (usually *n*-dodecylamine acetate). This solution is relatively unstable, and, when a document is placed in the PD reagent, silver ions (Ag^+) are reduced to silver metal (Ag^0) (Figure 9). The silver preferentially deposits on the fingerprint deposit to produce dark gray to black ridges on a light gray background. The ferrous (Fe^{2+}) ions participate in the reduction of the silver ions, while the ferric (Fe^{3+}) ions slow the reaction down. Citric acid is necessary to complex with the ferric ions and to maintain a low pH. The surfactant traps randomly generated silver particles to inhibit further premature deposition of silver. While the mechanism of fingerprint development with PD has not been conclusively elucidated, several theories have been proposed [49, 50].

$$Fe^{2+}_{(aq)} + Ag^{+}_{(aq)} \rightleftharpoons Fe^{3+}_{(aq)} + Ag^{0}_{(s)}$$

Figure 9 The redox reaction that forms the basis of fingermark development by PD

Once immersed in the PD reagent, fingermarks take anywhere from 10 to 60 min to develop on the substrate. It is important to remove the item and rinse off the excess reagent as soon as good contrast is observed as it is not possible to reverse overdevelopment. Some paper substrates contain alkaline binders and fillers that react strongly with PD, resulting in high background development. In this case, an acid prewash (typically a solution of maleic acid) can be used to neutralize these alkaline components and improve fingermark development [49–52]. A systematic study of acid prewashes has demonstrated the influence of paper pH on successful fingermark enhancement with PD [53]. In some cases, PD-developed fingermarks on dark paper substrates (e.g., brown paper) can be further enhanced by treatment with a sodium hypochlorite solution (dilute household bleach). This lightens the background and darkens the fingerprint due to the formation of silver oxide [49, 50, 54].

The PD technique presents a number of challenges that make its use unfavorable in many operational laboratories. It is time consuming to prepare and employ. PD is also very sensitive to environmental factors and contamination, which means that good-quality reagents and clean glassware must be used, and the technician preparing and applying the technique must to be experienced in its use. It is relatively expensive because the working solution has a short half-life (generally less than two weeks). The PD process is also destructive, which means that no further fingerprint treatments can be used, and other forensic evidence such as inks and handwriting are also likely to be compromised [1].

Several modifications to the PD formulation have been proposed [55–57]; however, the new formulations are still subject to the same robustness issues that plague the traditional PD [58]. Despite its limitations, PD remains the best technique to detect latent fingermarks on porous surfaces that are or have been wet. In addition, on substrates that have not been wet, PD can sometimes develop fingermarks that go undetected using amino acid reagents.

Oil Red O (ORO). While PD has long been considered the only routine technique that is able to develop latent fingermarks on items that have been wet, promising research has been conducted using a relatively new reagent based on the lipid dye Oil Red

O (ORO). ORO stains the lipid (sebaceous) component of the latent fingermark deposit to produce a dark red to brown image on a light pink background. Treated fingermarks begin to appear on substrates after 5 min, with the best results being obtained if the item is agitated in the solution for up to 1 h. Despite the development time involved, the technique is less labor intensive than PD as the item can be left immersed in the solution and an automatic rocker used to agitate. Overdevelopment does not readily occur with ORO. This reagent has the potential to be used in conjunction with PD for the development of fingermarks on items that have been wet or items that have been previously treated with amino acid-sensitive reagents [59–61]. Several attempts have been made to improve the original ORO formula; however, the original formulation has so far proven to be superior [61, 62].

Salama and coworkers [61] conducted an extensive study where latent fingermarks on a variety of paper types were immersed in water for varying lengths of time and aged over periods up to eight weeks. For fingermarks aged up to four weeks, ORO development was as effective as PD, sometimes more effective. However, better results were achieved with PD for eight-week-old latent fingermarks.

ORO can be used after amino acid–sensitive reagents, provided their formulations do not contain solvents with a dielectric constant of less than 24, as such solvents appear to remove the lipid fraction targeted by ORO [63]. ORO does not replace PD but can be used before PD treatment to maximize the exploitation of any latent fingermark evidence that may be present [61].

Detection Techniques for Thermal Paper. Thermal paper is a porous surface that presents particular challenges for latent fingermark development. The active coating that gives this paper its thermosensitivity reacts with polar organic solvents that are present in traditional reagents such as ninhydrin to produce a dark gray or black color that obscures any developed fingermarks [64]. While acetone can be used to remove the active layer of the paper [65], the results are still not optimal and the technique is very destructive. Promising results have been obtained using a modified 1,2-indanedione formulation that contains a lower concentration of polar components that are in a nonpolar carrier solvent [66]. Preliminary experiments with low heat from a hair dryer [67] or acetic acid fuming [68] have produced good results on certain types of thermal paper.

An extensive study [65] has compared the development of a printed alanine standard with several different solutions: NPB (15 g ninhydrin dissolved in 100 ml ethyl alcohol and made up to 2.5 l with petroleum ether), INON (4 g 2-hydroxy-2-(3,5,5-trimethyl-hexyl-oxy)-indan-1,3-dione which is a hemiketal of ninhydrin and isononanol dissolved in 1 l petroleum ether), ThermaNin (4 g ThermaNin dissolved in 1 l petroleum ether), and DFO (50 mg DFO in 2 ml glacial acetic acid and 4 ml methyl alcohol made up to 100 ml with HFE7100). Various whitening methods were also tested. The best results were obtained with a combination of NPB with a whitening mixture "G3", which contains 12.5 mmol each of 4-pyrrolidino-pyridine, oenantholactam, 1-octyl-2-pyrrolidone, and 1-cyclohexyl-pyrrolidone.

Recommended Detection Sequence. The recommended sequence for porous surfaces is given in Figure 10 [1, 40]. This should be used as a general guide, and modified sequences or the integration of other techniques should also be considered.

Detection Techniques for Nonporous Surfaces

Cyanoacrylate Fuming. Cyanoacrylate fuming is the most common technique applied to smooth, nonporous surfaces that are dry. It is also possible to treat items that have been wet as long as they are dried first. Cyanoacrylate esters (generally the ethyl ester) are sold commercially as rapid, high-strength glues like "superglue". Cyanoacrylate vapors react with certain components in the latent fingermark deposit (e.g., moisture and some sebaceous components) to produce a hard, white polymer known as *polycyanoacrylate* (Figure 11). Recently, several studies have been conducted in an attempt to more clearly elucidate the chemistry behind the reaction and the factors that contribute to the process [69, 70].

Generally, the fuming technique involves placing or suspending the item in a sealed chamber and heating a small amount of the cyanoacrylate liquid to around 80–100 °C to produce a sufficient concentration of vapor. The development of fingermarks needs to be closely monitored to ensure that overdevelopment does not occur.

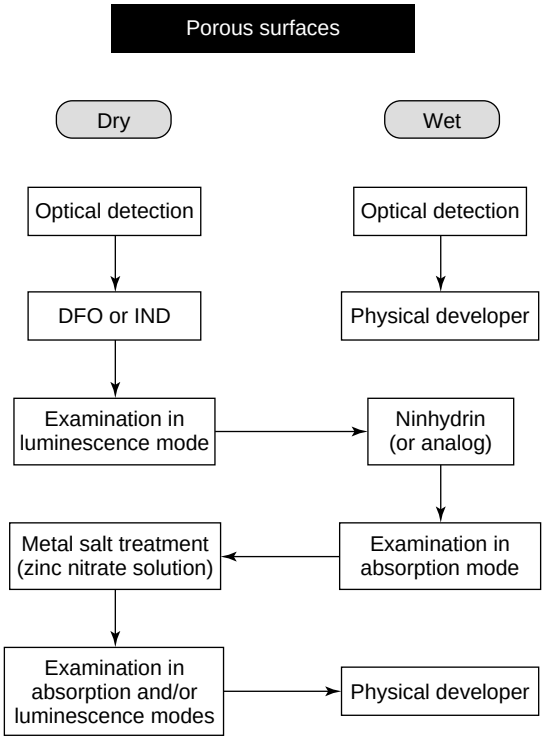


Figure 10 The recommended sequence for fingermark development on porous surfaces [Reproduced from Ref. 1. © Taylor & Francis Group, 2004.]

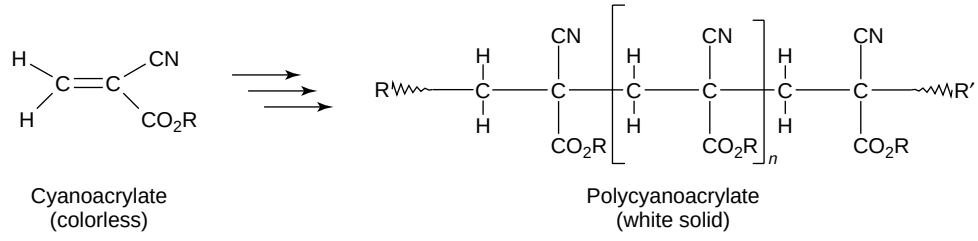


Figure 11 The polymerization of cyanoacrylate to give polycyanoacrylate [Reproduced from Ref. 1. © Taylor & Francis Group, 2004.]

Many commercial fuming units are available to automate the process, resulting in more effective control of temperature and vapor circulation, and the safe purging of the chamber of cyanoacrylate fumes before the items are removed. Some systems also control the humidity within the chamber. Portable cyanoacrylate fuming kits have also been investigated for use in the field. For example, when fuming large areas such as the inside of motor vehicles, a portable case can be used to generate cyanoacrylate fumes. These fumes are then pumped through a hose into the vehicle being processed [71]. A similar system can be employed to fume large items by building a plastic tent around the item and pumping the fumes into the tent. Commercial cyanoacrylate fuming systems are now available that can be safely and reliably used to process evidence in the field.

Vacuum cyanoacrylate (VCA) fuming involves placing the items in a chamber with a small amount of liquid cyanoacrylate and reducing the pressure

to approximately 200 mTorr using a rotary pump [72, 73]. The resulting treated fingermarks are more evenly developed and overdevelopment is unlikely. The developed fingermarks also have sharper ridge and pore detail than those developed by the conventional procedure [74–77]; however, the marks are less visible to the naked eye than with atmospheric pressure systems. Items such as unopened cans and bottles of soft drink are not suitable for VCA fuming as they may explode under reduced pressure.

Fingermarks treated with cyanoacrylate can be recorded using oblique lighting or coaxial illumination for white or multicolored surfaces. A reflected ultraviolet imaging system (RUVIS) can also be used. The contrast between cyanoacrylate-treated fingermarks and the substrate can be enhanced by powdering or applying a colored or luminescent stain once the cyanoacrylate polymer has been allowed to harden overnight. Many stains have been evaluated for the posttreatment of cyanoacrylate-developed fingermarks. Gentian violet is a colored stain that is useful for enhancing fingermarks on light colored substrates. However, luminescent stains are generally preferred, with the most popular being rhodamine 6G (luminescence excitation at 450–550 nm, observation at 550–600 nm) [78], Ardrox 970-P10 (luminescence excitation at 250–500 nm, observation at 450–650 nm) [79–82], and basic yellow 40 (luminescence excitation at around 445 nm, observation near 495 nm) [83, 84]. Other stains that have been evaluated include basic red 28 (luminescence excitation at 495 nm, observation at 585 nm) [85], safranine O (luminescence excitation at 520 nm, observation at 560 nm) [84], 4-(4-methoxybenzylamino-7-nitrobenzofurazan) (MBD); luminescence excitation at 350 nm, observation at 614 nm) [86], Nile red [87], and thenoyl europium chelate [88–90]. Combinations of various dyes have also been proposed for certain surfaces [85, 91, 92]. Cyanoacrylate stains are generally applied by dissolving the colored or luminescent material in a solvent mixture and then dipping or spraying the evidence. Background staining is then removed by washing with water. For substrates that are not suitable for staining (for example, semiporous substrates that cannot be effectively destained), fingerprint powders should be used to enhance cyanoacrylate-treated fingerprints. Black magnetic powder is the recommended powder, although luminescent powders can also be used. These powdered fingermarks can then be lifted, or simply photographed.

Vacuum Metal Deposition (VMD). Vacuum metal deposition (VMD) is a technique that can be used to develop latent fingermarks on a variety of nonporous and semiporous substrates [93–95]. Under vacuum, gold is evaporated and subsequently condenses on the substrate in a layer so thin that it is invisible to the naked eye. Zinc is then applied in a similar way and preferentially deposits on the areas of exposed gold where there is no fingermark deposit (Figure 12). Cadmium could also be used instead of zinc; however, cadmium is rarely used because of its toxicity. This technique generally produces negative fingermarks where the ridges remain transparent while the background is metalized (i.e., opaque); however, for certain types of polymer substrates and under particular conditions, reverse development occurs where the zinc deposits on the fingermark ridges and not on the background [94, 96, 97].

Jones and coworkers have extensively studied VMD development of fingermarks on a variety of polymer surfaces [98–101]. They found that VMD development is dependent on the type of polymer substrate and the amount of gold deposited; and that no one set of conditions is universally applicable. Guidelines for optimizing VMD development on various substrates, including problematic polymer banknotes, have been provided by this

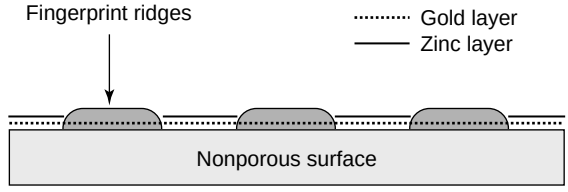


Figure 12 Fingermark development by vacuum metal deposition [Reproduced from Ref. 1. © Taylor & Francis Group, 2004.]

research group [100, 101]. More recent studies have investigated the gold agglomerates formation during the VMD process using transmission electron microscopy (TEM) imaging [102].

Levin Elad and coworkers [103] studied the surface properties of different polymers and examined their correlation with VMD development of latent fingermarks. The surface energy did not affect fingermark development; however, there was a correlation with the degree of roughness of the surface.

VMD can be used successfully after cyanoacrylate development and before the application of a luminescent stain [104–108]. Some studies have suggested that VMD may, in fact, be a more sensitive technique than cyanoacrylate, particularly for older latent fingermarks [109]. Despite the advantages of using this technique, the high cost of purchasing a VMD unit and the expertise required to obtain optimum results means that its application is generally limited to major cases [109] and, in particular, those involving "difficult" surfaces such as leather, smooth fabric, and polymer banknotes [101].

Recommended Detection Sequence. The recommended sequence for nonporous surfaces is provided in Figure 13. Where possible, a wet nonporous item should be dried at room temperature before being treated as a dry item. If this is not possible, SPR should be applied.

Miscellaneous Techniques

Multimetal Deposition (MMD). With multimetal deposition (MMD), the item to be treated is immersed in a colloidal gold solution (pH approximately 2.7) followed by a modified physical developer (MPD) [110]. The negatively charged gold colloids (prepared by treating tetrachloroauric acid with sodium citrate and a detergent) are attracted to organic residues in the fingermark deposit that are positively charged because of the strongly acidic conditions. The colloids then acts as nucleation sites that induce the precipitation of silver from the MPD solution. Developed fingermarks are visible as light gray to black ridges against a gray background. MMD can be used to develop fingermarks on many types of substrates, including "difficult" semiporous surfaces

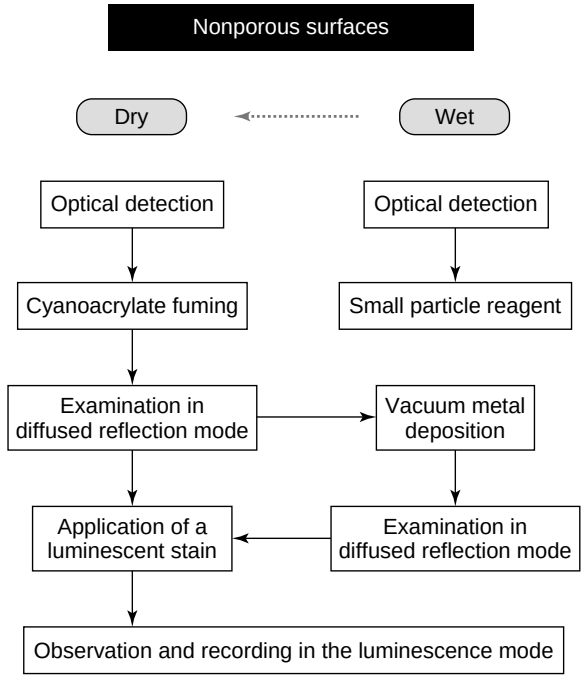


Figure 13 The recommended sequence for fingermark development on nonporous surfaces [Reproduced from Ref. 1. © Taylor & Francis Group, 2004.]

such as masking tape, glossy bottle labels, and rubber gloves [111].

Multi-metal deposition II (MMDII) is an optimized version of the technique that was developed to improve the reagent's sensitivity and specificity of latent fingermark enhancement [112, 113]. With MMDII, the use of silanized glassware and 14 nm colloidal gold particles is recommended, and the PD is based on silver acetate and hydroquinone rather than silver nitrate and a ferrous/ferric redox system. MMDII produces consistently better results than the MMD formulation and can be successfully applied to semiporous surfaces such as latex and nitrile gloves, polystyrene, and waxed paper [98]. However, MMD (and MMDII) still suffers from some of the disadvantages of conventional PD: the procedure is laborious, the reagents are delicate and expensive to prepare, significant experience is required for optimum results, and the technique is destructive.

Other Techniques Based on Nanoparticles. Becue and coworkers [114] have investigated modifications to the MMD method to reduce its operational limitations, namely cost, shelf life of reagents, and the labor-intensity nature of the process. Recently, a single-metal deposition (SMD) technique was proposed as another alternative to MMD (and MMDII) [115]. The method involves keeping the original gold colloid deposition and replacing the silver PD step with a gold-based amplification process. Results obtained with SMD were largely comparable to MMD, with the benefits of reduced cost, longer shelf life of reagents, and a simplified application procedure.

Choi and coworkers [116] have investigated nanoparticle binding to latent fingermarks in the MMDII process using scanning electron microscopy (SEM). They confirmed that the gold nanoparticles bind preferentially to the fingermark ridges under standard MMDII conditions, and that surfactant concentration influences background deposition but not ridge development. Conversely, pH variation influences ridge binding but not background development [116]. This research group has developed gold and silver nanopowders using oleylamine as a stabilizer and studied the fingermark development obtained by brushing the powder over latent fingermarks deposited on various surfaces. The long alkyl chain stabilizing groups were found to impart a lipophilic character to the gold nanoparticles, which increases the selectivity of the technique. Further studies will involve developing a water-dispersive gold nanoparticle to improve the MMD technique and investigating other nanoparticles that may improve latent fingermark development on different surfaces [117].

Almog and coworkers [118] investigated the enhancement of fingermarks treated with silver PD using various nanoparticles. Gold nanoparticles stabilized by *n*-alkanethiols and CdSe/ZnS nanoparticles stabilized by *n*-alkaneamine were found to adhere preferentially to the ridges of latent fingermarks on nonporous surfaces. CdSe/ZnS nanoparticles have fluorescent properties under UV illumination.

Iodine/Benzoflavone. Iodine fuming can be applied to a wide variety of porous and nonporous substrates. The iodine vapor is absorbed by the lipids in the latent fingermark deposit to produce a yellow–brown image. The process is simple, rapid, and economic; however, the fingermark development is a reversible physical process, which means that, unless the image is captured or chemically fixed, the contrast is generally poor and transitory. The technique is also only sensitive enough to develop latent fingermarks that are less than one week old.

Several methods have been proposed to capture or fix the image produced by iodine development, including silver plate transfer [119–121] (which is limited in application but has been proposed for the detection of fingermarks on human skin). The best results have been obtained with a solution of 7,8-benzoflavone (α-naphthoflavone), which chemically fixes iodine fumed fingermarks as a relatively stable dark blue–purple image [122, 123]. A system that involves mixing a solution of iodine with a solution of 7,8-benzoflavone immediately prior to use, and then spraying the mixture onto the surface of interest, has proven to be the particularly effective for fingermark detection at the crime scene [119, 120, 124].

Osmium Tetroxide and Ruthenium Tetroxide (RTX). Osmium tetroxide (OsO_4) is an oxidant that reacts with the unsaturated organic components of the latent fingermark deposit to give a dark gray–black product. As OsO_4 is volatile, the item to be treated is exposed to the vapor given off by crystals of the reagent in a closed glass container until a gray–black image of the fingermark is formed [125]. Development times range from 1 to

12 h. The technique can detect fingermarks on both porous and nonporous surfaces; however, exposure to OsO_4 can be fatal if inhaled, ingested, or absorbed through the skin. For this reason, the technique is no longer considered a viable alternative for fingermark detection.

Ruthenium tetroxide (RuO_4) reacts in the same way as OsO_4; however, it is not volatile at or near room temperature. The compound decomposes explosively at 108 °C, therefore fuming methods are extremely hazardous [125]. A safer method of generating RuO_4 vapor (known as *RTX*) involves reacting equal volumes of 0.1% ruthenium(III) chloride hydrate solution and 11.3% ceric ammonium nitrate solution at room temperature in a closed container [126]. The fumes generated produce a dark gray image in around 10–20 min or longer, depending on the surface. The fumes can be directly applied in a chamber enclosing the evidential items or by placing the reactants in a wash bottle and gently squeezing the generated RTX vapors through the nozzle that is directed onto the surface of interest. Nonporous items can also be directly immersed in the RTX reaction mixture for almost instantaneous development [127].

RTX works well when the latent fingerprint deposit contains oily material from the sebaceous secretions, but is generally ineffective on eccrine deposits. A modification of the procedure involves preparing a developing solution of 0.25% (w/v) RTX dissolved in a HFC solvent [124]. The item is sprayed or immersed in the solution, and fingermarks appear as brownish-black images. Fingerprints developed with RTX can be removed with a 4% hypochlorite solution. RTX is toxic by inhalation or skin contact [128], and must be used with great care.

A less expensive version of RTX solution that uses HFE7100 as a carrier solvent was proposed and compared with conventional RTX and iodine-benzoflavone for the development of latent fingermarks at the crime scene [129]. Recommendations are made in this study for the best sequencing of methods on various surfaces and for various ages of fingermarks.

Detection on Semiporous Surfaces

Fingerprint detection on semiporous surfaces such as waxed paper, glossy paper, matte-painted surfaces, and some rubber/latex gloves can often be difficult as the techniques available are generally suited to either porous or nonporous surfaces. Iodine/benzoflavone and MMD are excellent techniques for specific nonporous surfaces; however, the best approach for general fingermark detection on semiporous surfaces is to use a combination of techniques (Figure 14).

Detection on Human Skin

Human skin is an extremely challenging substrate to develop latent fingermarks on because the natural secretions comprising the latent deposit are also found on the surface of the skin. Therefore, when a detection technique is applied, the background develops along with any fingermarks, resulting in a very poor signal-to-noise ratio. Fingermarks on warm skin are particularly difficult to detect because the deposited secretions remain in a liquid state and diffuse very quickly. Despite the inherent difficulties, in certain cases, an attempt to develop latent fingermarks on human skin may be justified. For example, in a case of manual strangulation, fingermark development should be attempted on the neck of the victim.

Cadavers should ideally be processed for fingermarks at the crime scene to maximize chances of detection and recovery. As with all fingermark examinations, the sequence should begin with a thorough optical examination using different lighting conditions. Visible fingermarks should be photographed (and a silicone cast made if a molded impression is observed) before applying any other detection method. The appropriateness of subsequent techniques will depend on the time delay and whether the body has been stored in a cold room. Common techniques include the following: powdering [130, 131], iodine or iodine/benzoflavone [119, 120, 131, 132], RTX [124, 126, 131, 133], cyanoacrylate fuming [88, 134–136], and various transfer techniques [131, 137–142]. Cyanoacrylate fuming is precluded if it is more than 24 h after death or in the case of a cadaver that has been stored in a cold room because of the moisture that is present on the skin surface.

Sampson and Sampson [143] have published a review based on personal experience, published accounts, interviews, case reports, and student questionnaires and suggest guidelines for the processing of human skin for latent fingermarks. In addition to these recommendations (Figure 15), experience and experimentation should be used to determine the best sequence for a particular set of circumstances.

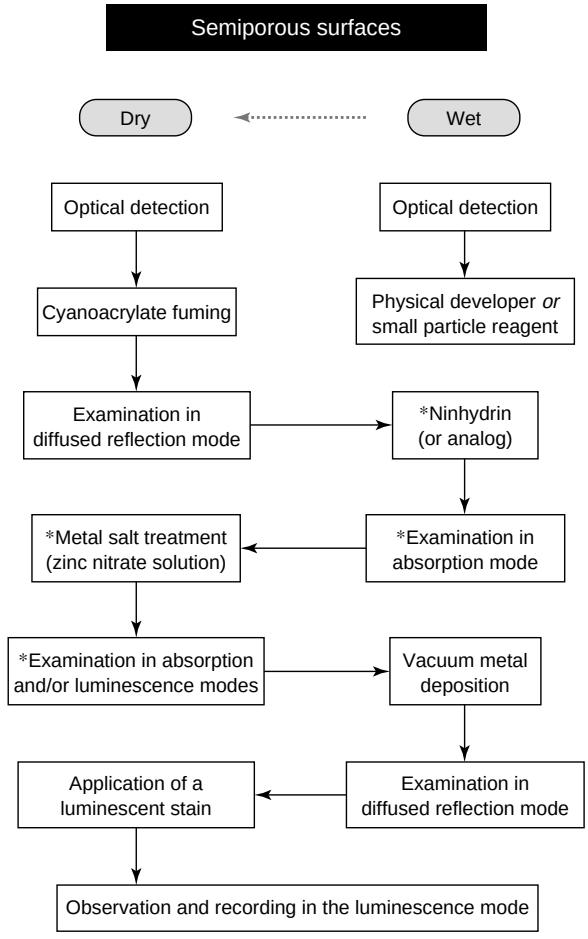


Figure 14 The recommended sequence for fingermark development on semiporous surfaces (*skip if item has been wet) [Reproduced from Ref. 1. © Taylor & Francis Group, 2004.]

Detection on Adhesive Surfaces

Adhesive surfaces such as the sticky side of adhesive tapes and labels are promising substrates for the development of fingermarks because such surfaces essentially pull secretions and skin debris from the finger when contact is made. In addition, a perpetrator is likely to remove their gloves to more easily manipulate the adhesive tape.

It is often necessary to remove adhesive tape from another surface (or from itself) before processing. Several techniques have been proposed for achieving this, including freezing, heating, and the application of a solvent [144–146]. Schwartz and coworkers [147] have shown that labels and stamps can be successfully removed with the product Un-Do® (a heptane-based adhesive neutralizer) without adverse effects on subsequent fingermark detection.

A range of techniques is available for the development of latent fingermarks on adhesive surfaces, and these have been extensively reviewed by Midkiff and coworkers [148–150]. Techniques used to develop latent fingermarks on adhesive surfaces must be combined with methods used to process any nonadhesive surface that is also present. For example, an adhesive tape must be examined on both the adhesive and nonadhesive sides, and the techniques used for each side must be compatible (i.e., not mutually exclusive). Normally, the best approach is to develop

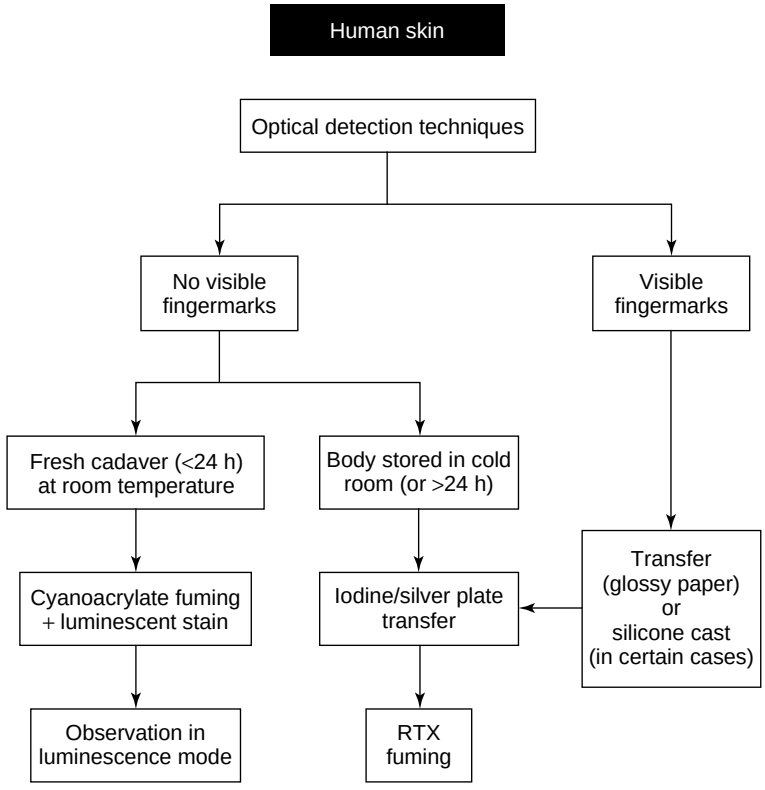


Figure 15 The recommended sequence for fingermark development on human skin [Reproduced with permission from Ref. 133. © IAI, 1994.]

and photograph fingerprints on the nonadhesive side before treating the adhesive side.

Gentian violet (crystal violet) is a reagent that stains the sebaceous material in the latent fingermark deposit to produce a dark purple image against a light purple background. The technique is very effective for the treatment of pressure-sensitive adhesive tapes such as clear Sellotape, paper masking tape, and polyvinyl chloride (PVC) insulation tape. The reagent is a stain solution that generally contains phenol, which facilitates the absorption of the stain into the fatty fingerprint deposit. It can be applied by dipping or using a pipette; or, alternatively, floating the tape (adhesive side down) on the surface of the reagent. The tape is then washed with water to remove excess stain. Observation of the treated fingermarks is generally made in the visible region, with coaxial illumination [151] or capture of the deep red to near-IR luminescence [152] being other options for enhancement. Fingermarks developed on dark-colored adhesive surfaces (e.g., black electrical tape) can be transferred onto fixed and washed photographic paper. Owing to the extreme toxicity of phenol, some phenol-free gentian violet formulations have been developed.

Sticky-side powder is a suspension of black fingerprint powder in a detergent solution that produces dark gray to black fingermarks on adhesive surfaces [153, 154]. A slurry is prepared by mixing equal parts of Photoflo® detergent (Kodak) and water with a small amount of black powder. This is then painted onto the adhesive surface using a soft brush (e.g., a camel-hair fingerprint brush) and gently rinsed off after 10–15 s under running tap water. The process can be repeated if only weak development is obtained the first time. Sticky-side powder produces good results on a range of adhesive surfaces, including duct tape, masking tape, surgical tape, paper-backed labels, clear plastic tapes, and reinforced packing tapes [154].

Black electrical tape is a particularly difficult surface to develop latent fingerprints on due to its dark coloring. White or ash gray powder can be used to replace the black powder in the sticky-side powder formulation to overcome this difficulty [155]. Fluorescent or white fingerprint powders can also be combined with Liqui-nox™ and water [156]. In fact, a variety of conventional fingerprint powders can be used in suspension to process adhesive surfaces for latent fingermarks [157]. Titanium dioxide has been proposed as a white pigment that can be used in sticky-side powder [158] or SPR [159] formulations to detect latent fingermarks on dark surfaces. An evaluation of the different methods that may be used to develop latent fingermarks on black electrical tape was conducted by Schiemer and coworkers [160]. Their recommended sequence is cyanoacrylate fuming and a BY40/BR28 combined fluorescent stain followed by a white powder suspension (1 ml Photo-Flo, 1.5 ml Citron detergent, and 2 ml water). Gentian violet can then be used followed by a transfer onto photographic paper.

The recommended sequence (Figure 16) does not include MMD; however, MMD may be considered as it is capable of developing fingerprints on both adhesive and nonadhesive surfaces. Again, experience and experimentation should be taken into consideration before applying techniques to evidential items as there is no universal procedure for adhesive surfaces.

Detection on Cartridge Cases and Firearms

While latent fingermarks may be developed on unfired cartridge cases using a variety of techniques, fired cartridge cases present a much greater challenge. This is most likely due to friction between the surface of the cartridge case and the chamber at the moment of ejection, although other contributing factors include friction at other stages of the firing process (e.g., loading of the magazine or of the live round into the chamber), the high temperature and pressure generated in the cartridge case at the time of firing, and exposure of the surface of the cartridge case to combustion gases and discharge residues generated at the time of firing [161]. The friction causes

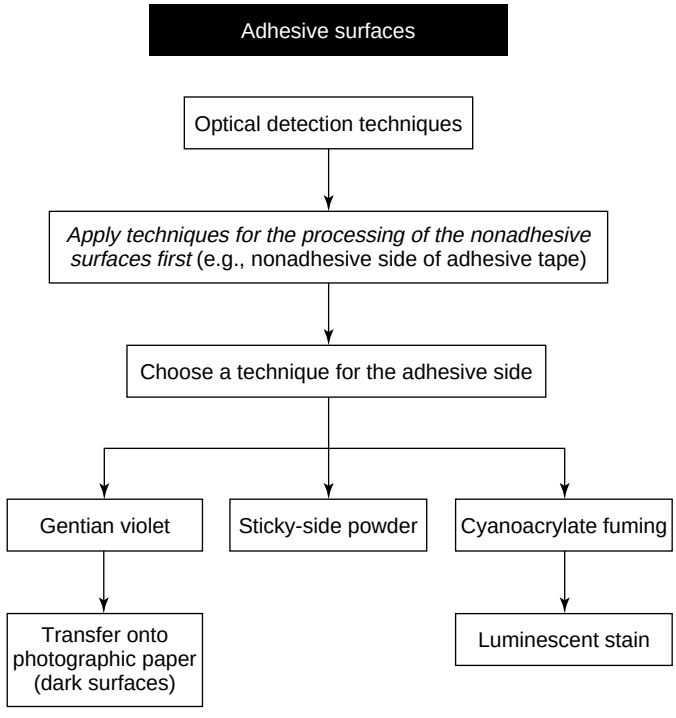


Figure 16 The recommended sequence for fingermark development on adhesive surfaces [Reproduced from Ref. 1. © Taylor & Francis Group, 2004.]

smearing of the fingermarks and, therefore, regardless of the detection technique employed, detection success rates on fired cartridge cases are generally low.

Latent fingermarks on brass cartridges may develop spontaneously due to a process of differential tarnishing; therefore, a thorough visual examination should be conducted before any other technique is applied.

Firearms can also present difficulties for latent fingermark development because of the nature of the surface (e.g., metal, wood or plastic), the presence of surface finishes and patterns, or surface contamination (e.g., gun oil and discharge residues), as well as the way that the firearm has been handled and packaged. Despite these inherent problems, identifiable fingermarks may be developed around 10% of the time [162].

Cyanoacrylate fuming produces mixed results on cartridge cases according to the type of material they are comprised of. Good results have been obtained on nickel casings treated with cyanoacrylate, while poor results are generally obtained on brass casings [163]. Satisfactory results have been achieved on plastic shotgun cartridges, with limited results on metallic casings [164]. Fingermark detection using VCA was investigated on a variety of different firearms. Conventional cyanoacrylate fuming was found to produce a whiter, more visible polymer than that obtained by the vacuum method. However, overdevelopment can be avoided with vacuum fuming, and therefore it is considered a better choice for processing firearms [165].

Gun blue is a reagent containing the active ingredients selenious acid (H_2SeO_3) and a cupric salt in an acid solution. The selenious acid and cupric ions oxidize certain metals (e.g., zinc, aluminum and iron) to form a black copper-selenide coating. The reagent selectively reacts on clean metal substrates, so the presence of greasy or oily material (such as a latent fingermark contaminated with sebaceous material) inhibits the reaction [166, 167]. A dilute gun blue solution can therefore be applied to metal surfaces such as brass cartridge cases to develop latent fingermarks as transparent images against a dark background. It is important to consider toolmark evidence before using gun blue, as the technique can destroy the fine detail present in such marks [168].

A comparison of etching (dissolution of the metal surface with an acid), blueing (dissolution of the metal with an acid and formation of a dark metal complex), and MMD showed that gun blue was the most effective treatment for the development of latent fingermarks on brass cartridge cases and that MMD was the best technique for aluminum cartridge cases. Lacquered steel cartridge cases were found to be a problematic surface, and cyanoacrylate fuming was recommended in this instance [169].

Other techniques that can be used to develop latent fingermarks on cartridge cases include palladium deposition techniques [166, 170, 171] and camphor fuming [172].

Enhancement of Fingermarks in Blood

Fingermarks in blood are often present at scenes of violent crime. Such marks can be very powerful evidence when the offender's fingermarks are found in the blood of the victim (as determined by DNA profiling). It is therefore important to preserve the DNA evidence present when enhancing fingermarks in blood. Nondestructive techniques such as optical enhancement should be used before other more destructive techniques. Heavily bloodstained areas that are unlikely to produce clear fingerprint detail should be sampled for DNA testing before fingermark development techniques are applied.

Fingermarks in blood should be considered at the same time as latent fingermarks, as the enhancement techniques for one type of evidence are not necessarily compatible with the other. For example, techniques that develop only fingermarks in blood, such as amido black or diaminobenzidine (DAB), are likely to destroy latent fingermarks.

In some instances, it may be useful to determine whether a fingermark in blood was formed by a contaminated finger (i.e., blood on the finger) or if it was a latent fingermark already on the surface that was subsequently developed by the blood. A number of studies have been conducted to determine whether this is in fact possible. Creighton [173] found that, when blood was allowed to flow over a series of latent fingermarks, the tendency was for the fingermark to repel the blood. Huss and coworkers [174] found that blood does not develop a previously deposited eccrine (sweaty) fingermark, whereas it is possible for blood to develop sebaceous (greasy) fingermarks, although the development is reversed because the greasy ridges

repel the blood, which consequently accumulates in the furrows.

Fingermarks in blood created by a contaminated finger can also occur as a negative image, where the furrows are stained with blood and the ridges are clear. This phenomenon was originally assumed to be a function of the pressure and contact involved during deposition of the fingermark; however, recent studies have refuted this [175]. Reverse development actually occurs when there has been a time lapse between contamination of the finger and deposition of the fingermark. When the blood present on the ridges is dry, it is not transferred. Rather, the fluid blood remaining in the furrows is deposited, resulting in a negative image.

Untreated blood can be enhanced in the absorption mode (recommended for lightly colored or luminescent surfaces) or in the diffused reflection mode (recommended for dark or shiny surfaces). The best conditions for visualization in the absorption mode are created with a strong band of light at around 415 nm (band width approximately 40 nm) in a dark room, as blood exhibits a strong absorbance at this wavelength [176]. Under these conditions, the fingermark appears almost black, creating good contrast with a light colored or luminescent background. Fingermarks captured on film or using a digital camera have better contrast than those perceived with the human eye due to our relative insensitivity in the violet region of the light spectrum.

For diffused reflection, an illumination wavelength other than the violet region is employed. For a colored surface, the best contrast is obtained with a wavelength that is opposite to the substrate color. The fingermark in blood is observed perpendicular to the surface, without the use of a barrier filter, while working under dark conditions. The incident light angle is varied until the best contrast is obtained. The incident light is strongly absorbed for dark surfaces or specularly reflected for shiny surfaces, while the blood itself exhibits diffuse reflectance. The end result is light ridges observed against a dark background.

UV luminescence techniques can be used where excitation is in the short-wave UV region with observation in the long-wave UV region to enhance blood, semen, and saliva marks [177]. However, short-wave UV light may be detrimental to subsequent DNA profiling, so care must be exercised when using this method (*see* **Friction Ridge Skin: Interaction between Fingerprint Detection and DNA/Biological Material**). In addition, UV imaging systems are expensive and the technique labor intensive when large surfaces are being screened.

Fingermarks in blood must be "fixed" to prevent the blood from diffusing or washing away when a chemical treatment is applied. For example, items can be immersed in either methanol or 5-sulfosalicylic acid solution for 5 min (or longer for heavier deposits) to fix the blood [178, 179].

Protein stains such as amido black [83, 178, 179], coomassie blue [180], and Hungarian red [181], or chemicals such as DAB or ABTS (2,2′-azino-di-(3-ethyl-benzthiazolinesulfonate) diammonium salt) [182] can be subsequently employed to enhance weak fingermarks in blood.

DFO is effective for the enhancement of both bloody and latent fingermarks on porous surfaces to give strong room-temperature luminescence [176]. The use of DFO or ninhydrin does not preclude the use of protein stains such as amido black; however, the converse is not true. DFO or ninhydrin is recommended as the first treatment because latent fingermarks are not developed by protein stains.

Leuchomalachite green, phenolphthalein, fluorescein, tetramethylbenzidine, and leuco rhodamine 6G are heme-reacting chemicals, which have been proposed for the enhancement of fingermarks in blood [183–187]. Reagents that react with heme are generally extremely sensitive. Fixed black-and-white photographic paper can be successfully used to transfer fingermarks in blood, which have been treated with either leucomalachite green (LMG), leucocrystal violet (LCV), but not DAB) [188]. This could be useful where the fingermarks are situated in locations that are difficult to photograph or where the contrast is poor.

Sears and coworkers [189] have systematically studied a number of techniques for the development of fingermarks in blood and have provided recommendations for the best sequences on a variety of surfaces. A recommended sequence for the detection and enhancement of fingermarks in blood is provided in Figure 17. Techniques for the development of latent fingermarks should be integrated appropriately in such sequences for the maximum exploitation of any fingerprint evidence that may be present.

Bleay and coworkers [190] investigated the development of fingermarks in blood that have been

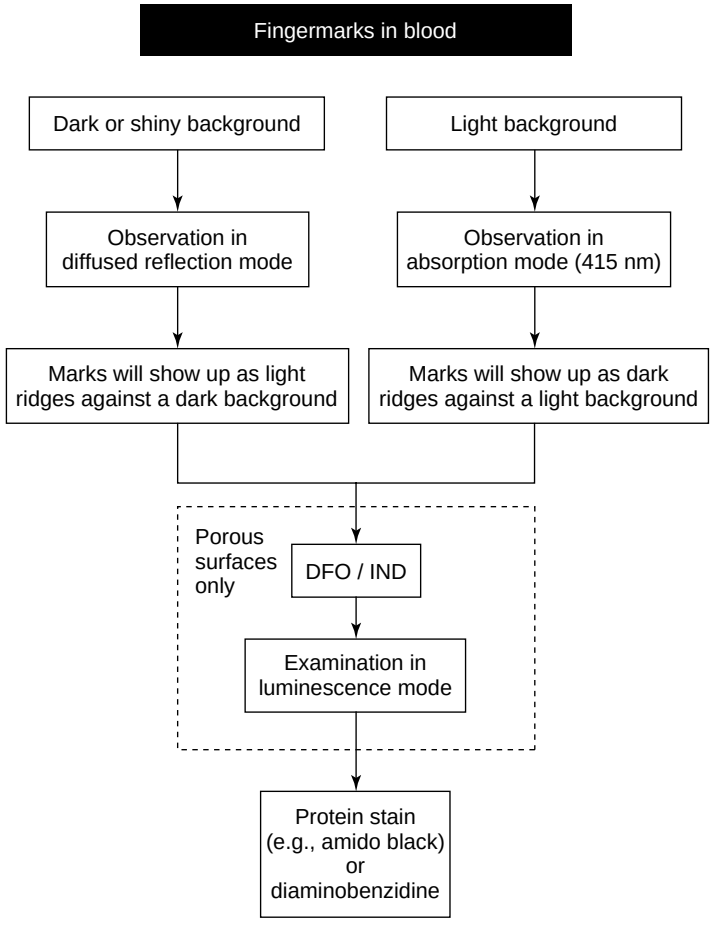


Figure 17 The recommended sequence for the enhancement of fingermarks in blood [Reproduced from Ref. 1. © Taylor & Francis Group, 2004.]

exposed to the conditions associated with fire scenes. They found that blood fingermarks could survive high temperatures and fume-filled environments. Protein dyes are ineffective on blood fingermarks that have been exposed to temperatures above 200 °C; however, black powder suspensions, cyanoacrylate fuming, and VMD can still be used.

References

[1] Champod, C., Lennard, C., Margot, P. & Stoilovic, M. (2004). *Fingerprints and Other Ridge Skin Impressions*, CRC Press LLC.

[2] Ong, S.K., Seah, L.K., Murukeshan, V.M. & Ong, L.S. (2004). Enhanced fluorescence in time-resolved imaging of latent fingerprints, *Journal of Forensic Identification* **54**(3), 281–295.

[3] Menzel, E.R. & Menzel, L.W. (2004). Ordinary and time-resolved photoluminescence field detection of traces of explosives and fingerprints, *Journal of Forensic Identification* **54**(5), 560–571.

[4] Seah, L.K., Dinish, U.S., Phang, W.F., Chao, Z.X. & Murukeshan, V.M. (2005). Fluorescence optimisation and lifetime studies of fingerprints treated with magnetic powders, *Forensic Science International* **152**, 249–257.

[5] Seah, L.K., Wang, P., Murukeshan, V.M. & Chao, Z.X. (2006). Application of fluorescence lifetime imaging (FLIM) in latent finger mark detection, *Forensic Science International* **160**, 109–114.

[6] Menzel, E.R., Schwierking, J.R. & Menzel, L.W. (2005). Functionalized europium oxide nanoparticles

for fingerprint detection: a preliminary study, *Journal of Forensic Identification* **55**(2), 189–195.

[7] Exline, D.L., Wallace, C., Roux, C., Lennard, C., Nelson, M.P. & Treado, P.J. (2003). Forensic applications of chemical imaging: latent fingerprint detection using visible absorption and luminescence, *Journal of Forensic Sciences* **48**(5).

[8] Payne, G., Reedy, B., Lennard, C., Comber, B., Exline, D. & Roux, C. (2005). A further study to investigate the detection and enhancement of latent fingerprints using visible absorption and luminescence chemical imaging, *Forensic Science International* **150**, 33–51.

[9] Tahtouh, M., Kalman, J.R., Roux, C., Lennard, C. & Reedy, B.J. (2005). The detection and enhancement of latent fingermarks using infrared chemical imaging, *Journal of Forensic Sciences* **50**(1).

[10] Lee, H.C. & Gaensslen, R.E. (2001). Methods of latent fingerprint development, in *Advances in Fingerprint Technology*, 2nd Edition, H.C. Lee & R.E. Gaensslen, eds, CRC Press LLC, Massachusetts, pp. 105–175.

[11] Saroa, J.S., Sodhi, G.S. & Garg, R.K. (2006). Evaluation of fingerprint powders, *Journal of Forensic Identification* **56**(2), 186–197.

[12] Singh, G.D., Sodhi, G.S. & Jasuja, O.P. (2006). Detection of latent fingerprints on fruits and vegetables, *Journal of Forensic Identification* **56**(3), 374–381.

[13] Sodhi, G.S. & Kaur, J. (2003). Fingerprint detection using phloxine B dye, *Journal of Forensic Identification* **53**(1), 8–13.

[14] Sodhi, G.S. & Kaur, J. (2005). Proflavin-based fingerprint dusting composition, *Fingerprint Whorld* **31**(122), 239.

[15] Sodhi, G.S., Kaur, J. & Garg, R.K. (2004). Fingerprint powder formulations based on organic, fluorescent dyes, *Journal of Forensic Identification* **54**(1), 4–8.

[16] Sodhi, G.S., Kaur, J., Garg, R.K. & Kobilinsky, L. (2003). A fingerprint powder formulation based on Rhodamine B dye, *Journal of Forensic Identification* **53**(5), 551–555.

[17] James, J.D., Pounds, C.A. & Wilshire, B. (1991). Flake metal powders for revealing latent fingerprints, *Journal of Forensic Sciences* **36**(5), 1368–1375.

[18] Van Oorschot, R.A.H., Treadwell, S., Beaurepaire, J., Holding, N.L. & Mitchell, R.J. (2005). Beware of the possibility of fingerprinting techniques transferring DNA, *Journal of Forensic Sciences* **50**(6), 1417–1422.

[19] James, J.D., Pounds, C.A. & Wilshire, B. (1991). Obliteration of latent fingerprints, *Journal of Forensic Sciences* **36**(5), 1376–1386.

[20] James, J.D., Pounds, C.A. & Wilshire, B. (1991). Magnetic flake fingerprint technology, *Journal of Forensic Identification* **41**(4), 237–247.

[21] Bandey, H. (2007). Powder effectiveness on surfaces commonly encountered at scenes of crime. *6th Biennial International Fingerprint Research Group Meeting*, Canberra, Australia.

[22] Goode, G.C. & Morris, J.R. (1983). *Latent fingerprints: A review of their origin, composition and methods for detection*. Atomic Weapons Research Establishment Report No 022/83.

[23] Haque, F., Westland, A.D., Milligan, J. & Kerr, F.M. (1989). A small particle (iron oxide) suspension for detection of latent fingerprints on smooth surfaces, *Forensic Science International* **41**, 73–82.

[24] Irrausch, F. (1990). *L'utilisation de L'oxyde de fer dans le Réactif de Microparticules*, Research Seminar, Institut de Police Scientifique et de Criminologie, Université de Lausanne, Switzerland.

[25] Frank, A. & Almog, J. (1993). Modified SPR for latent fingerprint development on wet, dark objects, *Journal of Forensic Identification* **43**(3), 240–244.

[26] Springer, E. & Bergman, P. (1995). A fluorescent small particle reagent (SPR), *Journal of Forensic Identification* **45**(2), 164–168.

[27] Bleay, S., Sears, V.G., Ciuksza, T. & Lawrie, K. (2007). Powder suspensions – not just for adhesive tapes. *6th Biennial International Fingerprint Research Group Meeting*, Canberra, Australia.

[28] Ruhemann, S. (1910). Triketohydrindene, *Journal of the Chemical Society* **97**, 2025.

[29] Grigg, R., Malone, J.F., Mongkolaussavaratana, T. & Thianpatanagul, S. (1986). Cycloaddition reactions relevant to the mechanism of the ninhydrin reaction: X-ray crystal structure of protonated Ruhemann's purple, a stable 1,3-dipole, *Journal of the Chemical Society: Chemical Communications*, 421–422.

[30] Lennard, C.J. (1986). *New Amino Acid Specific Reagents*. PhD Thesis. Australian National University, Canberra, Australia.

[31] Joullié, M.M., Thompson, T.R. & Nemeroff, N.H. (1991). Ninhydrin and ninhydrin analogs. Syntheses and applications, *Tetrahedron* **47**(42), 8791–8830.

[32] Lennard, C.J., Margot, P.A.J.-L., Sterns, M. & Warrener, R.N. (1987). Photoluminescent enhancement of ninhydrin developed fingerprints by metal complexation: structural studies of complexes formed between Ruhemann's purple and Group IIb metal salts, *Journal of Forensic Sciences* **32**(3), 597–605.

[33] Pounds, C.A., Grigg, R. & Mongkolaussavaratana, T. (1990). The use of 1,8-diazafluoren-9-one (DFO) for the fluorescent detection of latent fingerprints on paper. A preliminary evaluation, *Journal of Forensic Sciences* **35**(1), 169–175.

[34] Wilkinson, D. (2000). Study of the reaction mechanism of 1,8-diazafluoren-9-one with the amino acid, $_L$-alanine, *Forensic Science International* **109**(2), 87–103.

[35] Stoilovic, M. (1993). Improved method for DFO development of latent fingerprints, *Forensic Science International* **60**(3), 141–153.

[36] Masters, N., Morgan, R. & Shipp, E. (1991). DFO, its usage and results, *Journal of Forensic Identification* **41**(1), 3–9.

[37] Didierjean, C., Debart, M.-H. & Crispino, F. (1998). New formulations of DFO in HFE7100, *Fingerprint Whorld* **24**, 163–167.

[38] Hauze, D.B., Petrovskaia, O., Taylor, B., Joullié, M.M., Ramotowski, R. & Cantu, A.A. (1998). 1,2-indanediones: new reagents for visualizing the amino acid components of latent prints, *Journal of Forensic Sciences* **43**(4), 744–747.

[39] Wiesner, S., Springer, E., Sasson, Y. & Almog, J. (2001). Chemical development of latent fingerprints: 1,2-indanedione has come of age, *Journal of Forensic Sciences* **46**(5), 1082–1084.

[40] Wallace-Kunkel, C., Lennard, C., Stoilovic, M. & Roux, C. (2007). Optimisation and evaluation of 1,2-indanedione for use as a fingermark reagent and its application to real samples, *Forensic Science International* **168**, 14–26.

[41] Ramotowski, R. & Bicknell, D. (2007). Evaluation of the new 1,2-indanedione-zinc reagent. *6th Biennial International Fingerprint Research Group Meeting*, Canberra, Australia.

[42] Merrick, S., Gardner, S.J., Sears, V.G. & Hewlett, D.F. (2002). An operational trial of ozone-friendly DFO and 1,2-indanedione formulations for latent fingerprint detection, *Journal of Forensic Identification* **52**(5), 595–605.

[43] Gardner, S.J. & Hewlett, D.F. (2003). Optimization and initial evaluation of 1,2-indandione as a reagent for fingerprint detection, *Journal of Forensic Sciences* **48**(6).

[44] Stoilovic, M., Lennard, C., Wallace-Kunkel, C. & Roux, C. (2007). Evaluation of a 1,2-indanedione formulation containing zinc chloride for improved fingermark detection on paper, *Journal of Forensic Identification* **57**, 4–18.

[45] Patton, E. (2007). *Investigations into 1,2-Indanedione as a Reagent for the Detection of Latent Fingermarks*, Curtin University of Technology.

[46] Wilkinson, D. (2000). Spectroscopic study of 1,2-indandione, *Forensic Science International* **114**(3), 123–132.

[47] Almog, J., Cohen, Y., Azoury, M. & Hahn, T.-R. (2004). Genipin – a novel fingerprint reagent with colorimetric and fluorogenic activity, *Journal of Forensic Sciences* **49**(2).

[48] Levinton-Shamuilov, G., Cohen, Y., Azoury, M., Chaikovsky, A. & Almog, J. (2005). Genipin, a novel fingerprint reagent with colorimetric and fluorogenic activity. Part II: Optimization, scope and limitations, *Journal of Forensic Sciences* **50**(6).

[49] Cantu, A.A. (2001). On the composition of silver physical developers used to visualize latent prints on paper, *Forensic Science Review* **13**, 29–64.

[50] Cantu, A.A. & Johnson, J.L. (2001). Silver physical development of latent prints, in *Advances in Fingerprint Technology*, 2nd Edition, H.C. Lee & R.E. Gaensslen, eds, CRC Press, Boca Raton, FL, pp. 241–274.

[51] Ramotowski, R. (1996). Importance of an acid prewash prior to the use of physical developer, *Journal of Forensic Identification* **46**(6), 673–677.

[52] Ramotowski, R. (2000). A comparison of different physical developer systems and acid pre-treatments and their effects on developing latent prints, *Journal of Forensic Identification* **50**(4), 363–384.

[53] Wilson, J.D., Cantu, A.A., Antonopoulos, G. & Surrency, M.J. (2007). Examination of the steps leading up to the physical developer process for developing fingerprints, *Journal of Forensic Sciences* **52**(2), 320–329.

[54] Phillips, C.E., Cole, D.O. & Jones, G.W. (1990). Physical developer: a practical and productive latent print developer, *Journal of Forensic Identification* **40**(3), 135–147.

[55] Burow, D., Seifert, D. & Cantu, A.A. (2003). Modifications to the silver physical developer, *Journal of Forensic Sciences* **48**(5).

[56] Burow, D. (2003). An improved silver physical developer, *Journal of Forensic Identification* **53**(3), 304–314.

[57] Yapping, L. & Yue, W. (2004). A new silver physical developer, *Journal of Forensic Identification* **54**(4), 422–427.

[58] Cantu, A.A. (2005). Letter to the editor: re: a new silver physical developer, *Journal of Forensic Identification* **55**(3), 289–290.

[59] Beaudoin, A. (2004). New technique for revealing latent fingerprints on wet, porous surfaces: Oil Red O, *Journal of Forensic Identification* **54**(4), 413–421.

[60] Rawji, A. & Beaudoin, A. (2006). Oil Red O versus physical developer on wet papers: a comparative study, *Journal of Forensic Identification* **56**(1), 33–54.

[61] Salama, J., Aumeer-Donovan, S.B., Lennard, C. & Roux, C. (2008). Evaluation of Oil Red O as a replacement for physical developer, *Journal of Forensic Identification* **58**(2), 203–237.

[62] Neumann, C., Mack, N. & Wilson, G. (2007). Evaluation of Oil Red O. *6th Biennial International Fingerprint Research Group Meeting*, Canberra, Australia.

[63] Cantu, A.A., Burow, D. & Wilson, J. (2007). *On Some Properties of Oil Red O Latent Print Visualisation Reagent*, Canberra, Australia, 26–30 March, 2007.

[64] Stimac, J.T. (2003). Thermal and carbonless papers: a fundamental understanding for latent friction ridge development, *Journal of Forensic Identification* **53**(2), 185–197.

[65] Schwartz, L. & Klenke, I. (2007). Enhancement of ninhydrin- or DFO-treated latent fingerprints on thermal paper, *Journal of Forensic Sciences* **52**(3), 649–655.

[66] Stimac, J.T. (2003). Thermal paper: latent friction ridge development via 1,2-indanedione, *Journal of Forensic Identification* **53**(3), 265–271.

[67] Wakefield, M. & Armitage, S. (2005). The development of latent fingerprints on thermal paper using a novel, solvent-free method, *Journal of Forensic Identification* **55**(2), 202–213.

[68] Ma, R. & Wei, Q. (2006). Chemical fuming: a practical method for fingerprint development on thermal paper, *Journal of Forensic Identification* **56**(3), 364–373.

[69] Czekanski, P., Fasola, M. & Allison, J. (2006). A mechanistic model for the superglue fuming of latent fingerprints, *Journal of Forensic Sciences* **51**(6), 1323–1328.

[70] Wargacki, S.P., Lewis, L.A. & Dadmun, M.D. (2007). Understanding the chemistry of the development of latent fingerprints by superglue fuming, *Journal of Forensic Sciences* **52**(5), 1057–1062.

[71] Tissier, P., Didierjean, C., Prud'Homme, C., Prichard, J. & Crispino, F. (1999). A "cyanoacrylate case" for developing fingerprints in cars, *Science and Justice* **39**, 163–166.

[72] Watkin, J.E. & Misner, A.H. (1990). Fluorescence and crime scenes in the 90's, *RCMP Gazette* **52**(9), 1–5.

[73] Campbell, B.M. (1991). Vacuum chamber cyanoacrylate technique evolution, *RCMP Gazette* **53**(12), 12–16.

[74] Neri, R. (1992). *Détection des Empreintes Digitales au Cyanoacrylate: Technique "sous Vide" vs. Traditionelle*, Research Seminar, Insitut de Police Scientifique et de Criminiologie, Université de Lausanne, Switzerland.

[75] LeRoy, H. (1993). *Metal Vacuum Deposition and Vacuum Cyanoacrylate Development of Latent Prints. International Symposium on the Forensic Aspects of Latent Prints*, FBI Academy, Quantico, VA.

[76] Bentsen, R.K., Brown, J.K., Dinsmore, A., Harvey, K.K. & Kee, T.G. (1996). Post firing visualisation of fingerprints on spent cartridge cases, *Science and Justice* **36**, 3–8.

[77] Bessman, C.W., Nelson, E., Lipert, R.J., Coldiron, S. & Herrman, T.R. (2005). A comparison of cyanoacrylate fuming in a vacuum cabinet to a humidity fuming chamber, *Journal of Forensic Identification* **55**(1), 10–27.

[78] Menzel, E.R., Burt, J.A., Sinor, T.W., Tubach-Ley, W.B. & Jordan, K.J. (1983). Laser detection of latent fingerprints: treatment with glue containing cyanoacrylate ester, *Journal of Forensic Sciences* **28**(2), 307–317.

[79] Lennard, C.J. & Margot, P.A. (1988). Sequencing of reagents for the improved visualization of latent fingerprints, *Journal of Forensic Identification* **38**(5), 197–210.

[80] McCarthy, M.M. (1990). Evaluation of ardrox as a luminescent stain for cyanoacrylate processed latent impressions, *Journal of Forensic Identification* **40**(2), 75–80.

[81] Olenik, J. (1992). Ardrox: an alternate solvent system, *Journal of Forensic Identification* **42**(6), 513–516.

[82] Gamboe, M. & O'Daniel, L. (1999). Substitute Ardrox formula, *Journal of Forensic Identification* **49**, 134–141.

[83] Kent, T. (1986). *Manual of Fingerprint Development Techniques*, 1st Edition, Home Office Scientific Research and Development Branch, Brightman & Stratton Ltd, London.

[84] Hardwick, S.A., Kent, T. & Sears, V.G. (1990). *Fingerprint Detection by Fluorescence Examination: A Guide to Operational Implementation*, Home Office Police Scientific Development Branch, Sandridge, UK.

[85] Mazzella, W.D. & Lennard, C.J. (1995). An additional study of cyanoacrylate stains, *Journal of Forensic Identification* **45**(1), 5–18.

[86] Cummings, H., Hollars, M., Peigare, F. & Trozzi, T. (1990). BBD and MBD: multipurpose reagents for latent print detection, *Journal of Forensic Identification* **40**(6), 334–340.

[87] Day, K.J. & Bowker, W. (1996). Enhancement of cyanoacrylate developed latent prints using nile red, *Journal of Forensic Identification* **46**(2), 183–187.

[88] Misner, A.H., Wilkinson, D. & Watkin, J. (1993). Thenoyl europium chelate: a new fluorescent dye with a narrow emission band to detect cyanoacrylate developed fingerprints on non-porous substrates and cadavers, *Journal of Forensic Identification* **43**(2), 154–165.

[89] Wilkinson, D. & Watkin, J.E. (1993). Europium aryl-β-diketone complexes as fluorescent dyes for the detection of cyanoacrylate developed fingerprints on human skin, *Forensic Science International* **60**, 67–79.

[90] Lock, E.R.A., Mazzella, W.D. & Margot, P. (1995). A new europium chelate as a fluorescent dye for cyanoacrylate pretreated fingerprints – EuTTAPhen: europium thenoyltrifluoroacetone orthophenanthroline, *Journal of Forensic Sciences* **40**(4), 654–658.

[91] Cummings, H., Hollars, M. & Trozzi, T. (1993). Getting the most from cyanoacrylate dyes, *Journal of Forensic Identification* **43**(1), 37–43.

[92] Olenik, J.H. (1997). A simple three dye blend, *Journal of Forensic Identification* **47**(5), 530–533.

[93] Theys, P., Turgis, Y., Lepareux, A., Chevet, G. & Ceccaldi, P.F. (1968). Nouvelle technique de révélation de traces papillaires latentes (sur le papier) par métallisation sous vide, *Revue Internationale Police Criminelle* **23**(217), 106–108.

[94] Kent, T., Thomas, G.L., Reynoldson, T.E. & East, E.W. (1976). A vacuum coating technique for the development of latent fingerprints on polythene, *Journal of the Forensic Science Society* **16**, 93–101.

[95] Thomas, G.L. (1978). The physics of fingerprints and their detection, *Journal of Physics Series E: Scientific Instruments* **11**, 722–731.

[96] Grant, H., Springer, E. & Ziv, Z. (1996). *Vacuum Metal Deposition Inhibition on Polythene Bags. International Symposium on Fingerprint Detection and Identification*, Israel National Police, Ne'urim, Israel, pp. 203–215.

[97] Masters, N.E. & DeHaan, J.D. (1996). Vacuum metal deposition and cyanoacrylate detection of older latent prints, *Journal of Forensic Identification* **46**(1), 32–45.

[98] Jones, N. (2002). *Metal Deposition Techniques for the Detection and Enhancement of Latent Fingerprints on Semi-porous Surfaces*, University of Technology, Sydney, Sydney, Australia.

[99] Jones, N., Mansour, D., Stoilovic, M., Lennard, C.J. & Roux, C. (2001). The influence of polymer type, print donor and age on the quality of fingerprints developed on plastic substrates using vacuum metal deposition, *Forensic Science International* **124**, 167–177.

[100] Jones, N., Stoilovic, M., Lennard, C. & Roux, C. (2001). Vacuum metal deposition: factors affecting normal and reverse development of latent fingerprints on polyethylene substrates, *Forensic Science International* **115**(1–2), 73–88.

[101] Jones, N., Kelly, M., Stoilovic, M., Lennard, C. & Roux, C. (2003). The development of latent fingerprints on polymer banknotes, *Journal of Forensic Identification* **53**(1), 50–77.

[102] Dai, X., Stoilovic, M., Lennard, C. & Speers, N. (2007). Vacuum metal deposition: visualisation of gold agglomerates using TEM imaging, *Forensic Science International* **168**, 219–222.

[103] Levin Elad, M., Papo, P., Azoury, M. & Almog, J. (2007). VMD: the relationship between print quality and surface properties. *6th Biennial International Fingerprint Research Group Meeting*, Canberra, Australia.

[104] Yong, A.S.J. (1986). *Detection of Latent Fingerprints with Cyanoacrylates: New Techniques Involving Coloured and Photoluminescent Compounds*. PhD thesis. Australian National University, Canberra, Australia.

[105] Yong, A.S.J., Margot, P.A. & Warrener, R.N. (1986). Visual enhancement of superglue developed fingerprints. *9th Australian International Forensic Science Symposium*, Melbourne, Australia.

[106] Taroni, F. (1990). *Cyanoacrylate et Déposition Métallique: Comparaison de Leur Efficacité dans la Révélation des Empreintes Digitales*, Research Seminar, Institut de Police Scientifique et de Criminologie, Université de Lausanne, Switzerland.

[107] Taroni, F., Lennard, C.J. & Margot, P.A. (1990). Latent fingerprint detection on non-porous surfaces: comparison between metal deposition and cyanoacrylate fuming. *12th Meeting of the International Association of Forensic Sciences*, Adelaide, Australia.

[108] Murphy, M.P. (1991). A vacuum metal identification, *Journal of Forensic Identification* **41**(5), 318–320.

[109] Misner, A.H. (1992). Latent fingerprint detection on low density polyethylene comparing vacuum metal deposition to cyanoacrylate fuming and fluorescence, *Journal of Forensic Identification* **42**(1), 26–33.

[110] Saunders, G.C. (1989). Multimetal deposition technique for latent fingerprint development. *74th IAI Educational Conference Proceedings*, Pensacola, FL.

[111] Allman, D.S., Maggs, S.J. & Pounds, C.A. (1992). *The Use of Colloidal Gold/Multimetal Deposition for the Detection of Latent Fingerprints – A Preliminary Evaluation: Central Research and Support Establishment*, Home Office Forensic Science Service, UK.

[112] Schnetz, B. (1999). *La Révélation des Empreintes Digitales par l'or Colloïdal: l'amplification par des Techniques Biochimiques*. PhD Thesis. Université de Lausanne, Switzerland.

[113] Schnetz, B. & Margot, P. (2001). Technical note: latent fingermarks, colloidal gold and multimetal deposition (MMD). Optimisation of the method, *Forensic Science International* **118**(1), 21–28.

[114] Becue, A., Champod, C. & Margot, P. (2007). Use of gold nanoparticles as molecular intermediates for the detection of fingermarks, *Forensic Science International* **168**, 169–176.

[115] Stauffer, E., Becue, A., Singh, K.V., Thampi, K.R., Champod, C. & Margot, P. (2007). Single-metal deposition (SMD) as a latent fingermark enhancement technique: an alternative to multimetal deposition (MMD), *Forensic Science International* **168**, e5–e9.

[116] Choi, M.J., McBean, K.E., Wuhrer, R., McDonagh, A.M., Maynard, P.J. & Lennard, C. (2006). Investigation into the binding of gold nanoparticles to fingermarks using scanning electron microscopy, *Journal of Forensic Identification* **56**(1), 24–32.

[117] Choi, M.J., McDonagh, A.M., Maynard, P.J., Wuhrer, R., Lennard, C. & Roux, C. (2006). Preparation and evaluation of metal nanopowders for the detection of fingermarks on nonporous surfaces, *Journal of Forensic Identification* **56**(5), 756–768.

[118] Almog, J., Sametband, M. & Mandler, D. (2007). Application of nanoparticles for the enhancement of latent fingerprints. *6th Biennial International Fingerprint Research Group Meeting*, Canberra, Australia.

[119] Adcock, J.M. (1977). The development of latent fingerprints on human skin: the iodine-silver plate transfer method, *Journal of Forensic Sciences* **22**, 599–605.

[120] Gray, C. (1978). The detection and persistence of latent fingerprints on human skin: an assessment of the iodine-silver plate method, *Journal of the Forensic Science Society* **18**, 47–52.

[121] Arndt, C.B. (1985). Iodine silver plate transfer method, *RCMP Gazette* **47**(5), 19–21.

[122] Haque, F., Westland, A. & Kerr, F.M. (1983). An improved non-destructive method for detection of latent fingerprints on documents with iodine-7,8-benzoflavone, *Forensic Science International* **21**, 79–83.

[123] Mashiko, K. & Hizaki, M.I. (1977). Latent fingerprint processing: iodine 7,8-benzoflavone method, *Identification News* **27**(11), 3–5.

[124] Mashiko, K. & Miyamoto, T. (1998). Latent fingerprint processing by the ruthenium tetroxide method, *Journal of Forensic Identification* **48**(3), 279–290.

[125] Olsen, R.D. (1978). *Scott's Fingerprint Mechanics*, C.C. Thomas, Springfield, IL.

[126] Mashiko, K., German, E.R., Motojima, K. & Colman, C.D. (1991). RTX: a new ruthenium tetroxide fuming procedure, *Journal of Forensic Identification* **41**(6), 429–436.

[127] Benzoni, M. (1994). *Evaluation du RTX pour la Révélation des Empreintes Digitales*, Research Seminar, Institut de Police Scientifique et de Criminologie, Université de Lausanne, Switzerland.

[128] Blackledge, R.D. (1998). Letters to the editor: re: "Latent print processing by the ruthenium tetroxide method", Journal of Forensic Identification 48 (3), *Journal of Forensic Identification* **48**(5), 557–559.

[129] Flynn, K., Maynard, P., Du Pasquier, E., Lennard, C., Stoilovic, M. & Roux, C. (2004). Evaluation of iodine-benzoflavone and ruthenium tetroxide spray reagents for the detection of latent fingermarks at the crime scene, *Journal of Forensic Sciences* **49**(4).

[130] Haslett, M. (1983). Fingerprints from skin using the Magnabrush technique, *Fingerprint Whorld* **8**, 118–119.

[131] Hébrard, J. & Donche, A. (1994). Fingerprints detection methods on skin: experimental study on 16 live subjects and 23 cadavers, *Journal of Forensic Identification* **44**, 623–627.

[132] Wilkinson, D., Watkin, J.E. & Misner, A.H. (1996). A comparison of techniques for the visualisation of fingerprints on human skin including the application of iodine and α-naphthoflavone, *Journal of Forensic Identification* **46**, 432–453.

[133] Dolci, D. (1992). *Révélation des Empreintes Digitales sur la Peau*, Institut de Police Scientifique et de Criminologie, Université de Lausanne, Switzerland.

[134] Delmas, B.J. (1988). Postmortem latent print recovery from skin surfaces, *Journal of Forensic Identification* **38**(2), 49–56.

[135] Fortunado, S.L. & Walton, G. (1998). Development of latent fingerprints from skin, *Journal of Forensic Identification* **48**, 704–717.

[136] Hamilton, J. & Dibattista, J. (1985). Cyanoacrylate ester – latent print from murdered body, *Fingerprint Whorld* **11**(18–19).

[137] Sampson, W.C. (1992). Glass recovery investigative technique: G.R.I.T, *Journal of Forensic Identification* **42**, 96–100.

[138] Sampson, W.C. (1995). Glass recovery investigative technique: G.R.I.T, *Fingerprint Whorld* **21**, 55–56.

[139] Hoyser, R.L. (1992). An alternative G.R.I.T. technique, *Journal of Forensic Identification* **42**, 219–220.

[140] Guo, Y.-C. & Xing, L.-P. (1992). Visualization method for fingerprints on skin by impression on a polyethylene terephthalate (PET) semigrid sheet, *Journal of Forensic Sciences* **37**(2), 604–611.

[141] Reichardt, G.J., Carr, J.C. & Stone, E.G. (1979). A conventional method for lifting latent fingerprints from human skin surfaces, *Journal of Forensic Sciences* **24**(1), 135–141.

[142] Hamm, E.D. (1988). A latent from human skin or is it? *Fingerprint Whorld* **14**, 56–57.

[143] Sampson, W.C. & Sampson, K.L. (2005). Recovery of latent prints from human skin, *Journal of Forensic Identification* **55**(3), 362–385.

[144] Choudhry, M.Y. & Whritenour, R.D. (1990). A new approach to unraveling tangled adhesive tape for potential detection of latent prints and recovery of trace evidence, *Journal of Forensic Sciences* **35**, 1373–1383.

[145] Campbell, B.M. (1991). Separation of adhesive tapes, *Journal of Forensic Identification* **41**, 102–106.

[146] Stimac, J.T. (2000). Adhesive tape separation with Un-Du®, *Fingerprint Whorld* **26**, 153–157.

[147] Schwartz, R.L., Higginbotham, L.C. & Smith, D.R. (2003). The effect of Un-Du® on latent fingerprint development, *Fingerprint Whorld* **29**, 66–73.

[148] Midkiff, C.R. (1994). Development of prints on tape, *Fingerprint Whorld* **20**, 5–7.

[149] Midkiff, C.R. & Codell, D.E. (1995). Development of latent prints on tape. Part II, *Fingerprint Whorld* **21**, 21–26.

[150] Midkiff, C.R., Codell, D.E. & Chapman, J. (1997). Development of prints on tape. Part III, *Fingerprint Whorld* **23**, 83–86.

[151] Ziv, Z. & Springer, E. (1993). More applications of coaxial illumination in fingerprint detecting and photography, *Journal of Forensic Identification* **43**(4), 362–367.

[152] Bramble, S.K., Cantu, A.A., Ramotowski, R.S. & Brennan, J.S. (2000). Deep red to near infrared (NIR) fluorescence of gentian violet-treated latent prints, *Journal of Forensic Identification* **50**(1), 33–49.

[153] Yamashita, A.B. (1994). Use of a benchtop desiccator for vacuum cyanoacrylate treatment of latent prints, *Journal of Forensic Identification* **44**(2), 149–158.

[154] Burns, D.S. (1994). Sicky-side powder: the Japanese solution, *Journal of Forensic Identification* **44**(2), 133–138.

[155] Martin, B.L. (1999). Developing latent prints on the adhesive surface of black electrical tape, *Journal of Forensic Identification* **49**(2), 127–129.

[156] Parisi, K.M. (1999). Getting the most from fingerprint powders, *Journal of Forensic Identification* **49**(4), 494–498.

[157] Kimble, G.W. (1996). Powder suspension processing, *Journal of Forensic Identification* **46**(3), 273–280.

[158] Wade, D.C. (2002). Development of latent prints with titanium dioxide (TiO_2), *Journal of Forensic Identification* **52**(5), 551–559.

[159] Williams, N.H. & Elliott, K.T. (2005). Development of latent prints using titanium dioxide (TiO_2) in small particle reagent, white (SPR-W) on adhesives, *Journal of Forensic Identification* **55**(3), 292–305.

[160] Schiemer, C., Lennard, C., Maynard, P. & Roux, C. (2005). Evaluation of techniques for the detection and enhancement of latent fingermarks on black electrical tape, *Journal of Forensic Identification* **55**(2), 214–238.

[161] Wiesner, S., Springer, E. & Argaman, U. (1996). A closer look at the effects of the shooting process on fingerprint development on fired cartridge cases, *Proceedings of the International Symposium on Fingerprint Detection and Identification*, Israel National Police, Ne'urim, Israel, pp. 161–178.

[162] Barnum, C.A. & Klasey, D.R. (1997). Factors affecting the recovery of latent prints on firearms, *Journal of Forensic Identification* **47**(2), 141–149.

[163] Sampson, W.C. (1993). An inquiry into the methodology of preserving and developing latent prints on expended cartridge casings, *Journal of Forensic Identification* **43**(1), 4–12.

[164] Donche, A. & Musy, C. (1994). Development of latent fingerprints on cartridge casings, *Fingerprint Whorld* **20**, 13–19.

[165] Klasey, D.R. & Barnum, C.A. (2000). Development and enhancement of latent prints on firearms by vacuum and atmospheric cyanoacrylate fuming, *Journal of Forensic Identification* **50**(6), 572–580.

[166] Migron, Y. & Mandler, D. (1997). Development of latent fingerprints on unfired cartridges by palladium deposition: a surface study, *Journal of Forensic Sciences* **42**(6), 986–992.

[167] Cantu, A.A., Leben, D.A., Ramotowski, R., Kopera, J. & Simms, J.R. (1998). Use of acidified hydrogen peroxide to remove excess gun blue from gun blue-treated cartridge cases and to develop latent prints on untreated cartridge cases, *Journal of Forensic Sciences* **43**(2), 294–298.

[168] Vouk, D.M. (1996). The effects of gold gun blue on individual characteristics on cartridge casings, *AFTE Journal* **28**, 256–257.

[169] Schütz, F., Bonfanti, M. & Champod, C. (1999). La révélation des traces papillaires sur les douilles par les techniques de etching et de blueing et comparaison avec la déposition multimétallique, *Journal of the Canadian Forensic Science Society* **33**, 65–81.

[170] Migron, Y., Hocherman, G., Springer, E., Almog, J. & Mandler, D. (1998). Visualization of sebaceous fingerprints on fired cartridge cases: a laboratory study, *Journal of Forensic Sciences* **43**(3), 543–548.

[171] Migron, Y., Mandler, D., Frank, A., Springer, E. & Almog, J. (1996). Is a fingerprint left on a fired cartridge? The development of latent fingerprints on metallic surfaces by palladium deposition, *International Symposium on Fingerprint Detection and Identification*, Israel National Police, Ne'urim, Israel, pp. 203–215.

[172] Sturelle, V., Cominotti, C., Henrot, D. & Desbrosse, X. (2006). The use of camphor in the development of latent prints on unfired cartridge casings, *Journal of Forensic Identification* **56**(5), 694–705.

[173] Creighton, J.T. (1997). Visualisation of latent impressions after incidental or direct contact with human blood, *Journal of Forensic Identification* **47**, 534–541.

[174] Huss, K., Clark, J. & Chisum, W.J. (2000). Which was first – fingerprints or blood? *Journal of Forensic Identification* **50**, 344–350.

[175] Langenberg, G. (2007). Analysis of factors during the deposition of friction ridge skin impressions in a blood matrix. *6th Biennial International Fingerprint Research Group Meeting*, Canberra, Australia.

[176] Stoilovic, M. (1991). Detection of semen and blood stains using Polilight as a light source, *Forensic Science International* **51**, 289–296.

[177] Springer, E., Almog, J., Frank, A., Ziv, Z., Bergman, P. & Gui Qiang, W. (1994). Detection of dry body fluids by inherent short wavelength UV luminescence, *Forensic Science International* **66**(164–168).

[178] Hussain, J.I. & Pounds, C.A. (1988). *The Enhancement of Marks in Blood, Part I, 5-Sulphosalicylic Acid: A Convenient and Effective Fixative for Marks Made in Blood*, Central Research and Support Establishment, Aldermaston, UK.

[179] Sears, V.G. & Prizeman, T.M. (2000). Enhancement of fingerprints in blood, Part I: the optimization of amido black, *Journal of Forensic Identification* **50**, 470–480.

[180] McCarthy, M.M. & Grieve, D.L. (1989). Preprocessing with cyanoacrylate ester fuming for fingerprint impressions in blood, *Journal of Forensic Identification* **39**(1), 23–31.

[181] Sears, V.G., Butcher, C.P.G. & Prizeman, T.M. (2001). Enhancement of fingerprints in blood, Part 2: protein dyes, *Journal of Forensic Identification* **51**, 28–38.

[182] Caldwell, J.P., Henderson, W. & Kim, N.D. (2000). ABTS: a safe alternative to DAB for the enhancement of blood fingerprints, *Journal of Forensic Sciences* **45**, 785–794.

[183] Lee, H.C. (1984). TMB as an enhancement reagent for bloody prints, *Identification News* **34**(3), 10–11.

[184] Lee, H.C. (1984). Benzidine or o-tolidine, *Identification News* **34**(1), 13–14.

[185] Shipp, E., Fassett, M., Wright, R. & Togneri, E. (1994). Tetramethylbenzidine to the rescue, *Journal of Forensic Identification* **44**, 159–164.

[186] Cheeseman, R. & DiMeo, L.A. (1995). Fluorescein as a field-worthy latent bloodstain detection system, *Journal of Forensic Identification* **45**, 631–646.

[187] Yapping, L. & Yue, W. (2004). Bloody latent fingerprint detection using LeuR6G, *Journal of Forensic Identification* **54**(5), 542–546.

[188] Jaret, Y., Heriau, M. & Donche, A. (1997). Transfer of bloody fingerprints, *Journal of Forensic Identification* **43**, 38–41.

[189] Sears, V.G., Butcher, C.P.G. & Fitzgerald, L.A. (2005). Enhancement of fingerprints in blood. Part 3. Reactive techniques, acid yellow 7, and process sequences, *Journal of Forensic Identification* **55**(6), 741–763.

[190] Bleay, S., Moore, J. & Deans, J. (2007). The recovery of fingerprints in blood from arson scenes. *6th Biennial International Fingerprint Research Group Meeting*, Canberra, Australia.

SHAHEEN AUMEER-DONOVAN, CHRIS LENNARD AND CLAUDE ROUX

Friction Ridge Skin: Interaction between Fingerprint Detection and DNA/Biological Material

Introduction

A single exhibit can hold many forms of forensic evidence. For example, a knife used as a murder weapon may hold blood and fibers from the victim on the blade, fingerprints and trace DNA of the offender on the handle, and soil and pollen traces when it was discarded in a garden bed as the offender fled the scene. A forensic examiner will employ a sequence of methods to exploit all available information from the exhibit.

DNA and fingerprints are considered to be two of the most valuable forms of forensic evidence, given their potential to individually link a person to a trace. Researchers demonstrated the ability to recover DNA profiles from fingerprints and skin traces in 1997 [1]. This development highlighted the importance of combining fingerprint enhancement and DNA analysis from the same piece of evidence. Given that DNA sampling methods would destroy any fingerprints present, the effect, if any, that fingerprint enhancement methods have on subsequent DNA profiling becomes a concern.

Numerous fingerprint enhancement techniques have been investigated to assess their impact on DNA profiling of fingerprints and biological stains exposed to the reagents used. These techniques are discussed in the following sections, and guidelines provided as to the sequencing of fingerprint and DNA evidence. Detailed explanations of DNA profiling or fingerprint reagents are not discussed in this section. This information can be found in **DNA**; **Short Tandem Repeats**; **DNA: Sources of**; **Friction Ridge Skin: Fingerprint Detection and Recovery Techniques**; **Friction Ridge Skin: Morphogenesis and Overview**.

Fingerprint Enhancement Techniques

Light Sources

Alternative Light Sources and Lasers. Argon ion lasers at up to 9 W for 20–30 min have been applied to bloodstains, and no difference in the restriction fragment length polymorphism (RFLP) DNA profiles between treated and untreated samples was detected [2, 3]. Other alternative light sources such as the Polilight (Rofin, Australia) have been successfully tested, showing no deleterious effects on the DNA profiling of biological stains and fingerprints [3–5].

UV Light. Ultraviolet (UV) light can be divided into two types: shortwave, with wavelengths below 300 nm, and longwave, with wavelengths between 300 and 400 nm. These two types have vastly different effects. When fingerprints or other biological fluids are exposed to longwave UV, successful DNA profiles have been achieved [3, 4, 6]. Shortwave UV, however, was found to prevent DNA profiling of bloodstains exposed for more than 30 s [3]. The DNA molecule is a strong absorber of UV radiation, which leads to gene damage and mutations [7], and at a wavelength of 254 nm, UV light will break the molecular bonds in the DNA. Given the short exposure time limits, which are not feasible for forensic examinations, shortwave UV light is not recommended if DNA profiling is required.

Powders and Physical Methods

Powders: Black, White, Magnetic, Metallic, and Stickyside. The application of powder is the most ubiquitous method in fingerprint development, given its ease of use, low cost, and effectiveness at crime scenes. The types of powders available are consequently numerous, from basic carbon and titanium granular powders to metallic flake and magnetized powders, each with different benefits and applications.

Many research groups have tested the effect of various powders on DNA recovery [6, 8–12]. All were able to recover profiles from powdered fingerprints or biological stains; however, one study found a decline of approximately 25% in the amount of DNA recovered from saliva stains powdered with white or black powder. An extensive study of 11 different powders found that most allowed DNA recovery; however, two metallic powders (BVDA special silver and Faurot aluminum bronze) totally inhibited DNA analysis [8]. However, the Home Office Scientific Development Branch (United Kingdom) [13] has recovered DNA from aluminum powdered prints up to one year after their deposition. For any use of powder, it is recommended that DNA extraction methods

are chosen so as to completely remove the powder prior to amplification. This is particularly important for metallic and magnetic powders.

Adhesive Lifters. After powdering fingerprints, a crime scene examiner may choose to "lift" the developed fingerprint with adhesive tape for preservation. DNA can be recovered from both the tapelift and from the remnants of the fingerprint on the surface [13, 14]. This may be useful for cold cases, where lifted fingerprints have been archived.

Small Particle Reagent (SPR), Physical Developer (PD), and Iodine/Benzoflavone. Small particle reagent (SPR) and physical developer (PD) are useful methods for developing fingerprints on surfaces that have been wet, as they react to the insoluble components of fingerprints. The physical developer method utilizes the preferential deposition of silver onto the fingerprint ridges.

Iodine fuming is one of the oldest methods of fingerprint development, and operates by the absorption of iodine by the fatty material in the fingerprint to create a brown deposit. The absorption is transient, but can be fixed by 7,8-benzoflavone application.

Iodine and SPR have been found to have no negative effects on the profiling of bloodstains [6, 15], provided adequate cleanup is employed during the extraction. Iodine, however, was found to reduce the quantity of DNA recovered from fingerprints [16]. Conversely, physical developer has been found to adversely affect the chance of recovering DNA, in that after 48 h profiles may be recovered, but the chance declines rapidly the longer a sample is left posttreatment [13]. Physical developer involves the use of silver nitrate and washing steps that could contribute to the loss of the DNA in fingerprints.

Chemical Reagents

For Use on Porous Surfaces (Paper and Cardboard)

Ninhydrin, 1,8-Diaza-9-Fluorenone (DFO), and 1,2-Indanedione. These reagents chemically react with the amino acids present in fingerprint traces to produce a colored and/or fluorescent deposit. Ninhydrin was the first to be employed in fingerprint development, and its effect on the recovery of DNA profiles from treated fingerprints and blood or saliva prints has been investigated [6, 16–19]. This research shows that ninhydrin does not significantly affect the recovery of profiles; however, the quantity of DNA may be reduced after treatment. Given the negligible effect it has on the recovery of DNA, ninhydrin has been tested as a screening tool to determine if swabs of skin cell traces contain sufficient DNA for a profile [19]. While 120 of the 158 swabs that gave a reaction to ninhydrin also produced a DNA profile, several false negative results led to the conclusion that this method may be useful in volume crime investigations rather than serious offenses.

The effect of 1,8-diaza-9-fluorenone (DFO) and 1,2-indanedione on the subsequent DNA profiling of treated fingerprints, blood prints, and saliva has been examined by various researchers [6, 11, 18, 20–22]. DFO produced results similar to ninhydrin profiles may be obtained; however, the quantity of DNA recovered will be lower. DNA profiling after indanedione treatment was also successful; however, samples should be extracted quickly after treatment, as those left for six days proved negative [21].

For Use on Nonporous Surfaces (Glass and Plastic)

Cyanoacrylate fuming (CAF) and subsequent staining. The action of cyanoacrylate fuming (CAF) in the presence of fingerprints causes the ester to polymerize as a white residue layer. This layer can then be stained to enhance the visibility or fluorescence of the fingerprint. Several studies have tested whether CAF and various combinations of stains will affect DNA recovery. Full profiles were recovered after CAF and, variously, Basic Yellow 40, Crystal Violet, and Rhodamine 6G staining [5, 6, 10, 11, 23, 24]. However, cyanoacrylate has been found to inhibit the polymerase chain reaction (PCR) process in at least one study [24]. Extraction methods must therefore adequately remove the fingerprint reagents prior to amplification. It should also be noted that the layer of polymer must be adequately removed during sampling, so that the DNA-containing cells can be collected. Vigorous swabbing, or, if possible, cutting the substrate into pieces may assist this process.

Gentian Violet. Gentian violet is used to develop fingerprints on the adhesive side of tape. It has been found to have some affect on DNA recovery from

fingerprints with partial profiles resulting [13], but no affect on DNA recovery from saliva or bloodstains [6]. Phenol in the gentian violet formulation may be a reason for the DNA degradation; however, Paraben K has been proffered as an alternative [6].

Vacuum Metal Deposition (VMD). Vacuum metal deposition (VMD) is considered to be the most sensitive method for developing fingerprints on nonporous and semiporous surfaces, for example, polymer banknotes and leather [25]. The method is a two-stage process under vacuum, where a monolayer of gold is deposited on the surface but does not cover the fingerprint ridges, followed by a thicker layer of zinc, leaving the ridges clear against a silver-colored background. VMD may follow cyanoacrylate fuming in a sequence, and full profiles have been recovered from fingerprints on plastic treated in this sequence, indicating that VMD does not inhibit DNA analysis [4].

Other Examination Methods

Reagents for the Enhancement of Blood. Leucocrystal Violet (LCV), Amido Black, Coomassie Blue, Diaminobenzidine (DAB), Luminol, Fluorescein, Hungarian Red. It is often necessary at a crime scene to enhance marks (fingerprints, shoemarks, spatter patterns) in blood. The above-listed chemicals can increase the contrast between the blood and the substrate by producing a dark color, luminescence, or fluorescence when in contact with blood. These reagents, with the exception of diaminobenzidine (DAB), do not prevent the recovery of DNA profiles from treated bloodstains [6, 13, 18, 26, 27]; however, they can reduce the amount of DNA recovered, and may cause the need for extra purification during the extraction process. DAB has been found to inhibit DNA analysis, particularly if left in contact with the stain for longer than a few minutes [13]. The research indicated that DAB may have a degradative effect on the DNA molecule rather than inhibiting the extraction or amplification.

Electron Beam Irradiation. The use of pathogenic organisms such as anthrax as a terrorist weapon has become an issue in recent years. Items believed to be contaminated with a biological agent may be exposed to electron beam irradiation to neutralize the organism and render the item safe. The item would then be subjected to standard forensic examinations. DNA may still be recovered after this irradiation process, as profiles were recovered from licked envelopes exposed to 51.6-kGy irradiation, up to 56 days after the treatment [28]. Some evidence of DNA degradation was observed in the treated samples in terms of lower recovered quantities.

Powder Brush DNA Transfer. It is possible to recover DNA profiles from the applicators used to brush the powder over fingerprints and biological stains, which may be present at crime scenes [8, 12, 29–31]. Conceivably then, it would seem plausible that DNA could be transferred from one scene to another via the powder applicator. This phenomenon has been demonstrated in research [12, 30], but would seem less likely to occur in casework, given that the proportion of contaminant DNA "added" to a biological trace would be small and difficult to detect. However, the possibility exists and should be considered at every crime scene. To prevent such a chance the following recommendations are given [12]:

- For serious cases, consider the use of disposable or cleanable brushes, and smaller aliquots of powder.
- Use magnetic powder, whereby the "wand" applicator does not touch the surface. It should be noted that additional cleanup methods may be required during DNA extraction to remove all magnetic powder prior to amplification.
- Powdering wet areas and biological stains facilitates DNA transfer and should be avoided.
- Swab areas of an object unsuitable for fingerprints prior to fingerprint examination.
- Consider the use of other fingerprint methods, such as CAF.

Summary

Given the random nature of trace DNA sampling, definitive answers to the factors affecting its successful profiling can be difficult to give accurately. However, due the volume of research on the subject, certain guidelines can be offered with respect to the sequencing of fingerprint and DNA examinations.

With a few exceptions, fingerprint reagents will not prevent the subsequent DNA profiling of the treated substrate. The exceptions include shortwave UV light, physical developer, and DAB. Although,

the examiner must be aware that the quantity of DNA recovered will be lower after any treatment, potentially because of physical actions during the fingerprint treatment such as washing steps. Fuming with cyanoacrylate may protect DNA-containing skin cells with a layer of polymer and therefore lower the chance of loss; however, this may also make the cells difficult to remove during sampling.

Care should be taken by staff performing the DNA extraction to ensure that all traces of the fingerprint reagent have been removed before amplification. Extra cleanup methods may be required. Just 5 μg of fingerprint powder in the amplification mix can inhibit the PCR reaction [32].

Therefore, if DNA and fingerprint analysis is required of the same item, the following actions are suggested:

- Consider swabbing areas on the item that are unsuitable for fingerprint examination, for example, rough areas of a gun, the edge of a trigger, or the rim of drinking glass.
- If fingerprints that have been powdered are lifted, consider swabbing the remainder of the mark left on the substrate.
- Conduct one single treatment for fingerprints first, preferably a method with fewer steps. The single treatment can include the use of CAF followed by one stain.
- Do not use shortwave UV light, DAB, or physical developer where possible.
- Limit the time the item is in contact with the reagent.
- Use caution to avoid contamination during the fingerprint examination. For example, use fresh solution or a spray method, use fresh powder and brushes, and wear face masks and protective clothing.
- Sample and extract the DNA as soon as possible after the fingerprint treatment.
- Clearly detail on laboratory submission forms in which treatments have been applied, for the information of the DNA analysts.
- Laboratory staff should employ an extraction method that will adequately clean up treated samples.

References

[1] Van Oorschot, R.A. & Jones, M. (1997). DNA fingerprints from fingerprints, *Nature* **387**(6635), 767.

[2] Shipp, E., Roelofs, R., Togneri, E., Wright, R., Atkinson, D. & Henry, B. (1993). Effects of argon laser light, alternate source light, and cyanoacrylate fuming on DNA typing of human bloodstains, *Journal of Forensic Science* **38**(1), 184–191.

[3] Anderson, J. & Bramble, S. (1997). The effects of fingermark enhancement light sources on subsequent PCR-STR DNA analysis of fresh bloodstains, *Journal of Forensic Science* **42**(2), 303–306.

[4] Raymond, J.J., Roux, C., Du Pasquier, E., Sutton, J. & Lennard, C. (2004). The effect of common fingerprint detection techniques on the DNA typing of fingerprints deposited on different surfaces, *Journal of Forensic Identification* **54**(1), 22–44.

[5] Zamir, A., Springer, E. & Glattstein, B. (2000). Fingerprints and DNA: STR typing of DNA extracted from adhesive tape after processing for fingerprints, *Journal of Forensic Science* **45**, 687–688.

[6] Grubwieser, P., Thaler, A., Kochl, S., Teissl, R., Rabl, W. & Parson, W. (2003). Systematic study on STR profiling on blood and saliva traces after visualisation of fingerprints marks, *Journal of Forensic Science* **48**(4), 733–741.

[7] de Gruijl, F.R., van Kranen, H.J. & Mullenders, L.H.F. (2001). UV-induced DNA damage, repair, mutations and oncogenic pathways in skin cancer, *Journal of Photochemistry and Photobiology B: Biology* **63**(1–3), 19–27.

[8] Van Hoofstat, D.E.O., Deforce, D.L.D., De Pauw, I. & Van den Eeckhout, E.G. (1999). DNA typing of fingerprints using capillary electrophoresis: Effect of dactyloscopic powders, *Electrophoresis* **20**(14), 2870–2876.

[9] Pesaresi, M., Buscemi, L., Alessandrini, F., Cecati, M. & Tagliabracci, A. (2003). Qualitative and quantitative analysis of DNA recovered from fingerprints, in *Progress in Forensic Genetics 9*, Elsevier, pp. 947–951.

[10] Leemans, P., Vandeput, A., Vanderheyden, N., Cassiman, J.-J. & Decorte, R. (2006). Evaluation of methodology for the isolation and analysis of LCN-DNA before and after dactyloscopic enhancement of fingerprints, *International Congress Series* **1288**, 583–585.

[11] Bever, R.A., Gross, N. & Currence, S. (2002). DNA typing analysis from chemically processed fingerprints, *13th International Symposium on Human Identification*, Promega Corporation.

[12] Van Oorschot, R.A., Treadwell, S., Beaurepaire, J., Holding, N.L. & Mitchell, R.J. (2005). Beware of the possibility of fingerprinting techniques transferring DNA, *Journal of Forensic Science* **50**(6), 6.

[13] Kent, T., Sears, V., Bandey, H., Hart, A., Gibson, A. & Fitzgerald, L. (2003). *Fingerprint Development and Imaging Update*, Home Office Scientific Development Branch, London, Report No.: 6/2003.

[14] Schulz, M.M. & Reichert, W. (2002). Archived or directly swabbed latent fingerprints as a DNA source for STR typing, *Forensic Science International* **127**(1–2), 128–130.

[15] Zamir, A., Oz, C., Leifer, A. & Geller, B. (2002). The effect of small particle reagent employed as a fingerprint enhancement technique on subsequent STR typing from bloodstains, *Journal of Forensic Identification* **56**(6), 691–695.
[16] Balogh, M.K., Burger, J., Bender, K., Schneider, P. & Alt, K.W. (2003). STR genotyping and mtDNA sequencing of latent fingerprint on paper, *Forensic Science International* **137**(2–3), 188–195.
[17] Schulz, M.M., Wehner, H.D., Reichert, W. & Graw, M. (2004). Ninhydrin-dyed latent fingerprints as a DNA source in a murder case, *Journal of Clinical Forensic Medicine* **11**(4), 202–204.
[18] Fregeau, C.J., Germain, O. & Fourney, R.M. (2000). Fingerprint enhancement revisited and the effects of blood enhancement chemicals on subsequent *Profiler Plus*™ Fluorescent short tandem repeat DNA analysis of fresh and aged bloody fingerprints, *Journal of Forensic Science* **45**(2), 354–380.
[19] Anslinger, K., Selbertinger, U., Bayer, B., Rolf, B. & Eisenmenger, W. (2004). Ninhydrin treatment as a screening method for the suitability of swabs taken from contact stains for DNA analysis, *International Journal of Legal Medicine* **118**(2), 122–124.
[20] Zamir, A., Oz, C. & Geller, B. (2000b). Threat mail and forensic science: DNA profiling from items of evidence after treatment with DFO, *Journal of Forensic Science* **45**(2), 445–446.
[21] Azoury, M., Zamir, A., Oz, C. & Wiesner, S. (2002). The effect of 1,2-indanedione, a latent fingerprint reagent on subsequent DNA profiling, *Journal of Forensic Science* **47**(3), 586–588.
[22] Yu, P. & Wallace, M. (2007). Effect of 1,2-indanedione on PCR-STR typing of fingerprints deposited on thermal and carbonless paper, *Forensic Science International* **168**(2–3), 112–118.
[23] Wickenheiser, R.A. & Challoner, C.M. (1999). Suspect DNA profiles obtained from the handles of weapons recovered at crime scenes, *Tenth International Symposium on Human Identification*, Promega Corporation, Madison, Wisconsin.
[24] Von Wurmb, N., Meissner, D. & Wegener, R. (2001). Influence of cyanoacrylate on the efficiency of forensic PCRs, *Forensic Science International* **124**(1), 11–16.
[25] Stoilovic, M. & Lennard, C. (2006). *Fingerprint Detection and Enhancement*, 3rd Edition, Australian Federal Police, Canberra.
[26] Garofano, L., Pizzamiglio, M., Marino, A., Brighenti, A. & Romani, F. (2006). A comparative study of the sensitivity and specificity of luminol and fluorescein on diluted and aged bloodstains and subsequent STRs typing, *International Congress Series* **1288**, 657–659.
[27] Barbaro, A., Cormaci, P., Teatino, A. & Barbaro, A. (2004). Validation of forensic DNA analysis from bloodstains treated by presumptive test reagents, *International Congress Series* **1261**, 631–633.
[28] Withrow, A.G., Sikorsky, J., Upshaw Downs, J.C. & Fenger, T. (2003). Extraction and analysis of human nuclear and mitochondrial DNA from electron beam irradiated envelopes, *Journal of Forensic Science* **48**(6), 1–7.
[29] Sutherland, K.B.W., Cordiner, S.J., Bright, J. & Walsh, S.J. (2002). Commentary on, *Journal of Forensic Science* **47**(3), 442–450.
[30] Proff, C., Schmitt, C., Schneider, P.M., Foerster, G. & Rothschild, M.A. (2006). Experiments on the DNA contamination risk via latent fingerprint brushes, *International Congress Series* **1288**, 601–603.
[31] Wickenheiser, R.A. (2003). Trace DNA: a review, discussion of theory and application of the transfer of trace quantities of DNA through skin contact, *Journal of Forensic Science* **48**(2), 467.
[32] Van Oorschot, R.A., Phelan, D.G., Furlong, S., Scarfo, G.M., Holding, N.L. & Cummins, M.J. (2003). Are you collecting all the available DNA from touched objects? *International Congress Series* **1239**, 803–807; (Progress in Forensic Genetics 9).

JENNIFER J. RAYMOND, CLAUDE ROUX AND SIMON J. WALSH

Friction Ridge Skin: Morphogenesis and Overview

Anatomy

Outer Morphology of the Skin

The friction ridge skin occurs on the palmar sides of the hands and the plantar sides of the feet of all primates, including humans. As the name implies, the skin is ridged to produce friction. The friction aids in grasping and walking. The friction ridge skin also contains numerous sweat glands, and the pore openings are located along the tops of the ridges. The friction ridge skin does not contain sebaceous (oil) glands or hair follicles. In addition to the ridges, creases are found cress crossing through the friction ridge skin. These creases, in addition to the furrows between the ridges, permit flexibility of the skin. Figure 1 shows the friction ridge skin on a foot, palm, and finger; arrows point to ridges (R) and creases (C).

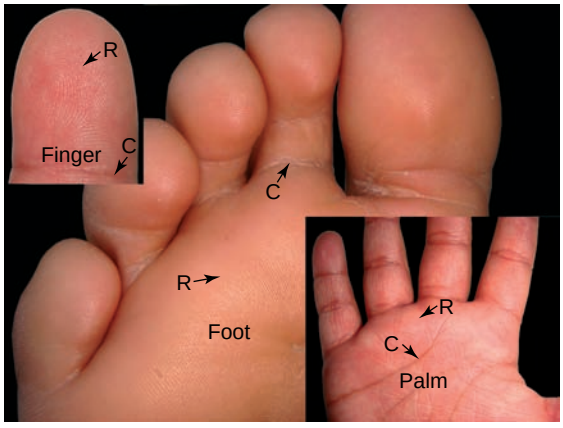


Figure 1 Friction ridge skin of the foot, palm, and finger showing the ridges (R) and creases (C)

Regions of the Hands and Feet

Areas of the hands and feet are described in relation to their orientation to the body. Proximal describes an area located toward the core of body while distal describes an area located away from the center of the body. For instance, when describing the hand, the wrist is located proximally (closer to the core of the body) and the tips of the fingers are located distally (further from the core of the body). The sides of the hands and feet are described relative to the bones of the forearm and lower leg, respectively. The thumb side of the hand is referred to as the *radial side* and the little finger side (blade of the palm) is referred to as the *ulnar side*. The great toe side (instep) of the foot is called the *tibial side* and the little toe side (outer edge) is called the *fibular side*.

The palmar side of the hand and soles of the feet are divided into different regions and contain prominent ridge flows delineated by deltas. A delta is the triradius formed where three ridge fields meet (Figure 2).

Regions of the Hand. The hand is divided into three main regions: interdigital, thenar, and hypothenar. As shown in Figure 3, the radial side of the palm is called the *thenar* (includes area labeled I) and the ulnar side of the palm is called the *hypothenar*. At the base of the palm, between the thenar and hypothenar, is the carpal (proximal) delta [1]. The interdigital region of the palm lies proximal to the fingers and contains the areas labeled II, III, and IV in Figure 3. The interdigital region also contains the digital deltas; there are typically four digital deltas (a, b, c, and d in Figure 3). The fingers are composed of three phalanges: distal, medial, and proximal. The thumb lacks a medial phalange. The phalanges are delineated by creases, which correspond to the underlying joints of the fingers.

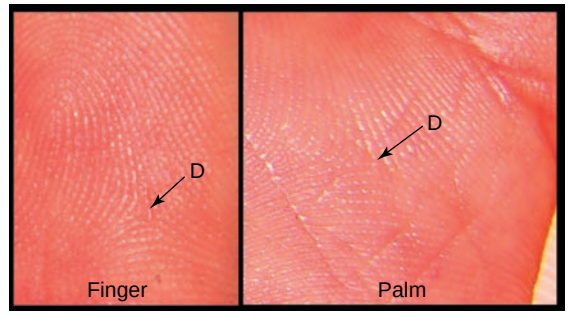


Figure 2 Deltas (D) on a finger and a palm

Regions of the Foot. The foot is divided into five primary regions: interdigital, hallucal, thenar, hypothenar, and calcar. The foot typically contains five digital deltas and a proximal delta. As shown in Figure 4, the interdigital area includes the areas marked II, III, and IV and digital deltas a, b, c, and d. The hallucal area corresponds to the ball of the foot and contains digital delta e. The interdigital and hallucal areas are demarcated from the rest of the foot by the ridge flows established by the proximal delta. The tibial side of the foot is called the *thenar* and the fibular side of the foot is called the *hypothenar*. The heel of the foot is called the *calcar area* [1]. Like the thumb, the great toe has a proximal and distal phalange. The remaining toes have proximal, medial, and distal phalanges.

Creases of the Hands and Feet

The general position of major creases is very robust between individuals due to their purpose – affording the hands and feet mobility. The major creases occur on all individuals with normal hand and foot morphology [2]. Figure 5 shows the normal configuration of the major creases of the hand. The only major creases of the foot are the creases associated with the toes. Additional creases on the hands and feet exist; however, these may not be present in all individuals.

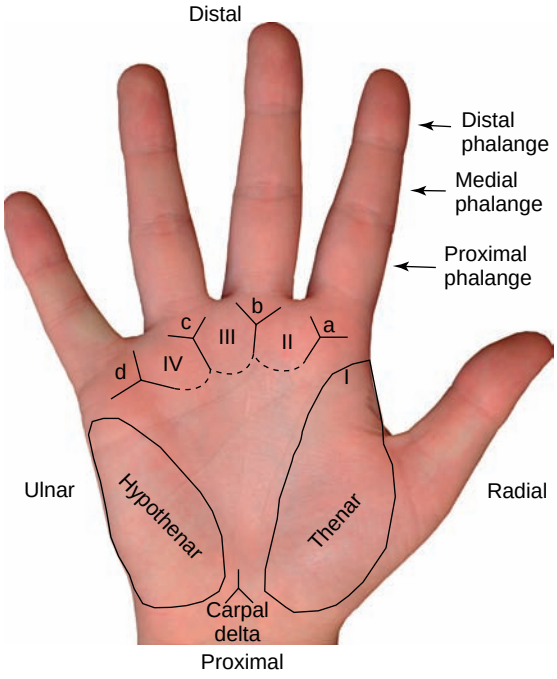


Figure 3 Areas of the palm showing the primary digital deltas (a, b, c, and d) and the carpal (proximal) delta. The thenar area is on the radial side of the palm and includes the region marked I; the hypothenar is on the ulnar side of the palm. The interdigital region contains areas II, III, and IV and the digital deltas. The phalanges of the fingers are marked as the distal, medial, and proximal

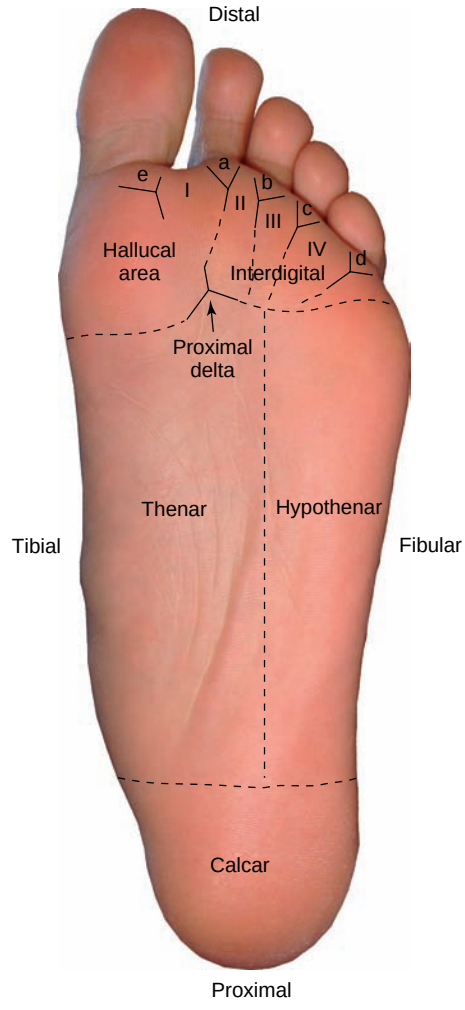


Figure 4 Areas of the foot showing the primary digital deltas (a, b, c, d, and e) and the proximal delta at the base of the ball of the foot. The hallucal region is the area under the great toe (including digital delta e); the interdigital region includes areas II, III, and IV and digital deltas a, b, c, and d; the fibular side of the foot is called the *hypothenar*; the tibial side of the foot is called the *thenar*; and the heel is called the *calcar area*

Anatomy of Friction Ridge Skin

The skin is composed of three principle layers: epidermis, dermis, and hypodermis. The superficial epidermis is a rapidly regenerating layer composed of stacked cells. The dermis is a supportive layer of connective tissue providing flexibility to the skin. The inner hypodermis is an insulating layer that cushions the skin.

The ridges and furrows seen on the outer surface of the epidermis reflect the complex structure beneath the surface of the skin (Figure 6). While the surface of the epidermis has alternating ridges and furrows, the bottom of the epidermis (where it is attached to the dermis), has alternating primary and secondary ridges. The primary ridges of the epidermis correspond to the surface ridges. The secondary ridges of the epidermis correspond to the surface furrows. The areas of dermis between adjacent primary and secondary ridges are called *dermal papillae.*

The sweat glands are appendages of the epidermis. As shown in Figure 6, the sweat glands are located in the primary ridges of the friction ridge skin. The coiled, secretory portions of the sweat glands are buried in the dermis. Sweat passes from the

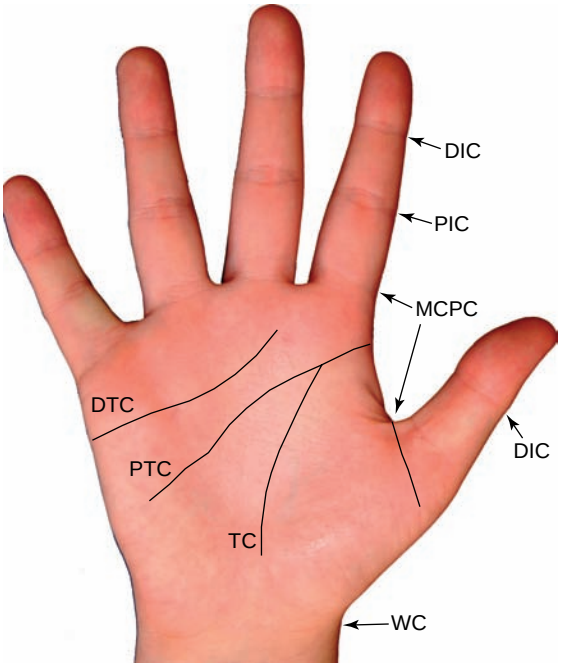


Figure 5 Major creases of the hands. Digital creases include the digital interphalangeal crease (DIC), the proximal interphalangeal crease (PIC), and the metacarpophalangeal crease (MCPC). The palmar creases include the distal transverse crease (DTC), the proximal transverse crease (PTC), thenar crease (TC), and the wrist crease (WC)

coiled portion, through the ducts in the primary ridges, and onto the surface of the ridge via a pore (Figure 7).

A fibrous sheet called the *basement membrane* tightly binds the primary and secondary ridges of the epidermis to the underlying dermis. This tight bond ensures that the ridges and furrows are locked into their configuration over the life of an individual, barring injury or disease.

Epidermis. The epidermis is a layered tissue composed primarily of skin cells called *keratinocytes*. The layers of the epidermis represent the biochemical changes that keratinocytes undergo as they are pushed from the bottom of the epidermis to the surface. As shown in Figure 8, the primary layers of the epidermis include the following: stratum basale (generating layer), stratum spinosum, stratum granulosum, stratum lucidum, and stratum corneum. There is also a specialized population of rapidly dividing cells in a suprabasal layer in the primary ridges. The

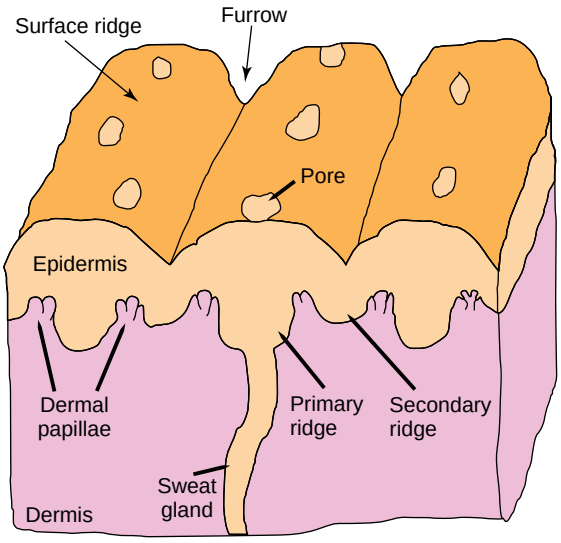


Figure 6 Cross section of the friction ridge skin

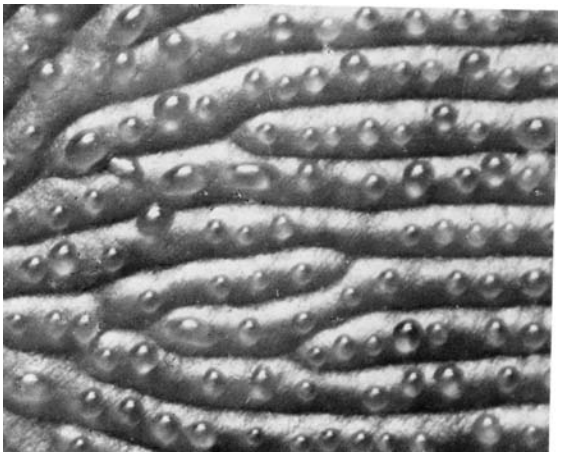


Figure 7 Pores of the friction ridge skin emitting sweat. [Reprinted from Montagna, p. 381) [3]

keratinocytes of the stratum basale are responsible for generating all the new cells that are pushed toward the surface to maintain the protective outer stratum corneum.

The keratinocytes of the epidermis are tightly bound to one another via intercellular junctions called *desmosomes*. This network of intercellular attachments prevents the cells from migrating laterally; the cells can only be pushed toward the surface of the skin. Movement toward the surface without lateral

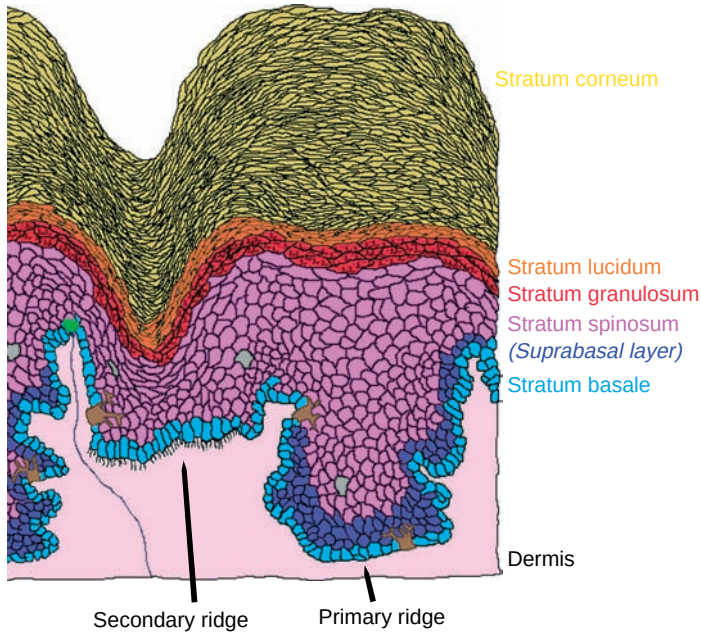


Figure 8 Layers of the epidermis

migration ensures that the configuration of ridges and furrows persists.

Other cells of the epidermis include melanocytes, Langerhans cells, and Merkel cells. The melanocytes provide pigmentation to the skin. The Langerhans cells are an extension of the immune system and Merkel cells are an extension of the nervous system [4].

Dermis. The dermis is the connective tissue layer that supports the epidermis. The dermis is composed primarily of fibers (collagen and elastin) and a gelatinous ground substance. The primary cell of the dermis is the fibroblast; it is responsible for maintaining both the fibrous and gelatinous structure of the dermis. Several immune response cells also populate the dermis: monocytes, macrophages, dermal dendrocytes, and mast cells [4]. The dermis contains capillary loops to provide nourishment to the epidermis and neural networks to provide sensitive touch discrimination.

Hypodermis. The hypodermis is the innermost layer of the skin. The hypodermis is composed primarily of subcutaneous fat. This fat cushions the skin, insulates the body, and serves as an energy reserve. It also permits mobility of the skin over underlying muscle and bone [4].

Physiology

Keratinization

The sheet of keratinocytes forming the stratum basale is responsible for dividing and pushing cells toward the surface of the epidermis. As the newly generated cells are pushed from the basal layer toward the surface, they undergo a process called *keratinization*. Keratinization prepares the cells for their ultimate purpose: providing a protective layer of dead, cornified cells on the exterior of the body. Figure 9 shows the cornified keratinocytes on the outer surface of the skin. The layers of the epidermis previously described represent the various stages of keratinization as the cells progress toward the surface of the skin.

Keratinocyte Proliferation

The keratinocytes of the stratum basale (basal cells) are responsible for continually dividing; pushing previously generated cells toward the surface to

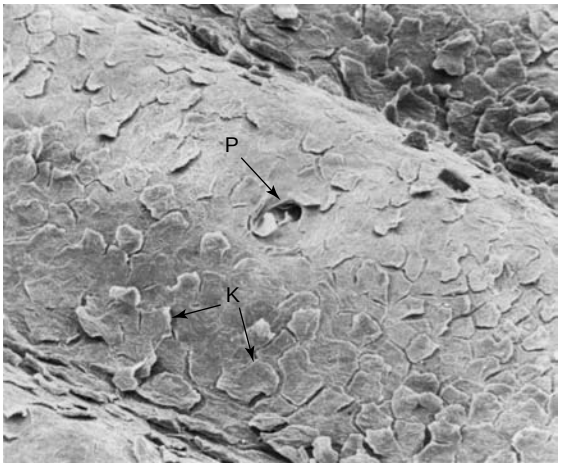


Figure 9 Surface of the friction ridge skin showing cornifed keratinocytes (K) and a sweat pore (P). [Reprinted from Montagna, p. 25]

replace those lost at the surface during exfoliation. The rate at which the basal cells divide must be equal to the rate at which cells are sloughing at the surface to maintain the appropriate skin thickness. The basal cells rely on chemical signals from the upper layers of the epidermis and from the dermis below. These signals either induce or inhibit cell division, enabling the skin to maintain the outer protective barrier [4]. If there is too much abrasion at the surface (resulting in a thinning of the stratum corneum), the basal cells will divide faster to provide additional replacement cells. If the skin is too thick, the cells will slow down the rate of division until enough cells at the surface slough off and the proper thickness is achieved. The skin's inherent drive to maintain the epidermis (ridges and furrows) contributes to the persistency of the friction ridge skin.

Persistence

The configuration of the friction ridge skin persists due to a combination of physical attachments and constant regulation of basal cell division in the epidermis. The basal keratinocytes proliferate in unison and are a template for the surface ridges and furrows. The intercellular attachments of the keratinocytes ensure that the cells migrate to the surface in concert, faithfully reproducing the surface features. The structure of the primary and secondary ridges of the epidermis and the protective dermis and hypodermis also provide durability to the skin.

Embryology

The friction ridges and major creases of the hands and feet develop on the growing fetus from approximately 8–18 weeks estimated gestational age (EGA). The shape and growth stresses of the fetal hand at the time of formation of creases and ridges have a direct impact on the position of creases and the general ridge flows (patterns) in the different regions of the hands and feet. While the shape and growth stresses of the fetal hand certainly guide the ridges and creases across the volar surface, the actual paths and shapes of the ridges and creases are left up to random events. This randomness imparts the friction ridge skin with its unique features. Figure 10 shows several ridge paths and a crease path.

Hand and Foot Development

The hands and feet undergo a similar developmental sequence; however, the development of the feet typically lags 0.5 week behind the development of the hands [5]. Between 6 and 13 weeks EGA, the hands and feet transform from flat paddles to recognizable hands and feet [2]. From 8 to 12 weeks EGA, however, there are special transient structures on

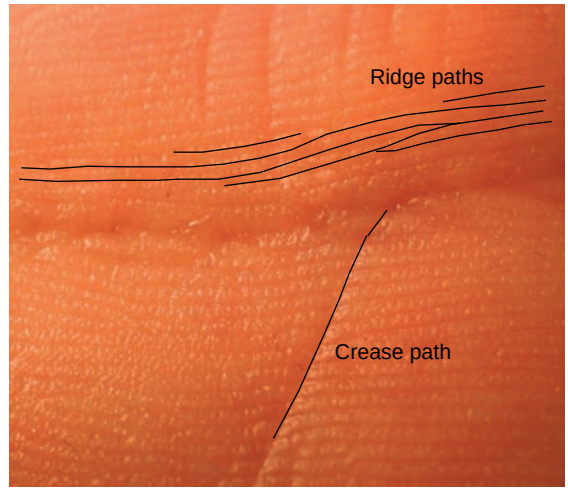


Figure 10 Ridge paths and a crease path on the palm of a hand

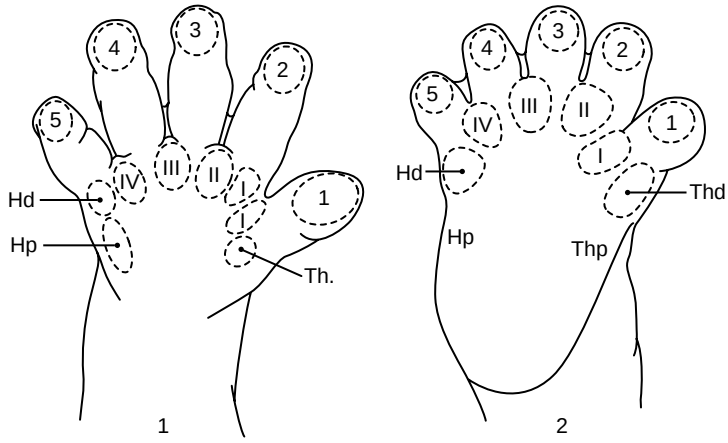


Figure 11 The volar pads of the hands and feet. The digital pads are marked as 1, 2, 3, 4, and 5. The interdigital pads are marked as I, II, III, and IV. The hypothenar pads are Hd (hypothenar distal) and Hp (hypothenar proximal). The thenar pads are marked as Th. (thenar), Thd (thenar distal), and Thp (thenar proximal) [Reprinted from Cummins 1929, p. 114.] [6]

the hands and feet that have a significant impact on the ridge flows and major creases seen on the hands and feet; these structures are called *volar pads*. The volar pads are localized swellings that occur on the hands and feet. As shown in Figure 11, there are prominent volar pads associated with distal phalanges, the interdigital regions, the hypothenar regions, and thenar regions of both the hands and feet [6].

Crease Formation. The major creases of the hand form concurrent with the volar pads, beginning at approximately 8 weeks EGA [2]. The thenar crease encircles the thenar pad (Th.) and the proximal portion of digital pad I. The proximal transverse cuts through the center of the palm and typically ends at the hypothenar pad (Hp). The distal transverse crease follows along the proximal edge of the interdigital volar pads III and IV and typically ends at interdigital pad II. The digital creases form at the joint locations of the fingers and thumbs. The digital creases of the feet start forming at approximately 9 weeks EGA [2].

Volar Pad Formation. The volar pads reach their maximum size at different times across the hands and feet. The volar pads of the palm peak in size at approximately 10 weeks EGA; the volar pads on the fingers (distal phalanges) peak at approximately 11 weeks EGA; and the volar pads of the feet peak at approximately 12 weeks EGA [5]. After the maximum size is attained, the volar pads discontinue growth and begin to disappear into the rapidly growing hand or foot. The size and shape of any volar pads present at the time of formation of the friction ridges have a direct impact on the general ridge flows in the hands and feet [5].

Friction Ridge Formation

The friction ridges typically form on the hands of the fetus at approximately 10 weeks gestation, and on the feet shortly thereafter. The general flow of ridges across the hands and feet is established by the growth stresses present on the hand or foot at the time of formation of the friction ridges. The predominant growth stress of the hands and feet is longitudinal (wrist to fingertip and heal to toe, respectively), causing the main ridge flows to cross the hand and foot laterally (side to side). If the developing ridges encounter a volar pad, the ridges are redirected based on the growth stresses established by the volar pad [7].

The fingers and toes tend to have intense patterns, as the volar pads are often quite prominent when the ridges are forming. Figure 12 shows the three primary pattern types that develop on the fingers: whorls, loops, and arches. These patterns reflect the size and shape of the volar pads at the time of formation of the friction ridges. Whorls tend to form on very high, rounded volar pads. Loops tend to form on

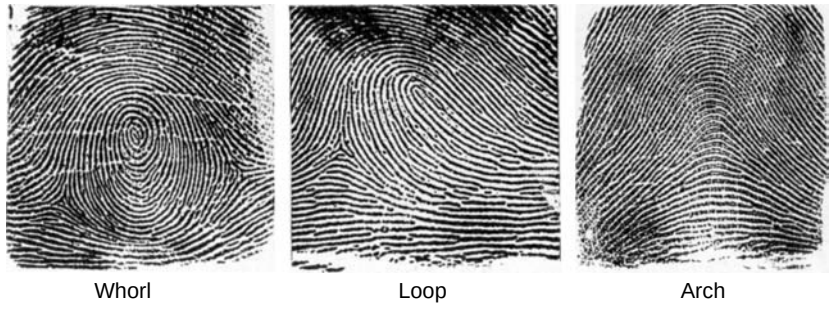


Figure 12 Three primary finger pattern types [Reprinted from Cummins 1943, p. 32 [8].]

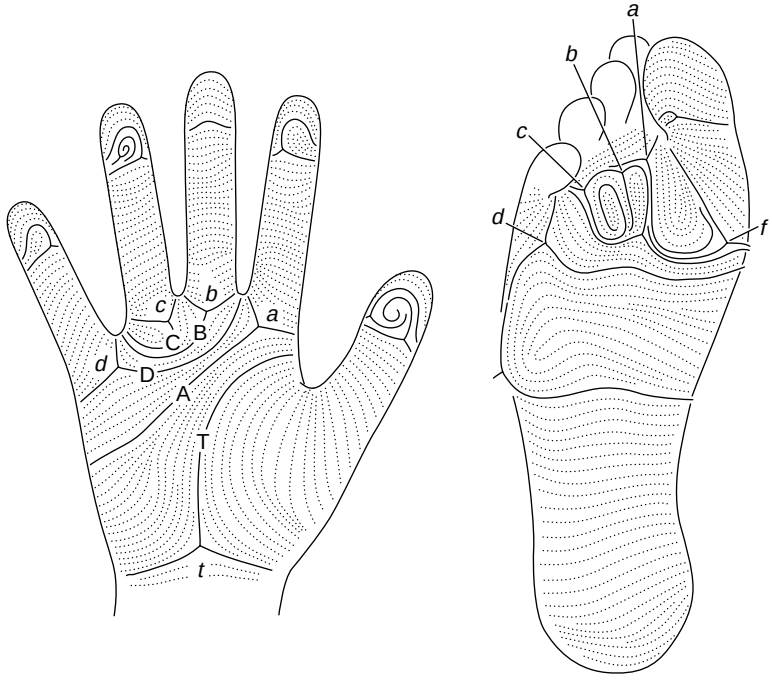


Figure 13 Predominant ridge flows on the hands and the feet. Digital deltas of the hands and feet are marked a, b, c, and d. The proximal delta of the hand is t and the proximal delta of the foot is f. On the hand, A, B, C, D, and T are the main ridge flows [Reprinted from Holt pp. 16, 24, respectively [1].]

asymmetrical volar pads that lean to one side. Arches tend to form on very low volar pads [5].

The palms and the feet may also have prominent volar pads when the friction ridges form. The palms and feet tend to have looping patterns in the interdigital region, including the hallucal area of the foot, that correspond to the interdigital pads. Occasionally, the thenar and hypothenar areas of the hands and feet have patterns; these patterns occur if the hypothenar or thenar pads are still prominent at the time of formation of friction ridge. Figure 13 shows the common ridge flows on the hands and feet.

Uniqueness

Although there is certainly variation to the general flows (patterns) of ridges on the palms and feet, this variation is limited by the general growth stresses

on the hands and feet. For instance, one would not find a hexagonal pattern of ridges since there is no fetal hand structure that would permit such a pattern. Within these general ridge flows, however, the specific paths and shapes of the ridges show infinite variability. This variability exists because the final path of each ridge and final three-dimensional shape of each ridge is not dictated during development.

During the development of any organism, the genetic code is responsible for ensuring that the organism has the proper body plan; genes are turned on and off in a regimented sequence to ensure that the limbs and organs have the appropriate form and functionality. There are many fine details of an organism; however, that are not programmed in the genetic code. For instance, the human body is programmed by the genetic code to develop hands and feet containing friction ridge skin on the volar surfaces. The overall ridge and crease patterns vary in a limited manner because the shape and function of the hand or foot is consistent between normal individuals.

The exact placement of each ridge and crease, however, is left up to the chance events that take place during development. These chance events are referred to as "developmental noise" [9] or "noise in gene expression" [10]. Essentially, during the developmental process of an organism, there are features that are not hardwired into the genetic code. Consequently, these features are at the mercy of random formation. The ridges and creases of the friction ridge skin are such features. The classic example is identical (monozygotic) twins who share the same genetic code but have different finger, palm, and footprints. Figure 14 is the left middle and index fingers from monozygotic twins.

As the ridges are formed on the fetal hands and feet, the hands and feet continue to grow, pulling existing ridges apart. As the ridges reach a critical separation distance, new ridges form in between to ensure the skin is continually ridged [11]. The process is dependent upon the growth of the fetus and the stresses on the epidermis during the development of the friction ridge skin. As a result, the length and path of each ridge is not determined by the genetic code of an individual, but instead by the unique forces at play on any particular area of the hand or foot during development.

The ridges continue to form until the fetus is approximately 18 weeks EGA. At this time, the friction ridge system stops adding new ridges. The ridges are locked into their configuration. The hands of an adult are obviously much larger than the hands of a fetus, but the relationship of the ridges to each other is conserved during growth to adulthood.

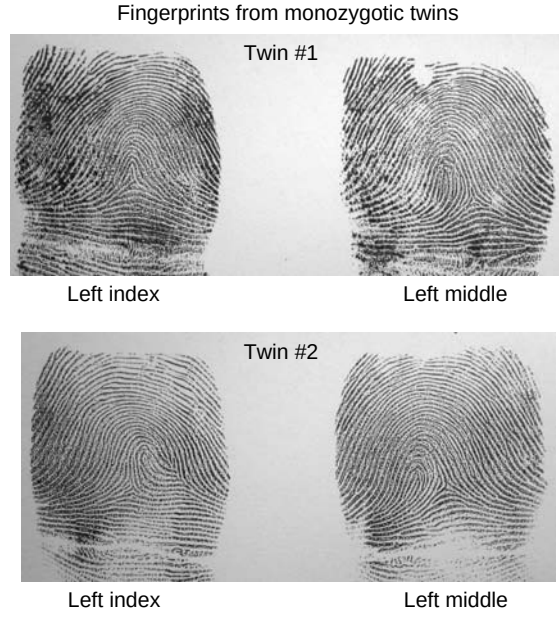


Figure 14 Left index and left middle fingerprints from monozygotic twins

References

[1] Holt, S. (1968). *The Genetics of Dermal Ridges*, Charles C. Thomas, Illinois.

[2] Kimura, S. (1991). Embryological development of flexion creases, *Birth Defects: Original Article Series* **27**(2), 113–129.

[3] Montagna, W. & Paul, P. (1974). *The Structure and Function of Skin*, Academic Press, New York.

[4] Freinkel, R. & David, W. (2001). *The Biology of Skin*, Parthenon Publishing, New York.

[5] Babler, W. (1991). Embryonic development of epidermal ridges and their configurations, *Birth Defects: Original Article Series* **27**(2), 95–112.

[6] Cummins, H. Carnegie Institute of Washington (1929). The topographic history of the volar pads (walking pads; tastaballen) in the human embryo, *Contributions to Embryology* **113**, 103–126.

[7] Mulvihill, J. & David, S. (1969). The genesis of dermatoglyphics, *The Journal of Pediatrics* **75**(4), 579–589.

[8] Cummins, H. & Charles, M. (1943). *Fingerprints, Palms, and Soles: an Introduction to Dermatoglyphics*, The Blakiston Company, Philadelphia.
[9] Lewontin, R. (1995). *Human Diversity*, Scientific American Library, New York.
[10] Raser, J. & O'Shea, E. (2005). Noise in gene expression: origins, consequences, and control, *Science* **309**, 2010–2013.
[11] Hale, A. (1952). Morphogenesis of volar skin in the human fetus, *The American Journal of Anatomy* **91**, 147–173.

ALICE V. MACEO

Frye v. United States

Importance of the Opinion

This case is one of the landmark decisions on the admission of novel forensic opinion evidence in the United States. It was also the first well-reasoned court decision on that issue. Handed down in 1923, *Frye* dealt with the admissibility of testimony based on the purportedly scientific systolic blood pressure deception test developed by Dr Marston. The systolic blood pressure deception test was, thus, the forerunner of the modern polygraph. The *Frye* principle requiring (*see* **General Acceptance Test for Novel Expert Evidence**) came later to be used as the polestar for admissibility of a wide variety of other types of expert opinion evidence.

While it was nominally deprived of its vitality in 1993 when the US Supreme Court held, in (*see* ***Daubert v. Merrell Dow Pharmaceuticals***), that the 1975 passage of (*see* **Federal Rule of Evidence 702**) superseded the *Frye* decision's requirements of general acceptance in federal trials, it remains important for several reasons. First, one of the *Daubert* mandates on admissibility retains, as factor in judging "reliability" of the evidence, whether a test is generally accepted in the profession wherein it belongs. Second, *Frye* remains the law in a significant number of American states that have decided to remain faithful to it and have eschewed an opportunity to follow the federal model of *Daubert*. Despite the importance of the decision, its text is relatively brief and the opinion does not contain citations to other authorities or precedents.

To this day, when the admissibility of opinion testimony is challenged in a pretrial hearing on a motion (*see* ***In Limine*** **Motions and Hearings**) asking the court to exclude the evidence for lack of reliability, forensic scientists may be required to testify in a special hearing to establish the reliability of techniques or methodologies on which they have based their opinion. While being purely a rule of American legal jurisprudence, the *Frye* decision has also been cited on the issue of expert opinion testimony by courts in other countries. All forensic scientists should be cognizant of its existence.

Text of the Opinion

Frye v. United States
Circuit Court of Appeals for the
District of Columbia
54 App. D.C. 46, 293 F. 1013 (Decided
December 13, 1923)

Before Smyth, Chief Justice, Van Orsdel, Associate Justice, and Martin, presiding Judge of the United States Court of Customs Appeals.

Van Orsdel, Associate Justice.

Appellant, defendant below, was convicted of the crime of murder in the second degree, and from the judgment prosecutes this appeal.

A single assignment of error is presented for our consideration. In the course of the trial counsel for defendant offered an expert witness to testify to the result of a deception test made upon defendant. The test is described as the systolic blood pressure deception test. It is asserted that blood pressure is influenced by change in the emotions of the witness, and that the systolic blood pressure rises are brought about by nervous impulses sent to the sympathetic branch of the autonomic nervous system. Scientific experiments, it is claimed, have demonstrated that fear, rage, and pain always produce a rise in systolic blood pressure, and that conscious deception or falsehood, concealment of facts, or guilt of crime, accompanied by fear of detection when the person is under examination, raises the systolic blood pressure in a curve, which corresponds exactly to the struggle going on in the subject's mind, between fear and attempted control of that fear, as the examination

touches the vital points in respect of which he is attempting to deceive the examiner.

In other words, the theory seems to be that truth is spontaneous, and comes without conscious effort, while the utterance of a falsehood requires conscious effort, which is reflected in the blood pressure. The rise thus produced is easily detected and distinguished from the rise produced by mere fear of the examination itself. In the former instance, the pressure rises higher than in the latter, and is more pronounced as the examination proceeds, while in the latter case, if the subject is telling the truth, the pressure registers highest at the beginning of the examination, and gradually diminishes as the examination proceeds.

Prior to the trial the defendant was subjected to a deception test, and counsel offered the scientist who conducted the test as an expert to testify to the results obtained. The offer was objected to by counsel for the government, and the court sustained the objection. Counsel for defendant then offered to have the proffered witness conduct a test in the presence of the jury. This also was denied. Counsel for defendant, in their able presentation of the novel question involved, correctly state in their brief that no cases directly in point have been found. The broad ground, however, upon which they plant their case, is succinctly stated in their brief as follows:

"The rule is that the opinions of experts or skilled witnesses are admissible in evidence in those cases in which the matter of inquiry is such that inexperienced persons are unlikely to prove capable of forming a correct judgment upon it, for the reason that the subject matter so far partakes of a science, art, or trade as to require a previous habit or experience or study in it, in order to acquire a knowledge of it. When the question involved does not lie within the range of common experience or common knowledge, but requires special experience or special knowledge, then the opinions of witnesses skilled in that particular science, art, or trade to which the question relates are admissible in evidence".

Numerous cases are cited in support of this rule. **Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs**.

We think the systolic blood pressure deception test has not yet gained such standing and scientific recognition among physiological and psychological authorities as would justify the courts in admitting expert testimony deduced from the discovery, development, and experiments made thus far.

The judgment is affirmed.

Related Articles

Daubert v. Merrell Dow Pharmaceuticals
Discovery: Depositions
Discovery: Discovery Motions
Discovery of Expert Findings
Discovery in the United States: Civil Cases
Discovery in the United States: Criminal Cases
Expert Opinion: United States
Expert Opinion in Court: a Comparison of Approaches
Expert Opinion in Court: Civil Law Jurisdictions (France, Germany, Italy, and Spain)
Expert Opinion: United Kingdom, Canada, and Australia
General Electric v. Joiner
***In Limine* Motions and Hearings**
Weisgram v. Marley

ANDRE MOENSSENS

FTIR Microscopy *see* Microscopy: FTIR

Future Dangerousness: Determining *see* Dangerousness: Risk of



Gay Panic *see* Homosexual Panic

Gender Differences in Aggression *see* Aggression: Gender Differences in

General Acceptance Test for Novel Expert Evidence

The "general acceptance test" is a legal principle that requires expert opinion evidence of a novel nature to have been generally accepted as a reliable method in the scientific or technical field in which it belongs. When such general acceptance has been shown to exist, expert testimony based thereon then may become admissible as evidence in a court of law.

The principle finds its origin in American jurisprudence in the case of *Frye v. United States* (*see* ***Frye v. United States***) [1], 293 F. 1013 (D.C.Cir. 1923). While the *Frye* case was concerned solely with the admissibility of opinion evidence based on systolic blood pressure (*lie detector*) test, the *Frye* principle of *general acceptance* became, thereafter, the polestar to guide courts in arriving at a decision on whether to admit opinion testimony based on almost any form of other novel scientific evidence in court.

While the principle became general legal authority in the United States for some time, the *Frye* decision has also been referred to in some court decisions in Canada, the United Kingdom, and in other countries as well, including some civil law system court decisions. The universal appeal of the rule of general acceptance as a test for admissibility became eroded in some jurisdictions after the 1993 decision in *Daubert v. Merrell Dow Pharmaceuticals* (*see* ***Daubert v. Merrell Dow Pharmaceuticals***) [2, 3], because *Daubert* demoted it to a mere factor in the making of the admissibility decision rather than as an absolute requirement.

While conceived as a legal standard, the *Frye* principle of general acceptance has found widespread recognition as an appropriate standard among forensic scientists as well.

References

[1] Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).
[2] Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579, 113 S.Ct. 2786 (1993).
[3] Kumho Tire Co. Ltd. v. Carmichael, 526 U.S. 137, 119 S.Ct. 1167 (1993).

Related Articles

Expert Opinion: United States
***In Limine* Motions and Hearings**

ANDRE MOENSSENS

General Electric v. Joiner

Facts and Issues Summarized

An electrician (Joiner), employed by a public utility who suffered from lung cancer brought an action against the manufacturer of polychlorinated biphenyls (PCBs) and the manufacturers (General Electric) of electrical transformers for job-related exposure to PCBs. The trial court determined that plaintiff's expert testimony on the causal connection between PCBs and lung cancer was unconvincing under *Daubert* criteria and excluded the evidence.

The intermediate appellate tribunal – the Court of Appeals for the Eleventh Circuit – reversed the trial judge's ruling. It held that the Court should have applied a more lenient standard of review because the Federal Rules of Evidence exhibit a clear preference in favor of admitting expert testimony.

A further appeal was taken by the defendants to the United States Supreme Court, which had to decide whether the Court of Appeals, in reversing the trial court's *Daubert* decision, applied the correct standard of review. The Supreme Court held that "abuse of discretion" is the proper standard of review of the district court's ruling excluding expert evidence on *Daubert* grounds. The trial court was found not to have abused its discretion in this case. The Court of Appeals was said to have applied overly stringent criteria in reviewing the trial judge's ruling; in so doing it "failed to give the trial court the deference that is the hallmark of abuse of discretion."

Importance of the Opinion for Forensic Science

This case was the first one after *Daubert* (*See* ***Daubert v. Merrell Dow Pharmaceuticals***; **Expert Opinion: United States**; **Federal Rule of Evidence 702**) in which the Supreme Court dealt with issues of admissibility of scientific evidence. *Joiner* reiterated the Court's admonition in *Daubert* that the trial judge is allowed considerable discretion.[a] As long as the judge's decision is supported by the evidentiary record, the appeals court should affirm the decision to admit or exclude evidence. Only if the trial court's opinion is wholly unsupported by the record – an "abuse of discretion," the lower court's ruling ought to be reversed by the reviewing court.

The Supreme Court found that the trial court had followed the *Daubert* guidelines properly while rejecting expert opinions based on studies proffered by the plaintiff's experts. The witnesses, according to the Supreme Court, cited animal studies that "were so dissimilar to the facts presented ... that it was not an abuse of discretion for the District Court to have rejected the experts' reliance on those studies" [1].

In the process of developing the reasons for its holding, the Court inserted some comments that are of particular importance to all forensic scientists. The first one of these is that the Federal Rules of Evidence allow trial courts "to admit a somewhat broader range of scientific testimony than would have been admissible under *Frye*." Thus, *Daubert* is to be read as admitting some scientific evidence that might not be admissible under *Frye*'s more rigid "general acceptance" test.

Despite this more lenient admissibility posture, the Supreme Court also issued the warning that trial courts cannot abandon their "gatekeeping" function. While the focus of the *Daubert* opinion was on the principles and methodologies used by experts in arriving at their opinions, and not on the conclusions that they have generated, the distinction between principles, methodologies, and conclusions is not always easily drawn. The Court stated [2]:

> [C]onclusions and methodology are not entirely distinct from one another. Trained experts commonly extrapolate from existing data. But nothing in either *Daubert* or the Federal Rules of Evidence requires a [trial] court to admit opinion evidence which is connected to existing data only by the *ipse dixit* of the expert. A court may conclude that there is simply too great an analytical gap between the data and the opinion proffered.

The trial court had found that the gap between data and conclusions was too wide, and its finding that the studies offered by plaintiff's experts did not support their conclusions that Joiner's exposure to PCBs contributed to his cancer was not an abuse of discretion.

End Notes

[a.] While *Joiner* was the first article to expound on the meaning of *Daubert*, it is not the only one. Two other Supreme Court cases are noted. (*see* ***Kumho Tire v. Carmichael***; ***Weisgram v. Marley***; **General Acceptance Test for Novel Expert Evidence**).

References

[1] General Electric v. Joiner, 522 U.S. at 145, 118 S.Ct. at 518.
[2] General Electric v. Joiner, 522 U.S. at 146, 118 S.Ct. at 519.

Related Articles

Expert Opinion: United States
Federal Rule of Evidence 702
***Ipse Dixit* Testimony**
Kumho Tire v. Carmichael
Weisgram v. Marley

ANN C. SMITH

Genetic Factors in Behavior ***see*** Genomics and Behavioral Evidence

Genetics: Population ***see*** Hardy-Weinberg Equilibrium

Genocide ***see*** Homicide: Multiple (Behavior)

Genomics and Behavioral Evidence

Genetics and Behavior

It has been thought for hundreds of years that human behavior is driven by some combination of nature (heredity, family history, genetics, genomics), nurture (parental upbringing, influence of peers, good and bad life experiences), and free will. Scientists, philosophers, mental health professionals, and lawyers debate the relative importance of nature, nurture, and free will, but the answer may be, "It depends." In some cases, it depends on the specifics of a particular individual and a particular set of circumstances to know whether the person's behavior was determined by heredity, life experience, or a personal decision – or, more likely, some interaction among these three forces. In general, the relationship between nature and nurture is that environment affects the way genes express themselves; for a detailed discussion of this topic see reference [1].

Brief History of Genetics and Crime

Physicians since the eighteenth century thought that criminal and violent behavior was caused partly by one's family background, partly by unfortunate life experiences, and partly by purposeful malicious intentions. For example, Benjamin Rush, the "Father of American Psychiatry," published *Medical Inquiries and Observations upon the Diseases of the Mind* in 1812. He thought a predisposing cause of mental illness was "a peculiar and hereditary sameness of organization of the nerves, brain, and blood vessels [that] sometimes pervades whole families." However, the more immediate causes of mental illness included traumatic life experiences such as "terror from fire," "earthquakes," and "shipwreck."

In 1881, the trial of Charles Guiteau for the assassination of President James Garfield highlighted the debate among alienists and neurologists regarding the concept of moral insanity. The proponents of moral insanity thought that some forms of antisocial behavior were inherited. However, the jury rejected Guiteau's insanity defense and he was hanged. Sigmund Freud published "Heredity and the Neuroses"

in 1896. Freud said that some mental illnesses were the result of an interaction among hereditary preconditions, "concurrent causes" such as exhaustion and trauma, and "specific causes" such as harmful sexual experiences.

The relationship between a person's physical characteristics and criminal behavior was studied systematically. Cesare Lombroso, sometimes called the "Father of Criminology," published *L'Homme Criminel* in 1895. Lombroso's thesis, influenced by Charles Darwin, was that "primitive" physical characteristics such as a strong jaw, heavy brows, thick lips, and projecting ears predicted criminality. William H. Sheldon published *Varieties of Delinquent Youth* in 1949. Sheldon thought the mesomorphic or muscular body type was predictive of criminal behavior.

The eugenics movements in the United States (about 1910–1935) and in Europe were popular but misguided attempts to apply new knowledge regarding genetics to social engineering. More than 30 states in the United States adopted laws providing for forced sterilization of individuals who were thought to carry a serious hereditary disorder. This practice was approved by the US Supreme Court in the 1927 case of *Buck v. Hall*, in which Justice Oliver Wendell Holmes said that an allegedly retarded woman should be subjected to forcible sterilization because "three generations of imbeciles are enough." In 1934, the Nazi government in Germany used the eugenics movement as the basis for the Law for the Prevention of Genetically Diseased Offspring, which provided for forced sterilization on individuals with diseases that were presumed to be hereditary such as schizophrenia, manic-depressive illness, epilepsy, blindness, and deafness. The horrible consequences of the eugenics movements of the early twentieth century have continued to worry mental health professionals and legal scholars in the twenty-first century. There are concerns that government will use behavioral genomic testing to limit or control individuals because of their genetic makeup.

Although the three-dimensional structure of DNA was discovered by Watson and Crick in 1953, it was not until 1956 that Tjio and Levan announced that the correct diploid count for human chromosomes is 46, including sex chromosomes that are usually either XX (in females) or XY (in males). After the karyotyping of chromosomes became both feasible and accurate, a theory developed that men with an XYY karyotype were predisposed to violence. In the 1970s, there was a good deal of interest in the legal community and also in popular culture about the proposal that there was a genetic explanation for criminal behavior. When more detailed studies were conducted, it was found that XYY men were more likely to be arrested and incarcerated than typical or XY males, but were not more likely to be violent. It was suggested that XYY men were more likely to be arrested because of their learning disabilities or low intelligence [2].

Since the 1920s, there has been much research regarding the relationship between heredity and antisocial and violent behavior. The traditional genetic epidemiological research – based on family, twin, and adoption studies – was summarized by Anderson [3]. Although the studies are not always consistent, there is evidence that both antisocial behavior and violent behavior are heritable. For a more detailed discussion of the philosophical and legal aspects of genetics and criminal behavior, see Reference [4]. For an overview of behavioral genetics and a discussion of its future prospects, see Reference [5].

Traditional genetic research has been helpful in establishing the heritability of psychiatric conditions and certain types of behavior, but it is not very helpful when it comes to testifying about a particular individual. Testimony regarding behavioral genomics will be more meaningful when it is known that a particular gene or a group of interacting genes combined with particular types of life experiences are highly associated with violence or some other specific behavior. Then, the defendant being evaluated can be assessed for the relevant gene(s) and life experiences and testimony will be based on the genetic and environmental interactions for that individual.

Types of Genomic Information

Regarding terminology, the word "genetic" has traditionally referred to a person's genetic makeup or genotype at a specific locus. For example, "Huntington's disease is a genetic disorder." "Genomic" is currently used to indicate how all the DNA material in a cell, both nuclear and mitochondrial, interacts over time to influence biological pathways and physiology. For example, "The more we learn about the human genome, the more complex and dynamic it becomes." Also, "chromosomal" refers to circumstances in which entire chromosomes are missing

or duplicated. For example, "Down's syndrome is a chromosomal disorder caused by the presence of all or part of an extra chromosome 21."

There are three ways in which a specific individual's genetic makeup may be relevant to his or her behavior: (i) the person's genotype may exactly designate a psychiatric or medical diagnosis that clearly explains the person's abnormal behavior. (ii) The person's genotype may support a psychiatric diagnosis that has been made on clinical grounds. (iii) The person's genotype may help to explain a person's violent or criminal behavior.

Genotype Designates the Diagnosis

In some circumstances, the introduction of genetic information at a trial is easy to understand and readily fulfills *Daubert* and *Frye* criteria for reliability and acceptance in the professional community. That is the case when the genetic information is used to show that a specific individual has a specific genetic or chromosomal disorder. For instance, evidence regarding two well known disorders – Down's syndrome (a chromosomal disorder) and Huntington's disease (a genetic disorder) – may arise in a legal context. There are highly specific and reliable methods for genotyping individuals who may have these conditions. Of course, simply having a diagnosis is usually not sufficient all by itself to justify a legal conclusion. In addition to making the diagnosis, it is necessary to show that the defendant's mental functioning or behavior was impaired to a significant degree. The nature of the impairment depends on the legal issue being addressed – which might be competency to stand trial, the insanity defense, diminished capacity, or mitigation.

Most genetic disorders that directly cause cognitive and behavioral symptoms – such as mental retardation and violence – are identified in childhood. For example, Down's syndrome occurs in about one out of 800 live births. Individuals with Down's syndrome may be severely retarded and may commit a criminal offense out of ignorance or naïveté. These individuals may become irritable and frustrated and commit a violent act. Persons with Down's syndrome have a serious mental defect that affects their ability to appreciate the wrongfulness of their behavior, and may lack the mental capacity to participate in legal procedures because of their mental retardation. For example, they may not be competent to stand trial or testify as a witness at a trial.

Some individuals are born with a genetic disorder but the diagnosis is not made until adulthood. It may be that the person's genetic condition is first manifested by some form of aberrant behavior, which leads to his arrest. In this set of circumstances, it is possible for the mental health professional who conducts the pretrial forensic evaluation to be the first person to accurately diagnose the person's medical and psychiatric condition. Huntington's disease is an adult-onset, autosomal, dominant, neurodegenerative genetic disorder. Huntington's disease causes psychosis, dementia, and sometimes violence. The diagnosis of Huntington's disease – based on the defendant's genotype – may have a bearing on the outcome of a criminal case. In this circumstance, the genotype determines the diagnosis and there is a distinct causal relationship between the genotype and the behavior.

There are other genetic or chromosomal disorders that may be manifested by psychological or behavioral symptoms. For example: fragile X syndrome, Prader–Willi syndrome, Wilson's disease, and velocardio-facial syndrome.

G × E Interaction May Support Diagnosis

The inheritance of most psychiatric disorders – including schizophrenia and bipolar disorder – is not mediated by single genes or Mendelian genetics. Investigators of complex human behaviors have proposed complex genomic models such as: interaction among several genes; interaction among genes and epigenetic and other intracellular factors; and interaction among genes and the environment. The interaction between genetic and environmental factors has been called *G × E interaction*. Forensic psychiatrists and psychologists can evaluate defendants in terms of G × E interactions and their findings may be appropriate for testimony in criminal trials. A G × E interaction may support the diagnosis that has been made on clinical grounds and may help explain the defendant's behavior [6].

For example, in some criminal trials, there is a dispute over whether the defendant had a specific mental illness or perhaps any mental illness at the time of the alleged offense. For instance, the defense may assert that the defendant had major depressive disorder, severe, with psychotic features, and that was

why he killed his baby. The prosecution may claim the defendant did not have a significant mental illness at all, but simply pretended to have severe depression and command hallucinations after his arrest in order to avoid responsibility for the baby's death. In this circumstance, the forensic psychiatrist or psychologist may want to evaluate the defendant for a particular G × E interaction because this finding may support the diagnosis of major depressive disorder. A person who is homozygous for the short allele of the *SLC6A4* gene is more likely to become depressed and suicidal after stressful situations than a person who is homozygous for the long allele of that gene. If a person has this genotype and stressful life experiences, it would support the contention that the person had a predisposition to a major depressive disorder and suicidal states at the time of the alleged offense. In this circumstance, the genotype does not "make" the diagnosis of severe depression, but it "supports" the diagnosis that was made on clinical grounds.

Avshalom Caspi and his colleagues were the first to demonstrate psychiatrically relevant G × E interactions [7, 8]. In one of their studies [8], the genetic factor was the polymorphism of the serotonin transporter gene, which is referred to as *5-HTT* or *SLC6A4*. The environmental factor was severe psychosocial stressors. Caspi *et al.* [8] addressed "why stressful experiences lead to depression in some people but not in others." The authors found that a particular allele of the serotonin transporter gene appeared to protect individuals from the harmful biopsychosocial impact of multiple stressors. For this research to be usable in expert testimony in the US legal system, it must be replicated. Thus far, the work of Caspi *et al.* [8] has been replicated by several other research teams, which used varying definitions of psychosocial stressors, psychological distress, and genetic risk. Also, a few published studies did not replicate the findings of Caspi *et al.* [8]. Research regarding the *SLC64A* polymorphism, stressful life events, and depression was summarized by Zammit and Owen [9].

G × E Interaction May Explain Behavior

In some circumstances, the defendant's genotype may help to explain his violent or criminal behavior. For example, a man who has the low activity allele of the monoamine oxidase A (*MAOA*) gene and who experienced serious child maltreatment is more likely to manifest violent and antisocial behavior as an adult than a man who has the high activity allele of this gene. The concept is that these genetic and environmental factors may interact and, later in life, predispose a person to have antisocial, violent behavior. In this kind of case, the genotype does not make a specific diagnosis or support a specific diagnosis, but it does help to explain that a particular person may have a predisposition to maladaptive behaviors.

Avshalom Caspi and his colleagues also studied this type of G × E interaction. In the study of Caspi *et al.* [7], the genetic factor was the *MAOA* gene and the environmental factor was severe childhood abuse. They "studied a large sample of male children from birth to adulthood to determine why some children who are maltreated grow up to develop antisocial behavior, whereas others do not." In selecting the *MAOA* gene, the Caspi team knew that both humans and mice that lacked the gene altogether (*MAOA* "knockouts") became violent and aggressive, respectively. The *MAOA* gene, located on the X chromosome, encodes the MAOA enzyme, which metabolizes neurotransmitters, such as serotonin, norepinephrine, and dopamine. Regarding subsequent research, the work of Caspi *et al.* [7] has been replicated by several other teams that used varying definitions of child maltreatment, violent behavior, and genetic risk. Also, a few published reports failed to replicate the research of Caspi *et al.* [7].

Applications of Forensic Genotyping in Criminal Trials

The relevance of this testimony regarding behavioral genomics to criminal trials depends on the exact nature of the genetic findings and their role in the person's case, that is, whether the genotyping is intended to establish a specific diagnosis, to support a diagnosis made on clinical grounds, or simply to explain a person's criminal acts. Also, the significance of this testimony depends on what legal issue is being addressed, that is, whether it relates to competency to stand trial or engage in some other aspect of the legal procedures, sanity, capacity to achieve the mental state required for particular offenses, or mitigation. Information regarding the defendant's genotype might be relevant in several ways, as summarized in this section of the article.

Insanity Defense

Most jurisdictions in the United States have some version of the insanity defense, although there are various definitions of what constitutes insanity. A typical definition is that "at the time of the commission of the acts constituting the offense, the defendant, as a result of a severe mental disease or defect, was unable to appreciate the nature or wrongfulness of such defendant's acts." In some cases, genetic testing may help to establish that the defendant has one of the criteria for an insanity defense, i.e., "a severe mental disease or defect." For example, Huntington's disease is considered a severe mental disease or mental illness within the psychiatric community. Down's syndrome typically causes a severe mental defect or mental retardation. The diagnosis of a genetic condition does not, however, prove the second criterion for an insanity defense, i.e., the defendant's inability "to appreciate the nature or wrongfulness of such defendant's acts." To determine if the defendant had the capacity to appreciate the wrongfulness of his actions at the time of the alleged offense, a forensic evaluator must collect information regarding the person's functional abilities by relying upon: interview and psychological testing of the defendant; interviews of collateral sources, such as family members, who observed the defendant around the time of the alleged offense; and the investigation by law enforcement personnel regarding the circumstances of the crime and the defendant's behavior and statements when he was arrested.

Diminished Capacity

Most crimes are defined by a particular *mens rea* (guilty mind) and *actus reus* (guilty act). For example, the *actus reus* for first degree murder is the killing of another person. Typically, the *mens rea* for first degree murder is that the killing of another person was done in a premeditated and intentional manner. "Diminished capacity" refers to a defendant's inability to achieve the mental state required for a particular crime. For example, a person who is very intoxicated on cocaine or a person who is extremely depressed might not be able to exercise reflection and judgment and might lack the capacity to commit first degree murder.

Some states allow for the introduction of testimony by mental health professionals regarding diminished capacity, so information regarding the defendant's genotype might be relevant. As in the discussion regarding the insanity defense, the diagnosis of Huntington's disease or Down's syndrome help to explain why a defendant's mental abilities were severely impaired at the time of the alleged offense. The diagnosis by itself would not prove diminished capacity, but it would be part of the testimony to show why the person lacked the capacity to form a particular mental state.

Also, suppose that a defendant who is accused of first degree murder had a history of severe child abuse and a genotype that included the low activity allele of the *MAOA* gene. Because of his genetic makeup and life experiences, the defendant is at a high risk to commit violent acts. Suppose the circumstances of the alleged offense were such that the defendant might have been in a state of excitement or passion. Suppose also that the defendant was using cocaine and marijuana at the time of the alleged offense. One can see how all of this information – including the *MAOA* genotyping – could be considered by the trier of fact to decide that the defendant did not have the mental capacity to commit first degree murder.

Mitigation

In the guilt phase of the trial, the jury or the judge decides whether the defendant committed the crime. If the defendant is found guilty, additional evidence is presented in the penalty phase of the trial and the jury or the judge decides what the defendant's sentence will be. During the penalty phase, the defense attorney can present a very broad range of information to influence the jury and judge to be lenient toward the defendant. If the defendant was found guilty of a capital crime such as first degree murder, the jury may forego the death penalty and sentence the person to life in prison.

In the penalty phase of the trial, the defense often presents testimony about the defendant's difficult childhood, such as experiencing chronic illness, child maltreatment, or poverty. It may be relevant to present testimony that a person's genetic makeup – along with other factors like physical or sexual child abuse – predisposed a person to commit a violent, antisocial act. The defense attorney would argue that the defendant did not ask to have a particular genetic makeup. He did not ask to be physically or sexually abused as a child. But these factors – without

his desire, knowledge, or awareness – made it more likely he would commit a violent act later in life.

Dangerousness

The prosecution also has an opportunity at the penalty phase of a trial to present additional evidence. In some circumstances, the prosecution may argue that behavioral genomic testimony is not mitigating, but is an aggravating factor that should increase the length of a sentence. For instance, the prosecution may say that because of his genetic makeup and his childhood experiences, the defendant has violent tendencies and should have a longer sentence in order to protect society for a longer time.

Whether testimony regarding behavioral genomics favors the defense or the prosecution depends on the circumstances. In a case of aggravated robbery, for instance, it may be logical for the prosecution to argue that a person's genotype and bad life experiences mean he should be imprisoned longer in order to protect society. In a case of capital murder, on the other hand, the defendant is never going to live outside prison and threaten society. In that situation, it may be logical for the defense to argue that the person's genotype and bad life experiences mean he should have a life sentence rather than the death penalty.

Malingering

In some trials in which psychiatric or psychological testimony is offered, there is a dispute about whether the defendant is malingering mental illness. If the prosecution can show that the defendant was malingering, the jury is less likely to find the person insane or eligible for a reduced sentence. The defense may be able to use behavioral genomic testimony to bolster the argument that the defendant has an actual mental disorder.

In the future, it is likely that researchers in behavioral genomics will show that groups of genes or various G × E interactions greatly increase the risk that a particular person will have a serious psychiatric condition, i.e., schizophrenia, bipolar disorder, obsessive-compulsive disorder, autism, etc. It is likely that this kind of information will be one aspect of what the forensic psychiatrist or psychologist considers in assessing whether a defendant is malingering. It would raise suspicions of malingering if a person had no genetic or environmental risks for schizophrenia, but sometimes manifested symptoms that seemed consistent with schizophrenia. Of course, genetic testing would only be one part of a comprehensive evaluation, and not determinative all by itself. It is possible for a person to have the genetic risk factors for schizophrenia, but not have the illness. Also, it is possible for a person with actual schizophrenia to also be malingering.

Competency

Mental competency is a very important issue in many aspects of the criminal justice system. Mental health professionals are frequently asked to assess whether a defendant was competent in the past when he waived Miranda rights and made a statement to investigators. They are asked to evaluate whether a defendant is currently competent to stand trial or competent to waive his right to have an attorney represent him at the trial. After conviction, there may be an issue of whether an inmate is competent to waive his appeals or even competent to be executed.

In conducting a competency evaluation, the mental health professional would ordinarily assume the person is competent unless there is some reason to think otherwise. To conclude that an individual is not competent, the evaluator would have to identify some mental disorder that is causing the impairment in mental functioning. As in the previous discussion – regarding insanity, diminished capacity, and malingering – genetic testing may help establish that a person has a mental illness that would cause enough impairment to compromise the individual's competency. Of course, making the diagnosis of mental illness or mental retardation does not in itself mean the person is not competent, and that conclusion would require additional assessment of the person's actual functioning.

Juvenile Court

Since most juvenile courts are organized around the principle of rehabilitation, juvenile court judges are usually very interested in having a full understanding of the youngster's biological, psychological, and social background. Juvenile court judges – in order to arrive at an appropriate disposition of each case – take into consideration lengthy social histories, the

results of psychological testing, and psychiatric evaluations. There is a search for why this particular boy or girl got into so much trouble in the past, and then an effort to devise an intervention or corrective action to reduce the likelihood of more trouble in the future.

Preliminary research regarding G × E interactions has started to explain why some children manifest oppositional defiant disorder and conduct disorder, while others do not. Of course, juvenile court judges are not interested merely in research and theories, but they want solutions, interventions, and rehabilitation that actually works. As both genetic and environmental risk factors and their interactions become better understood, practitioners will develop more specific and more effective treatments. For example, it will be determined that some forms of adolescent violence are treatable with medication, while other forms of violence require lengthy residential treatment. In the future, an assessment of a youngster's risk factors based on G × E interactions will be a common feature of biopsychosocial evaluations conducted for juvenile courts.

Causation and Correlation

This article discusses a wide range of testimony regarding behavioral genomics. In some of the examples cited (such as Down's syndrome and Huntington's disease), there is an obvious relationship between the genetic abnormality and the behavioral outcome that is universally accepted by the scientific community. Both scientists and legal scholars would agree that having three copies of chromosome 21 is a sufficient cause for the clinical phenotype known as *Down's syndrome*.

The relationship between the G × E interactions (*SLC6A4* and psychosocial stressors; *MAOA* and childhood maltreatment) and behavior discussed in this article is not so obvious. There certainly is a correlation between the G × E interactions and the behavior, but one cannot say definitely there is a causal relationship. On the basis of the replicated research, the interaction of the *SLC6A4* gene and psychosocial stressors is a probabilistic cause of depression and suicidality, which means the G × E interaction increases the chance that depression and suicidality will occur. Likewise, the interaction of the *MAOA* gene and childhood maltreatment is a probabilistic cause of violence.

Implications for the Future

The goal of future research is to establish linkages along the following lines: individuals with a combination of allelic polymorphisms and certain life experiences have specific changes in brain functioning, manifested by functional magnetic resonance imaging (fMRI); there is a physiological relationship between these changes, such as modification in the activity of a particular neurotransmitter; these changes in the brain cause specific alterations in emotional and/or cognitive functioning; the functional alterations cause a consistent pattern of behavior; and the pattern of emotional, cognitive, and behavioral changes may constitute a diagnosis.

Genotyping is not a test that can be interpreted or presented all by itself. If genotyping is conducted on a criminal defendant, it should be part of a comprehensive psychiatric or psychological forensic evaluation. A pretrial forensic evaluation typically consists of many parts including a review of medical records, a review of the investigation of the crime, interviewing the defendant, psychological testing, neuropsychological testing, interviewing family members and other collaterals, and sometimes other investigations such as brain scans, electroencephalograms, and consultation with other medical specialists. As a small part of this elaborate evaluation, genotyping for *MAOA*, *SLC6A4*, and other genes might included.

Studying G × E interactions with behavioral and psychiatric implications is a popular topic and this type of research will flourish in coming years. It is likely that many G × E interactions will be highly correlated with major depressive disorder, bipolar disorder, schizophrenia, posttraumatic stress disorder, obsessive-compulsive disorder, panic disorder, substance abuse, and violent sexual predation. These interactions will usually not prove that a person has a particular diagnosis, but they will support and explain the diagnosis that has been made through traditional evaluation methods. Also, if the defendant is suspected of malingering, an investigation of the person's genetic makeup and relevant life experience may be used to support the claim that he or she actually has a serious mental disorder.

References

[1] Ridley, M. (2003). *Nature Via Nurture: Genes, Experience, and What Makes Us Human*, Harper Collins, New York.

[2] Gotz, M.J., Johnstone, E.C. & Ratcliffe, S.G. (1999). Criminality and antisocial behavior in unselected men with sex chromosome abnormalities, *Psychological Medicine* **29**, 953–962.
[3] Anderson, G.S. (2007). *Biological Influences on Criminal Behavior*. CRC Press, Boca Raton, FL.
[4] Wasserman, D. & Wachbroit, R. (eds) (2001). *Genetics and Criminal Behavior*, Cambridge University Press, Cambridge, UK.
[5] Parens, E., Chapman, A.R. & Press, N. (eds) (2005). *Wrestling with Behavioral Genetics: Science, Ethics, and Public Conversation*, Johns Hopkins University Press, Baltimore.
[6] Bernet, W., Vnencak-Jones, C., Farahany, N. & Montgomery, S. (2007). Bad nature, bad nurture, and testimony regarding *MAOA* and *SLC6A4* genotyping at criminal trials, *Journal of Forensic Sciences* **52**, 1362–1371.
[7] Caspi, A., McClay, J., Moffitt, T.E., Mill, J., Martin, J., Craig, I.W., Taylor, A., Poulton, R. (2002). Role of genotype in the cycle of violence in maltreated children, *Science* **297**, 851–854.
[8] Caspi, A., Sugden, K., Moffitt, T.E., Taylor, A., Craig, I.W., Harrington, H., McClay, J., Mill, J., Martin, J., Braithwaite, A., Poulton, R. (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene, *Science* **301**, 386–389.
[9] Zammit, S. & Owen, M.J. (2006). Stressful life events, 5-HTT genotype and risk of depression, *The British Journal of Psychiatry* **188**, 199–201.

Related Articles

DNA

Insanity: Defense

Mitigation Testimony

WILLIAM BERNET

Genomics and Mental Illness *see*

Psychopathology: Terms and Trends

Genotyping: Virus *see*

Geographical Identification by Viral Genotyping

Geographical Identification by Viral Genotyping

Introduction

Major genetic differences between human populations must have evolved when they expanded, originating from Africa, across the Earth in the course of the last 100 000 years [1, 2]. The theory of isolation by distance, i.e., the decrease of genetic similarity with increasing geographic distance has been supported by data on genetic polymorphism [3, 4]. In the course of this evolution, humans have carried many parasites that have coevolved with humans resulting in different types of relationship, i.e., commensal, mutualistic, or pathogen to humans [5]. Therefore, also evolution and genomic diversity of viruses has been used to gain better insight into patterns of ancient human migration [6]. Studies with human papillomavirus (HPV-16), polyomavirus JC (JCV), and also the bacterium *Helicobacter pylori* have shown a similar pattern of evolution to humans with origins in Africa and subsequent migration of ethnic groups to the other continents in prehistoric times [7–10].

This association of parasitic genomic diversity with geographic area has also attracted forensic scientists. Recently, Ikegaya *et al.* reported a novel method applying genetic diversity of the JC virus to forensic case work [11]. After that report, several other parasites have been examined to find out whether their genotype and geographical region correlate to such an extent that would allow them to be used as aids to the process of human identification.

Relations between Parasitic Genotype and Geographic Area

JCV, a member of the *Polyomaviridae* family, was first isolated in 1971 from the brain of a patient with progressive multifocal leukoenchephalopathy (PML) and was eventually found to be ubiquitous in the human population [12, 13]. After primary asymptomatic infection in childhood, JCV persists in the renal tissue of most adults, excreting its

Table 1 Distribution of JCV genotypes in the old world

Genotype	Geographic region where indicated genotype is mainly detected[a]
EU-a	Europe
EU-b	Southern Europe
EU-c	Northeast Siberia, Arctic circle
B1-c	Europe, Mediterranean
Af1	Central and western Africa
Af2-a	Southern Africa
Af2-b	Northern Africa, western Asia
Af3	Central Africa
B1-a	China
B1-b	Central and western Asia
B1-d	Saudi Arabia, Greece
B2	India, Mauritius
CY-a	Korea
CY-b	North-eastern China, southern Japan
SC	Southeast Asia, southern China
MY	Northern Japan, Korea
8A	New Guinea, Oceania
2E	Oceania

[a] Areas of geographic distribution indicated according to reports by Yogo *et al.* and Takasaka *et al.*

progeny virus in urine. JCV strains worldwide can be classified into more than 30 genotypes from 18 main genotypes by phylogenetic analyses [14] of viral 5100p DNA sequences [15]. Each genotype is distributed in relatively small distinct areas of different parts in the world (Table 1).

BK virus (BKV) belongs to the same family of *polyomaviridae* as JCV, and was first detected from a renal transplant patient in 1971 [16]. Primary infection with this virus usually occurs in childhood and the seroprevalence rate reaches the adult rate (65–90%) at the age of 5–10 years [17]. BKV is also known to persist in the kidneys and to excrete progeny virus in urine [18]. BKV strains worldwide can be classified into seven main genotypes according to phylogenetic analyses of viral DNA sequences [19, 20]. Each genotype occupies distinct areas of distribution in different parts of the world (Table 2).

So far, a correlation between the geographic area and genotypes has also been noted in other parasites, such as Epstein Barr virus (EBV) [20], human herpes virus type 1 (HHV-1), *Helicobacter pylori* [21], and *Candida albicans* [22]. Figures 1 and 2 show the phylogenetic tree of the JC virus and HHV-1 virus, respectively. These phylogenetic trees, constructed from worldwide isolates, show close correlations to

Table 2 Distribution of BKV genotypes in the Old World

Subtype	Subgroup	Geographic region where an indicated genotype is mainly detected[a]
I	Ia	Africa
	Ib-1	Southeast Asia
	Ib-2	Europe
	Ic	Northeast Asia
IV		Continental Asia
II, III		Rarely detected

[a] Areas of geographic distribution indicated according to Zheng *et al.*

the geographic areas; however, the relation between the area and genotypes is much closer in JC virus than HHV-1 virus. The reason for this is that the evolutional history of JCV is almost synchronized to the history of humans, whereas HHV-1 is not, as a host change may have occurred; therefore, not only geographic correlations but also evolutional history of the parasite in question has to be analyzed.

Detection of Parasitic Genotypes from Forensic Samples

Polymerase Chain Reaction (PCR) amplification of a 610 bp region (VT intergenic region: IG region) of the viral genome was used for JCV detection from forensic autopsy cases. The detection rate was 45% when 200 mg of renal tissue was used as template and somewhat lower, 33%, from 5 ml urine. The detection rate was higher (>50%) with elderly subjects. Multiple samples taken from the same kidney and the contralateral one from the same cadaver showed identical sequences. The JCV detection rate was not related to the cause of death and was hardly affected by postmortem decomposition [23]. Detection of JCV from formalin-fixed and paraffin-embedded tissue has also been reported [24]. In formalin-fixed, paraffin-embedded samples, not only the original sequences but also those with 1% of base substitution were detected and no genotype change was found. In formalin-soaked samples, the original sequences and those with more than 1% of substitution causing genotype change were detected; thereafter, genotype was determined only using specimens in a frozen state or that had been formalin-fixed for a short time. However, JCV was detectable from 50 μl of urine [25]. When 200 μl of urine was used for

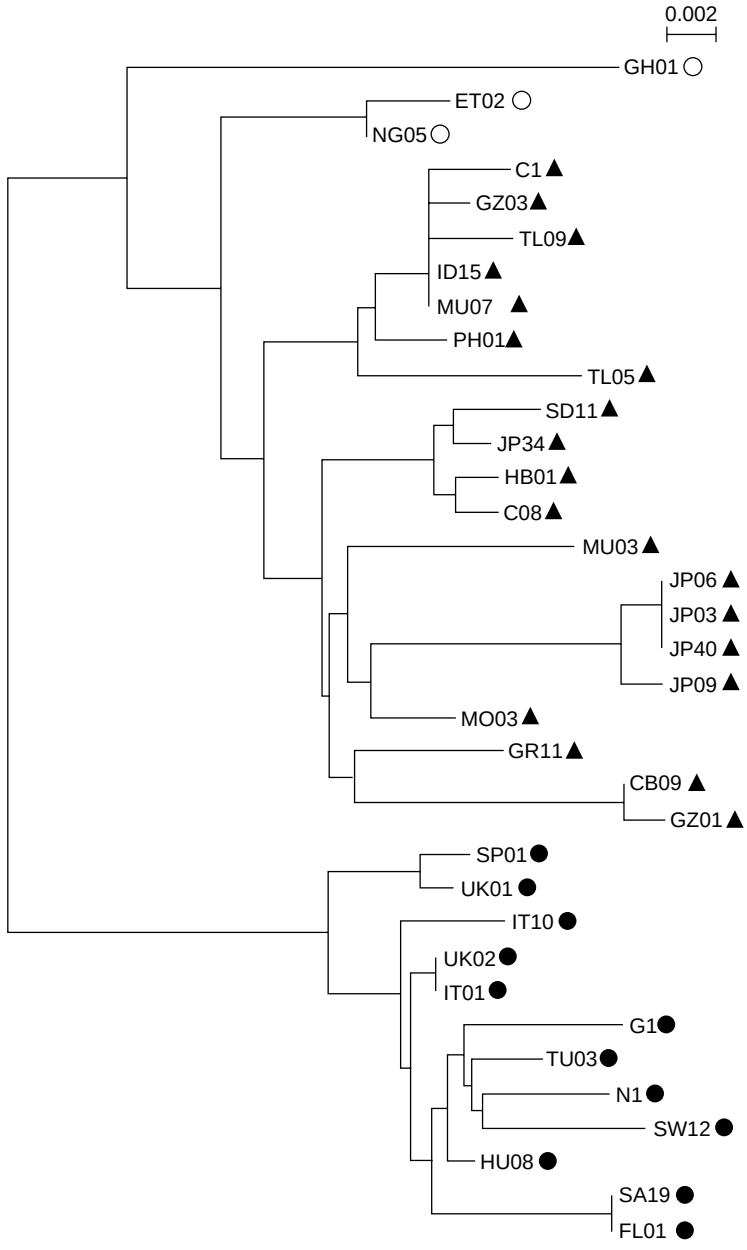


Figure 1 Phylogenetic analysis of JCV genome. 610 bp hypervariable region (IG region) of JCV genome was phylogenetically analyzed. Open circle, African isolates; closed circle, European isolates; and closed triangle: Asian isolates

detection, 54.9% was JCV positive, and JCV DNA was highly detectable from minute urine stains prepared 3 months earlier using 100 μl of JCV-positive urine. These findings show that this method could be used for urine stain samples left at the scene and small amounts of urine taken from autopsies. Although there are several chemical or immunological methods to identify urine stains, it can be problematic to confirm that they are of human origin. However, detection of JCV or BKV from urine



Figure 2 Phylogenetic analysis of HHV-1 genome. 666 bp hypervariable region of HHV-1 genome was phylogenetically analyzed. Open circle, African isolates; closed circle, European isolates; and closed triangle: Asian isolates

stains automatically confirms that they are human, because these viruses only infect the human urinary tract.

BKV DNA was also detectable in 30.5% of postmortem cases [26]. In addition, JCV DNA was detected in 47.2% of the same postmortem cases. Overall, BKV or JCV was detectable in 69.8% of postmortem cases.

As EBV infects human leukocytes, it might, at least in principle, be a good tool to identify

using blood samples, however, the detection rate from postmortem blood samples was only 16% [20]. Also HSV-1 DNA has been detected postmortem in forensic cases from the trigeminal nerve by PCR method with a detection rate of 61% [27], as well as Helicobacter Pylori from gastric mucosa, with a detection rate of 56% [28].

Estimation of Geographic Origin

After JCV infection, the same viral strains are maintained in the kidneys and urine throughout life, irrespective of the host's relocation; therefore, the estimated geographical origin from the JCV genotype is closely related to the area where the host grew up. Among all the parasites described above, the JCV genotype has been classified in most detail. The applicability of these methods depends in particular on the distribution of the genotype, subtypes, and subgroups within the geographic area in question. Especially in Asia, many genotypes and their subtypes are found in small domains; for instance, mainly two genotypes are found in Japan, with one genotype prevailing in the north and the other in the south. We can therefore estimate from which area a cadaver originated from and the probability of a cadaver originating from the place where it was found. Japanese police started to use JCV method in 2005 and BKV method in 2007. So far, these methods have been applied to more than 10 cases and in some of them with good results. For other parasites, it is not clear whether double infection or genotypic change might occur during life, therefore, further studies are necessary.

Potential New Approaches toward Geographical Identification Using Human Parasites

In addition to the genomic diversity of human parasites, another approach that may have potential to be used in similar way in the future, but has to our knowledge not been investigated, is the impact of antibiotics. The widespread use of antimicrobial agents worldwide has led to an increasing prevalence of drug-resistant strains of bacteria that can show distinct patterns both geographically and over time. In 1992 established Alexander Project is a continuing surveillance study examining the antimicrobial susceptibility of bacterial pathogens of community-acquired respiratory tract infections with participating centers in many countries in Europe, Africa, Latin America, Middle East, Far East, and the United States [29]. As this type of bacterial resistance is under continued evolution, it would presuppose the existence of a global online database for comparison and additionally a sufficient stability of the bacterial strains to be detected in postmortem material.

References

[1] Hinkes, M.J. (2008). Migrant deaths along the California-Mexico border: an anthropological perspective, *Journal of Forensic Sciences* **53**(1), 16.

[2] Cattaneo, C., Ritz-Timme, S., Schutz, H.W., Collins, M., Waite, E., Boormann, H., Grandi, M. & Kaatsch, H.J. (2000). Unidentified cadavers and human remains in the EU: an unknown issue, *International Journal of Legal Medicine* **113**(3), N2.

[3] Cavalli-Sforza, L.L., Menozzi, P. & Piazza, A. (1993). Demic expansions and human evolution, *Science* **259**(5095), 639.

[4] Ramachandran, S., Deshpande, O., Roseman, C.C., Rosenberg, N.A., Feldman, M.W. & Cavalli-Sforza, L.L. (2005). Support from the relationship of genetic and geographic in human populations for a serial founder effect originating in Africa, *Proceedings of the National Academy of Sciences of the United States of America* **102**(44), 15942.

[5] Dethlefsen, L., McFall-Ngai, M. & Relman, D.A. (2007). An ecological and evolutionary perspective on human-microbe mutualism and disease, *Nature* **449**(7164), 811–818.

[6] Cavalli-Sforza, L.L. & Feldman, M.W. (2003). The application of molecular genetic approaches to the study of human evolution, *Nature Genetics* **33**(Suppl), 266–275.

[7] Ho, L., Chan, S.Y., Burk, R.D., Kiviat, N., Lancaster, W., Mavromara-Nizos, P., Labropoulou, V., Mitrani-Rosenbaum, S., Norrild, B., Pillai, R.M., Stoerker, J., Syrjaenen, K., Wheeler, C.M., Williamson, A.L. & Bernard, H.U. (1993). The genetic drift of human papillomavirus type 16 is a means of reconstructing prehistoric viral spread and the movement of ancient human populations, *Journal of Virology* **67**(11), 6413.

[8] Sugimoto, C., Kitamura, T., Guo, J., Al-Ahdal, M.N., Shchelkunov, S.N., Otova, B., Ondrejka, P., Chollet, J.Y., El-Safi, S., Ettayebi, M., Grésenguet, G., Kocagöz, T., Chaiyarasamee, S., Thant, K.Z., Thein, S., Moe, K., Kobayashi, N., Taguchi, F. & Yogo, Y. (1997). Typing of urinary JC virus DNA offers a novel means of tracing

human migrations, *Proceedings of the National Academy of Sciences of the United States of America* **94**(17), 9191.

[9] Covacci, A., Telford, J.L., Del Giudice, G., Parsonnet, J. & Rappuoli, R. (1999). Helicobacter pylori virulence and genetic geography, *Science* **284**(5418), 1328.

[10] Ikegaya, H., Zheng, H.Y., Saukko, P.J., Varesmaa-Korhonen, L., Hovi, T., Vesikari, T., Suganami, H., Takasaka, T., Sugimoto, C., Ohasi, Y., Kitamura, T. & Yogo, Y. (2005). Genetic diversity of JC virus in the saami and the finns: implications for their population history, *American Journal of Physical Anthropology* **128**(1), 185.

[11] Ikegaya, H., Iwase, H., Sugimoto, C. & Yogo, Y. (2002). JC virus genotyping offers a new means of tracing the origins of unidentified cadavers, *International Journal of Legal Medicine* **116**(4), 242–245.

[12] Padgett, B.L., Walker, D.L., ZuRhein, G.M., Eckroade, R.J. & Dessel, B.H. (1971). Cultivation of papova-like virus from human brain with progressive multifocal leucoencephalopathy, *Lancet* **1**(7712), 1257–1260.

[13] Kunitake, T., Kitamura, T., Guo, J., Taguchi, F., Kawabe, K. & Yogo, Y. (1995). Parent-to-child transmission is relatively common in the spread of the human polyomavirus JC virus, *Journal of Clinical Microbiology* **33**(6), 1448–1451.

[14] Yogo, Y., Sugimoto, C., Zheng, H.Y., Ikegaya, H., Takasaka, T. & Kitamura, T. (2004). JC virus genotyping offers a new paradigm in the study of human populations, *Reviews in Medical Virology* **14**(3), 179–191.

[15] Cole, C.N. & Conzen, S.D. (2001). Polyomaviridae: the viruses and their replication, in *Field's Virology*, D.M. Knipe & P.M. Howley, eds, Lippincott Williams & Wilkins, Philadelphia, Vol. **2**, p. 3280.

[16] Gardner, S.D., Field, A.M., Coleman, D.V. & Hulme, B. (1971). New human papovavirus (B.K.) isolated from urine after renal transplantation, *Lancet* **1**(7712), 1253–1257.

[17] Knowles, W.A. (2001). The epidemiology of BK virus and the occurrence of antigenic and genomic subtypes, in *Human Polyomaviruses: Molecular and Clinical Perspectives*, K. Khalili & G.L. Stoner, eds, Wiley, p. 704.

[18] Zhong, S., Zheng, H.Y., Suzuki, M., Chen, Q., Ikegaya, H., Aoki, N., Usuku, S., Kobayashi, N., Nukuzuma, S., Yasuda, Y., Kuniyoshi, N., Yogo, Y. & Kitamura, T. (2007). Age-related urinary excretion of BK polyomavirus by nonimmunocompromised individuals, *Journal of Clinical Microbiology* **45**(1), 193–198.

[19] Zheng, H.Y., Nishimoto, Y., Chen, Q., Hasegawa, M., Zhong, S., Ikegaya, H., Ohno, N., Sugimoto, C., Takasaka, T., Kitamura, T. & Yogo, Y. (2007). Relationships between BK virus lineages and human populations, *Microbes and Infection* **9**(2), 204–213.

[20] Ikegaya, H., Motani, H., Sakurada, K., Sato, K., Akutsu, T. & Yoshi, M. (2008). Forensic application of Epstein-Barr virus genotype: correlation between viral genotype and geographical area, *Journal of Virological Methods* **147**(1), 78–85.

[21] Kersulyte, D., Mukhopadhyay, A.K., Velapatiño, B., Su, W., Pan, Z., Garcia, C., Hernandez, V., Valdez, Y., Mistry, R.S., Gilman, R.H., Yuan, Y., Gao, H., Alarcón, T., López-Brea, M., Balakrish Nair, G., Chowdhury, A., Datta, S., Shirai, M., Nakazawa, T., Ally, R., Segal, I., Wong, B.C., Lam, S.K., Olfat, F.O., Borén, T., Engstrand, L., Torres, O., Schneider, R., Thomas, J.E., Czinn, S. & Berg, D.E. (2000). Differences in genotypes of Helicobacter pylori from different human populations, *Journal of Bacteriology* **182**(11), 3210–3218.

[22] Xu, J. & Mitchell, T.G. (2003). Geographical differences in human oral yeast flora, *Clinical Infectious Diseases* **36**(2), 221–224.

[23] Ikegaya, H., Iwase, H. & Yogo, Y. (2004). Detection of identical JC virus DNA sequences in both human kidneys, *Archives of Virology* **149**(6), 1215.

[24] Ikegaya, H., Iwase, H., Zheng, H.Y., Nakajima, M., Sakurada, K., Takatori, T., Fukayama, M., Kitamura, T. & Yogo, Y. (2005). JC virus genotyping using formalin-fixed, paraffin-embedded renal tissues, *Journal of Virological Methods* **126**(1–2), 37–43.

[25] Sakurada, K., Ikegaya, H., Motani, H., Iwase, H., Sekiguchi, K., Akutsu, T., Yoshino, M., Takatori, T. & Sakai, I. (2005). JC virus genotyping can be used to narrow down the native place of persons from urine stains, *Japanese Journal of Forensic Science and Technology* **10**(2), 111–117.

[26] Ikegaya, H., Motani, H., Saukko, P., Sato, K., Akutsu, T. & Sakurada, K. (2007). BK virus genotype distribution offers information of tracing the geographical origins of unidentified cadaver, *Forensic Science International* **173**(1), 41–46.

[27] Motani, H., Sakurada, K., Ikegaya, H., Akutsu, T., Hayakawa, M., Sato, Y., Yajima, D., Sato, K., Kobayashi, K. & Iwase, H. (2006). Detection of herpes simplex virus type 1 DNA in bilateral human trigeminal ganglia and optic nerves by polymerase chain reaction, *Journal of Medical Virology* **78**(12), 1584.

[28] Nagasawa, A., Azuma, K., Motani, H., Hayakawa, M., Yajima, D., Kobayashi, K. & Iwase, H. (2007). Detection of helicobacter pylori from postmortem gastric mucosa. – application to the geographic identification of unidentified cadavers, *76th Kanto meeting of Japanese Society of Legal Medicine*, Yokohama.

[29] Jacobs, M.R., Felmingham, D., Appelbaum, P.C. & Grüneberg, R.N. (2003). The Alexander project 1998–2000: susceptibility of pathogens isolated from community-acquired respiratory tract infection to commonly used antimicrobial agents, *The Journal of Antimicrobial Chemotherapy* **52**(2), 229.

Further Reading

Jernberg, C., Löfmark, S., Edlund, C., Jansson, J.K. (2007). Long-term ecological impacts of antibiotic administration on the human intestinal microbiota, *ISME Journal* **1**(1), 56–66.

Related Articles

Anthropology

Disaster Victim Identification

HIROSHI IKEGAYA, PEKKA J. SAUKKO, YOSHINAO KATSUMATA AND TAKEHIKO TAKATORI

Glass

Introduction

Glass is ubiquitous and has been used for many centuries in a range of objects such as jewelry, drinking vessels, vases, and windows. The success of glass and its use can be attributed to its robust, inert, and decorative qualities.

Glass Technology

Glass is a noncrystalline or amorphous mixture of compounds, which when combined in the molten phase and cooled, exhibits a solidified version of its liquid structure. Hence, it can be regarded as being a supercooled liquid [1]. The main constituent of glass is silicon dioxide or silica (SiO_2). To reduce the high melting point of the silica [2] and to improve its properties, such as transparency when solid, several oxides can be added to it and heated. Typically, sodium oxide or soda (Na_2O) and calcium oxide or lime (CaO) are added to form the basis of the "soda–lime–silica" glasses used in most modern windows and containers [3]. Many other oxides are also added to alter the properties of the glass, such as iron oxide (Fe_2O_3), which is added to the mixture to produce a brown or green tint [3].

Nowadays, the molten mixture is molded, shaped, or blown to form container glassware, or drawn through rollers to produce patterned window and wired glasses [3, 4]. However, most modern windows are manufactured by the float process, whereby the molten mixture is poured onto molten tin, left to cool in a controlled atmosphere, to produce flat distortion-free panes [3, 4]. Methods can be used to strengthen glass. For example, glasses can be toughened or tempered by reheating and cooling rapidly, which creates a strong type of safety glass that forms cubes on breaking rather than shards [3, 4]. This type is used in the side windows of vehicles and glass doors. Laminated glass is another type of safety glass, which comprises two or more layers of plain flat glass with a thin sheet of plastic in between each layer [5]. This is used in vehicle windscreens and provides the basis of bulletproof screens used in banks and building societies.

Window glasses have been manufactured by various methods over the years. For example, the molten glass mixture was passed through rollers, cooled, and polished to create plate glass or a "gob" of molten glass was spun on the end of a rod to form a large flat disc as in the case of Crown glass [4]. Evidence of a type of manufacture can often be seen by the forensic scientist when examining the physical and chemical properties of a glass sample in the laboratory.

Breaking Glass

While glass is an inert and robust material, if enough force is applied the glass may crack and completely shatter, creating many hundreds and thousands of fragments, some of which may land on the clothing, footwear, and hair of the individual breaking it or to an individual standing close by. This aspect is important to forensic scientists. In the late 1960s, Nelson and Revell demonstrated the backscatter or backward fragmentation of glass fragments on to a person breaking a window [6]. They found that a large number of fragments were propelled in the direction of the force, away from the breaker, however, a large number of fragments were scattered back toward the origin of the force. Subsequent studies have confirmed this process [7–10]. In addition, researchers have attempted to estimate the number of fragments transferred to an individual breaking a window under varying conditions [7–11]. In summary, the number of fragments transferred by backward fragmentation is very variable, and can range from none to many tens, as this can depend upon how the window is broken and how far an individual is from the breaking

glass. Particles of glass can also be transferred by other means. For example, if the person leans in or walks through, a hole created in the broken glass [12], or by secondary transfer, via an item, that is, exposed to the breaking glass and is subsequently handled [13, 14]. While many studies tend to consider the transfer of glass from a breaking or broken window, other glass objects, such as bottles, vehicle windows and lenses, and drinking vessels, can also be broken at the scene. As a consequence, fragments from these broken objects can also be transferred to the victim or perpetrator, although the number of fragments involved, tend to be lower than that for a breaking window.

Research has shown that once fragments of glass are transferred, to the surfaces of clothing, these can be lost over a matter of hours with normal wear [9, 15, 16]. Typically, the larger fragments are lost first with the smaller, microscopic particles being lost later. The type of garment can also affect how long glass fragments may be retained on the surfaces [9, 15, 16]. For example, particles tend to remain on the fibrous surface of a fleece jacket longer than that of a smooth cotton tee-shirt.

In the laboratory, the forensic scientist may recover any particulate material by initially examining the exhibit by eye and picking off larger pieces using forceps. However, in the majority of cases, because most glass particles tend to be invisible to the naked eye, items are shaken vigorously over brown paper, or a large inverted cone, so as to dislodge any fragments. The resulting debris is collected into a small pot, which is examined under a low-power stereo microscope for any particles, the number and appearance of which are recorded.

Analysis and Comparison

A comparison is usually made of a sample of fragments of glass recovered from the suspect's (alleged perpetrator) or victim's clothing, footwear, or hair, with a control sample of the broken glass item from the crime scene. Several methods of comparison are used by forensic scientists. Two of the most commonly employed techniques have been considered here. These and others are usually performed in a sequence or stepwise process. Thus, if recovered fragments are found to be different from the broken window or object (i.e., the control glass sample), at any stage, then the examination will be stopped and an interpretation relating to any differences made accordingly.

The measurement of refractive index (RI), an optical property of glass (and other materials), is used extensively. The RI is a measure of the bending of light as it moves from one medium (e.g., air) into another (e.g., glass). The theory of RI and the technology relating to the measurement of fragments of glass can be found in several publications [17, 18]. In essence, a fragment (usually <1 mm) is immersed in silicone oil on a glass slide, which is then placed on to a hot stage, sited under a microscope. The RI of the oil changes with temperature, whereas the RI of the glass remains stable. The hot stage is controlled by a computer to heat and cool the fragment in the oil. During the heating and cooling cycles, there will be a temperature at which the RI of the oil matches that of the glass and the fragment disappears momentarily. The computer records these temperatures, which are averaged and converted to RI. The RI is determined for a sample of the control glass and a selection of fragments recovered from the suspect or victim. The refractive indices of any recovered fragments should be within certain parameters in order to ascertain if these are indistinguishable (match) from the control glass sample. These parameters may include comparing each individual recovered value against the range of the control glass values, or extending the range of the control by ± 3 standard deviation. Other methods used involve "grouping" the recovered results and comparing the mean of these with the mean of the control glass sample values, using statistical methods [16]. If the recovered values fall outside the parameters selected, then the fragments may be regarded as being different. Most modern window glasses have an RI in the range of 1.45–1.53 [18], although some other glass objects may fall outside this range.

If the recovered fragments are found to be indistinguishable from the control glass sample these are analyzed further, often by an analytical technique that will examine elemental composition. Many techniques can be used, such as scanning electron microscopy with energy dispersive X-ray detector (SEM-EDX) [19, 20]. A sample of the matching or indistinguishable recovered glass, and a sample of the control glass are analyzed and the chemical components determined, and compared to one another. The

elemental composition of any recovered fragments should also be within defined parameters, such as plus or minus standard deviation of each element or other statistical/multivariate limits of the control glass sample [16], in order to ascertain if these are indistinguishable in composition from the control glass sample. If not, then the fragments may be regarded as being different. As only one or two recovered fragments are analyzed elementally, the results will impinge on all the recovered fragments whose RI has been determined previously.

Once the recovered fragments have been examined and compared with the control sample, the results can be reviewed by the scientist, from which he can draw certain conclusions about the findings. While the determination of RI and elemental composition are highly discriminating techniques, it is not currently possible to determine conclusively that a recovered fragment, with a matching RI and composition, has come from the control or broken item at the scene.

Interpretation of Findings

Studies have shown that the presence of glass fragments on the surfaces of clothing of an individual from the general population is highly unusual [21, 22]. Moreover, glass particles matching a control glass sample is also significant, even in the presence of nonmatching glass [23, 24]. Therefore, if the scientist has found a group of recovered fragments that are indistinguishable from (matching) the control sample in the tests performed, then he is able to conclude, depending upon the circumstances that there is "some support" for the proposition that the individual has smashed the window concerned. The type of phraseology used and the strength of the evidence will depend upon the circumstances of the case, the findings, and any assessment made. In the United Kingdom, scientists tend to use Bayesian inference for the interpretation of glass evidence, which incorporates the use of at least two competing propositions and the estimation of the likelihood ratio (LR) [16, 25]. The strength of the evidence is usually conveyed in cases involving glass (and other evidence types) using a logarithmic verbal scale [26] going in stages from 'no support' to 'extremely strong support' for a proposition (e.g., LGC Forensics uses a scale comprising inconclusive, no support, limited support, moderate support, moderately strong support, strong support, very strong support and extremely strong support).

References

[1] Shelby, J.E. (1997). *Introduction to Glass Science and Technology*, RSC, paperbacks ISBN 0-85404-533-3.

[2] David, R.L. (2006). *CRC Handbook of Chemistry and Physics*, Taylor and Francis Group, CD-ROM.

[3] Copley, G.J. (2001). The composition and manufacture of glass and its domestic and industrial applications, in *Forensic Examination of Glass and Paint*, B. Caddy, ed, Taylor and Francis, Chapter 2, ISBN 0-7484-0579-8.

[4] Burgoyne, I. & Scoble, R. (1983). *Two Thousand Years of Flat Glass Making*, Printing & stationery department Pilkington Plc, St Helens.

[5] http://www.glassonweb.com/glassmanual/topics/index/laminated.htm. (2008).

[6] Nelson, D.F. & Revell, B.C. (1967). Backward fragmentation from breaking glass, *Journal Forensic Science Society* **7**, 58–61.

[7] Luce, R.J.W., Buckle, J.L. & McInnis, I. (1991). A study of the backward fragmentation of window glass and the transfer of glass to individual's clothing, *Journal of Canadian Society of Forensic Science* **24**, 78–89.

[8] Pounds, C.A. & Smalldon, K.W. (1978). The distribution of glass fragments in front of a broken window and the transfer of fragments to individuals standing nearby, *Journal Forensic Science Society* **18**, 197–203.

[9] Hicks, T., Vanina, R. & Margot, P. (1996). Transfer and persistence of glass fragments on garments, *Science and Justice* **36**, 101–107.

[10] Allen, T.J. & Scrannage, J.K. (1998). The transfer of glass – part 1. Transfer of glass to individuals at different distances, *Forensic Science International* **93**, 167–174.

[11] Locke, J. & Unikowski, J.A. (1991). Breaking of flat glass part 1: size and distribution of particles from plain glass windows, *Forensic Science International* **51**, 251–262.

[12] Allen, T.J., Hoefler, K. & Rose, S.J. (1998). The transfer of glass – part 2. A study of the transfer of glass to a person by various methods, *Forensic Science International* **93**, 175–193.

[13] Allen, T.J., Hoefler, K. & Rose, S.J. (1998). The transfer of glass – part 3. The transfer of glass from contaminated person to another uncontaminated person during a ride in a car, *Forensic Science International* **93**, 195–200.

[14] Allen, T.J., Cox, A.R., Barton, S., Massam, P. & Lambert, J.A. (1998). The transfer of glass – part 3. The transfer of glass fragments from the surface of an item to the person carrying it, *Forensic Science International* **93**, 201–208.

[15] Pounds, C.A. & Smalldon, K.W. (1977). *The efficiency of searching for glass on clothing and the persistence of*

glass on clothing and shoes, Forensic Science Service, CRSE Report Vol. 280, pp. 1–12.
[16] Curran, J.M., Hicks, T.N. & Buckleton, J.S. (2000). *Forensic Interpretation of Glass Evidence*, Taylor and Francis, ISBN 0-8493-0069-X.
[17] Scientific working group for materials analysis (SWGMAT) (2005). Glass refractive index determination July 2004, *Forensic Science Communications* **7**(1), 1–9.
[18] Hamer, P.S. (2001). Microscopic techniques for glass examination, in *Forensic Examination of Glass and Paint*, B. Caddy, ed, Taylor and Francis, Chapter 3, ISBN 0-7484-0579-8.
[19] Scientific working group for materials analysis (SWGMAT) (2005). Elemental analysis of glass July 2004, *Forensic Science Communications* **7**(1), 1–19.
[20] Almirall, J.R. (2001). Elemental analysis of glass, in *Forensic Examination of Glass and Paint*, B. Caddy, ed, Taylor and Francis, Chapter 4, ISBN 0-7484-0579-8.
[21] McQuillan, J. & Edgar, K.A. (1992). Survey of the distribution of glass on clothing, *Journal of the Forensic Science Society* **32**, 333–348.
[22] Petterd, C.I., Hamshere, J., Stewart, S., Bruich, K., Masi, T. & Roux, C. (1999). Glass particles in the clothing of members of the public in south-eastern Australia – a survey, *Forensic Science International* **103**, 193–198.
[23] Lambert, J.A., Satterthwaite, M.J. & Harrison, P.H. (1995). A survey of glass fragments recovered from clothing of persons suspected of involvement in crime, *Science and Justice* **35**, 273–281.
[24] Coulson, S.A., Buckleton, J.S., Gummer, A.B. & Triggs, C.M. (2001). Glass on clothing and shoes of members of the general population and people suspected of breaking crime, *Science and Justice* **41**, 39–48.
[25] Cook, R., Evett, I.W., Jackson, G., Jones, P.J. & Lambert, J.A. (1998). A model for case assessment and interpretation, *Science and Justice* **38**, 151–156.
[26] Evett, I.W., Jackson, G., Lambert, J.A. & McCrossan, S. (2000). The impact of the principles of evidence interpretation on the structure and content of statements, *Science and Justice* **40**, 233–239.

Related Articles

Evidence Interpretation: a Logical Approach

Glass Evidence: Bayesian Approach to

Sampling Trace Evidence

Trace Evidence: Transfer, Persistence, and Value

TINA J. LOVELOCK

Glass Evidence: Bayesian Approach to

Introduction

Glass examiners are often asked to answer questions such as “Does the glass recovered from the suspect come from the crime scene window?” or “Did this person break that window?”. Such questions deal with unknown events and therefore their answers are inherently probabilistic or statistical in nature. For this very reason the analysis will never answer such questions definitively. However, statistical treatments do provide a framework which can ensure objectivity and consistency. Excluding DNA, glass is considered by some to be the field of evidence interpretation which makes the most use of statistics [1]. It is a forerunner of many ideas applied to other evidential material [2]. It is therefore not surprising to find that glass is the field, after DNA, where one can find the most publications showing how probabilities and statistics can help the scientist’s assessment of the significance of the evidence.

Glass is one of the most common types of trace evidence: in New Zealand, for example, in 2005, 18.6% of all physical evidence casework in NZ involved glass (Curran personal communication). The most common scenario where glass evidence may arise is when a window is broken in order to gain entry to a building or vehicle. In this chapter, we will assume that a window has been broken, for illustrative purposes. However, the methodology can be used in any investigation involving broken glass, such as a “hit and run”, assault or even laboratory accidents. If a window is broken, tiny fragments of glass may be transferred to a person’s clothing, footwear and headgear. If there are suspicious circumstances, the crime scene examiner will take a sample from the broken window. This sample can be referred to as the *control sample*, the reference, the known, the source, the bulk, or the crime sample [2]. If the police arrest a suspect, clothing, footwear, and headgear may be seized to search for glass fragments (*see* **Short Tandem Repeats: Interpretation** or [3]). The clothing will be examined for fragments if the case preassessment indicates that useful evidence is likely to be obtained. The fragments recovered from the suspect’s

clothing or person have been referred to as *unknown*, *questioned*, *transferred*, *recovered*, or *suspect* fragments. Here, we will use the terms *recovered* and *source* fragments. We will show how the case can be preassessed (formally or not), how to sample the evidence depending on the case, how examination results can be interpreted and how the value of evidence using the two-stage and continuous approaches can be assessed.

Establishing Working Propositions and Preassessment

Suppose that a witness sees a man break a house window with some sort of tool and attempt to enter. The witness shouts and scares off the breaker before he enters the premises. Half an hour later, Mr G, who fits the witness's description, is arrested in the vicinity of the crime scene. He denies any involvement. His clothing is seized and glass from the broken window is taken. In order to preassess (and assess) the case, the expert needs to evaluate the scientific evidence given two (or more) propositions. The importance of establishing working propositions cannot be overstated: it is a very critical stage, and is probably the most difficult stage of evidence interpretation. One has to be careful to work propositions rather than explanations [4]. Three key principles help the scientist [2, 5, 6]:

1. The scientist must interpret findings in the light of the nonscientific evidence. That is why the scientist needs to know something about the *circumstances* of the alleged incident.
2. The scientist, given *two mutually exclusive* propositions, must consider the findings: here they will be denoted H_p and H_d (for propositions proposed by prosecution and defense).
3. The scientist, *given* the *proposition*, should address the probability of the *evidence*.

A classification (or *hierarchy*) of the different kinds of propositions has been proposed [6]: level 1 (source level), level 2 (activity level) and level 3 (offense level). In order for the uncapitalise Court to address questions at the offense level, assistance from the forensic scientist will be needed. He/she will, depending on the circumstances, address level 1 or level 2. Propositions for glass evidence evaluation at different levels could be the following:

Level 1

H_p : The glass recovered on the jacket of Mr G. came from the broken window.

H_d : The glass recovered on the jacket of Mr G. came from some other source.

Level 2

H_p : Mr G. is the person who broke the window.

H_d : Mr G. had nothing to do with the incident.

Level 3

H_p : Mr G. is the offender.

H_d : Mr G. had nothing to do with the incident.

An example of explanation could be:

Mr G. has been in recent contact with broken glass.

It must be remembered that in order to go from level 1 to level 3, level 2 needs to be addressed. The higher we go in the hierarchy, the higher the value, the greater the requirement for expertise and the greater the need for background information. It may be a difficult task, but we believe that the best person to address this level is the forensic glass examiner (as opposed to a scientist in general or the court).

Once the propositions have been established, the glass examiner can preassess the case either informally or formally [7]. The first option will consist of thinking about what we would expect to find if the propositions were true. For example, would we find matching and/or nonmatching glass, how many fragments would we expect given each proposition? Formal preassessment [7] using the likelihood ratio (LR) approach (see [5] or Section "Interpretation: Source Level 1") consists in assessing six different probabilities using published or subjective data, before even looking at the evidence:

1. The probability of finding no matching glass fragments if the suspect is the person who broke the window.
2. The probability of finding a few (1–2) matching glass fragments if the suspect is the person who broke the window

3. The probability of finding many (3 onwards) matching glass fragments if the suspect is the person who broke the window
4. The probability of finding no matching glass fragments if the suspect had nothing to do with the incident
5. The probability of finding a few (1–2) matching glass fragments if the suspect had nothing to do with the incident
6. The probability of finding many (3 onwards) matching glass fragments if the suspect had nothing to do with the incident.

It should be noted that all of these statements contain level 2 propositions. This is for convenience only. These probabilities are not always easy to evaluate because of the lack of information and because of the variability of the different parameters. Sensitivity analysis [5, 8] with, for example, Bayesian Networks ([9]; Section by Taroni) can be used in order to estimate the influence of the uncertainty on the available parameters. Once the propositions have been established and the case preassessed, a discussion between the forensic scientist and the person asking for the analysis should take place in order to evaluate the case and its possible contribution. It is also possible (e.g., [8]), if and only if expectations regarding the number of fragments to be found given each proposition have been made, to search the suspect items for glass and then only address two of the six probabilities above. If we recover six fragments for example, we would then preassess the case using probabilities *3* and *6*. This would give the customer more value for a little additional cost.

In conclusion, as shown by [2], preassessing the case involves collecting all relevant information, discussing the propositions, identifying the parameters necessary to evaluate the evidence, discussing the search strategy, [3] and estimating the overall value of a case.

Interpretation: Source Level 1

When addressing source level propositions, such as H_p: The glass recovered on the jacket of Mr G. came from the broken window and H_d: If the glass recovered on the jacket of Mr G. came from some other source, then one is interested in determining if the recovered and the source fragments are different (or how different or similar they are) and how common/rare the analyzed characteristics are. The analyzed characteristics can be univariate (a good example would be refractive index measurements, for information on glass analysis, see Section on or [3]) or multivariate (e.g., results of elemental analysis). The evaluation can take place in one stage (the continuous LR approach, see [3, 10]) or two stages. As presented earlier, a very useful approach to interpretation is the so-called Bayesian Approach. This approach is not new Aitken (*see* **Interpreting Expert Opinions: History of**) and has been applied in many different areas such as DNA, disputed paternity or uncapitalise Court presentation. Let E denote the evidence, H_p and H_d the propositions and I the background information we have on the case (e.g., breaking method, time elapsed ...). The odds form of Bayes theorem states:

$$\underbrace{\frac{\Pr(H_p|E,I)}{\Pr(H_d|E,I)}}_{\text{Posterior odds}} = \underbrace{\frac{\Pr(E|H_p,I)}{\Pr(E|H_d,I)}}_{\text{Likelihood ratio}} \times \underbrace{\frac{\Pr(H_p|I)}{\Pr(H_d|I)}}_{\text{Prior odds}} \quad (1)$$

The odds form of Bayes theorem lends itself to a very simple method for updating prior belief: we may rewrite (1) as

$$\text{Posterior odds} = \text{likelihood ratio} \times \text{prior odds} \quad (2)$$

This formulation is convenient because it separates the probabilities relating to the evidence from those relating to the propositions. As mentioned earlier, the former usually concerns the forensic scientist and the latter the court. The LR approach can be applied at the comparison stage [10, 11] or after comparison. The continuous LR approach is more complex but more efficient than the two-stage approach. However, it is useful to begin with the two-stage approach [12] in order to understand the LR thinking.

The Two-Stage Approach

Let's take our example where a window was broken, and say that, after case preassessment, the decision was made to search for glass on Mr G's jacket and that six fragments were recovered. The refractive indices (RIs) of each of those six fragments were measured, as well as the RIs for a selection of fragments from the window broken at the crime scene.

Interpretation Based on Refractive Index Measurements. The first stage of the two-stage

approach consists in determining if the measurements made on the two samples are statistically different or not. Not surprisingly, early statistical interpretations of glass refractive index measurements used formal hypothesis tests for the difference of two population means [13, 14]. Before that, glass experts either used graphical displays of their results, range tests or the so-called 2 or 3 sigma, or standard deviation (SD) tests. An example of those tests with real measurements is given below: it is shown here and it has been shown before that the performance of these tests is poor compared to other methods. These methods have been superseded. A traditional statistical test for the difference in two population means with unknown variances is Student's t-test. Evett [13, 14] proposed the use of Student's t-test for comparing the mean RIs of the recovered and source samples. Walsh *et al.* [15] suggested that Welch's modification of Student's t-test was perhaps more appropriate as it allows the unequal variances in the two populations. This relaxation is quite useful, as it has been shown that measurements on recovered fragments present an SD larger (often 50% larger) than measurements made on a large sample of a window [3]. Both tests require measurements to be grouped before testing. Several grouping algorithms have been presented (divisive and agglomerative grouping methods, [3]). The divisive methods [16] out-performs agglomerative methods [17]. However, the main point is not so much *how* refractive index measurements are grouped but *that* they are grouped at all.

Once, the RI measurements of the recovered fragments have been grouped and the significance level chosen, one can proceed with the test. Let us assume that as in our example, all recovered fragments are grouped and that there has been only one window broken (one source only). If the fragments do not match, then the LR at source level is 0 (this is precisely one of the limitations of the two-stage approach!). If the fragments match, the LR is as follows:

$$\text{LR source level} : \frac{1}{f} \tag{3}$$

The forensic scientist will then estimate the frequency f (which could also be called *match probability*) of the analyzed characteristics. It goes without saying that the frequency must be estimated using the same procedure that was used for comparison.

Table 1 RI Measurements from casework (Courtesy of the Forensic Science Service, FSS London Laboratory Glass Workshop, January 2003)

Hat	Jacket	Window
1.51827	1.51826	1.51828
1.51826	1.51821	1.51828
1.51820	1.51826	1.51826
1.51827	1.51827	1.51827
1.51824	1.51819	1.51825
1.51826	1.51824	1.51826
		1.51827

Example : Comparing RI Measurements. As an example we will take casework data shown in Table 1 (Courtesy of the Forensic Science Service, *FSS London Laboratory Glass Workshop, January 2003*) from fragments recovered on a jacket and on a hat. We will not present the range test and the 3SD test as these have serious limitations and flaws.

If the fragments on each garment are grouped together and then compared using the t-test or the Welch test, the P-values using Student's t-test are respectively 0.14 and 0.04 for the hat and the jacket. With the Welch test, the P-values are 0.19 and 0.07. If our significance level is 0.01, then we would fail to reject the null-hypothesis using either test. That is we would say that we were unable to determine any significant difference between these samples on the basis of their means and that the observed differences can be explained by random measurement error alone. Therefore using the Student's t-test or the Welch test, the samples are not differentiated. To estimate the frequency of occurrence, the expert will then search the relevant database with the same tests that were used for comparing the data. Examples and formulae for this procedure can easily be found (e.g., in [3]). The means, SD and critical value of the Student's t-distribution (using Welch's formula for the degrees of freedom) are used along the appropriate-formula (e.g., in [3], p.151) to give a "match window", which can then be used to determine the frequency. In our example case, using a significance level of 0.99, gives us a match window from 1.51821 to 1.51829. Using this window with the Lambert *et al.* survey [18], yields a frequency of 2%. The corresponding LR would be 50.

Interpretation Based on Elemental Analysis Measurements. Researches have found that elemental

analysis enables to differentiate glasses with a similar RI (see Section on Trace Evidence). Traditional treatment of the data involves determining the mean concentration and the SD for each element. The 3 sigma rule can then be used to determine if the fragments match or not. If any of the range fails to overlap, the fragments are considered not to match. This approach suffers from two problems: the multiple comparison problem (approximately the false rejection rate of 0.1% is multiplied by the number of elements tested, so for 10 elements the false rejection rate would not be 0.1% but 9.6%, [3]) and the fact that this rule fails to take into account any estimated correlation between the elements. As a solution, a multivariate analog to Student's t-test – Hotelling's T^2 – has been suggested [3, 19] and is highly recommended. Hotelling's T^2 may be used on its own as a hypothesis test, or combined into a continuous LR approach. Aitken [20] proposed a formulation based on kernel density estimation. One advantage of Aitken's approach is that it drops the (untestable) assumption of multivariate normality (MVN) of the elemental concentration means. This advantage is perhaps more psychological than practical. The assumption of MVN is actually guaranteed by the Central Limit Theorem. However, Aitken and Lucy [21] and Aitken *et al.* [22, 23] proposed an aesthetically appealing way of looking at the evaluation of trace evidence. They take a "variance components" (or random effects if you are Bayesian) approach to the treatment of the measurements. Aitken *et al.* explicitly model three sources of variation. The first source of variation is the variation that arises from the measurement process. For example, if you buy a digital thermometer or set of scales, you'll usually find in the accompanying documentation a statement like, "this device is accurate to $\pm 2\,^{\circ}\mathrm{C}$ or $\pm 0.5\,\mathrm{g}$". This variation is purely an artifact of the measurement device. The second source of variation is the variation that is inherent in the item itself. For example, if we are interested in the measurement of the concentration of calcium in a piece of glass, we might regard it as being homogenous throughout. This does not mean however, that if I measure the concentration at two different sites on the glass I will get exactly the same concentration (ignoring measurement error). What it means is that the concentration varies uniformly throughout the piece of glass. By explicitly modeling this we attempt to understand the natural variation within the same source. Finally, the third source of variation is the variation that arises between items from different sources. This variation is the one that forensic scientists naturally focus on. The Bayesian approach also focuses on the first and second sources (either separately or together). Aitken *et al.* provide software for the computation of their statistic, as does Curran *et al.* It is probably safe to say that the statistics for elemental concentration caseworks are underdeveloped and could withstand substantial improvement (*see* **Glass**).

Disadvantages of the Two-stage Approach. The two-stage approach suffers from several drawbacks: first it assumes that if the fragments come from the broken window, then they would match with probability 1. There is however a possibility that certain fragments would not match the crime scene sample, for example because of the float manufacturing process or/and sampling procedures [15]. Another disadvantage of the two-stage approach is the significance test and the so-called fall-off-the cliff effect (label given by Ken Smalldon). If we imagine a situation where the P-value is 0.009, the samples would be declared to not match. However, if the P-value changed by $0.002-0.011$, then the samples would be declared to match. Such a reversal would be difficult to justify to the court, and would certainly trouble the forensic scientist if multiple matching fragments, with a rare RI close to the window, had been recovered in that case. Hypothesis testing fails to incorporate relevant evidence such as the relative frequency of the glass and the number of glass fragments that have been recovered.

The Continuous Approach

Abandoning a match/not match approach [10] is one of the greatest advances in interpretation. It enables to give a different value to a good match (close) and a "poor" match, there is no fall-off-the cliff effect, and it allows the scientist to weigh all the evidence in one process rather than in a step-by-step fashion. Moreover, it is less prone to be influenced by departures from normality [Bennett *et al.* [24], [1]] than significance testing is. The approach is described in several books (e.g., [1, 2]) and has been applied to different evidence type (e.g., inks, speaker recognition, fingerprints). To distinguish the LR applied for the assessment of evidence and for assessing the value of a comparison, we will use

LR for the global assessment of glass evidence and lr continuous for the assessment of the value of a comparison.

Interpretation Based on Refractive Index Measurements. In 1996, a case was presented [15] where a pharmacy window was broken. Fragments recovered on two suspects were compared to the broken window based on refractive index measurements followed by a t-test. According to the test, the samples did not match. There were however contradictory elements: a lot of glass fragments (rare event), some coming from a float surface (rare event), were recovered from the suspects' clothing. The refractive index measurements of the glasses were close. And, paint flakes, indistinguishable from the crime scene window frame had also been found on one of the suspects. Thus, despite the results of the t-test, the evidence supported the proposition that the suspects were at the crime scene when the window was broken. It was therefore suggested to replace the match/non match approach with a continuous likelihood based scale. This involves replacing $1/f$ with

$$lr_{\text{cont}} = \frac{f(\bar{X} - \bar{Y}|S_X, S_Y)}{\hat{g}(\bar{Y})} \tag{4}$$

where $\hat{g}(\bar{Y})$ is the probability density for float glass measured at the mean of the recovered sample. This value is usually obtained from a kernel density estimate. $f(\bar{X} - \bar{Y}|S_X, S_Y)$ is the value of the probability density for the difference of the two sample means. This is approximated by an unscaled t-distribution using Welch's modification to Student's t-test. Aitken [20] also showed that the assumption of normality could be dropped and the numerator of the LR could be replaced by a univariate kernel density estimate.

Interpretation Based on Elemental Analysis Measurements. When glass evidence is quantified by elemental composition, an analogous equation proposed by Curran [3, 19, 25] can be used. As previously noted, Aitken *et al.* [21, 23] (200x) have a treatment based on kernel density estimation.

Interpretation: Activity Level 2

When addressing activity level propositions, important forensic parameters such as the number of glass fragments, where they have been recovered (e.g., hair, garments, shoes), and the method of breaking are taken into account. As mentioned earlier, we think that the expert should always try to address activity level propositions: the higher in the hierarchy the greater the value added by the evidence. And for the Court to address the offense level, these parameters have to be taken into account and who is the best person to evaluate these parameters, if it is not the forensic scientist?

Example: One Group, One Control

Consider a case where the fragments have been grouped and where one window has been broken. When preassessing the case the following propositions have been chosen:
H_{p} : Mr G. is the person who broke the window.
H_{d} : Mr G. had nothing to do with the incident.
To evaluate the evidence E, two questions have to be answered:

> *What is the probability of the evidence given that the prosecution proposition is correct and given the background information.* (background information is a key element, if information changes, probabilities have to be reassessed, (*see* **Evidence Interpretation: a Logical Approach**).

And

> *What is the probability of the evidence given that the defense proposition is correct and given the background information.*

These two probabilities form the numerator and denominator of the LR.

$$LR = \frac{\Pr(E|H_{\text{p}}, I)}{\Pr(E|H_{\text{d}}, I)} \tag{5}$$

If we consider the denominator first, the question, in more detail, is:

> *What is the probability that one would find a single group of m matching fragments on the surface of the fleece jacket if Mr G. has not broken the window at the scene, given what is known about the incident and the suspect.*

As mentioned earlier, the information that we have on the suspect's activities is provisional and may change. Regardless of what we know about Mr G, it is inarguable that he has come to the attention

of the police in connection with a breaking offense. Casework clothing surveys (see below) are surveys of people who have also come to police notice in that way. Thus, it is believed that this is the most relevant survey in this case. If Mr G had said that he worked on a demolition center, then the most relevant survey would have been surveys done on persons working on demolition centers. But, in this particular case, he has given no explanation. If, using this information, we rephrase our question it becomes:

> *If we examine the fleece jacket of a man who has come to police's notice on suspicion of a breaking offence, yet he is unconnected with the offense, what is the probability that one would find a single group of m fragments which matches the control in that particular case?*

Casework clothing surveys distinguish between groups of glass that match the casework crime scene samples and those that do not. If we assume that the matching glass did come from the incident investigated, then the nonmatching glass can be considered as the background.

Let $b_{g,m}$ be the probability of g groups of size $(m_1, \ldots, m_g) = m$ being found. (In the literature, the terms P_i and S_j have been used to denote the probabilities relating to the number of groups P and the size S of the groups. This is equivalent to the one term $b_{g,m}$). Let f denote the probability that a group of glass fragments on the surface of such a person's clothing will match in RI the window in this particular case (lr continuous will be used instead of f if using the continuous lr to assess the value of a comparison). We make two assumptions:

> *A1: There is no association between the number of groups of glass found on a person's clothing and the size of these groups and*
> *A2: There is no association between the frequency of a given RI on clothing with either the number or the size of the group.*

It is unlikely that either of these assumptions is exactly true, but it is believed that they are correct at least as any first order approximation [3]. By invoking these two assumptions, we can estimate the denominator and answer our question *what is the probability that one would find a single group of m fragments which matches the control in that particular case on a person unconnected with the crime?*

$$\Pr(E|H_{\mathrm{d}}, I) = b(1, m) \cdot f$$

or if using the continuous lr

$$\Pr(E|H_{\mathrm{d}}, I) = b(1, m) \cdot lr_{\mathrm{cont}} \quad (6)$$

Estimating the numerator is slightly more complex as we have to allow for at least two possible explanations (actually, there are $m + 1$ explanations, see [3]) for the evidence given H_{p}.

1. Either, the group of fragments was transferred from the scene window, and Mr G had no glass beforehand.
2. Or no glass was transferred from the window and/or recovered, but Mr G had already one group of glass on his fleece.

If we invoke the assumptions A1 and A2, and let t_m denote the probability that given H_{p} and I, m glass fragments would be recovered on Mr G's fleece, then the denominator is:

$$\Pr(E|H_{\mathrm{p}}, I) = t_0 \cdot b(1, m) \cdot f + t_m \cdot b(0) \quad (7)$$

Now that we've defined the numerator and denominator terms, we can calculate the LR:

$$LR = \frac{\Pr(E|H_{\mathrm{p}}, I)}{\Pr(E|H_{\mathrm{d}}, I)} = t_0 + \frac{b(0) \cdot t_m}{b(1, m) \cdot f} \quad (8)$$

If we are using the continuous approach to glass comparison then this becomes

$$LR = t_0 + \frac{b(0) \cdot t_m}{b(1, m)} \cdot lr_{\mathrm{cont}} \quad (9)$$

There are infinite variations around this scenario [2, 3], however this example shows how probabilities can help the forensic scientist to assess the value of glass. This framework allows taking into account all factors that are intuitively important.

Estimation of Background and Transfer Probabilities

To assess the value of glass evidence at activity level, in addition to the lr continuous (or f), the expert has to estimate background and transfer probabilities. Clothing surveys enable the estimation of the former. Two types of clothing surveys have been performed: some on general population [26–31], and some on suspects [8, 18, 31]. Depending on the case and on the defense proposition, the expert will use one or the other. For example, when the suspect does not

have any explanation for the presence of glass on his/her clothing, the surveys on suspects' clothing will be used. If the recovered glass has been found on a victim, then general clothing surveys may be more relevant. Surveys have shown that more glass is recovered as background on persons suspected of breaking offenses, than on general population. When, choosing the relevant survey, it is important to consider the population of interest, as well as where and how the fragments have been searched for and on which garment. For instance, if the fragments have been recovered from the pockets of a person who is the victim of a crime, then a survey of a general population (e.g., [26]) where the scientists looked for fragments on pockets and turn-ups should be used. If the fragments were recovered from the outer surfaces of the clothing of a suspect, then surveys on suspect populations (e.g., from casework) ought to be used [8, 18, 31]. There are also surveys which looked at the question of background glass on shoes [32], Harrison *et al.* [33], [30] or in hair [34]. Curran *et al.* [3] have a comprehensive chapter on background surveys which includes all of the references here except for [31]. This latter study is a combination of survey data and extensive modeling of the $b_{(g,m)}$ terms and should be regarded as an important contribution to this subject. "Transfer probabilities" incorporate three distinct processes: the initial transfer, persistence and recovery. It has been argued that they depend on too many factors and are too difficult to estimate. Whereas it is true that transfer is a variable phenomenon, it remains essential to estimate transfer probabilities and a lot of research has been dedicated to fill in the gaps in our knowledge (for a review, see [3]). Experiments have been performed on glass transferred onto the ground in order to identify the parameters involved in initial transfer [35–40]. Researchers have studied transfer [38, 39] and persistence of glass recovered on garments [28, 41–47]. And last, but not least, because of the modular nature of transfer and persistence, graphical models (a superclass of Bayes Nets) have been suggested as a tool for the estimation of transfer probabilities [48]. As transfer is a variable phenomenon, we suggest that before searching the garments (when preassessing the case), the forensic scientist estimates a range of transfer probabilities. This will allow him/her to perform sensitivity analysis and show how uncertainty on transfer probabilities influences the overall value of glass evidence in that particular case.

Concluding Remarks

Many of the statistical techniques presented here are very difficult to implement without the aid of computer software. Commercial and noncommercial packages exist: some allow the estimation of transfer probabilities (TRANSFER), others measurement comparisons using a two-stage and/or a continuous approach (statistical analysis of glass (STAG) and fragment data system (FDS) allows grouping and comparing RI data, ELEMENTAL allows to compare data from elemental analysis). The computer aided glass evaluation (CAGE, CAGE2000) software allows the expert to assess glass evidence as a whole. It provides an interactive guide to estimate transfer and background probabilities by storing experimental and survey data (*see* **Use of Knowledge-Based Systems in Forensic Science**). The UK Forensic Science Service is considered as a pioneer in this field and is at the origin of FDS and CAGE. However, recently most products have been developed during PhD researches ([25], TRANSFER, STAG, ELEMENTAL; [8], CAGE2000).

In conclusion, we would like to say that there are very few alternative approaches to glass interpretation and that even if the LR approach is not a panacea, it provides a very good framework. It could be argued that because glass characteristics change a lot, then one should not attempt to estimate its frequency of occurrence. The same argument has been applied to transfer or background probabilities because they are hard to estimate. Proponents of such arguments believe that the solution is to analyze the glass with more discriminating techniques so that the glass would become "unique". However, we do not believe that such considerations will answer the Court's question. Assessing the strength of the match at source level is only one part of the process. As we have seen it is essential to address questions relative to activity level. Statistical tools enable us to deal with the uncertainty inherent in these questions and those uncertainties will remain regardless of whether we can determine a "match" between two samples of glass with 100% accuracy.

References

[1] Curran, J.M. (2003). The statistical interpretation of forensic glass evidence international statistical review, *International Statistical Review* **71**(3), 497–520.

[2] Aitken, C.G.G. & Taroni, F. (2004). *Statistics and the Evaluation of Evidence for Forensic Scientists*, 2nd Edition, John Wiley & Sons, Chichester.

[3] Curran, J.M., Hicks, T.N. & Buckleton, J.S. (2000). *Forensic Interpretation of Glass Evidence*, CRC Press LLC, Boca Raton, FL.

[4] Evett, I.W., Jackson, G. & Lambert, J.A. (2000). More in the hierarchy of propositions: exploring the distinction between explanations and propositions, *Science and Justice* **40**(1), 3–10.

[5] Evett, I.W. & Weir, B.S. (1998). *Interpreting DNA Evidence – Statistical Genetics for Forensic Scientists*, Sinauer Associates, Sunderland.

[6] Cook, R., Evett, I.W., Jackson, G., Jones, P.J. & Lambert, J.A. (1998). A hierarchy of propositions: deciding which level to address in casework, *Science and Justice* **38**(4), 231–240.

[7] Cook, R., Evett, I.W., Jackson, G., Jones, P.P. & Lambert, J.A. (1998). A model for case assessment and interpretation, *Science and Justice* **38**(3), 151–156.

[8] Hicks, T. (2003). *De l'Interprétation des Fragments de Verre en Criminalistique*, Ph.D. Thesis, Université de Lausanne, Lausanne, Suisse.

[9] Taroni, F., Aitken, C., Garbolino, P. & Biedermann, A. (2006). *Bayesian Networks and Probabilistic Inference in Forensic Science*, John Wiley & Sons, Chichester.

[10] Lindley, D.V. (1977). A problem in forensic science, *Biometrika* **64**(2), 207–213.

[11] Evett, I.W (1986). A Bayesian approach to the problem of interpreting glass evidence in forensic science casework, *Journal of Forensic Science Society* **26**(1), 3–18.

[12] Evett, I.W. & Buckleton, J. (1990). The interpretation of glass evidence – a practical approach, *Journal of Forensic Science Society* **30**, 215–223.

[13] Evett, I.W. (1978). The interpretation of refractive index measurements II, *Forensic Science International* **12**, 37–47.

[14] Evett, I.W. (1977). The interpretation of refractive index measurements, *Forensic Science International* **9**, 209–217.

[15] Walsh, K.A.J., Buckleton, J.S. & Triggs, C.M. (1996). A practical example on the interpretation of glass evidence, *Science and Justice* **36**(4), 213–218.

[16] Triggs, C.M., Curran, J.M., Buckleton J.S. & Walsh, K.A.J. (1997). The grouping problem in forensic glass analysis: a divisive approach, *Forensic Science International* **85**(1), 1–14.

[17] Evett, I.W. & Lambert, J.A. (1982). The interpretation of refractive index measurements III, *Forensic Science International* **20**, 237–245.

[18] Lambert, J.A., Satterthwaite, M.J. & Harrison, P.H. (1995). A survey of glass fragments recovered from clothing of persons suspected of involvement in crime, *Science and Justice* **35**(4), 273–281.

[19] Curran, J.M., Triggs, C.M., Almirall, J.R., Buckleton, J.S. & Walsh, K.A.J. (1997). The interpretation of elemental composition measurements from forensic glass evidence I, *Science and Justice* **37**(4), 241–244.

[20] Aitken, C.G.G. (1995). *Statistics and the Evaluation of Evidence for Forensic Scientists*, John Wiley & Sons, Chichester.

[21] Aitken, C.G.G. & Lucy, D. (2004). Evaluation of trace evidence in the form of multivariate data, *Applied Statistics* **53**(4), 109–122.

[22] Aitken, C.G.G., Shen, Q., Jensen, R. & Hayes, B. (2007). The evaluation of evidence for exponentially distributed data, *Computational Statistics and Data Analysis* **51**(12), 5682–5693.

[23] Aitken, C.G.G., Zadora, G. & Lucy, D. (2007). A two-level model for evidence evaluation, *Journal of Forensic Science* **52**(2), 412–419.

[24] Bennett, R.L., Curran, J.M., Kim, N.D., Coulson, S.A. & Newton, A.N. (2003). Spatial variation of refractive index in a pane of glass, *Science and Justice* **43**(2), 71–76.

[25] Curran, J.M. (1996). *Forensic Applications of Bayesian Inference to Glass Evidence*, University of Auckland.

[26] Pearson, E.F., May, R.W. & Dabbs, M.D.G. (1971). Glass and paint fragments found in men's outer clothing – report of a survey, *Journal of Forensic Science* **16**(3), 283–300.

[27] McQuillan, J. & Edgar, K. (1992). A survey of the distribution of glass on clothing, *Journal of Forensic Science Society* **32**(4), 333–348.

[28] Hoefler, K., Hermann, P. & Hansen, C. (1995). A Study on the Persistence of Glass Fragments on Clothing After Breaking a Window, *Australian and New Zealand Meeting*, Sydney, Australia.

[29] Roux, C., Kirk, R., Benson, S., van Haren, T. & Petterd, C.I. (2001). Glass particles in footwear of member of the public in south-eastern Australia – a survey, *Forensic Science International* **116**(2), 149–156.

[30] Lau, L., Beveridge, A.D., Callowhill, B.C., Conners, N., Foster, K., Groves, R.J., Ohashi, K.N., Sumner, A.M. & Wong, H. (1997). The frequency of occurrence of paint and glass on the clothing of high school students, *Journal of the Canadian Society of Forensic Science* **30**(4), 233–240.

[31] Coulson, S.A., Buckleton, J.S., Gummer, A.B. & Triggs, C.M. (2001). Glass on clothing and shoes of members of the general population and people suspected of breaking crimes, *Science and Justice – Journal of the Forensic Science Society* **41**(1), 39–48.

[32] DeHaan, J.D. & Davis, R.J. (1977). A survey of men's footwear *Journal of Forensic Science*, **17**(4), 271–185.

[33] Harrison, P.H., Lambert, J.A. & Zoro, J.A. (1985). A survey of glass fragments recovered from clothing of persons suspected of involvement in crime, *Forensic Science International* **27**, 171–187.

[34] McQuillan, J. & McCrossan, S. (1987). *The Frequency of Occurrence of Glass Fragments in Head Hair Samples-A Pilot Investigation*, Northern Ireland Forensic Science Laboratory. (572).

[35] Nelson, D.F. & Revell, B.C. (1967). Backward fragmentation from breaking glass, *Journal of Forensic Science Society* **7**(2), 58–61.

[36] Locke, J. & Unikowski, J.A. (1991). Breaking of flat glass – part 1: size and distribution of particles from plain glass windows, *Forensic Science International* **51**(2), 251–262.
[37] Locke, J. & Unikowski, J.A. (1992). Breaking of flat glass – part 2: effect of pane parameters on particle distribution, *Forensic Science International* **56**(1), 95–106.
[38] Pounds, C.A. & Smalldon, K.W. (1978). The distribution of glass fragments in front of a broken window and the transfer of fragments to individuals standing nearby, *Journal of Forensic Science Society* **18**, 197–203.
[39] Luce, R.J.W., Buckle, J.L. & McInnis, I.A. (1991). Study on the backward fragmentation of window glass and the transfer of glass fragments to individual's clothing, *Journal of the Canadian Society of Forensic Science* **24**, 79–89.
[40] Locke, J. & Scranage, J.K. (1992). Breaking of flat glass – part 3: surface particles from windows, *Forensic Science International* **57**, 73–80.
[41] Pounds, C.A. (1977). *The Efficiency of Searching for Glass on Clothing and the Persistence of Glass on Clothing and Shoes*, Central Research Establishment (HOCRE). (208).
[42] Brewster, F., Thorpe, J.W., Gettinby, G. & Caddy, B. (1985). The retention of glass particles on woven fabrics, *Journal of Forensic Science* **30**(3), 798–805.
[43] Hicks, T., Vanina, R. & Margot, P.A. (1996). Transfer and persistence of glass fragments on garments, *Science and Justice* **36**(2), 101–108.
[44] Allen, T.J., Hoefler, K. & Rose, S.J. (1998). The transfer of glass. Part 2. A study of the transfer of glass to a person by various methods, *Forensic Science International* **93**, 175–193.
[45] Allen, T.J. & Scranage, J.K. (1998). The transfer of glass – part 1: transfer of glass to individuals at different distances, *Forensic Science International* **93**, 167–174.
[46] Allen, T.J., Hoefler, K. & Rose, S. (1998). The transfer of glass – part 3: the transfer of glass from a contaminated person to another uncontaminated person during a ride in a car, *Forensic Science International* **93**, 195–200.
[47] Allen, T.J., Cox, A.R., Barton, S., Messam, P. & Lambert, J.A. (1998). The transfer of glass – part 4: the transfer of glass fragments from the surface of an item to the person carrying it, *Forensic Science International* **93**, 201–208.
[48] Curran, J., Triggs, C.M., Buckleton, J.S., Walsh, K.A.J. & Hicks, T. (1998). Assessing transfer probabilities in a Bayesian interpretation of forensic glass evidence, *Science and Justice* **38**(1), 15–21.

JAMES M. CURRAN AND TACHA N. HICKS

Globes: Light, Examination *see* Light Bulbs and Filaments: Examination of

Grouping: Blood *see* Blood Grouping

GSM Analysis and PDAs

General Introduction to Cellular Phone Networks

Cellular Phone Network Global Architecture

Cellular phone networks (*see* **Analysis: Computer Network**) may be based on various technologies and standards, but they all share a common philosophy: to allow users to have wireless phone calls (and advanced services) while being able to move on wide distances without being disconnected.

Cellular phone networks are wide: they may extend over a country or even over several continents. They rely on both an air interface (radio link between cellular phone and a base station) and a terrestrial interface (edge and core network equipment). Area covered by a single radio base station is called a *cell*, hence the name "cellular network/cellular phone". Cell shape and cell area vary depending on emitted radio signal strength, air condition (rain, snow, etc.), physical obstacles (such as hills, forests, and buildings), wave-reflecting or wave-absorbing surfaces, (lake, clouds, etc.), and various other factors. Distance between a cellular phone and its radio base station may extend up to several dozen kilometers.

"Handover" is the ability for a cellular phone to switch from one radio base station to the next radio base station[a] without being disconnected from the network. "Roaming" is the ability for a cellular phone to use different cellular networks (e.g., when traveling abroad).

Other Radio Communication Networks

Cellular phones should not be confused with plain cordless phones (such as digital enhanced cordless telecommunications (DECTs) phones). The latter may be considered only as cordless extension to wired phones (such as public switched telephone network (PSTN) phones). They allow users to have wireless phone calls only in a geographically limited environment (e.g., within a building, at home, or at the office), though distances can go up several hundred meters thanks to additional "relays". Coarsely compared to Internet Protocol (IP) networks, cellular phone networks may be considered autonomous system (AS)[b] within the Internet, while cordless phones may be considered part of a (wireless) local area network (W)LAN.

Cellular phone networks shall neither be confused with terrestrial trunked radio TETRA[c] networks (radio terminals such as walkie-talkies used in professional emergency networks such as police forces, fire brigades, etc.) nor with satellite phone networks (where terminal connects to a satellite constellation – such as Iridium, Inmarsat, or Globalstar – *via* the air). Note, though, that satellite phone handsets may share similarities with cellular phone handsets (functions and services, potential use of a chip card to connect to the network, etc.).

Cellular Phone Network Technologies

In order to allow several users accessing the network from the same place at the same time, cellular phone networks must use radio resource-sharing technologies. Main technologies used are TDMA (time division multiple access), FDMA (frequency division multiple access), CDMA (code division multiple access), or combinations of these.

- In TDMA, each user is given a time-slot.
- In FDMA, each user is given a frequency spectrum portion.
- In CDMA, each user is given a code.

This article does not intend to give further details, so interested readers may look for books on radio technology.

Cellular phone networks shall convey voice, data, and signals. Voice and data are both user defined, and user data might be anything from text messages to pictures. Signaling is nonuser-defined data (technical data/service data) that is required to establish a communication channel between user A and user B. Note, though, that user data (such as short message service (SMS)) may be conveyed in signaling channels.

Since user data may also be website-browsed content, streamed videos, or streamed television, it may not be (strictly speaking) user defined, but rather user dependant.

On a historical perspective, cellular phone networks are commonly classified into three categories:

- 1G (first generation): first generation cellular phone networks were based on analog standards (such as Nippon Telegraph and Telephones (NTT) in Japan; Nordic Mobile Telephones (NMT), C-Netz, Radiocom 2000, and Total Access Communication System (TACS) in Europe; Advanced Mobile Phone System (AMPS) in the United States, etc.). Nowadays, they have almost completely disappeared, even though some operators have not yet discontinued their service.
- 2G (second generation): second generation cellular phone networks shifted to digital standards (such as personal digital cellular (PDC) in Japan; Interim Standard-95 (IS-95) in South Korea and in the United States; Global System for Mobile (GSM) in Europe; Digital Advanced Mobile Phone System (D-AMPS) and Integrated Digital-Enhanced Network (iDEN) in the United States, etc.).
- 3G (third generation): third generation cellular phone networks are supposed to offer higher data rates and enhanced compatibility, along with multimedia services. Freedom of mobile multimedia access (FOMA) (in Japan), CDMA2000 (such as in the United States) and Universal Mobile Telecommunication System (UMTS) (such as in Europe) are all compliant with the International Mobile Telecommunications-2000 (IMT-2000) standard issued by International Telecommunication Union ITU.

Some intermediate technology networks (such as general packet radio service (GPRS)) are commonly referred to as *2.5G networks*, since they offer extended capacities compared to 2G networks, while not fully complying with 3G standards. Future "full-IP" networks, which would be even more

Internet-linked and with enhanced roaming capabilities, are commonly referred to as *4G networks*, though no standard has yet been defined.

Compatibility between the aforementioned standards (PDC, IS-95, GSM, iDEN, D-AMPS, FOMA, CDMA2000, UMTS, etc.,) varies and depends upon several factors (among which are frequency used, need for a subscriber identity module (SIM) card, etc.). This article does not intend to give a full compatibility chart to cellular phone users traveling abroad.

The following points may be noted:

- In spite of common beliefs, a user has no need for a 2.5G or 3G network to be able to send data (such as binary data) through a cellular phone network: "GSM-data" subscription (e.g.), based on circuit switched data (CSD) technology, may be sufficient. Bank card counterfeiting based on skimming sometimes uses "GSM-data" subscription to send card numbers abroad.
- Wireless data protocols such as I-mode or Wireless Application Protocol (WAP) should not be considered on the same level as 1G/2G/2.5G/3G/4G technologies: they only allow web-content presentation on cellular phones; they do not define any underlying (physical) transportation technology.

Mobile Phone Handset Architecture

Some handset manufacturers have a worldwide market share (such as Nokia, Sony Ericsson, Samsung), while others cover a narrower geographical area (such as Alcatel and Sagem – mainly in France, Siemens – mainly in Germany, Motorola – mainly in Northern America).

Hardware Architecture

Hardware architecture may be divided into three parts: handset itself, data storage memory card, (optional), and SIM card (depending on network technology).

Much like many other mobile devices, a handset may be considered a small-size computer. It is composed of a motherboard with a processor, random access memory (RAM), mass memory storage device (MMSD), power source, and additional modules, all interconnected through a bus and packaged in a case. The case bears the keyboard, screen, and connectors (power, data storage memory card, etc.).

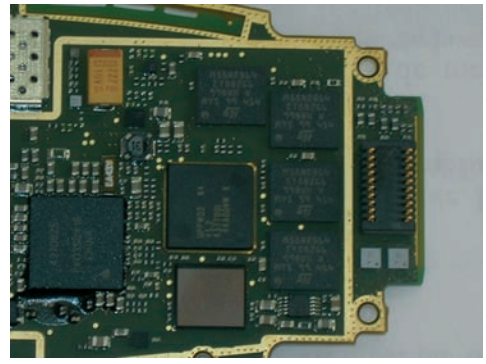


Figure 1 Nokia mobile phone embedding four flash memories

Mass storage generally relies on Flash memory (either NAND or NOR). There may be one single Flash component or several of them (up to four on some Nokia phones, for example (Figure 1)). Some handsets embed both NAND Flash and NOR Flash: in such cases, NOR Flash is used to store firmware (because it allows execute in place), and NAND Flash is used to store user data (because more read/write cycles are possible). On the most recent handset models, Flash memory is sometimes replaced by 1-in. microsize hard-disk drive (Hitachi and Cornice[d] are major manufacturers of such microdrives).

Power is supplied by main the battery, which is removable and rechargeable. Without such main battery, handset cannot be powered on without any external power supply (e.g., cable and plug). Many modern handsets also use an additional microsize button battery, used to avoid date and time setting erasure when main battery is removed (Figure 2).

On modern handsets, the antenna (and the whole radio module) is embedded in motherboard and cannot be disconnected easily.

Apart from the GSM connectivity module, additional modules are optional. Most of the time, they consist of optics (digital still camera or video camera) (*see* **Image Processing and Analysis**) or they offer extra wireless capabilities, such as infrared data association (IrDA) or Bluetooth; some handsets even have WiFi capabilities (Institute of Electrical and Electronics Engineers (IEEE) 802.11 standard), such as Nokia

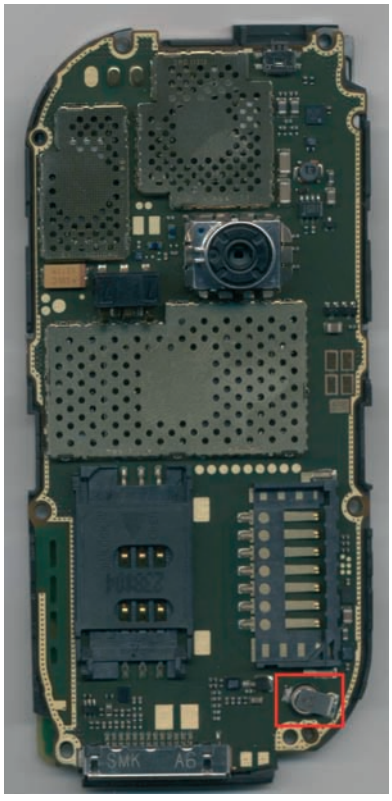


Figure 2 Additional microsize button-battery on a Nokia 7610 mobile phone

6136, which can be used along with a "Unik subscription" from the French operator Orange, to switch seamlessly from a voice over Internet protocol (VoIP) phone call through a WiFi access point to a GSM network. Global positioning system (GPS) may also be found (such as on Pharos handsets, or on some Hewlett Packard iPAQ models).

Data storage memory card is optional, though quite common. Card format may be virtually anything existing on the market (dozens of different format exist). Nevertheless, most common formats are Multi Media Card (MMC) (MMC and MMC micro), Secure Digital (SD), (SD, miniSD, and microSD) and Memory Stick (MS) (MS and MSmicro/M2), because microsize versions of these formats exist, which fit well the mobility and light-weight requirements. Such memory cards rely on Flash memory.

One interesting feature of the MMC system specifications is password protection: using the CMD42 LOCK_UNLOCK command, one can totally prevent MMC card access (either writing or reading). Since SD derives from MMC and is supposed to partially comply with MMC system specification, some SD cards also support password protection. Password protection on MMC and SD, though, is largely vendor dependent and remains quite rare in the mobile phone world. At the time of writing, it seems that Symbian S60 phones supported memory card password protection, as well as Windows Mobile 5.0 (or higher) phones embedding BSquare SDIO Now 2.0, or higher application.

According to various fora,[e] it seems that the following phones would support memory card password protection at the time of writing: Nokia N71/73/75/80/93, Nokia 6230/6230i/6233/6265i/6270/6275/6280/6600/6670/7373/7610/9210i/9300, Nokia E-series, Motorola A780, and Motorola ROKR E6.

SIM chip card comes from GSM standard. Its role is to authenticate the user on the network so as to check his/her credentials (using Ki cryptographic key), and to provide Kc cryptographic key to allow voice encryption over air interface. One single SIM card may be used in different handsets by same user, so that one does not have to shift telecommunication operator subscription when shifting handset (and vice versa). When no SIM card is used (as in PDC networks in Japan), users have to buy a new handset when swapping telecommunication operators.

SIM cards have data storage capacities, but those are mostly standardized: user with no technical knowledge can only store specified sets of data (such as text messages, address book, etc. see section "In Mobile Phone Handset"). Storage size is rapidly increasing.

Some SIM cards look like one single physical chip card, whereas they actually embed several SIM card data (including several sets of cryptographic keys). A legacy application for such SIM cards is to allow using one single SIM card with both a personal phone number and a corporate phone number (with separate billing), or to have a two-in-one SIM card for domestic use and roaming use. SIM Max[f] software can be used to "build" such "multi SIM" cards, if not proposed off-the-shelf by telecommunication operator (such as Tele2 in the Netherlands, or the Tunisiana "switch service"[g] in Tunisia).

Mass-storage SIM cards have been created recently: a memory card and a SIM card incorporated into a single case. Addressed via MMC or SD protocol,

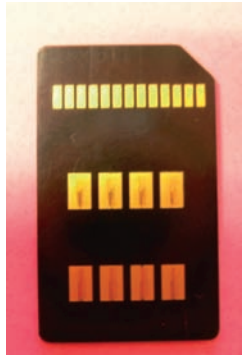


Figure 3 Orange SIM+card

the card behaves like an MMC/SD memory card; addressed via ISO7816 interface, the card behaves like a SIM card. A compatible phone is needed to fully support both features. Examples of such cards are "multimedia SIM" (jointly produced by Gemalto, Oberthur Card Systems, and Sagem Orga), "SIM+" (by Orange) (Figure 3), "MegaSIM" (by Msystems), "S-SIM" (by Samsung).

Software Architecture

As with any other "active device" (either mobile or not), a cellular phone handset has an operating system (OS), often called *firmware*. Data (either OS or user data) is stored as files according to a file system, which may or may not be OS-specific.

Various OS exist in the mobile phone market as follows:

- Created jointly by Nokia, Motorola, and other manufacturers, Symbian OS now equips most Nokia, Sony Ericsson, Motorola, Panasonic, Sharp, Siemens, LG, and other phones. Symbian OS supports FAT32, as well as other file systems.
- Microsoft Windows CE shall not be confused with Microsoft Windows Mobile (which is a platform including Windows CE OS and many applications). Companies such as HTC, Hewlett-Packard, and Acer produce Windows CE mobile phones. There was no dedicated file system initially (user data was stored in RAM), before the Persistent Storage file system was introduced in 2002.
- Linux-based mobile phones abound, even in major manufacturer catalogues: Motorola (e.g., A-series), Samsung (e.g., SGH-i858), NEC, and Panasonic all produce several Linux mobile phones. Trolltech Qtopia Linux stack for mobile devices is worth mentioning. Interested readers may refer to www.linuxdevices.com. Examples of file systems are Yet Another Flash File System (YAFFS), Journalized Flash File System (JFFS2), Samsung Robust File System (RFS), etc.
- Palm OS history is quite difficult to follow and may cross the Linux one. Handspring/Palm (Treo and Tungsten), Sony (Clié), Kyocera, and Samsung have all produced Palm OS mobile phones. There was no dedicated file system initially (user data was stored in RAM), before Nonvolatile File System (NVFS) was introduced in version 5.4 of the OS.
- Nucleus RTOS is used in various Motorola, Samsung, LG, Nec, etc. mobile phones. Nucleus RTOS supports FAT file systems.
- Research In Motion (RIM) OS has become very popular with Blackberry phones. RIM OS file system is proprietary.
- Mac OS X has been introduced in the phone market since mid-2007 with Apple iPhone.

As opposed to hard disk drives, where sector 0 always remains sector 0,[h] data stored in Flash memory follow a two-level structure as follows:

- "Upper level" is logical: this is the file system. It may be FAT, FAT-like (e.g., transactional file system (TFS4) on some Samsung mobile phones) or any other file system (see above), documented or not.
- "Lower level" is physical: this is the (electronic) block organization. By designing, each block has a limited read/write cycle lifetime. As a consequence, frequently used data tend to speed up their storage block aging. This led Flash manufacturers to imagine block-supervision features. Block supervision often means bad block management and wear leveling: when a block is more used than others, the Flash controller flags it as "used" and moves data to another block to force the used block "to have a rest"; when other blocks have filled their "age delay", the used block is flagged back to "normal", which means data may be stored there again. When a block is definitely overused, it is flagged as "bad" and it cannot be

reused anymore, and data is moved to another block to avoid loss.

Briefly, (logical) sector 0 may not always be stored on (physical) block 0. It may be stored on block 123 at time t_1, then on block 456 at time t_2, then again on block 123 at time t_3, then on block 789 at time t_4, and so on, though OS does not see any of these modifications and always address sector 0.

A Flash Translation Layer (FTL) is needed to operate such block shifts without interfering with file system. FTL is manufacturer dependent and may or may not be documented; Samsung XSR (eXtended Sector Remapper) is a well-documented FTL.

Where and How to Find Data?

In Mobile Phone Handset

As hinted in section "Hardware Architecture", useful data may be found in several locations: in mobile phone handset embedded Flash memory, in additional Flash memory removable card and in SIM card (when applicable).

In SIM Card. GSM standard defines hundreds of specifications for the SIM card. Shortly said, data is stored on SIM card according to files ("elementary file" – EF) and directories ("dedicated file" – DF). Each of these has read/write access rights[i]. Almost all investigation-interesting data on SIM card is read-protected through a code:

- PIN (personal identifying number) may be considered first-level code; after three failed PIN attempts, SIM card is locked, and PIN unlocking key (PUK) is needed.
  (GSM standards allows for several PIN codes (e.g.: PIN1 and PIN2), though "additional" PIN codes (e.g., PIN2) are seldom used.)
- PUK may be considered "master code"; after 10 failed PUK attempts, SIM card is totally "frozen". (As for PIN, GSM standard allows for several PUK codes (e.g, PUK1 and PUK2), directly linked to the PIN codes (e.g., PIN1 and PIN2); since "additional" PIN codes (e.g., PIN2) are rarely used, so are "additional" PUK codes (e.g., PUK2).)

Critical data, such as cryptographic keys or International Mobile Equipment Identity (IMSI) number cannot be read or modified without additional privileges: an administrator code is required.

Without knowledge of any code, it is still possible to retrieve data, though Integrated Circuit Card Identifier ICCID might be the only piece of data really interesting (printed ICCID may also be read on SIM card plastic media, though some telecommunication operators will print a number slightly different from ICCID[j]). ICCID is standardized and gives indication about mobile phone operator. Depending on local jurisdiction, mobile phone operator may be required to release PUK code to investigators, allowing further SIM card analysis.

Among standardized files, the most interesting ones are the following:

- address book: name, phone number, and additional info (such as email) in "extensions"
- call registry
- SMS messages: sent SMS and received SMS. Erased SMS may often be retrieved, since they are most of the time only "flagged as erased". Many types of SMS messages exist, including acknowledgement of receipt, multisegmented SMS, Enhanced Messaging Service (EMS), etc.
- Geographical information, such as the last location area code (LAC) used, forbidden Public Land Mobile Network (PLMN), last broadcast control Channel BCCH used. Note, though, that some mobile network operators have SIM cards with prestored PLMN (in France in 2004, some Orange SIM cards had a Hungarian network and a Thaï network as prewritten Forbidden PLMN).
- According to GSM standard, a UMTS SIM card (USIM card) is able to store MMS messages; at the time of writing, no real case had been found with MMS on SIM card.

Nonstandard files may also be retrieved, using the "SELECT" command and browsing through the entire hierarchical SIM card structure on a "brute-force" process. Totally mobile-network-operator dependent, such nonstandard files may be difficult to decode or to interpret; note, though, that technically aware users may hide personal data in such nonstandard zones.[k] SIMbrush[l] application claims to retrieve all files on a SIM card, including such nonstandard files.

Mass-storage SIM cards (see section "Hardware Architecture") shall follow a two-step analysis: one

using ISO7816 interface and GSM standard, one using MMC/SD protocol and computer forensics tools (*see* **Computers**).

"MultiSIM" cards remain difficult to identify and may present a risk for investigators: typing in PIN A allows access to SIM card A, giving investigators the false impression "there is nothing else to look at"; actually typing in PIN B allows access to SIM card B, typing in PIN C allows access to SIM card C, and so on.

In Additional Flash Memory Removable Card. Additional Flash memory removable cards should be analyzed through standard computer forensics process (*see* **Computers**). When password protected, MMC/SD cards might be unlocked:

- On Symbian S60 third edition and Nokia ESeries phones, using remote lock code: remote lock option locks the phone remotely with a password by sending a dedicated SMS from a third-party mobile phone, as well as it locks memory card with same password.
- On older Nokia phones, using password stored in MMCstore file on the mobile phone handset used along with memory card; such a file may be accessed through FExplorer application (see above), which may not be considered forensically sound depending on local jurisdiction. Another option is to spy data sent by the phone to the memory card using a MMC/SD bus analyzer (password is sent clear).
- Depending on the make and model, more advanced techniques might be used, such as clock attacks.

In/On Mobile Phone Handset. International Mobile Equipment Identity (IMEI) is handset serial number. Unless scratched or erased, it can be read on a sticker at the handset rear, and it can also be retrieved from user interface by typing a dedicated command (most of the time *#06# – note some handsets such as some Motorolas use different commands). IMEI may bring useful information if handset is supposed to have been stolen, and it can be used to get the list of SIM cards that were once inserted in the phone (see "IMEI/IMSI association" in section "In Cellular Network Core (telecommunication operator premises)".

Data found in handset itself ranges from address book to call registry, SMS, MMS, pictures, videos, etc. (*see* **Image Processing and Analysis**). New applications and services available on the most recent handsets, such as Internet browsing, emailing, push-to-talk (GSM handsets are used both ends as walkie-talkies), instant messaging (e.g., Yahoo Go Mobile for Symbian), peer-to-peer (P2P) applications,[m] Voice over Internet Protocol/Telephone over Internet Protocol (VoIP/ToIP) on 3G networks (e.g., Skype for Windows Mobile), etc., may quickly broaden the range of data to be retrieved.

Classical methods to retrieve such data are connecting the phone to a computer using data cable, IrDA interface, or Bluetooth interface, and launching a synchronization application. Dozens of such tools exist: Micro Systemation. XRY, Forensic Science Service CellDek, CelleBrite UME-36 Pro Forensics, MobilEdit Forensics, Oxygen Phone Manager, DataLifter Cell Phone Analyzer, Paraben Device Seizure, Guidance Software Neutrino, Wellphone, etc. Most of them are commercial tools, with the notable exception of open-source TULP2G,[n] originally developed by the Netherlands Forensic Institute.

All these solutions require powering on handset and booting it up, which may either not be possible if handset is broken, or which may not be considered forensically sound in some jurisdictions. As explained in section "Software Architecture", bad block management and wear-leveling may rearrange physical blocks anytime when phone is on, thus modifying evidence and raising the risk of overwriting erased data. In addition, some phones have event logs: Samsung A800, for example, creates a new logfile each time the OS is booted up and turned off.

When connecting a computer to the phone, Bluetooth ID of computer may appear in phone Bluetooth device history, thus modifying evidence.

Another limitation of synchronization solutions is the need for a SIM card inserted in the phone (at least in a GSM environment): in most cases, application cannot access phone content if SIM card is not inserted and SIM card PIN/PUK typed in; reprogramming a "blank" SIM card with same IMSI and ICCID is a way around, provided one knows last IMSI used (which may not be always possible, especially when a foreigner's handset is seized: foreign telecommunication operators may not be forced to release their International Mobile Equipment Identity/International Mobile Equipment Identity (IMEI/IMSI) association logs). The latter prevents phone connecting to the network, which

would modify telecommunication operator logs and possibly modify evidence (e.g., if voice calls or SMS are received while analyzing phone); using a Faraday cage or a network jammer (depending on local jurisdiction) also prevents the phone connecting to the network.

When a handset security code is set up by user, or when multimedia directory is password-protected, synchronization solutions generally fails to retrieve data.

At the time of writing, no synchronization solution could claim fully supporting 100% of the past and present existing phones. In some cases, it is still necessary to perform a manual analysis, browsing phone menus through normal user interface.

Lastly, most of these synchronization solutions only allow "active" data retrieval (i.e., data that can still be browsed and seen through normal user interface). As for computer forensics, deleted data are generally not really deleted: they still remain physically present for some time (depending on the handset use and further data overwriting "deleted" data).

More advanced methods such as the following allow circumventing many of the limitations presented above:

- File system "explorers" (such as FExplorer[o] for Symbian 60 Series phones, totalcmd[p] with various phone file systems plugins, etc.) may allow to browse phone file system freely, even when handset security code or multimedia directory password is set up. Nevertheless, such applications are often closed-source and not commercial products, so they may not be validated in some jurisdictions.
- "Dump software" and "Flashboxes" (such as RedBox, Vygis, Alibababox, Tornado, etc.) are tools originally designed by hackers to flash handset firmware or modify handset IMEI. They also provide unlocking codes, display various codes set up by user, and allow dumping (copy) embedded Flash memory. Once dumped, Flash memory can be analyzed with an hexadecimal editor or various forensic tools: that way, it may possible to retrieve erased data, as well as IMSI lists. They are often closed-source and are not industrial products, so they may not be validated in some jurisdictions. SIM card does not have to be inserted.
- Test and maintenance Joint Test Action Group (JTAG)[q] interface may be used on some mobile phone handsets to access and dump embedded Flash memory, through specific Boundary-scan description language (BSDL) commands. A JTAG interface card has to be installed on the forensic officer computer, and a JTAG interface cable has to be connected between the computer and the mobile phone Printed Circuit Board (PCB) (provided mobile phone PCB has JTAG plots).
- Dismantling handset, desoldering Flash memory, mounting it on a component reader, dumping it, and analyzing raw hexadecimal result (see above) are probably the ultimate solution. It does not require the handset to be fully functional and it does not have the handset powered on; submerged phones, crushed phones, or burnt phones might be analyzed following this method. This may be the most forensically sound solution, but also the most expansive one and the most difficult one! Solder is either Thin small-outline packages (TSOP) or Ball grid array (BGA): in the latter case, a dedicated desoldering station is needed (either hot-air, infrared, or laser). If desoldering is not processed properly and carefully by specifically trained staff, risk of breaking or over-heating Flash memory is nonzero (thus corrupting or erasing data), and there is no guaranty that the mobile phone will function properly after resoldering.

In Cellular Network Core (Telecommunication Operator Premises)

Depending on local jurisdiction, telecommunication (*see* **Analysis: Computer Network**) network operator may be forced (by way of subpoena or court order) to disclose various logs and information they keep for technical as well as financial and legal reasons. Content of such logs and duration of preservation may vary, but they generally include the following:

- subscriber identification (identity, bank details, and subscribed options);
- telephone number portability (i.e., when subscriber has changed operator while keeping his previous phone number);
- PUK code (when SIM card PIN is unknown/undisclosed);

- extended call registry (full dialling and dialled numbers, calls timestamping and duration, and used Cell ID, IMEI/IMSI associations, etc.);
- WAP/Internet gateway logs generated at the Gateway Mobile Switching Center (GMSC) level.

Lawful interceptions are generally conducted with the assistance of the suspect's telecommunication operator, at the Mobile Switching Center (MSC) level.

In addition, some telecommunication operators provide address book and SMS directories (web)storage services to their customers.

In Mobile Phone Handset User Computer

Since many mobile phone users transfer data from mobile phone to computer, one should not forget to look for mobile-phone-related data on the suspect's computer. Dedicated software may be used to operate such transfers, so one should focus on directories created and used by such applications. Looking for phone make/model like directories may also help, as well as investigating Bluetooth-related directories (Bluetooth is often used to transfer data from phone to computer).

On Windows systems, investigating registry may help find previously computer-connected Bluetooth devices (and mobile phones in particular).

Last, analyzing potential mobile-phone-generated JPEG pictures for EXIF metadata or 3GP video files for mp4 metadata may be worth a try (*see* **Image Processing and Analysis**).

Computer forensics is not the scope of this article. Further details may be found in dedicated books (*see* **Computers**).

Other Cell-phone-Related Forensics

Other Devices with GSM Capabilities[r]

Personal digital assistants (PDAs) with GSM capabilities are sometimes called *smartphones* and categorized in a different family from "plain" mobile phone handsets.

Some laptop computers embed a Global System for Mobile/General Packet Radio Service (GSM/GPRS) module (such as Siemens MC75) along with a SIM card slot.[s] With a bit of luck, this might be a way to identify and locate a suspect stealing a laptop and using his own SIM card instead of the victim's SIM card. Laptop computers may also connect to cellular networks using a GSM/GPRS PC Card.

Last but not least, computers may access the Internet through a cell phone used as a Bluetooth or IrDA modem.

Mobile Handset Hacking

Virtually any OS or any of its modules or installed applications may one day present vulnerability, so virtually any mobile handset may once be "hacked" (*see* **Analysis: Computer Network**).

This is especially true about mobile handsets with Bluetooth capabilities. Several viruses and worms for mobile phones, either "proof of concept" or "real and wild", surfaced in the past (e.g., Cabir, "the first virus for mobile phones", could spread from phone to phone *via* Bluetooth), so that the antivirus industry now boasts dozens of antivirus applications for mobile handsets. Similar malware spreading through IrDA file sharing were also announced (e.g., Liberty and Phage for PalmOS). As of Comwarrior, it was supposed to be the first mobile phone virus spreading through MMS service. Last but not least, Mosquitos virus was described as sending premium-rate SMS (see also section "Various Other Frauds").

Bluetooth hacking, using various attacks such as BlueSnarf, BlueBug, BlueSmack, BlueSpoof, etc., may allow either illegitimate access to user data, launching AT commands on attacked device, and sending calls or SMS[t], spoofing a Bluetooth handset identifier (see also section "Spoofing") in order to connect to other Bluetooth devices in the shoes of spoofed device, or provoking a Bluetooth deny of service on attacked device. Similarly, WiFi enabled mobile handsets may be subject to WiFi attacks; it is worth noticing that various firewall applications for mobile handsets are now commercially available.

Such attacks may soon bring the famous "Trojan horse defence/Virus made me do it defence" from the computer world to the mobile phone world.

Other types of mobile phone hacking have also been reported: they all intend at least to crash the phone remotely, at best to take full control over the phone remotely.

Historically, the first published attack may be the "SMS crash attack" by Job de Haas from ITSX

company;[u] since then, Job de Haas has conducted much work on the field of mobile phone security.[v]

Wilfried Hafner from SecurStar company claimed at the Systems 2006 IT Security Area show in Munich (Germany) that he could send a "Rexspy Trojan" to a GSM phone,[w] using both SIM Toolkit Application Programming Interface (API) and Over The Air (OTA) protocol; SIM Toolkit API allows sending commands to the GSM handset, while OTA protocol allows telecommunication operators to send "user-invisible" SMS updating either SIM card or handset firmware. The "Rexspy Trojan" was supposed to let the hacker eavesdrop any voice and SMS victim's communication, as well as downloading a copy of the victim's address book. Though this undocumented demonstration raised doubts and critics, it definitely opened perspectives in the GSM security field.

Strange as it appears, "spyphones" are officially designed and sold in order to spy on the user's "physical" audio environment, as a hidden microphone would do, or to eavesdrop his communications. Hardware solutions include handsets such as Siemens S55 Premium Spy, Siemens C55 FancySpy, and Siemens M50 EconomySpy, whereas software solutions (such as Call Magic, FlexiSpy,[x] or Wireflex Soft products[y]) require a specific application being (secretly) installed on the victim's handset. Legitimate use of such solutions may range from child monitoring to phone bill checking.

Spoofing

Spoofing is the act of using a fake identifier on a network (IP address, Media Access Control (MAC) address, telephone number, etc.) (*see* **Analysis: Computer Network**).

SMS spoofing is the act of sending an SMS with a fake sender telephone number (either existing or not). Many software solutions and websites offer such services, either paid or free. Privacy protection might be a legitimate use for SMS spoofing.

IMEI spoofing is the act of permanently changing IMEI on a (stolen) mobile phone. European GSM operators have set up the Central Equipment Identify Register (CEIR) database, registering all stolen mobile phones in order to ban them on the participating European GSM networks. Whereas GSM Association recommends that IMEI should not be modified in any way by any means (it should not be modified, overwritten or shunted to another IMEI), most of the mobile phones are still vulnerable to IMEI modification, especially through the use of "Flash-boxes" (see section "In/On Mobile Phone Handset").

Caller ID spoofing is possible at least through two methods: Bluetooth hacking (see section "Mobile Handset Hacking") and GSM gateways (also called *SIM Boxes*, *Mob boxes*, etc.). GSM gateways interface cellular networks and PSTN networks and can also take advantage of (Vo)IP networks, in order to reduce costs of international and/or fixed mobile voice calls; these gateways are also used as "anonymizing proxies" (as Internet anonymizing proxies do for computers IP address when connecting to a website), either for privacy protection or for marketing purposes. GSM gateways may be illegal in some jurisdictions.

Various Other Frauds

GSM voice mailbox access and settings may be PIN-protected, but default PIN may be easy to guess (e.g., 0000 or the last four telephone number digits). Hence GSM voice mailbox "hacking" may be made easier for fraudsters. Depending on the telecommunication operator security rules, using a victim's voice mailbox PIN may allow the "phreaker" to modify the welcoming voice mailbox, as well as calling any telephone number in the world at the victim's expense (when a voice message is left, some telecommunication operators allow the user to call back a telephone number of his choice, possibly different from the caller telephone number).

Premium-rate fraud remains as popular in the mobile phone world as it has always been in the fixed world. Most famous frauds consist in sending a "luring SMS" to the victim, leading him (e.g.,) to answer by SMS, or in having his phone ring only once, leading him to call back. Premium-rate fraud may also be based on some mobile phone virus (see section "Mobile Handset Hacking").

SMS and MMS services are not free from spam. Dedicated software can easily be found on the Internet to send thousands of SMS/MMS at low costs. Telecommunication operators' web gateways generally limit the number of SMS one can send through their website.

Geographical Positioning

Geographical positioning (*see* **Analysis: Computer Network**) may either be conducted on a cooperating

mode (target GSM user knows his geographical position is tracked) or on a noncooperating mode (e.g., police trying to locate a suspect). Geographical positioning may be tracked real time (e.g., when user wants to find a restaurant, when emergency services try to find a missing person, etc.) or afterward (e.g., investigators trying to locate the place where a mobile phone call was received in a murder case, based on the Cell ID obtained in detailed call logs from the telecommunication operator) (see section "In Cellular Network Core (telecommunication operator premises").

Real-time geographical positioning is either based on Cell ID (but this may lack precision), on distance triangulation (based on GSM signaling parameters, such as timing advance or power level), or on goniometry. IMSI catchers (also called *IMSI grabbers*) may be used to locate a subscriber independently from the telecommunication operator.[z]

Afterwards geographical positioning is generally used by investigators to determine whether the suspect was really where he pretends to be when he was having a phone call. It may be based on radio coverage maps provided by the telecommunication operator, but the latter are often not precise enough (they may be estimated with simulation software without any real field checking) (Figures 4 and 5). In such cases, cell site forensics with mobile test tools (such as Sagem OT490 or GSM receiver), GPS device, and cartographic software are required.

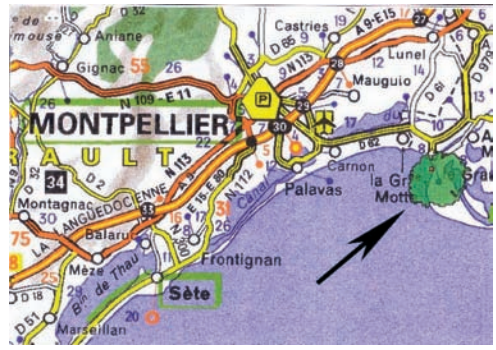


Figure 5 Example of lack of precision (Montpellier and Sete are 31 km away from each other)

End Notes

[a.] "Next" radio base station may not be the immediate geographical neighbor of the previous radio base station.

[b.] Autonomous System: a part of the Internet being administered by an Internet Service Provider (ISP)

[c.] Terrestrial Trunk Radio/Trans European Trunk Radio Access

[d.] During the writing of this article, Cornice company ceased its activity.

[e.] www.s60tips.com, www.allaboutsymbian.com, www.motorolafans.com

[f.] http://sim-max.biz (may be illegal in some jurisdictions)

[g.] http://www.tunisiana.com/publish/viewarticle.jsp?ID = 247

[h.] Logical Block Addressing (LBA) and P-List/G-List remappers on hard-disk drives raise an exception on this statement. But capabilities of these features still remain far from those of Flash memories, as described in this paragraph.

[i.] For concise information on SIM card, see www.simagine.axalto.com/summer2006pdf/SIM%20Overview.pdf

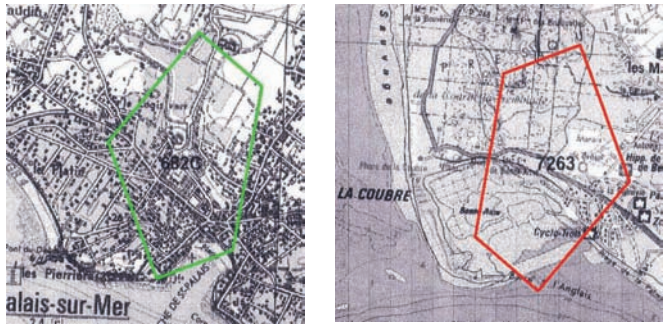


Figure 4 Two different GSM cells displaying same radio coverage shape, obviously simulated

[j.] In France, last digits of the printed SIM Serial Number may be different from last digits of the chip-embedded ICCID.
[k.] *Data Hiding in SIM/USIM cards: a steganographic approach*, by Savoldi and Gubian, in Second International Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE'07), pp. 86–100, 2007.
[l.] *SIMbrush: an open-source tool for GSM and UMTS forensics analysis*, by Casadei, Savoldi and Gubian, in First International Workshop on Systematic Approaches to Digital Forensic Engineering, 2005. Tool can be downloaded at http://asterix.ing.unibs.it:8080/site001/research/simbrush.tar.gz
[m.] E.g. MobileMule and µTorrent mUI are "only" remote controls of computer P2P clients for mobile handsets, while ipra*cool by Ipracom seems to be a "real" P2P file-sharing application.
[n.] http://tulp2g.sourceforge.net
[o.] www.gosymbian.com
[p.] www.totalcmd.net/directory/fsplugin.html
[q.] Joint Test Action Group; for more information on JTAG, see

– *Forensic Data Recovery from Flash Memory*, by Breeuwsma, de Jongh, Klaver, van der Knjiff and Roeloffs, in Small Scale Digital Device Forensic Journal, vol.1 number 1, 2007.
Forensic imaging of embedded systems using JTAG (boundary scan), by Breeuwsma, in Digital Investigation, vol.3 issue 1, March 2006, pp. 32–42.

[r.] This paragraph does not intend to give a full list of all existing devices with GSM capabilities.
[s.] For example, Flybook, Gobook and Hummer Itronix series, Sony Ericsson GC79, Sony Vaio VGN-T340P, Eurocom M190S, Packard Bell Easy Note H5310 Phone Pack ...
[t.] Such calls and SMS will be billed to the user of attacked device. See also sections "Spoofing and Various Other Frauds", as well as Adam Laurie's, Martin Herfurt's and Marcel Holtmann's various works, such as

– *Hacking Bluetooth enabled mobile phones and beyond – Full Disclosure*, by Laurie, Herfurt and Holtmann, 21st Chaos Communication Congress in Berlin (Germany), 2004 (available on http://trifinite.org/Downloads/21c3_Bluetooth_Hacking.pdf)
– *Bluetooth*, by Adam Laurie, in Digital Investigation, vol.3 issue 1, March 2006, pp. 17–19.
www.thebunker.net/resources/bluetooth

[u.] *Mobile Security: SMS and WAP*, by Job de Haas (ITSX), presented at Black Hat Briefings'01, Amsterdam (The Netherlands), 2001, available on http://www.itsx.com/hal2001/hal2001-itsx.ppt
[v.] See, for example, *Symbian Phone Security*, by Job de Haas (ITSX), presented at Black Hat Europe 2005 in Amsterdam (The Netherlands), available on www.blackhat.com/presentations/bh-europe-05/BH_EU_05-deHaas.pdf
[w.] At the time of writing, a video of this demonstration in German was available at www.it-sa.de/itsa_aspx.php?file = RO_Mi_16_30_Hafner& year = 2006
[x.] www.flexispy.com
[y.] At the time of writing, "spysms" and "spyphone" were available at www.spy-sms.com
[z.] Such equipment may have restricted use in some jurisdictions.

NICOLAS DUVINAGE

Guardianships of Adults

In the United States, all adults are considered legally capable of making and carrying out decisions regarding their personal and financial affairs unless a civil court of law, usually a Probate Court, determines otherwise. In those instances, after reviewing evidence, a judge appoints a person or agency to act as the decision maker for the person with diminished capacity. That person or agency is then termed a *guardian* or a *conservator* depending on the law of the state where the appointment is made. A guardianship or conservatorship can be terminated when the person regains capacity. In reality, however, for most people, the loss of capacity is permanent and the guardianship or conservatorship exists as long as the person lives.

Who are the People in Guardianship?

Some adults with diminished capacity, such as those with developmental disabilities never had the capacity to care for themselves. Other adults lose capacity during adulthood because of a catastrophic accident or illness, mental illness, war injuries, or substance abuse. Most people who are in guardianship or conservatorship are over the age of 65. This group of Americans particularly those over 75 years of age is the most likely group to lose the capacity to take care of themselves and to handle their affairs because of physical and/or mental impairments resulting from chronic conditions associated with aging [1].

Who are the Guardians?

A wide variety of agencies and individuals serve as guardians. Most guardians are family members and indeed, laws of most states reflect a clear preference for family members to be appointed as guardians [2]. When family members cannot serve as a result of incapacity or alleged abusive behavior or if there is no family available, other people such as neighbors, friends, and long-time professional associates such as accountants or attorneys may request the court for appointment as guardian. Or, an agency or private professional guardian may petition the court for appointment.

Private Professional Guardians

In the last three decades in the United States, individuals have emerged who are setting up private practices as guardians. They are termed *private professional guardians*. Most people starting private practice as guardians already have a professional identity that is related, such as social worker, nurse, trust officer, public guardian employee, accountant, or attorney. Others would have served as guardians for family members and found the work rewarding and now want to do it for a living. People entering guardianship practice usually have well-developed people skills or well-developed financial skills. It is the rare individual who has both. In large part, this new role is unregulated except for monitoring by the court, which may or may not be possible because of scant resources. Only a few states have licensing requirements, Florida, Arizona, California, and Washington being among them. In other states, practice is unregulated and anyone may petition the court to serve as a private professional guardian, regardless of skill, knowledge, or background.

Volunteer Guardians

Some states or jurisdictions seeking to provide more guardianship services, especially for people with low incomes and no one else to help them, are turning to volunteers. Typically, volunteers provide services connected with guardianships of persons. The volunteers may be individually appointed as guardians or the sponsoring agency may be appointed. Volunteer guardians cannot function without the support of a paid staff that is knowledgeable about working with volunteers [3]. Ideally, volunteers are treated as if they are staff members. This means there is an application process that includes reference checks and criminal background checks, job descriptions with clear expectations, training and supervision, and evaluations. There also must be a policy and protocol that permits the release or "firing" of volunteers who cannot do the work appropriately.

Agencies as Guardians

Agencies also serve as guardians. The agency may be public, for-profit, or nonprofit. Nonprofit agencies are usually motivated by religious ideals, charitable intent, or the social good. They often serve people who have few, if any assets. Their funding derives from United Way, public funds, foundations, endowments, grants, fees, fundraisers, Medicaid, or funds provided by the parent organization if there is one. For-profit organizations are competitive in the market place and therefore must pay taxes on their profit. They must consider cost effectiveness and financial bottom lines. Many accept only guardianships of estate and provide no personal services [4]. Others contract with professional case managers who then provide the personal care the person in guardianship needs.

Every state has a form of public guardianship. Public guardians serve when no one else will or can. They are often the first choice when family members are in conflict and cannot agree on anything or if the person with diminished capacity is physically dangerous to his or her family but not to anyone else [5]. Public guardian agencies are typically under funded

and frequently expected by county government to support themselves with fees from the estates of the people in guardianship, often a near impossibility.

History of Guardianships in the United States

Ancient Greek and Roman documents mention what could be considered the earliest form of guardianships. Mental disabilities or what we would now term *psychiatric disorders* were the focus. Concerns centered on the assets of the person and protecting them. Protection of the person was presumed to be the responsibility of the family [6]. A similar outlook prevailed in England until the fourteenth century when the statute, "De Praerogativa Regis" (the royal prerogative) was enacted. This law recognized the duty of the King to protect his subjects who could not protect themselves [7]. The doctrine *parens patriae* is based on this law. Implementation of the law resulted in formal proceedings for both "idiots" (those who never had capacity) and "lunatics" (those who had lost capacity) during which the Lord Chancellor held inquisitions to determine if a committee should be appointed for the protection of the person and property of the adult [6]. This process was available only to people who had assets because it was expensive. The main focus was on protecting property [7].

The doctrine of *parens patriae* prevailed in Colonial America and later, the United States. Protection of the person as well as the property began to emerge as a focus. Even so, protection of property remained the stronger focus. Gradually guardianship statutes and procedures were enacted in every state providing a variety of protections and due process provisions depending on the state. Very little attention was paid to individual rights although the courts, in keeping with state law, granted guardians sweeping powers over the person and estate of the individual. The concern was to protect the impaired person from harm, not to empower the person. This emphasis began changing in the 1960s, most likely because of the sweeping civil rights movements in the United States that focused on the rights of minorities, women, those with disabilities, and elders. There is now recognition that guardianship of the person is at least as important as guardianship of the estate. Various legal groups including the American Bar Association's Commission on Law and Aging have published scholarly articles and promulgated national conferences where like-minded reformers met and developed strategies. Newspapers still frequently run articles that expose abuses in guardianships. Gradually, state laws began focusing on due process rights of the person who is the subject of a guardianship as well as closer monitoring of guardianships by courts. In addition, the concept of global incompetency has lost currency. The awareness grew (and is growing) that people who have lost capacity in one area of functioning have other areas of functioning that are not impaired and those areas should be respected and promoted. The focus now is to recognize the still existing capabilities and tailor make the guardianship to fit the individual. For instance, many people in guardianship have the ability to make a will, marry, vote, or handle a small bank account. Implementation of the changes in state laws has been uneven for many reasons: lack of financial resources for courts, stereotypes about the disabled and elderly, and inertia.

Changing Demographics

The United States, like many other countries in the world, is experiencing an unprecedented growth in the number of people over age 65. There are greater number of older people than ever before and they are living longer as a result of medical advances and healthier life styles. In 2005, people over 65 represented 12.4% of the US population. That group had increased by 9.4% since 1995. By 2030, the number of people over 65 will account for roughly 20% of the US population. Of those over 65, the group over 85 is growing the fastest and will increase by 20% between 2000 and 2010 [8]; [9]. This is the group that anecdotally is most susceptible to guardianships because of one or more chronic illnesses that have worsened and have had an impact on cognition and the ability to carry out activities of daily living.

The acknowledgment and acceptance of the existence of elder abuse and neglect, beginning in the late 1970s is also having an impact on guardianships (*see* **Elder Abuse: Policy**; **Elder Abuse: Risk**). Many guardianships are sought as a way to address an abuse or neglect situation by removing the alleged abuser or neglector from decision making for the impaired adult and to substitute a responsible person or agency. According to the best available estimates,

between 1 and 2 million Americans, age 65 or older have been injured, exploited, or otherwise mistreated by someone, usually a family member, on whom they depend for care or protection. This figure does not include those who are victimized by strangers or those who self-neglect [10]. The number of cases of elder mistreatment will most likely grow as the elder population increases, which in turn will also result in more guardianships.

Alternatives to Guardianship

Guardianship can result in the removal of constitutional rights and can be complex, time consuming, and costly. It can also result in a loss of privacy because the guardianship court file may be public in part or whole. In reality, few people ever look at a guardianship file but the potential remains. For the above reasons and because the concept of the least restrictive alternative is now imbedded in guardianship law in most states, alternatives to guardianship are important to consider. Alternatives are "various legal tools, social services, and government programs that may delay or prevent the appointment of a guardian for a person who is not capable of making decisions on his or her own behalf" [11].

None of the alternatives is exactly equal to guardianship. Each alternative has a feature(s) of guardianship and, if tailored individually and appropriately to the individual, may delay guardianship or eliminate the need for it completely. Each has advantages and disadvantages just as guardianships do. One alternative may be appropriate at one stage of an illness and another more appropriate at a later time, especially with physical and cognitive impairments caused by progressively debilitating diseases. Appointment of a guardian does not necessarily represent a failure of the other alternatives. It may mean that the alternatives were not suitable or available or there was no one to serve as the surrogate decision maker [1].

The main alternatives for the health care of the individual are healthcare directives including durable powers of attorney for healthcare and living wills, healthcare or family consent laws, single court actions, and measures for alone elders. Alternatives for financial decision making include specific and general powers of attorney, durable powers of attorney for finances, trusts, and joint ownership of property, representative payee, and single court actions.

Alternatives for Healthcare

Healthcare Directives. These are alternatives that are planned by the person who has capacity in case he has decisional incapacity in the future. They consist of durable powers of attorney for healthcare and living wills. Some states combine the two concepts into one document called *Healthcare Advance Directives*. Either way, it is important to keep the document(s) current.

Living wills are documents that give directions to physicians that tell them what to do if the person who executed the document is in a terminal condition, or, in some states, a persistent vegetative state. Most states have enacted living wills but the provisions differ from state to state. The disadvantages of the living will are that often it cannot be located when and where it is needed. Moreover, it is simply a piece of paper unless someone actively advocates for the stated wishes of the person (Wood, E. (2004). Personal Communication).

A durable power of attorney for healthcare is a written, witnessed document between two people that authorizes one person to act or represent the other in healthcare decision making. The person who delegates the authority is termed the *principal* and the person accepting the authority is termed the *agent*. The person who grants the power must be fully capable of making decision in his or her own best interest at the time the document is signed or, it is invalid. A durable power of attorney usually takes effect once the person has lost decision-making capacity, which is usually determined by a medical doctor. The most important part of preparing this document is selecting the person who will be the decision maker. It should be someone familiar with the values of the person executing the document.

This alternative is inexpensive since the document can be found in stationary stores. Durable powers of attorney for healthcare can exist along with guardianships and in fact, provide guidance to the court as to what the now incapacitated person would want. Disadvantages include lack of supervision of the agent and the possible invalidity from one state to another. The document is frequently not available at the exact time and place when it is needed. Durable powers of attorney for healthcare do not cover giving consent

for living arrangements either, a common concern with the care of elders.

Healthcare or Family Consent Laws. Over 35 states and the District of Columbia have enacted statutory authorization for healthcare consent by family members (or others who know the patient) when there is no prior written document by the incapacitated person [12]. The decision maker acts in accordance with the statute. Most statutes provide for a hierarchy of decision makers, beginning with a guardian if there is one, and then family members in order of relationship. Friends are usually at the end of the list after family members. Physicians are not necessarily aware of these laws, however.

Single Court Transactions. Many state laws provide for single transactions that involve court action. An attorney for a healthcare facility, usually an acute care facility, brings the petition to the court. Depending on state law, another attorney is appointed to represent the patient. The judge then considers the evidence from both attorneys, including the statements of the person who may need the medical care. Single court transactions may also be allowed for moving a person from one setting to another.

Measures for Alone Elders. Medical decision making for older people who do not have capacity and have no one to make decisions for them has become a focus for policy makers [13]. The term *unbefriended elders* has been coined to describe these people, who may have been societal loners and who have left few traces of how they lived their lives or what their medical preferences might be. Some had close relationships and family but they would have died or become incapacitated themselves. Emergency situations and routine care are generally covered by common law and thus medical practitioners can carry them out without legal interventions. Usually the issue of decision-making for alone elders comes up in hospitals or nursing homes. The decisions to be made include end-of-life decisions, surgery, the administration of psychotropic medication, placement of feeding tubes and use of antibiotics, ventilator support, and transfers to acute facilities from nursing homes [13]. Existing measures to address the issue of decision making are scant and not uniform. Some states designate physicians, sometimes in combination with ethics committees, to make the decision. Three states have unique external committees of trained volunteers to make the decision. Some states have established public guardianship programs to serve this vulnerable population. In some instances, those professionals involved may decide to use the single court transaction option.

Existing institutional practices include ethics committees (which are usually advisory only and may not actually make the final decision) and other established procedures. Sometimes, by default, physicians serve as ad hoc guardians and do what they think is best within the ethical dictates of their profession [13].

Financial Decision-Making Alternatives

There are several alternatives to managing the financial affairs of someone who has lost decision-making capacity.

Powers of Attorney. A power of attorney is a written witnessed document between two people that authorizes one person to act for or represent the other. The person granting the power must be fully capable of making decisions in his or her own best interest at the time the document is executed. There are three types of powers of attorney for handling finances: limited or specific, general, and durable.

The "limited or specific" power of attorney is for one situation only. For instance, a person traveling to another country may delegate his or her real estate agent to sign refinancing documents while he or she is away. Or, a person with severe arthritis may delegate responsibility for handling a specific bank account to someone else so that routine bills can get paid.

A "general power" of attorney usually gives the agent the right to handle all the financial affairs of the person granting it. Both the specific power of attorney and the general power of attorney become invalid when the principal loses decisional capacity. In practice, however, many people, unaware that it is no longer legal, continue to use the power of attorney long after the principal begins manifesting diminished capacity [14].

"Durable" powers of attorney for finances designate someone to manage all assets and financial transactions in case of future decisional incapacity. It is an attempt to avoid guardianship.

Powers of attorney have the advantage of being private. They can be drawn up and signed without an attorney, which means they are cost efficient. Forms

are readily available in stationary stores. The powers to be delegated can be tailored to the individual's wishes. There are also disadvantages. Powers of attorney can be easily abused because the agent is not supervised by anyone. There is no bonding, which means that if the agent mishandles the assets or uses them for himself or herself, there is no "insurance" mechanism to reimburse the principal.

Trusts

Trusts are written legal agreements that provide for property management, which in some cases include provisions for distribution of assets after the person has died. State and federal laws govern trusts. The person who establishes a trust is termed the *trustor*, grantor, or settler. The person or institution that manages the trust is termed the *trustee*. The person who receives funds from the trust is termed the *beneficiary*. The trust document sets forth the exact nature of the trust, when it takes effect, its revocability, and its funding guidelines, as well as stating directions as to the care and disposal of the assets [15]. A guardianship may supplement the trust in order to manage properties that are not in the trust or to address the medical and other care decisions that may be needed.

There are three major types of trusts: testamentary, special needs and revocable living.

A "testamentary" trust takes effect on the death of the trustor. It is provided for in the will and dictates the distribution process. It is used when a testator does not want to make outright distribution to a beneficiary or beneficiaries after his or her death. They can be used to provide income to a spouse or partner with the remaining assets going to the rest of the family on the death of the spouse or partner. A testamentary trust is a tax savings plan.

"Special needs" trusts are irrevocable and are established for the sole benefit of a disabled adult, usually one under the age of 65. The usual special needs trust is pursuant to federal law created in 1993 which allows a disabled person to receive public benefits while at the same time trust assets are dispersed under strict definitions. Repayment to the state is required if there are assets left in the trust after death [16].

"Revocable Living" trusts are the most common types of trusts. This trust takes effect during a person's lifetime and may help avoid guardianship. Most people name themselves as the trustee initially and, at the same time, name another person or institution to serve as successor trustee should they lose decisional capacity. This type of trust has several advantages: it provides for financial and property management in the event of incapacity, it permits changes at any time by the trustor, it specifies the disposal of the trustor's assets should the incapacity be extended or permanent. The disadvantages include the lack of bonding and hence protection for the trustor when a successor trustee takes over, there is no reduction in taxes although "trust mill" purveyors the opposite, and no one monitors the actions of the successor trustee.

Joint Ownership. Real property and cash assets can be jointly owned. There are many types of joint bank accounts. Most states allow the type where the contents of the bank account(s) go to the person who survives the other regardless of whose money it was in the first place. Still others require both signatures on withdrawals. State laws and local bank practices dictate which options are available. Advantages of this alternative include ease of execution and assistance to those who cannot get to the bank. Disadvantages include loss of control of the assets or property, exposure to each other's creditor claims and tax consequences. Also, there is no bonding if one person takes all the money.

Representative Payee. Several federal agencies are authorized to appoint a person or institution such as a bank or a nursing home to receive federal benefits on behalf of the recipient. These agencies include the Social Security administration, Veteran's Administration, Railroad Retirement Board, Department of Defense, and the Office of Personnel Management. The person or institution receiving the check is termed a *representative payee*. Physician statements as well as those by relatives and friends are accepted as evidence. Anyone can serve as a representative payee [17]. Although annual accountings are required, enforcement is not rigorously enforced [1].

Single Court Transactions. Many states have laws that provide for single transactions pertaining to finances. One of the most dramatic is the one that allows courts to freeze bank accounts upon being provided sufficient evidence. This is particularly important when an abuser is draining a bank account. Some states allow or require that an attorney be

appointed to represent the person whose assets are to be frozen.

When is a Guardianship Appropriate?

It is obvious that many of the alternatives to guardianship can be very helpful. The next question is when should a guardianship petition be filed? The following basic criteria must all be met:

1. The guardianship state law is applicable.
2. The person in question has diminished capacity to make decisions.
3. Alternatives to guardianship have been carefully considered but found inadequate to the situation. Or, one of the alternatives has been used to abuse the person with diminished capacity.
4. There is objective evidence demonstrating the lack of adequate decision making on the part of the person with diminished capacity. This can include physician declarations, functional evaluations, and cognitive evaluations, perhaps by a neuropsychologist. In addition, there are several examples demonstrating that the person with diminished capacity cannot provide proper support for food, clothing shelter, medical care, and the handling of financial resources. Evidence may be nonpayment of rent and utility bills, failing to keep medical and dental appointments, marked confusion as to time and place, missing fund, or substance abuse causing medical problems.
5. There is risk of personal or financial injury or damage to the impaired adult that only a guardianship can prevent. A less restrictive alternative will not suffice.
6. Court authority is needed to stop abuse, protect the victim, and seek restitution, or to take some type of legal action, such as: settle a lawsuit, represent the person in an existing divorce proceeding, or file a civil lawsuit for personal injury, fraud, interference with an expected inheritance, false imprisonment, negligence, intentional or reckless infliction of emotional distress, institutional abuse, professional malpractice, or civil rights violations because of disability or age.
7. There is an appropriate person or agency willing to serve as guardian.
8. There are signs that undue influence is being exerted even though the person does not have obvious signs of mental impairment. Such signs include the following:
   (a) The elder signs documents and takes actions that run counter to his or her long-time values and beliefs.
   (b) Someone systematically isolates the elder from people who care about him or her.
   (c) The elder suddenly changes long-time providers such s physicians, dentist, or attorneys.
   (d) The elder suddenly moves into someone else's home under the guise of needing care, or someone suddenly moves into the elder's home.
   (e) The elder places unusual trust in newfound acquaintances when there is a history of mistrust in the family, especially with financial affairs.
   (f) The elder begins writing checks made out to "cash", always in round numbers, and frequently in large amounts.
   (g) Someone makes promises of lifelong care in exchange for the elder's property when he or she dies [14, 1].

Obviously, filing for guardianship is not to be taken lightly. Filing is only the first step in a guardianship. It is the start of the court process, which may or may not lead to the creation of a guardianship for an impaired person.

Court and Guardianships

Guardianships are considered to be matters for the civil justice system. The "guardianship court" may have a different name depending on the state. It could be Probate Court, County Court, District Court, Circuit Court, Court of Common Pleas (e.g., Pennsylvania, USA), Chancery Court, Surrogate Court, or Superior Court [18]. Some states handle guardianships of the estate in the Probate Division and guardianship of the person in the Family Court.

The process of guardianship in the court unfolds in a predictable fashion: petition filed, emergency guardianship (in some cases) court hearing re: the general guardianship, ongoing guardianship, and termination of the guardianship. Each stage of the guardianship process is or should be accompanied by due process protections. The Fourteenth Amendment to the United States Constitution requires that

due process protections be afforded to anyone who is threatened with the loss of liberty or property because of state action. While people in guardianship retain ownership of their assets, they lose control over them. They are not imprisoned but they lose control over where they will live and the freedom to choose the services that will be provided. There needs to be thoughtful procedures that are fundamentally fair, strictly defined and adhered to, and are taken step by step. There needs to be full disclosure of the steps to the person affected and opportunities for him or her to object and to give input with each step. This is called *due process* [19]. Implementation of due process is summarized as follows:

1. The person who is alleged to need a guardian is given written notice of the pending proceeding in a way that maximizes the change so that he or she will understand it.
2. There is sufficient time for the person to respond and prepare a defense, if desired.
3. The person is furnished with the evidence that supports the petition.
4. The person is informed of the right to attend the hearing, to object to the proceedings up to and including a trial by jury, to have someone else act as guardian, and to have an attorney.
5. An attorney is appointed to represent the person if he or she asks for it, if the court thinks it is in the best interests of the person, or if the state law requires it.

Unfortunately, in many jurisdictions, even when state law provides for due process protections, they do not take place evenly because of a lack of resources, because no one thinks they are really necessary as a result of ageism or biases against autonomy for people with diminished capacity, or because family and professionals think they know what is best for the person with diminished capacity [20, 21].

Court Stages of Guardianship

Petition Filed and Hearing. The first formal step in the guardianship process is the filing of the guardianship petition with the court. It is usually filed in the county where the person with diminished capacity lives or has real property. Most states have standard forms. The more complete the petition is, the better the judge can grasp the situation and the more carefully the guardian's powers can be tailored to meet the needs of the person with diminished capacity. At the time the petition is filed with the court, a date will be set for a hearing on the general guardianship. It is usually 6–8 weeks hence. After filing the petition for guardianship, the attorney of record must give notice of the filing including the time and place of the hearing, to a variety of people including the person affected and usually relatives within the first (children and parents) and second degree (grandchildren and siblings).

During the time before the hearing, a court appointed person may go out to interview the person alleged to need a guardianship and investigate the circumstances of the case. The person may or may not be a court employee. A written evaluation may be filed or perhaps an oral report. The labels of such professionals are guardian ad litem, court investigator, and court visitor. The judge takes the reports of these professionals quite seriously because they are neutral parties who do not stand to benefit whether the guardianship is granted or not.

Before the hearing, the judge reviews and considers the information that has been filed re: the proposed guardianship. The hearing is the time when the judge formally listens to comments from the person who is thought to need a guardianship and the proposed guardian. Their attorneys also speak. Anyone else can also speak about the case. Hearings are generally public matters and anyone can attend. Judges take care not to reveal the contents of any confidential documents that have been filed. Usually the judge will give his or her decision at the hearing in open court.

Temporary or Emergency Guardianships. It does happen that that there are situations where urgent action is needed and it would be a disservice to wait for the hearing on the general guardianship. There are two general situations where guardianship powers are needed urgently: the person with diminished capacity is in physical danger or assets are being actively dissipated or wasted. For example,

1. Someone is trying to move the person with diminished capacity to an undisclosed location.
2. Assets are actively being drained from the person's bank accounts and they need to be frozen until further investigation can be made.
3. The person's house has been put up for sale and escrow is closing.

4. The person has an urgent need for medical attention.
5. Court permission is needed to move the person, who is demented, wandering, and getting lost frequently to a secure facility.

When a temporary guardianship is granted, it is reviewed by the court appointees and the judge at the time of the hearing on the general guardianship.

Ongoing Guardianship. There are many activities during the "lifetime" of a guardianship. Official documents must be filed by the guardian including inventories and appraisement of all assets as well as accountings as required by state law if there is a guardianship of estate. Guardians of person must file General Plans in some jurisdictions, which detail what they expect to be doing during the guardianship. They may also be required to file Status Reports annually. While the reports that the guardian files are self-reporting and as such will not probably be incriminating, they are useful in the course of any investigations that are done and they serve to remind the guardian that the court is involved in what he or she does. Forward-looking courts make efforts to educate guardians with handbooks and videotapes. Most courts provide careful monitoring but many courts do not have the resources to do so. Close monitoring, to see that required documents are filed, that they comport with the state law, and that they are serving the person in guardianship, is important to detecting and preventing abuse and neglect.

Termination of Guardianships. Guardianships are terminated because the person in guardianship is restored to a state of full functioning, the person dies or moves to another state, or a trust is established and the guardianship is no longer needed. It does happen that people recover from difficult situations such as closed head injuries. An attorney also may seek to terminate a guardianship of estate if all the assets have been used.

Conclusion

Until recent years, guardianships and guardianship process have not been well understood in the United States. As more of our population ages and lives longer, guardianships will be more common. State laws have by and large been reformed but implementation is not uniform. The growing numbers of older adults and the emergence of elder abuse and neglect make it imperative that all those concerned about elders and disabled people view guardianships in their reality and work to make them effective and respectful for the people in guardianship.

References

[1] Quinn, M.J. (2005). *Guardianships of Adults: Achieving Justice, Autonomy, and Safety*, Springer, New York.

[2] Bayles, F. & McCartney, S. (1987). *Guardians of the Elderly: An Ailing System*, Associated Press.

[3] Teaster, P.B., Schmidt, W.C., Abramson, H. & Almeida, R. (1999). Staff service and volunteer staff service models for public guardianship and "alternatives" services: who is served and with what outcomes? *Journal of Ethics, Law, and Aging* **5**(2), 131–152.

[4] Barnes, A. (2002). The virtues of corporate and professional guardians, *Stetson Law Review* **XXXI**(3), 941–1046.

[5] Teaster, P. (2005). Public guardians, in *Guardianships of Adults: Achieving Justice, Autonomy, and Safety*, M.J. Quinn, ed, Springer, New York, pp. 96–104.

[6] Brakel, S., Parry, J. & Weiner, B. (1986). *The Mentally Disabled and the Law*, 3rd Edition, The University of Chicago Press, Chicago.

[7] Wood, E. (2005). History of guardianship, in *Guardianship of Adults: Achieving Justice, Autonomy, and Safety*, M.J. Quinn, ed, Springer, New York, pp. 17–48.

[8] Administration on Aging (2006). *A Profile of Older Americans: 2006*, Retrieved 10/23/07, from www.aoa.dhhs.gov.

[9] Centers for Disease Control and Prevention and The Merck Company Foundation (2007). *The State of Aging and Health in America 2007: Executive Summary*, The Merck Company Foundation, Whitehouse Station.

[10] Bonnie, R.J. & Wallace, R.B. (2003). *Elder Mistreatment: Abuse, Neglect, and Exploitation in an Aging America*, (eds) National Academies Press, Washington, DC.

[11] Stiegel, L. (1992). *Alternatives to Guardianship: Substantive Training Materials and Module for Professionals Working with the Elderly and Persons with Disabilities*, American Bar Association Commission on Legal Problems of the Elderly, Washington, DC.

[12] Sabatino, C. (2002). *State Health Decisions Legislative Update-2000*, Retrieved February 28, 2004, from http://wwwabanet.org.

[13] Karp, N. & Wood, E. (2003). *Incapacitated and Alone: Health Care Decision-making for the Unbefriended*

Elderly, American Bar Association Commission on Law and Aging, Washington, DC.
[14] Quinn, M.J. & Tomita, S.K. (1997). *Elder Abuse and Neglect: Causes, Diagnosis, and Intervention Strategies*, 2nd Edition, Springer, New York.
[15] Kapp, M. (1992). *Geriatrics and the Law: Patient Rights and Professional Responsibilities*, 2nd Edition, Springer, New York.
[16] Bernstein, R.M. & Bernstein, M.S. (1999). The use of special needs trusts in the settlement of personal injury cases, *Elder's Advisor* **1**(1), 58–61.
[17] Komlos-Hrobsky, P. (1989). *Representative Payee Issues in the Social Security and Supplemental Security Income Programs*, National Clearinghouse for Legal Services.
[18] Benton, F. (1998). The courts, in *Guardianship of the Elderly: Psychiatric and Judicial Aspects*, G.H. Zimny & G.T. Grossberg, eds, Springer, New York, pp. 74–85.
[19] Atkinson, G. (1980). Towards a due process perspective in conservatorship proceedings for the aged, *Journal of Family Law* **18**(4), 819–845.
[20] Sabatino, C. (1996). Competency: refining our legal fictions, in *Older Adults' Decision-making and the Law*, Smyer, M., Warner Schaie, K. & Kapp, M.B., eds, Springer, New York, pp. 1–28.
[21] Sabatino, C.P. & Basinger, S.L. (2000). Competency: refining or removing our legal fictions, *Journal of Mental Health and Aging* **6**(2), 119–144.

Further Reading

Wood, E. (2003). Alternatives to Guardianship. *Presentation for Continuing Legal Education to Virginia Guardians Ad Litem*. Roanoke.

MARY J. QUINN

Gudjonsson Suggestibility Scales

see Interrogative Suggestibility

Gunshot Residues ***see*** Firearm Discharge Residue: Analysis of

Gunshot Wounds

Introduction

Firearm injuries are regarded as a special form of blunt trauma. The damage to the organism is caused by the impact of a single projectile (or a multitude of pellets) propelled from a barrel by high-pressure combustion gases and striking the body at a high velocity. Gunshot wounds, in a broader sense, are also lesions caused by blank-cartridge weapons as well as injuries due to livestock stunners, stud guns used in the construction industry, and similar devices.

Weapons and Ammunition

Weapon Types

Firearms enable the user to hit a target from a distance. Depending on the weapon type (*see* **Firearms: Overview**), a distinction is made between handguns (short-barreled firearms for use with one hand: pistols, revolvers) and long arms (portable long-barreled firearms for use with both hands: rifles, submachine guns, machine guns, and shotguns).

Modern pistols carry a magazine in the handle; after firing a cartridge, its case is ejected as the bolt is sliding back. Revolvers have a rotating cylinder holding the cartridges; after firing the shot, the case remains in the chamber.

Weapons intended to be fired from the shoulder have either a rifled barrel (see below), e.g., hunting rifles (firing single projectiles) and military assault rifles, or they have a smooth barrel such as shotguns (firing lead pellets, shotgun slugs, or rubber/plastic pellets).

Rifled and Smooth Barrels

The inside surface of a rifled barrel has a spiral pattern of elevations (lands) and grooves in between. As the bullet passes through the barrel, the lands cut into the cylindrical part of the projectile, thus causing it to rotate around its longitudinal axis ("twist"). This gyroscopic spin stabilizes the bullet on its trajectory and improves the accuracy of the shot.

Shotguns have smooth-bore barrels without grooves and lands. They are designed for firing shells

containing mostly lead pellets or shotgun slugs. Shotgun shells loaded with rubber or plastic pellets are designed for self-defense or for police riot-control purposes [1]. The cylindrical barrel may be slightly tapered toward the muzzle end (choke bore) in order to keep the discharged shot in a tighter spread over a longer distance.

Caliber

On the one hand, the term *caliber* represents the diameter of the bore, and, on the other hand, the bullet diameter. For metric calibers, the caliber specifications are nominal values usually based on the diameter of the lands (maximum diametrical distance between the lands in the barrel). Most projectiles have a somewhat larger diameter. Example: The caliber 7.65 mm Browning has a land diameter of 7.63 mm, whereas the bullet diameter is 7.85 mm. The Anglo-American caliber specifications (in inches with 1 in. corresponding to 25.4 mm) are based rather on the bullet diameter.

In pistols, the predominant calibers are 6.35 mm, 7.65 mm, 9 mm, and .45 in, whereas in revolvers the main calibers are .32, .357, .38, and .44 in.

Cartridges for .22 caliber rimfire weapons can be fired not only from handguns but also from long rifles (LRs).

In cartridges designed for hunting rifles, it is usual to indicate not only the caliber but also the length of the case in millimeters (e.g., 7×64 mm, 8×57 mm). The same applies to the cartridges of military rifles (e.g., 5.56×45 mm, 7.62×51 mm).

The caliber of shotgun barrels is not identical with its internal diameter. It indicates the number of spherical lead balls collectively making up one English pound (1 lb = 453.6 g) in weight; the diameter of these lead balls corresponds to the internal diameter of the barrel (e.g., 18.2 mm for shot caliber 12 and 16.8 mm for shot caliber 16). An exception to this rule is the .410 shotgun, which is named after the barrel diameter in inches.

Cartridges

The cartridges of handguns and rifles consist of a case with a primer in its base, the propellant (gunpowder) above, and a bullet (projectile).

The case is usually made of brass. Cartridges for autoloading pistols have a groove just above the base so that the fired case can be extracted and ejected by the recoiling slide. Revolver cartridges have a base plate that is wider than the cartridge case.

The explosive compound is located in the base of the cartridge case. When the trigger is pulled, the firing pin strikes the base of the case, causing the primer compound to detonate and initiate the burn of the gunpowder.

The primer consists of an explosive compound sensitive to percussion. Initially, the primer was made of mercury fulminate, which was later replaced by a combination of chemical compounds typically containing the elements lead, barium, and antimony. Therefore, these elements play an important role in the detection of gunshot residues (GSR). On the other hand, nowadays, lead-free primers are also available (e.g., "Sintox" with the main elements zinc and titanium).

Above the primer, the case is filled with gunpowder (propellant**),** which does not detonate, but burns rapidly. "Black powder", which has been known for centuries and is a mixture of 75% potassium nitrate, 15% charcoal, and 10% sulfur, is rarely used as a propellant nowadays (for historical or reproduced muzzle loaders or some blank-cartridge weapons). Because of the high amount of solid combustion residues, a lot of dense smoke is produced [2, 3].

Currently, mostly smokeless powder is used as propellant because of its higher efficiency of combustion compared to that of black powder. It either consists of nitrocellulose (NC) alone ("single-base powder") or of NC dissolved in nitroglycerine ("double-base powder"). The powder grains are of pale green color or have a shiny silver-black appearance (if coated with graphite). The powder particles resemble thin platelets (flake or disk powder), small spheres (ball powder), cylinders, or short tubes (tubular powder). The individual powder particles have a diameter ranging from several tenths of a millimeter up to more than 1 mm. In Flobert cartridges (caliber 6 or 9 mm), the primer also serves as propellant.

The deflagration of the propellant produces an ample amount of gases (CO_2, CO, H_2, N_2, nitrous gases, water vapor), which are under high pressure and accelerate the projectile on its way through the barrel. The processes inside the weapon itself are called *interior ballistics* (*see* **Firearm Examination: Ballistics**).

In handguns and .22 caliber rimfire rifles, the muzzle velocity typically ranges from 250 to

400 ms^{-1}, whereas in hunting and military rifles it is considerably higher (approximately 700–1000 ms^{-1}).

Generally, ammunition designed for autoloading pistols and military rifles has full-jacketed bullets with a lead core and a jacket made of steel or a copper alloy covering the front and lateral parts of the projectile but not its base.

For hunting, most cartridges are equipped with a semijacketed bullet. As the jacket is open at the tip, the projectile with its soft-point or hollow-point lead core deforms more readily, thus transferring more energy to the animal body.

Standard .22 caliber rimfire ammunition usually has unjacketed lead bullets. Revolvers mostly fire lead bullets, but also semijacketed projectiles.

According to the bullet's head configuration, there are different shape categories: round nosed, flat nosed, pointed, cylindrical, and truncated cone.

Apart from the shape and the design of the projectile, its mass and velocity are also of importance for its potential effect. From the mass (m) and the velocity (v), the kinetic energy E_k can be calculated ($E_k = 1/2mv^2$); it is indicated in Joule (J). As examples, we may mention the mass of a .22 LR projectile (2.55 g) and a 9 mm Parabellum projectile (8.0 g); in the first instance (.22 LR), the bullet energy is about 140 J, and, in the second instance (9 mm Para), about 430 J. The bullet energy of military and hunting cartridges is many times higher. Note that the energy increases exponentially with the velocity of the projectile.

Shotgun Ammunition

Conventional shotgun shells are fired from smooth-bore shotgun barrels. Instead of a single projectile, they mostly contain numerous spherical pellets made from hard lead. The pellet diameter of birdshot varies between 2 and 4.5 mm, and the pellet velocity on leaving the barrel is about 300 ms^{-1}. The larger the pellet, the greater is its range, given the same velocity.

In addition to birdshot, cartridges with especially large pellets ("buckshot") are also available, with single shotgun slugs (e.g., Brenneke type, Foster, Smith & Wesson sabot) and with rubber/plastic pellets.

The cartridge cases are made of cardboard or plastic and the floor (mostly consisting of brass) contains the primer. The space above it is filled with a (smokeless) nitro powder, which is separated from the lead shot by intermediate layers (felt or plastic wad). Nowadays, the cardboard and felt wads have been largely replaced by plastic wads in connection with shot cups for the pellets.

Wound Ballistics

Exterior ballistics deals with the behavior of the projectile *after* leaving the barrel (trajectory, velocity etc.), whereas terminal ballistics covers the interaction between the projectile and the target (*see* **Firearm Examination: Ballistics**). If the target is a human or animal body, one speaks of wound ballistics.

Fundamentals of the Wounding Capacity

The wounding capacity of a projectile is partly due to the direct destruction of anatomical structures along the bullet track by crushing, punching, and tearing. Another type of lesion is caused by changes in pressure and displacement of tissue (with stretching and shearing) around the permanent wound channel. The extent of mechanical damage depends on the amount of kinetic energy (E_k) released in the tissue.

When a projectile penetrates the tissue, this is displaced laterally (radially) – i.e., at right angles to the bullet path – thus forming a temporary wound cavity, whose diameter can be considerably larger than the bullet. The radially displaced tissue then moves back in the opposite direction toward the geometric bullet path.

The process just described is especially marked with high-energy projectiles such as those fired from military and hunting rifles. In fluid-filled organs (heart, urinary bladder) or in the skull, the radial expansion may lead to a "hydrodynamic explosive effect" with bursting of the encasing structures. Cases in which the brain is completely flung out of the cranial cavity are referred to as *exenteration shots*. Even a shot with a smaller transfer of kinetic energy may cause indirect lesions away from the wound track, e.g., skull fractures, cerebral contusions, and stretch-mark-like tears of the facial skin.

The "permanent" wound channel represents the destructive passage of the bullet itself. The path is filled with blood and is surrounded by a more or less wide zone in which the tissue was temporarily stretched, thus suffering structural damage ("zone of extravasation").

By firing test shots at "simulants" such as gelatin or glycerin soap, the transfer of kinetic energy in biological soft tissue can be visualized, as the density of these materials is similar to that of muscle. In contrast to elastic gelatin, soap shows an almost plastic deformation. The bullet path and the volume of the cavitation remaining after firing the shot is proportionate to the energy transferred [4].

A stable bullet with a low deformation potential (full-jacketed projectile, as used in military ammunition) produces a "narrow channel" in the simulant at first, which then opens into the larger temporary cavity as the projectile moves in a sideways position (when tumbling) and imparts more energy to the surrounding tissue. An expanding deformation of the projectile also increases the effect of radial displacement and thus the volume of cavitation. In deformation projectiles (e.g., semijacketed hollow-point bullets as used in civilian hunting ammunition), the cavity starts forming immediately after penetration. A fragmentation of the projectile results in a multitude of wound tracks.

When projectiles of identical design, head configuration, and mass are fired, the transfer of energy in a dense medium and thus the extent of the temporary cavity essentially depend on the velocity of the bullet.

Wound Track

In gunshot injuries, the bullet may remain lodged in the body, or perforate it completely (through-and-through shot), or hit only the surface tangentially (graze wounds). In the first-mentioned type of gunshot wound, there is an entrance, but no exit wound. Rather often, projectiles traveling at a low velocity remain lodged under the tough and resilient skin on the side of the body opposite to the entrance wound, where their final position can be recognized by a hematoma and/or a palpable resistance (Figure 1). Radiological examination is advisable in any case to determine and document the localization of bullets and bullet fragments retained in the body.

For the determination of the angle of fire (in relation to the horizontal, sagittal, and frontal plane of the body), it is imperative that the length of the wound track and the localization of the entry and exit wound or, in shots with the projectile retained



Figure 1 Nonperforating gunshot injury to the chest (.22 caliber rimfire rifle, homicide). The spent lead bullet got stuck in the subcutaneous tissue of the back where a hematoma indicates its position (a). The projectile could be exposed by carefully incising the overlying skin (b)

inside the body, the final position of the bullet are exactly measured and recorded (height from the plantar plane and the lateral distance from the median plane). Statements as to the bullet's trajectory before entering the individual are possible only if additional facts (such as the position of the victim and/or the shooter) are known; in the presence of a through-and-through gunshot wound, it is essential to know any secondary sites of impact or penetration (on the floor or the walls, in the furniture etc.) to reconstruct the trajectory.

In most cases, the wound track in the body is linear. Full-jacketed rifle bullets, however, may produce a curved bullet path if its length in the body is longer than 20–30 cm. The deviation from the straight line begins when the projectile first moves into a lateral position, i.e., in the region of the first cavitation, as the pressure gradient along the projectile becomes asymmetrical, thus creating a force component lateral to the direction of the movement.

Nonlinear bullet paths are often caused by internal ricochet. Inside the cranial cavity, such ricochets are seen in 10–25% of the cases and occur if the projectile is deflected from the internal table of the skull with a low residual energy. Bullets either ricochet back into the brain at an acute angle or pass along the inner surface of the skull producing a curved wound track in the underlying brain.

The latter type of ricochet not only occurs on the concave side of the cranium but also on other inner surfaces (e.g., ribs) if a concave boundary surface continuously changes the direction of the bullet.

In a perforating shot, the projectile produces an exit wound as it leaves the body. Bullets with a low residual energy are sometimes no longer capable of perforating the clothing covering the exit wound. After having passed through one part of the body (e.g., the upper arm), the bullet may reenter in another part (e.g., the thorax).

Graze shots produce groovelike lesions on the body surface, occasionally accompanied by short tears along the wound edges. If a projectile strikes the body with low residual energy, but does not penetrate, the affected skin may show an excoriation and/or a hematoma.

Intermediate Targets, Deflection of the Projectile

For the interpretation of a gunshot wound, it may be essential to know if the projectile struck the human body primarily or if it interacted with an intermediate object before. For example, if the bullet passes through an intermediate target such as a door first, the typical "ring of dirt" on the site where it entered the clothing or body is missing, because the grayish-black depositions adhering to the bullet surface were already wiped off at the primary target.

When a bullet passes through the dense medium of an intermediate target, it loses its gyroscopic stability, resulting in a rotation around a lateral axis. If such a bullet then hits the body in an oblique or sideways position, the entrance wound is elongated and there is a higher loss of energy in the initial section of the wound track. Analogous effects may be seen when the bullet was already deformed at the primary target.

If a projectile is deflected by an intermediary target instead of penetrating it, one speaks of a ricochet. A change of direction may happen, for example, if the bullet strikes stone, concrete, or asphalt. In such cases, the bullet often shows a flattened, mirrorlike surface. The deformation and/or fragmentation of the ricochet bullet may cause an atypical entrance wound with no, or an incomplete ring of dirt. Owing to the loss of velocity and the instability of the ricochetting projectile, its depth of penetration is less than in primary hits after an undisturbed trajectory [5].

Lethal Gunshot Injuries

According to statistical investigations, about 20% of gunshot injuries are primarily lethal, i.e., the victims die before receiving medical care.

Generally, fatal consequences of a gunshot wound also have to be expected if weapons are used, which lay people would not consider very dangerous (e.g., air guns, blank-cartridge weapons, and .22 caliber rimfire rifles). For example, bullets fired from conventional air guns (with a common barrel diameter of .22 or .177 in, that is, 5.6 and 4.5 mm, respectively) may perforate the thin temporal squama or penetrate into the cranium via the orbital cavity (eye socket). The gas jet of blank-cartridge guns has repeatedly caused penetrating skin lesions, bone fractures, and lethal injuries of vessels or organs, when fired from a very short distance [6]. Of course, projectiles with a low energy fired from .22 caliber rimfire weapons may also produce fatal injuries if major organs or great vessels are hit along the bullet track.

In gunshot injuries with a fatal outcome, the direct lethal effect may be due to various causes. A special case is the gunshot-related "exenteration" of the brain from the skull [3]. When the shot strikes the nape of the neck or the occipital region, it may directly destroy vital centers of the brain stem. More often, it is not the cerebral lesion as such, but the subsequent increase in the intracranial pressure (due to intracerebral, subarachnoidal, and subdural bleeding, sometimes associated with cerebral edema) that is responsible for the lethal outcome.

Gunshot fractures of the bony skull base are often followed by a hemorrhage into the nasopharynx; if the victim is unconscious, fatal aspiration of blood will result. Gunshot-related lacerations of the venous sinuses may act as entrance sites for air bubbles, possibly leading to death from venous air embolism.

Injuries to the heart, great vessels, or parenchymatous organs cause massive internal bleeding with consecutive hemorrhagic shock. Gunshots to the lung with traumatic pneumothorax are an acute threat because of the impaired respiration – especially if both sides are involved. Inflammatory complications are potential causes of delayed death.

Ability to Act

It is often wrongly believed that a gunshot to the head or the trunk always incapacitates the victim immediately. This opinion is disproved by a multitude of well-documented cases in which gunshot victims performed surprisingly differentiated actions even after severe traumatization of vital organs [7].

If a victim becomes unable to act, this is usually due to functional impairment of the central nervous system caused either by tissue lesions directly or indirectly by insufficient oxygen supply. Immediate incapacitation is to be expected if the bullet destroyed parts of the brain essential for physical activity – with exenteration of the entire organ in extreme cases. Targets of immediate incapacitation are the upper cervical spinal cord, the brain stem, the cerebellum, the basal ganglia, the motor areas of the cerebral cortex, and the large motor nerve tracts [8]. The bullet need not necessarily pass through these cerebral regions directly, as the gunshot-related pressure and shearing forces can also damage nerve structures and impair functions away from the bullet path.

Cerebral hypoxia with consecutive unconsciousness following gunshots to the chest is mostly due to massive loss of blood. However, even if the bullet strikes the heart, the aorta, or other large arteries, blood circulation will hardly cease immediately and even then the oxygen reserves left in the brain may be sufficient for simple and short actions. Consequently, rapid but not immediate incapacitation is to be expected after gunshot injuries of the heart, the aorta, and the pulmonary artery. On the other hand, victims go down immediately if struck in the spinal cord.

The pathophysiological considerations just described have significant implications for the assessment of suicides in which several shots were fired. Continued ability to act after a cerebral gunshot injury is observed especially if low-energy ammunition was used and/or the bullet track did not involve the above-mentioned structures of immediate incapacitation (upper cerebral spinal cord, brain stem, motor cortex areas, and large motor pathways). In most suicides with more than one shot to the head, only the frontal lobe(s) or one of the temporal lobes of the brain is involved. Multiple gunshots to the cardiac region are seen more often than multiple suicidal shots to the cerebral cranium.

Stopping Power

The term *stopping power* is used to characterize the potential biological effect of a projectile, in particular, its capacity to prevent a person from moving or attacking [4]. Actually, the idea conveyed by movies and TV films that the impact of a bullet stops or even knocks down the affected person is not true in real situations. Bullets do not have the potential to throw people off their feet. Otherwise, the person who shoots the gun would be knocked over, as action and reaction are equal and opposite.

In fact, the effectiveness of the projectile depends on the amount of energy transferred to the body, leading to local displacement and destruction of tissue. In this context, the shape of the bullet is essential for its effectiveness: If the bullet head is blunt, deceleration and energy transfer are larger. However, in real cases, the effect of a bullet not only results from its effectiveness but also to a large extent from the point of impact, i.e., from the affected region and the relevant anatomical structures.

Embolism of Projectiles

The rarely seen transport of bullets or shot pellets within the vascular system is called *embolization*. Most of the embolized projectiles are of smaller caliber and low velocity, which is sufficient only to penetrate the artery or vein, but not to exit the vessel again, so that the foreign body, which is now localized inside the vessel, may be moved to a region of the body away from the bullet path where it can be easily visualized by radiography [9].

Bullet and pellet embolization is mostly seen in the arterial system (entry via the heart or the aorta and transport, e.g., to the leg arteries). In rare cases, a projectile may enter a vein and travel from there to the (right) ventricle or to the branches of the pulmonary artery [10, 11].

Delayed Effects

In survived gunshot injuries with retained bullets or pellets, the question arises, if this may cause chronic lead poisoning. Generally, the risk is assessed as being very low. Most cases reported in the literature refer to patients with projectiles lodged in the joints or bones. The latency period until an intoxication becomes manifest ranges from a few months to several decades.

Criminalistic Aspects

The purpose of clinical examination or autopsy of persons with gunshot injuries is to answer the following questions [3]:

- Do the findings confirm the assumption of a gunshot injury?
- Number of hits?
- Did a striking projectile pass through the body or is it lodged in the body or did it produce a graze wound?
- What was the direction and angle of fire (trajectory)?
- Are there any clues as to the type of weapon and ammunition used?
- From what distance was the shot fired (contact shot, close-range shot, distant shot)?
- Do the wound characteristics in connection with the traces at the scene suggest self-infliction or involvement of another party?
- Did the gunshot injury result in immediate incapacitation?

Entrance and Exit Wounds

In order to determine the direction of fire, it is imperative that entrance and exit wounds are interpreted correctly.

Characteristics of Entrance Wounds

Typical features of an entrance wound in the skin are

- punched-out hole (i.e., a central tissue substance defect that cannot be closed by approximation of its edges);
- marginal zone without epidermis (abrasion ring);
- grayish-black ring of dirt (provided the projectile did not pass through another target first).

When the shot was fired either with the muzzle in contact or at close/intermediate range, the respective signs can be regarded as further evidence of a bullet entry wound.

Entrance Hole

The central entrance defect is roundish (if the projectile strikes at a right angle) or oval (if it strikes at an oblique angle). The diameter is usually smaller than that of the bullet.

The discrepancy between the caliber of the projectile and the diameter of the permanent entrance hole can be explained by the elastic behavior of the skin: On impact of the bullet head, the edges of the defect temporarily move centrifugally due to radial forces causing a reversible widening of the bullet entrance hole [12]. When the deformation forces cease, the elastic skin resumes its former shape so that the permanent entrance defect may be much smaller than the diameter of the bullet (this discrepancy is particularly marked on the palms of the hands and soles of the feet [9, 13, 14]). The size of the skin wound therefore does not allow to draw accurate conclusions as to the caliber of the projectile.

The reason for the skin defect remaining at the entrance site is essentially that the projectile transports tissue particles into the depth of the wound track. Moreover, at the moment of impact, small skin

particles are flung back against the direction of fire [12, 15, 16].

Abrasion Collar (Abrasion Ring/Margin/Rim)

The central entrance hole is usually surrounded by a circumferential loss of epidermis (and its natural pigmentation [17]), forming a moist, reddish margin when fresh and later assuming a brownish color due to the drying of the unprotected corium (Figures 2, 5a and 5b).

With the help of high-speed photography, Sellier was able to prove already in 1967 that the epidermis-free margin of the entrance wound is not caused by any major indenting with consecutive overstretching and local friction [12, 18]. When the bullet head strikes the skin, backspatter of marginal tissue particles is induced by the pressure exerted on the entrance site. The former idea that the bullet head indents the skin before penetration, thus causing marginal abrasion, is not correct. It also neither results from the bullet being hot nor from its rotating movement.

In the peripheral parts of the abrasion collar, the epidermis is often torn and detached like wallpaper, so that parching can progress beyond the epidermis-free zone after prolonged exposure to air. When the bullet strikes at an oblique angle, the abrasion ring is elliptic and eccentric, being wider on the side from which the shot was fired. A unilateral widening of the abrasion collar thus gives an indication of the direction in which the bullet was traveling.

If the skin of the entrance region is under water, no abrasion ring is formed. The same is true for shots to palms and soles [9, 18]. Entrance wounds from high-velocity centerfire rifles may lack a typical abrasion collar, but show small splits radiating from the edges (so-called microtears).

Bullet Wipe-off ("Ring of Dirt", "Grease Ring")

The criminalistic importance of the bullet wipe is due to the fact that – at least on the primary target – this finding is a reliable sign of a bullet entrance.

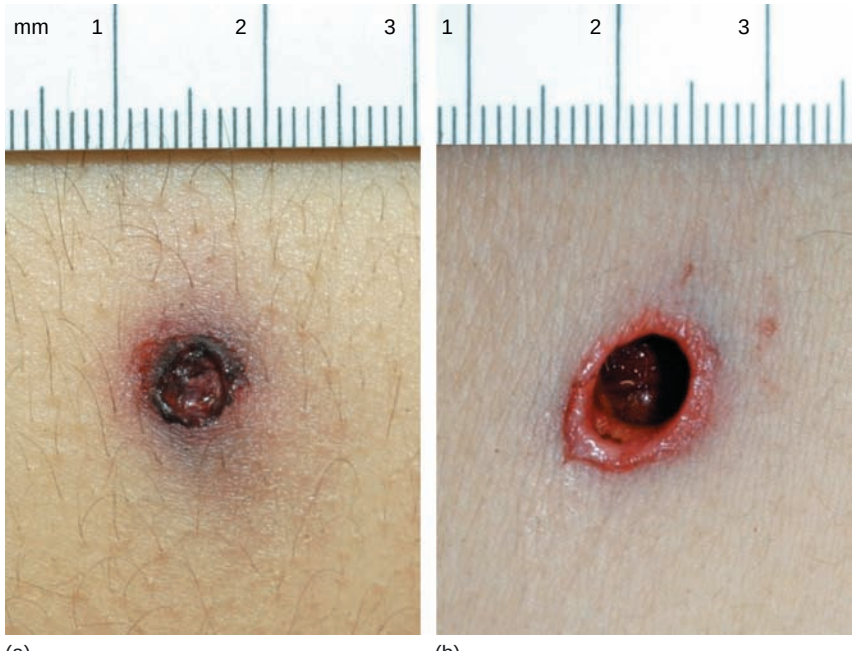


Figure 2 (a) Entrance wound (from a .22 LR projectile) in the epigastric region, which was covered only by a thin T-shirt. A vague grayish-black ring of dirt partly overlies the dried abrasion collar. (b) Entrance wound (from a 7.62 mm pistol bullet). The skin was covered by several layers of clothing. Therefore, the central defect is surrounded by a characteristic circumferential abrasion ring, but no bullet wipe-off

The term *bullet wipe* refers to the mechanism of formation: when the projectile hits a skin region not covered by clothing, sooty remnants and other residues deposited on the bullet's head are transferred to the wound margin, so that a grayish–blackish ring (partly) overlies the abrasion collar. More often, this "ring of dirt" is seen on the uppermost textile layer, but not (or only vaguely) on the margin of the entrance wound. In oblique gunshots, the bullet wipe is eccentrically enlarged on the side from which the shot was fired. The bullet wipe is *not* a sign of a close-range or a contact shot, as it also occurs in distant shots.

Exit Wounds

The exit wound presents as a slitlike or stellate severance of tissue (Figure 3a, b). In typical cases, there is – contrary to the entry wound – no real hole, i.e., no tissue defect. This means that the wound usually can be closed by bringing the edges into apposition. An exit wound produced by a bullet passing sideways through the skin may be slitlike and therefore mistaken for a stab wound.

Often, though not always, the size of the exit wound is larger than that of the entry wound. In practice, the uncritical application of this unreliable "rule" often leads to misinterpretations [5, 19]. Thus, contact shots fired to the head may show "stellate entrance wounds" with long radial tears; in such cases, the exit wound may be much smaller (Figure 4). In cases of splinter injuries (e.g., by fragments of explosive weapons), the entry wound is always larger than the corresponding exit. The differentiation between entrance and exit should never be made on the basis of simply comparing the wound dimensions.

The size of the exit wound mainly depends on the diameter of the temporary cavity at the site where the bullet leaves the body. In some cases, bone splinters carried along may also contribute to a larger exit hole. Many projectiles leave the body deformed and/or tumbling, which may also influence the shape of the exit wound.

It goes without saying that exit wounds cannot have a bullet wipe. Occasionally, the margins of the exit wound are abraded (shored) when a firm object (e.g., tight-fitting clothes, floor, wall or back of a chair) is pressed against the body at the site of the exiting projectile (Figures 3c and 5c). Under such circumstances, the skin around the exit is abraded by the supporting surface. In contrast to the "original" abrasion ring around the entry wound, in "shored" or "supported" exits the area of abrasion is not concentric, but irregular or lopsided and often disproportionately large [17].



Figure 3 (a) Slitlike bullet exit wound in the buccal region resembling a stab (.22 caliber revolver, homicide). (b) Stellate exit wound in the scalp, which was partly shaved before autopsy (7.65 caliber pistol, suicide). (c) Shored exit wound on the back. In the moment of discharge, a heavy backpack pressed against the bullet exit site (7 mm caliber hunting rifle, suicide)



(a) (b)

Figure 4 (a) Suicidal contact shot to the right temple with a revolver (caliber .357 magnum). The entrance wound shows a central defect with partially blackened edges (→) and radial lacerations of different length. Note the superficial stretch tear at the inner angle of the right eye, which is located several centimeters away from the wound channel (hyperextension of the facial skin by the expanding gases). (b) Irregularly shaped exit wound close to the left ear. The maximum diameter of the exit wound is much smaller than the large split at the entry site



(a) (b) (c)

Figure 5 (a) Five gunshot entrance wounds in the anterior thoracic region, which was originally covered by clothing. (b) Roundish entrance holes with circular abrasion rings (after removing the body hair). (c) Corresponding exit wounds on the back, which was in contact with the ground ("supported exits"). A 7.62 mm caliber pistol, homicide

Classification of Entrance Wounds in Relation to the Range from Muzzle to Target

To understand the different features of gunshot entry wounds, it is necessary to be familiar with the major processes occurring when a firearm is discharged. As the trigger is pulled, the firing pin is released and strikes the primer at the base of the cartridge case. The detonating primer ignites the propellant. The subsequent burning (deflagration) of the gunpowder generates a large amount of expanding gas, which is under high pressure and propels the projectile down the barrel. The gas is composed of carbon monoxide, carbon dioxide, oxides of nitrogen, and other compounds. A small percentage of the powder grains remains unburned or only partly burned.

Already, before the projectile leaves the barrel, a cloud of gunsmoke exits the muzzle. The term *gunsmoke* refers to the grayish-black combustion products of the powder that has not fully converted to gases. Essentially, gunsmoke consists of carbon in the form of soot.

Apart from the combustion gases and the finely dispersed soot, there are always unburned and partly burned powder grains expelled along with the projectile. The cloud of powder soot rapidly decelerates so that smoke soiling is to be expected only relatively close to the muzzle. The larger powder grains (having a diameter of at least several tenths of a millimeter) can also reach more distant targets.

The diameter of the spread and the density of soot and/or powder particles on a target are not only dependent on the range of fire but also on the cartridge type and the weapon (length of the barrel). Consequently, the range of discharge can only be evaluated by firing test shots with the respective weapon and ammunition [5].

In handguns, macroscopically visible traces of gunsmoke are to be expected up to a range of several centimeters. Depending on the weapon and ammunition, gunpowder grains may reach targets several decimeters away (in rifles even more than 1 m). The use of silencers strongly reduces the deposition of soot and powder particles, thus creating the false impression of a larger range of fire [20].

When a gun is discharged, two different light phenomena can be observed: first, the flame – a short, mostly dark red jet of fire caused by the not yet completely finished combustion of the powder particles; second, the muzzle flash – a glaring fire ball some distance away from the barrel end caused by the reaction of the incompletely oxidized combustion gases with the oxygen in the air.

With nitro powder, the extremely short impact of the muzzle flame is usually not sufficient to cause substantial burns on the clothing or skin. Sometimes, frizzing may be seen on the hair near the entry wound. Thermal damage is possible in shots with nitro ammunition fired from a very short distance (near contact) if textiles made of thermolabile synthetic fibers melt on the underlying skin [14, 15, 21]. If black powder ammunition is used, close-range shots may cause impressively large burns.

In forensic medicine, three ranges of fire are distinguished according to morphological criteria (*see* **Shooting Distance: Estimation of**):

- contact range
- short/close and medium/intermediate range
- long/distant range.

Contact Shots

The term *contact shot* means that the muzzle was held against the body surface at the time of discharge. In contact shots, soot-containing combustion gases are propelled into the depth of the entry wound. They expand beneath the skin and blacken the initial section of the wound track (pocketlike undermining, “powder cavity” containing soot and gunpowder particles). As the combustion gases have a high content of carbon monoxide (up to 50%), the surrounding tissue often assumes a bright cherry-red color. In tight (hard) contact shots, all the combustion products enter the wound, whereas in loose, angled, or incomplete contacts some soot may escape between the muzzle and skin so that the adjacent surface is blackened (Figure 6).

The entrance region is bloated by the inrushing powder gases and balloons backward against the muzzle end of the weapon, which is imprinted on the skin, causing a “muzzle abrasion” (“barrel marking”, “muzzle contusion”). Mechanically, the muzzle imprint is a patterned pressure abrasion with a tendency to parching after exposure to air [18]. Apart from the barrel end (or its contours), other constructional parts situated near the muzzle, such as the front sight and/or the recoil spring guide, may also be imprinted (Figure 7).

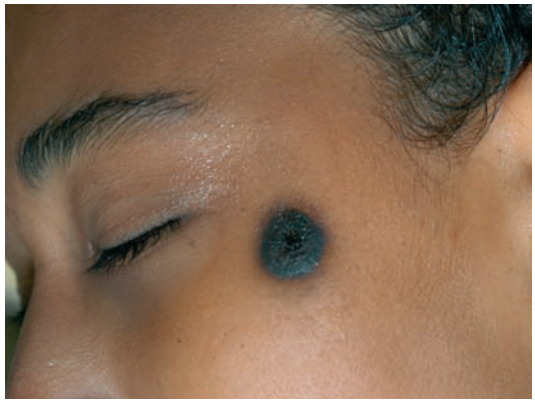


Figure 6 Near/loose-contact shot to the left zygomatic region (.22 caliber revolver, homicide). The entrance hole is surrounded by a wide zone of intense powder soot blackening

The muzzle contusion allows to draw significant conclusions:

- The weapon was in contact with the skin at the instant of discharge.
- The configuration of the imprint mark corresponds with the constructional elements being in line with the muzzle or just behind. Therefore, the imprint mark can characterize the type of weapon used (e.g., revolver or pistol) or even a specific make or model.
- The imprint configuration may provide information as to the way in which the weapon was held at the moment of discharge. For example, an imprint of the front sight below the bullet entrance means that the weapon had been held upside down (i.e., with the handle pointing upward).

If the entrance wound is above a bony support (e.g., in the frontal and temporal region), the subcutaneous expansion of the penetrating combustion gases may cause radial skin tears due to overstretching, resulting in a stellate wound of entrance. This additional sign of a contact shot is facultative: Shots fired with low-energy ammunition such as .22 LR do not necessarily cause stellate lacerations even at the sites having a bony support. Away from the entrance



(a) (b)

Figure 7 Hard-contact gunshot wounds to the right temple in two suicide cases with imprint marks mirroring the front end of the recoil spring guide and its position in relation to the muzzle. (a) "Normal" position (with the recoil spring guide being below the barrel and the grip pointing downward). The muzzle became blocked by soil when the pistol fell down after discharge. Brain substance protrudes through the bullet entrance wound. (b) The weapon is turned around the barrel's longitudinal axis so that the imprint is lateral to the bullet entrance hole

wound stretch-mark-like tears of the facial skin may occur, especially in shots to the forehead and the submental region (cf. Figure 4a, [22]).

Close and Intermediate-Range Shots

In close-/intermediate-range shots, GSR (soot and/or powder particles) are deposited around the entry wound. Usually a distinction is made between close-range shots and medium (intermediate)-range shots.

Close-range shots are defined by the presence of a zone of powder soot soiling surrounding the bullet entrance (often associated with additional powder tattooing). The grayish-black soot leads to skin or textile discoloration of a cloudy structure, whose intensity decreases with growing firing distance. Shots fired at an oblique angle result in an asymmetrical soot pattern with unilateral extension on the side of the shooter or away from it (depending on the angle at which the shot was fired and the range of fire). Interfering objects such as clothing or body parts (hand) may partially filter out the gunsmoke. Flash suppressors, which are often used in military rifles, have lateral smoke outlets, which may produce a flowerlike pattern of soot with several radial petals corresponding to the number of slits [9, 23]. In revolvers, there is a gap between the cylinder and the barrel, which allows the combustion gases to emerge sideward. The soot and the powder grains escaping from the gap may produce a characteristic linear or L-shaped mark if the skin or a fabric is in close proximity at the time of discharge [9, 24].

The term *medium-range shot* is used if no zone of powder soot blackening is discernible around the entry wound any more, but there are unburned or partially burned gunpowder grains deposited on, or forced into, the skin or clothing. The penetrating capacity of the powder grains depends on the propellant (flake or ball powder, grain size), the weapon, the range of fire, and the surface properties of the target. At short shooting distances, grains may be driven through thin textiles and cause stippling on the underlying skin [9, 17].

On the skin, the powder particles cause either superficial epidermal lesions (with subsequent drying) or – if deposited beneath the epithelium – petechial dermal hemorrhages (Figure 8). According



Figure 8 (a, b) Close-range shot to the left preauricular region (7.62 mm caliber pistol, homicide). The entrance wound (arrow) is surrounded by a dense dispersal pattern of unburned/partly burned powder grains associated with a small amount of soot around the bullet hole

to some authors, the term *tattooing* should be used to describe forceful in-driving, whereas the term *stippling* means the mere presence of impact markings.

The distribution pattern of powder tattooing/stippling varies according to the angle of fire: only perpendicular shots produce a radially symmetrical picture; in most other cases, the affected skin area is elliptic in shape. The entry wound may be localized outside the tattooing if the powder grains were partly filtered out by clothing or other primary targets (Figure 9, [17]). Pseudo tattoo marks can be due to fragments of an intermediate target such as the window of a car. In such cases, fragments of glass may produce irregular stippling lesions on the person seated behind the perforated window [9, 15, 21].

Distant-range Shots

In forensic usage, the term *distant shot* means that the weapon was discharged from such a distance that no soot and no powder grains could reach the body surface (skin or clothing in covered body regions). The minimum range for this type of gunshot wounds varies not only depending on the weapon and ammunition but also on the sensibility of the investigation method used.

The special methods of determining the range of fire cannot be discussed here. For securing and adequately preserving any GSR that may be present on the clothing or in the vicinity of entry wounds, close cooperation with the responsible experts is necessary (*see* **Firearm Discharge Residue: Analysis of**; **Shooting Distance: Estimation of**). In any case, the relevant findings should be documented by photographs and also by X rays, whenever possible.



Figure 9 Two intermediate-range shots to the sternal region. As the victim was clothed, the most of the unburned powder particles were filtered by the textile layers. Only some peripheral powder grains produced stippling of the uncovered neck. 7.62 mm caliber pistol, homicide [Reproduced from Ref. 17. © Hodder Arnold, 2000.]

Shotgun Injuries

Shotguns are hunting or sporting weapons intended to be fired from the shoulder (*see* **Firearms: Overview**). They are either single- or double-barreled, the latter ones being arranged either side by side or "up and under". A so-called pump gun has a pipe magazine under the barrel, which can take up several cartridges. Usual shotgun shells with birdshot or buckshot contain a multitude of pellets, which first travel together for a short distance and then separate more and more. The increasing dispersal of shot improves the hunter's chance of striking a moving target such as a hare or a flying duck. The criminal use of shotguns on humans is common, and improvisations, such as sawn-off barrels, facilitate the handling and hiding of the weapon [11, 14]. Shortening the barrel results in an increased spread of the pellets in midrange and distant-range shotgun discharges. Shot pellets are made from lead, which is easily deformed when striking dense tissue.

A shotgun contact wound is produced when the muzzle is placed tightly against the body surface. The entrance wound roughly corresponds to the gauge and is of circular shape in most body regions, but stellate over the bone (due to the expansion of the inrushing gases with consecutive backward ballooning of the skin). Sometimes, there is a clear imprint abrasion mark, e.g., from the front sight or – in double-barreled weapons – from the nonfiring muzzle. The wound edges may be blackened, but most of the soot enters the body and is deposited in the depth of the wound (Figure 10). The surrounding muscle is often colored cherry-red due to carbon monoxide. An intraoral discharge (Figures 11 and 12a) or a contact wound to the head (forehead, temple, under the jaw) leads to massive destruction of the skull ("bursting of the head", "shooting off the face") and is occasionally associated with evisceration of the brain [3, 9, 14].

Even in cases with extreme splitting of the face and the scalp, careful approximation of the wound edges helps in finding the entrance site



(a) (b)

Figure 10 Contact-range discharge to the left side of the neck (12 G shotgun, homicide). (a) Entrance wound with longitudinal splitting laceration toward the angle of jaw and soot staining in the depth. (b) Exit site on the right lateral neck

(Figure 12b). Loose-contact discharge allows the escape of sooty combustion gases, staining the skin around the entrance hole (Figure 13). Intermediate-range shots are characterized by the presence of stippling/tattooing from unburned propellant (up to 1 m, depending on the type of powder). In close-distance discharges, additional smoke soiling is seen.

The injury pattern of shot ammunition is mainly influenced by the distance between the muzzle and the target: As the range increases, the initially circular entrance hole shows scalloping of the wound edge ("nibbling", "crenation"); from a distance of approximately 2 m, peripheral pellets produce satellitelike holes outside the central entry defect. Ranges of several meters are characterized by a sievelike wound pattern [3, 5, 9, 14, 21, 24, 25]. If shots are fired from short distances, wads or plastic cups may either penetrate the body together with shot (cf. Figure 11) or cause excoriations of characteristic shape on the victim's skin ("wad abrasion") [3, 5, 9, 14, 24, 25].

Shotgun slugs are large, single lead projectiles destined for smooth-bore shotguns and also for those having a choke at the end of the barrel. The hitting accuracy of shotgun slugs is considerably lower than that of rifle projectiles; as a consequence, they should not be used for distances beyond 35–50 m in hunting. Some shotgun slugs are designed according to the arrow principle (heavy front part, light rear part).

Internal Findings

The gunshot lesions of the inner organs can only be briefly discussed here. Special mention should be made, however, of bullet holes in the flat bones of the skull. On the entrance side, the bone defect in the outer table is sharp edged and its minimum diameter roughly corresponds to the caliber of the projectile, whereas the inner table is beveled out in a conelike manner [1, 3, 5, 9, 14, 17, 21, 24]. This characteristic widening on the exit side allows to determine the direction of fire even in an isolated bone. For gunshot exit holes of the skull, the opposite is true: The outer table shows a craterlike defect ("outward beveling"). Projectiles striking at an acute angle produce keyhole-shaped entrance defects in flat bones with partial cratering of the outer table at the side away from the shooter. So the manner in which the bone breaks may



Figure 11 Intraoral discharge of a 12-gauge shotgun, resulting in facial disruption (suicide). The spent plastic piston and the lead pellets were recovered from the depth of the wound

indicate the inclination of the bullet path. Beveling is not restricted to the skullcap, as it is seen likewise in other flat bones such as the sternum, the pelvis, and the ribs.

In contact shots to the cerebral cranium (frontal, temporal, parietal, and occipital region), soot deposits are found not only under the skin but also around the bone defect and on the underside of the lifted-off periosteum (Figure 14), often even on the outer surface of the dura mater [14, 24, 26].

Frequently, radial fractures may extend from gunshot holes of the skull. According to Puppe's rule, a secondary fracture line ceases when it meets a pre-existing fracture line, which may help in determining the sequence of the shots.

In soft tissue, the wound track collapses and/or is filled with blood. Postmortem probing of the bullet path involves the risk of causing artifacts and should therefore be avoided. Parenchymatous organs such as the liver, the kidneys, and the spleen may show large stellate wounds at the sites of entry and exit.

Whole projectiles or bullet fragments removed from the body by surgical intervention or during autopsy must be preserved for further laboratory investigation, including ballistic comparison. A recovered projectile provides information regarding the caliber, twist direction, number and width of lands



Figure 12 (a) Intraoral shotgun wound with extreme facial splitting (suicidal discharge of a 12-gauge shotgun). (b) Reconstruction by reapproximating and sewing together the originally gaping lacerations. Note the superficial tears from overstretching along the nose. (c) Fragment of the skullcap (inner aspect) with shallow imprints (arrow) and one deformed pellet still adhering to the inner table of the bone. (d) Shotgun shell of the type used in the suicide case



Figure 13 Near contact-range discharge beneath the chin with intense soot deposition around the entrance hole (12-gauge pump-action shotgun, suicide)

and grooves, as well as of individual characteristics imparted by the inner surface of the barrel (*see* **Firearms: Bullet and Cartridge Case Identification**).

Livestock Stunners ("Humane Killers", "Slaughterer's Guns"), Stud Guns, Blank-Cartridge Guns

Stunning devices are guns firing a cylindrical steel bolt about 10 cm deep into the brain of animals intended for slaughter, which usually results in immediate unconsciousness [27]. Humane killers are loaded with a blank cartridge. After the captive bolt has been fired, a recoil spring brings it back into its initial position.

In forensic practice, injuries from livestock stunners are seen primarily in suicide cases; homicides and accidents with a fatal outcome are very rare. Suicides are mostly committed by persons with a pertinent professional experience (farmers, slaughterhouse workers, butchers).

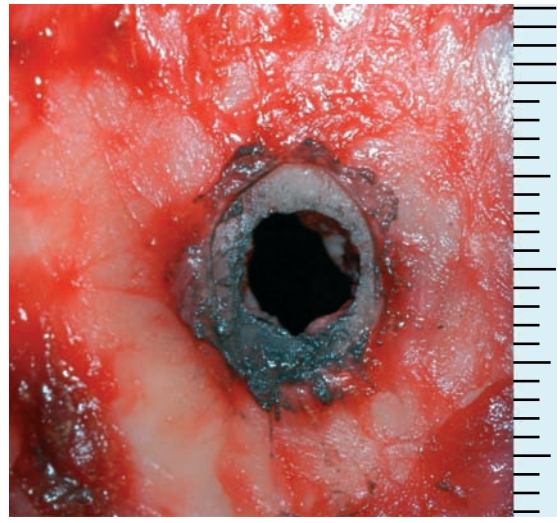


Figure 14 External surface of the skull (right temporal region) showing a roundish bone defect with soot deposited on the margin. The periosteum is detached and reflected. Suicidal contact shot, 6 mm caliber single-shot pistol

The steel bolt has a diameter of about 10–12 mm and a conically grooved distal end with a sharp edge and thus produces a punch injury (Figure 15a, b and c). As the bolt returns into the instrument, no projectile, but only punched-out material (skin, bone) is left in the depth of the wound track (Figure 15d). Because of the limited forward movement of the bolt, no exit wound is to be expected in shots to the head. Some types of livestock stunners have gas outlets in the muzzle plane. Corresponding to the openings of the smoke conduits, the entrance wound may be accompanied by two or, less often, four small zones of powder soot blackening located in pairs opposite each other (Figure 15a).

Cartridge-operated stud guns and nail guns are used to shoot fasteners into walls, wood, concrete etc. [3, 14]. In contrast to livestock stunners, the fasteners leave the device and may fly freely, thus causing an accident. For safety reasons, a shot can be fired only when the front plate of the instrument is pressed firmly to the object fired at.

Blank firing pistols and revolvers are detailed facsimiles of "real" handguns. The blank cartridges destined for these weapons contain gunpowder, but

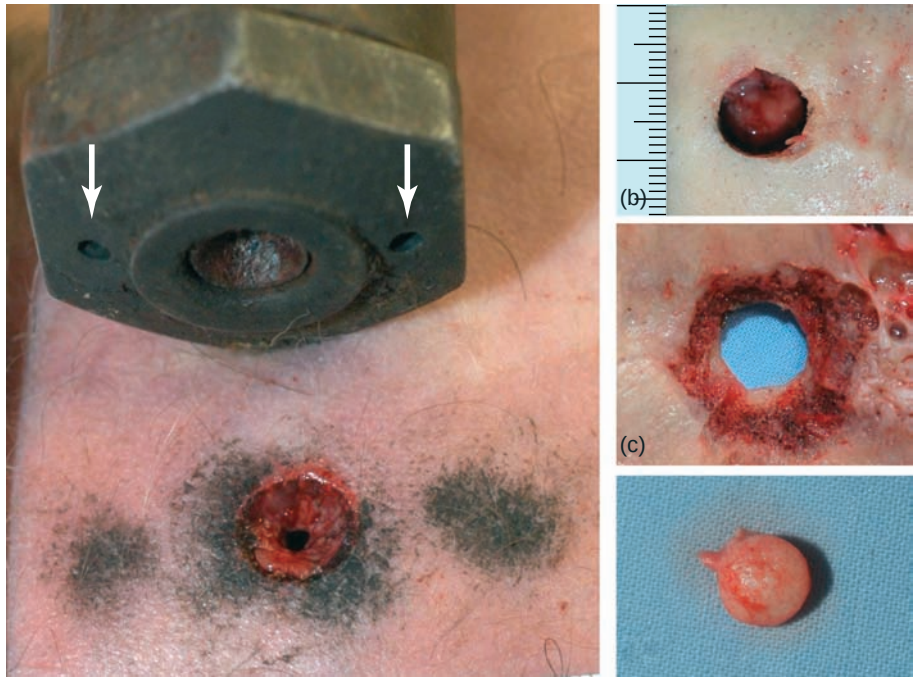


Figure 15 Suicidal captive-bolt injury to the parietal region from a livestock stunner (type Kerner). (a) Circular, sharply defined punch lesion of the skin. On both sides of the skin defect, roundish zones of powder soot blackening (corresponding to the opposite openings of the two smoke conduits in the muzzle end, arrows). (b) Circular entrance defect in the outer table of the skullcap, approximately corresponding to the diameter of the steel bolt. (c) Inner aspect of the skullcap with cone-shaped widening of the hole in the direction of the shot. (d) Punched-out bone fragment recovered from the depth of the wound channel

no projectile. If the muzzle is held in close proximity to the body surface or even in contact with it, the gas jet from the blank gun is capable of penetrating the skin and causing potentially fatal injuries [6, 28].

Forensic Examination and Documentation

In all firearm fatalities, careful documentation is of utmost importance [3, 5, 29]. This includes taking photographs and close-up views of each wound using a scale. The clothing must be preserved, as the uppermost layer may exhibit the bullet wipe around entrance holes and depositions of soot and/or powder particles in close- and medium-range shots, respectively.

Whenever possible, X rays should be taken before autopsy in two planes (anteroposterior as well as lateral views). They are not only a useful means of permanent and objective documentation but also help in exactly locating and characterizing all bullets and any metal fragments including separated jackets. Radiographs are also a valuable tool to find projectiles lodged in body regions, which are hardly accessible during autopsy (e.g., within the vertebral column). In addition, X rays provide evidence that a bullet might have been deflected or embolized [30].

All wounds have to be described exactly with regard to their location using fixed landmarks such as the base of the heels, the midline, the height above the buttocks, and the distance from the top of the head. The documentation should also mention the size and shape of each wound, the features of the wound margins and their surroundings, the presence or absence of GSR such as soot or stippling on the clothing and/or skin, the total length of the wound

tracks, and, of course, the injuries to the internal organs.

It is important to recover any bullets or major parts thereof from the body of the victim. Subsequent laboratory investigation may help in identifying the bullet type and to assign the fired bullet to a specific weapon (*see* **Firearms: Bullet and Cartridge Case Identification**).

Manner of Death

To classify a gunshot wound as suicidal, homicidal, or accidental, a synoptic evaluation of the scene and the circumstances of the case, the evidence obtained from the injuries, the victim's clothing, and the laboratory investigations concerning the weapon, ammunition, and range of fire has to be made [9, 30]. An easy access to weapons due to a permissive legislation is associated with increased rates of firearm homicide and suicide.

From the medicolegal point of view, the question has to be answered whether the entry wound is localized in a region typical for suicides [9, 31, 32]: temple (cf. Figures 4, 7 and 14), mouth (Figure 16; cf. Figures 11 and 12), cardiac region, forehead, and submental region (cf. Figure 13). In almost all cases of suicide, the muzzle is held against the body or inserted into the oral cavity. In shots to the chest, the skin is seldom bared before [33]. In more than 20% of the suicides committed with pistols or revolvers, the weapon is found clutched in the firing hand [9, 34].

The examination of hands to detect GSR, especially lead, antimony, and barium originating from the primer, can only be briefly mentioned here (*see* **Firearm Discharge Residue: Analysis of**). These residues mainly escape from the cylinder–barrel gap (in revolvers) or from the ejection port (in automatic pistols) and come to rest on the skin and/or clothing where they can be collected for subsequent chemical analysis. The presence of GSR is detected by flameless atomic absorption spectroscopy (FAAS) or by scanning electron microscope-energy dispersive X-ray spectrometry (SEM-EDX) [19, 35].

Sometimes, clues suggesting suicide are found on examination with the naked eye only (Figure 17), e.g., spray of blood or tissue deposits on the firing hand ("backspatter" from the entry wound [9, 36, 37]), traces of soot on thumb and index finger (if the muzzle end was held against the entrance site with one hand), or injuries from the edges of the recoiling slide. Direct and prolonged contact of the skin with steel parts of the weapon in a moist environment promotes the formation of brownish rust stains [9].



(a) (b)

Figure 16 Intraoral shot from a .22 caliber rimfire rifle (suicide). (a) Near contact entrance wound in the palate with eccentric soot deposition around the bullet hole. (b) Exit wound on the top of the head after reapproximation of the edges



(a) (b)

Figure 17 Suicidal shot to the right temple with a .38 caliber revolver. The left hand (a) was used to steady the barrel. Corresponding to the position of the cylinder–barrel gap, black soot is deposited on the palm (arrow). The trigger was operated with the right hand (b). The radial aspects and the back of both hands were spattered with a spray of blood

This phenomenon is found especially in suicides, although it is no proof that the shot was self-inflicted.

Injuries Caused by Explosives

In peacetime, injuries and fatalities due to the detonation of explosives are mostly seen in connection with politically motivated or terrorist attacks against persons (letter or parcel bombs), vehicles, and buildings (*see* **Explosions: Scene Investigation**; **Explosion Debris: Laboratory Analysis of**). Accidents are mostly due to natural gas ignitions, chemical explosions or improper handling of explosives, fireworks, etc. In countries without terrorist activities, suicides using explosives are rare and usually restricted to persons with a pertinent professional experience.

The injury pattern is often characterized by a complex combination of different lesions [1, 38, 39]. Mechanical tissue destruction up to traumatic amputation and evisceration is caused by blunt force (exploding device and other objects impacting the body, pressure wave), often associated with penetrating injuries from splinters, burning, and soot blackening of the skin. The internal examination may reveal organ and vascular damage, skeletal fractures, ruptured tympanic membranes, and acute pulmonary emphysema. Complete body X-ray examination should be performed whenever possible.

References

[1] Saukko, P. & Knight, B. (2004). *Knight's Forensic Pathology*, 3rd Edition, Arnold, London, pp. 245–280.

[2] Karger, B. & Teige, K. (1998). Fatalities from black powder percussion handguns, *Forensic Science International* **98**(3), 143–149.

[3] Dodd, M.J. (2006). *Terminal Ballistics: A Text and Atlas of Gunshot Wounds*, CRC Press, Boca Raton.

[4] Sellier, K.G. & Kneubuehl, B.P. (1994). *Wound Ballistics and the Scientific Background*, Elsevier, Amsterdam.

[5] Lew, E., Dolinak, D. & Matshes, E. (2005). Firearm injuries, in D. Dolinak, E.W. Matshes & E.O. Lew, eds,

Forensic Pathology: Principles and Practice, Elsevier Academic Press, Burlington, pp. 163–200.

[6] Rothschild, M.A. & Vendura, K. (1999). Fatal neck injuries caused by blank cartridges, *Forensic Science International* **101**(2), 151–159.

[7] Karger, B. (1995). Penetrating gunshots to the head and lack of immediate incapacitation. II. Review of case reports, *International Journal of Legal Medicine* **108**(3), 117–126.

[8] Karger, B. (1995). Penetrating gunshots to the head and lack of immediate incapacitation. I. Wound ballistics and mechanisms of incapacitation, *International Journal of Legal Medicine* **108**(2), 53–61.

[9] Di Maio, V.J.M. (1999). *Gunshot Wounds: Practical Aspects of Firearms, Ballistics, and Forensic Techniques*, 2nd Edition, CRC Press, Boca Raton.

[10] Yoshioka, H., Seibel, R.W., Pilhai, K. & Luchette, F.A. (1995). Shotgun wounds and pellet emboli: case reports and review of the literature, *Journal of Trauma* **39**(3), 596–601.

[11] Pollak, S., Ropohl, D. & Bohnert, M. (1999). Pellet embolization to the right atrium following double shotgun injury, *Forensic Science International* **99**(1), 61–69.

[12] Sellier, K. (1969). Bullet entry studies of the skin, *Beiträge zur Gerichtlichen Medizin* **25**, 265–270.

[13] Pollak, S. (1980). Morphology of bullet entrance wounds in palms and plantae, *Zeitschrift fur Rechtsmedizin. Journal of Legal Medicine* **86**(1), 41–47.

[14] Spitz, W.U. (2006). Injury by gunfire, in W.U. Spitz, ed, *Spitz and Fisher's Medicolegal Investigation of Death*, 4th Edition, Thomas, Springfield, pp. 607–746.

[15] Pollak, S. (1982). Macro- and micromorphology of bullet wounds caused by handguns, *Beiträge zur Gerichtlichen Medizin* **40**, 493–520.

[16] Grosse Perdekamp, M., Vennemann, B., Mattern, D., Serr, A. & Pollak, S. (2005). Tissue defect at the gunshot entrance wound: what happens to the skin? *International Journal of Legal Medicine* **119**(4), 217–222.

[17] Besant-Matthews, P.E. (2000). Examinations and interpretation of rifled firearm injuries, in J.K. Mason & B.N. Purdue, eds, *The Pathology of Trauma*, 3rd Edition, Arnold, London, pp. 47–60.

[18] Pollak, S. & Rothschild, M.A. (2004). Gunshot injuries as a topic of medicolegal research in the German-speaking countries from the beginning of the 20th century up to the present time, *Forensic Science International* **144**(2–3), 201–210.

[19] Smock, W.S. (2000). Evaluation of gunshot wounds, in J.A. Siegel, P.J. Saukko & G.C. Knupfer, eds, *Encyclopedia of Forensic Sciences*, Vol. 1, Academic Press, London, pp. 378–384.

[20] Missliwetz, J., Denk, W. & Wieser, I. (1991). Shots fired with silencers – a report of four cases and experimental testing, *Journal of Forensic Sciences* **36**(5), 1387–1394.

[21] Pollak, S. & Saukko, P. (2003). *Atlas of Forensic Medicine (CD-ROM)*, Elsevier, Amsterdam.

[22] Faller-Marquardt, M. & Pollak, S. (2002). Skin tears away from the entrance wound in gunshots to the head, *International Journal of Legal Medicine* **116**(5), 262–266.

[23] Grosse Perdekamp, M., Braunwarth, R., Schmidt, U., Schmidt, W. & Pollak, S. (2003). Contact shots from infantry weapons fitted with a flash-suppressor, *Archiv für Kriminologie* **212**(1–2), 10–18.

[24] Dana, S.E. & Di Maio, V.J.M. (2003). Gunshot trauma, in J. Payne-James, A. Busuttil & W. Smock, eds, *Forensic Medicine: Clinical and Pathological Aspects*, Greenwich Medical Media, London, pp. 149–168.

[25] Cassidy, M. (2000). Smooth-bore firearm injuries, in J.K. Mason & B.N. Purdue, eds, *The Pathology of Trauma*, Arnold, London, pp. 61–74.

[26] Faller-Marquardt, M., Bohnert, M. & Pollak, S. (2004). Detachment of the periosteum and soot staining of its underside in contact shots to the cerebral cranium, *International Journal of Legal Medicine* **118**(6), 343–347.

[27] Pollak, S. & Saukko, P.J. (2000). Humane killing tools, in J.A. Siegel, P.J. Saukko & G.C. Knupfer, eds, *Encyclopedia of Forensic Sciences*, Academic Press, London, Vol. 2, pp. 939–944.

[28] Grosse Perdekamp, M., Peuten, M., Sequenc, A., Schmidt, U. & Pollak, S. (2001). Fracture of the lower jaw by a contact shot with a blank cartridge weapon (starter's pistol), *Archiv für Kriminologie* **208**(3–4), 88–95.

[29] Kirk, G.M. (2005). Firearm injuries, in J. Payne-James, R.W. Byard, T.S. Corey & C. Henderson, eds, *Encyclopedia of Forensic and Legal Medicine*, Elsevier, Oxford, Vol. 3, pp. 110–118.

[30] Naidoo, S.R. (2005). Ballistic trauma, overview and statistics, in J. Payne-James, R.W. Byard, T.S. Corey & C. Henderson, eds, *Encyclopedia of Forensic and Legal Medicine*, Elsevier, Oxford, Vol. 1, pp. 271–283.

[31] Sellier, K. (1986). Death: accident or suicide by use of firearm, in A. Maehly & R.L. Williams, eds, *Forensic Science Progress*, Springer, Heidelberg New York Tokyo, Vol. 1, pp. 91–115.

[32] Karger, B., Billeb, E., Koops, E. & Brinkmann, B. (2002). Autopsy features relevant for discrimination between suicidal and homicidal gunshot injuries, *International Journal of Legal Medicine* **116**(5), 273–278.

[33] Karger, B., Kersting, C. & Brinkmann, B. (1997). Prior exposure of the entrance wound region from clothing is uncommon in firearm suicides, *International Journal of Legal Medicine* **110**(2), 79–81.

[34] Krauland, W. (1984). Evaluation of fatal gunshot injuries (the weapon in the hand), *Archiv für Kriminologie* **174**(1–2), 1–22.

[35] Heard, B.J. (1997). *Handbook of Firearms and Ballistics: Examining and Interpreting Forensic Evidence*, Wiley, Chichester, pp. 179–200.

[36] Karger, B., Nüsse, R., Schroeder, G., Wüstenbecker, S. & Brinkmann, B. (1996). Backspatter from experimental close-range shots to the head. I. Macrobackspatter, *International Journal of Legal Medicine* **109**(2), 66–74.

[37] Betz, P., Peschel, O., Stiefel, D. & Eisenmenger, W. (1995). Frequency of blood spatters on the shooting hand and of conjunctival petechiae following suicidal gunshot wounds to the head, *Forensic Science International* **76**(1), 47–53.

[38] Crane, J. (2005). Explosive injury, in J. Payne-James, R.W. Byard, T.S. Corey & C. Henderson, eds, *Encyclopedia of Forensic and Legal Medicine*, Elsevier, Oxford, Vol. 3, pp. 98–100.

[39] Spitz, W.U. (2006). Medicolegal considerations of bomb explosions, in W.U. Spitz, ed, *Spitz and Fisher's Medicolegal Investigation of Death*, 4th Edition, Thomas, Springfield, pp. 777–782.

Related Articles

Explosions: Scene Investigation
Explosion Debris: Laboratory Analysis of
Firearm Examination: Ballistics
Firearms: Bullet and Cartridge Case Identification
Firearm Discharge Residue: Analysis of
Firearms: Overview
Shooting Distance: Estimation of

STEFAN POLLAK AND PEKKA J. SAUKKO

Hair: Animal

Introduction

The earliest publications acknowledging the significance and value of animal hairs in the forensic arena date back to the nineteenth century [1, 2]. In 1894, Alfred Swaine Taylor documented a case, which relied on establishing innocence by determining hairs on an alleged murder weapon to be animal, not human, in origin [3]. By the early 1900s, microscopical examination of hair was well established; in 1931 Professor John Glaister published his work on the study of mammalian hairs from the medicolegal aspect, which became a standard reference work [4].

Technological advances, in the late 1930s, for the textile industry prompted detailed hair structure studies by the Wool Industries Research Association (WIRA, England). Wildman published a booklet in 1940 on animal fibers of industrial importance and discussed their origin and identification. This work was subsequently enlarged upon, republished in 1954 [5], and became generally accepted as the definitive work on hair structure. However, Wildman's work concentrated on a detailed description of fibers of commercial importance and no attempt was made to produce a key or system of identification for unknown hairs. A further account of Wildman's work was subsequently given by Appleyard [6].

Descriptive keys to facilitate the identification of mammalian hairs were published by Mathiak [7], Mayer, [8] and Moore [9]. Brunner and Coman [10] acknowledged these works and their value but believed that a number of problems still existed in the identification of unknown animal hairs, which could not be addressed solely by a key based system. The authors attempted to address these issues and minimize the problems of inter and intraspecies variation and overlap by using a photographic reference system.

For almost a century, the compound and comparison microscopes remained the only reliable tools for the microscopical identification and characteristics found in animal hair. However, the advent of nuclear DNA (nDNA) and mitochondrial DNA (mtDNA) analyses in the forensic arena provided a degree of individualization of animal hair, which was not possible *via* microscopy alone [11–14].

Transfer and Persistence of Animal Hair

The forensic community generally accepts that whenever two objects or surfaces come into contact, a transfer of material occurs. Trace evidence material that is transferred may be used to associate objects, individuals, or locations. Once transferred, animal hairs may be recovered from evidentiary items using a variety of techniques that include removal with forceps or adhesive tape.

Until the mid-1990s, published material did not exist regarding the transfer and persistence of animal hair. Prior to this, forensic hair examiners relied on their experience and a plethora of published data detailing the transfer and persistence of textile

and natural fibers to determine if recovered animal hairs were transferred *via* primary or secondary transfer, i.e., animal sheds hairs onto a surface or object (primary transfer), which then become transferred onto a different surface or object (secondary transfer). As a result of these studies and based on the experience of the examiner, hair examiners based their conclusions on the tenet that the greater the number of hairs recovered from an item, the more likely that the transfer arose as a result of primary transfer rather than secondary transfer.

Robertson [15] addressed the issue of the transfer and persistence of hairs thus: "... there are elements common to fibers and hairs (after all, hairs are just natural fibers and the factors which affect their transfer and persistence appears similar).... Most of the work on transfer and persistence relates to studies with textile fibers. Hairs could be expected to behave like woolen textile fibers because both have scaled outer layers. Hence, many of the conclusions reached for fibers may also be applicable to hairs".

In 1998, D'Andrea *et al.* [16] published their findings from their preliminary studies on the transfer and persistence of animal hair during simulated burglaries and assaults. They concluded that the results "confirm the existence of an important transfer and persistence of this type of evidence". Furthermore, these authors determined that "the numbers presented ... are consistent with the kinetics of disappearance, well studied in the context of textile fibers". Their empirical data supported the premise on which hair examiners previously based their opinions regarding the transfer and persistence of animal hairs.

In 2008, Robertson *et al.* [17] investigated the transfer and persistence of cat and dog hairs in relation to wool and nonwoolen fabrics. The results indicate that these animal hairs exhibited an almost linear decay for hair loss from acrylic and woolen fabrics and a classic exponential loss with cotton and polyester fabrics.

Microscopical Examination of Animal Hairs

The morphology and composition of animal hair is comparable to human hairs in that there are essentially three components to the examination of animal hairs:

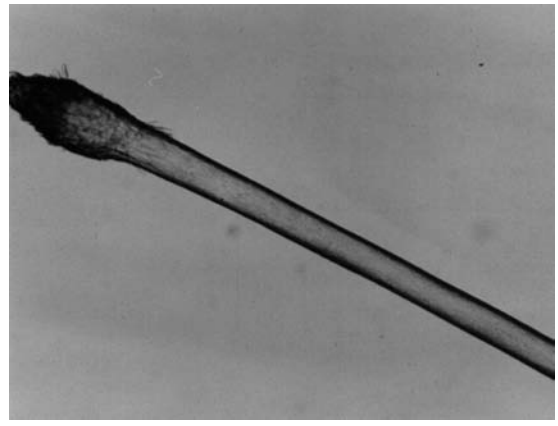


Figure 1 Human scalp hair that bears a club-shaped root

- The determination of the hair as animal versus human in origin.
- The identification of the animal hair as originating from a particular family or species.
- The use of the comparison microscope for the simultaneous examination of microscopic characteristics found in questioned and known animal hairs.

If the questioned animal hairs are sufficient in number and/or hair types, distinctive enough, and in good enough condition, the above goals may be readily achieved.

Animal Hair versus Human Hair

The identification of a hair as being animal versus human in origin is generally easy to accomplish and may be achieved at the macroscopic level. In general, human hair roots are club-shaped (bulbous appearance) or flattened (Figures 1 and 2). In contrast, animal hair roots may exhibit a variety of shapes and forms, which may be characteristic of the species as depicted in figures 3 and 4.

Human hair possesses a uniform flattened (imbricate) cuticular scale pattern (Figure 5) and, if present, an amorphous medulla, which constitutes less than a third of the shaft width (Figure 6).

In contrast, animal hairs exhibit an array of different scale patterns and medullae forms, which, in

Table 1 Characteristics that may be used to differentiate between human and animal hairs

Feature	Human hair	Animal hair
Color	Relatively consistent along hair shaft	Often showing naturally occurring abrupt, profound color changes known as *banding*
Medulla	Less than 1/3 of the hair shaft amorphous appearance, most discontinuous when present	Usually greater than 1/3 of the width of the hair shaft Continuous, often varying in appearance along the shaft, defined structure with a variety of configurations
Pigment distribution	Even, slightly more toward the cuticle	Central or denser toward the medulla
Cuticular scale	Imbricate, similar along the shaft from root (proximal end) to tip (distal end)	Variety of patterns often showing variation in structure from root to tip
Root	Usually bulbous (club-shaped) and indistinct	Variety of shapes and forms, usually distinct



Figure 2 Human scalp hair that bears a flattened (ribbon) root

general, occupy greater than a third of the shaft width. Figures 7–12

Table 1 details the major morphological characteristics of human and animal hairs that may be present and assist in distinguishing human hair from animal hair.

The most obvious characteristic feature exhibited by animal hairs is the naturally occurring

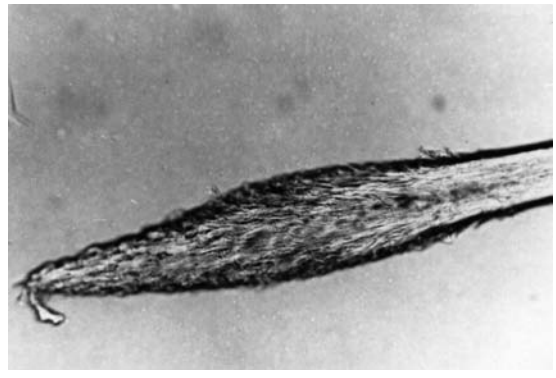


Figure 3 Photomicrograph depicting the spade-shaped root characteristic of dog hairs

phenomenon known as *banding*. Banding is an abrupt and profound color change in the same hair and can be seen with the naked eye. This naturally occurring phenomenon does not occur in human hairs.

Identification of Animal Hairs

Prior to attempting to identify an animal hair to a particular family or species, it is necessary to characterize the hair to a particular "type" to assess if characterization is possible. The hairs comprising the pelt of an animal may be identified on their gross morphological appearance.

Close examination of animal pelts reveals that some sparsely distributed hairs are distinctly longer than the remainder of the hairs. These longer hairs are called *overhairs*, for species identification purposes these hairs are not particularly useful as many



Figure 4 Photomicrograph depicting the wineglass-shaped root characteristic of deer hairs

of the characteristics required to determine species may not be present or visible if the hairs are heavily pigmented.

The guard hairs are coarser and larger than the remainder of the hairs forming the pelt. The largest of the guard hairs, the primary guard hairs, are of paramount importance in the identification of species as they exhibit the most diagnostic features. The secondary guard hairs bear characteristics, which are a hybrid of the coarse guard hairs and the finer underhairs, that is they are almost as coarse and of the same length as the primary guard hairs and wavy like the underhairs. These secondary hairs also bear diagnostic features and are just as important as the primary guard hairs for identification purposes.

The underhairs are shorter and much finer than the overhairs and guard hairs. These hairs are usually wavy and usually bear insufficient features upon which a reliable identification can be made.

The types of hairs that may be found on an animal pelt and examples of hair profiles that may be present are illustrated in Figures 13 and 14.

To further assist in the identification of animal hair to a particular species Brunner and Coman [10] introduced the reader to diagrammatic representations of some of the morphological features, which may be seen in animal hairs to assist in determining possible species of origin (Figures 15–17).

Cuticular Scale Patterns. Unlike human hairs, animal hairs bear a variety of cuticular scale patterns, which may change along the hair shaft. These patterns may be visualized by the use of a compound microscope (transmitted light) if the hair has prominent scale patterns or is not opaque. However, the use of a scanning electron microscope (SEM) greatly enhances the ability to view the cuticle. Another more routinely used method to view the scale patterns of

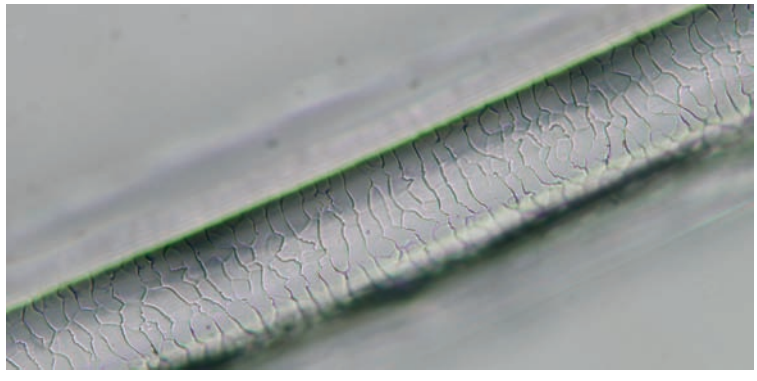


Figure 5 Imbricate scale pattern characteristic of human hairs

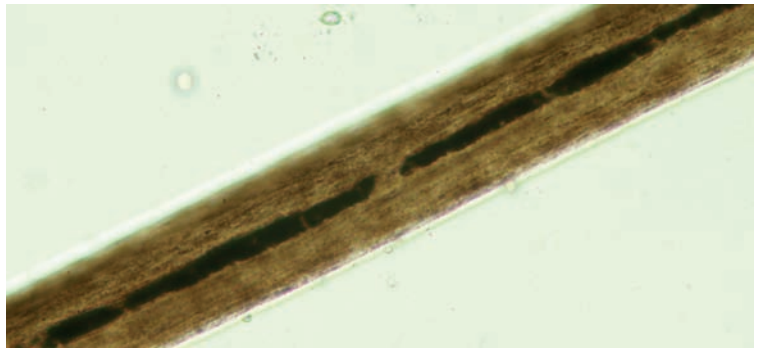


Figure 6 Amorphous medulla, which may be seen in human scalp hairs

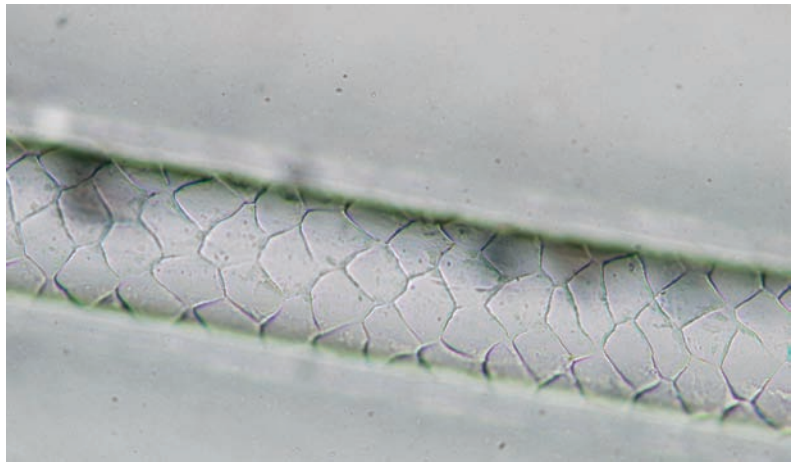


Figure 7 Example of scale pattern, which may be seen on cat hair

animal hair involves coating a glass microscope slide with a thin layer of clear adhesive or nail polish, placing the hair (with the tip slightly protruding from the slide to facilitate removal) in the adhesive, which is allowed to dry and upon removal of the hair the imprint of the scale (cast) pattern is revealed; this imprint may be visualized with the use of a compound microscope.

Figures 18 and 19 depict examples of cuticular scale patterns that may be seen in animal hairs. Rabbit hair shows a chevron scale pattern (Figure 18); whereas the dog shows a broader irregular mosaic scale pattern (Figure 19) as seen in scale cast patterns.

Medullae Patterns and Configurations. Hairs originating from members of the deer family characteristically exhibit a very broad honeycomb (wide lattice) medulla that virtually comprises the whole cortex (Figure 20), while the medulla of the dog hair is a much narrower simple medulla (Figure 21).

Root Shapes. Certain animal species bear roots of a certain morphology that is very characteristic of the species. For example, the forensic community widely accepts that dog hairs bear distinctive spade-shaped roots resembling a closed umbrella (Figure 3) and that deer hairs bear a wineglass-shaped root (Figure 4).

Cross Sections. Although the majority of family/species identification is based on cuticular scale



Figure 8 Example of the scale pattern, which may be seen on rat hair

patterns, root and medullae morphology, cross-sectional examination of animal hairs may be a useful tool in the preliminary grouping of mammal species for a rapid identification of hairs. However, cross-sectioning are not species specific and as this method is a destructive one, it is not recommended as a routine procedure.

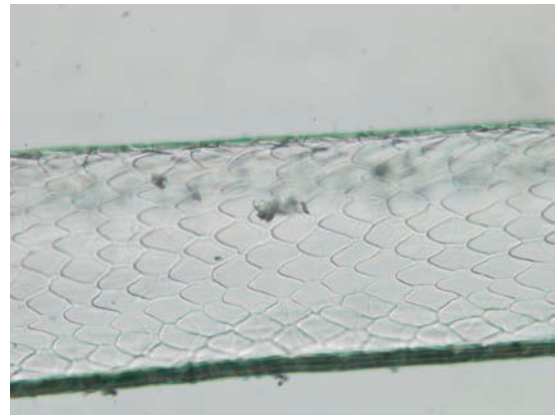


Figure 9 Example of the scale pattern, which may be seen on seal hair

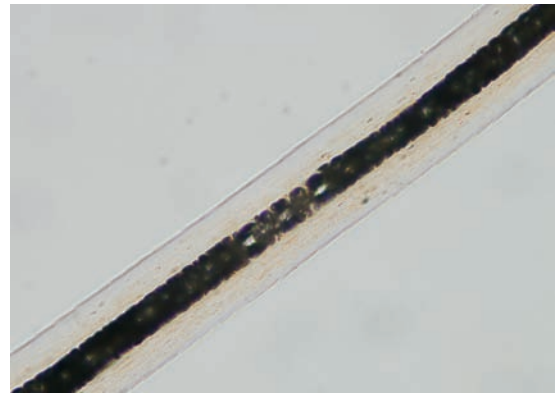


Figure 10 Example of the medulla exhibited by dog hair



Figure 11 Example of the medulla exhibited by rat hair

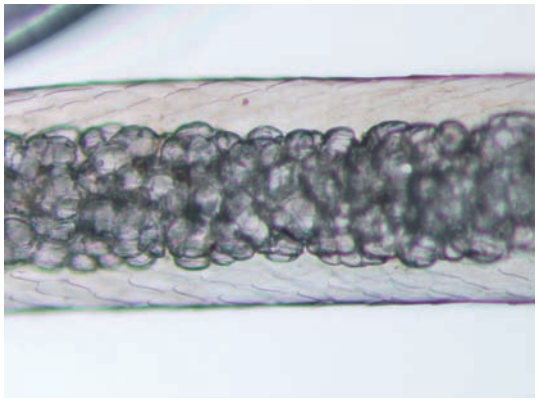


Figure 12 Example of the medulla exhibited by seal hair

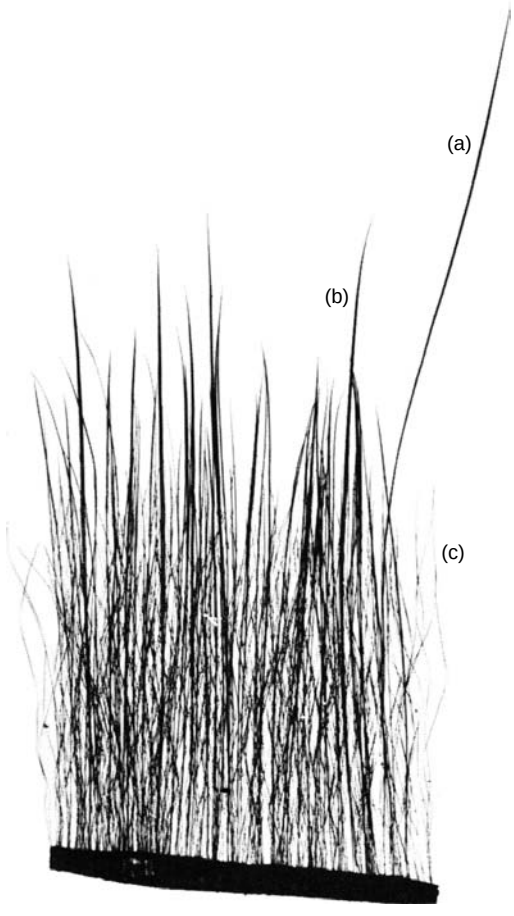


Figure 13 Examples of hair types, which may be found on the pelage of animals. A tuft of hair from the brown rat (*Rattus norvegicus)* showing an overhair (a), guard hairs (b), and underhairs (c)

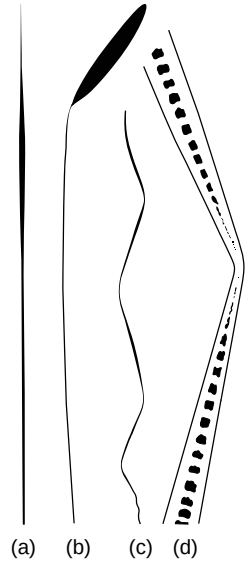


Figure 14 Examples of hair profiles: Guard hairs (a and b); underhair (c); magnified view of a constriction in an underhair (d)

The forensic analyst is often presented with only a few animal hairs to examine; family/species identification of animal hair is possible without having to resort to cross sectioning.

Thus, the use of keys [8, 9, 18] in conjunction with standard reference works [5, 6, 10] may assist the examiner to identify the family/species of origin of a particular animal hair.

Differentiation and Identification of Cat and Dog Hairs

Cat and dog hairs are probably the animal hairs most frequently encountered by forensic analysts, probably because of the popularity of these animals as domestic pets and the ease in which they shed their hairs on upholstery and clothing. Peabody *et al.* [19] described a method based upon a statistical difference between the medullary index (MI) of cat and dog hairs. The MI is the ratio of the width of the medulla in relation to the width of the hair shaft at a particular section of the hair. Peabody found that when the MIs are potted against shaft width, dog hairs predominantly lie below a line whereas cat hairs lie above it as illustrated in Figure 22.

The MI used in conjunction with other characteristics such as cuticular scale pattern and root shape may

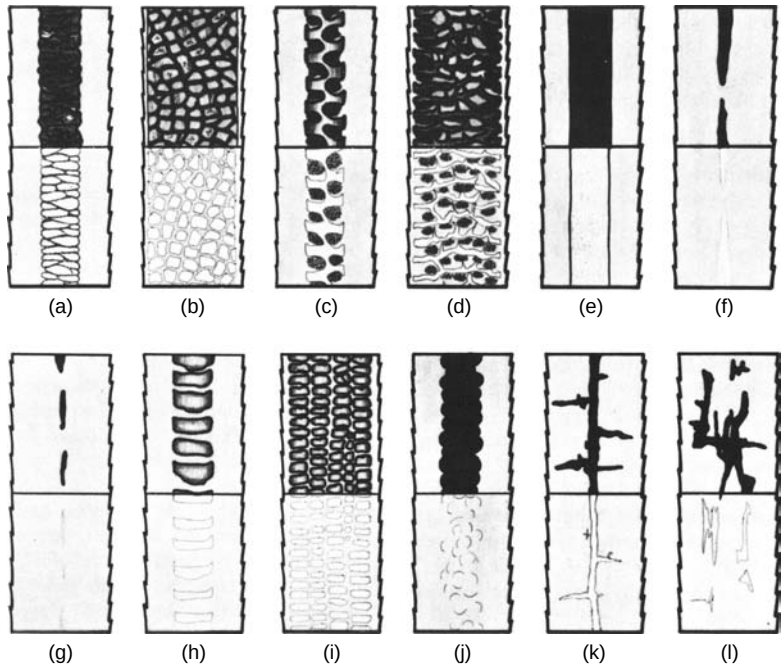


Figure 15 Medullae types exhibited by animal hairs as classified by Brunner and Coman. (a) Narrow medulla lattice; (b) wide medulla lattice; (c) narrow aeriform lattice; (d) wide aeriform lattice; (e) simple; (f) interrupted; (g) fragmental; (h) uniserial ladder; (i) multiserial ladder; (j) globular; (k) stellate; (l) intruding

further facilitate the identification and differentiation of cat and dog hairs.

The appearance of the root shape or form may assist in the identification of an unknown animal hair. Hicks [20] stated that cat hairs possess a fibrillar root (the appearance resembles a short bristled artist's brush) and dog hairs possess a spade-shaped root that resembles a closed umbrella, a tenet generally accepted by the forensic community. However, studies [21] revealed that cat hairs may also exhibit a spade-shaped root, which is shorter than those found on dog hair as illustrated in Figure 23.

The results of the study indicated that if the determination on whether a hair is of cat or dog in origin is based solely on the appearance of a spade-shaped root, an erroneous identification may be made.

Comparison Microscopy of Animal Hairs

The comparison microscope enables simultaneous examination of questioned and known hairs.

Although the comparison microscope is often used for the examination of human hairs, this is not the case for the comparison of animal hairs. Prior to 1988, no published literature existed regarding the significance of animal hair comparisons. Suzanski [22] conducted a number of comparisons of German Shepherd dog hairs to determine whether hairs from these animals possessed sufficient variation of microscopic characteristics to permit some degree of individualization. Suzanski performed the comparisons on the basis of the "unknown' dog hairs being selected, in his absence, from the known samples. The results indicated that with this particular breed of dog some degree of individualization was possible.

Suzanski [23] subsequently enlarged upon this work using the previously used known and" unknown" dog hair collection techniques by comparing hairs from a variety of purebred dogs and dogs of mixed breeds; the results supported his previous findings. In his 1988 publication, Suzanski noted that it is possible for two different dogs to have hair that is indistinguishable, only guard hairs and

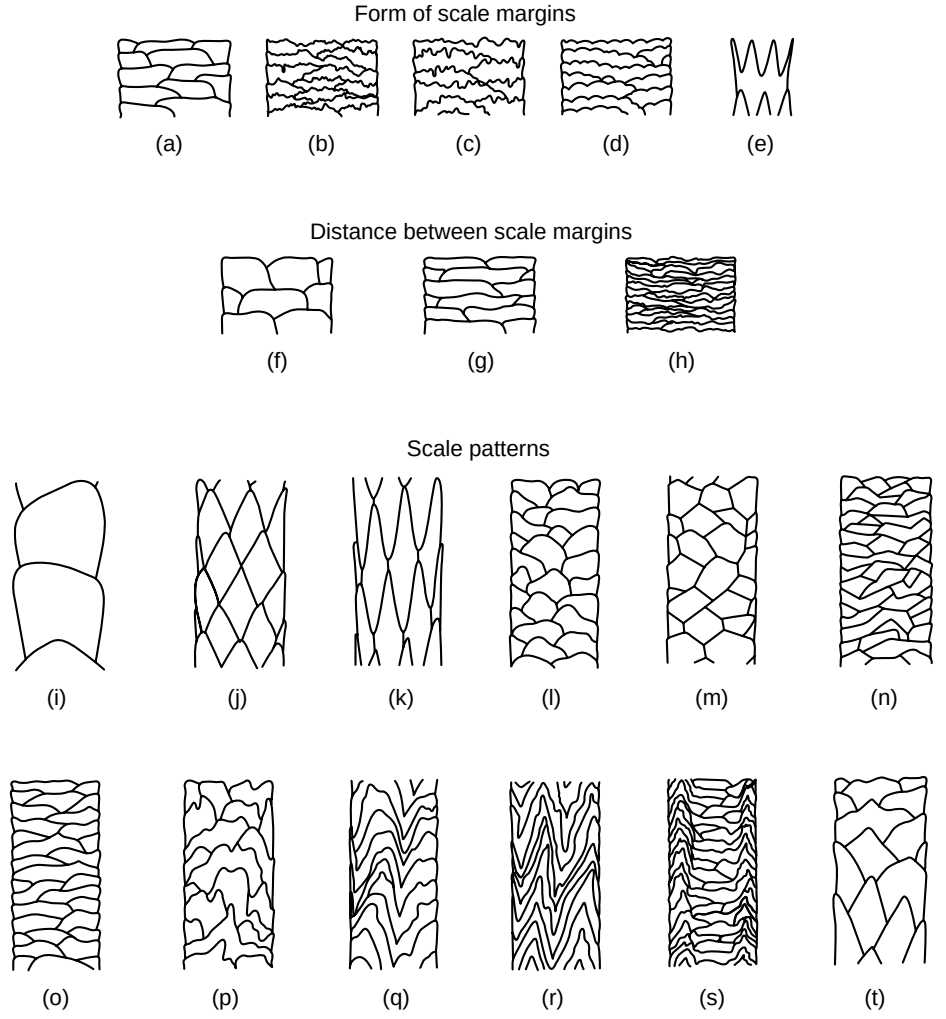


Figure 16 Cuticular scale patterns exhibited by animal hairs as classified by Brunner and Coman. (a) Smooth; (b) crenate; (c) rippled; (d) scalloped; (e) dentate (f) distant; (g) near; (h) close; (i) simple coronal; (j) diamond petal; (k) narrow diamond petal (l) broad petal; (m) regular mosaic; (n) flattened irregular mosaic; (o) regular wave; (p) irregular wave; (q) single chevron; (r) double chevron; (s) streaked; (t) transitional

intermediate hairs (secondary hairs) are suitable for comparison purposes and the principles of dog hair comparisons are generally the same as for human hair comparison although the relative weights placed on certain characteristics would be somewhat different.

DNA Analyses on Animal Hairs

Until the advent of DNA profiling, the hair examiner relied solely on microscopy to determine some level of individualization. DNA profiling of animal hairs has been able to place some statistical weight on the degree of individualization, which was erstwhile not possible. The following example illustrates this premise. In 1996 a North American forensic analyst presented testimony in which a suspect was linked to a homicide through cat hairs. A discarded jacket stained with the decedent's blood had several white cat hairs lodged in the lining; the suspect owned a white cat. The DNA profile obtained from one of

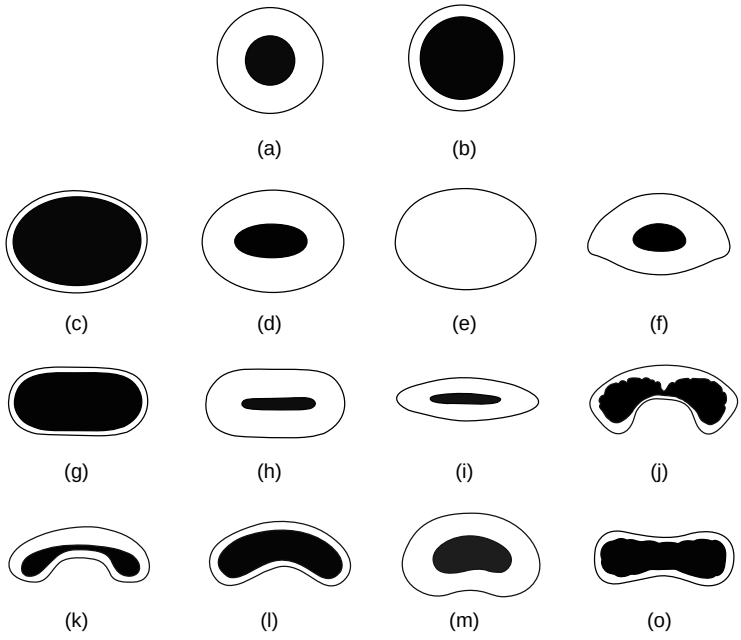


Figure 17 Cross-sectional shapes exhibited by animal hairs as classified by Brunner and Coman. (a) Circular medium size medulla; (b) circular large medulla; (c) oval large medulla; (d) oval medium size medulla; (e) oval medulla absent (f) eye-shaped; (g) oblong large medulla; (h) oblong medium size medulla; (i) cigar-shaped; (j) concavo-convex divided medulla; (k) concavo-convex bilobed medulla; (l) concavo-convex large medulla; (m) reniform; (n) dumb-bell shaped

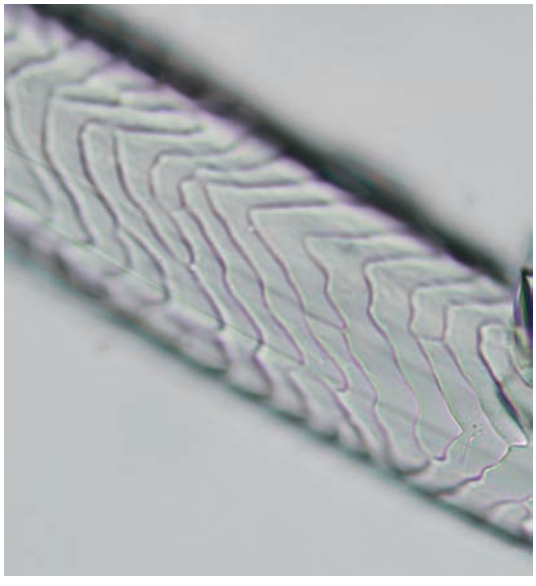


Figure 18 Chevron cuticular scale pattern as seen on rabbit hair

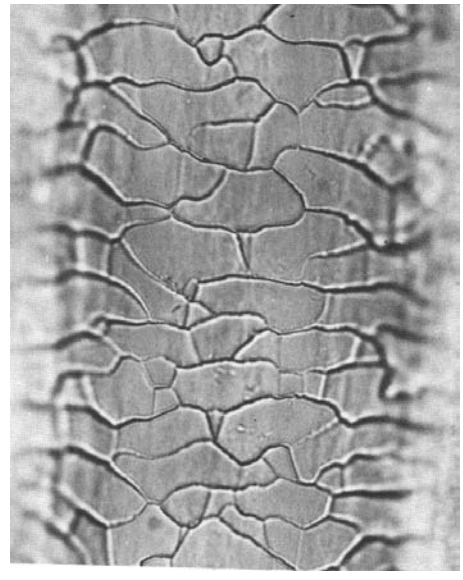


Figure 19 Irregular mosaic cuticular scale pattern as seen on dog hair

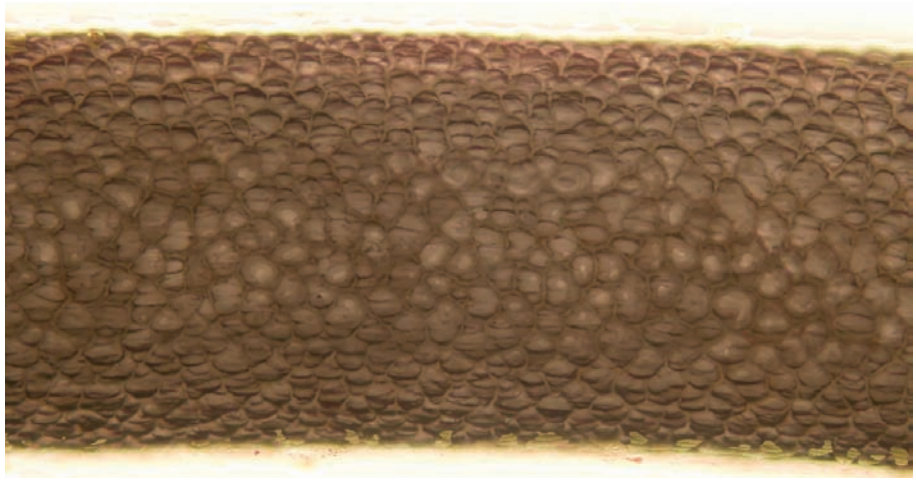


Figure 20 Example of a wide lattice (honeycomb) medulla as seen in deer hair

the roots of the recovered cat hairs was compared to the DNA profile obtained from the blood of the suspect's cat. The DNA profile from the recovered cat hair and the DNA profile from the suspect's cat matched. This evidence was presented at the trial and the jury subsequently found the suspect guilty of second-degree murder. This case represented legal precedence for the introduction of DNA profiling of pet animal hairs in forensic cases that associates the suspect in capital cases [11]. Thus, the advent of DNA profiling heralded a new era, not only for the forensic analysis of body fluids and human hairs, but also in the examination of animal hairs. As technology improves, the combination of microscopy and DNA analyses in the examination of animal hair is likely to become routine. However, until such times that the DNA profiling of animal hairs becomes routine, the only recourse available to the forensic analyst is the reliance on the "traditional" examination of animal hairs.

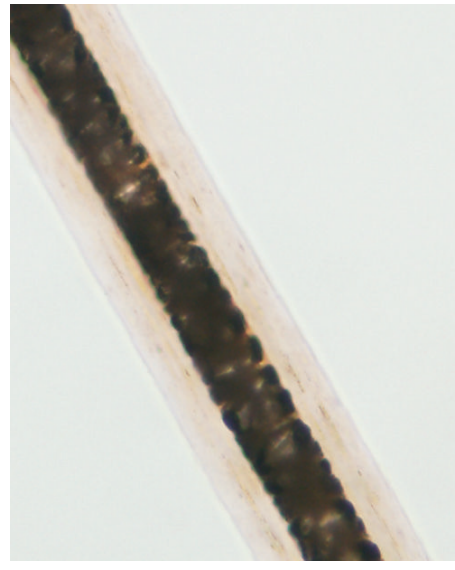


Figure 21 Example of a simple, narrow medulla as seen in dog hair

The Examination of Animal Hairs in Forensic Casework

The forensic examination of animal hairs is eclectic being conducted on, but not limited to, cases involving the illegal trafficking of animals, poaching, and killing of protected species, cruelty to animals, food contaminants, and in crimes against the person in which animal hairs may provide the nexus between offender and crime scene and/or victim.

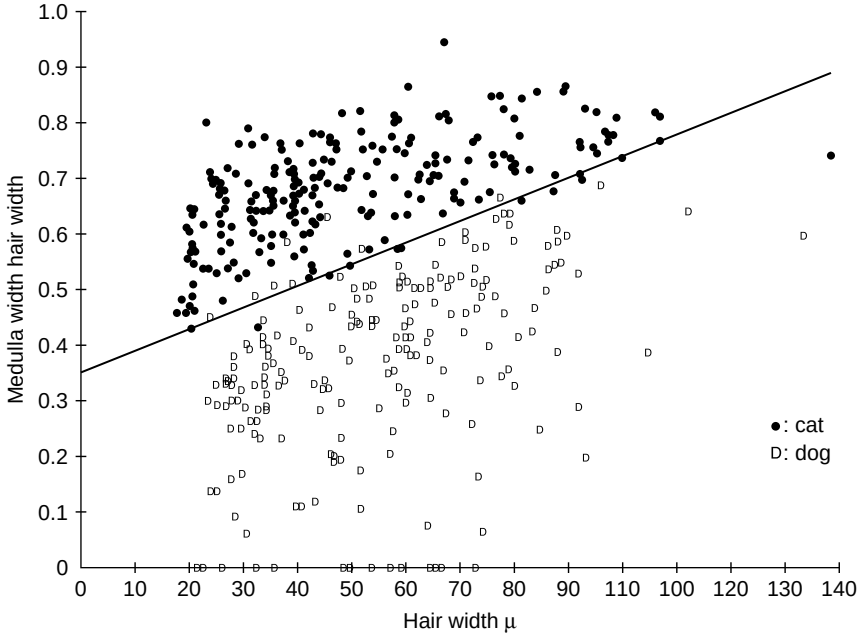


Figure 22 Medullary fractions (indices) derived from cat and dog hairs

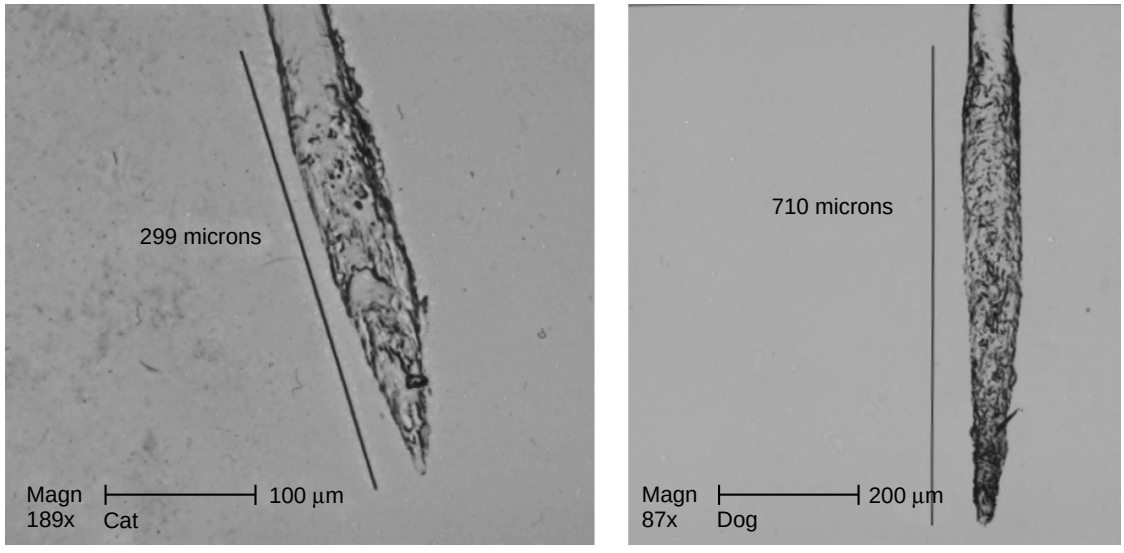


Figure 23 Spade-shaped roots depicted by cat (a) and dog (b). Primary guard hairs as viewed by the scanning electron microscope

Acknowledgment

My gratitude to the publishers and individuals for granting permission to reproduce their work, in particular, Dr Hans Brunner and John Terlet. I would also like to thank Ms Tahnee Dewhurst, Steve Fowler (VPFSC), and Liz Brooks (AFP) for their time and patience in assisting with the figures.

References

[1] Brown, S.E. & Erickson, N.E. (1978). A chronology of hairs and fibres as evidence in Canada, *Canadian Society Forensic Science Journal* **11**, 185.

[2] Gerber, S.M. & Saferstein, R. (eds) (1997). *More Chemistry and Crime – From Marsh Arsenic Test to DNA Profile*, American Chemical Society, Washington, DC.

[3] Taylor, A.S. (1894). *The Principles and Practice of Medical Jurisprudence*, 4th Edition, Churchill, London.

[4] Glaister, J. (1931). *A Study of Hairs and Wools (Belonging to the Mammalian Group of Animals, Including a Special Study of Human Hair, Considered from the Medico-Legal Aspect)*, Misr Press, Cairo.

[5] Wildman, A.B. (1954). *The Microscopy of Animal Textile Fibres*, Wool Industries Research Association (WIRA), Leeds.

[6] Appleyard, H.M. (1978). *Guide to the Identification of Animal Fibres*, Wool Industries Research Association (WIRA), Leeds.

[7] Mathiak, A.A. (1938). A key to hairs of the mammals of Southern Michigan, *Journal of Wildlife Management* **2**, 253.

[8] Mayer, W.V. (1952). The hair of California mammals with keys to the dorsal guard hairs of California mammals, *The American Midland Naturalist* **48**, 480.

[9] Moore, J.E. (1978). A key for the identification of animal hairs, *Journal of the Forensic Science Society* **28**, 335–339.

[10] Brunner, H. & Coman, B. (1974). *The Identification of Mammalian Hairs*, Inkata Press, Melbourne.

[11] Menotti-Raymond, M., David, V.A. & O'Brien, S.J. (1997). Pet cat hair implicates murder suspect, *Nature* **386**, 774.

[12] Menotti-Raymond, M., David, V.A., Stephens, J.C., Lyons, L.A. & O'Brien, S.J. (1997). Genetic individualization of domestic cats using feline STR loci for forensic applications, *Journal of Forensic Science* **42**, 1039.

[13] Savolainen, P., Rosen, B., Holmberg, A., Leitner, T., Uhlen, M. & Lundeberg, J. (1997). Sequence analysis of domestic dog mitochondrial DNA for forensic use, *Journal of Forensic Science* **42**, 593.

[14] Savolainen, P. & Lundeberg, J. (1999). Forensic evidence based on mtDNA from dog and wolf hairs, *Journal of Forensic Science* **44**, 77.

[15] Robertson, J. (2003). Transfer and persistence of hairs, in *The Forensic Examination of Hairs and Fibres*, I. Freckleton & H. Selby, eds, The Law Book Company, Sydney, Chapter 88, p. 4057.

[16] D'Andrea, F., Fridez, F. & Coquoz, R. (1998). Preliminary experiments on the transfer of animal hair during simulated criminal behaviour, *Journal of Forensic Science* **43**, 1257.

[17] Robertson, J., Brooks, E., Boehme, A., Robertson, K. & McNevin, D. (2008). Recent trends in the forensic examinatin of hairs, *Global Forensic Science Today Scientific and Investigation Techniques* **4**, 8–18.

[18] Petraco, N. (1987). A microscopical method to aid in the identification of animal hair, *Microscope* **35**, 83–92.

[19] Peabody, A.J., Oxborough, R.J., Cage, P.E. & Evett, I. (1983). The discrimination of cat and dog hairs, *Journal of the Forensic Science Society* **23**, 121–129.

[20] Hicks, J.W. (1977). *Microscopy of Hairs – A Practical Guide and Manual*, Federal Bureau of Investigation, Washington, DC.

[21] Tridico, S.R. (2004). Hair of the dog: a case study, in *Trace Evidence Analysis – More Cases in Mute Witnesses*, M.M. Houck, ed, Elsevier Academic Press, Burlington, pp. 27–52.

[22] Suzanski, T.W. (1988). Dog hair comparisons: a preliminary study, *Canadian Society Forensic Science Journal* **21**, 19–28.

[23] Suzanski, T.W. (1989). Dog hair comparison: purebreds mixed breeds, multiple questioned hairs, *Canadian Society Forensic Science Journal* **22**, 299–309.

SILVANA R. TRIDICO

Hair: Microscopic Analysis

Introduction

Hairs are a ubiquitous trace material in forensic investigation. Few trace materials give rise to such passionately held and differing views as to their value as evidence. Hair cannot be ignored as it is almost invariably present at a crime scene and there are (still) many laboratories conducting some level of hair examination. The aim of this short article is to describe the basics of hair examination, with what can, and what cannot, be safely concluded when examination of hairs is conducted according to appropriate standards, and finally to describe the role of DNA testing of hairs as part of that protocol.

Hair Biology and Physiology

The biology and physiology of hairs has been comprehensively studied by scientists from many areas of the biological and other sciences where most of our

knowledge of hair growth, basic structure, and genetic control comes from nonforensic-based research. As some hairs, notably wool from sheep, are used as textile fibers, there is a large body of literature dealing with sheep hair, from which much of our broader knowledge is derived.

Hair is only found in the tissue of mammals. The living part of hair is the hair follicle that develops as an invagination of the epidermis or skin. The development of the hair follicle has been extensively studied and reported on (see [1] and Figure 1). A basic understanding of this structure is necessary to

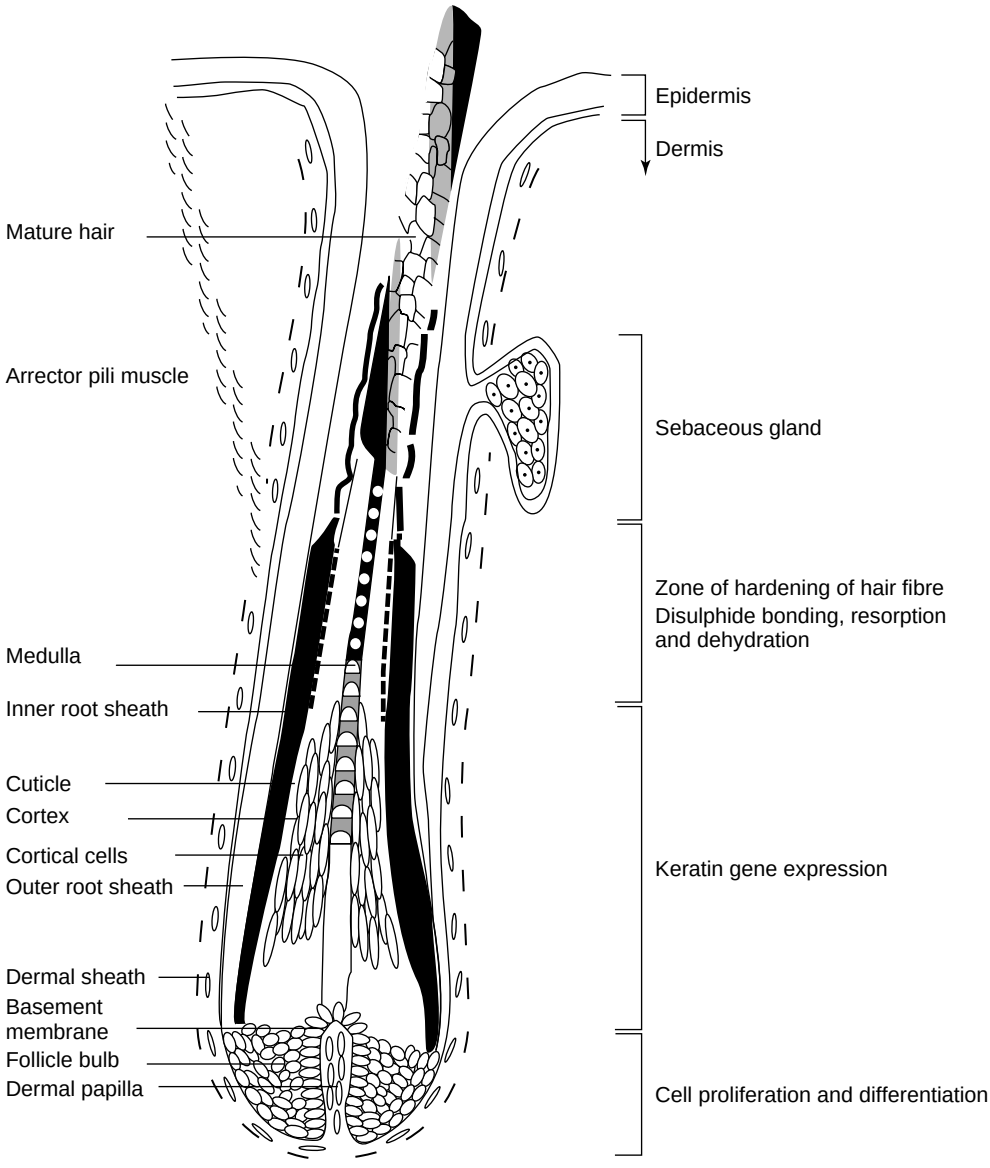


Figure 1 Schematic diagram of the pilosebaceous unit showing various main features and the regions where the main events of cell proliferation and keratinization take place. The cells in the bulb region move into different streams that give rise to the cortex, the cuticle, the medulla, and the inner root sheath [Reproduced with permission from Birkhauser Verlag Basel. © 1997.]

appreciate the various stages of hair growth and, from a particular forensic interest, hair "shedding". The basal part of the hair follicle, or bulb, can be considered to have two regions: a lower region of actively dividing cells and an upper region in which the cells differentiate and form the various cell types that comprise the mature hair shaft. The hair is quite firmly anchored in the dermis by two layers of sheath cells.

Hairs are composed of a protein called *keratin* and this soon hardens in the hair shaft through the formation of disulfide bonds. The hair shaft beyond the basal area is a dead structure and hair fibers are robust and resistant to degradation.

All five million hair follicles in a human are formed during fetal development. These follicles can produce different types of hair during the lifetime of a human. Before birth the follicles produce "lanugo" hairs. These are sometimes seen in a premature baby but are generally replaced at about the eighth month of pregnancy by short, fine "vellus" hairs. The main hairs of interest to forensic scientists are "terminal" hairs. These include primary and secondary hairs. The latter, sometimes called *sexual hairs*, develop during puberty and are subject to hormonal control.

The terminal hairs of humans are equivalent to the underhairs, fur, and wool of other mammals. Typically, a human will have in the order of 100,000 scalp follicles and about 2 million over the whole body area. These are all present at birth!

The Hair Cycle

The human hair cycle is a continuous process. Unlike with other mammals, where many hair follicles are at the same growth phase in a periodic fashion (hence, molting seen in domestic animals), human hair follicles are independent of each other. For convenience, the hair growth cycle is broken up into three phases, "anagen", "catagen", and "telogen" (Figure 2).

There is considerable variation in the reported duration of the anagen phase, but it is at least around 3 years and may be as long as 8 years for scalp hair. As scalp hair grows at about 1 cm per month, in theory a single growth cycle could result in hair of 1 m in length. As there are documented examples of hair being longer than this, hairs may go through more than one growth cycle.

The catagen phase occupies a relatively short period of time during which the follicle regresses. The telogen, or resting phase, can then last for 3–4 months. Telogen hairs lack the hair sheath cellular "anchor" of the anagen hair and are consequently easily removed during normal "growing" activities.

Although between 80 and 90% of scalp hairs at any point in time are anagen hairs and 10–20% telogen, "most" (95%) of hairs recovered in forensic work are telogen hairs because these are loosely held and more easily dislodged. Only 1–2% of hairs are catagen hairs as this is a short transition phase. The length of these growing phases,

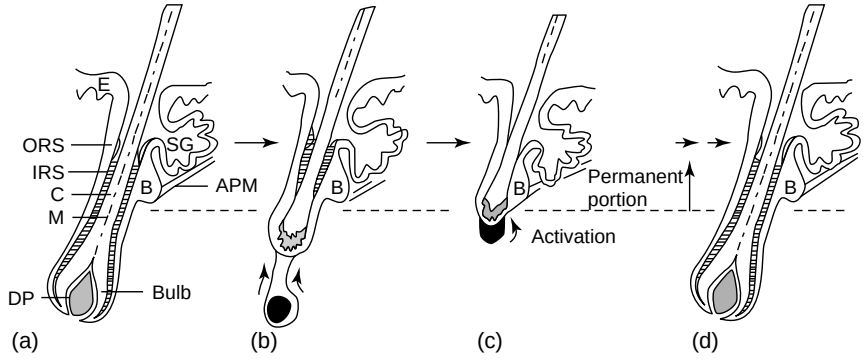


Figure 2 Diagrammatic representation of hair follicles at different stages of the growth cycle. Anagen is the active growth phase during which follicle development takes place and the hair fiber is produced. Catagen is the regression phase in which tissue changes occur as the follicle approaches telogen, the resting phase. The next anagen is thought to be initiated by an interaction between dermal papilla cells and pluripotent stem cells located in the bulge region. B, bulge; APM, arrector pili muscle; SG, sebaceous gland; C, hair cortex and cuticle; M, medulla; DP, dermal papilla [Reproduced with permission from Elsevier. © 1990.]

the percentage of hairs in each phase, and hair growth rates for other types of human hair are different [1].

Hair Transfer and Persistence

Hairs are not generally "shed" but are usually removed by normal everyday activities such as combing or brushing hair. This primary loss leaves hairs on clothing, bedding, car seats, and so on. Most recovered hairs are the result of a secondary transfer from the latter types of surfaces. Once transferred, hairs are not static and redistribute. Hence, great care needs to be taken in reading too much into the location of recovered hairs, although in some cases this is critical information.

Work on trace materials, especially fibers, has demonstrated a classic decay curve with a very high initial loss of trace material followed by a much slower loss over time. Until recently, hairs were thought to behave in a similar way. Work by Dachs *et al.* [2] and Boehme *et al.* [3] have shown that for both human and animal hairs, there is an initial loss but hairs are not lost as quickly as fibers, and the loss follows a more linear decay pattern. This is especially so with coarser or more open structured wool and acrylic recipient fabrics where the hairs appear to become physically "trapped".

Hair Morphology and Microscopic Structure

During the differentiation phase in the growth of a hair, three distinct cellular components are formed:

- the central "medulla" or core;
- the main body of the hair or "cortex"; and
- the outer "cuticle" comprising overlapping scales.

The development of these cell types in the follicle is described in Harding and Rogers [1]. Cortical cells are elongate or fusiform (spindle shaped) and have a complex chemistry based on filaments with different forms of keratin in macro- and microfibrils. In human hairs cortical cells are 80–100 μm long and 5–10 μm at their widest point. The total width of human scalp hair varies between 20 and 120 μm.

Hairs are visually and microscopically colored due to the presence of "melanin" pigment granules. There are two main types of melanin pigment, "eumelanin", which are dark brown to black in color, and "phaeomelanin", which are reddish yellow. All naturally colored hair contains various proportions of these two pigment types. Melanin synthesis takes place in melanosomes that are subcellular particles present in cells called *melanocytes*. These cells secrete the pigment granules into the cortical cells as they develop in the cells at the apex of the dermal papilla in the hair root. Pigment granules can be found in the medulla, but are not usually found in the cuticle.

The medulla comprises loosely packed cells with an open structure. In human hairs there are periods when the medulla is not produced in the follicle resulting in the medulla being absent from the hair shaft. The overall appearance of the medulla in human hair is amorphous or lacking a well-defined structure. When present it is often not continuous and it occupies less than one-third of the shaft diameter. By contrast, in nonhuman animal hairs, the medulla is most often continuous and broad, occupying more than one-third of the hair shaft. The width of the medulla over the width of the hair shaft is termed the *medullary index* (MI). In nonhuman hairs the medulla cellular structure usually has a well-defined structure that can be easily described and classified. Typical medulla types include multiserial ladder and wide aeriform lattice (Figure 3).

Finally, the cuticle forms the outer layer of the hair shaft. The cuticle is composed of flattened and overlapping (imbricate) scale cells. The scales slope outward with their edges pointing toward the tip of the hair. Although a single layer of cells in the follicle, they overlap such that, in human hairs, the effective thickness is about six cells. In nonhuman hairs the effective cell thickness can exceed 10 cell layers. The appearance of these overlapping scales, when viewed microscopically, can be described and used to differentiate animal hairs. All human scalp hairs have the same basic imbricate or flattened scale pattern and this feature cannot be used to differentiate between human hairs.

The Forensic Process

Hicks [4] describes the process of hair examination as a "procedural forensic taxonomy", based on the questions below:

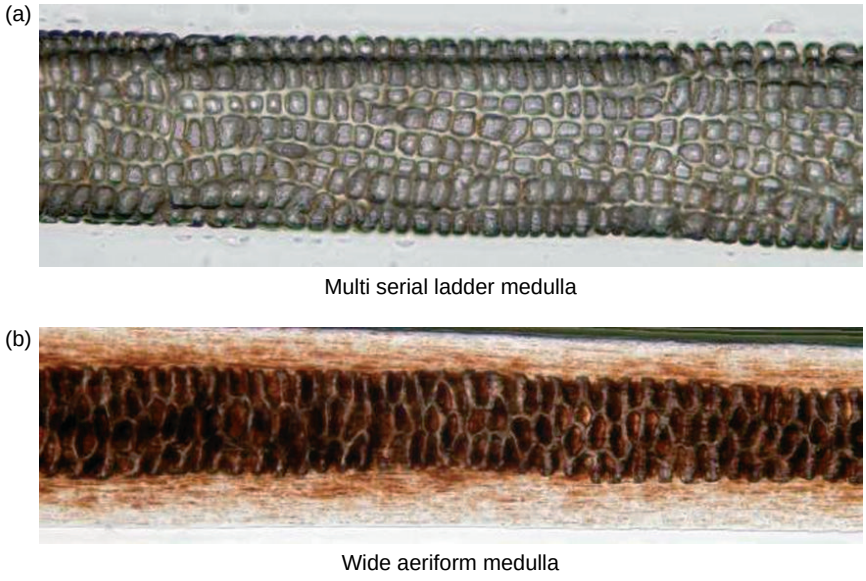


Figure 3 Examples of animal hair medulla: (a) multi serial ladder medulla and (b) wide aeriform lattice medulla

1. Is the object a hair?
2. If it is, is it human or animal?
3. If it is an animal hair, what kind of animal did it come from?
4. If it is a human hair, the hair is examined for the following information:
   (a) somatic area of origin (body area);
   (b) estimation of ancestry (race, ethnicity);
   (c) evaluation of root (phase of growth, presence of cellular material);
   (d) evaluation of the tip;
   (e) other damage, disease, or treatment; and
   (f) assessment of the hair's suitability for microscopic comparison.

To answer these questions, typically the forensic examiner has one or more (often many) hairs that have been "recovered" from a crime scene or following examination of clothing or other items. It is important, if the hair examiner has not personally recovered these hairs, that the examiner is fully aware of all information which may be relevant to assessing evidential significance. Hair examination is "not" merely an exercise in individualization.

These "recovered" hairs become the "questioned" hairs in the forensic process. The specific location of recovered hair(s) may be the critical factor. In some case situations, it may be sufficient to confirm that the recovered material are indeed hairs, and then whether or not they are human or nonhuman. In the absence of exemplar or "known" hair samples, it may be possible to individualize a recovered hair through DNA testing.

Routinely, the forensic hair examiner has samples of hair that have been collected from a deceased person, a victim of an alleged crime, from one or more suspects, or from other relevant persons for elimination purposes. For animal hairs, the known sample may come from an animal relevant to the alleged circumstances of the case.

It is essential that known samples meet appropriate standards in that there are sufficient numbers and the hairs are representative of any variation in the donor. Too often, known samples do not meet these basic requirements. For human scalp hair, the recommended number is in the order of 100 hairs. For human pubic hairs, which are less variable, in the order is 40–50 hairs. A mix of anagen and telogen hairs is preferable. For this reason, combing is the best technique as it yields more telogen hairs typical of what is seen in recovered hairs. Pubic areas of victims should first be combed to recover loose hairs which may be extraneous and of evidential value. Hairs should only be cut as a last resort and this must be at skin level. For animal hairs, where identity is

the issue, laboratories should have relevant reference sample collections.

A subset of the known hair sample is selected on the basis of visual examination, and, sometimes, after examination using a low magnification microscope with incident illumination.

At this level of examination it is usually possible to separate nonhuman and human hairs, but this needs to be confirmed by microscopic examination of mounted hairs. For nonhuman hairs, the microscopic scale features are very important in identifying an animal source. It is usual to take a caste or impression of the scale pattern before mounting animal hairs for microscopic examination. For human hairs, the subset of selected hairs (usually 10 hairs) are mounted by placing individual hairs onto glass microscope slides with a permanent mountant. There are several suitable mountants. The key factor is that the mountant should have a refractive index (RI) of 1.52–1.54, similar to the RI of the hair cortex at 1.55.

Table 1 lists the features of human *versus* nonhuman hair.

The subsequent examination of animal hairs involves systematically and thoroughly describing the microscopic detail at low magnification and higher magnification microscopy. Features assessed include hair shape or profile, cuticle and scale features, medulla appearance, and color. The most common animal hairs seen in forensic laboratories are domestic cat and dog hairs. These can be surprisingly difficult to differentiate. The easiest and surest method to distinguish cat from dog hair relies on measuring the MI and plotting this against hair width (see [5]).

For human hair, the forensic examination process starts with a detailed examination of mounted hairs using incident illumination and a low magnification ($\times 10$ to $\times 40$), stereomicroscope. At this magnification, features assessed include hair length, the shaft profile, the visual color, and the appearance of the root and tip. The literature contains numerous schemes to describe these features. The author's scheme for human hair features is included as Appendices 1 and 2. This classification of hair features is discussed by Robertson [5].

At the stereomicroscope level, the examiner is usually able to comment on somatic origin (body area) and may be able to estimate ancestry. Neither of these categories is as clear cut as one might expect as there are intermediate hair types and we live in a world of increasing mixed ancestry.

The most important element of the stereo examination of questioned hairs is to assess their suitability for "routine" nuclear (n)DNA testing. The critical requirement is to be able to accurately identify telogen hairs as these are "not" suitable for routine nDNA testing. Anagen hairs are suitable for nDNA testing and some success may be obtained with early stage catagen hairs. A common mistake made by inexperienced examiners is to select hairs on the basis of attached cellular material remaining from the hair follicle. Often telogen hairs have a small tag of cellular material but this does not contain DNA. Figure 4 shows hairs suitable and not suitable for nDNA testing. Recent work in any laboratory has shown that it may be possible to obtain in DNA from some telogen hair when a sufficient number of nuclei are present (Brooks, personal communication).

The appearance of the tip end may also yield important information including the individuals grooming habits and, sometimes, health condition.

Table 1 Human compared with nonhuman hair features

Feature	Human	Nonhuman
Color	Relatively consistent along shaft	Often showing profound color changes and banding
Cortex	Occupying most of the width of shaft – greater than medulla	Usually less than the width of medulla
Distribution of pigment	Even, slightly more toward cuticle	Central or denser toward medulla
Medulla	Less than 1/3 width of shaft amorphous, mostly not continuous when present	Greater than 1/3 width of shaft continuous, often varying in appearance along shaft, defined structure
Scales	Imbricate similar along shaft from root to tip	Often showing variation in structure along shaft from root to tip

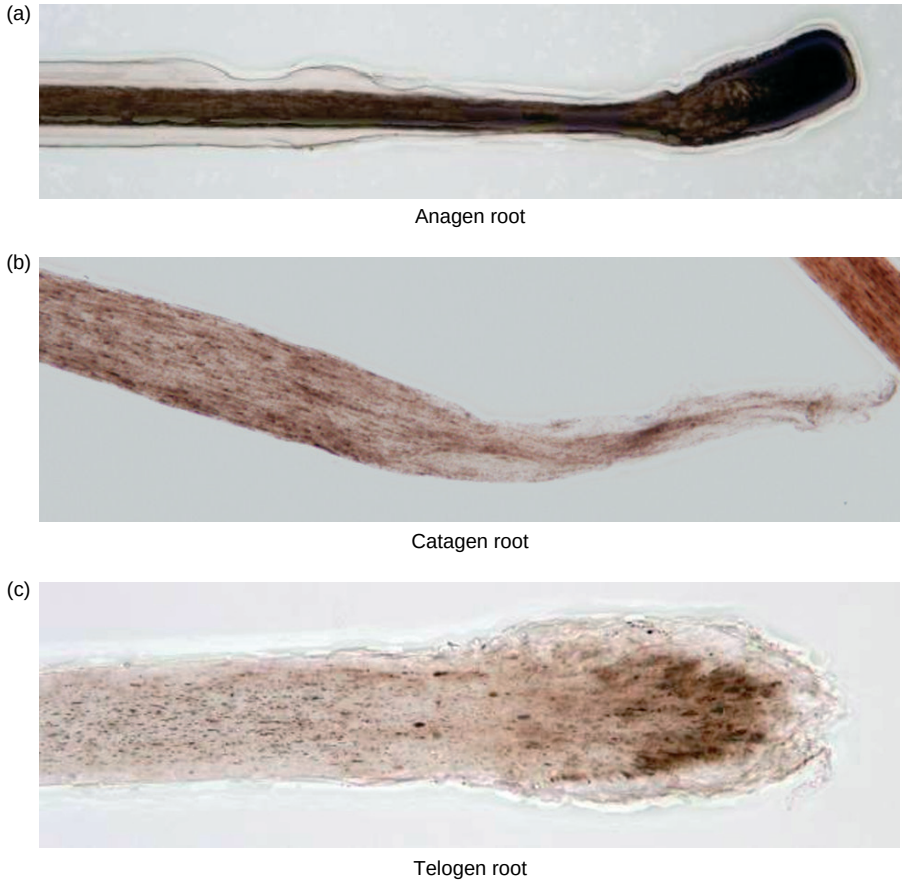


Figure 4 Hair root growth phases: (a) anagen root, (b) catagen root, and (c) telogen root

Color is an important feature initially assessed at stereo level. Here, an early indication of artificial coloring is often visible.

In the forensic process, examination may not be necessary beyond the stereo microscope level. It is critical to emphasize that hair cases require careful management focused on what are the aims of the examination process. The complexity of cases obviously varies but the fundamental aim is consistent, that is, to answer the question as to whether or not a questioned hair could have originated from a nominated individual. Provided a good quality known hair sample is available, the examiner can determine the variation present in that sample and the variation between samples. If a questioned sample is different from a known sample, then it can be eliminated. Simple as this sounds the reality is a little more complex. This is because no two hairs are identical in every respect and too often the known samples have limitations. Nonetheless the essential, and critical element, is that the examination focuses on meaningful differences and not on similarities. A similarity is a small difference!

If a questioned hair cannot be eliminated at the stereo level, the forensic process moves to the use of transmitted light microscopy with a normal total magnification range of $\times 100$ to $\times 400$. At this level it is possible to see the internal features of the hair including the medulla (when present), the cortex and in particular, the pigmentation and other features such as cortical fusi and ovoid bodies. Because hairs are essentially round (there is some racial variation in cross sectional shape), it has traditionally proved very difficult to produce photographic images that accurately convey what is seen by the examiner. The examiner constructs a mental three-dimensional

image of the hair by focusing through the hair. The recent use of automontage techniques, which allows the examiner to build up an image of an optical section of the hair, has enabled the examiner to now produce exceptional "in focus" images of hair features (Brooks, 2007). These images can be used for numerous purposes and hold the future promise that it may be possible to conduct numerical and objective assessment of the microscopically observable features.

At this time, hair examiners rely on personal judgment to "call" a particular feature. It is beyond the scope of this article to attempt to describe the subclassification of the detailed microscopically observable features (see [5]). Suffice to say that many of the features examined deal with pigmentation and how it is organized, or presents, to the examiner. Good practice must include the examiner making a detailed, systematic, and accurate record of the features seen. The use of data sheets as shown in Appendices 1 and 2 promotes this objective. However, it would be an inappropriate use of such a data sheet to expect that they could be "overlayed" to show a "match". A questioned and known hair "must" be compared side by side using a "comparison" microscope (see Figure 5). This allows the examiner to compare the overall pattern of the hairs and the precise detail at comparable points along the hair shaft. If there are no meaningful differences the examiner moves from the area of elimination to that

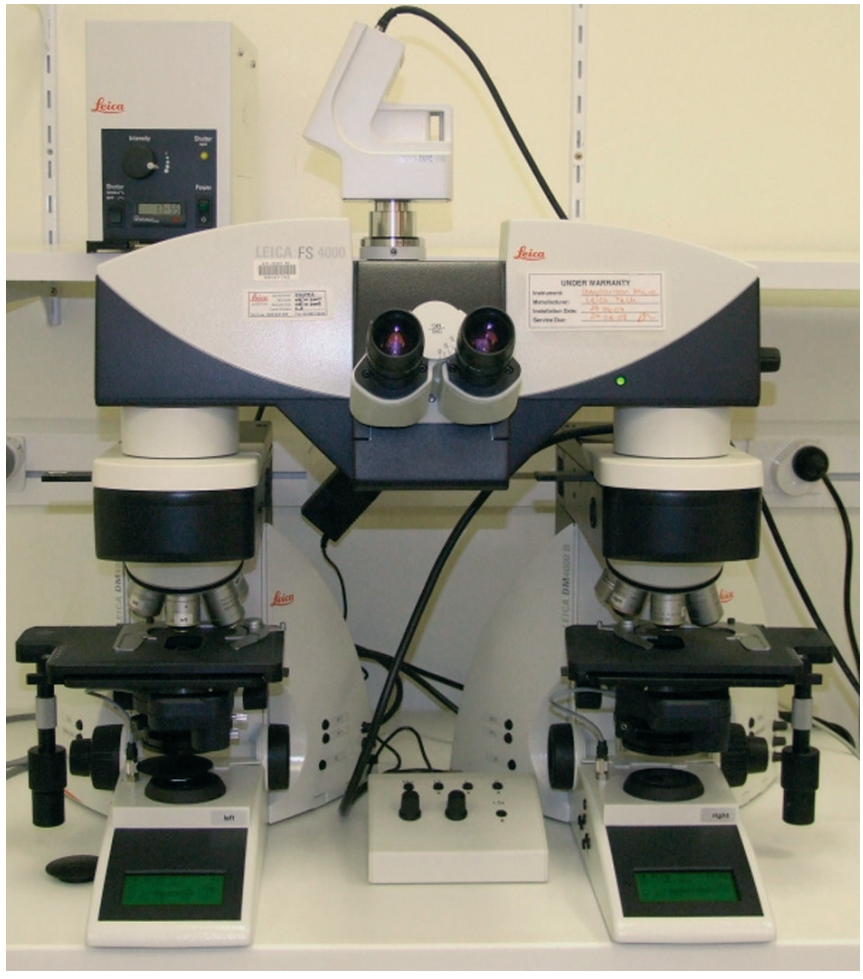


Figure 5 Comparison microscope

of inclusion. What does inclusion mean in forensic context and how far can a hair examiner go to individualize a hair?

Hair Individualization

As stated above, the primary focus of the hair examinations should be elimination. It is often possible to say that a questioned hair could not have come from a nominated individual(s). It is important to recognize that in many case situations, there are many recovered hairs and possibly many donors. It is often not possible to "assign" every recovered hair. By eliminating obviously different hairs the examiner can focus on the hairs that fall within the range of one or more known samples. Conceptionally hairs should be considered as "packages of information". A relatively colorless hair, or indeed an opaque black hair, have very low observable information. A hair with clear and well-defined pigmentation and other microscopically observable features would have a much higher information content. Many hairs are obviously different at the stereo microscope level, but a higher level of discrimination can be achieved with higher magnifications and transmitted light microscopy. If no meaningful differences are seen when two well-featured hairs are compared at high magnification this is a meaningful inclusion. However, it has to be accepted that the conclusions drawn by the examiner are subjective and result from judgment borne of experience. The word "experience" here fails to convey what is really meant. Someone could have 20 years experience and be a hopeless examiner for a whole spectrum of reasons. Experience, regrettably, is not always a key indicator of competence. Nonetheless, it is also inescapable that for the competent examiner it is the experience of simply dealing with numerous hair examinations, which assists in assessing the significance of a hair comparison.

Gaudette and colleagues have attempted to assist forensic examiners to place some numerical weight on their conclusions (Gaudette, 1999). Ignoring the possibility of an inconclusive result, Gaudette points out that there are two possible conclusions that can arise from a hair comparison:

(a) that the unknown or recovered hair could have come from the same person from whom a known hair sample was obtained or
(b) that the recovered hair could not have come from such origin.

(The author notes that he used "could have come from" and not "is consistent with" as used by Gaudette.)

Similarly there are two possible states of reality:

(1) the hair did originate from the same person as the known sample or
(2) the hair did not originate from the same person as the known sample.

If the true state of nature is (1) and conclusion (a) is reached, then the examiner is correct. If the true state of nature is (2) and conclusion (b) is reached, then the examiner is also correct. However, if the truth is (1) above but the examiner concludes (b), then an error is committed. The examiner has wrongly eliminated a person as being the source of the recovered hair. Gaudette refers to this as a "Type I error".

The final possible combination is that the truth is (2) above but the examiner concludes (a). This is a "Type II error" and has far more serious consequences because the examiner has wrongly included a person as the possible source of a hair.

Because the implications of a type II error are so serious, hair examiners have always tended toward a conservative approach to minimize this type of error, probably at the cost of an increase in type I errors.

When it is stated that a recovered hair could have come from a known source there are two possibilities: either the hair actually originated from that person "or" there was a coincidental match. It is well recognized in the scientific literature that it is possible for two different people to have hairs that are indistinguishable by present methods. Thus, coincidental matches occur. Gaudette aimed to quantify this possibility, but the interpretation of his figures is greatly debated. Many observers accept these studies do show that the chance of a coincidental match can be quite small. It is not possible to quantify this figure for general application because it depends on a number of factors. Paramount is the ability of the examiner to discriminate between hairs. For any examiner, the characteristics of the individual hair must be considered. Clearly, as previously stated, hairs with well-defined pigment and other features contain much more information potential than relatively colorless hairs.

Any reasonable hair examiner is willing to admit that hairs are not unique and no sensible examiner allows himself or herself to quote a statistical figure.

The advent of nDNA analysis means that in some instances it is possible to individualize and give a statistical estimate of the likelihood of a hair having come from a nominated individual. As previously stated "normal" mtDNA analysis is only possible with hairs in the anagen or early catagen growth phase.

Several studies have looked at extending nuclear DNA testing to telogen hairs [6] with some studies reporting limited success with telogen hairs using modified DNA analytical techniques. My own group are of the view that the DNA being detected is not inside (endogenous) the hair but is rather on the outer layer of the hair (exogenous) and that it may be better to view "hairs" as a surface on which trace DNA may be found. Regardless of the above debate, the current situation is that success rates with genuine telogen hairs are quite low and the profiles obtained are often incomplete and difficult to interpret.

It is also possible to analyze for mitochondrial (mt) DNA. There are only a small number of laboratories worldwide offering mtDNA testing which, for hairs, is time consuming and quite costly. The technique requires 1–2 cm of hair, preferably with the root present. Because of its cost and the fact it is a destructive technique, hairs should "always" be first subjected to microscopic examination from an elimination perspective. Even if a mtDNA result is obtained, it is not possible to separate brothers and sisters from their mother due to maternal inheritance of mtDNA. Statistical databases for mtDNA remain relatively small and some courts have only accepted mtDNA as having exclusionary value.

A recent study of microscopic hair comparisons and subsequent mtDNA analysis has confirmed the value of a dual approach of microscopy followed by mtDNA testing. Not all hairs are suitable for meaningful microscopic comparison and mtDNA testing is especially useful in these instances [7].

From time to time other nonmicroscopic approaches to hair examination have been proposed. Over the decades these have included elemental analysis using neutron activation analysis (NAA), protein analysis, ABO typing [5] and, most recently, it has been suggested that isotope ratio mass spectrometry (IRMS) may yield useful information. The scientific basis for the latter is that the hair is an environmental monitor and that isotopes reflect the geographic history of the donor.

None of these approaches has, or will, replace the core position of microscopic examination.

Admissibility and Evidential Value of Hair Comparisons

Rules of evidence and admissibility vary according to country and legal system. However, the Frye Test and subsequent Daubert Test in the United States have provided a useful framework to discuss whether or not hair examinations can meet relevant standards. Houck *et al.* [8] recently discussed these issues and concluded that hair comparisons can be reliable when such comparisons are conducted by a competent examiner working under an appropriate quality assurance framework. The Scientific Working Group for Hairs (SWGhair) continues to develop recommendations for such standards and for the training of hair examiners. Houck and Bisbing [9] have discussed what might constitute an appropriate training program. This must include the evaluation of testimony skills and comprehensive competency testing. The latter can be problematic given the lack of good quality and appropriate external proficiency tests for hair.

An area of concern to many forensic scientists is the apparent willingness of courts to accept scientists as qualified to give expert-opinion evidence "without" a rigorous examination of their qualification and relevant experience to offer an opinion. A difficult question in the area of hair examination is whether there is a relevant scientific community. Houck *et al.* [8] argue that there is a long history of scientific research in comparative biology and physical anthropology dating back to the eighteenth century. While this may be true, the difficulty faced by lawyers today may be in finding an expert who has experience and credibility.

Concluding Remarks

In many cases hairs are only one of a number of trace materials recovered. Often the condition of such hairs is important. For example, where a person has been assaulted by a particular type of implement or weapon. Sometimes

the number and location of hairs can be useful in attempting to reconstruct events. As demonstrated above, where hairs are not examined, this can leave unanswered questions that make it more difficult for juries to determine the circumstances and facts of a case.

The reality is that hairs are a very common, if not the most common, trace material encountered

Appendix I

Australian Federal Police

Forensic Services, laboratory services Page of

Examination record – sheet 1 Forensic case no:

Item no:

Hair number										
Macroscopic feature		r	t	r	t	r	t	r	t	r	t
Approx. length	cm									
Shaft profile	Straight Wavy Curly Peppercorn									
Colour[1]	Colourless Yellow Brown Reddish Black									
Root	Absent Anagen Catagen Telogen Sheath/follicular tag									
Tip	Natural taper Cut Rounded Frayed or abraded Split Crushed or broken Singed									

General description and comments:

[1]A assessed with a stereo microscope using standardised illumination

B basic colour to be qualified by shade or depth of colour. Light (L), Mid (M) or Dark (D)

C note artificial colouring

Examined by: _____

Notes by: Day: Date: Time:

F-LAB-BC-091-V02

Appendix II

Australian Federal Police

Forensic Services, laboratory services Page of

Examination record - Sheet 2 Forensic case no:

Microscopic feature		Hair number 1 r	1 t	2 r	2 t	3 r	3 t	4 r	4 t	5 r	5 t
Shaft diameter	Max units at x 40										
Pigment density	None Light Medium Heavy Opaque										
Pigment Distribution (across hair shaft)	Uniform Towards medulla Towards cuticle To one side										
Pigment Aggregate Shape	Streaked Clumped oval Clumped round										
Pigment granule Shape	Fine Oval/Oblong										
Pigment Aggregate Size	Fine Medium Coarse										
Pigment Granule Size	Fine Medium Coarse										
Ovoid bodies[2]											
Medulla Distribution	None Medulla < Space Medulla > Space Continuous										
Medulla type	Opaque Translucent										
Cortical fusi[2]											
Cortical Texture	Not visible or smooth Visible or coarse										
Cuticle	Thickness units x 40 Colour										
Cuticle outer	Smooth Serrated Ragged Cracked Looped										

[2] Where present their shape, size and distribution both along and across the hair shaft may have some value when used as comparative features.

Examined By: _____ Checked By: _____

Notes by: Day: Date: Time:

F-LAB-BC-089-V01

in forensic work. To ignore them is facile and may jeopardize an investigation, and ultimately a "just" outcome.

There remains a need for a concerted research effort to provide a more solid basis for the interpretation of hair evidence. It is also time for laboratory systems and scientists that are either unable or unwilling to embrace hair examinations with the necessary investment in time and commitment to withdraw altogether from the field. Half-hearted commitment can only continue to do damage to the credibility of hair examination.

Acknowledgment

Photographs are courtesy of Elizabeth Brooks.

References

[1] Harding, H. & Rogers, G. (1999). Physiology and growth of human hair, in *Forensic Examination of Hair*, J. Robertson, ed, Taylor and Francis, London, pp. 1–77.
[2] Dachs, J., McNaught, J.J. & Robertson, J. (2003). The persistence of human scalp hair on clothing fabrics, *Forensic Science International* **138**, 27–36.
[3] Boehme, A., Brooks, E., Robertson, J., Anson, J. & McNaught, I. (2007). The persistence of animal hairs in a forensic context. *Australian Journal of Forensic Sciences* **41**, in press.
[4] Hicks, J.W. (1977). *Microscopy of Hairs – A Practical Guide and Manual*, Federal Bureau of Investigation, Washington, D.C.
[5] Robertson, J. (1999). Forensic and microscopic examination of human hair, in *Forensic Examination of Hair*, J. Robertson, ed, Taylor and Francis, London, pp. 79–154.
[6] McNevin, D., Wilson – Wilde, L., Robertson, J., Kyd, J. & Lennard, C. (2005). Short tandem repeat (STR) genotyping of telogen hair. Part 1. Review of current status and knowledge gaps, *Forensic Science International* **153**, 237–246.
[7] Houck, M.M. & Budowle, B. (2002). Correlation of microscopic and mitochondrial DNA analysis of hairs, *Journal of Forensic Sciences* **45**, 1–4.
[8] Houck, M.M., Bisbing, R.E., Watkins, T.G. & Harman, R.P. (2004). Locard exchange: the science of forensic hair comparisons and the admissibility of hair comparison evidence. Frye and Daubert considered, *Modern Microscopy Journal* (www.modernmicroscopy.com/main.asp?article=36).
[9] Houck, M.M. & Bisbing, R.E. (2005). Forensic human hair examination and comparison in the 21st century, *Forensic Science Review* **17**, 52–65.

JAMES ROBERTSON

Hair: Toxicology

Introduction

Using hair as a medium to analyze drug exposure has been receiving increased attention during recent times because of the less embarrassing circumstances of collection, its long drug detection window, and because it does not decompose like other body tissues after death.

In 1979, Baumgartner *et al.* [1] published the pioneering report on using radioimmunoassay (RIA) to detect morphine in the hair of heroin abusers. This article was followed by a great number of studies, which mostly included RIA and/or gas chromatography–mass spectrometry (GC/MS). Today, chromatographic procedures, especially those coupled to MS or tandem MS, represent the gold standard for the identification and quantification of drugs in hair, owing to their separation ability and their low limits of detection.

Hair analysis is now routinely used as a tool for the detection of xenobiotics (drugs of abuse, pharmaceuticals, environmental contaminants, doping agents, etc.) in forensic science, traffic medicine, occupational medicine, and clinical toxicology.

Biology of Hair

Hair is a product of differentiated organs in the skin of mammals composed of protein (65–95%, keratin essentially), water (15–35%), lipids (1–9%), and minerals (<1%). The hair shaft consists of an outer cuticle that surrounds a cortex. In some types of hair, the cortex surrounds a central medulla. The hair shaft develops in a follicle closely associated with the sebaceous and apocrine glands. The growth of hair occurs in cycles, alternating between periods of growth (*anagen* phase) and periods of quiescence (*catagen* and *telogen* phases). About the one million hair follicles of the adult scalp, approximately 85% of the hair is in the growing phase and the remaining 15% is in a quiescent stage. Hair is produced during four to eight years for head hair (<6 months for nonhead hair) at a rate of approximately 0.22–0.52 mm per day or 0.6–1.42 cm per month [2] for head hair. The

growth rate depends on the type of hair, physiological factors, and anatomical location.

The exact mechanism by which chemicals are bound in hair is not known, but it is generally considered that xenobiotics can enter into hair by at least three mechanisms: from the blood during hair formation, from sweat and sebum, and from the external environment.

Collection procedures for hair analysis of drugs have not been standardized. In most published studies, the samples are obtained from random locations on the scalp, but hair is best collected from the area at the back of the head, called the *vertex posterior*. Compared with other areas of the head, this area has less variability in the hair growth rate, the number of hairs in the growing phase is more constant, and the hair is less subject to age- and sex-related influences. Pubic hair, arm hair, and axillary hair are possible alternative sources for drug detection when scalp hair is not available.

Hair Analysis

The most crucial issue facing hair analysis is the avoidance of false-positive results caused by passive exposure in the environment to the drug, causing it to be present on the surface of the hair rather than incorporated in it. In most laboratories, hair analysis starts with a wash step to remove external contamination.

Hair analysis involves at least five steps:

1. decontamination of the hair
2. preparation of the hair: pulverization, segmentation in short pieces
3. incubation: in methanol, acid, sodium hydroxide, buffer
4. extraction: liquid/liquid, solid phase, solid phase microextraction
5. analysis immunoassay screening (ELISA) and/or chromatography (gas, liquid) coupled to MS (tandem).

Cut-off concentrations from the Society of Hair Testing (SoHT) and expected concentrations in hair for drugs of abuse are presented Table 1.

These cut-offs were established to avoid false-positive results due to external contamination. They are used both for ELISA screening (enzyme linked immunosorbent assay q.v.) and chromatographic confirmations in workplace drug testing.

Table 1 SoHT cut-off concentrations (when tested by GC/MS) and expected concentrations for drugs of abuse in hair

Drug	GC-MS cut-off concentration	Expected concentrations
Heroin	0.2 ng mg^{-1} of 6-acetylmorphine, morphine	0.5–100 ng mg^{-1}, in most cases <15 ng mg^{-1}
Cocaine	0.5 ng mg^{-1} of cocaine and 0.05 ng mg^{-1} of benzoylecgonine and cocaethylene	0.5–100 ng mg^{-1}, in most cases <50 ng mg^{-1}, in crack abusers >300 ng mg^{-1} is possible
Amphetamine, MDMA	0.2 ng mg^{-1} for each drug	0.5–50.0 ng mg^{-1}
Cannabis	0.1 ng mg^{-1} for THC	THC: 0.05–10 ng mg^{-1}, in most cases <2 ng mg^{-1}
	0.2 pg mg^{-1} for THC-COOH	THC-COOH: 0.5–50 pg mg^{-1}, in most cases <5 pg mg^{-1}

MDMA, methylenedioxymethamphetamine; THC, tetrahydrocannabinol; THC-COOH, tetrahydrocannabinol-11-oic-acid

In most cases, they are not applied in forensic cases, where sometimes a single exposure has to be demonstrated.

Applications

Verification of Drug Use History

By providing information on exposure to drugs over time, hair analysis may be useful in verifying self-reported histories of drug use in any situation in which a history of past rather than recent drug use is desired. During control tests of hair, a drug user is not able to hide their drug abuse, as the window of detection is often very long, i.e., months. Urine or blood tests are not useful to detect drug use when drugs are consumed every few days even when the tests are repeated [3].

The advantages and disadvantages of drug detection in urine and hair are presented in Table 2.

Hair analysis can also provide a retrospective calendar of an individual's drug use. For this, multisectional analysis is required and involves taking a length of hair and cutting it into sections to measure drug use during shorter periods of time. The hair must be cut as close as possible to the scalp and particular care is also required to ensure that the individual hairs in the cut-off tuft retain the position they originally had beside one another.

The most extensive study on sectional analysis for drugs of abuse has involved patients in rehabilitation centers. Segmental hair analysis is used to verify both their previous drug history and their recent enforced abstinence. To verify abstinence, the lowest drug concentration is found in the segments nearest the root.

The differentiation of heroin users from individuals exposed to other sources of morphine alkaloids can be achieved by identifying directly heroin or 6-acetylmorphine [4].

Hair analysis is suitable for accurately monitoring relative changes in drug intake in the same individual. In several studies, it has been shown that a doubling of drug dose in the same individual results in a doubling of the drug content of the hair, which is not the case for blood and urine drug level. Patients who received constant methadone doses under close supervision showed considerable constancy in their methadone levels in hair.

Several studies have suggested that the hair color (or the melanin content) may be the major determinant of drug binding and, consequently, may result in color bias in hair testing. Researchers have demonstrated that different hair types incorporate differing amounts of drugs when exposed under identical conditions. The higher accumulation of alkaline substances (such as cocaine or heroin) in black hair as compared with blond hair will need to be addressed (or controlled) in order not to discriminate specific population due to their hair color, resulting in inequity during investigation [5].

Table 2 Main characteristics and performance of analyses in urine and hair

	Urine	Hair
Drugs of abuse and pharmaceuticals detected[(a)]	All	All
Main compounds	Metabolites	Parent drugs
Detection time window	2–3 days	Months, years
Analytical techniques	Immunoassays, followed by chromatography/mass spectrometry	Chromatography/mass spectrometry
Specificity	Family diagnosis, then specific confirmation	Specific identification
Analysis duration	+	+ + +
Type of measurement	Incremental	Cumulative
Sample collection	±invasive	Noninvasive
Adulteration	Possible	Impossible
Preservation	−20 °C	Ambient temperature

[(a)] only doping hormones erythropoietin(EPO), growth hormoneGH etc.) are not incorporated in hair

Alcohol Abuse

Considering the large dimension of alcohol-associated problems, the diagnosis of excessive alcohol consumption is an important task from medical point of view. Markers of ethanol consumption in hair are ethylglucuronide and fatty acid ethyl esters (FAEE). Detection of ethylglucuronide in hair is always associated with alcohol consumption, whereas a negative result does not unambiguously exclude the alcohol abuse [6]. Investigations on the FAEE have been proposed by Pragst *et al.* [7] to monitor alcohol consumption. FAEE are formed in presence of ethanol and free fatty acids, triglycerides, lipoproteins, or phospholipids by an FAEE synthase found not only in liver but also in hair roots. FAEE determination is of interest as they appear responsive to alcohol-induced organ damage. Four FAEE (ethyl myristate, ethyl palmitate, ethyl oleate, and ethyl stearate) are suitable markers for the detection of heavy alcohol consumption and show differential concentrations in hair of children, adult teetotalers, and social drinkers in comparison with FAEE concentrations found in hair of alcoholics.

Verification of Doping Practices

Athletes use both endogenous and exogenous anabolic steroids because it has been claimed that they increase lean body mass, strength, and aggressiveness and lead to a shorter recovery time between workouts.

The greatest use of hair in this context may be in identifying false-negatives results in negative test results on blood or urine due to recent abstinence of a drug a few days before the competition. Hair can also indicate the history and frequency of drug intake as repetitive use can be demonstrated by segmental analysis along the hair shaft.

Unlike testosterone in urine, the interpretation of concentration findings in hair can be difficult. The range between physiological concentrations of testosterone and those found in abusers seems to be rather small. Therefore, to complement testosterone determination, the identification of unique testosterone esters in hair enables an unambiguous determination of doping because the esters are exogenous substances and therefore could not come from endogenous testosterone [8].

Another advantage of hair analysis is the possible discrimination between nandrolone and abuse of other 19-norsteroids (norandrostenedione and norandrostenediol), which lead to the same urinary metabolites (norandrosterone and noretiocholanolone). This is obviously not possible in urine, where hair can identify the exact identity of the parent compounds [9].

In the case of longitudinal surveys of athletes, hair analysis appears as the solution of choice to document doping practices.

β-Adrenoceptor agonists are banned in competition sports because of their sympathomimetic properties (stimulant effects) and their activity as anabolic agents at higher dosages. However, salbutamol is permitted only by inhalers and must be declared prior to the competition. As the drug is permitted for specific therapeutic purposes, together with a medical prescription, it is relatively easy to evade a positive urine test. Again, segmental hair analysis would document unambiguously doping by the athletes, when the drug (such as a corticoid or an anabolic compound) is found in several consecutive hair sections. This can also be controlled in the case of long-term abuse of corticoids.

Regranting of Driving Licenses

The major practical advantage of hair testing compared with urine and blood testing for drugs is its larger detection window, which is weeks to months, depending on the length of the analyzed hair shaft, against a few days for urine. There is a reasonable agreement that the qualitative results obtained from hair analysis are valid and that hair can provide long-term histories of drug use. In some countries, persons whose driving license has been refused, revoked, or suspended for addiction to psychoactive drugs or for driving “under the influence” can obtain a license after a medical committee has confirmed the actual and complete abstinence from illicit drugs by the use of hair testing and has excluded any additional risk of future relapse of drug abuse. To provide objective evidence of abstinence from drugs with an acceptable chronological window in order to support the clinical decision of this medical committee, hair analysis has been included in a panel of clinical and laboratory tests, aimed at investigating retrospectively any drug-associated behaviors of subjects. Comparison between hair analysis and urinalysis shows a much higher diagnostic sensitivity for hair tests [10].

Drug-Facilitated Crimes

The use of a drug to modify a person's behavior for criminal gain is not a recent phenomenon. However, the recent increase in reports of drug-facilitated crimes (sexual assault, robbery, incapacity, etc.) has caused alarm in the general public. Drugs involved can be pharmaceuticals, such as benzodiazepines (flunitrazepam, lorazepam, etc.), hypnotics (zopiclone, zolpidem), sedatives (neuroleptics, some antihistamines), or anesthetics (γ-hydroxybutyrate (GHB), ketamine), drugs of abuse, such as cannabis, ecstasy, or LSD, or more often ethanol. Owing to their low dosage, excepted for GHB, a surreptitious administration into beverages such as coffee, soft drinks, or alcoholic cocktails is relatively simple.

Most of these substances possess amnesic properties and therefore the victims are able to less accurately recall the circumstances under which the sexual offence occurred. As they are generally short acting and short lived, they impair an individual rapidly and are then excreted or metabolized quickly. In these situations, blood or even urine can be of little assistance in demonstrating the use of drugs. Hair was suggested as a valuable specimen in situations where, as a result of a delay in reporting the crime, natural processes may have eliminated the drug from typical biological specimens. The introduction of the use of liquid chromatography (LC)-MS/MS allows the detection of a single dose of most sedatives [11].

The discrimination between a single exposure and long-term use can be documented by multisectional analysis. In the absence of migration of the drug or its metabolite along the hair shaft, a single spot of exposure must be present in the segment corresponding to the period of the alleged event, using a growth rate for hair of 1 cm per month. As this growth rate can vary from 0.7 to 1.4 cm per month, the length of the hair section must be calculated accordingly and this variability accounted within the estimates of time of use. A delay of three to five weeks between the offense and hair collection for sectional analysis (2-cm segments) is recommended so as to analyze the hair shaft at the point of exposure. The hair must be cut as close as possible to the scalp. Particular care should also be taken to ensure that the individual's hair in the strand retains the position it originally had beside every other strand.

Conclusion

It appears that the value of hair analysis for the identification of drug users is gaining recognition. This can be seen from its growing use in preemployment screening, in forensic sciences, and in clinical applications. Hair analysis may be useful adjunct to conventional (urine and blood) drug testing in toxicology. Specimens can be more easily obtained with less embarrassment, and hair can provide a more accurate history of drug use.

Although there are still controversies on how to interpret the results, particularly concerning external contamination, cosmetic treatments, ethnic bias, or drug incorporation, pure analytical work in hair analysis has reached a sort of plateau, having solved almost all the analytical problems. Although GC and LC-MS are the methods of choice in practice, GC and LC-MS/MS are today used in several laboratories, even for routine cases, particularly to target low-dosage compounds.

References

[1] Baumgartner, A.M., Jones, P.F., Baumgartner, W.A. & Blank, C.T. (1979). Radioimmunoassay of hair for determining opiate-abuse histories, *Journal of Nuclear Medicine* **20**, 748–752.

[2] Saitoh, M., Uzaka, M., Sakamoto, M. & Kobori, T. (1969). Rate of hair growth, in *Advances in Biology of Skin: Hair Growth*, A. Montana & M. Dobson, eds, Oxford, pp. 183–194.

[3] Baumgartner, W.A., Hill, V.A. & Blahd, W.H. (1989). Hair analysis for drugs of abuse, *Journal of Forensic Sciences*, **34**, 1433–1453.

[4] Goldberger, B.A., Caplan, Y.H., Maguire, T. & Cone, E.J. (1991). Testing human hair for drugs of abuse. III. Identification of heroin and 6-acetylmorphine as indicators of heroin abuse, *Journal of Analytical Toxicology*, **15**, 226–231.

[5] Joseph, R.E., Su, T.P. & Cone, E.J. (1996). In vitro binding studies of drugs to hair: influence of melanin and lipids on cocaine binding to Caucasoid and Africoid hair, *Journal of Analytical Toxicology*, **20**, 338–344.

[6] Jurado, C., Soriano, T., Giménez, M.P. & Menéndez, M. (2004). Diagnosis of chronic alcohol consumption. Hair analysis for ethyl-glucuronide, *Forensic Science International*, **145**, 161–166.

[7] Pragst, F., Auwaerter, V., Sporkert, F. & Spiegel, K. (2001). Analysis of fatty acid ethyl esters in hair as possible markers of chronically elevated alcohol consumption by headspace solid-phase microextraction (HS-SPME) and gas chromatography-mass spectrometry (GC-MS), *Forensic Science International*, **3095**, 1–13.

[8] Kintz, P., Cirimele, V., Jeanneau, T. & Ludes, B. (1999). Identification of testosterone and testosterone esters in human hair, *Journal of Analytical Toxicology*, **23**, 352–356.
[9] Kintz, P., Cirimele, V. & Ludes, B. (2000). Discrimination of the nature of doping with 19-norsteroids through hair analysis, *Clinical Chemistry*, **46**, 2020–2022.
[10] Sachs, H. (1996). Hair analysis as a basic for driving ability examination, *Toxicorama*, **6**, 11–17.
[11] Kintz, P., Villain, M., Cirimele, V. & Ludes, B. (2004). Testing for the undetectable in drug-facilitated sexual assault using hair – analyzed by tandem mass spectrometry – as an evidence, *Therapeutic Drug Monitoring*, **26**, 211–214.

Further Reading

Kintz, P. (ed) (1996). *Drug Testing in Hair*, CRC Press, Boca Raton, pp. 1–293.
Kintz, P. (ed) (2006). *Analytical and Practical Aspects of Drug Testing in Hair*, CRC Taylor & Francis, Boca raton, pp. 1–382.
Madea, B. & Musshoff, F. (eds) (2004). *Haaranalytik*, Deutscher Artze-Verlag, pp. 1–395 (in German).

PASCAL KINTZ

Hallucinations

Introduction

Hallucinations are sensory perceptions experienced while awake with a compelling sense of reality in the absence of external stimuli. Hallucinations can involve one or more of the five primary sensory senses of hearing (auditory), vision (visual or optical), smell (olfactory), taste (gustatory), and touch (tactile). Auditory hallucinations consist of sound, usually voices, although noises and music are not uncommon. Visual hallucinations may consist of complex images such as living creatures as well as simple images such as flashes of light. Olfactory and gustatory hallucinations involve usually unpleasant false perceptions of smell and taste, respectively. Tactile hallucinations involve the sense of touch, such as in formication (the sensation of insects crawling under the skin).

Hallucinations differ from illusions, which are misperceptions or misinterpretations of objective sensory stimuli. Insight as to whether the hallucination represents a false sensory experience may or may not be present. A hallucination is a symptom or a manifestation of a mental or medical condition resulting from a variety of etiologies. Hallucinations often co-occur with delusions (*see* **Delusions**).

The Origin of Hallucinations

Research has shown that during the experience of hallucinations, the part of the brain responsible for the actual sensory perception is active. For example, when experiencing auditory hallucinations, the auditory cortex of the brain is active as if actually hearing an external sound [1].

Hallucinations have been associated with various biological causes. Pathology of the central nervous system (CNS), i.e., the brain and related structures, has been associated with the development of hallucinations. Traumatic brain injury, cerebrovascular accidents (strokes), infectious or inflammatory CNS illnesses, tumors or other space-occupying lesions, and seizure disorders have been associated with hallucinations.

With advancing age, the likelihood of developing dementia increases. Although the central feature of dementia involves cognitive impairment, visual and auditory hallucinations frequently appear in individuals suffering from the commonly seen dementias arising from Alzheimer's disease, Parkinson's disease, and dementia with Lewy bodies [2–5].

Medical illnesses that lead to metabolic disturbances, such as endocrine, vitamin deficiency, hepatic, or renal disorders that have CNS effects, can also lead to the development of hallucinations. Recreational drugs (e.g., methamphetamine, cocaine, cannabis (marijuana), mescaline, lysergic acid diethylamide (LSD), phencyclidine, psilocybin (hallucinogenic mushrooms), and 3,4-methylenedioxymethamphetamine (MDMA or Ecstasy) frequently give rise to hallucinations with their use. Alcohol has been associated with hallucinations, though the most common manifestation has been visual hallucinations associated with alcohol withdrawal syndrome. Several prescribed medications have also been implicated in the development of hallucinations, including medications prescribed to treat mental disorders [6–8].

Hallucinations can also arise in association with intense stress, sleep deprivation, and sensory deprivation. In acute grief reactions, the recently deceased

loved one may be perceived by the survivor as having communicated via auditory or optical means. Extreme psychosocial stressors, such as experiencing events involving actual or threatened death or serious injury, can lead to the development of posttraumatic stress disorder (PTSD). A principal feature of PTSD involves recurrent and intrusive flashbacks of the traumatic event that involves one or more of the primary senses.

The most common manifestation of hallucinations occurs in the context of psychosis, particularly schizophrenia and related psychoses, mood disorders with psychotic features, and substance-induced psychotic disorders [9]. Nonetheless, despite various scientific explorations regarding the biological origin of hallucinations, our knowledge at this level remains in the embryonic stages [10, 11].

Treatment of Hallucinations

When there is an identified biologic causation to the hallucinations, such as a vitamin deficiency giving rise to beriberi, the treatment consists of treating the medical condition with the vitamin thiamine. Similarly, when treating a substance-induced psychotic disorder with hallucinations, abstinence from the offending drug should, but not always, quell the hallucinations and other psychotic symptoms. Treatment of hallucinations associated with dementia, even with a known illness such as Alzheimer's disease, Parkinson's disease, or dementia with Lewy bodies, has a lower success rate in view of the limitations of the therapeutic interventions for these medical illnesses.

In treating auditory hallucinations in the cases of schizophrenia, mood disorders with psychotic features, or other psychoses, use of antipsychotic medications has resulted in reduction of auditory hallucinations, delusions, and/or other manifestations of psychosis. Nonetheless, the proportion of individuals whose hallucinations or other psychotic symptoms fail to respond to clinical intervention is not insignificant.

Command Hallucinations

When the content of hallucinations involve a command or order for the individual to act, such hallucinations have come to be known as *command* hallucinations. Command hallucinations may be the most commonly known form of hallucinations by the general public whether in the venue of a high profile criminal trial, or when family members, neighbors, or strangers are thought to be experiencing them in the community. Command hallucinations can take one of two forms, auditory or visual. Visual hallucinations involving text (known as *orthographic* hallucinations) generally are comprised of individual letters, words, or nonsense letter strings. Orthographic hallucinations generally appear in the context of medical or neurologic disease. Only one case of command visual (orthographic) hallucinations has been reported in the literature in the past quarter of a century [12].

Command auditory hallucinations alert the psychiatrist or other mental health professional of a potential for harmful or destructive acts. Individuals have been known to castrate or enucleate themselves in the presence of command auditory hallucinations [13–15]. Delusions generally accompany the command auditory hallucinations that result in such self-mutilation. Command auditory hallucinations have also been associated with homicidal violence toward others such as the case of a man with a command auditory hallucination and delusions who killed his infant daughter [16]. Nine percent of a sample of "White House cases", i.e., individuals who had been psychotically preoccupied with prominent political figures who had come to the attention of the Secret Service, experienced command auditory hallucinations [17]. Clinical research into command auditory hallucinations using sample populations and not just single case reports or a small case series dates back only about two decades to a study based on data collected from a general psychiatric inpatient service at an urban academic medical center [18]. Subsequently, samples of individuals with command auditory hallucinations from a variety of settings, inpatient and outpatient, clinical and forensic, and public and private settings, in a variety of countries (United States, United Kingdom, Australia, Germany, New Zealand, and Singapore) have since been reported in the anglophonic literature [19–30].

The limited literature on command auditory hallucinations does not find this type of hallucinations to be more prevalent in forensic than in nonforensic samples [31]. The various studies find a large range (7–70%) for the prevalence of command auditory hallucinations containing harmful content with a

median prevalence of about 40% [31]. Harmless commands are not of clinical or forensic concern, even if obeyed. Besides the harmfulness of the content of the command auditory hallucination, the extent of the individual's compliance with the command also assumes clinical and/or forensic significance. The presence of delusions and the individual's affective (emotional) state play a substantial role in whether an individual complies with a harmful command. Some recent statistical studies on command auditory hallucinations [30, 32] do not give much guidance as to who complies and only suggests that command auditory hallucinations with harmful content signal a need for careful clinical assessment.

Malingering of Hallucinations

The central feature of malingering involves the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives [[9], pp. 739–740]. The frequency of feigned psychosis is unknown [33]. Research has not yet arrived at a foolproof method to determine the presence of genuine verses manufactured psychosis.

From reviewing the literature and his clinical experience, forensic psychiatrist Philip Resnick has listed six areas of inquiry regarding the genuineness of reported auditory hallucinations. These six areas include vocal characteristics, source of the hallucination, characteristics of the hallucination, relationship to the hallucination, belief/response to hallucinations, and associated characteristics [33]. However, on closer inspection of Resnick's inquiry, Resnick's method involves comparing the individual's description of the auditory hallucinations with characteristics of auditory hallucinations described by actual patients. Although the survey of the literature on the presentation and content of auditory hallucinations shows likely characteristics as to what constitutes a genuine hallucination, the clinical research has shown considerable variability across studies. Of course, during the course of ordinary and routine clinical practice of inpatient psychiatry, individuals may misrepresent the presence or extent of hallucinations to resist or gain admission, remain in the hospital, or seek release from the hospital. Resnick uses a similar strategy to assess the genuineness of visual hallucinations, though in the case of visual hallucinations the frequent association with specific biologic causes such as medical illness and (recreational or prescribed) drugs may complicate the task of assessing reported visual hallucinations.

At this point, keeping abreast of the clinical and research psychiatric literature may be the best strategy for psychiatrists and other clinicians to know the characteristics of a typical hallucination. For example, in a recent study of individuals with visual hallucinations in the context of mental disorders, the subjects described the hallucinations as generally humanoid in content, with only a minority consisting of animals or objects. The visions almost invariably arose without warning. Most reported feeling overwhelmed and frightened, though the visions lasted only seconds or minutes [34].

The determination of the genuineness versus the manufacture of a reported hallucination often does not present as a simple dichotomous choice. Individuals who have had previously experienced hallucinations, whether from a mental disorder, medical illness, and/or recreational or prescribed drug use, have first hand knowledge of the characteristics of a hallucination, thereby complicating the task.

Probably the most frequent occurrence of manufacture or exaggeration of hallucinations involves requests for admission to psychiatric facilities, whether by individuals with no history of mental illness or by individuals who have previously experienced actual hallucinations. Also the most frequent manifestation of denial of actual hallucinations takes place when individuals seek to avoid psychiatric hospitalization, or if already hospitalized, to obtain release.

Hallucinations in Forensic Settings

Outside the civil commitment setting, hallucinations arise in a variety of legal contexts. In a nutshell, forensic evaluators attempt to determine the presence of hallucinations with regard to the legal issue involved and if present, the effect(s) of hallucinations on the legal matter at issue. Outside of the treatment-related civil commitment realm, the three most common forensic evaluations in which hallucinations may be especially germane involve disability, competency to stand trial, and the mental state at the time of the offense.

Functional Capability versus Disability

Hallucinations can be a significant factor in determining a person's work ability or degree of compensable disability. Command auditory hallucinations may order the person to stay away from others and prevent the person from working in a job that requires interpersonal interaction. The distraction caused by hallucinations could interfere with a person's concentration or attention and so job tasks cannot be accomplished.

Competence to Stand Trial

A pretrial defendant who experiences hallucinations may have significant impairment in the capacities to communicate with defense counsel to prepare a defense or to participate in the courtroom proceedings. Such impairment could jeopardize the defendant's competence to stand trial. The presence of auditory hallucinations does not automatically confer incompetence as some individuals can proceed with the case if the defendant can ignore the hallucinations. Nonetheless, exaggeration, manufacture, or denial of hallucinations can arise depending on the defendant's goal.

Mental State at the Time of the Offense

Hallucinations may form part of the basis for a clinical-legal opinion supporting the insanity defense or diminished capacity. Hallucinations may also be viewed as a potential mitigating factor for the trier of fact to consider during the sentencing phase of a criminal trial. With regard to the insanity defense, command auditory hallucinations commonly play a crucial role. An adjudication of not guilty by reason of insanity can be made if the trier of fact finds that the command hallucinations negated the defendant's cognitive or volitional capacities as outlined by the jurisdictional insanity statute. A forensic evaluation would involve arriving at a hypothesis regarding the genuineness of the hallucination with regard to the alleged crime and the extent of the defendant's compliance with the hallucinated command.

References

[1] Ropohl, A., Sperling, W., Eistner, S., Tomandi, B., Reulback, U., Kaltenhauser, M., Kornhuber, J. & Maihofner, C. (2004). Cortical activity associated with auditory hallucinations, *Neuroreport: For Rapid Communication of Neuroscience Research* **15**, 523–526.

[2] Factor, S.A. & Molho, E.S. (2004). Threatening auditory hallucinations and Cotard syndrome in Parkinson disease, *Clinical Neuropharmacology* **27**, 205–207.

[3] Wilson, R.S., Krueger, K.R., Kamenetsky, J.M., Tang, Y., Gilley, D.W., Bennett, D.A. & Evans, D.A. (2005). Hallucinations and mortality in Alzheimer disease, *American Journal of Geriatric Psychiatry* **13**, 984–990.

[4] Fenelon, G., Goetz, C.G. & Karenberg, A. (2006). Hallucinations in Parkinson disease in the prelevodopa era, *Neurology* **66**, 93–98.

[5] Mosimann, U.P., Rowman, E.N., Partington, C.E., Collerton, D., Littlewood, E., O'Brien, J.T., Burn, D.J. & McKeith, I.G. (2006). Characteristics of visual hallucinations in Parkinson disease dementia and dementia with Lewy bodies, *American Journal of Geriatric Psychiatry* **14**, 153–160.

[6] Bourgeois, J.A., DeJuan, T., Johansen, T. & Walker, D.M. (1998). Visual hallucinations associated with fluoxetine and sertraline, *Journal of Clinical Psychopharmacology* **18**, 482–483.

[7] Solhkhah, R., Finkel, J. & Hird, S. (2000). Possible risperidone-induced visual hallucinations, *Journal of the American Academy of Child and Adolescent Psychiatry* **39**, 1074–1075.

[8] Gross-Tsur, V., Adina, J. & Shalev, R.S. (2004). Hallucinations during methylphenidate therapy, *Neurology* **63**, 753–754.

[9] American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition, Text Revision (DSM-IV-TR), American Psychiatric Association, Washington, DC.

[10] Stebbins, G.T., Goetz, C.G., Carrillo, M.C., Bangen, K.J., Turner, D.A., Glover, G.H. & Gabrieli, J.D.E. (2004). Altered cortical visual processing in PD with hallucinations: an fMRI study, *Neurology* **63**, 1409–1416.

[11] Ukai, S., Yamamoto, M., Tanaka, M. & Takeda, M. (2004). Treatment of typical Charles Bonnet syndrome with donepezil, *International Clinical Psychopharmacology* **19**, 355–357.

[12] Ffytche, D.H., Lapin, J.M. & Philpot, M. (2004). Visual command hallucinations in a patient with pure alexia, *Journal of Neurology, Neurosurgery, and Psychiatry* **75**, 80–86.

[13] Hall, D.C., Lawson, B.Z. & Wilson, L.G. (1981). Command hallucinations and self-amputation of the penis and hand during a first psychotic break, *Journal of Clinical Psychiatry* **42**, 322–324.

[14] Tobias, C.R., Turns, D.M., Lippmann, S., Pary, R. & Orophilla, T.B. (1988). Evaluation and management of self-mutilation, *Southern Medical Journal* **81**, 1261–1263.

[15] Field, H.L. & Waldfoget, S. (1995). Severe ocular self-injury, *General Hospital Psychiatry* **17**, 224–227.

[16] Pontius, A.A. (1990). Infanticide in limbic (?) psychotic trigger reaction in a man with Jacksonian and petit mal

(?) seizures: "kindling' by traumatic experience, *Psychological Reports* **67**(3 Part 1), 935–945.
[17] Shore, D., Filson, C.R., Johnson, W.E., Rae, D.S., Mehrer, P., Kelley, D.J., Davis, T.S., Waldman, I.N. & Wyatt, R.J. (1989). Murder and assault arrests of White House cases: clinical and demographic correlates of violence subsequent to civil commitment, *American Journal of Psychiatry* **146**, 645–651.
[18] Hellerstein, D., Forsch, W. & Koenigsberg, H.W. (1987). The clinical significance of command hallucinations, *American Journal of Psychiatry* **144**, 219–221.
[19] Junginger, J. (1990). Predicting compliance with command hallucinations, *American Journal of Psychiatry* **147**, 245–247.
[20] Rogers, R., Gillis, J.R., Turner, R.E. & Frise-Smith, T. (1990). The clinical presentation of command hallucinations in a forensic population, *American Journal of Psychiatry* **147**, 1304–1307.
[21] Thompson, J.S., Stuart, G.L. & Holden, C.E. (1992). Command hallucinations and legal insanity, *Forensic Reports* **5**, 29–43.
[22] Junginger, J. (1995). Command hallucinations and the prediction of dangerousness, *Psychiatric Services* **46**, 211–214.
[23] Zisook, S., Byrd, D., Kuck, J. & Jeste, D.V. (1995). Command hallucinations in outpatient with schizophrenia, *Journal of Clinical Psychiatry* **56**, 462–465.
[24] Kasper, M.E., Rogers, R. & Adams, P.A. (1996). Dangerousness and command hallucinations: an investigation of psychotic inpatients, *Bulletin of the American Academy of Psychiatry and the Law* **24**, 219–224.
[25] Beck-Sander, A., Birchwood, M. & Chadwick, P. (1997). Acting on command hallucinations: a cognitive approach, *British Journal of Clinical Psychology* **36**(Part 1), 139–148.
[26] Wong, M., Fenwick, P., Fenton, G., Lumsden, J., Maisey, M. & Stevens, J. (1997). Repetitive and non-repetitive violent offending behaviour in male patients in a maximum security mental hospital–clinical and neuroimaging findings, *Medicine, Science, and the Law* **37**, 150–160.
[27] Read, J. & Argyle, N. (1999). Hallucinations, delusions, and thought disorder among adult psychiatric inpatients with a history of child abuse, *Psychiatric Services* **50**, 1467–1472.
[28] Erkwoh, R., Willmes, K., Eming-Erdmann, A. & Kunert, H.J. (2002). Command hallucinations: who obeys and who resists them? *Psychopathology* **35**, 272–279.
[29] Fox, J.R.E., Gray, N.S. & Lewis, H. (2004). Factors determining compliance with command hallucinations with violent content: the role of social rank, perceived power of the voice and voice malevolence, *Journal of Forensic Psychiatry and Psychology* **15**, 511–531.
[30] Lee, T.M.V., Chong, S.A., Chan, Y.H. & Sathyadevan, G. (2004). Command hallucinations among Asian patients with schizophrenia, *Canadian Journal of Psychiatry* **49**, 838–842.
[31] Shawyer, F., Mackinnon, A., Frahall, J., Trauer, T. & Copolov, D. (2003). Command hallucinations and violence: implications for detention and treatment, *Psychiatry, Psychology and Law* **10**, 97–107.
[32] McNiel, D.E., Eisner, P.F. & Binder, R.L. (2000). The relationship between command hallucinations and violence, *Psychiatric Services* **51**, 1288–1292.
[33] Resnick, P.J. (1997a). Malingered psychosis, in *Clinical Assessment of Malingering and Deception*, 2nd Edition, R. Rogers, ed, Guilford Press, New York, pp. 47–67.
[34] Gauntlett-Gilbert, J. & Kuipers, E. (2003). Phenomenology of visual hallucinations in psychiatric conditions, *Journal of Nervous and Mental Disease* **191**, 203–205.

GREGORY B. LEONG, J. ARTURO SILVA AND ROBERT WEINSTOCK

Handwriting and Signatures, Comparison of

Introduction

The forensic examination of handwriting is considered by some to be the oldest of the forensic sciences [1]. The discipline appeared to develop into a formalized study aided by the publication of Albert S. Osborn's "Questioned Documents" textbook in 1910. A second edition of this text appeared in 1929 and the work is still considered by many to be the pivotal doctrine [2]. Since then, other texts have appeared that provide detail of the principles and processes of forensic handwriting text and signature examinations [3–7].

In contemporary legal arenas, forensic handwriting examiners (FHEs) are regularly called upon to provide their opinions regarding the authorship of writings that are either in question or in dispute. As a subdiscipline of forensic document examination, forensic handwriting and signature examination is one of the most commonly accepted lines of behavioral identification evidence in existence today. Despite this, handwriting examination has in recent times attracted a healthy variety of academic criticism [8–10]. The first of these articles criticized the poor scientific character of the handwriting identification discipline [8]. This position was confirmed in

a subsequent court ruling [9]. Since the mid 1990s, the debate has continued to develop, both in the literature [10–16] and the courts (see review of relevant court rulings in [1]). It is clear that the basis and probative power of the opinions that FHEs regularly table will remain the subject of scrutiny in future. For the overview presented here, we must set aside this controversy to some extent. In practice, handwriting evidence continues to be led in courts of law internationally and, in the vast majority of instances, is accepted into evidence. This article explores, in general terms, the basis, the structure, and the characteristics of forensic handwriting and signature investigations.

Handwriting of individuals can find its way into a variety of criminal and civil investigations. Theoretically, it is convenient to think of questioned handwriting as either comprising a population of text characters that are designed to be read, or a signature. Signatures are developed and executed by individuals as their "mark", which may or may not exhibit textual elements. Essentially, the FHE applies their expertise to the task of comparing the population of features associated with the questioned writing, to a comparable population of features associated with the exemplar (known) writing of a particular individual or person suspected of writing the entries. Propositions that may account for similarities or dissimilarities between the exemplar and questioned writings are then considered by the FHE. On the basis of their detailed knowledge of handwriting features, simulation (copying) processes, and disguise strategies, the FHE, where they deem the evidence to be sufficient, may express an opinion as to whether or not there is support for the proposition that the exemplar writer wrote the questioned entries.

It is not difficult to imagine that for any particular case, the populations of characters and writing features, and the quality and quantity of comparable features will differ with respect to previous cases that FHEs have had exposure to. Herein rests one of the key elements of the handwriting identification paradigm. The instrument performing the comparison task is the FHEs cognitive machinery and, although a variety of techniques may be used to assist the examination process (for example, magnification and image-enhancement techniques), the approach remains essentially internalized and contains an element of subjectivity. Objective or FHE assisted computer comparison approaches have been reported and are mainly used to search large databases of writers [17–20]. Examples of other algorithmic-based objective comparison techniques that have been used to research aspects of writing dynamic or static characteristics can be found [21–27]. Despite a relatively rich interest in computer comparisons, objective techniques are not routinely applied to casework tasks.

The subjective element in handwriting and signature examination should not necessarily alarm client groups and the courts. It is a fact that all studies comparing FHEs with lay persons opinions, although limited in their number and scope, have shown that FHEs are significantly more skilled than their lay counterparts in handwriting and signature examination tasks [28–32]. Sadly, however, only a fraction of practicing FHEs has participated in such studies. The studies published to date have been limited in terms of their number and extent to which they can address the wide range of writing behaviors thought to exist in the population.

What is Handwriting?

Handwriting is a complex overlearnt motor behavior that is carried out, at least to some extent, by most members of the population. A major occupation of primary school children is the development of their writing skills. Children are generally taught a copybook system of writing. These systems provide the repertoire of characters and instructions as to the appropriate process by which each character is to be formed. Years of guided, informal, and sustained practice is then invested by individuals in order to fluently capture the concatenation of strokes that form the elements of each character. Particular energy is focused on each formation in terms of its construction, its internal proportions, and its connections. This is carried out in order for the individual to satisfactorily perform the myriad of coordinated muscle contractions that will participate in forming characters and words. It is these patterns of movements that drive the writing implement that will ultimately lay down the richness of spoken language in the visual domain.

There are many copybook systems taught worldwide. At least 76 different systems are taught in North America alone [7]. Signatures, a common identifier of individuals, develop later as a visual mark. These marks may be stylized to the extent that the final form may have little relationship to their handwritten text characteristics.

Basis of Identification

FHEs embrace that handwriting behavior, at least in the skilled adult population, possesses two key qualities, which makes it a potentially useful form of identification evidence in the forensic sciences. Although the behavior (like nearly all learned pattern movements) does vary from one performance to another (referred to as "natural" or "intrawriter variation"), it is found to be relatively stable within the adult individual. This relative feature invariance within an individual's writing results from the complex process by which handwriting commands are laid down, retrieved, and played out from our long term motor memories [33, 34]. Such is the robustness of the behavior in the skilled writer that it may remain relatively unchanged over the best part of adult life. It is also found that the richness of features associated with an individual's handwriting (whether these be character constructions, proportions, the relationship between characters, the placement of words, the quality of the line, or the relative pressure that is exerted to form the line trace) markedly varies from one person to another. This is evident even when individuals are initially taught an identical copybook model. This phenomenon is referred to as *interwriter variation.* These two key qualities of handwriting behavior are used to mount identification and exclusionary lines of evidence by FHEs. The practical examination of handwriting does become considerably more complex when unnatural writing behaviors (such as disguise and mimicking (simulation behaviors) as well as environmental and/or circumstantial factors (illness, age, etc.,)) are taken into account; however, we shall deal with these specific types of activity later in the article.

In the very simplest of forms, the process of handwriting comparison and identification can be described as follows [35]. FHEs are provided with a handwriting sample (whether text based or signatures) whose author is considered unknown or disputed. This sample is termed "*questioned*". The questioned writing is compared with another sample of writing (a "comparison sample"). This comparison sample could be writing submitted as having been written by a particular person (an "exemplar") or it could be another sample of questioned handwriting (an approach termed a "*common authorship*" examination). Exemplar writings are the most frequently encountered form of comparison sample. These may be either "requested", where the content of the handwriting is dictated to the individual, or "collected", where an investigator locates samples of an individual in day-to-day writing. The comparison sample is examined and the pictorial, structural, and line quality features that the FHE believes to characterize the handwriting are assessed and documented. These features are compared to the features associated with the questioned sample. If it is found that the questioned features are similar to the specimen features, then propositions are advanced that could account for the similarities. Should the examiner be of the opinion that there is evidence to support the proposition that the similarities result from both the questioned and specimen images having been produced by the one writer then the handwriting is said, with some level of confidence, to be "identified". Alternatively, if the features are found to be dissimilar to the questioned images, explanations are proposed that could account for the dissimilarities, and in some instances the examiner will form the opinion that there is evidence to support the proposition that the handwriting was not written by the specimen writer.

Handwriting is a behavioral artifact. It is what is left behind in the physical, visual domain after our bodies have cognitively processed and executed the commands required to perform a specific sequence of ordered muscle contractions. As a behavior, the evidence is characterized by a number of potentially limiting factors. It is the case that there is not "an amount" of handwriting that can fully characterize the extent of the variation in the behavior or provide sufficient information to describe definitively the characteristics of the movement system responsible for the storage, retrieval, and control of the movement instructions that form the characters as they finally appear in physical form. Databases are still not commonly used to estimate objectively the relative frequency of any particular handwriting feature in the population and as such "chance-match" probabilities, the very numbers that bestow upon DNA sciences its "gold status" in the identification sciences, remain subjectively assessed by the FHE. Elements of the behavior may change over time. The fluency of handwriting movements may alter over the course of different writing events, and the features of writing may change naturally over time or in response to more significant events (for example, a purposeful intervention by the writer, in response to altered writing conditions or demands [36, 37], or in response to

other factors including illness or drug-induced states [38–42]). Complicating the dynamics of the writing behavior itself is that the forensic identification technique is based on learnt human (the FHEs) perceptual and cognitive strategies. There is the potential for normal human variation to play a part in any opinions that are ultimately formed. Handwriting "identification" is better referred to as "*opinion identification evidence*" whereby opinion refers to the FHEs belief in the extent to which the evidence supports that the writer of the comparison sample did or did not write the questioned writing. Although this may be at odds with the beliefs of some practitioners in the field, this approach will fit more comfortably with the behavioral scientists practicing within the discipline.

Many concepts, of varying practical importance, are not addressed here. For example, the extraction of writer demographic features such as age and gender. Although there has been some controlled research conducted for the determination of such characteristics [43, 44], the results either do not support that the techniques employed are sufficiently reliable for their routine usage, or that the techniques themselves can not be applied to casework samples due to the limitations associated with the format of commonly encountered written documents.

Exemplars for Comparison

The sample of handwriting that is in question is not within the control of the investigator or the FHE. What is potentially within the realm of control of at least the investigator is the extent to which the writing behavior of any suspect, or subject, might be sampled. The most critical aspect of any forensic handwriting investigation is sourcing suitable known examples of an individual's normal writing for comparison. Exemplar material is also referred to as "*known*", "*standard*" or "*specimen*" writing by different FHEs internationally. The success of any case involving disputed or questioned handwriting will be heavily dependent on both the quality and quantity of the specimen (known) handwriting, submitted for comparison, in conjunction with the quality and quantity of the questioned writings. The terms "*quality*" and "*quantity*" with respect to any particular case and any particular FHE may vary and these factors remain hidden within the subjective processing of the evidence by FHEs. Although there are exceptions, FHEs are generally not responsible for sourcing exemplars or validating that the exemplar material submitted is the known writing of a particular individual. The onus of proof regarding the source of exemplars usually remains with the submitting investigator. The history and development of the US rules compelling suspects to provide handwriting samples has been reported [45].

Exemplar handwriting can either be "requested" or "collected". As a guide, the process by which these two populations of writing may be sourced is described.

Requested Exemplars. These samples are written by an individual at the request of another and are specifically performed for the purpose of providing a set of known writings for forensic comparison. The following approach is generally used to generate material of this type:

1. The subject is allowed to sit comfortably and is provided with all material required to produce the exemplars (writing instruments and writing medium).
2. The subject is not permitted to view the questioned writing prior to taking of the exemplars (in order to avoid any allegation that the individual in any way copied either the form of the characters or the formatting of the questioned writings).
3. Multiple documents, similar in format to the document on which the questioned writing appears, are prepared and held by the investigator. For example, if the questioned writing is text on a lined foolscap page, similar lined foolscap pages would preferentially be used.
4. A writing instrument similar to that used to form the questioned handwriting is provided. Most exemplars are generated using ball point pens. Should the questioned writing be generated using a less common writing implement (such as pencil or crayon), then the subject may be requested to repeat the writings using these devices.
5. The subject is provided with the first prepared document. All writing instructions are provided verbally. The content of the exemplar to be dictated will depend on the form and format of the questioned writing. If the questioned writing is a signature, then the subject will be asked to sign their name (one to a page). If the questioned writing is uppercase handwritten text, then the subject will be asked to write in uppercase and the content will be dictated to them. As the

subject completes each exemplar, the investigator signs and dates the document, and it is removed from the subjects view. This process is repeated for the remaining exemplar document repetitions that have been prepared.

In some cases, *pro forma* exemplar documents either take the place of case-specific prepared documents, as described above, or are added to the prepared documents to be provided to the subject. *Pro forma* documents provide the subject with written prompts as to what to write and in what format the entries are to be written. Documents of this type are generally designed to capture as many of the subjects handwriting characters and their combinations as is practical. This approach provides a broad sample of the subjects writing behavior to the FHE for any comparison purposes that might be required. Occasionally this approach is used in preference to case-specific exemplars as the investigator may not wish for the subject to know the content of the questioned document.

Requested exemplars offer the advantage that the writing behavior will be directly comparable to the questioned material (that is, the exemplars will exhibit the same format of document, and the same letters, letter combinations, word forms, and sentence structures). The disadvantages associated with requested exemplars is that they are unlikely to exhibit the full extent of the subjects range of natural writing variation (since they were executed at a single point in time), they may be disturbed due to the inordinate attention placed on the writing act, and they do offer the subject the opportunity to disguise their normal writing habits. For this reason, requested exemplars are normally supplemented with "collected" exemplars.

Collected Exemplars. Collected exemplars are writings that are made by an individual during the course of their day-to-day activities. These exemplars are considered by many FHEs to be the most desirable for forensic comparisons. The collected writings are unlikely to be the product of disguise (particularly those collected prior to the time that a questioned sample of handwriting was written) and are likely to show the range of normal variation in an individuals' writing behavior.

When collected exemplars are sourced, particular attention is paid to the format of the writings. In general terms, signatures can only be compared with signatures (unless an individual has signed using their normal handwritten text forms), uppercase can only be compared to uppercase writing, cursive can only be compared to cursive, and printed can only be compared to printed writing. As such, when an investigator collects a subject's handwriting, the format of the questioned text needs to be represented in the final body of collected exemplar material.

Timing of the Exemplar Handwriting. One of the most important elements of the exemplar collection process is the relationship of the date(s) associated with the questioned writing with the date(s) associated with the exemplar writings (this is particularly true when dissimilarities are detected between the questioned and exemplar writings). The preference of FHEs is for the exemplar material to "time bracket" the questioned writing. Ideally, the exemplars should be drawn from normal writing activity both before and after the date(s) associated with the questioned writings. Repetitions of exemplar writings should also attempt to be sourced as close in time to any date associated with the questioned writings as possible. In practice, this is often a challenge.

Sources of Collected Exemplar Writing. Exemplar handwriting can be sourced from a wide variety of document types [7]. Commonly submitted "collected" handwritten text and signature exemplars are documents associated with charge accounts, change of address forms, affidavits, business agreements, credit and insurance applications, membership applications, passport applications, work and school assignments, attendance records, banking documents, general business correspondence, recipes, credit card documents, declarations, grocery lists, guest registers, hospital records, identification cards, leases, mortgages, personnel records, greeting cards, postcards, tax returns, time sheets, and wills.

Although it is not the task of the FHE to validate the source of exemplars, it is the case that examiners do inspect the body of writings submitted as exemplars for any potential contamination by another writer. Generally, any spurious entries in minority are excluded prior to any detailed examination being carried out. Should there exist uncharacteristic variation between different exemplar documents the investigator may be quizzed as to the reliability of the sample and further exemplars may be called for.

Handwriting Examination Theory and Practice

The extent to which the FHE supports an identification or exclusionary conclusion is referenced by their use of opinion level scales. A variety of these scales are in use internationally and the wording differs from place to place. A common form is the verbal probability scale and makes use of nine levels (four opinion strength levels either side of an inconclusive opinion) [46]. Included here is a five-level scale that is used for international validation and proficiency testing and shows how both verbal probability (probability of a proposition – prior odds) and Bayesian reporting language (likelihood ratio assisting the expert to quantify his or her support for a proposition against an alternative proposition) is used.

Level 1 – Identification

The writer of the specimens wrote the questioned writing. This is an opinion that can be thought of as a "moral" certainty (rather than a scientific certainty) and is a categorical opinion that the writer of the specimen handwriting wrote the questioned handwriting [47]. Some examiners also define this opinion as "very strong support for the proposition that the writer of the exemplars wrote the questioned writing."

Level 2 – Qualified identification

There are indications that the writer of the specimens wrote the questioned writing.

This opinion level is used when there is an identifiable limitation on the comparison process (e.g., a limited amount of questioned and/or exemplar writing/nonoriginal questioned and or exemplar writings). This opinion has also been defined as "moderate-to-strong" support for the proposition that the writer of the exemplars wrote the questioned writing.

Level 3 – Inconclusive

Owing to limitations to the structure or content of the case, the FHE is not able to form an opinion as to whether or not the writer of the exemplars wrote the questioned handwriting.

Level 4 – Qualified exclusion

There are indications that the writer of the specimens did not write the questioned writing. This opinion level is used when there is an identifiable limitation on the comparison process (e.g., a limited amount of questioned and/or exemplar writing/nonoriginal questioned and/or exemplar writings). This opinion has also been defined as "moderate-to-strong" support for the proposition that the writer of the exemplars did not write the questioned writing.

Level 5 – Exclusion

The writer of the specimens did not write the questioned writing. This is an opinion described as a "moral" certainty and is a categorical opinion that the writer of the specimen handwriting did not write the questioned handwriting. Some examiners also define this opinion as "very strong support for the proposition that the writer of the exemplars did not write the questioned writing."

Handwriting Characteristics

Clearly, it is the character of the visual artifacts that FHEs rely upon to perform their comparisons. Underlying any opinion formed by the FHE regarding the concept of individualization must be a detailed understanding as to the extent of intra and interwriter variation and how these might interact with each other. Intrawriter variation might be considered as a major source of "noise" in the identification paradigm. But this is an overly simplistic view. The way that an individual's handwriting varies can provide the FHE with features that might be considered to assist in discriminating between writers (that is added to a body of features that can be used to discriminate between writers). FHEs' understanding of intrawriter variation is likely to be finely tuned through their constant exposure to the phenomenon due to their detailed examinations of enormous amounts of exemplar documents. This exposure provides an environment where their cognitive systems learn what features may change naturally within a writer, and what dissimilarities between questioned and exemplar writings should evoke particular suspicion.

For FHEs to form any opinion at all, there must be a firm belief that there is a wide range of interwriter variation present in any particular population of writers. Evidence for the extent of interwriter variation can be drawn from a variety of sources. At the most basic-level interwriter variation is what allows individuals to easily recognize both their own handwriting and the writing of others to which they have

had prolonged exposure. Further empirical evidence comes from the study of closely related individuals. It may be expected that genetically close individuals, or those with similar family backgrounds may exhibit handwriting features that have the potential to be indistinguishable. This is not found to be the case [48]. Large-scale handwriting investigations have been reported where the correct writers were ultimately identified from surveys of multitudes of individuals' exemplars [49, 50]. A qualitative approach was used by Huber (2000) on nearly 1000 samples of writing. This author found that for all bar two of the samples, the balance of the authors could be discriminated [51]. Database searching systems would be unable to function with any measure of success without there being a marked degree of interwriter variation. More recently, Srihari and his colleagues (2002), using computer algorithms, were able to validate handwriting individuality in a sample population of 1500 individuals with a 96% confidence [24]. Further lines of empirical support for the large degree of interwriter variation have been reported [52–55].

Ultimately, it is all the elements that form handwriting images that FHEs' cognitive machinery compares between the exemplar and questioned line traces. These are described in various, but essentially similar ways within the major texts in the field. As a summary, a holistic approach to the issues of what constitutes handwriting features is described here and is based on feature descriptors that have been described [7]. It is the case that, due to a combination of the intrawriter variation phenomenon coupled with the continuous signal that handwriting presents to us, FHEs do not "count" features. There are no objectively based minimum thresholds to define when an FHE will form an opinion that the evidence supports a particular level of identification or exclusion. This characteristic of the evidence can result in variations in the strength that different FHEs will form opinions, even given the same material to examine. This can also result in FHEs supporting propositions at odds with each other. The extent that this occurs remains essentially unreported.

Huber and Headrick (1999) find it helpful to classify handwriting features into four general types [7].

Type 1 – Elements of style

This includes the relative placement of text, the uniformity of the margins, the interline spacing, the parallelism of the lines of writing, the character, position and frequency of interlineations, the depth of any indentations (for example, when the writer commences new paragraphs), the paragraphing format, the use of numerals and symbols in monetary amounts, the location of handwriting relative to any printing on the document, the class of character (whether cursive, printed, manuscript (disconnected), or composites of these styles), the connections (between both characters and words), the characters design and construction, the proportion of features both within and between characters, writing size, the slant of writing, and intra/interword spacing habits.

Type 2 – Elements of execution

These features include the use of abbreviations, the orientation of the handwriting relative to real or imaginary baselines, the length, direction, path and taper of commencing and terminating strokes, the presence, style and location of diacritics and punctuation, the presence of embellishments, the legibility of the writing, the presence of pen stops and pen lifts, the quality of the line, and pen control characteristics (as evidenced by pen pressure differentials in the line trace).

Type 3 – Consistency or natural variation

This comprises an assessment of the extent of imprecision with which the writer executes the movements, which results in variations in the handwriting features on repeated occasions.

Type 4 – Lateral expansion and word proportions

These factors contribute to the pictorial character of the writing and refer to a combination of the horizontal dimensions of a group of successive letters and words, the relative horizontal to vertical character of the writing and the size and spacing of the writing components.

As FHEs assess these types of writing components, and compare the features between the exemplars and questioned entries, their cognitive machinery is likely to be interpreting the weight of the similarities and dissimilarities that they observe in terms of the extent to which they might support either an identification or exclusionary conclusion. Only limited studies have been reported, which attempt to explore how FHEs might cognitively approach the

task and so the rationalization of how this hidden process may be achieved remains anecdotal [56].

For some FHEs, the theoretical basis on which identification opinions are mounted is centered on what is generally described in the accepted texts as "*Class and Individual Characteristics Theory*". Stated simplistically, it is accepted that handwriting is comprised of features that are shared by populations of writers and are of little or no value to identification processes (in many cases, these are remnants of the copybook system that the writer was originally taught). Although not useful for identification, these characteristics may be very useful in terms of writer exclusion. When inexplicable repeated dissimilarities in characteristics (including class elements) are observed between a population of naturally written exemplar and questioned writings, this would support the proposition that the population of writings were written by different individuals. Identification opinions are mounted on the basis of individual characteristics. On the basis of their experience, FHEs who support this theory claim that they can detect features in handwriting that are individual to a particular person. Other FHEs, while embracing the belief that handwriting is composed of class characteristics and characteristics that vary from "class", do not claim to be able to detect characteristics that are unique to a particular writer. "Complexity Theory" is an example of such an approach [57]. This theory embraces the notion that there is a wide range of interwriter variation in the population. It proposes that there are two basic relationships that exist, which underpin all handwriting identifications. The first is that, as the amount of writing available for analysis and comparison increases, the more likely that writing is to contain features that are deviant when compared with other individuals' handwriting (that is, the probability of a chance match with respect to all features decreases). For example, at the simplest level, if we were to collect random samples of handwriting exhibiting identical text and compare them letter by letter, we should find increasing sample to sample deviations as the comparison proceeds. The second relationship in this theory is that as the amount of writing increases, the more difficult it will be for another individual to simulate (copy or mimic) it successfully. Although relevant to handwriting text evaluations, simulation behaviors will be dealt with in more detail in the signature section of the article.

There are at least anecdotal accounts of varying interpretations and strength of belief in these theories. Ultimately, to differing extents, both theories suffer due to the subjectivity associated with the interpretive processes coupled with the lack of detailed empirical investigations into their veracity. FHEs have varying personal views as to the basic theoretical process. The more pressing issue is, however, to provide validating information regarding the probative value of their opinions. This is for many being addressed. This occurs through the medium of small externally set proficiency testing (for example, those provided through Collaborative Testing Services (USA)), through between laboratory collaborative trials or resulting from participation in large-scale blind tests that have been reported [58–63].

When FHEs perform their comparisons, detailed notes are made. The form that the documentation of observations takes can vary quite markedly. In many instances, for the exemplar material, each character, character combination and any word formations shared between the exemplar and questioned material are carefully drawn out. In some cases, the FHE will extract each of these elements digitally so as to avoid the inaccuracies of the drawing process. The same process may be undertaken for the questioned material. In this way, an ordered set of the features of characters and their variations are documented, a detailed knowledge of the characteristics of the writings are laid down in the mind of the FHE and the comparison process is facilitated. Other methods are also used. A FHE may mark on reproductions of the handwriting the features that are similar or dissimilar between the populations of exemplar and questioned writings. What remains is for the FHE, on the basis of all their observations of the feature similarities and/or dissimilarities, to address the ultimate set of propositions; whether the evidence supports the proposition that the exemplar writer did or did not write the questioned material. Documented methods that articulate the key elements of the process are likely to be able to be secured from most laboratories offering handwriting evidence. One such treatment of method has been reported [35].

The identification and exclusion approach summarized here is extremely effective when FHEs are provided with naturally written skilled handwriting. Practically, in their deliberations, FHEs are confronted with the casework reality that, even given

suitable naturally written exemplar writings, the questioned writing may not be naturally written. Questioned writing may be the product of disguise (either by the exemplar writer or by an individual other than the exemplar writer) or may be the product of simulation behavior (written by a writer copying the handwriting characteristics of the exemplar writer for the purpose of implicating them in the preparation of a document). In both cases, when the significance of the features in terms of authorship is being addressed, these unnatural writing behaviors are taken into consideration.

Disguised and Simulated Handwriting

Much has been written on the subject of disguised writing behaviors [64–70]. In essence, the writer is attempting to purposefully change elements of their normal handwriting output while simultaneously suppressing their natural handwriting feature habits. This is typically done to avoid being identified through the medium of any handwriting comparison process. Disguised writing behavior is most commonly observed on threatening correspondence. The approach and effectiveness of the disguised strategy adopted does vary markedly from person to person. Some individuals feel that simply altering the construction of one character will be sufficient. Others adopt more complex approaches that may include simplifying their writing style to modify characters into a series of straight line constructions. Still others will use their nondominant hand to which may alter the pictorial character of the writing as well as bestowing on the writing the character of reduced skill.

Testing of FHEs on blind trials has discovered that, in general, disguise strategies adopted by individuals are effective in cloaking their identity. FHEs, when considered as a group, exhibit high rates of inconclusive opinions on disguised writings and disproportionately high errors when compared to the natural writings of individuals [62]. There are examples of individual FHEs that do appear to have more advanced skills in determining the authorship or nonauthorship of disguised writing; however, only a fraction of practicing FHEs expose themselves to large-scale testing and the number of these tests to date do not provide any firm foundations to state convincing claims as to expertise for questioned writings of this type.

Simulated text writings are found to result in slower movement velocities and a greater number of movement disfluencies [71]. Simulated writings have also proven to be problematic for FHEs. Although the frequency of occurrence of these writing types in the casework population is likely to be low, it does not alter the fact that the probative value of any identification opinions is dependent on the skill of the FDE at not incorrectly opining that a sample of simulated writing was written by the writer of the exemplars. A similar pattern of inconclusive and error scores are found for the simulated writing group as for the disguised writings. The most current research strongly suggests that when FHEs, particularly those from a nonblind testing culture, are confronted with questioned writings exhibiting a mixture of similar and dissimilar features to the exemplar writings, then the appropriate opinion to reach is no opinion at all [62].

Signature Examinations

One of the major fields of study of the FHE is signature behavior. Signatures appear as personal identifying marks on all manner of documents. Although the control mechanism underlying their performance is similar to that associated with handwritten text, their role is not necessarily to be a readable representation of the name of the signer. As a consequence, signatures do vary from a readable text-based representation of an individual's name all the way through to a stylized pattern where no indication may be present as to the identity of the component characters.

The purpose of a signature clearly is for the signer to acknowledge the content of the document on which it appears. For skilled signatures at least, the complexity of the underlying movement control system guiding their production bestows, upon any act of copying such a movement, some considerable difficulty. The process of copying a signature is described by FHEs as "simulation behavior". The more common, but inaccurate, term used in the lay environment is "forgery". A considerable amount of information and research is available on simulations in both book form [2–7] and research literature [72–77].

Simulations form part of a wider family of signing behaviors encountered in forensic casework. The following provides a brief summary of the types of questioned signatures that FHEs encounter.

Genuine Signatures

FHEs, subsequent to an analysis and comparison between the features associated with the exemplar and questioned signature, may find support for the proposition that a signature is genuine. In these instances, the questioned signature is usually naturally and fluently written and exhibits no significant dissimilarities with respect to its constituent features when compared to the exemplar group.

Spurious Signatures

Questioned signatures of this type may appear to have no relationship to features of the genuine form other than those that could exist by chance. The writer of the spurious signature either did not have access to, or did not choose to directly simulate the genuine form. These signatures may be stylized (which excludes any opportunity to compare its features with exemplar text of any suspect) or may be text based. Text-based spurious signatures may be able to be compared directly to the collected or requested handwriting exemplars of an individual.

Disguised Signatures

These signatures are written by the exemplar writer; however, an attempt is made to alter the characteristics of their normal signature for the purposes of potential denial of the signing act at a later point in time. It has been reported that typically obvious characters are altered (for example, the initial letter in any component forms) [78–81]. The reality is that the strategy that the exemplar writer may use is limited only by their imagination, their motor abilities, and the extent to which the signature might need to pass initial inspection when the documentation on which it appears is first passed.

Freehand Simulated Signatures

The simulator, having access to a genuine model signature, attempts to replicate the form either by directly referring to a model or by referring to their memory of its appearance. The simulator may embark on a detailed study of the genuine signature in order to practice its component parts. Alternately, a simple drawing of the form may be the only strategy that the simulator attempts. Typically, simulated signatures show a range of defects when compared to the exemplar signature. Tremor in the line, a reduction in shading patterns (which in natural writings are caused by smooth changes in pen speed and pen pressure) and blunt initial and terminal strokes may be evident. Uncharacteristic pauses or pen lifts may be detected. These are thought to be associated with the simulator either assessing the progress of their copy or indeed halting the movement in order to refer back to the genuine model signature. Touching-up, adding strokes, or "patching" of the signature may be evident where the simulator, unsatisfied with the overall reproduction, attempts to improve the form through the application of additional pen strokes. Dissimilarities with respect to overall size of the signature, the relative proportion of constituent features, and the slope of the form may also be detected. Alternatively, the simulator may wish to capture the fluency characteristics of the genuine signature. The speed-accuracy trade-off will, in many instances, result in a simulation that displays inaccuracies in fine movement detail, construction, and relative proportions of features.

Traced Simulated Signatures

One of the ways that a simulation can be produced (either of handwritten text or a signature) is to use a model of a genuine form as a physical guide [82, 83]. Traced signatures are related to the freehand simulated signatures with respect to all but the defects associated with size and relative proportions. Owing to the additional demands of following a guide, these signatures tend to show greater deficiencies than do freehand simulations. Examples of tracing methods are as follows:

- A document bearing a genuine signature is placed against a light source (for example, a window). The document on which the simulator wishes to trace the signature is placed over the genuine document and aligned according to where the tracing is to appear. Depending on the paper, sufficient gross information may be seen through the medium for the simulator to follow the line of the genuine formation. This technique is relatively effective at reproducing the gross pictorial characteristics but as one might imagine any fine detail in the genuine signature features may be missed, the sequence of pen strokes may be incorrectly executed and the line quality is generally found to be poor due to the very

slow speeds that the simulator might execute the movements in order to faithfully follow the trajectory of the line.

- The document on which the simulation is to be made is placed under a document where the genuine model signature appears. The simulator, using a writing implement such as a ballpoint pen, firmly follows the line of the genuine signature. The pressure that this creates causes an indentation line on the underlying document in the shape of the genuine signature. The uppermost document is removed and the simulator then uses a writing implement to follow the indentation line. This technique will still result in the defects associated with the tracing process described previously. Additional evidence will be left in the form of the indentation line which, due to the inaccuracies typically associated with the simulators movements, will not be overwritten in its entirety by the inking medium. Indentation lines in close proximity to the questioned signature, however not always parallel to the trajectory of the written trace, will typically be in evidence. There are a number of variations on this technique. The simulator may use carbon paper between the two documents. The simulator then is only required to exert minimal pressure to create a weak visual outline of the genuine signature that appearing on the top document onto the document below it. This outline can then be overwritten with any given writing implement in an attempt to obscure it. Alternately, a graphite (pencil) layer can be placed on the rear of the document directly over the back of a genuine signature. This document is placed over the document where the simulation is to appear. Pressure is applied over the genuine signature in the same way as described previously. This transfers some of the graphite from the rear of the uppermost document onto the front surface of the document below it. Again, the simulator can then follow this outline using a writing implement of their choice. Any fragments of the graphite line can then be removed using an eraser. What is often observed, with this technique of tracing, is not only partial lines of indentation caused by the pressure of the transfer process but also fragments of graphite resulting from incomplete erasure. Smudging and paper fiber disturbance caused by the physical trauma to the writing medium itself may also be in evidence.

Traced simulations are more likely to deposit additional evidence for the FHE to base their opinions on. On occasions, the model used for the tracing can been found. The traced simulation can be found to be sufficiently superimposable on the submitted model such that the FHE opines that it is unlikely that the tracing shares such close association by chance. Alternatively, additional ink lines, indentation lines (caused by drawing over the genuine signature with a noninking pen) or graphite on the rear of the document may be detected in close association with the genuine signature. Traced simulation behavior can also assist in the forensic investigation of cases involving multiple questioned signatures. One model may have been used by the simulator for their repeat tracings. This will result in questioned signatures not only bearing the typical deficiencies of the simulation process but also exhibiting such a close correspondence in their features that FHEs will opine that they were traced from the same model.

Machine-Generated Signatures

The quality, accessibility and ease of use of contemporary photocopiers, printers and digital image capture, and manipulation techniques provide yet another means for a simulator to reproduce handwriting, and in particular signatures. The older “cut and paste” approach to signature reproduction simply comprised cutting out a genuine signature (or a reproduction of it) from a document, placing it on a photocopier platen and feeding the document on which it is to appear either through the manual feed mechanism or placing it in an existing paper tray cartridge. Typically, some attempts are made such that the reproduced signature appears in the desired position. Alternately, the cutout signature may be stuck to a fabricated document and photocopied directly. Evidence of this process is detectable by the FHE. Microscopic examination will reveal that the line of the signature is not comprised of ink, but rather photocopier toner particles. The toner usually appears on top of the paper fibers as differential pressure is not applied in its production (we observe pressure fluctuations in the performance of genuine signatures). Often, there might also exist a toner shadow where

the edge of the paper cut, associated with the questioned signature, meets the surface of the document onto which it has been stuck.

More advanced approaches involve the utilization of a scanner, a computer with some image-enhancement/manipulation software and a printer (for example, color laser or high-resolution inkjet). Genuine signatures can be scanned into appropriate software, manipulated if required (for example, the slant may be changed, the internal features stretched proportionately or disproportionately, the internal constructions can be altered, etc.,) and either digitally cut and pasted or imported into a document for printing. Excluding the use of robotic arms as an output medium [84, 85], the FHE still has the means to determine that the final product is not the result of human movement using typical writing implements but rather a machine-generated process.

Other Signatures

There are yet other signatures behaviors that are less well understood and which have attracted few reported empirical studies (with respect to the FHEs claims to validity) than those described above. These include signing behaviors associated with various forms of ill health and degenerative neural conditions, those affected by both legal and illegal drugs affecting fine manipulative movements, and the effect of unusual signing positions and backing surfaces. Fortunately, these types of signing behaviors are less commonly encountered in routine casework and, in many instances, the robustness of the approach may compensate for any of these internal or external conditions having an effect on the overall reliability of the opinion. The potential for alterations to normal motor output resulting from factors associated with these behaviors is generally considered by the FHE, particularly with respect to securing appropriate exemplars for comparison.

Few large-scale signature validation studies have been reported [31, 32]. In general, these studies show that the expertise claimed by FHEs in discriminating between genuine and simulated signatures is real. However, little is known regarding the potential error rate associated with the application of the technique across wider casework problems. The most recent report suggests that, in a similar way as has been shown for handwriting text, questioned signatures displaying combinations of similar and dissimilar features remain the most problematic in terms of FHEs attribution of authorship [86]. These are the disguised and simulated signatures (only freehand simulations have been used in studies, to date). The identification of genuine signatures attracts only small numbers of erroneous opinions and FHEs used in the reported study call these signatures confidently (with only small inconclusive scores being recorded). Disguised and simulated signatures attract large inconclusive scores and, at least for disguised signatures, significantly high errors. FHEs opinions, at least with respect to freehand simulated signatures, are still useful if we take into account their first stage observations. Some FHEs provide opinions that questioned signatures are “nongenuine”. Nongenuine questioned signatures refer to formations that are either the product of a simulation or disguise process. These FHEs, recognizing the problematic nature of exclusionary opinions on questioned signatures, offer this opinion without proffering any opinion as to whether or not the exemplar writer wrote the formation. Limited empirical studies, to date, show that, in order to avoid serious error, this is the most desirable position for the majority of FHEs to take. It leaves the court with the ultimate decision as to the likelihood of disguise behavior (which, in many instances, due to the courts access to the full circumstances of the case, they will be far better placed to make a ruling on).

In some instances, there is general agreement that it is appropriate for an FHE to support an exclusionary opinion. These are cases where the exemplar material is ideal, where it time-brackets the alleged date of signing of the questioned signature, and where the skill of the exemplar writer is far lower than the skill at which the questioned signature was written. These cases are, however, relatively rare.

Who are FHEs?

FHEs emerge from a wide variety of backgrounds. Some are police or investigators that have undergone advanced training or apprenticeships in the forensic handwriting discipline. Still others are individuals with science qualifications exhibiting a variety of academic specialties (reflecting the requirements of the wider document examination field within which handwriting examination is often placed). These individuals are also required to undergo training in the discipline. A variety of individuals, from all walks of life, do claim expertise in the

field based on a variety of mentored training programs. Their mentored relationship is usually coupled with forensic document examination society affiliations. All practitioners tend to gravitate toward the same set of dedicated textbooks [2–7], and there are a variety of scientific journals reporting research in the forensic handwriting field (many of these are listed in the reference section of this article).

Forensic handwriting examination is essentially a deregulated discipline. Many individuals may support their claim to expertise based on their affiliations to particular organizations or groups. Examples of such organizations are The American Academy of Forensic Sciences, The American Society of Questioned Document Examination, The Southwestern Association of Forensic Document Examiners (USA), The Southeastern Association of Forensic Document Examiners (USA), The Forensic Science Society (UK), The European Network of Forensic Handwriting Experts, The Gesellschaft fur Forensische Schriftuntersuchung (Germany), the National Association of Documents Examiners (USA), and The Association of Forensic Document Examiners (USA). Some organizations offer the FHE the opportunity for certification (usually in the broader discipline of Forensic Document Examination). As an example, The American Board of Forensic Document Examination offers certification to applicants with a combination of appropriate character and professional standing, with at least a bachelors degree (or equivalent), who have completed at least a two year full-time training program in a laboratory recognized by the Board, and who can demonstrate a profile of professional activities deemed appropriate. More information about training and certification in criminalistics can be found in **Training and Certification (in Criminalistics)**.

Besides their membership to groups or associations, FHEs may further support their claims to expertise through their years of experience in the field. Unfortunately, studies to date have not shown that there is any relationship between years of experience once FHEs are trained and their performance on blind trials [1]. There are still no reported studies comparing the relative abilities of different group memberships or even comparisons of FHEs opinions where their training process is uncommon. That forensic handwriting examination is a human skill that does exist is not essentially in dispute. The difficulty for the courts, in the absence of a strong scientific culture focused on testing of both underlying belief and practitioner skill, is the assessment of the suitability of any particular specialist to have their opinions heard. There is no doubt that professional memberships, certification, proficiency test records and rules of evidence admissibility (such as the Daubert checklist [87]) do go some way to address this dilemma; however, the environment within which FHEs practice still can be argued to rally its strength from historical acceptance rather than from scientifically based empirical studies [10]. This, at least in the very short term, is unlikely to change. Inevitably the so called Paradigm Shift [16] in the identification sciences will begin to impact appropriately on this discipline. The shift is likely to be accelerated through the ever increasing number of scientist entering the field. Their knowledge of scientific culture and, hopefully, their enthusiasm to question underlying theory and apply appropriate research skills, including advance statistics, will result, in so far as behavioral identification will allow, in many of the shortfalls of this science being addressed.

References

[1] Risinger, D.M. (2006–2007). Handwriting identification, in *Modern Scientific Evidence: The Law and Science of Expert Testimony*, D.L. Faigman, D.H. Kaye, M.J. Saks & J. Sanders, eds, Thomson West, pp. 377–524.

[2] Osborn, A.S. (1929). *Questioned Documents*, 2nd Edition, Nelson-Hall, Chicago.

[3] Harrison, W.R. (1958). *Suspect Documents: Their Scientific Examination*, Praeger, New York.

[4] Conway, J.V.P. (1959). *Evidential Documents*, Charles C Thomas, Illinois.

[5] Hilton, O. (1982). *Scientific Examination of Questioned Documents*, Science Publishing, Elsevier, New York.

[6] Ellen, D. (1989). *The Scientific Examination of Documents: Methods and Techniques*, Ellis Horwood, West Sussex.

[7] Huber, R.A. & Headrick, A.M. (1999). *Handwriting Identification: Facts and Fundamentals*, CRC Press, Boca Raton.

[8] Risinger, D.M., Denbeaux, M.P. & Saks, M.J. (1989). Exorcism of ignorance as a proxy for rational knowledge: the lessons of handwriting identification "expertise", *University of Pennsylvania Law Review* **137**, 731–792.

[9] United States v. Starzecpyzel, 880 F.Supp. 1027 (S.D.N.Y. 1995).

[10] Mnookin, J.L. (2001). Scripting expertise: the history of handwriting identification evidence and the judicial construction of reliability, *Virginia Law Review* **87**(8), 1723–1846.

[11] Risinger, D.M. & Saks, M.J. (1996). Science and nonscience in the courts: daubert meets handwriting identification expertise, *Iowa Law Review* **82**, 21–74.

[12] Moensens, A.A. (1997). Handwriting evidence in the post-Daubert world, *UMKC Law Review* **66**(2), 251–343.

[13] Risinger, D.M., Denbeaux, M.P. & Saks, M.J. (1998). Brave new "post-Daubert world"- a reply to professor Moenssens, *Seton Hall Law Review* **29**, 405–485.

[14] Galbraith, O., Galbraith, C.S. & Galbraith, N.G. (1995). The principle of the "Drunkard's Search" as a proxy for scientific analysis: the misuse of handwriting test data in a law journal article, *International Journal of Forensic Document Examination* **1**, 7–17.

[15] Saks, M.J., Risinger, D.M., Rosenthal, R. & Thompson, W.C. (2003). Context effects in forensic science: a review and application of the science of science to crime laboratory practice in the United States, *Science & Justice* **43**(2), 77–90.

[16] Saks, M.J. & Koehler, J.J. (2005). The coming paradigm shift in forensic identification science, *Science* **309**, 892–895.

[17] Philipp, M. (1992). Efficiency control study of the FISH system under real world conditions, *Proceedings of The 3rd European Conference For Police And Government Handwriting Experts*, Rome Italy, October 5–7.

[18] van Erp, M., Vuurpijl, L., Franke, K. & Schomaker, L. (2004). The WANDA measurement tool for forensic document examination, *Journal of Forensic Document Examination* **16**, 103–118.

[19] Franke, K. & Koppen, M. (2001). A computer based system to support forensic studies on handwritten documents, *International Journal on Document Analysis and Recognition* **3**(4), 218–231.

[20] Srihari, S. & Leedham, G. (2003). A survey of computer methods in forensic document examination, in *Proceedings of the 11th International Graphonomics Society Conference*, H.L. Teulings & A.W.A. vanGemmert, eds, Nijmegen, The Netherlands, pp. 278–281.

[21] Found, B., Rogers, D. & Schmittat, R. (1998). Matrix analysis: a technique to investigate the spatial elements of handwriting, *Journal of Forensic Document Examination* **11**, 51–71.

[22] Franke, K. & Grube, G. (1998). The automatic extraction of pseudodynamic information from static images of handwriting based on marked gray value segmentation, *Journal of Forensic Document Examination* **11**, 17–38.

[23] Plamondon, R. & Srihari, S.N. (2000). On-line and off-line handwriting recognition: a comprehensive survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**, 63–84.

[24] Srihari, S.N., Cha, S.H., Arora, H. & Lee, S. (2002). Individuality of handwriting, *Journal of Forensic Sciences* **47**, 856–872.

[25] Ueda, K., Matsuo, K. & Nakamuru, Y. (2004). A computer-based system to support forensic analysis of Japanese writing, *Journal of Forensic Document Examination* **16**, 41–56.

[26] Leedham, G., Pervouchine, V. & Tan, W.K. (2004). Quantitative letter-level extraction and analysis of features used by document examiners, *Journal of Forensic Documents Examination* **16**, 21–39.

[27] Ostrum, B. & Tanaka, T. (2006). Another look at handwriting movement, *Journal of the American Society of Questioned Document Examiners* **9**(2), 57–67.

[28] Found, B., Sita, J. & Rogers, D. (1999). The development of a program for characterising forensic handwriting examiners' expertise: signature examination pilot study, *Journal of Forensic Document Examination* **12**, 69–80.

[29] Kam, M., Wetstein, J. & Conn, R. (1994). Proficiency of professional document examiners in writer identification, *Journal of Forensic Sciences* **39**, 5–14.

[30] Kam, M., Fielding, G. & Conn, R. (1997). Writer identification by professional document examiners, *Journal of Forensic Sciences* **42**, 778–786.

[31] Kam, M., Gummadidala, K., Fielding, G. & Conn, R. (2001). Signature authentification by forensic document examiners, *Journal of Forensic Sciences* **46**, 884–888.

[32] Sita, J., Found, B. & Rogers, D. (2002). Forensic handwriting examiners' expertise for signature comparison, *Journal of Forensic Sciences* **47**(5), 1117–1124.

[33] van Galen, G.P. (1980). Handwriting and drawing: a two stage model of complex motor behavior, in *Tutorials in Motor Behavior*, G.E. Stelmach & J. Requin, eds, North Holland, Amsterdam, pp. 567–578.

[34] Thomassen, A.J.W.M. & van Galen, G.P. (1992). Handwriting as a motor task: experimentation, modeling and simulation, in *Approaches to the Study of Motor Control and Learning*, J.J. Summers, ed, Elsevier Science Publishers BV, North Holland, pp. 113–144.

[35] Found, B. & Rogers, D. (eds) (1999). Documentation of forensic handwriting comparison and identification method: a modular approach, *Journal of Forensic Document Examination* **12**, 69–80.

[36] DeAngelis, C.M. (1997). Effects of writing surface and author's position on handwriting, *Journal of Questioned Document Examination* **6**(2), 10–17.

[37] Miller, L.S. & Harralson, H. (2005-06). An examination of tremor and distortion caused by extrinsic handwriting conditions, *Journal of Forensic Document Examination* **17**, 65–82.

[38] Hilton, O. (1969). Consideration of the writer's health in identifying signatures and detecting forgery, *Journal of Forensic Sciences* **14**, 157–166.

[39] Behrendt, J.E. (1984). Alzheimer's disease and its effect on handwriting, *Journal of Forensic Sciences* **29**, 87–91.

[40] Beck, J. (1985). Handwriting of the alcoholic, *Forensic Science International* **28**, 19–26.

[41] Galbraith, N.G. (1986). Alcohol; its effect on handwriting, *Journal of Forensic Sciences* **31**, 580–588.

[42] Walton, J. (1997). Handwriting changes due to aging and Parkinson's syndrome, *Forensic Science International* **88**, 197–214.

[43] Hecker, M.A. (1995). Gender identification through handwriting, in *Proceedings of the Seventh Biennial Conference of the International Graphonomics Society*, Phylmar Associates, Ontario, Canada, pp. 74–75.

[44] Haines, K., Phillips, J.G., Rogers, D. & Found, B. (2001). The ability to distinguish handwriting samples on the basis of writers' age and gender, *Journal of Forensic Document Examination* **14**, 31–51.

[45] Hammond, D. (1997). The history and development of laws compelling suspects to provide handwriting exemplars: a look into federal, state and military courts, *International Journal of Forensic Document Examination* **3**(1), 45–48.

[46] McAlexander, T. (1991). The standardization of handwriting opinion terminology, *Journal of Forensic Sciences* **36**, 311–319.

[47] Evett, I. (1998). Towards a uniform framework for reporting opinions in forensic science casework, *Science & Justice* **38**(3), 198–202.

[48] Ramsey Lines, S. & Franck, F. (2003). Triplet and sibling handwriting study to determine degree of individuality and natural variation, *Journal of the American Society of Questioned Document Examiners* **6**(2), 48–55.

[49] Baxendale, D. & Renshaw, I.D. (1979). The large scale searching of handwriting samples, *Journal of Forensic Document Examination* **19**, 245–251.

[50] Harvey, R. & Mitchell, R.M. (1973). The Nicole Brazier murder: the role of handwriting in large-scale investigation, *Journal of the Forensic Science Society* **13**, 157–168.

[51] Huber, R.A. (2000). The heterogeneity of handwriting, *Journal of the American Society of Questioned Document Examiners* **3**(1), 2–10.

[52] Horton, R.A. (1996). A study of the occurrence of certain handwriting characteristics in a random population, *International Journal of Forensic Document Examiners* **2**, 95–102.

[53] McClary, C.R. (1997). A study of baseline alignment in signatures and handwritten sentences, *International Journal of Forensic Document Examiners* **3**, 35–44.

[54] Ahola, N.M. (2000). Classification and frequency of occurrence of specific number styles, *Canadian Society of Forensic Science Journal* **33**(1), 13–22.

[55] Marquis, R., Schmittbuhl, M., Mazella, W. & Taroni, F. (2005). Quantification of the shape of handwritten characters: a step to objective discrimination between writers based on the study of the capital character O, *Forensic Science International* **150**, 23–32.

[56] Dyer, A., Found, B. & Rogers, D. (2006). Visual attention and expertise for forensic signature analysis, *Journal of Forensic Sciences* **51**(6), 1–8.

[57] Found, B. & Rogers, D. (1998). A consideration of the theoretical basis of forensic handwriting examination, *International Journal of Forensic Document Examiners* **4**(2), 109–118.

[58] Found, B., Rogers, D. & Herkt, A. (2001). Comparison of document examiners' opinions on original and photocopied signatures, *Journal of Forensic Document Examiners* **14**, 1–13.

[59] Found, B., Rogers, D. & Herkt, A. (2001). The skill of a group of forensic document examiners in expressing handwriting and signature authorship and production process opinions, *Journal of Forensic Document Examiners* **14**, 15–30.

[60] Found, B. & Rogers, D. (2003). The initial profiling trial of a program to characterise forensic handwriting examiners' skill, *Journal of the American Society of Questioned Document Examiners* **6**(2), 72–81.

[61] Found, B. & Rogers, D. (2005). The relative strength of forensic document examiners' identification and elimination opinions on individuals' natural writings, *Proceedings of the International Graphonomics Society*, Salerno, Italy.

[62] Found, B. & Rogers, D. (2005). Problem types of questioned handwritten text for forensic document examiners, *Proceedings of the International Graphonomics Society*, Salerno, Italy.

[63] Found, B. & Rogers, D.K. (2005). Investigating forensic document examiners' skill relating to opinions on photocopied signatures, *Science & Justice* **45**(4), 199–206.

[64] Harris, J.J. (1953). Disguised handwriting, *Journal of Criminal Law, Criminology and Police Science* **43**, 685–689.

[65] Alford, E.F. (1970). Disguised handwriting, *Journal of Forensic Sciences* **15**, 476–488.

[66] Webb, F.E. (1977). The question of disguise in handwriting, *Journal of Forensic Sciences* **23**, 149–154.

[67] Franck, F.E. (1988). Disguised writing: chronic or acute, *Journal of Forensic Sciences* **33**, 727–753.

[68] Konstantinidis, S. (1987). Disguised handwriting, *Journal of the Forensic Science Society* **27**, 383–392.

[69] van Gemmert, W.A., van Galen, G.P., Hardy, H.J.J. & Thomassen, A.J.W. (1996). Dynamical features of disguised handwriting, *Presented at the 5th European Conference for Police and Government Handwriting Experts*, The Hague, The Netherlands, November.

[70] Tsui, D.C.K. (1997). Disguised Chinese handwriting by Chinese-Canadian writers, *Canadian Society of Forensic Science Journal* **30**(3), 137–150.

[71] van Galen, G. & van Gemmert, A.W.A. (1996). Kinematic and dynamic features of forging another person's handwriting, *Journal of Forensic Document Examination* **9**, 1–25.

[72] Muehlberger, R.J. (1990). Identifying simulations: practical considerations, *Journal of Forensic Sciences* **35**, 368–374.

[73] Leung, S.C., Cheng, Y.S., Fung, H.T. & Poon, N.L. (1993). Forgery I–simulation, *Journal of Forensic Sciences* **38**, 402–412.
[74] Black, J.A. (1995). Simulated signatures – forgery by imitation, *Journal of Forensic Identification* **45**, 283–297.
[75] van Gemert, A.W.A. & van Galen, G.P. (1996). Dynamic features of mimicking another persons writing and signature, in *Handwriting and Drawing Research: Basic and Applied Issues*, M.L. Simner, C.G. Leedham & A.J.W.M. Thomassen, eds, IOS Press, Amsterdam, pp. 459–471.
[76] Black, D., Found, B. & Rogers, D. (2003). The frequency of the occurrence of handwriting performance features used to predict whether questioned signatures are simulated, *Journal of Forensic Document Examination* **15**, 17–28.
[77] Schuetzner, E.M. (2006). A study of attempted simulated signatures by teenage writers, *Journal of the American Society of Questioned Document Examiners* **9**(2), 95–103.
[78] Michel, L. (1978). Disguised signatures, *Journal of the Forensic Science Society* **18**, 149–154.
[79] Herkt, A. (1986). Signature disguise or signature forgery, *Journal of the Forensic Science Society* **26**, 257–266.
[80] Mohammed, L.A. (1993). Signature disguise in Trinidad and Tobago, *Journal of the Forensic Science Society* **33**, 21–24.
[81] Wendt, G.W. (2000). Statistical observations of disguised signatures, *Journal of the American Society of Questioned Document Examiners* **3**(1), 19–27.
[82] Leung, S.C., Fung, H.T., Cheung, Y.S. & Poon, N.L. (1993). Forgery II - tracing, *Journal of Forensic Sciences* **38**, 413–424.
[83] Kullman, D. (2002). Impression by traced forgery, *Journal of the American Society of Questioned Document Examiners* **5**(1), 28–38.
[84] Schneider-Pieters, H., ten Camps, C. & Hardy, H. (1996). The computer–friend or foe? *Presented at the 5th European Conference for Police and Government Handwriting Experts*, The Hague, The Netherlands, November.
[85] Franke, K. & Schomaker, L. (2004). Robotic writing trace synthesis and its application in the study of signature line quality, *Journal of Forensic Document Examination* **16**, 119–141.
[86] Found, B. & Rogers, D. (2007). The probative character of forensic document examiners' identification and elimination opinions on questioned signatures, *Proceedings of the 13th Biennial Conference of the International Graphonomics Society*, Melbourne, Australia, November.
[87] Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S.579, 113 S.Ct. 2786, 125 L.Ed. 2d 469 (1993).

BRYAN FOUND

Handwriting and Signatures, Interpretation of Comparison Results

Introduction

The comparison of handwriting is one of the oldest forms of identification evidence. Since ancient times the signature has been recognized as a reliable way of ensuring that the correct authority has been given by the correct person and history is littered with examples of handwriting being used to identify an individual, from wills and wedding certificates to the Royal Ascent and the Magna Carta. There are a number of books describing the application of handwriting to forensic problems [1–3]. However, with the advent of DNA and the use of electronic devices, handwriting has started to take a back seat in both forensic identification and personal identification and people have begun to take a more critical look at the science behind handwriting comparison. Used carefully, and in the right hands, handwriting comparison still has a lot to offer the investigator but its interpretation must be updated to cope with the challenges of the modern world.

A Note on Graphology

Before discussing the interpretation of handwriting comparison, the reader should be aware of the other branch of handwriting examination known as *Graphology*. Graphology is the process of determining character traits from handwriting – whether a person is extrovert, introvert, a good communicator, shy, a leader, etc. While graphology has its place in society, it is not forensic handwriting comparison and should not be confused with the process of trying to determine who wrote a specific document. There are practitioners who claim to be able to do both forensic handwriting comparison and graphology, but proficiency at one skill does not imply proficiency in the other. This article relates only to forensic handwriting comparison.

Why Handwriting Can Be Used to Identify an Individual

Handwriting comparison is based on the premise that no two people write exactly the same.[a] This cannot be proved but, as with fingerprints, the theory has never been disproved and therefore stands. A person's handwriting is the result of their physiological makeup, their motor control (their ability to control intricate motions) and their experience in writing. Basic forms of writing are taught in schools, providing the underlying style of the writing, but a person very quickly moves away from the taught style to suit their own personal circumstances. By the age of 18 a person's writing style is a habitual form, which they can use fluently without thinking and which is very difficult for them to alter without sacrificing speed. The habitual writing form for most people is so complex and there are so many variables that the assumption that "no two people's writing is ever exactly the same is reasonable and in accord with long experience.

Overlying the physiological and experiential influences on a person's handwriting, which are in-built, are the immediate physical influences such as the health of the individual, the influence of drink and drugs, writing surface, writing implement, the ambient temperature, clothing, external movement, and so forth, all of which may influence the appearance of the written work on the page.

Handwriting is a very complex image to analyze and the writing of one person contains a degree of natural variation that must be assessed and taken into account when making a handwriting comparison. Handwriting has often been described as simply pattern recognition, and likened to fingerprint comparison or footwear comparison. However, it is more complex than that, because it can be deliberately varied and deliberately simulated. These factors are difficult to build into an interpretative model and are what makes the comparison a highly skilled process that should not be underestimated in its difficulty. Nevertheless the application of an appropriate interpretation framework (on the basis of the concept of likelihood ratio) to the interpretation of handwriting comparison helps to mitigate most of the major errors. Moreover, it makes the interpretative process more transparent. In a series of papers by Evett *et al.* [4], three general principles emerge. They are as follows:

1. To evaluate evidence it is necessary to consider at least two propositions.
2. Evidence is evaluated in the light of the other (relevant) information in the case.
3. It is necessary to consider questions like "what is the probability of the evidence"?

A Short History of Handwriting

Handwriting comparison has been instrumental in many high profile prosecutions; the prosecution of Jeffery Archer for perjury, the detection of the Harold Shipman murders through the identification of a forged will, the conviction of the Brighton bomber, and so forth. However, the reliability of handwriting as a forensic tool is increasingly being called into question on both sides of the Atlantic more by perception than by reality [5, 6]. To understand the difficulties in the interpretation of handwriting and the need for a more modern approach, we must first look at the history behind handwriting comparison. There are many literary examples of the use of handwriting to identify an individual. Some are successful, such as that carried out by Sherlock Holmes in *The Sign of Four*, but many revolve around the false identification of an individual. In William Shakespeare's Twelfth Night, Malvolio discovers a letter and states:

> By my life, this is my lady's hand! These be her very C's, her U's and her T's; and thus makes she her great P's. It is, in contempt of question, her hand

The letter is in fact a careful simulation of Olivia's writing, but his erroneous conclusion results in his humiliation, spending the rest of the play grinning and wearing cross-gartered stockings. In the Winslow Boy by Terrence Rattigan the court case hangs on the testament of the handwriting expert, who is subsequently discredited, thus causing the case against the Winslow boy to collapse. There are also real-life examples of errors of interpretation resulting in miscarriages of justice and humiliation of the individuals concerned (the case against Alfred Dreyfus for example [7]). Perhaps the most famous of recent years is the Hitler Diaries, which were announced as genuine by several eminent historians and scientists. Unfortunately they proved to be forgeries, the mistake being made because some of the specimen writing was in fact not the writing of Hitler but of the forger. These examples show where the inherent weaknesses of handwriting comparison lie. Firstly,

handwriting examiners are too ready to "transpose the conditional" and infer that matching handwriting implies that the individual is the author, and secondly that the ability of someone to simulate another's writing is very difficult to estimate in scientific terms.

Recent Criticism of the Interpretation of Handwriting

Two factors have emerged in the last 20 years and this has meant that handwriting examiners have had to re-evaluate the way they interpret their findings: the Starzepyzel judgment in the United States and the emergence of DNA as an identification science.

In 1993 there was a trial in the United States in which the right of the handwriting examiner to give evidence was questioned. The gist of the argument was that handwriting comparison was not a science and therefore the identification evidence gained through handwriting comparison was inadmissible; consequently the examiner should not have expert status. The judge held a "Daubert" hearing to determine *inter alia* whether handwriting comparison was a scientific endeavor. The trial judge's conclusion was that it was not, but that handwriting comparison was more practical in character. He likened handwriting examiners to harbor pilots, guiding ships into dock. They were much more familiar with the rocks and shallows of the harbor than the captain of the ship and therefore they had an expertise that was very useful, but only in certain circumstances: put them in another harbor, and their knowledge is useless. He allowed forensic document examiners to testify to their examinations and comparisons, but prohibited them from testifying to the ultimate conclusion they had reached [8]. While the opinion of the *Starzecpyzel* judge is not followed by the majority of other courts, it is nevertheless legitimate to argue in the United States that handwriting is not appropriate in certain circumstances. This judgment led to some more fundamental criticism of handwriting comparison in the United States by a group of eminent lawyers. Faigman summarized it all in 1999 when he criticized handwriting examiners as not moving with the times [9] and Risinger, Denbeaux, and Saks [5] have written a number of articles exploring the weaknesses of handwriting comparison, the most famous of which is equating handwriting comparison to witchcraft [5, 6]. The main criticisms can be summarized as the lack of reproducibility, the lack of consistent and universal standards, and the lack of formal qualifications to practice. Some of these issues have subsequently been addressed. To answer reproducibility, Kam and others in the United States [10, 11] have published a series of papers looking at the performance of handwriting examiners and Bryan Found of the University of La Trobe has developed a testing regime for State examiners in Australia so that the skills of examiners can be assessed against known results. This has led to written standards and formal qualifications [12, 13]. In Europe, the European Network of Forensic Institutes (ENFSI) working group for handwriting has also been working on a set of written standards and in the United States there are formal qualifications to be gained from the American Society of Questioned Document Examiners (ASQDE) and the American Board of Document Examiners (ABFDE), and again written standards are to follow. In the United Kingdom all state examiners are expected to gain admittance to the Council for the Registration of Forensic Practitioners (CRFP), which ensures that their work is regularly reviewed by other practitioners. However, the fundamental question of whether the examiner has interpreted the results in the correct way still remains, because the traditional regime relied on a holistic assessment based on the examiner's training and experience.

Handwriting and DNA Interpretation

Handwriting examiners have always interpreted their findings in terms of authorship, reaching such conclusions as "Mr X was the writer of the check" or "Mrs Y probably wrote the mortgage application". The emergence of DNA as an interpretive framework in the 1990s led to questions about the way scientific evidence generally was interpreted. In the UK court of appeal (R. v. Doheny and Adams 1 Cr App Rep 369, 375) the court ruled that it was the function of the scientist to inform the court of the likelihood of the evidence (i.e. the similarity in the DNA of the questioned and specimen samples) if the DNA came from a suspect or, alternatively, if the DNA came from someone other than the suspect (the *random occurrence ratio* as named by the court). However. the scientist had no right to tell the court that the DNA came from the suspect (*which would be transposing the conditional*). Fundamentally it is the court's role to interpret the scientific evidence in

the light of all the evidence in front of it, not the scientist's. The implications for handwriting comparison are that, however good the match, the scientist's interpretation can only be probabilistic and therefore only support the view of authorship.

Handwriting Comparison – the Basics

Handwriting comparison for forensic purposes is the comparison of the letterforms in one piece of writing with that in another. It is done in three stages: investigation, comparison, and evaluation. We deal with each stage in turn, although in reality an experienced examiner often performs all three stages at the same time.

Investigation

A handwriting examiner often states that the handwriting comparison is not dependent on outside influences and in this sense is a pure comparison. However, this may not be true. Peirce [14] gives many examples of the ways in which a forensic scientist may reason at the scene of a crime. At this stage, there may be no suspect and part of the role of the scientist is in advising the investigator about how the incident took place. Many hypotheses may be informally generated by the scientist and most are discarded as they become incompatible with subsequent observation and reasoning: others survive and may become refined by the particular circumstances of the case. The process of hypothesis generation and refinement characterizes the scientist as *investigator*.

Later in the same case, we find the scientist at the laboratory supervising the examination of materials from the scene and, perhaps, carrying out comparisons with other materials taken from a suspect for the crime. Here the process of hypothesis refinement may continue, culminating wherever possible in a pair of propositions that might represent prosecution and defense positions at court. This stage clearly represents the scientist as *evaluator*.

It is possible to see these two roles emerging as a document examiner first approaches a handwriting comparison.

Consider a case that consists of a passage of questioned handwriting and request specimens from a person, Mr X. The request is, did Mr X write the note? When an examiner first examines the questioned and the specimen material, they have in their mind the two possible alternative propositions, which are either of the following two:

- The documents were written by the same person.
- The documents were written by different people.

The likelihood ratio will be given by the probability of the evidence given the first statement is true divided by the probability of the evidence given that the second statement is true, or

$$\mathrm{LR} = \frac{\begin{array}{c}\text{The probability of the evidence}\\ \text{given Mr X wrote the note}\end{array}}{\begin{array}{c}\text{The probability of the evidence}\\ \text{given that someone other than}\\ \text{Mr X wrote the note}\end{array}} \quad (1)$$

At this stage, what the "evidence" comprises is yet to be determined. Before the examiner can address these alternative propositions, they must first address the joint issues of disguise and simulation, and to do this they must refine the propositions.

Refining Propositions

If we translate this back into a body fluids case, the likelihood ratio is as follows:

$$\mathrm{LR} = \frac{\begin{array}{c}\text{The probability of the DNA evidence}\\ \text{given Mr X was the source}\end{array}}{\begin{array}{c}\text{The probability of the DNA evidence}\\ \text{given that someone other than}\\ \text{Mr X was the source}\end{array}} \quad (2)$$

where the source of the DNA is a recovered blood stain. In this case the DNA expert can simply determine the DNA profile of the stain, assign probabilities in the two propositions and establish the likelihood ratio. With handwriting comparison, the process is not so simple. For example, if we asked the examiner, before they looked at the handwriting, "do you expect to find many similarities and a few differences between the two sets of handwriting, if the prosecution proposition is true" then they would rightly say that the number of similarities and differences depends on many factors about which she has no knowledge. If the two pieces of writing are naturally written and both are extensive then the examiner would expect to find many similarities and a few significant differences. However, if the specimen writing is disguised the examiner would expect to find many differences, but also a few significant similarities. Similarly, if the questioned writing is simulated, another set of expectations would arise. The examiner, therefore, has to act as an investigator to refine the propositions before they can move on

to the comparison and evaluation stages. We could think of a range of ways in which the prosecution proposition could be refined to make the question easier to answer; one such method is to break the main propositions into component parts (a concept known as *nested* propositions):

Mr X wrote the questioned handwriting

H_{p1}: in his normal natural hand;
H_{p2}: in a disguised hand;
H_{p3}: while suffering from a temporary debilitating illness;
H_{p4}: while under the influence of drinks or drugs;
H_{p5}: while under the influence of some environmental factor (e.g., while in a moving vehicle; in an awkward position; against an unsuitable surface);
H_{p6}: with his unaccustomed hand.

Likewise, the examiner's expectations would be easier to assess if the defense examination were also broken down. Possible alternatives include the following:

Someone other than Mr X wrote the questioned handwriting:

H_{d1}: in their normal style;
H_{d2}: as a simulation of Mr X's handwriting;
H_{d3}: in a disguised hand.

To ease the assessment of a relevant likelihood ratio, the examiner must refine the propositions. She can do this in a variety of ways. The simplest procedure would be to have a quick look at the documents *before* speaking to the customer. She may look for evidence of disguise or simulation (such as a lack of fluency or amended or inconsistent letterforms) and she would check to ensure that both the specimen writings and the questioned writing are consistent with being the work of one person. During this investigative phase, she may speak to the investigator to gather information about the way the samples were taken or the location where the note was found (e.g., if it was found in the back of a car, then movement could account for inconsistencies; if it was written in front of a bank clerk, then movement is very unlikely to be the cause).

By conducting this investigative phase, some possibilities can be eliminated. Assume that, as a result of this first examination, the examiner observes that both the questioned and specimen writings are fluent and that both are written in a very distinctive hand. As a consequence the evidence allows refinement of the propositions and they now become as follows:

Mr X wrote the questioned handwriting:

H_{p1}: in his normal natural hand.

Someone other than Mr X wrote the questioned handwriting:

H_{d1}: in their normal style;
H_{d2}: as a simulation of Mr X's handwriting.

A full and detailed comparison of the writing can now be conducted, which should reveal whether any differences present between the writings are the result of natural variation, different writers, or simulation. In most cases, this is straightforward and leads the layman into believing that handwriting comparison is simply a matter of looking for similarities (e.g., Malvolio); they do not look for the differences.

If the investigative phase reveals that the questioned writing is very hesitant while the specimen writing is very fluently written and there are many obvious differences between the writings, then the expectations of the examiner will alter, and the propositions are refined in a different way. They become as follows:

Mr X wrote the questioned handwriting:

H_{p2}: in a disguised hand;
H_{p3}: while suffering from a temporary debilitating illness;
H_{p4}: while under the influence of drink or drugs;
H_{p5}: while under the influence of some environmental factor (e.g., while in a moving vehicle; in an awkward position; against an unsuitable surface);
H_{p6}: with his unaccustomed hand.

Someone other than Mr X wrote the questioned handwriting:

H_{d1}: in their normal style;
H_{d3}: in a disguised hand.

The most important factor to note here is that the examiner would not now look for lots of similarities (as they know that they do not exist) but would assess the nature of the differences and whether they would be expected to occur if any of the subsets of the main proposition is true. Sometimes they are much more likely to occur if one particular nested proposition is true, and this can shift the likelihood ratio in one direction or the other, but on other occasions (particularly simple signatures) no one proposition

stands out, and the likelihood ratio may remain near to one.

Two important points emerge from this reasoning: first, it is very difficult to place actual numbers of the probabilities we are considering here, and therefore a likelihood is always qualitative and is based on the experience of the examiner. While some studies have been conducted into the occurrence data for individual letterforms, which would give values that could be put into a likelihood ratio, this would only be applicable if the alternatives to be considered were as follows:

Mr X wrote the questioned handwriting

H_{p1}: in his normal natural hand.

Someone other than Mr X wrote the questioned handwriting:

H_{d1}: in their normal style.

This is rarely the case, but again is often assumed to be the case by the layman. It is much more difficult to estimate the probability of the evidence given that the writing is disguised or simulated, although qualified examiners will be able to make an estimate of this on the basis of their experience.

The second is that the investigative exercise does not eliminate any nested propositions but only informs which ones are going to be most influential on the likelihood ratio. Therefore it is difficult to justify the definite opinion on this argument.

Comparison

The next part of the examination is to look at the writing in detail and assess how similar the questioned and specimen writing is, in the light of the investigative findings. This is generally done by the systematic comparison of each character in turn, comparing its structure (the way the pen moves in constructing the character) its shape, its variation (some people have two different forms of letter or vary the letter according to its position in the word) and the characters' proportions with respect to adjacent characters (e.g., some people always write a "P" much bigger than a "O"). Even a small piece of handwriting contains a wealth of information and it is therefore not surprising that the degree of match between the questioned and known characters is often a point of debate between two handwriting examiners. However, the disagreements are usually about magnitude and range of variation rather than fundamental differences of opinion and are influenced by the experience of the examiner. Once this assessment has been done for all the characters and other features present in both, the questioned and specimen writing, then the examiner will have a list of the similarities and the differences between the writings. They then assess the significance of the match or mismatch in the light of the previously conducted investigative phase.

Interpretation

As has been previously described, there are some inherent problems to be overcome in the interpretation of evidence gained from a handwriting comparison. It is not sufficient to simply say "these writings are similar" or "these writings are different" because both observations can occur whether or not the writings are by the same person. In terms of the harbor pilot, making observations, without deciding upon a likely course, leaves the ship still at sea.

The first challenge the handwriting examiner is faced with is inherent variability of one person's writing. If they see a difference between one piece of writing, does this mean that the writing is by a different person, or simply that the writer has a wider range of variation than is demonstrated by the specimen writing? This assessment often comes down to a matter of experience and this is why comparisons in scripts which are not intimately known by the examiner, should be treated with caution. The judge in the Starzecpyzel case said that the handwriting examiner was more like a harbor pilot guiding a ship into port than a scientific expert. By implication, they are more likely to be reliable when they are in waters they are familiar with, but that is not to say that they cannot be of some use in other arenas, as they will be able to interpret the signs better than a layman.

The second challenge is deliberate disguise; if the two pieces of writing appear very different, then disguise of one or both pieces of writing must be considered as a possible explanation for the differences. In larger amounts of writing disguise is often difficult to maintain but in smaller amounts it has to be considered as a real possibility and the results of these examinations are often inconclusive.

Signatures

The treatment described earlier, while being focused on handwriting, can equally be applied to signatures. However, in this case the examiner is now dealing with a small amount of writing, often written in a very personalized way. The effect of this is to alter the importance and influence of some of the subpropositions described earlier. For instance, assume we have a signature that is so personalized that one is unable to distinguish individual letters. Consider the nested prosecution and defense propositions described earlier:

Mr X wrote the questioned signature:

H_{p1}: in his normal natural style;
H_{p2}: in a disguised hand;
H_{p3}: while suffering from a temporary debilitating illness;
H_{p4}: while under the influence of drink or drugs;
H_{p5}: while under the influence of some environmental factor (e.g., while in a moving vehicle; in an awkward position; against an unsuitable surface);
H_{p6}: with his unaccustomed hand.

Someone other than Mr X wrote the questioned signature:

H_{d1}: in their normal style;
H_{d2}: as a simulation of Mr X's handwriting;
H_{d3}: in a disguised hand.

The examiner quickly undertakes the investigative stage of the examination, and finds that the questioned and specimen signatures look the same in general appearance. The possibility of this occurring is much higher for some of the subpropositions than the others, and the propositions, for practical purposes, can be reduced to the following:

Mr X wrote the questioned signature:

H_{p1}: in his normal natural style.

Someone other than Mr X wrote the questioned signature:

H_{d2}: as a simulation of Mr X's handwriting.

The chance of a coincidental match is so small as to be eliminated, and the crux of the examination becomes whether or not someone would have the skill to be able to simulate the signature in question. To answer this, the examiner then has to draw on their experience as to the complexity of the signature and the ability of people to simulate this complexity of signature this well. It can be seen that signature comparison draws more heavily on the skill and experience of an examiner than handwriting comparison, but the principles for comparison and evaluation are the same.

Summary

In summary, the concept of a likelihood ratio can be applied to handwriting comparison in much the same way as to other branches of forensic science. However, because the examiner has an investigative role to play in making a handwriting comparison there is an added complexity to the examination. This can be addressed by using the concept of "nested" or subpropositions.

End Notes

[a.] The same here refers to a holistic view of the writing, rather than the exact reproduction of each written line. In fact, even one person never writes "exactly the same" on two different occasions. The premise is that "within person variation is smaller than between person variation".

References

[1] Huber, R.A. & Headrick, A.M. (1999). *Handwriting Identification: Facts and Fundamentals*, CRC Press, Boca Raton.

[2] Ellen, D. (1997). *The Scientific Examination of Documents – Methods and Techniques*, 2nd Edition, Taylor and Francis.

[3] Hilton, O. (1982). *Scientific Examination of Questioned Documents*, Revised Edition, Elsevier.

[4] Cook, R., Evett, I.W., Jackson, G., Jones, P.J. & Lambert, J.A. (1998). A model for case assessment and interpretation, *Science and Justice* **38**, 151–156.

[5] Risinger, D.M., Denbeaux, M.P. & Saks, M.J. (1898). Exorcism of ignorance as a proxy for rational knowledge: the lessons of handwriting identification "expertise", *University of Pennsylvania Law Review* **137**, 731–792.

[6] Saks, M.J. & Risinger, D.M. (1966). Science and non science in the courts: Daubert meet handwriting identification expertise, *Iowa Law Review* **82**, 21–74.

[7] Champod, C., Taroni, P. & Margot, P. (1999). The Dreyfus case – an early debate on conclusions, *International Journal of Questioned Document Examiners* **5–44**, 6–59.

[8] United States v. Starzecpyzel, 880 F.Supp. 1027 (SDNY 1995).

[9] Faigman, D.L. (1999). *Legal Alchemy – The Use and Misuse of Science in the Law*, W. H. Freeman, New York.
[10] Kam, M., Fielding, G. & Conn, R. (1997). Writer identification by professional document examiners, *Journal of Forensic Science* **42**, 778–786.
[11] Kam, M., Gummadidala, K., Fielding, G. & Conn, R. (2001). Signature authentication by forensic document examiners, *Journal of Forensic Sciences* **48**, 884–888.
[12] Found, B., Sita, J. & Rogers, D. (1999). The development of a program for characterizing forensic handwriting examiners expertise: signature examination pilot study, *Journal of Forensic Document Examination* **12**, 69–80.
[13] Sita, J., Found, B. & Rogers, D.K. (2002). Forensic handwriting examiners expertise for signature comparison, *Journal of Forensic Sciences* **47**, 1117–1124.
[14] P. Weiss, C. Hartshorne, eds, (1934). *Collected Papers of Charles Sanders Peirce*, Harvard University Press.

STEPHEN P. DAY

Handwriting Comparison *see* Handwriting and Signatures, Comparison of

Hardy-Weinberg Equilibrium

The Hardy–Weinberg equilibrium is a state in which the allelic frequencies do not change. Its forensic significance is that in the process of calculating the expected frequencies of genotypes in the population, this equilibrium, or steady state, is assumed as a starting basis.

Given that any person will normally have two alleles at each locus (although both may be the same), the allelic frequencies can be used to calculate genotype frequencies (paired combinations of alleles) by multiplication.

In a simple example, if there are only two alleles (A or B) possible at a locus then there are only three types of individuals: AA, BB, and AB. If the frequency of allele A is p, and the frequency of allele B is q, it is possible to calculate the frequency of each type of person in the population *assuming the Hardy–Weinberg rules*. One of these is "random mating". The alleles exist in pairs within men and women, but the pairs separate in the formation of sperm and egg so, in effect, the next population is a sample of drawing two alleles together, one from the male and one from the female. For p As and q Bs, it is quite simple algebra to calculate the result of one round of mating where a child will receive (A or B) and (A or B):

$$(p + q) \times (p + q) = (p + q)^2 = p^2 + 2pq + q^2 \quad (1)$$

This gives the frequency of AA people as p^2, AB people as 2pq, and BB people as q^2.

Thus, an allelic frequency database can be used to calculate a random match probability (RMP) to an individual: the probability that one could pick a person who would have the same genotype (profile) at random from a population of unrelated individuals. In forensic applications, the loci chosen will have many more than just two alleles, i.e., A, B, C, D, etc. with frequencies p, q, r, s, etc. The principle for the calculation of genotype frequencies is the same as above, just involving a larger equation (1) for the whole population. Then, if you want to know the RMP for a person who has the genotype of BD, the algebra dictates that it would be calculated as 2qs, the RMP for a person who is CC would be r^2, and so on. This is the value for one locus. For a group of loci as used in a profiling multiplex, the RMP is derived by calculating the combined genotype frequencies using the product rule such that the genotype frequency for each locus is multiplied together.

The full set of "rules" of the Hardy–Weinberg Equilibrium for a population are that it must be large, there must be random mating of individuals, no mutations, no migration out of or into the population, and no selection acting on any of the individuals. If these rules are not met, the population is not in equilibrium and the allele frequencies will be changing.

For the purposes of forensic DNA profiling, the most important consideration when doing the RMP calculation using the equations above, is the effect of selective interbreeding of small populations in the real world, whether geographically or culturally isolated to some degree. This nonrandom mating will alter the frequency with which some alleles will be

seen compared with the frequencies of a wider population from which a sample database may have been collected. There is some debate on how to accommodate this, and other possibilities, into calculating the statistics of the random match probability for a profile (*see* **Short Tandem Repeats: Interpretation**; **Mixture Interpretation: DNA**). Most commonly a correction factor is incorporated into the calculation for genotype frequencies to allow for population substructures [1–2].

References

[1] Balding, D.J. & Nichols, R.A. (1994). DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands, *Forensic Science International* **64**, 125–140.

[2] National Research Council Committee on DNA Forensic Science (1996). The Evaluation of Forensic DNA Evidence. National Academy Press, Washington, D.C.

Further Reading

Butler, J.M. (2005). Forensic DNA Typing 2nd Edition, Elsevier.

Rudin, N. & Inman K. (2002). An introduction to Forensic DNA Analysis, 2nd Edition, CRC Press.

SCOTT BADER

Hare Psychopathy Checklists *see* Psychopathy Checklists

HCR-20 *see* Risk Assessment: Patient and Detainee

Head Injury: Neuropsychological Assessment

Traumatic brain injury (TBI) is defined by the International Neuropsychological Society (INS) Dictionary of Neuropsychology as injury to the brain resulting from "an external force such as a blow to the head, concussive forces, acceleration–deceleration forces, or projectile missile (e.g., bullet)" ([1], p. 160). Most incidents of TBI are caused by falls, with motor vehicle crashes and violent assaults also causing significant percentages of TBI as well. The National Center for Injury Prevention and Control estimates that approximately 1.4 million Americans sustain a TBI every year resulting in nearly 50 000 deaths, a quarter of a million hospital stays, and over a million treat and release cases handled by emergency departments throughout the country. These statistics are very likely to be a significant underestimation of TBI's prevalence, as many persons who experience a head injury never seek medical care for their ailments [2]. The Center for Disease Control and Prevention (CDC) estimates that nearly 2% of the country's population, over 5 million citizens have long-term deficits resulting from TBI [3]. The cost to the country's economy resulting from TBI is staggering with estimates as high as 50+ billion dollars yearly [4]. Of course, this does not take into consideration the personal toll that TBI takes on those afflicted and their loved ones and dependents.

After sustaining a TBI, regardless of its cause, first responders and medical professionals typically assess the severity of the injury with behavioral rating scales such as the Glascow Coma Scale (GCS). This rating scale has three individual components on which the patient's ability to respond is rated across three different areas: eye opening response; verbal response; and motor response. The sum of the three scales ranges from a minimum score of 3 to a maximum score of 15 [5]. GCS ratings are taken upon initial contact with the TBI patient, and often ratings will be taken at various points in time after initial evaluation. For example, ongoing assessment of consciousness may be needed to gauge changes arising from the neurological insult (e.g., a subacute, subdural hematoma), or other elements not directly related

to the injury itself such as states of intoxication at the time of the injury that will affect the patient's level of consciousness as time passes and sobriety sets in. It is generally accepted that an individual has sustained a "severe" TBI when their GCS score is in the 3–8 range, a "moderate" TBI with scores in the 9–12 range, and a "mild" TBI with a score in the 13–15 range. In emergent contexts, assessment with the GCS often dictates the next course of assessment and treatment taken by medical professionals. For example, severe TBI patients, as measured by GCS, may require a critical care physician to address issues relevant to appropriate airway management and to assist with resuscitation. A request for neuroimaging (e.g., computed tomography (CT)) may be the first assessment procedure required for persons with GCS ratings in the moderate to mild range. Moreover, scales such as the GCS may be relevant to predicting outcome [6]. However, one must always consider the circumstances surrounding the time period in and around the TBI, the specific nature of the physical injuries sustained at the time of the TBI, and any other treatment issues relevant to the case (e.g., if the person had to be sedated because of agitation) before drawing conclusions about GCS scores as relevant predictors of adjustment to the TBI.

The CDC estimates that nearly three-quarters of TBI in a given year are due to concussions or other mild forms of TBI resulting in changes across the neurocognitive spectrum, in particular, attention and memory. The Diagnostic and Statistical Manual of Mental Disorders Fourth Edition Text Revision (DSM-IV-TR) [7] addresses the sequelae of mild TBI by proposing a new diagnostic category for this phenomenon, postconcussional disorder. Concussion typically involves an incident resulting in loss of consciousness for a brief period of time (e.g., 5 min), loss of memory after the traumatic event, and possibly posttraumatic seizures within 6 months of the injury. However, loss of consciousness may not be a necessary condition for postconcussive symptoms to emerge [8, 9]. Symptoms associated with postconcussional disorder may include becoming fatigued more easily, sleep disturbance, headaches, bouts of dizziness, irritability, changes in mood such as depression, disinhibition, and reduced motivation. The changes in cognition may be observed after the traumatic incident, typically a closed head injury, and for many persons may resolve overtime. Of course, more serious TBI cases effect change in various aspects of thinking (attention, memory, and reasoning), sensation (touch, smell, and taste), language (expression and reception), and emotion (anxiety, mood, personality functioning, and social skills). All of these functional areas can be formally assessed and compared with premorbid abilities.

The Context of TBI Forensic Evaluation

There are many reasons for which a formal forensic assessment of the TBI patient may be requested (*see also* **Neuropsychological Assessment**). In civil law cases, psychologists, psychiatrists, neurologists as well as rehabilitation medicine specialists with specific training in TBI are often needed to inform the trier of fact. Certainly, expert testimony in tort cases for both plaintiff and defendant is often called for to help understand the impact of TBI upon the alleged victim, and its likely course overtime. Forensic evaluators give expert opinion on both sides of the adversarial process in courts, often explaining the severity of injury, its specific manifestations, and amenability to treatment. Some cases may not involve tort issues *per se*, but custodial matters and/or civil liberties. For example, a person's capacity to be a responsible parent may be called into question after a TBI, or one's ability to live independently, control one's finances, or engage in contracts (including making a will) may be the relevant issue of inquiry.

In the criminal jurisprudence arena, there are a number of areas in which a TBI defendant's neurocognitive capacities are questioned and may be relevant. This includes a defendant's ability to comprehend verbal information such as Miranda warnings, to rationally engage with his or her defense counsel, and to understand the criminal charges and possible implications (competency to stand trail). In severe cases of TBI, there may be an issue of whether the defendant can ever be restored to competency to stand trial, pursuant to *Jackson v. Indiana*. There may be issues of criminal culpability that surface and as such, a formal assessment of the defendant's mental state at the time of the alleged criminal incident will be investigated. These criminal responsibility (i.e., insanity) evaluations may play a pivotal role in how the court decides a case, that is, whether or not the defendant will be held criminally responsible for his or her actions. Neuropsychological impairment could certainly impact the defendant's *mens rea*, or capacity to form the specific intent for the crime charged, and

how the perpetrator of the criminal event thought about his or her actions at the time. Such information is also relevant to the sentencing phase of a criminal trial, that is, what may happen to the defendant after the guilt phase is completed (e.g., incarceration vs. hospitalization). Forensic assessment of TBI may determine in what type of correctional setting a guilty defendant will be placed, or even whether TBI symptoms are considered aggravating or mitigating factors that the trier of fact (a jury and/or a judge or panel of judges) considers before deciding to sentence a defendant to death in capital cases [10].

Elements of the TBI Forensic Evaluation

Before one can begin to understand how TBI has impacted the allegedly brain injured patient's life (if at all), it is necessary to collect information from collateral sources to gain a functional understanding of how the patient was getting along prior to the critical incident. The forensic evaluator must get a thorough family history, educational background, occupational history, medical history, mental health history, substance abuse history, and history of contacts with the criminal justice system (juvenile and adult). It is preferable to obtain written records to substantiate any reported information. The evaluator should also contact and interview persons who are familiar with the TBI patient and his or her life history. This may include family, friends, previous employers (e.g., supervisors), teachers, etc. Some indication of the TBI patient's premorbid neurocognitive status is also needed if the evaluator is to reliably assess for the presence of deficits. Forensic neuropsychological assessment mostly involves intraindividual comparison of abilities such that premorbid functioning and current performance on tests are compared, and differences between these ability levels are determined and analyzed. Neuropsychological testing must also be compared to appropriate normative samples. If direct measures of premorbid abilities (e.g., achievement tests, intelligence tests, school grades) are not available, one may have to estimate a person's premorbid abilities. This usually includes testing an individuals reading ability (or test of verbal capacities), and combining the obtained scores with demographic information (e.g., age, sex, race, education, and occupation). Reading test scores are noted to be relatively unaffected by TBI and are referred to as *hold scores*. This, of course, is not true if a person has sustained a TBI affecting those areas of the brain specifically utilized in language ability (for a full discussion of different methods for estimating premorbid abilities, please read the article entitled "The Rationale of Deficit Measurement" [10]).

There are far too many neuropsychological assessment tools to mention in this brief overview. For a comprehensive presentation of this field of study, the reader is referred to updated texts such as Lezak, Howieson, & Loring's *Neuropsychological Assessment* [11], and Mitrushina, Boone, Razani, and D'Elia's *Handbook of Normative Data for Neuropsychological Assessment* [12]. The following areas should be thoroughly assessed: the patient's level of consciousness at the time of the assessment(s); orientation to person, place, time, and circumstance; attention and concentration; memory for both verbal and nonverbal material across short and long time spans; receptive and expressive language skills; constructional abilities; executive functions; judgment and insight; affect and mood. Each of these areas is briefly described below.

Orientation

A good neuropsychological assessment must first account for the how alert the patient is at the time of the assessment procedures. There are times when patients arrive for an assessment after taking prescribed narcotic analgesics for pain they report experiencing subsequent to an injury. Patients have arrived intoxicated on alcohol or seemingly hung over, yet they report that they are prepared to proceed with the evaluation. In serious cases of documented TBI, it is not unusual to observe patients who fatigue rapidly while working on challenging neuropsychological tasks. These observations are critical for the assessment. However, under such circumstances it is nearly impossible to determine if deficits produced on testing are a product of the altered state of consciousness which may be a reversible condition, or some underlying deficit sustained as a result of the TBI.

One's orientation to self refers to not only knowing who the person is (e.g., one's name), but other pertinent biographical information as well (e.g., date of birth, place of birth, names of siblings, children, etc.). Orientation to place accounts for present knowledge of the patient's whereabouts, where the patient presently resides. Time orientation is typically assessed by asking the patient the date, day of the

week, season, and time of the day. Finally a patient should be asked how he or she understands the reason for his or her being with you, the evaluator. This is an excellent time to correct any misunderstanding the patient may have of the role of the evaluator, and to inform the patient that there may be limitations to confidentiality entailed in the assessment. For example, the evaluation may be court ordered, or requested by the opposing side in a tort case.

Attention and Concentration

Assessment of attention is certainly critical to understanding how well an individual responds to demands in his or her environment. Assessment involves "tuning in" what is important, while "tuning out" extraneous information not specifically relevant to the completion of some demanded task. Patients may be requested to repeat strings of numbers, varying in length, both forwards and backwards. Concentration typically involves more sustained effort across a specific time span. For example, the patient may be asked to cross out a specific letter on a page filled with letters and numbers and to work as quickly as one can. Looking at how accurately one can do such a task as well as how quickly it can be completed tells the evaluator much information about an individual's concentration skills and efficiency of thinking.

Memory

Memory tasks are critically important because they tap one's ability to learn new information, thereby indicating how well one can adapt to new challenges. Verbal learning tests often look at an individual's ability to learn a list of words read aloud multiple times, or a story (a complex prose paragraph). Typically, these tests require the information to be verbally reported immediately after the information is presented, and then after a time delay, usually at least 20 min or more. There are nonverbal learning tasks as well that involve committing to memory drawn figures or pictures depicted on a page and displayed to the examinee for a specific period of time. After a free recall component following a time lapse, many tests will require the examinee to recognize and discriminate what he or she was asked to commit to memory. For example, the examinee may be told to choose the words he or she learned previously from a longer list of words, some that were previously presented and others that were not (distractors). It is not uncommon to find TBI patients who do poorly when attempting to freely recall information, but they can recognize what they had been asked to learn with very fine discrimination (called a *retrieval deficit*). Memory tasks are designed to assess types of amnesia that may afflict persons with TBI.

Language

An individual's ability to communicate can be adversely impacted by TBI. The entire spectrum of speech from production (e.g., motor skills, word knowledge, syntax, vocal inflection, and facial expressions) to receptive abilities (e.g., carrying out verbal and/or written commands) should be assessed. Reading ability should be tapped as well. Particular attention should focus on the person's capacity to think abstractly.

Constructional Ability

Constructional tests usually require that the examinee reproduces pictures (drawing) or puts together designs of varying complexity with different color blocks. Careful attention to the examinee's approach to the task often reveals how one perceives the test's challenge, plans to complete the tasks, and then carries out the plan with efficiency and speed. These tasks are usually referred to as *visual-motor*.

Executive Functioning

Executive functions have been described as the "most complex of behaviors", which are "intrinsic to the ability to respond in an adaptive manner to novel situations" having four components: volition (being aware of the goal of the act); planning (identifying and organizing the steps to attain the goal); purposive action (carrying out the plan and if need be, adjusting the plan to changing circumstances and contingencies); and effective performance (self-monitoring and self-correcting) [11, pp. 611–638].

Judgment

Many neuropsychologists utilize tests that tap into social judgment. Questions may be posed that require the examinee to generate socially acceptable solutions to a problem (e.g., "What should one do if he or she discovers a fire in a crowded building?"). The

ability to generate appropriate solutions to interpersonal problems can be negatively impacted by brain injury, especially injuries affecting the frontal lobes. In addition, it has long been known that TBI patients will sometimes get extremely frustrated in stressful encounters with others. As a consequence, they may exhibit exceptionally poor responses to stress including acting out with episodes of verbal aggression or even violence [13]. Arrest rates for criminal acts in the TBI population are well above the arrest rate in the population at large [14], and the percentage of inmates in prisons who have reported sustaining a TBI is well above the prevalence rate of TBI in the nonincarcerated population [15].

Insight

It is also important to assess how the brain injured examinee views his or her functional capacities since the TBI itself. The ability to recognize that one has sustained deficits in functioning is a critical step in the rehabilitation of the individual. Yet, there are some persons who are incapable upon reflecting and evaluating their own abilities, showing poor insight into their condition since the TBI. If one is unable to recognize under what conditions one is incapable of completing a specific task, he or she will be unable to implement newly learned skills or utilize mental prosthetic devices so often an integral part of the rehabilitation plans.

Mood and Affect

Some indication of a person's level of anxiety and mood at the time of the examination process is necessary. An examinee may be highly anxious at the time of the assessment for various reasons. The adversarial nature of court proceedings and the reasons for the testing could produce great anxiety in some examinees. The tasks themselves are often challenging and may produce stress that the examinee feels ill-equipped to cope with. Highly anxious examinees may perform poorly or the anxiety levels may negatively impact test performance [15]. Depressed TBI patients, especially those with vegetative signs of depression, often exhibit a variety of cognitive deficits that may be reversible if the depression remits after treatment; this is particularly true of older patients [16].

Effort and Malingering

Finally, it is known that a considerable number of individuals being assessed in a forensic context will either fabricate or exaggerate symptoms. One must assess the individual's level of effort, and rule-out the possibility of malingering during these evaluations. Malingering is defined as "the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives" ([7]; p. 739). Incentives may include the avoidance of criminal prosecution, financial gain from a tort case, or receiving disability benefits. Measuring the effort one gives during the neuropsychological examination is considered by clinicians to be core of the assessment [17]. There are a number of valid and reliable measures to assess level of effort and/or malingering of neurocognitive impairment, with good sensitivity and specificity. It is well accepted that assessment for malingering is a necessary component of the TBI evaluation. The National Academy of Neuropsychology issued a position paper on this topic in which they wrote: "... the assessment of symptom validity is an essential part of a neuropsychological evaluation. The clinician should be prepared to justify a decision not to assess symptom validity as part of a neuropsychological evaluation (p. 421)" [18].

Interpretation of the TBI Forensic Evaluation

The neuropsychological evaluation results are assessed together with results from a thorough physical examination, usually performed by a neurologist, including an analysis of the alleged TBI patient's gait and posture, cranial nerves, motor skills, tendon reflexes, coordination, sensory functioning, and autonomic system. Neuroimaging studies, such as CT or magnetic resonance imaging (MRI) scans may also be included. The examiner must make rational connections between test results and purported deficit states that exist and are attributable to the TBI. Moreover, the ecological validity of the findings must be discussed in full. Ecological validity is defined as "the functional and predictive relationship between the patient's performance on a set of neuropsychological tests and the patient's behavior in a variety of real-world settings (e.g., at home, work, school, community)" [19]. At this stage of

the evaluation, there is another important issue that must be considered. The fact that a TBI patient did poorly on a small number of neuropsychological tests (assuming adequate effort was given) might not necessarily mean that the TBI patient had significant deficits in the area tapped by the test in which he or she did poorly. If you give a large number of neuropsychological tests to someone, whether he or she has brain impairment or not, there will be a few tests in which poor performance is expected to occur by chance alone. This issue must be addressed in the interpretation section of the assessment report [19, 20].

Future Trends in TBI Evaluation

As emergency procedures become increasingly more sophisticated at stabilizing patients, and new generation medications are discovered that quickly decrease swelling in the brain, more persons who have sustained TBI will survive, where just a few years ago many such injuries would likely have resulted in death. New assessment procedures with sophisticated computerized methods of interpretation and more refined images will emerge. Measures of hemodynamic response of neural activity (functional MRI) observed while a person performs a specific task will allow stronger inferences to be drawn between TBI and functional deficits. More subtle abnormalities often associated with closed head injury can now be studied with diffusion tensor imaging [21]. The usual array of TBIs including closed head injuries and penetrating head injuries will challenge examiners. Blast-related injuries from improvised explosive devices (IEDs) used in warfare and terrorism are resulting in a sizeable number of veterans and other victims who suffer concussion and contusion from sudden increases and decreases in air pressure near the blast. These symptoms can be combined with posttraumatic stress disorder making it very difficult to tease apart what is causing the blast victim's reported deficits [22]. For example, a recent study of veterans returning from combat duty in the Iraq war found that those who suffered mild TBI (approximately 5% of the large sample surveyed) caused by loss of consciousness as a result of exposure to an explosive device, also had symptoms of posttraumatic stress disorder and depression. When these psychological injuries were controlled in the statistical analysis, symptoms usually associated with TBI such as concentration difficulties and memory impairment were no longer significantly associated with the TBI. Only the symptom of headaches remained as a significant correlate with TBI [23]. Simply stated, there is evidence to support that the psychological injury may be far more debilitating than the actual neuropsychological impairment. A more refined approach studying this phenomenon is certainly needed [24].

References

[1] Loring, D.W. (1999). *INS Dictionary of Neuropsychology*, Oxford University Press, New York.

[2] Langlois, J.A., Rutland-Brown, W. & Thomas, K.E. (2004). *Traumatic Brain Injury In The United States: Emergency Department Visits, Hospitalizations, And Deaths*, Center for Disease Control and Prevention, National Center for Injury Prevention and Control, Atlanta.

[3] Thurman, D., Alverson, C., Dunn, K., Guerrero, J. & Sniezek, J. (1999). Traumatic brain injury in the United States: a public health perspective, *Journal of Head Trauma and Rehabilitation* **14**(6), 602–615.

[4] Finkelstein, E., Corso, P. & Miller, T. (2006). *The Incidence and Economic Burden of Injuries in the United States*, Oxford University Press, New York.

[5] Teasdale, G. & Jennett, B. (1974). Assessment of coma and impaired consciousness: a practical scale, *Lancet* **2**, 81–84.

[6] Fleming, J., Tooth, L., Hassell, M. & Chan, W. (1999). Prediction of community integration and vocational outcome 2–5 years after traumatic brain injury rehabilitation in Australia, *Brain Injury* **13**, 417–431.

[7] American Psychiatric Association (2000). *The Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition, Text Revision; DSM-IV-TR, American Psychiatric Association.

[8] Anderson, S.D. (1996). Postconcussional disorder and loss of consciousness, *Bulletin of the American Academy of Psychiatry and the Law* **24**(4), 493–504.

[9] Hyman, H.A. (2001). Neurolitigation of the MTBI case without loss of consciousness: using somatic complaints to make your case, *NeuroRehabilitation* **16**(2), 103–108.

[10] Heilbrun, K., Marczyk, G.R. & DeMatteo, D. (2002). *Forensic Mental Health Assessment: A Casebook*, Oxford University Press, New York.

[11] Lezak, M.D., Howieson, D.B. & Loring, D.W. (2004). *Neuropsychological Assessment*, 4th Edition, Oxford University Press, New York.

[12] Mitrushina, M., Boone, K.B., Razani, J. & D'Elia, L.F. (2005). *Handbook of Normative Data for Neuropsychological Assessment*, 2nd Edition, Oxford University Press, New York.

[13] Brooks, N., Campsie, L., Symington, C., Beattie, A. & McKinlay, W.W. (1986). The five-year outcome of severe blunt head injury: a relative's view, *Journal of Neurology, Neurosurgery, and Psychiatry* **49**, 764–770.
[14] Hall, K., Karzmark, P., Stevens, M., Englander, J., O'Hare, P. & Wright, J., (1994). Family stressors in traumatic brain injury: a two-year follow-up, *Archives of Physical Medicine and Rehabilitation* **75**, 876–884.
[15] Buckelew, S.P. & Hannay, H.J. (1986). Relationships among anxiety, defensiveness, sex, task difficulty, and performance on various neuropsychological tasks, *Perceptual and Motor Skills* **63**, 711–718.
[16] Palmer, B.W., Boone, K.B., Lesser, I.M. & Wohl, M. (1996). Neuropsychological deficits among older depressed patients with predominantly psychological or vegetative symptoms, *Journal of Affective Disorders* **41**, 17–24.
[17] Green, P., Rohling, M.L., Lees-Haley, P.R., Green, P., Rohling, M.L., Lees-Haley, R. & Allen, L.M. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants, *Brain Injury* **15**(12), 1045–1060.
[18] Bush, S.B., Ruff, R.M., Troster, A.I., Barth, J.T., Koffler, S.P., Pliskin, N.H., Reynolds, C.R., & Silver, C.H. (2005). Symptom validity assessment: practice issues and medical necessity (NAN Policy & Planning Committee), *Archives of Clinical Neuropsychology* **20**, 419–426.
[19] Axelrod, B.N. & Wall, J.R. (2007). Expectancy of impaired neuropsychological test scores in a non-clinical sample, *International Journal of Neuroscience* **117**, 1591–1602.
[20] Miller, L.S. & Rohling, M.L. (2001). A statistical interpretive method for neuropsychological test data, *Neuropsychology Review* **11**(3), 143–169.
[21] Kraus, M.F., Susmaras, T., Caughlin, B.P., Walker, C.J., Sweeney, J.A. & Little, D.M. (2007). White matter integrity and cognition in chronic traumatic brain injury: a diffusion tensor imaging study, *Brain* **130**, 2508–2519.
[22] Taber, K.H., Warden, D.L. & Hurley, R.A. (2006). Blast-related traumatic brain injury: what is known? *Journal of Neuropsychiatry and Clinical Neurosciences* **18**, 151–145.
[23] Hoge, C.W., McGurk, D., Thomas, J.L., Cox, A.L., Engel, C.C. & Castro, C.A. (2008). Mild traumatic brain injury in U.S. soldiers returning from Iraq, *New England Journal of Medicine* **358**(5), 453–463.
[24] Koch, W.J., Douglas, K.S., Nicholls, T.L. & O'Neill, M.L. (2006). *Psychological Injuries: Forensic Assessment, Treatment, and Law*, Oxford University Press, New York.

SCOTT A. BRESLER AND DANIEL A. MARTELL

Hearsay Evidence

Hearsay Evidence Banned . . . Sometimes

There are a number of requirements that must be fulfilled before the hearsay exclusionary rule will bar the use of oral testimony or documentary evidence. First of all, the statement or writing must have been made at a time prior to its attempted use in a legal proceeding. The maker of the statement or the author of the writing, often referred to as the *declarant*, is usually not present at the proceeding when it is sought to be introduced as evidence. The content of the statement or writing is sought to be introduced as evidence by a person who, earlier, overheard the declarant uttering the words or executing the writing. Finally, the out-of-court communication is offered as proof of the truth of what it contains.[a]

If a statement or writing satisfies these requirements, the basic principle of the rule against hearsay evidence in a jurisdiction that follows the common law prohibits its use as evidence.[b]

The rule against hearsay evidence has no equivalent in civil law countries, where all evidence is admitted and considered by the tribunal.

Reasons for the Rule's Existence

The fact finder in a trial typically must be satisfied that any witness or evidence presented is reliable. In a jury trial, the jury is the finder of the disputed facts. The very basic proposition is that all persons who are expected to present evidence must be competent witnesses. When a litigant objects to the competency of a witness offered by the opposing side, or if the court has reservations about that competence, there may be a need to determine whether the person had the ability to perceive, remember, and communicate the events about which he offers to testify. All witnesses must possess a minimum of competence, perception, memory, and communication.[c]

Reliability of witness testimony may be particularly difficult to ascertain if the declarant of the out-of-court statement is not in court, and its content is sought to be conveyed by the testimony of someone who merely overheard the statement when it was being made or seeks to use a document prepared

out of court. There is a potential risk of untrustworthiness if the declarant is absent and cannot be cross-examined.

Where a litigant objects to evidence on hearsay grounds, he is arguing in essence that the statement ought not to be received in evidence, because to admit it would deprive the litigant of the opportunity to explore the circumstances under which the statement was allegedly made, and indeed to challenge whether it had been made at all. If the declarant is not present, and cannot be cross-examined, evidence of unknown validity may well mislead or distract the jurors.

The right to confront and cross-examine witnesses who give evidence against a litigant is one of the bedrocks of the adversary criminal justice system.

Rule Does Not Apply to All Out-of-Court Statements

It must be noted that the rule against the evidentiary use of hearsay does not apply to all out-of-court statements or documents, but only to those that are sought to be used as proof of the truth of their content. Thus, if the reason for introducing an out-of-court statement is not to prove its truth, but to prove that the statement – whether true or false – was made at all, the rule against hearsay does not apply. If the statement or document is probative of the issue on whether it was uttered, then it is not considered hearsay and will be admissible.

In addition to exempting all out-of-court statements from the rule against hearsay that are offered to prove some fact independent of the truthfulness of its content, the rule of exclusion also does not apply to a series of statements where, despite its purpose seeking to use it as proof of its truth, the system of justice nevertheless believes that, for policy considerations, exclusion ought not to occur.

Thus, FRE 801, which defines hearsay, specifically exempts a series of out-of-court statements that would, in the absence of the exemption from the hearsay rule, be considered "hearsay". Therefore, FRE 801 simply, and somewhat illogically, calls such statements "nonhearsay". FRE 801(d)(1) states that, in some cases, prior out-of-court statements by witnesses who are currently testifying and subject to cross-examination will not be considered hearsay statements.[d]

The same FRE 801, in its subsection (d)(2), also exempts from the prohibition against admitting hearsay evidence any admission of a party opponent or his employees and agents.[e]

It can be observed that this rule eliminates from the hearsay exclusionary concept's applicability a great number of out-of-court statements.

There are Many Exceptions to the Rule against Hearsay

Even statements that clearly fit the definition of "hearsay" may be admissible, because, in addition to assertions that are not considered hearsay at all, common law courts have, over the years, created a long list of exceptions to the rule against hearsay. These court-made hearsay exceptions are, today, also codified in codes of evidence such as the Federal Rules of Evidence (FRE) and other similar compilations of evidentiary principles.

Thus, a true hearsay statement, which is sought to be used as evidence of the truth of its content, may be admitted in evidence if it fits one of the many recognized hearsay exceptions.

Some legal commentators suggest that the prohibition against the use of hearsay evidence might just as well not exist anymore, because the exceptions have swallowed the rule. Despite this assertion, the rule continues to exist, perhaps with greater force in US courts than in other common law countries.

In the United States, the exceptions to the prohibition against hearsay are codified in FRE 803 and 804.

In FRE 803, an impressive list of 23 exemptions allows the admission of certain types of hearsay evidence. These exceptions apply whether or not the declarant of the out-of-court statements is available for testimony. In FRE 804, an additional five exceptions to the rule are noted, but, in these cases, the hearsay statement will be admissible only if the declarant of the statement is unavailable for testimony or cannot be found.

Exceptions to the rule against use of hearsay are fairly technical. There are many court decisions that describe, in great detail, the circumstances under which they will be applicable. Litigators are familiar with these court rulings. In many cases, these technicalities are unimportant to forensic experts, but at least six of these categories of hearsay exceptions do involve forensic experts with some regularity. They are discussed hereafter.

Important to forensic practitioners are these statements excepted from the rule against hearsay:

1. statements made for purposes of medical diagnoses or treatment to doctors or medical personnel[f];
2. records of regularly conducted activity (which would include laboratory or QA (Quality Assurance) entries and an expert's bench notes)[g];
3. the exception for public records and reports of public offices or agencies[h];
4. records of vital statistics[i];
5. statements contained in ancient document in existence 20 years or more[j]; and, perhaps most importantly for forensic experts
6. statements contained in learned treatises and professional publications.[k]

End Notes

[a.] In the United States, the FRE regulate the conduct of civil and criminal proceedings and the evidence which it is appropriate for a court to consider. These rules apply in all federal courts. The FRE also have their equivalent in the evidence codes in most of the US states, many of which have incorporated the FRE in whole or in part in their own state laws.

FRE 801(a) defines a hearsay statement as "(1) an oral or written assertion or (2) nonverbal conduct of a person, if it is intended by the person as an assertion". In FRE 801 (c) the definition further limits the concept's applicability to only those statements "offered in evidence to prove the truth of the matter asserted".

[b.] FRE 802 provides: "Hearsay is not admissible except as provided by these rules or by other rules prescribed by the [United States] Supreme Court pursuant to statutory authority or by Act of Congress". All states have an equivalent rule of evidence.

[c.] An additional historical requirement for testimonial capacity requires a commitment to be truthful. This commitment is sought to be safeguarded by requiring a witness take the "oath" to speak truthfully or, if a witness's religion prohibits oath-taking, by his affirmation of the commitment to tell the truth.

[d.] FRE 801(d)(1) provides that an out-of-court statement is not hearsay if the maker of the statement, referred to in the Rule also as *the declarant*, "testifies at the trial or hearing and is subject to cross-examination concerning the statement, and the statement is (A) inconsistent with the declarant's testimony, and was given under oath subject to the penalty of perjury at a trial, hearing, or other proceeding, or in a deposition, or (B) consistent with the declarant's testimony and is offered to rebut an express or implied charge against the declarant of recent fabrication or improper influence or motive, or (C) one of identification of a person after perceiving the person."

[e.] FRE 801(d)(2) states that the hearsay exclusionary rule also does not apply where "The statement is offered against a party and is (A) the party's own statement in either an individual or representative capacity or (B) a statement of which the party has manifested an adoption or belief in its truth, or (C) a statement authorized by the party to make a statement concerning the subject, or (D) a statement by the party's agent or servant within the scope of the agency or employment, made during the existence of the relationship, or (E) a statement by a coconspirator of a party during the course and in furtherance of the conspiracy. . . . "

[f.] FRE 803(4).

[g.] FRE 803(6). This used to be called the *Business Records* exception, but it is, today, much broader than it once was and applies also to not-for-profit institutions, professions, occupations, or callings of any kind.

[h.] FRE 803(8).

[i.] FRE 803(9).

[j.] FRE 803(16).

[k.] FRE 803(18). The "learned treatise" exception is much broader than its designation intimates. It permits the use as evidence of "published treatises, periodicals, pamphlets, on a subject of history, medicine, or other science or art, established as a reliable authority by the testimony or admission of the witness or by other expert testimony or by judicial notice". If a statement contained in such a source is admitted, the text may be read into evidence, but the publication or document itself may not be received as an exhibit.

Related Articles

Chemical, Biological, Radiological, and Nuclear Investigations

Chain of Possession of Tangible Evidence

Cross-Examination of Experts

Demonstrative Evidence

Direct Examination of Experts

Expert Opinion in Court: a Comparison of Approaches

Expert Opinion in Court: Civil Law Jurisdictions (France, Germany, Italy, and Spain)

Expert Opinion: United Kingdom, Canada, and Australia

Foundation Testimony

Hypothetical Question

***In Limine* Motions and Hearings**

Judicial Notice of Scientific Principles and Facts

Jury Instructions on Expert Testimony

Learned Treatises as Evidence

Therapeutic Jurisprudence

ANDRE MOENSSENS

Height Determination from Skeletal Remains *see*

Anthropology: Ancestry and Stature Determination

Histology

Short Historical Introduction into Microscopy and Histopathology

Although some medicolegal practices, e.g., the investigation into causes of sudden unexpected and violent deaths by barber-surgeons, were established early in medieval Europe, the idea of forensic histology became theoretically possible first with the improvement of light microscope and histological methods, and secondly after the cell theory by Schleiden, Schwann, and Purkinje in the first half of the 19th century and the concepts of pathogenesis of diseases by Carl von Rokitansky (1804–1878) and lastly cellular pathology by Rudolf Virchow (1821–1902) in 1858 [1, 2].

It took over 200 years from the invention of the first microscope in the early 1600s in the Netherlands until its use started to strike root as a scientific instrument useful for the study of pathology. Many obstacles hampered the beginning of microscopic pathology and in addition to poor optics, several other reasons have been mentioned as an explanation; e.g., secrecy of the art, high cost of the apparatus, technical difficulty, lack of ideas, and late acceptance by Universities [3, 4].

Interestingly, Marie-François Xavier Bichat (1771–1802), the first to introduce the concept of tissues as distinct entities and often called the *father of modern histopathology*, did not use microscopy. Likewise, René Théophile Hyacinthe Laennec (1781–1826), another famous French physician of the era, was against the use of microscope [4].

Through accumulation of knowledge of pathological processes and their histological features, histological examination of major organs became gradually an integral part of every academic autopsy, although daily routine practices may still vary considerably according to resources and cultural setting. It is noteworthy that already the Prussian decree on the procedure of a medicolegal autopsy (1875) recommended (§ 5) that the forensic doctor should have a microscope with two objectives and the minimum of 400-fold magnification. Also, Virchow pointed out in his book about dissection technique, with particular consideration for medicolegal practice, that certain pathological changes cannot be recognized by the naked eye, but only with the help of a microscope or a magnifying glass [5].

Indication for Forensic Histological Investigation

The procedure of the cause of death investigation in a medicolegal setting depends largely on jurisdiction, as in some countries, it is limited to clearly suspicious, crime-related deaths, whereas in others, the forensic pathologist deals, in addition to the previous ones, also with the whole spectrum of other types of deaths including sudden unexpected deaths due to presumed natural causes, accidents, suicides, occupational diseases, and iatrogenic causes i.e., deaths due to medical diagnostic or therapeutic procedures.

Medicolegal autopsy often differs from hospital autopsy by the lack of any preliminary knowledge of

the deceased, neither of possible illnesses nor circumstances of death and who may be unidentified and just found dead. Therefore the forensic pathologist ought to be prepared for every eventuality as presumed natural death may turn out to be a homicide.

In the vast majority of cases forensic histology deals with tissue samples retained for cause of death investigation purposes during autopsy. Additional questions that need to be dealt with include vitality, timing, and causes of injuries and their differentiation from postmortem changes.

In the broadest sense, if one includes methods other than just classical histology, forensic aspects of tissue investigation may concern other situations. In mass disasters, bodies may be fragmented and the human origin of isolated tissue fragments as well as the personal identity of the victims has to be established usually applying molecular biological methods. A question about the origin or identity of samples may also arise as a result of mix-up during tissue processing or other similar situations where the chain of evidence may have been broken. Occasionally the question of paternity first arises postmortem when the alleged father dies before the paternity issue has been investigated. In cases where there has been an autopsy and tissues have been retained for histology, the formalin fixed and paraffin-embedded tissue blocks can be used as material for DNA extraction. The extracted DNA can then be compared with the DNA of the child and the mother. Once we investigated paraffin-embedded tissue from a curettage, in which the patient had been diagnosed with endometrial cancer but the histological examination of the uterus that had been operatively removed did not show any signs of malignancy. The question arose whether the original cancer diagnosis had been incorrect and whether the patient should be given postoperative chemotherapy in case the diagnosis had been false, for example, owing to a sample mix-up. Extraction of DNA from the paraffin-embedded tissue from the scraping and comparison with the DNA sample from the patient was able to confirm that the original diagnosis had been correct.

Guidelines

Many autopsy guidelines express a view as to autopsy histology. As a rule, complete histological examination of all major organs is considered an essential part of every postmortem and this principle has been adopted by many guidelines:

A recommendation of the Council of Europe: Rec (99) 3 of the Committee of Ministers to Member States on the harmonization of medicolegal autopsy rules; Adopted by the Committee of Ministers on 2 February 1999, states in section 6. Sampling:

> . . ., the following minimum rules should be applied:
> a. in all autopsies, the basic sampling scheme includes specimens from the main organs for histology. [6].

The Royal College of Pathologists (United Kingdom) Guidelines on autopsy practice, September 2002, takes a similar position in section 9 Autopsy histology:

> 9.2 As best practice, sampling of all major organs for histology in all autopsies is recommended.

The importance of histology to the cause of death investigation is further emphasized.

> 9.6.2 . . .. If in advance of the autopsy, discussions with a Coroner indicates to the pathologist that retention of material that may be relevant to the investigation will not be permitted, the pathologist should decline the request to perform the autopsy [7].

The Code of practice and performance standards for forensic pathologists by the Home Office Policy Advisory Board for Forensic Pathology and The Royal College of Pathologists, November 2004 and Code of Practice and Performance Standards for Forensic Pathologists dealing with Suspicious Deaths in Scotland by Scottish Government, Crown Office Procurator Fiscal Service (COPFS) and The Royal College of Pathologists, November 2007, have the same standpoint to postmortem histology:

> A histological examination should be made, by the pathologists themselves, of the major organs (assuming that they are not heavily decomposed) in all suspicious deaths. Histology is of value in confirming, evaluating and sometimes revising the course of natural disease processes that may have contributed to the cause of the death. Other samples should be taken for histological examination depending on the circumstances of the case, e.g., for the purposes of aging injuries. The reasons behind any decision not to undertake a histological examination must be adequately recorded, in order that the pathologist may be in a position to defend this decision if required [8].

Methods

General Remarks. The choice and use of correct methods is paramount for reliable results. This starts with sampling of tissues, in which several aspects have to be considered. Samples should be taken from suspected pathological findings, or when no such findings are visible to the naked eye, representative samples should be taken from the main organs bearing in mind that special sampling procedures and protocols may apply, e.g., for neuropathology, the cardiac conduction system or pediatric cases.

Samples have to be cut using sharp instruments to minimize mechanical damage to the tissues that may cause artifactual changes and hamper the interpretation of the findings. Sites, where the samples have been taken from, have to be recorded and numbered in the autopsy protocol to enable topographical correlation and comparison with macroscopic findings and possible later review and quality assurance measures.

For routine histopathology most tissues are fixed using buffered formaldehyde to stop the autolytic processes and degradation by bacteria and to stabilize the proteins. The size and the thickness of the sample must be in compliance with the choice of the fixation method and appropriate fixation time to avoid fixation artifacts and to allow the use of special methods such as immunohistochemistry, as too long fixation may destroy the antigenic properties of the tissue and produce false negative results. Special procedures are often necessary for enzyme- or immunohistochemical methods or to demonstrate substances that are easily inactivated, destroyed, or removed during routine fixation or tissue processing procedures. In these cases frozen fresh tissue can be used that is cut using a cryostat. At every stage of tissue handling, the specimens have to be properly labeled to preserve the chain of evidence. After fixation the tissue samples are processed either manually or using automated tissue processors to impregnate them with solid medium, usually with paraffin wax, to enable sectioning with microtome. In transmission electron microscopy samples are usually embedded in epoxy resin and cut with ultramicrotome into semithin or ultrathin sections that are stained using heavy metals such as lead, uranium, or tungsten.

The choice of embedding material, optimal thickness of the sections when cutting, and the staining methods are all important factors contributing to the final result. One has to be aware that each of these individual steps is a potential source of artifacts that may hamper the correct interpretation of the findings [9].

Staining. Paraffin sections are usually stained to visualize the various tissue components and structures. The affinity to a given stain depends on complex physico-chemical interactions between the solvent, dye, and tissue. Many of the classical staining methods, such as hematoxylin-eosin, Mallory Phosphotungstic acid-hematoxylin (PTAH), or Van Gieson stain, among others, were introduced in the nineteenth century and quite a few of them are still in use today. Depending on their chemical composition and the characteristics of the staining solutions, they show characteristic affinity to different tissue components that help to recognize morphological features of normal tissue and to differentiate them from pathological changes. In addition to the capability to visualize tissue structures some staining methods show a shift in color when tissue becomes injured, e.g., the well-known phenomenon of increased cytoplasmic eosinophilia resulting from increased binding of eosin by cytoplasmic proteins with hematoxylin-eosin stain in myocardial injury or ischemic nerve cell injury. The conventional histological staining methods are mainly used to visualize morphological changes in the cells and tissues [9].

Special Techniques. In histochemistry, staining methods are used to demonstrate specific components in cells or tissues such as iron. In Perls Prussian or Berlin blue reaction the section, is treated with dilute hydrochloric acid to release ferric ions from binding proteins. These ions react with potassium ferrocyanide to produce an insoluble blue compound. Fat or lipids have to be stained using frozen sections as fixation and tissue processing usually removes them (Oil-red-O or Sudan stain).

Microorganisms can be demonstrated using special stains, e.g., bacteria using Gram-stain and Mycobacterium tuberculosis using Ziehl-Neelsen-stain. Periodic acid-Schiff or PAS-stain can be used to stain fungi. It also stains glycogen, mucin, mucoprotein, and glycoprotein. Viral antigens can be detected using immunohistochemistry by specific, typically monoclonal, antibodies against viral protein generated in laboratory animals. The sensitivity and specificity varies according to the type of the virus. In situ

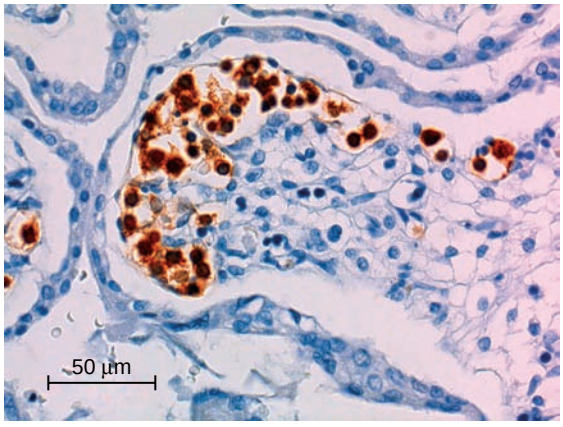


Figure 1 Localization of parvovirus B19 DNA in chorionic villi of placenta determined by *in situ* hybridization assay (dark brown cells). (Courtesy of Dr. Tytti Vuorinen, Department of Virology, University of Turku)

hybridization is a method that uses a labeled viral DNA or RNA probe to localize a specific DNA or RNA sequence of the virus within a tissue section (Figure 1).

Special histological techniques are applied in timing of injuries particularly in context with skin wounds and cerebral trauma. A prerequisite for using these is that both the pathologist and the laboratory have solid expertise with methodology preferably based on experimental research and proven in practical casework.

Enzyme histochemical methods are rarely applied to autopsy material for diagnostic purposes. When applied, fresh frozen cryostat sections have to be used as routine fixation and subsequent tissue processing usually destroys or diminishes the enzyme histochemical reactivity of most enzymes. It would be misleading to speak about enzyme activity in this context as, although the enzyme histochemical methods have been named after the individual enzymes, the enzyme histochemical reactivity need not necessarily correlate at all with the actual biochemical activity of the enzyme in question, suggesting that other factors, such as the presence of auxiliary enzyme systems or other cofactors of the histochemical reaction, may be rate-limiting [10].

This generates an analogous problem as with conventional histological methods in which a shift in color, e.g., when using Acid Fuchsin (AF) or Haematoxylin-Basic Fuchsin-Picric acid (HBFP) stain was claimed to demonstrate early myocardial ischemia but was proven to be unreliable; unless the scientific basis and limitations of any histological method is thoroughly understood, it is hazardous to use it for medicolegal purposes [11].

After the great advances of immunohistochemistry in basic research, clinical pathology and the availability of a huge number of new antibodies, the use of immunohistochemical methods was soon applied to medicolegal problems both in animal experiments and autopsy material to elicit new insight into questions of vitality, aging of wounds, and head injury and to find new markers of early myocardial injury. So far, the results have not always been unambiguous and their practicability and reliability has been difficult to assess because of diverging results and sometimes also because of shortcomings in experimental design and control material. It is obvious that in experienced hands with well-established methods, immunohistochemistry can be of advantage over conventional histochemistry [12].

Histological Findings as Forensic Evidence. In the administration of justice, a high degree of reliability is expected also from forensic histological evidence. Janssen [13, 14] contemplates the value and significance of histological data from the following aspects: “first, according to the evidential value of the tissue data itself; second, according to its position within the framework of evaluation, which results from the facts of the case that have come to light in combination with other findings”. He further grades the evidential value of forensic histology results in three groups (slightly modified from [14]):

Group 1: Comprises histological findings in which identification of their cause, character, and chronological pathogenesis is definite permitting conclusion about the etiology and causality independent of any further data. As examples serve, certain stages of tuberculosis and foreign body granulomas (Figure 2).

Group 2: Comprises findings that have forensically verifiable significance in association with further premises. The presence of additional facts is, therefore, prerequisite for the conclusiveness of such evidence.

Janssen describes, as an example, suspected homicide by means of an assault to the neck, where isolated fresh hemorrhages in the deep soft parts of

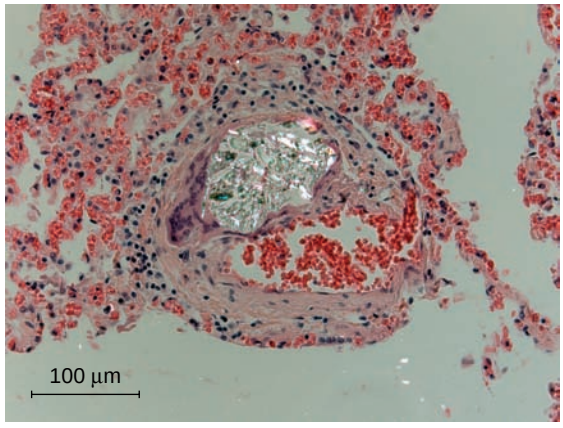


Figure 2 Microscopic view of a foreign body giant cell in polarized light showing birefringent material (probably talcum powder) from the use of adulterated drugs within pulmonary vessel wall of an intravenous drug addict

the victim's neck can be considered as evidence of strangulation only in combination with corresponding injuries of the overlying skin and larynx and further indications of asphyxiation and vascular congestion. As such, the hemorrhages alone are not unequivocal and not of conclusive significance.

Group 3: Comprises findings, which neither singly nor in combination with other facts, represent evidence that excludes all other possibilities of casual explanation. Examples are given as acute vascular congestion, pulmonary emphysema, petechial serous hemorrhages, cerebral edema, and the majority of all parenchymal necroses. Janssen considers these findings either inappropriate or only of limited use as forensic histological proof.

Problem Areas. Postmortem diagnosis of sudden cardiac death (SCD) can be taken as an example that illustrates the problems related to the way of thinking that is common in this context in scientific literature. Ischemic heart disease is the most common cause of sudden unexpected deaths in Western countries and in those jurisdictions where these belong to the domain of forensic pathology; they form, by far, the largest proportion of medicolegal autopsies. In the pathology of sudden cardiac death, sclerosis and stenosis of coronary arteries is not sufficient to diagnose SCD and obvious infarcts are the exception. As early morphological changes seen in light microscopy are unspecific, many pathologists have been actively searching for more sensitive methods to demonstrate early myocardial injury. This is logical as such, but leads to another problem, because, as a rule, the more sensitive the method, the more easily it detects agonal ischemia and along the autolytic changes these may be very difficult to differentiate from changes during life. Research on early myocardial injury may give us a better insight into the pathophysiology of cardiac injury but purely from the cause of death viewpoint it is nonsensical, as strictly speaking, even a fresh myocardial infarction visible to the naked eye does not prove that the person died of infarction unless other causes of death have been excluded. This was well demonstrated by the following case. A middle-aged man was found dead in his room in a mental hospital and the autopsy showed a large fresh myocardial infarction visible to the naked eye and comprising approximately 70% of the left ventricle muscle mass. If the pathologist had not been aware of the circumstances, the diagnosis might have been different, as the man did not die from natural causes he had hanged himself [12].

A recent survey of myocarditis deaths showed a high error rate as histopathologic reanalysis of samples taken during autopsy showed that only 32% of the 142 subjects met the Dallas criteria for myocarditis. The most evident cause of death in the Dallas-negative subjects was ischemic heart disease ($n = 78$, 55% of all cases), suggesting that myocarditis is overdiagnosed in routine autopsies, particularly in patients who have died suddenly or are found dead [15] (Figure 3).

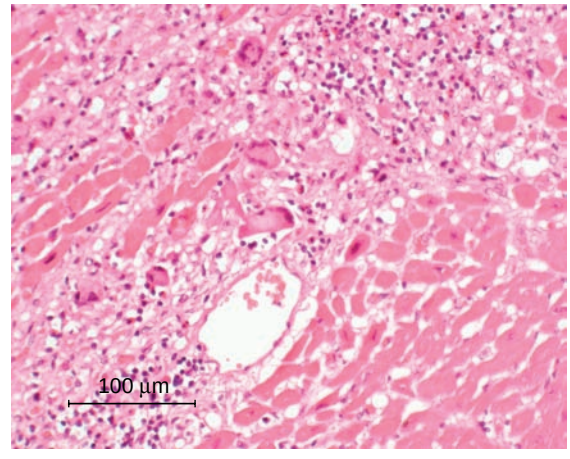


Figure 3 Giant cell myocarditis (Haematoxylin-Eosin)

Other problematic areas where special expertise and methodology is often necessary are, e.g., head injury, sudden death in epilepsy, and pediatric forensic pathology, in which forensic neuropathologist or forensic pediatric pathologist should be consulted and involved at an early stage of investigation. Unfortunately, such an expertise is rarely available worldwide.

Medicolegal autopsy with its ancillary investigations resembles the scene of crime investigation in so far, that if, e.g., histology samples have not been taken, the neglect can seldom be corrected afterwards. Considering the high requirement for reliability of forensic evidence and the widely accepted principles in many autopsy guidelines that histological examination is an integral part of every autopsy, it seems inconceivable that this is not consistently practiced. One reason for this may be the common miscomprehension about the degree of certainty as to the cause of death that can be achieved even after an autopsy. With the exception of disease processes causing major bleedings by rupture of an artery, such as the aorta, or even minor vessels in vital centers like the brain, or rupture of the heart wall because of infarction, or obstructing vessels which stopping a vital part of the blood circulation, as in pulmonary embolism, the cause of death can rarely be established with absolute certainty. Even in traumatic deaths, where the gross injuries are most obviously incompatible with life, they need not have anything to do with the cause of death as they may have been caused postmortem owing to a secondary accident triggered by a disease, for example, sudden cardiac death while driving a vehicle.

Even in cases where everything appears to be "clear" at first sight, new questions may arise later, which may remain unanswered if histology samples have not been taken. It is further not wise and unacceptable, though it happens in many jurisdictions, that a layperson should have the say whether or not histology should be taken. It is ignorance or self-deception to believe that, for instance, the cost of histology is the issue, as a single miscarriage of justice can cost the authorities more both in terms of money as well as in loss of credibility. An autopsy is not complete without histology and both sampling and microscopy should preferably be in the same hands and carried out by the same pathologist who performed the autopsy.

References

[1] Virchow, R. (1871). *Die Zellularpathologie*, Verlag von August Hirschwald, Berlin.

[2] Saukko, P. & Pollak, S. (2005). AUTOPSY – procedures and standards, in *Encyclopedia of Forensic and Legal Medicine*, Elsevier, Oxford, p. 166.

[3] Majno, G. & Joris, I. (1973). The microscope in the history of pathology. With a note on the pathology of fat cells, *Virchows Archiv. A: Pathology. Pathologische Anatomie* **360**(4), 273–286.

[4] Dhom, G. (2001). *Geschichte der Histopathologie*, Springer-Verlag, Berlin, Heidelberg.

[5] Virchow, R. (1893). *Die Sections-Technik im Leichenhause des Charité -Krankenhauses, mit besonderer Rücksicht auf gerichtsärztliche Praxis*, Verlag von August Hirschwald, Berlin.

[6] Council of Europe (2000). Recommendation no. r (99) 3 of the committee of ministers to member states on the harmonization of medico-legal autopsy rules, *Forensic Science International* **111**(1–3), 5–58.

[7] The Royal College of Pathologists (2002). *Guidelines on Autopsy Practice*, The Royal College of Pathologists, London, p. 61.

[8] Scottish Government, Crown Office Procurator Fiscal Service, and the Royal College of Pathologists, (2007). *Code of Practice and Performance Standards for Forensic Pathologists Dealing with Suspicious Deaths in Scotland*, Scottish Government, Edinburgh, p. 27.

[9] Bancroft, J.D. & Stevens, A. (1990). *Theory and Practice of Histological Techniques*, Churchill Livingstone, Avon.

[10] Hiltunen, J.K., Saukko, P. & Hirvonen, J. (1985). Correlations between enzyme histochemical reactions and respective enzyme activities in global ischaemic rat hearts, *British Journal of Experimental Pathology* **66**(6), 743–752.

[11] Saukko, P. (1983). *Evaluation of Diagnostic Methods for Early Myocardial Injury in Sudden Cardiac Deaths*, University of Oulu, Oulu.

[12] Saukko, P. & Knight, B. (2004). *Knight's Forensic Pathology*, Edward Arnold, London.

[13] Janssen, W. (1977). *Forensische Histologie*, Verlag Max Schmidt- Römhild, Lübeck.

[14] Janssen, W. (1984). *Forensic Histopathology*, Springer-Verlag, Berlin, Heidelberg.

[15] Kyto, V., Saukko, P., Lignitz E., Schwesinger, G., Henn V., Saraste, A. & Voipio-Pulkki L.M. (2005). Diagnosis and presentation of fatal myocarditis, *Human Pathology* **36**(9), 1003–1007.

Related Articles

Autopsy

Cardiac and Natural Causes of Sudden Death

Microscopy: Light Microscopes
Substance Abuse

PEKKA J. SAUKKO AND STEFAN POLLAK

History of Interpretation in Forensic Science *see* Interpreting Expert Opinions: History of

Homicide *see* Aggression; Homicide: Multiple (Behavior)

Homicide: Multiple (Behavior)

Murder is the common law offense of homicide; unlawful killing of another human being with malice aforethought. While definitions vary by jurisdiction, there are generally distinctions in legal definition such as between premeditated or without deliberation and premeditation; murder in the course of commission of a felony; and various degrees of manslaughter or homicide without malice aforethought. Murder is a serious crime and its investigation, and the dynamics of the incident and the perpetrator, vary from case to case and are the subject of an extensive literature [1, 2].

This discussion focuses upon relatively rare subsets of murder involving the death of multiple victims. Some of these multiple death situations are discussed in depth elsewhere and for these the reader is referred to **Serial Homicide**; **Suicide (Behavior)**; **Stalking**; **Battered Spouse Syndrome** for specialty discussions on specific topics such as elder abuse *see* **Elder Abuse: Policy**; **Elder Abuse: Risk**.

Serial killers may murder large numbers of victims, but the individual killings are separated by often-extended periods of time during which emotions cool and the perpetrator focuses upon other thoughts and activities. This discussion will deal specifically with instances in which the killer acts in continuous heat of passion. These are usually referred to as *mass murders* in which multiple victims are killed at the same time, and spree killings in which multiple victims are killed at two or more locations as the perpetrator moves from place to place. There are variations in each of these scenarios, and at times the distinctions can become blurred, but each is discussed separately.

Mass Murder

The US Department of Justice's Bureau of Justice Statistics [3] defines a mass murder as the murder of four or more victims at one location, within one event. Some researchers [4] have expanded this definition to include incidents where the perpetrator had a clear intention of killing a large number of persons, for example, carrying or firing a hundred or more rounds of ammunition, but nevertheless failed to kill four victims. On the other hand, some domestic killings may result in four or more deaths, but still possess the characteristics of a domestic homicide or murder/suicide (*see* **Suicide (Behavior)**). Although guns, knives, and blunt instruments are the most common weapons, mass murder can also be accomplished with bombs, collisions, fires, drugs or poisons, and through other destructive means.

Mass murders have been documented throughout history [5]. The Bible cites several examples including Samson's slaying of the Philistines (Judges 16 : 27–30) and King Herod's order for the murder of all male children age two years or less in the region of Bethlehem while trying to kill Jesus (Matt. 2 : 16). On August 10, 1810, at Ywahoo Falls, Kentucky, racist whites murdered over 100 Cherokee women and children. In 1993, close to 1 million Rwandans were slaughtered over a period of 100 days by an extremist faction of the Rwandan armed forces. Terrorist attacks on the World Trade Center in New York City and the Pentagon in Washington, DC, killed thousands of civilians on September 11, 2001. While mass murderers are often civilians; genocide, unprovoked killings during such activities are taking back

prisons and other facilities commandeered unlawfully [6], and violations of the rules of war by the military are also mass murder.

The Westroads Mall shooting on December 5, 2007, is representative of civilian mass murders. Nineteen-year-old Robert A. Hawkins killed nine people (including himself) and wounded four, two of them critically, when he entered the Von Maur department store in the Westroads Mall in Omaha, Nebraska, and opened fire with a stolen AK-47 rifle and two 30-round clips of ammunition. The entire incident took just 6 min. Hawkins had a history of mental difficulties, was estranged from his parents, had just lost his job at a fast food outlet for allegedly stealing $17, and had recently broken up with his girlfriend. He left a will and a long note including "I just want to take a few pieces of shit with me . . . just think tho [sic], I'm gonna be fuckin [sic] famous".

On the morning of April 19, 1995, Timothy McVeigh drove up to the front of the Alfred P. Murrah Federal Building in Oklahoma City in a rented Ryder truck in which he and Terry Nichols had assembled a 5000-pound bomb of ammonium nitrate fertilizer and nitromethane motor-racing fuel. McVeigh ignited a 2-min fuse and ran. When the bomb detonated at 9:02 am, it lifted McVeigh 2 in. off the ground. The explosion killed 168 people, and 450 were injured. Nineteen of the victims were small children in the day care center on the ground floor of the building. More than 300 buildings were damaged and 12 000 volunteers and rescue workers supported operations following the bombing.

McVeigh was an Army veteran and loner who planned the attack for the second anniversary of the Waco, Texas, siege of a religious compound as revenge for what he considered to be a tyrannical federal government. The deadliest act of terrorism within the United States prior to the attacks of September 11, 2001, this represents another form of mass murder and one which has escalated worldwide with increases in terrorist car bombings and suicide bombers.

Spree Murder

One of the best-known spree killers is the team of Charles Raymond Starkweather and his girlfriend Caril Ann Fugate. Starkweather was aged 17 in 1957, and Caril Ann 14, when Charles committed his first murder by killing a store clerk with a shotgun. Two months later, the pair killed 10 additional victims in a Nebraska and Wyoming road trip crime spree that has inspired at least six movies and influenced popular literature and music. [7]

Many of the crimes that we think of as mass murders are technically spree crimes. Charles Whitman's 1966 "Texas Tower" shooting rampage from the observation deck of the University of Texas, where he was a student, killed 14 people and wounded 31 others. However, this actually occurred after he had murdered his wife and mother at two separate locations, later, the day before. Similarly, the Virginia Tech massacre on April 16, 2007, which resulted in 32 deaths, the deadliest shooting incident by a single gunman in U S history, consisted of two separate attacks approximately 2 hours apart.

Causes and Circumstances

There has been considerable speculation, in both academic and popular literature, about what "causes" multiple murders. The topic itself, however, is so emotionally charged that interpretations are often shaped by personal fears and preconceptions. In reality, there are almost as many possible explanations as there have been perpetrators.

Physical illness is occasionally implicated. For example, Texas Tower shooter Charles Whitman was troubled by severe headaches for which he took extensive over-the-counter medication. The day before his death he requested that an autopsy be conducted on him, and asked that a part of his estate be donated to mental health research in an effort to prevent further tragedies of the type he knew that he was about to commit. His autopsy did disclose a highly cancerous glioblastoma that probably would have taken his life within a year. This condition, however, unlike Whitman's deadly rampage, is not unusual. Other than the fact that he was clearly uncomfortable, there is no reason to believe that the glioblastoma created a mental condition that set Whitman on an inevitable path to murder. In fact, there are few, if any, cases in which mass homicide can be traced to a physical determinate.

Mental illness is also frequently advanced as a cause. Whitman, however, is typical of most mass murders. He was rational, systematic, and his actions were clearly premeditated. He was not under the influence of a mind-altering drug, and even the notes that he wrote about his killings of his wife and

mother reflect love, compassion, and a belief that he was acting to spare them pain and grief. The systematic and carefully planned nature of virtually all multiple murderers precludes an insanity defense (*see* **Insanity: Defense**). We are limited by the fact that such a large percentage of mass murderers die in the commission of their crimes and, therefore, have not provided a sizable research population; but those who survive to be taken into custody seldom even attempt a mental illness or even diminished capacity defense (*see* **Behavioral Science Evidence**). Moreover, even if grounds could be found for such a defense, juries are unlikely to view a mass murder perpetrator as a candidate for rehabilitation or even mercy.

So what does drive the multiple murderers? Several studies have addressed this question, and the literature provides a range of diverse alternative explanation [8–10]. Many cite deficiencies in moral development [11, 12], but this often conflicts with the firm, if perverse, moral beliefs seen particularly among former military perpetrators including Whitman and McVeigh. Most perpetrators do exhibit some level of psychopathy (*see* **Psychopathy**). Often, young perpetrators such as Robert Hawkins have had experiences with substance abuse (*see* **Alcohol: Behavioral and Medical Effects**; **Substance Abuse**), although this exposure in itself may not have been satisfying, and few incidents are carried out while under the influence. Perpetrators may also have histories of personality disorders such as attention deficit, mood disorders, oppositional defiant disorders, and other conditions associated with authority figure conflicts.

Psychosis has been advanced in a small minority of mass murderers, but most of these cases fall short of legal insanity. Occasionally, psychotic mothers have killed their dependent children, but these crimes only meet the definition of mass murder because there happened to have been four or more children. Other diminished responsibility diagnoses have been reported, such as posttraumatic stress disorder (*see* **Posttraumatic Stress Disorder**), depression, masochism, and various levels of paranoia; but these personality disorders rarely come close to explaining the level of premeditated violence.

Thrill killing is perhaps most associated with the 1924 Chicago case of intelligent upper-class college students Nathan Leopold and Richard Loeb [13], who were convicted of murdering 14-year-old Bobby Franks for the experience of committing the act. A more recent similar case was the killing of a university couple in New Hampshire by Vermont teenagers Robert Tulloch and James Parker in 2001 [14]. While rare, documented cases of mass murder exist for which no other explanations can be offered.

Sexual homicide can be linked to at least two different types of motives. The first relates to power. The perpetrator is not driven by sexual gratification but by the power to terrify his victim. As the victim wears into numb exhaustion and fails to express a terrified response, she is killed, and a new victim sought [15]. While actually a form of serial homicide (*see* **Serial Homicide**), serial sexual homicide can also resemble spree killings.

Unlike with serial homicide, however, direct sexual motivation is rare in mass murderers. The leading exception is the Austrian Sylvestre Matuschka who apparently derived sexual pleasure from blowing up trains with dynamite, thus killing the passenger. His lethal sexual fetish claimed 22 lives before he was apprehended in 1932.

Ideological/disciple and the related *Political/hatred* murders are among the most pervasive and virulent. Reflected in the Nazi holocaust and radical Islamic genocide, as well as the Ywahoo Falls murders and the many instances of "racial cleansing" seen in Eastern Europe, Africa, and elsewhere, related dynamics are seen in individual acts such as suicide bombings, ideologically propelled murder rampages, and violent sprees by religious and political disciples. A noted variation was seen in the 1978 Jonestown incident that ended the lives of 918 Americans involved with the "Peoples Temple Agricultural Project" cult in Guyana [16, 17]. Jonestown is significant not only in that it provides one of our most recent demonstrations of ideological power and the lengths that disciples go to please their leader but also in that it involved people killing their own family and friends.

Perverted love is frequently seen in murder/suicide, and Charles Whitman appears to be a good example. Here, family members and loved ones are seen as an extension of the perpetrator. When the perpetrator sees no hope for survival, this belief is broadened to include loved ones who are killed in the belief that death will spare them future suffering. The perpetrator's extended sense of self prevents him from viewing others as having a life separate from his own.

In other instances, particularly when murder is not followed by suicide, the motivation is often more rooted in anger and loss. Ronald Gene Simmons, the father of his daughter's son, killed 14 members of his family on Christmas in 1987 when his wife threatened to divorce him. Julio Gonzalez became jealous of his girlfriend in 1990 and torched the Bronx's Happy Land Social Club, killing nearly all 97 persons inside. In still other instances, mass murder has been committed to protect a family member or members. During the Allen-Edwards feud in Hillsville, Virginia, on March 13, 1912, feudists entered a courtroom where a family member was being tried and killed the judge, sheriff, commonwealth attorney, a juror, and an innocent bystander.

Revenge/execution killings often involve "getting even" with persons who are seen as having inflicted harm or humiliation. Sometimes this is simply part of a criminal pattern, as with the Saint Valentine's Day "massacre" shootings of seven people in a garage as part of a 1929 Chicago gang war. Occasionally, multiple murders take place during crimes simply because the opportunity presents itself.

Other instances involve disgruntled employees, classmates, and even patients who rationalized their responsibility and transfer the blame to others (*see* **Threat Assessment: Workplace**; **Threat Assessment: School**; **Violence Risk Assessment for Mental Health Professionals**). Examples include David Burke, a fired airline employee, who followed his boss onto a plane on December 7, 1987, shot him, and caused the plane to crash killing 43 people. Mark Barton, a day trader, became angry after losing a great deal of money and on July 29, 1999, killed his family and entered two brokerage firms, slaying 9 and wounding 12. Dylan Klebold and Eric Harris committed suicide after their mass murder at Columbine High School in Littleton, Colorado, on April 20, 1999.

Other factors can also be cited for mass murders. A great many theories of mass murder have been advanced and will be discussed further. A frequent classification however, is "unexplained" that is used to characterize cases that do not fit obvious patterns. Some classification systems focus upon method of operation; some upon victim characteristics, and others upon perpetrator characteristics. In this last grouping, a few patterns do emerge. Perpetrators are often found to have been bullied as children. Many are characterized as "loners", often with violent fantasies, who feel socially excluded [18]. This sense of being cut off from others can lead to reduced empathy and emotional numbing and can mark a fairly abrupt change associated, for example, with a relocation or a change of environment [19].

Perpetrators are frequently from vulnerable age groups: adolescents and older workers who see their lives as offering few options and fear the loss of something seemingly minor but very important to them. Thomas Hamilton, after losing his volunteer position as a scoutmaster in Dunblane, Scotland, entered a primary school and killed 17 children. Mass murderers often have no significant criminal record, and others perceive him or her to be reasonable or at least harmless. Typically, many years of frustration have occurred before the fatal event.

Overall, retrospective analysis (*see* **Psychological Autopsy**) often discloses feelings of hopelessness and possibly despair or depression, in people who otherwise live quiet and even solitary lives, which have not attracted attention, but who decide to go out in a blaze of glory taking others with them. When they have a particular target in mind that is seen as responsible for their fate, they focus systematically upon that target, taking out bystanders if they simply wander into the way. If there is no specific target, victims are taken randomly.

Another recurrent finding is a critical mortality event. Charles Starkweather apparently had no intention of killing the store clerk that was his first victim. The clerk, however, attempted to snatch Starkweather's shotgun. There was a tussle, and when it was over Charles shot and killed the clerk. This was a transformative event for Starkweather – he had made the transition from ordinary citizen to murderer, but otherwise the world had not changed. Timothy McVeigh described his military experience as a similar transformation, removing the constraints that prevent most people from taking human life. This pattern is also seen in civilian massacres by the military, as well as in other situations where people became ideological or mission-driven executioners. For these individuals, once they have taken a life, the second and the third and all the rest that follow seem almost inconsequential.

These transformative events seem to have different impacts upon different people. Most confront the realities of life's fragility as just another insight. A rare few progress to mass murder. Others, including some military veterans and police officers, appear to

have been so impacted by the loss of the clear line between their commitment to the sanctity of life and the sudden ease with which they can end life that, perhaps protectively, they have killed themselves.

Social Influence

Human activity takes place within a social context. Even the solitary gunman is interacting with society. Therefore, it is impossible to look at violence outside of social contexts. With mass murder, these social factors are multiple and intricate.

First, mass murder is often carried out by pairs or groups of perpetrators. Terrorists, totalitarian hit men, and criminal mobs all are generally part of a social network. Charles Starkweather and Caril Ann Fugate were a pair, and Starkweather was reportedly astonished by Caril Ann's unleashed viciousness. Dylan Klebold and Eric Harris worked together to carry out the Columbine murders. John Allen Muhammad and Lee Boyd Malvo worked together as the Beltway Snipers to kill 10 people in the Washington, DC area in 2002. Catherine Wood and Gwen Graham worked together at a nursing home when they discovered that murdering patients together enhanced their lesbian sexual pleasures [20]. Shared secrets strengthen intimacies and reinforce fantasies. This powerful dynamic is seen repeatedly in case histories and no less so, in all of its combinations and permutations, in mass murder. In many instances, it is unlikely that involved individuals would have participated in crimes had it not been for their social reinforcements.

Social influence in the form of media coverage and general social fascination can also play a role in precipitating any deviant behavior [21]. Copycat behavior can be of two forms: it can inspire tactics to hide one crime against the backdrop of another, and it can ignite a social contagion that may increase the incidence of a publicized form of deviant behavior. It is widely speculated that John Allen Muhammad was motivated in part in his random shootings by plans to shoot his estranged wife and make it look as though it were part of someone else's unrelated crime spree. There are numerous instances in which people have taken advantage of a publicized series of crimes to carry out a killing in a similar manner to divert attention from the actual perpetrator – the person with the obvious motive to commit what would otherwise have been an isolated act of murder.

Greater attention has been devoted to the belief that publicity inspires others to commit atrocities. Certainly public recognition can spread bad ideas as well as good, but one of the strengths of open societies is their marketplace of ideas. This marketplace may often seem to be saturated by tasteless and absurd examples of impropriety. However, even when faced by irresponsible media, society constantly measures its own conduct against all other examples of human behavior and generally comes to an appropriate balance. The impressionable and those already leaning in a deviant direction can view bad examples as support for their predilections [22], but to conclude that the public forum "causes" mass murder is no more correct than the reality that social forces play a role in all behavior.

Occurrence and Demographics

Most research data describe mass murderers as having been more than 90% male, and about 70% white. This is consistent with the facts discussed earlier, but by no means establishes that other groups are not at risk [23]. Many believed that only whites were serial killers, until Wayne Bertram Williams was convicted in the 1984 serial murders of 29 black Atlanta youths. One of the lessons from the Muhammad/Malvo Beltway Sniper case is that police missed several opportunities to capture the shooters because they were relying on a "profile" of the shooter as a white male (*see* **Profiles: Psychological and Behavioral**). Similarly, Countess Bathori of Transylvania, who killed 650 girls between 1604 and 1611, probably holds the all-time record as serial killer, providing one good example of the error in assuming that violence is the sole province of males [24–26] (*see* **Aggression: Gender Differences in**).

The popular media would suggest that certain setting and certain types of mass killings are now occurring with disproportionate frequency. The American slang phrase for turning into a rampaging killer, "going postal", grew out of a few post office mass shootings, while the facts show clearly that post offices are still among the safer working environments [27]. Similarly, school shootings are often portrayed in the media as on the rise, while they have actually been decreasing.

Even academics have contributed to the belief that mass murder is a recent American phenomenon, and a growing threat [28]. A careful analysis of the facts

does not support this conclusion [29]. Lists of multiple murders disclose that such crimes occur even in the most peaceful countries, although different cultures have different names for and cultural interpretations of the phenomenon such as the concept of "running amok" in the Malay/Indonesian/Filipino culture [30].

Mass violence is a problem that predates recorded history. While the largest number of deaths occur during times of war and under totalitarian rule regardless of political leanings, individual civilian incidents are rare, fairly evenly distributed over time and geography, and are associated not only with a broad diversity of contributing factors; but furthermore arise out of common human emotions and traits that influence social interactions as a whole.

As human beings, we tend to feel that our current lives are at the center of events. We also seek to solve problems and take responsibility for preventing needless suffering. Unfortunately, this perspective can also cause us to lose sight of a larger picture, and to be convinced that current events or circumstances are the cause of all calamities. In the process, simple fundamental awareness such as never assuming that anyone is "harmless" and that all people should be approached with respect and consideration can too easily fade from view.

Conclusions

Mass murders do occur, but they are very rare. Nevertheless, their catastrophic and unpredictable nature, combined with a universal fascination with inner potentials for violence, lead to many distortions in its perception. Mass killers seldom meet the criteria for insanity or even diminished responsibility defenses. Most plan their crimes systematically and with premeditation.

Mass violence can take place anywhere in the world, increase during war and under totalitarian regimes, is seem in many forms, does not fit a single pattern, and has been traced to diverse causes and circumstances. Mass murderers often lead quiet lives and do not have criminal records. While techniques have improved for the assessment of threats made by those who display prior warning signs, prevention is difficult without infringing upon important human and civil rights. There is no proof that mass murder is a recent phenomena, or that it is increasing because of current social conditions. The dynamic of mass violence parallel the dynamics of human behavior in general, and as a problem, is best addressed not as a harbinger or alien concept but as an unfortunate aspect of human existence.

References

[1] Vold, G.B., Bermard, T.J. & Snipes, J.B. (2001). *Theoretical Criminology*, 5th Edition, Oxford University Press, New York.

[2] Geberth, V.J. (2006). *Practical Homicide Investigation*, 4th Edition, CRC Press, Boca Raton.

[3] http://www.ojp.usdoj.gov/bjs/. (accessed Oct 2008).

[4] Kelleher, M.D. (1997). *Flash Point: The American Mass Murderer*, Praeger, Westport.

[5] An unofficial compilation of mass murders by type and case status can be found at, http://en.wikipedia.org/wiki/Mass-Murder. Which also contains links to accounts of most of the actual crimes. (accessed Oct 2008).

[6] Bell, M. (1985). *The Turkey Shoot*, Grove Press, New York.

[7] The films include the 1973 Badlands directed by Terrence Malick and starring Martin Sheen and Sissy Spacek; and the 1994 Natural Born Killers. Horror writer Stephen King was strongly influenced by the case as a child and kept a scrapbook of news stories on the pair of killers; and Bruce Springsteen's "Nebraska" album is based upon the Starkweather/Fugate killing spree, http://www.imdb.com.

[8] Dietz, P. (1986). Mass, serial, and sensational homicide, *Bulletin of the New England Medical Society* **62**, 477–491.

[9] Holmes, R.M. & Holmes, S.T. (2000). *Mass Murder in the United States*, Prentice Hall, Upper Saddle River.

[10] Meloy, J.R., Hempel, A.G., Mohandie, K., Shiva, A.A. & Gray, B.T. (2001). Offender and offense characteristics of a nonrandom sample of adolescent mass murderers, *Journal of the American Academy of Child and Adolescent Psychiatry* **40**(6), 719–728.

[11] Kohlberg, L. (1984). *The Psychology of Moral Development: The Nature and Validity of Moral Stages*, Harper and Row, New York.

[12] Moral development is most frequently a factor among juveniles as discussed in Palermo, G.B. & Ross, L.E. (1999). Mass murder, suicide, and moral development: can we separate the adults from the juveniles? *International Journal of Offender Therapy and Comparative Criminology* **43**(1), 8–20.

[13] Baatz, S. (2008). *For the Thrill of It: Leopold, Loeb and the Murder that Shocked Chicago*, HarperCollins, New York. For a compilation of thrill murders with links to most of the cases see http://en.wikipedia.org/wiki/Thrill_killing.

[14] For a compilation of news coverage of the case http://topics.nytimes.com/top/reference/timestopics/

people/t/robert_w_tulloch/index.html. (accessed Oct 2008).
[15] Hazelwood, R. & Michaud, S.G. (2001). *Dark Dreams Sexual Violence, Homicide and the Criminal Mind*, St. Martin's Press, New York.
[16] Klineman, G. & Butler, S. (1980). *The Cult That Died*, G.P. Putnam's Sons, New York.
[17] Hall, J.R. (1987). *Gone from the Promised Land: Jonestown in American Cultural History*, Transaction Publishers, New Brunswick. The author is indebted to psychiatrist James K. Knoll for sharing with the author his extensive research files on Jonestown.
[18] DeWall, C.N. & Baumeister, R.F. (2006). Alone but feeling no pain: effects of social exclusion on physical pain tolerance and pain threshold, affective forecasting, and interpersonal empathy, *Journal of Personality and Social Psychology* **94**(1), 1–15.
[19] Fox, J.A., Burgess, A.W., Levin, J. & Wong, M. (2006). *Panel Report of the March 25, 2006, Capitol Hill Shooting*, Submitted to The Seattle Police Department, July 17, 2006.
[20] Ramsland, K. (2007). Women who kill together, *The Forensic Examiner*, September 64–66.
[21] Coleman, L. (2004). *The Copycat Effect: How the Media and Popular Culture Trigger the Mayhem in Tomorrow's Headlines*, Pocket Books, New York.
[22] Mullen, P.E. (2004). The autogenic (self-generated) massacre, *Behavioral Science and the Law* **22**, 311–323.
[23] Edwards, C.N. (2001). *Responsibilities and Dispensations: Behavior, Science, & American Justice*, Four Oaks Press, Dover.
[24] Segrave, K. (1992). *Women Serial and Mass Murders*, McFarland, London.
[25] Kelleher, M.D. & Kelleher, C.L. (1998). *Murder Most Rare: The Female Serial Killer*, Random House, New York.
[26] Segrave, K. (1992). *Women Serial and Mass Murders*, McFarland, London; Kelleher, M.D., & Kelleher, C.L. 1998. *Murder Most Rare: The Female Serial Killer*, Random House, New York; See also, for example, (1995). Serial killer gets life sentence, *Atlanta Journal & Constitution*, November 22, A10. Regarding Rosemary West. Roy Hazelwood's research with the wives of serial rapists provides graphic examples as to the extent women often accompany their husbands in violent, sexual crimes (personal communications, Dundee, Scotland, June 28, 1996).
[27] http://permanent.access.gpo.gov/lps12068/33994.pdf. The homicide rate for postal workers is 0.26 per 100,000; while retailers, who include convenience store operators who often face burglars, have a rate of 2.1 homicides per 100,000 workers; and taxi cab driver are at 31.54 per 100,000 workers. (accessed Oct 2008).
[28] Fox, J.A. & Levin, J. (1994). *Overkill: Mass Murder and Serial Killing Exposed*, Plenum, New York.
[29] Duwe, G. (2006). A circle of distortion: the social construction of mass murder in the United States, *Western Criminology Review* **6**(1), 59–78.
[30] http://en.wikipedia.org/wiki/Running_amok. (accessed Oct 2008).

CARL N. EDWARDS

Homicide: Serial *see* Serial Homicide

Homicide Followed by Suicide *see* Suicide (Behavior)

Homicide of Elderly *see* Elder Abuse: Policy

Homosexual Panic

Background and Definition

In its most general definition, "Homosexual Panic" (also referred to as *Gay Panic*) has been described as an adverse reaction, whether affective, behaviraol, or autonomic, to a perceived homosexual advance. Homosexual panic was first described by Edward J. Kempf in 1920 as "acute homosexual panic" and has alternately been referred to as *Kempf's Disease* [1]. Kempf defined acute homosexual panic as "panic due to the pressure of uncontrollable perverse sexual cravings" (p. 477) that present with physiological indicators of arousal (e.g., increased blood pressure), an overly defensive state against possible homosexual encounters, and sensory disturbance (e.g., visions, voices, and complaining of feeling "dopy") ([2], p. 514).

Kempf's conceptualization of acute homosexual panic was strongly influenced by Sigmund Freud's psychoanalytic theory. Acute homosexual panic was seen as the ego's response to the id's "lurking (homosexual) libidinous urges, over which the (individual) has no control" ([3], p. 296). Homosexual panic was similarly described by Glick [4] as an "acute episodic schizophrenic reaction" due to the individual's "unconscious wish to present himself as a homosexual object with the expectation of dire consequences" (p. 20).

As the definition of homosexual panic has evolved, it is now used less to describe an internal psychological condition and more as an acute reaction when one's heterosexuality is externally "threatened" by perceived homosexual advances. Indeed, some psychologists (e.g., [5]) continue to describe homosexual panic as an adverse psychological reaction that is singly rooted in internal causes; specifically, repressed homosexual desires and insecurity. However, the majority of current definitions describe homosexual panic as the occurrence of rage or violence due to an external stimulus (i.e., a perceived homosexual advance) combined with an individual's ambivalence toward homosexuality. Additionally, current descriptions of homosexual panic typically omit the occurrence of a psychosis.

Legal and Psychological Use

Homosexual panic has never been recognized as a disorder by the Diagnostic and Statistical Manual (DSM) of the American Psychiatric Association [6]. Legally, however, homosexual panic has been used in defense of acute violence both successfully and unsuccessfully. While the homosexual panic defense has resulted in sporadic cases of jury nullification (i.e., acquittals), it is most frequently used to justify charging defendants with a crime less serious than the potential maximum charge (e.g., charging aggravated assault instead of attempted murder) and in mitigating sentences.

In the highly publicized "Matthew Sheppard case" [7], a homosexual panic defense was raised. The defense alleged that the defendant, Aaron McKinney, went into a violent rage subsequent to being sexually propositioned by Matthew Sheppard. The Judge rejected the defense, stating that it amounted to a temporary insanity or diminished capacity defense, both of which were disallowed by Wyoming law. In the case of *West Virginia v. Bee* [8], a homosexual panic defense was successfully raised. In this bizarre case, Dean Ludwig Bee stabbed Dent Pickman to death after Bee accepting a ride home with Pickman from a bar. In his trial, Bee stated that he woke up at Pickman's house to find his body covered in various condiments and Pickman licking them off. Bee asserted that he "went crazy", leading to Pickman's death. Bee was acquitted of murder by a West Virginia jury.

While the *Sheppard* and *Bee* cases show that the homosexual panic defense has been both rejected and accepted by American courts, the most common result of such a defense is a "compromise verdict" (i.e., convicting a defendant of a less serious offense or passing a minimum rather than maximum sentence). One of the most well-known examples of homosexual panic and a compromise verdict is the 1995 "Jenny Jones case". In 1995, Scott Amedure went on *The Jenny Jones Show* (a then-popular talk show) to express his sexual attraction toward his friend Jonathan Schmitz. Soon after the taping of the show, Schmitz killed Amedure. In *Michigan v. Schmitz* [9], Schmitz admitted to the killing but claimed that he was humiliated by Amedure's sexual advances, exacerbated by the fact that the show aired on national television. Although Schmitz was charged with first-degree murder, he was ultimately convicted of a lesser crime of second-degree murder.

Homosexual Panic and *Daubert*

Currently, most conservative standards regarding the admissibility of scientific evidence in all federal and most state courts are set forth in the U S Supreme Court decision in *Daubert et al. versus Merrell Dow Pharmaceuticals, Inc*. [10]. This decision requires that, to be admissible in court, scientific evidence must: (i) be amenable and have been submitted to scientifically testing, (ii) have been peer reviewed, (iii) include a known error rate, (iv) have limits regarding its definition, and (v) have widespread acceptance within the relevant scientific community.

While some of the *Daubert* language is imprecise (e.g., widespread) and has been debated in relation to various scientific methods and concepts, its relation to the homosexual panic defense is clear. Readers of the clinical literature related to homosexual panic

will note its heavy psychoanalytic underpinnings, a theoretical perspective that is plagued by an inability to directly observe, thus validate, its central concepts. Moreover, homosexual panic has been characterized in many ways. In some literature it is defined by internal distress or anxiety. Others have described it as an acute panic to an external stimulus, leading to rage and violence.

With this lack of diagnostic specificity, homosexual panic does not meet the standard of having been or being amenable to scientific testing. Owing to this lack of rigorous study, there exists no known prevalence data, prognosis, or error rate for homosexual panic. Moreover, homosexual panic is clearly deficient in the area of "widespread acceptance", in that literature regarding the phenomenon is lacking and clinicians have varying opinions as to its validity. It must thus be concluded that homosexual panic does not meet the standards put forth by *Daubert*.

In a definitive legal review, Wall [1] examines the use of homosexual panic in evaluating criminal responsibility and diminished capacity. Wall concludes that, "... there is little empirical evidence to justify excusing or mitigating criminal behavior on the basis of anxiety stemming from a nonviolent homosexual advance" (p. 458). His recommendation that "... forensic evaluators should be wary of advocating for the potential relevance of homosexual panic because there is poor evidence to support its existence" is supported by this review, as is his statement, "... the poorly understood relationship between being the recipient of a homosexual advance and the carrying out of criminal behavior should not be elevated to the status of a syndrome or excuse by forensic evaluators" (p. 459).

Summary

Homosexual panic is currently described as a rage reaction to a perceived homosexual advance. As a defense to an acute violent episode, it is indeed problematic. Homosexual panic fails to meet standards set forth by *Daubert* in several ways. Homosexual panic lacks diagnostic specificity and widespread acceptance by clinicians. It has not been rigorously tested and lacks ability to be evaluated scientifically. As such, the incidence of homosexual panic has no known prevalence, prognosis, or associated error rates. It is thus a concept that lacks exculpatory value, although it continues to be used to explain extreme reactions to perceived homosexual advances.

References

[1] Wall, B.W. (2000). Criminal responsibility, diminished capacity, and the homosexual panic defense, *Journal of the American Academy of Psychiatry and the Law* **28**, 454–459.

[2] Kempf, E.J. (1920). *Psychopathology*, CV Mosby, St. Louis, MO, pp. 477–515.

[3] Cassity, J.H. (1941). Personality study of 200 murderers, *Paper Presented at the Annual Meeting of the American Psychiatric Association*, Cincinnati, OH.

[4] Glick, B.S. (1959). Homosexual panic: clinical and theoretical considerations, *Journal of Nervous and Mental Disease* **129**, 20–28.

[5] Walter, R. (1985). Homosexual panic and murder, *The American Journal of Forensic Medicine and Pathology* **6**, 49–51.

[6] American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders (DSM) [currently DSM-IV with DSM IV-R proposed 2000]*, American Psychiatric Association, Washington, D.C.

[7] Wyoming v. Aaron James McKinney, Criminal Action 6381, (2d Albany County 1999).

[8] West Virginia v. Bee, 85 F 2 (Richie County 1985).

[9] Michigan v. Schmitz (Cir. Oakland County 1996).

[10] Daubert *et al.*, v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 529 (1993).

Related Articles

Syndromes: Psychological

Temporary Insanity

TRACY A. THOMAS AND WILLIAM J. FREMOUW

Hormonal Treatment *see* Sex Offenders: Treatment of

Hostages *see* Policing and Critical Incident Teams

Hostages: Behavior of *see* Stockholm Syndrome

Human Factors: Industrial Incidents

Introduction

This is a review of the basic issues commonly encountered in product liability cases, with an emphasis on human factors (HFs). HFs (ergonomics) is the study of the capabilities and limits of the human, which must be considered in the design of a safe product. HF topics can be summarized by how we perceive, (vision, hearing, touch, taste, and smell), how we process information (cognitive functions, memory, learning, errors, motor, and control processes), and how we move in a physical world (physiology, anthropometry, and biomechanics).

Product liability litigation is based on the concept that an injured party (plaintiff) can seek to recover damages for personal injuries or loss from the manufacturer (defendant) and others in the chain of commerce, because the plaintiff believes that the damages were caused by an unreasonably dangerous and therefore defective product. Industrial incidents tend to take on two flavors; a dramatic accident with chemical releases, collapses, and explosions or smaller scale failures related to equipment failures, misuse, or design failures. Both flavors invariably have significant elements of HFs that directly or indirectly contributed to the event.

The courts are challenged to maintain a balance between the benefits of the products produced by the manufacturers and the risk to users of these products, namely the consumers, or end users in commerce. In effect, the court balances the responsibility (duty) of the manufacturer to provide a safe product and imposes on the ordinary consumer a similar responsibility to use the product in a responsible manner. This balance is played out in product litigation cases, and ultimately the jury decides the issues, with varying consistency. At the heart of the argument is always the reasonable and foreseeable use of the product, and the foreseeable misuse of the product, and in the case of a failure of a product, the manner in which it fails.

Human Users of Human-Designed Products

Many of the challenges confronting today's human senses did not exist a mere 200 years ago. The tasks of driving a complex vehicle, sitting and typing at a computer, using a cell phone, or following written instructions are all new environments for the human. Further, our abilities as individuals vary significantly, by stature, gender (yes, they are different), and age, not to mention individual variability in the summary elements just mentioned.

Because HF principles are so fundamental to the design and utility of all products, there are many instances of product liability actions (as well as many accident reconstruction cases) that involve a HF expert on behalf of the plaintiff or defendant. Almost always there are questions as to the adequacy of warnings and instructions. Operating controls, warning lights, and their location and functionality, visibility, and field of view are other common areas of interest. The further one investigates a near miss, material loss, injury accident, or fatality, the more insight one gains from considering the HF aspects at each stage. Eventually, we confront the fact that we live in an imperfect world and that errors can creep in at many levels. Therein lays the true challenge, to design in a manner that mitigates potential errors, by both the manufacturer and the user.

Product Liability

The primary purpose of designing and manufacturing consumer products is to put them to use by other humans, while recovering costs and making a profit. This simple fact applies to every object produced by the creative genius of mankind ranging from housing, transportation, and the myriad of tools and appliances we use in our lives. The degree to which these products are successfully used is directly linked to the skill of the designer, particularly in understanding the user population abilities and limitations.

One of the recurring themes in the forensic field is the finding that the designer (if one can be identified) rarely has any training in the application

of safety or Human Factors Engineering (HFE) or related fields. The result is that all too often known design principles are ignored, violated, or misapplied. For example, a common consumer reaction is "who designed this?" or "how is this thing supposed to work?" or "how am I supposed to get my finger in here?". One need look no further than the common appliances (TV/VCR/DVD controls and cell phones) we find in ordinary life. When consumer products exhibit injury or fatality failure modes linked to a lack of design foresight, the question is no longer one of inconvenience or frustration, but injuries or fatalities, and becomes the focus of investigations and litigation.

Hence there is the necessity for the designer and manufacturer to seriously address the needs and limitations of the consumer, whether in the private or industrial sector. The product must be physically robust to meet the conditions of use and functionally suited to be used safely by the consumer. Finally, the product must be produced free of defects, or if they are present, must fail in a safe mode. If any of these considerations fail, an injury can result.

Few product engineers or designers receive sufficient engineering training, and fewer yet complement their basic technical knowledge with study in the areas of HFs or safety engineering. All too often, the deliberate (often ignorant) choice is made to use warnings in order to mitigate a known design or manufacturing hazard (Be Careful of Sharp Edges). *Uloth v. Cititak Corp* (1970) established the duty of the designer or manufacturer to the consumer. A warning is not sufficient to avoid responsibility for negligence, when a design change could have mitigated or eliminated the hazard.

In other instances, a marketing directive supercedes or undercuts the engineering criteria. Marketing or business decisions may dictate an unsafe design feature, safety compromises, derating or reduction of appropriate safety factors, merely to compete with another equally unsafe design. Standards, which are often consensus standards, can become mired in committee politics that are by definition dominated by members with an interest to be on the committee, and all too often a vested interest.

To balance this duty of care or responsibility placed on the designer and manufacturer, the consumer's responsibility is to be informed of the instructions and warnings, to use the product in the intended manner, and to protect themselves and others from hazards that may be reasonably encountered in the intended use of the product. In the end, the benefit and utility of the product to the consumer must outweigh the risk of harm.

Contributing Sciences

HFs draw from a wide spectrum of technologies that have evolved as disparate fields of study. The study of humans and work (ergonomics) evolved in Central Europe with the coming of the industrial age. Its more technical offspring HFE originated in the developmental work during World War II aimed at improving the performance and survivability of complex weapons systems. Building on this foundation, space exploration and even nuclear ballistic missile technologies all added to our understanding of human performance limits.

In a dated but nevertheless illustrative diagram, F. Dukes-Dubos represented the roots and interrelationships of HFs (ergonomics) as the mathematical, physical, and biological sciences, including also social and behavioral Sciences. (ASSE Journal, Oct. 1972) In essence, all sciences are represented.

In HFs work, it is easy to recognize the contributions of statistics, fluid dynamics, physical anthropology or psychology or biomechanics. All these fields and their many related areas are combined to further our understanding of what works best for human use. The end product is seen in specialty fields, from sports medicine to bioenvironmental engineering or HFE.

Interaction of Human Factors, Ergonomics, Safety, and Safety Engineering

In product liability forensic practice, and for that matter in many industrial investigations, the key issues are always the same. What were the physical and human aspects of the situation that contributed to the event sequence, and what role did the human-designed product have in the outcome? These questions are not new. Over thousands of years of recorded history, the methods of answering these questions have evolved from simple direct retribution or thoughtful judgments by elders to today's expensive and all too often ineffective legal proceedings.

One can argue that the origin of law itself is rooted in the formalized retribution for perceived wrongs going back to the Code of Hammurabi. In

recent times, based on English common law, we have adopted statutory and regulatory law as a means to safeguard ourselves as well as important property. We invoke laws that are intended to make us safe, provide for a safe workplace, and safely transport us from place to place. We base these laws on our understanding of the cause and effect relationships between unsafe acts, behaviors, products, and injuries or deaths.

HFs provides the foundation for human safety by defining the boundaries or limits of human perception and performance, considering the five senses, and modifiers such as age effects or cultural stereotypes. HF answers questions such as how well can we see under specific light conditions, what is the human response to vibration, how do we respond to perceptual cues in the visual field, how to we perceive dangers and react in a timely manner, how do we respond to complex cognitive (decision) situations, how do we interpret warnings, amongst many. The practice of HFs has grown to include every aspect of human activity from cognitive modeling, to aging, aerospace systems, communications, ergonomics, computer systems, vision, medical and transportation systems, and so on.

Ergonomics, as practiced in the United States (as opposed to the more workplace-oriented practice in the European Union (strike EU)), provides guidelines for human performance optimization for routine interactions with our man-made environment. Ergonomics answers questions such as how can we design our products to minimize human fatigue or discomfort, how much can we carry without undue strain, and how long can we work under severe heat or cold conditions?

Safety programs evaluate this information, and with other sources, such as industrial hygiene (which studies the human in the atmospheric and chemical work environment) structure human behavior in the workplace, and transfers this experience to the consumer forum. Today's consumer safety consciousness is rooted in the development of safety programs in industry.

Safety engineering evolved from the realization that very complex systems such as modern weapons systems or combined human/machine/computer systems cannot be effectively safeguarded by conventional preventive safety programs. The development of system safety analysis techniques such as fault tree analysis (FTA), failure modes and effects analysis (FMEA) and preliminary hazard analysis (PHA), and their derivatives permitted the study of these complex systems in both quantitative and qualitative ways that minimized the mission risk. These safety engineering methods have now been used in the chemical, aerospace, and automotive industries. These methods are now applicable to the design and manufacturing of consumer products, and include HF issues, such as human error and foreseeable use and misuse. By extension, these same techniques can be used in the investigative process.

The emphasis on safety was institutionalized with the advent of the US Occupational Safety and Health Act (OSHA) of 1970. OSHA incorporates many lessons from safety engineering principles that apply not only to consumer goods but also to the industrial workplace.

HFs, HFE, ergonomics, safety, safety engineering, combined, summarize our knowledge of what constitutes human safety. All have a role in the forensic analysis of consumer product issues. They are complimentary elements of the study of the human, the human environment, and the interactions.

Product Lifetime

A consumer product typically follows the classic product development curve, with some notable exceptions. Products are characterized by discrete time frames such as product conception, research and development, design, prototyping, test marketing, manufacturing, distribution, useful lifetime, extended lifetime, salvage, or disposal. It is during the distribution and initial useful lifetime that many use and safety issues arise. Also, toward the end of a product's useful lifetime, more instances of product failure occur, which may raise the frequency of safety issues. In many cases, as a product reaches maturity in its cycle, another product or variant is introduced. Changes are often made for cost reduction, functionality, technology change, market needs, price points, market coverage, or just to give a product a new look. All these elements can adversely influence consumer products. HFs forensic issues often include questions of why changes were or were not made, what testing was performed, what usability, instruction, or warning aspects were affected, and when.

Product Need

Why is a product deemed necessary in the first place? Products are introduced to assist with work

or perform work not practical otherwise, to provide entertainment or recreation, to provide transportation, health care, and so on. Inevitably, the profit motive can be found in the mix of answers. Either the demand exists, and the manufacturer hopes to profit from selling into the demand, or a demand can be created through novelty, fad or astute marketing.

In either case, the product need has been identified and evaluated by someone, either empirically or through planned research or deliberate planning. Many products are introduced as enhancements or reincarnations of similar past products. Even in developed markets, similar competing products are introduced. Ultimately, the market judges the need for a product. Successful products are rarely associated with chronic safety problems in the consumer product arena. Similarly in the industrial environment, products, devices, or systems are vetted by their historical success, or discarded or redesigned after associations with lost time accidents, or OSHA fines. Successful products excel in areas of sound design, high-quality manufacturing, and appropriate instructions, warnings and warranty programs.

Products that convert, consume, produce, or transmit energy, in any form (electrical, mechanical, chemical, thermal, radiation, potential, or kinetic) are particularly susceptible to consumer product liability actions. Products that involve chemical reactions, fuels, and combustible gases must be justified from a utility standpoint, with safety as a prime consideration. HF issues can extend to the ability to smell a dangerous substance, adequacy of complex instructions regarding consumer assembly of products, fit of body parts in nip points, and length of hair in entanglement cases among others.

Legal Context of Products

Government Role: Consumer Product Safety Commission

Two serious safety problems stand out as historical motivators for US government action. With the advent of the steam boiler, (in factories and on ships) there came explosions and massive carnage. This new situation generated action in the form of the first industrial standards. The American Society of Mechanical Engineers (ASMEs) formulated boiler design standards, which were eventually adopted in1846 as Federal Government mandatory standards. Similarly, in the late 1800s and in the early 1900s, equally serious accidents caused by industrial and railroad mishaps led to the adoption of workmen's compensation programs (Theodore Roosevelt's Federal Employer's Liability Acts) and initiated the structured development of safety standards and regulations, laws that are with us to this day.

In 1970s, the OSHA was promulgated as the next major government push for industrial safety. Not long afterward, this trend came to consumer goods when Congress, in 1972, passed the Consumer Product Safety Act and created the Consumer Product Safety Commission. (CPSC).

This was the first significant safety legislation on behalf of the consumer, specifically in the area of product hazards and defects. CPSC maintains records of consumer reports of near miss or actual injuries, and on the basis of its investigations it can force the recall of consumer products. The CPSC database is augmented by the National Electronic Injury Surveillance System (NEISS), a nationwide injury reporting system from hospital emergency room reports.

Every year, the CPSC receives and investigates thousands of reports, issues product recalls, and can compel action against a manufacturer, or even preemptively ban a product deemed an imminent hazard. At any given time, the CPSC (www.cpsc.gov) lists hundreds of products (into the millions of units) that have been recalled in the areas of child products, toys, household products, outdoor products, sports and recreation products, and specialty products (but excludes automobiles, trucks, and motorcycles, which are under National Highway Traffic Safety Administration (NHTSA)/Department of Transportation (DOT)). There is no such specific oversight of industrial products; OSHA administers this issue with the overall requirement that the employer is responsible for all safety issues inside the plant or workplace boundary.

Another complimentary element of the government's presence in consumer product and industrial safety is the emergence of standard setting organizations (American Society for Testing and Materials (ASTM), American National Standards Institute (ANSI)) and industrial and product test laboratories such as Underwriters Laboratories Inc. (UL). Over the last 50 years, these organizations have developed the testing procedures, product, and performance consensus standards that have often been adopted by the

government regulators. As minimum consensus standards, there is always room to improve them, and they are not a guarantee of an adequate defense. On the other hand, failure to meet these minimum standards almost always assures that the plaintiff prevails.

Legal Theories

Strict Liability

Strict liability is one theory that the plaintiff can use to present his case. The plaintiff must prove that the product had a defective condition, the defective condition existed when the product left the control of the manufacturer, the condition made the product unreasonably dangerous, and the defect was the cause of the injury.

The defense will try to show that the defect was the result of some action or event after the product left the defendant's control, or that the plaintiff had knowledge of the existence of the hazard, and voluntarily assumed the risk.

In a HF context, again warnings and instructions are the primary issues. Those are the relatively soft targets. Secondary issues arise in the manner of use of the product, the perception of risk by the user, and the physical attributes that permit use or misuse.

Negligence

Another theory the plaintiff can use is that the defendant was negligent, and must prove that the product had a defective condition when it left the manufacturer's control, the defective condition caused the injury, and the harm was foreseeable to the manufacturer, who did not exercise due care to guard against the harm.

The defense will try to show contributory negligence on the part of the plaintiff or a failure to act as a "reasonable man under like circumstances". A variation includes comparative fault and similar fault apportioning schemes. In the modified comparative fault, the plaintiff will not recover at all if the percentage of fault is over 50% (varies by state).

This theory appears to favor the defense, in as much as any reduction of plaintiff's damages is considered a win. In a more cynical mode, the more doubt (real, smoke, or spin) that can be attributed to the plaintiff's circumstances or their actions, or in the case of minor children, their parent's actions or inactions, and even a finding of fault on the part of the defendant can totally negate any actual damage awards.

Types of Product Defects

Product Defects

Product liability issues would not exist without product defects. The basis of litigation is the identification of a defect, consisting of a design, manufacturing or warranty defect, or a combination thereof. Oftentimes, for a HF practitioner the primary focus is the analysis of warnings. However, design and manufacturing defects can contribute to the manner of use of the product, or failure mode, and so must not be ignored. Although this is the predominant arena of product liability, a fair number of industrial environment accidents result in third party lawsuits based on the premise that the product involved was defective. Otherwise, workmen's compensation insurance precludes a lawsuit against the employer.

Design Defect

A design defect exists in a consumer product if the potentially harmful aspects of the product are present as a part of the intended design of the product. A product can be unreasonably dangerous to an ordinary user by virtue of a design element, for example an unprotected gear sprocket on a go-cart. This feature, when combined with loose clothing or long hair, a reasonably foreseeable element of the user population, can result in scalping or fatal neck injuries. This is an example of a fundamental guarding design defect combined with an expected anthropometric population distribution and behavior that caused harm to the user.

A design defect can further be parsed into subcategories. These can be configurational (proper size, orientation, hazardous surfaces, and product instability), constitutional, (contains unnecessary toxic or flammable substances, inappropriate strength or flexibility or durability), functional (did not perform as intended, fostered inappropriate use, and did not perform as well as it should have), and prophylactic (lacked guards against abuse or misuse, lacked protection against injury enhancement, and lacked backup protection against failure).

Manufacturing Defect

A manufacturing defect exists in a consumer product if the potentially harmful aspects are present due to mistakes or omissions in the manufacturing process, and can include material defects, process-induced defects, and assembly errors. An unfortunately common occurrence is the omission of critical parts or bolts, which leads to the overstressing of other parts, resulting in eventual failure. Similarly, inclusions or other metal defects can lead to premature failures of springs or similar elements. The challenge to the HFs practitioner is to link the failure mode to the human error (missing assembly step) event, or inspection error, or quality control procedure. If a product does fail, is there a fail-safe modality that was included in the design, or was the manufacturing defect and its consequences not properly anticipated? The HF engineering expert must be able to dig into engineering documentation, and understand the design, testing, design change, manufacturing, and quality control processes. A comparative analysis of competing products from other manufacturers often helps to point out critical differences (e.g., steam iron, insulated wires, and terminations versus open connections in a water environment, or sealant application procedures that differ over model changes). This analysis combined with an understanding of how a product is actually used (yes, people really do test touch irons with a wet finger) can sort out seemingly unrelated events.

Marketing Defect

A marketing defect exists in a consumer product if the information accompanying the product contains potentially harmful aspects, including omissions, or errors that present an unreasonable risk to ordinary users of the product. Warnings are covered in detail elsewhere in this publication. Marketing issues include both warnings and instructions, as well as other communications that affect the usability or fitness for use of a product.

The cornerstone of warnings is the ability of the warning to elicit from the reader the safe behavior desired. Under no circumstance is a warning sufficient if the hazard could have been eliminated, or a design change could have eliminated or mitigated a hazard. Similarly, a warning is not an adequate substitute for an effective guard. Only when design and guarding (or safety device) issues have been thoroughly exhausted is a warning feasible (hierarchy of protection). For maximum effectiveness, the warning must take a specific form, and comport with HFs standards for warnings, or on a second tier, other applicable statutory standards.

Consumer Use, Foreseeable, and Unforeseeable

Foreseeability

A fundamental concept in product liability is the reasonably foreseeable use of a product, which should have and could have been anticipated by the manufacturer, and it is opposite, the unreasonable use of a product by the consumer, in the sense that such use was unintended and unforeseeable by the manufacturer. The courts do limit the definition or interpretation of “foreseeable”, to both protect the consumer and also prevent unreasonable claims against the manufacturer.

On the one hand, the manufacturer is tasked with the anticipation of all the reasonably foreseeable uses and misuses of a product, and providing a product that if it fails, does so in a passive or fail-safe manner, and that inherent (cannot be removed by design) hazards are clearly communicated to the user by appropriate means, such as warnings, instructions, or training.

The consumer is tasked with being sufficiently knowledgeable (through written instructions and warnings) about the product to avoid uses that would expose the user to obvious hazards, or situations where the user is exposed to hazards through deliberate neglect, lack of maintenance, care in use, or ignoring warnings or instructions.

Another complicating factor is the presence of children in the user environment. Assumptions made of adult users and foreseeability do not apply to children. The key differential between adults and children is experience and the knowledge that comes from it in evaluating potential hazards. Children simply have not developed the risk perception attributes of adults. Even teenagers, seemingly mature in many ways, are not adults, and are treated as such under the law.

This is an important aspect of HF forensic practice. Products from swimming pools, to playgrounds, to appliances in the garage, home or school are in the child environment. Children are creatures that taste, touch, feel, and explore their new world by

turning things upside down and inside out. They are not privy to those things that we as adults know as *hazards*, or the finality of injuries and death. There are good reasons that the dumb stunts we commonly see on television are almost exclusively the domain of adolescents. They are indeed stupid in the literal sense, and can be exploited by their peers for "fun" and others (adults) for profit. The message to other consumers from these stunts or for that matter similar product advertising stunts is that this seemingly irrational behavior is acceptable. Have you ever seen a snowmobile or go-cart or car or similar commercial where the vehicle is not speeding along? But then read the warnings on these same products, all will of course advise against speed. This aspect is a goldmine for the plaintiff, and a minefield for the defendant.

The often mentioned notion of "common sense" is predicated upon the false assumption that we humans in fact have a common understanding of safety. One need only travel to see regional, societal, and cultural differences in the notion of human safety, or the very value of human life. The manufacturer must anticipate not only the reasonably foreseeable actions of adults and children but also, in this global marketplace, the actions of all people in the international consumer market. Literacy affects the consumer's ability to understand basic technical concepts necessary to safely operate many modern products that use energy, chemicals, high pressures, or energy sources. A manufacturer must contend with this disparate user population, if they wish to tap into the potential profits of an extended marketplace.

The analysis of accidents has the benefit of hindsight. The task for the plaintiff's HF expert is usually to investigate the circumstances of use, and to determine (as a basis of their opinion) whether the consumer use in question was reasonably foreseeable and whether the manufacturer knew or should have known that significant hazards existed.

Safety is the result of a deliberate effort to analyze the need for the product, develop a complete and tested design, and provide the necessary quality control program to assure a defect-free product. The plaintiff expert's responsibility is to consider all the available information and testimony that deals with the use of the product, and develop the product performance and use characteristics with a particular emphasis on underlying causes of the injury or death. Although the accident occurred with certainty, it is important to try to be objective in the analysis of foreseeability. If the expert's conclusion is that the use was indeed foreseeable, especially based on prior cases or near misses, and that the consumer used reasonable precautions, the burden to prove otherwise falls on the defendant.

Any record of prior or repeated negative instances of circumstances that the manufacturer knew or should have known about strengthens the plaintiff's case. The plaintiff's expert must not overstate the case, or presume to know how the actual consumer plaintiff thought or perceived the conditions at the time of injury. The opinion basis, particularly in HF cases, when possible, should be derived from studies of a wide range of subjects, in similar circumstances.

As an example, forklift trucks (FLTs) are involved in thousands of injury or fatal accidents every year, almost exclusively in the industrial environment. Although they are very sophisticated machines, they are fundamentally overloaded as a matter of course, with minimal stability due to their three-point suspension and marginal counterweights. From the manufacturer's perspective, this is done to maximize mobility and to minimize transportation and operating costs, and thereby maximize sales. Although patents and literature regarding overloads go back 50 years, most FLTs are capable of lifting well in excess of their safe loads, and very few convey this fact to the operator through even rudimentary alarms. Although their counterparts, cranes, by law must indicate lifted weights to the operator, this is not the case with FLTs. The result is that detailed instructions exist in manuals, decals and warnings appear on the machines, rated capacity is printed on a plaque, but actually lifting the load amounts to guesswork by the operator. From a HF perspective, the key element, timely notice of a hazard not obvious to the operator, is missing, and can be provided at little or no cost without compromising the fundamental utility of the machine.

The defense expert will seek to find flaws in the plaintiff expert's understanding or representation of the case, or question whether the consumer's actions were foreseeable by the manufacturer. The key opportunity for the defense expert is to determine if indeed the event occurred as described. If a significantly different fact set or theory can be proved to be indeed true, the plaintiff's case may well evaporate. However, merely floating trial balloons of alternate unsubstantiated theories, or creating "smoke" is neither helpful to the court or to the case.

A faulty HF technique (plaintiff or defense) that occurs from time to time is to gather response data from seemingly random pools of subjects by the use of "loaded questionnaires". These are statistically false results that appear to support or refute the reasonableness of a consumer's behavior in the use of a product. HF experts must be particularly careful about such techniques. Working with questionnaires requires specific research experience, and going down to the local mall to do random interviews of the population does not necessarily qualify as a valid representation of how a plaintiff actions were reasonably foreseeable or not. One of the paramount difficulties with HF research is that dealing with subjects (participants) in rigorous studies is time consuming, expensive, statistically tenuous, and limited in the transferability of results to other seemingly similar situations. Many studies also suffer from the use of college age (mature?) subjects that may or may not be appropriate for the question at hand. For example, they may be appropriate for perception reaction visibility studies of drivers at night, but must be balanced with subjects of different age groups to better represent the actual user population.

Consumer Misuse

The defense can often prevail if consumer misuse can be clearly demonstrated. The plaintiff's knowledge of the existence and character of a dangerous condition and the voluntary assumption of the attendant risk can be a defense. In HFs issues, for example, given that well-written warnings and clear instructions are provided, and an obvious hazard (such as a power cutting tool) is present, misuse or misapplication by a consumer plaintiff will not make for a strong case. If, however, the hazard is not as clearly recognized, the outcome may be different. Consider a high-pressure water jet, where even playful misuse by a consumer can result in lethal injuries. The danger is not as clear as that of a cutting saw, and a consumer (adult or child) may not necessarily relate water flow to an extreme hazard, even if the warnings are adequate. Hence there may not be as clear an assumption of risk as in the preceding case. In this case, the presence of warnings is necessary to alert the user of the hazard.

If the consumer knowingly tolerates an unsafe condition, (a guard is removed) and continues to use the product, this can be construed as misuse. If however, the guard is of such poor design as to fall off on its own, or interferes with reasonable use of the device, or can be removed with little or no effort, the tables may be turned. Similarly, placing a washing machine interlock switch under the lid in plain view invites the user to place a finger on the switch, to keep the machine running, easily defeating the purpose of the switch.

Error

When humans operate complex machines or equipment, there is a significant probability of human error, particularly when unusual or unexpected conditions arise. For this reason, HFs engineering techniques must be used in the research, development, testing, and implementation of industrial or consumer products. Error or fault tolerance is important in a complex product, particularly in the area of controls. A control that can be actuated with a single motion, (shift lever, for example) is more prone to faults than the same control with an L-shaped path, requiring a deliberate sequence of two or more specific but different actions. Similarly, a fault-tolerant system must permit an easy return (undo) to the initial state. For example, a control lever on a motorized vehicle (in this case a vibrating compactor) combined directional control with vehicle acceleration and roller vibration, but did not have a detent or clearly defined neutral position, resulting in a loss of control fatality.

Another common source of errors is the rental equipment environment. As a very industrialized society, we have available to us a huge variety of capital intensive equipment for rent or lease. Things that most people would never dream or owning can be rented for industrial use, transportation, home use, or recreation. However, little or no training is provided. The written instructions may be furnished and the renter may listen to a short explanation and demonstration, and sign the fine print. But is that enough? The manufacturer and renter have a higher duty of care when knowingly providing equipment to a potentially naïve or ill-trained client. Errors will occur as the user becomes familiar with the safe operation of the equipment, and this must be reflected in the speed, response, and other characteristics of the machine. From a HF error perspective, the fact that the renter signed the rental contract with all the attendant disclaimers is not a particularly compelling defense when the renter is known not to be fully trained in the use of that particular device.

Emerging Situations

The advent of the "big box warehouse" stores has pushed products that were formerly industrial products into the home. From industrial cleaners to automotive repair equipment to massive barbeques, many such items can now be found in the home environment. Machine tools and construction equipment are no longer the domain of the farm or industrial shop. High-pressure (4500 psi) water washers, which can easily maim or amputate limbs, are used to clean driveways, and 20 horsepower chippers hog away at branches in the backyard. These types of powerful industrial appliances were not prevalent until the last decade or so. Many of these situations will involve HF experts, in the usual issues.

Another issue is the global distribution of goods. More and more pictographs are used to try to communicate important messages. Alternately, abbreviated brochures printed (in very small print) in up to 30 languages can be seen attached to products. How well this accomplishes the need to communicate effectively remains to be seen. At some point the cost savings of using universal pictographs is offset by the failure to effectively communicate important information. The safest, most direct method is to clearly identify controls and their functions in native language(s).

As the global economy continues, the Asian economy is producing more and more goods that flow to other markets. Of particular concern is the flood of cheap tools and machine parts from China and other sources. Private or industrial consumers are not necessarily aware that the original products when produced in the United States several decades ago were more conservatively rated (for example, hydraulic bottle jacks). In order to maximize profit, and compete on advertised ratings, the products themselves may have been weakened over time. Although warnings are provided, there is no attempt to convey to the user the critical information that the safety margins that existed at one time have been lowered. This is just one example of the generic problem of cheapening a product to the point where the fundamental risk–benefit balance is compromised. The beneficial aspects and, to some extent, high costs of United States and EU product safety programs have yet to reach around the globe.

Another interesting development is the emergence of a class of products that are not strictly consumer products but rather mission-specific law enforcement electronic control devices. These devices are used in a highly controlled environment by police agencies as less than lethal weapons, although lethal backup is recommended. These electronic immobilization devices are better analyzed from a system safety engineering perspective modeled after MIL STD 882 where the mission requirements are defined, and the risks/benefits are starkly different from the classic consumer product perspective. The users are highly trained, the devices are controlled, and the persons affected are usually being apprehended or arrested after other means of simple persuasion have failed. HF expert issues are somewhat different, as the usual relationship between a consumer product and an end user are masked by an employer/employee situation, which usually falls under OSHA or workplace constraints. The warnings issues are also somewhat more complicated as the training of instructors and classes of officers overrides the usual product labels and simple instructions.

These and similar issues will occupy HF experts in the years to come.

The Expert and the Court. The HF expert is presented as an individual who by training, education, and experience can offer specialized knowledge that will truthfully add to the understanding of the case circumstances by the court. The court includes the judges, lawyers, and the jury. The expert must present opinions based on the known facts of the case, deposition testimony, and the specific analysis performed by either the plaintiff or defense expert. Unfortunately, this noble goal is all too often subverted through overzealous representation, pressure from attorneys, or simple greed.

One of the problems with HF studies is that it is often not practical or affordable to replicate or even attempt full studies with a large subject pool. If it is not possible or feasible to conduct studies, then the HF expert must rely on similar topics from the literature, or a basic analysis that considers both contributing and mitigating factors in the case. For example, an analysis of a scene where visibility is an issue should list, compare, and contrast individual factors that facilitate or minimize an individual's ability to see a scene. On the basis of this analysis, founded on basic HF and HFE knowledge, the expert can present the court with an unbiased opinion.

HF is not as neat and clean as some other disciplines. For example, a metal fracture can be classified

as a brittle or fatigue failure. There are distinct characteristics that can be tested and absolutely agreed upon by opposing experts. HF and HFE disciplines are on similar firm grounds on the basics of human endurance parameters, some vision issues, and some standard warning issues. But the very nature of testing human subjects has to deal with within-subject, between-subject errors, and other errors in arriving at statistically significant conclusions in almost all fields of HF study. That is why warnings and their effectiveness are such a soft science compared to the earlier example of metal fracture. These aspects of HF research must be explained to the court to properly convey both the strengths and the limitations. One of the key elements to convey is that the HF expert can only present data that applies in a general sense for a specific population that has been studied. A HF expert should not and cannot present testimony that purports to characterize the specific response of an individual in a given situation. For example, only the actual driver of a vehicle knows what they saw in the instant before they hit the train. The HF expert can only testify to the field of view, to the ambient or luminaire lighting, to general perception reaction times, to the conspicuity of the scene, and so on. In other words, was the train visible to an average representative subject? The HF expert cannot testify as to what the driver actually saw or what they were looking at or when they saw it! It is the purpose of the court to determine the outcome of the litigation issue. The adversarial system can be used effectively to show distinctions between experts. All too often it is far from perfect.

There are things that the HF expert can do to improve their effectiveness. One is to be proficient in several complementary disciplines and the other is not to rely only on their individual experience, but rather to rely on the documented response of a large subject pool. A very important attribute is to have more than just academic experience. Working in the field and in real-world situations places theory in a far more practical perspective. Where possible, a report can be a very effective tool to sharpen one's focus on the case.

The HF experts, as all persons offering forensic testimony, are bound by their own technical abilities, knowledge of the field, and personal responsibility to offer only opinions they can truthfully swear to, which aid the court in determining the final outcome of the case.

Examples

These three cases are presented to illustrate home appliance, recreation vehicle, and construction equipment accidents, with HF elements, including risk perception, foreseeable behavior, and induced human error through defective design.

Home Washer/Arm Avulsion

A girl aged 14 was washing the family clothes. A 22-month-old child was in her care. In order to amuse the young child she decided to show him the wonderful colors as fabrics were spinning in the machine during the spin cycle. She lifted the washer lid and placed a dime into a slot in the frame next to the open washer (which she often did to keep the cycle going when adding bleach) that housed the interlock switch. The protruding plastic button was then held down by the dime, as if the lid were still closed. She then picked up the child so he could see into the spinning machine. Before she knew what happened, the child reached into the machine to touch the spinning fabric, and his arm was avulsed at the elbow as it tangled in the fabric.

Comment: There had been multiple prior instances of injuries or avulsions, even to adults, in similar circumstances. The switch was in an open and easily defeated location, negating its utility. The behavior of the girl was no different than that observed at the county fair, where small children are held up so they can see the colors on spinning paint platters mounted in drums. The case was settled on behalf of the plaintiff. HF issues included child-risk perception, reasonably foreseeable behavior to defeat the switch, and a record of past incidents. Identical washers furnished to commercial laundries had for some years been fitted with a recessed and hence protected switch, to prevent users from doing multiple loads on one paid cycle.

Go-Cart/Neck Fracture, Fatality

A young girl aged 13 asked to use her brother's new go-cart in a large enclosed corral. She was instructed about the use of the cart, and fitted with her father's helmet. She even tucked her long braided hair into

her T shirt, and put on the seatbelt. She proceeded to drive the go-cart in 33 meter (100 ft) diameter circles at a moderate rate. Her younger brother was in the area, but her parents had to leave on an emergency call. After a few minutes, the brother noted that the cart was stopped, and investigated. The hair that had been tucked away had apparently been jostled out of her shirt, and slipped behind the seat. In an instant, it wrapped around an improperly guarded shaft and drive sprocket, fracturing C2.

Comment: Because state law is a modified comparative negligence statute, the defense claimed that the parents and the decedent were negligent. Defense experts claimed the helmet size was too big, and that the parents and the girl were negligent in not sufficiently heeding the long hair warning. This latter claim was based on a questionnaire presented to adults in a mall, and teens in a classroom, wherein photos of the cart and instructions were presented and opinions were solicited as to the safety of the actions of the girl and her parents. The defense HF expert refused to answer questions related to the guard.

Plaintiff's expert testified that the product was defective, that testing (dangling a scarf behind the seat) was inadequate, that the warning was inadequate when a simple guard would have eliminated the hazard, that the helmet was irrelevant given the rapidity of the snap (fractions of a second), and that even if the parents were there the outcome would be the same. Plaintiff's expert further testified that the guard was easy and inexpensive to implement, and that the questionnaire data was not inappropriate. Anthropometric data was presented with respect to stature based on recent photos, and data was obtained documenting the prevalence of long hair in the general population.

A reconstruction of the accident identified the resting position of the decedent, which was masked in the coroner's photos by a sheet. Exemplar wigs worn on persons of varying stature in the go-cart demonstrated that the hair would readily get into the sprocket. Furthermore, evidence of eight or more similar past cases was presented, which involved young children and teens that suffered complete scalpings, or fatal neck fractures. The go-cart industry knew of these incidents, but resisted adopting a standard.

The jury found for the plaintiffs, the parents, and found the go cart product to be defective, but due to jury instructions attributed 50% comparative negligence (10% decedent, 20% each parent) to the plaintiffs. This jury finding under Idaho law precluded any monetary award. The case was appealed to the Idaho State Supreme Court, which affirmed the decision.

The CPSC however, on August 15, 2000, issued a recall for all 91 000 affected Manco Go-Carts, in *Woodburn v. Manco* (Idaho). Industry standards now call for a complete molded guard behind the driver's seat. Ironically, in 1996, the same firm had had to recall carts because of exposed rotating rear axles.

Compactor Rollover/Fatality

An experienced construction worker, who had previously used a rented vibrating compactor machine of a similar make and model, was operating on a pad being built adjoining a slope. The operator and machine went off the edge of the pad unseen, and overturned, crushing the driver.

Comment: The defense claimed that the operator was unsafe, and drove the machine off the pad through inattention, although there were no eyewitnesses. The plaintiff's expert testified that the machine-directional control switches were in fact changed after the incident and before the machine was rented again some months later. The control lever operated a cable control line that operated a cam, which actuated a pair of opposed switches that operated the drum vibrators and a neutral position switch on the dash. The cable also actuated the hydrostatic transmission in forward, reverse, and speed. Exemplar machines demonstrated that the switches were commonly defective, difficult to adjust, and unreliable. In effect, the control of the machine was easily compromised, giving the operator little feedback, and only gross control over the speed and direction. Furthermore, defective adjustment permitted the vibrators to continue to vibrate when the machine was supposedly stopped. Marks in the pad indicated that the machine was reversing close to the edge of the pad and that the machine had vibrated off the edge. The reconstruction of the event indicated to the plaintiff's expert that the operator likely tried to stop the machine, but the continued vibration walked the unit off the edge. This effect was observed in exemplar machines. The expert concluded that the design was defective, depriving the operator of sufficient feedback to prevent a possible fatal control error.

The case was settled in favor of the plaintiff.

Table 1 Forensic Engineering Case Expert Considerations

Primary factor	Considerations
Defects, design	Present as part of the intended design
	Can induce operator errors
	Not obvious to ordinary consumer
	Fails in hazardous mode
	Dangerous beyond consumer's ordinary expectation
	Risk of harm > benefits of design
Defects, manufacturing	Mistakes in manufacturing
	Human errors, omissions, reversals, substitutions
	Quality control deficiencies
Defects, warranty	Hidden hazard w/o warning
	Inadequate warning, high residual risk
	Warning instead of design remedy
	Not warning to protect marketability
Product not defective, design	Meets consumer expectations
	Risk reduced by safety features
Product not defective, warranty	Adequate warning of risk
	Safe use if warning is heeded
	Conspicuous, immediate reaction
	Choice made to accept risk
	Not feasible to further reduce risk
	Benefits > risks, cannot be obtained otherwise
Foreseeable use	Meets ordinary consumer's expectations
	Used within capability of product
	Children have lower or no risk perception
Foreseeable misuse	Can be reasonably expected in the environment of use
	Product lends itself to such use, even if unintended
	Caused by human error
Government databases	CPSC: Consumer Product Safety Commission
	NEISS: National Electronic Injury Surveillance System
	NHTSA: National Highway Traffic Safety Administration
	DOT: Department of Transportation
Standards	ANSI: American National Standards Institute
	ASTM: American Society for Testing and Materials
	Compliance with standards not adequate defense
Errors	
Product design	Engineering documents, change orders
	Hazard analyses or FMEA or FTA documents
	Communications regarding failure analysis, incident reports
Product testing	UL: Underwriter's Laboratories communications
	Manufacturer test records

Checklist

Main Factors to Consider, In Addition to Basic HF Issues

A checklist is presented in Table 1, Forensic Engineering Case Expert Considerations, which can be helpful in the formulation of investigative and analytic issues. The reader should bear in mind that additional checklists can be used in a matrix that considers individual contributions of case or scene elements, such as perception reaction time, conspicuity, age, glare, and other HF concepts. These can then be evaluated as positive, neutral, or

negative contributions to the case element being evaluated.

Definition of Relevant Terms

Consumer Expectation Test. Beyond the scope of this review, but important to note, is the definition of relevant terms. It is important to be aware that Forensic Engineering HF investigations often deal with terms that have specific definitions, some in law, some from related fields, such as Safety Engineering. For example, the definition of a defective product. Also known as the Consumer Expectation Test, a product is defective if it is "dangerous to an extent beyond that which would be contemplated by the ordinary consumer who purchases it, with the ordinary knowledge common to the community as to its characteristics". These definitions help focus the expert on the particular aspects of the case that need development.

Further Reading

Brauer, R.L. (1994). *Safety and Health for Engineers*, Van Nostrand Reinhold, New York.

Brown, S. (1991). *The Product Liability Handbook*, Van Nostrand Reinhold, New York.

Carper, K.L. (1989). *Forensic Engineering*, Elsevier Science, New York.

Hammer, W. (1989). *Occupational Safety Management and Engineering*, 4th Edition, Prentice Hall, New Jersey.

Kroemer, K.H.E., Kroemer, H.J. & Kroemer-Elbert, K.E. (1990). *Engineering Physiology: Bases of Human Factors/Ergonomics*, 2nd Edition, Van Nostrand Reinhold, New York.

Sander, M.S. & McCormick, E.J. (1987). *Human Factors in Engineering and Design*, 6th Edition, McGraw-Hill, New York.

Sekuler, R. & Blake, R. (1990). *Perception*, 2nd Edition, Mc Graw Hill, New York.

Related Articles

Adversary Systems of Justice

Chain of Possession of Tangible Evidence

Civil Law Systems of Justice

Daubert v. Merrell Dow Pharmaceuticals

Discovery of Expert Findings

Discovery: Depositions

Discovery in the United States: Civil Cases

Federal Rule of Evidence 702

Frye v. United States

Kumho Tire v. Carmichael

Reconstruction: Accident

ADAM K. ALEKSANDER

Human Remains and Identity

The scene of crime is a well-known forensic scenario. And when human remains are in fact the issue, anthropological and archaeological techniques are indispensable. Anthropologists are progressively and frequently involved in the retrieval of or, to a lesser extent, the search for human remains [1–4].

Physical anthropologists are key personnel in the case of skeletonized or partly skeletonized remains, badly charred bodies, and buried bodies. In the case of quasi-skeletal, skeletal, or charred remains found on the "surface", the person appointed to observe, register, recover, and generally "take care" of the remains must have experience in osteology. Lack of such experience may lead to serious errors such as nonretrieval of skeletonized remains scattered across the surface by fauna. In order to recover the entire skeleton, it is necessary to be able to quickly perform an inventory of all human bones and know how to identify them. In particular, when dealing with nonadult skeletal remains, very small bone pieces can be misidentified or simply not recognized at all. This implies loss of information. Similar or even more difficult problems are encountered in the case of charred bodies, where the specialist has to retrieve all fragments of burnt bone and distinguish them from other debris (Figures 1 and 2).

Sometimes, particularly in urban contexts, commingled remains are found. Once again, the presence of an anthropologist may be fundamental for immediately verifying whether they are human or not, or if fragments of several individuals are present. This is necessary in order to simplify procedure on-site. Discovery of animal bones or of the ancient origin



Figure 1 The inside of a burnt car, where debris can still be seen on the floor



Figure 2 Debris from which fragments of burnt human bones (prevalently cranium and long bones) were recovered after a meticulous collection and study of all material visible in Figure 1



Figure 3 Example of commingled human remains found in an urban context during gas works. In such cases, the presence of the anthropologist on-site will be useful for verifying the type of remains. In fact if remains are commingled, it is necessary to verify whether several individuals are present. The presence of numerous individuals, in an urban context, may help exclude forensic interest



Figure 4 A quick assessment of the minimum number of individuals (done by counting the number of the most frequently found bone) may help indicate whether one is considering the burial of a single individual or of several subjects. In this figure the presence of two right temporal bones (to the right), a part of an adult humerus (top left), and a child's humerus (lower left) indicate the presence of at least three individuals

of bones may exclude a forensic interest (Figures 3 and 4).

In the case of buried remains, these must be exhumed and collected with stratigraphical and archaeological strategies. Retrieval and registration of remains without archaeological methodology will entail loss of body parts, loss of stratigraphical information useful for determining the postmortem interval, and damage to bones, which will lead to difficulties in interpreting bone trauma.

Sometimes, however, the task is different: the remains still have to be found. This usually is the case of criminals who performed or witnessed a killing and burial and who, maybe after years decide to tell investigating authorities where the body was buried. A search must be performed in order to do this. Forensic archaeology is necessary in these cases. Forensic archaeology is the discipline of archaeology applied

to legal matters. It is the adaptation to the forensic scene of archaeological sciences, intended not only as the study of postdepositional events, but also as the application of archaeological methodology to excavate and recover human remains. Application of forensic archaeology is therefore crucial in reconstructing perimortal and postmortal events – and in particular when the site of burial has to be found [5–11].

It may in fact happen that a person has gone missing and that authorities suspect he or she has been killed and buried. Search for burial sites of human remains may be performed with different methods. Aerial photography is one method of detecting anomalous ground peculiarities which may indicate a burial site. Aerial photography allows one to have a large perspective of a suspected area and detect anomalies in geomorphology or vegetation which may suggest the presence of a burial site. All soil variations in fact influence the vegetation above. When a body is buried, soil and vegetation are drastically affected. Thus long-lasting alterations may be found. Aerial photography is however a costly means of search and frequently obstructed by the presence of trees in woody areas and buildings. Other more commonly used methods can be field walking, geophysical techniques, and cadaver dogs.

Field walking involves detailed observation of open areas performed by rows of experts walking in a line next to one another so that there may be a 20–30% superimposition of each participant's field of vision with his or her neighbor. Some researchers recommend using a T-bar which may test the consistency of soil. At times it may be convenient to test the area with a metal detector, particularly if the body is known to have been buried with metal artifacts. Geophysical methods are based on the use of specific traits of the soil, such as conductivity and resistivity. They indicate alterations in the underlying soil which may point to a burial site. Cadaver dogs may also be used. They are dogs that have been trained to detect decomposing remains. It is not well known if they can be species-specific, but, particularly in cases where postmortem interval is not too long, they may be useful in detecting the bodies [12–23]. All the above-mentioned methods, however, entail the risk of false positives (which, at the most, will oblige one to excavate an area that does not contain human remains) and false negative results. This is more risky. If these methods do not give a positive result, this does not mean that there are no human remains buried in that area. The only way to exclude a suspected area is to perform an archaeological excavation. This consists in excavating the area according to archaeological methods so that significant soil characteristics come to the surface. The archaeologist will therefore proceed to excavate carefully and in depth the entire area until he or she either finds the remains or can conclude, having excavated the area thoroughly, that a specific area has not been used as a burial site. All this has to be performed following archaeological methods and with appropriate tools. Archaeological excavation is carried out by the removal of individual layers and by the investigation of the contents of respective layers. Excavation by the archaeological method has the following main goals: to delineate stratification of the site; to preserve the context and identify the remains *in situ*; to avoid producing lesions on the human remains which may hinder interpretation of trauma; and to establish relationships between the different objects found. Frequently, human remains are excavated with inadequate methods – picks and shovels (Figures 5 and 6). This inevitably leads to loss of information and destruction of forensic evidence. Fundamental questions that magistrates may ask and need to be answered could be the following: the original depth and position of burial (in mafia cases for example it is important to verify the witnesses' reliability); cause and manner of death; identification; and time since



Figure 5 Improper excavation of human remains, with a shovel and pick, which has led to destruction of the skeleton and loss of information



Figure 6 Proper excavation of a skeleton, where the position of every bone has been preserved and no damage has been done to the skeleton



Figure 7 In 2001 a young man confessed to having killed and buried a male and female teenager six years prior to his confession. The woody area visible in the picture was the general site indicated



Figure 8 After removal of the initial layer of soil, an oval outline appeared in the soil, which was lighter in color than the surrounding area and of approximately 2 m × 1.5 m



Figure 9 The fill was fully excavated with archaeological methods



Figure 10 Human remains appeared at the bottom of the pit, at a depth of approximately 2 m

death. All these questions may remain unanswered if recovery is not performed appropriately.

The "tomb" must be excavated delicately layer by layer, with horizontal movements of a trowel. Once all the remains are exposed, a grid is then performed and correct positioning of every bone and object is performed (Figures 7–10). Newer technology, such as laser scanners, applied on typical scenes of crime, allow the operator to immortalize the original position of the skeleton.

References

[1] Hoshower, L.M. (1998). Forensic archaeology and the need for flexible excavation strategies: a case study, *Journal of Forensic Sciences* **43**(1), 53–56.

[2] Hunter, J., Roberts, C. & Martin, A. (2002). *Studies in Crime: An Introduction to Forensic Archaeology*, 1st Rev. Edition, BT Basford, London.

[3] Hunter, J. & Cox, M. (2005). *Forensic Archaeology*, Ist Edition, Routledge.

[4] Miller, P.S. (1996). Disturbances in the soil: finding buried bodies and other evidence using ground penetrating radar, *Journal of Forensic Sciences* **41**(4), 648–652.

[5] Moore, C.E., Davis, B.D. & Leney, M.D. (2002). Analysis of pilot-related equipment and archaeological strategy in the recovery of aircrew losses from the Vietnam War, *Journal of Forensic Sciences* **47**(6), 1210–1214.

[6] Owsley, D., Ubelaker, D.H., Houck, M.M., Sandness, K., Grant, W., Craig, E., Woltanski, T. & Peerwani, N. (1995). The role of forensic anthropology in the recovery and analysis of Branch Davidian Compound victims: techniques of analysis, *Journal of Forensic Sciences* **40**(3), 341–348.

[7] Rebmann, A., David, E. & Sorg, M.H. (2000). *Cadaver Dog Handbook. Forensic Training and Tactics for the Recovery of Human Remains*, CRC Press, Boca Raton.

[8] Rodriguez, W.C. & Bass, W.M. (1993). Decomposition of buried bodies and methods that may aid in their location, *Journal of Forensic Sciences* **30**(3), 836–852.

[9] Ruffell, A. (2005). Searching for the IRA "disappeared": ground-penetrating radar investigation of a churchyard burial site, Northern Ireland, *Journal of Forensic Sciences* **50**(6), 1430–1435.

[10] Sorg, M.H., David, E. & Rebmann, A.J. (1998). Cadaver dogs, taphonomy and postmortem interval in the Northeast, in *Forensic Osteology: Advances in the Identification of Human Remains*, K.J. Reichs, ed, Charles C Thomas Publisher, Springfield.

[11] Spenneman, D.H. & Franke, B. (1995). Archaeological techniques for exhumations: a unique data source for crime scene investigations, *Forensic Science International* **74**(1–2), 5–15.

[12] Ubelaker, D.H., Owsley, D.W., Houck, M.M., Craig, E., Grant, W., Woltanski, T., Fram, R., Sandness, K. & Peerwani, N. (1995). The role of forensic anthropology in the recovery and analysis of Branch Davidian Compound victims: recovery procedures and characteristics of the victims, *Journal of Forensic Sciences* **40**(3), 335–340.

[13] Webster, A.D. (1998). Excavation of a Vietnam-era aircraft crash site: use of cross-cultural understanding and dual forensic recovery methods, *Journal of Forensic Sciences* **43**(2), 277–283.

[14] Schulz, J.J., Wheler, S.M., Williams, J.W. & Dupras, T.L. (2005). *Forensic Recovery of Human Remains: Archeological Approaches*, CRC Press.

[15] Killam, E.W. (2004). *The Detection of Human Remains*, IInd Edition, Charles C Thomas Publisher.

[16] Brothwell, D.R. (1981). *Digging up Bones*, IIIrd Rev Edition, Cornell University Press.

[17] Haglund, W.D. & Sorg, M.H. (2001). *Advance in Forensic Taphonomy – Method, Theory and Archaeological Perspectives*, CRC Press.

[18] Wiltshire, P.E. (2006). Consideration of some taphonomic variables of relevance to forensic palynological investigation in the United Kingdom, *Forensic Science International* **163**(3), 173–182.

[19] Warren, M.W. & Schultz, J.J. (2002). Post-cremation taphonomy and artifact preservation, *Journal of Forensic Sciences* **47**(3), 656–659.

[20] Mildenhall, D.C., Wiltshire, P.E. & Bryant, V.M. (2006). Forensic palynology: why do it and how it works, *Forensic Science International* **163**(3), 163–172.

[21] Myller Coyle, H., Lee, C.L., Lin, W.Y., Lee, H.C. & Palmbach, T.M. (2005). Forensic botany: using plant evidence to aid in forensic death investigation, *Croatian Medical Journal* **46**(4), 606–612.

[22] Campobasso, C.P. & Introna, F. (2001). The forensic entomologist in the context of the forensic pathologist's role, *Forensic Science International* **120**(1–2), 132–139.

[23] Van Dijick, P.J. & Van de Voorde, H. (1984). Evaluation of microbial soil identity in forensic science, *Zeitschrift für Rechtsmedizin. Journal of Legal Medicine* **93**(2), 71–77.

Related Articles

Anthropology

Crime Scene Investigation

CRISTINA CATTANEO AND DOMINIC SALSAROLA

Human Rights Violations: Mental Health Support *see* Disaster Mental Health

Hypnosis *see* Deception: Truth Serum

Hypnosis and Memory

The legal admissibility of hypnotically recovered memories has stirred contentious debate (*see also* **Eyewitness: Suggestibility of**; **Memory: Repressed**). Court rulings regarding hypnosis have varied widely from rejecting hypnotically elicited recall to treating it as more reliable than actual physical evidence [1, 2]. Laws regarding the admissibility of hypnotic recollections diverge, with 27 states ruling hypnotically elicited recall *per se* inadmissible,[a] 4 states with precedents of *per se* admissibility,[b] and 13 states examining admissibility on a case-by-case basis.[c] In this article, we will assess the status of forensic hypnosis and evidence that pertains to disparate legal decisions.

In one of the first court cases involving hypnosis, *People v. Ebanks*, 117 Cal. 652, 665, 49 p. 1049, 1053 (1897), it was ruled that "The law of United States does not recognize hypnotism" (p. 1053). This decision was justified because at that time hypnosis was not recognized as a legitimate scientific tool by any professional organization. Because of this situation, much expert testimony about hypnotically elicited testimony was excluded by *Frye v. U.S.*, 293 F. 1013 (D.C. Cir, 1923), which held that scientific information must be "sufficiently established to have gained general acceptance in the particular field to which it belongs" (p. 1014).

However, opinions about hypnosis evolved as scientists increasingly accepted the idea that memory provides a comprehensive, accurate, permanent, and potentially accessible [1, 3] record of historical events. This "video recorder model" of memory supported the use of special recall enhancement techniques, including hypnosis, to access purportedly repressed, dissociated, or otherwise unavailable memories [3].

By 1968, hypnosis had gained enough acceptances that in *Harding v. State*, it was ruled that hypnotically elicited testimony was admissible because the hypnotist was a well-trained professional, and the procedures were neither leading nor suggestive. This ruling set the precedent for *per se* admissibility of hypnotic testimony, and began the heyday of forensic hypnosis, which lasted until it became apparent that many "recovered" memories were wildly inaccurate.

For example, in *State v. Mack*, 292 N.W.2nd 764 (Minn., 1980), a hypnotized person recalled eating pizza in a restaurant that did not serve pizza, seeing tattoos on someone who did not have tattoos, and being stabbed with scissors or a knife where there was no physical evidence that a weapon had been used. Such inaccuracies prompted researchers and triers of fact to reexamine their earlier conclusions. In *People v. Shirley*, 723 P. 2nd 1354 (Cal., 1982), the California Supreme Court adopted a *per se* inadmissible rule, and set into motion a tidal change of judicial opinion against *per se* admissibility [1]. Professional organizations such as the American Psychological Association [4, 5] and the American Medical Association [6] followed suit with statements that discouraged the use of hypnosis for memory recovery.

Nevertheless, at the present time, judicial opinions about hypnosis are mixed, and some individuals [7] have strongly defended the value of hypnosis in the forensic arena. Below we review the relevant empirical evidence for each of the three main stances that courts have taken regarding hypnotically elicited testimony: *per se* inadmissible, *per se* admissible, and admissibility determined on a case-by-case basis.

Per se Inadmissible

The majority of states that have ruled on the issue of the admissibility of hypnotically elicited testimony have applied the guidelines set forth in *Frye v. U.S.*, 293 F. 1013 (D.C. Cir, 1923) or *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579 (1993), and have adopted a *per se* inadmissible rule.[a] Some courts have gone as far as to equate hypnosis with tampering with or manufacturing evidence, and others have excluded all testimony from previously hypnotized people, even testimony regarding memories recalled before being hypnotized [8, 9].

Research provides support for *per se* inadmissibility. Many studies indicate that memory, in general, is reconstructive, incomplete, and imperfect. Numerous reviews [10–15] converge in the opinion that

hypnosis does not offer recall advantages beyond waking conditions, and can result in an increase in accurate memories that is often more than negated by an increase in inaccurate memories. Steblay and Bothwell [11], for example, reported that relative to waking conditions, hypnosis produces more recall errors, more intrusions of uncued errors, and higher levels of memories for false information relative to nonhypnotic methods.

False memories are associated with hypnotic responsiveness, such that highly suggestible subjects (15–20% of the population) tend to report more false memories than low hypnotizable persons, and highly suggestible individuals are particularly prone to memory errors in response to misleading information [11, 16]. However, even relatively nonsuggestible participants (lowest 15–20% of the population), report false memories [17]. Accordingly, the vast majority of people are vulnerable, to some extent, to hypnotically elicited false memories.

The confidence expressed by a witness is the single most important factor in persuading jurors that a witness correctly identified the criminal [18]. Therefore, judiciary parties have been very concerned about claims that hypnosis can increase confidence in inaccurate memories and have the potential to distort testimony. Significantly, more than two-third of the studies that have examined confidence in recollections have shown inflated confidence rates for memories as a product of hypnosis [19]. More specifically, 23 studies have shown that hypnosis either increases confidence relative to a nonhypnotic group, or participants confidently report hypnotic pseudomemories of events that they earlier denied occurred when they were not hypnotized. Nine published studies have found no difference in confidence expressed in hypnotic versus nonhypnotic situations. However, in five of these studies, hypnosis produced more errors or less-accurate information on some or all measures, and in all of the remaining studies, with one exception, there were no differences in memory accuracy across hypnotic and nonhypnotic conditions. The single study in which differences emerged had design flaws that likely accounted for the discrepant findings [19].

Only two studies have assessed confidence in recollections in terms of the breakdown of testimony among hypnotized *versus* nonhypnotized participants in a cross-examination situation. Spanos, Gwynn, Comer, Baltruweit, and deGroh [20] determined that hypnotizable individuals who viewed a videotape of a crime misattributed a substantial number of suggested characteristics to the offender and frequently misidentified a mug shot of the offender. Hypnotic and nonhypnotic interrogations with leading questions produced equivalent effects across the two conditions. During later cross-examination, the hypnotized and nonhypnotized participants were equally likely to "breakdown" during questioning, and disavow their earlier misattributions and misidentifications. However, of the eight participants who selected a mug shot during interrogation, a total of 50% maintained their selection during cross-examination. In contrast, only one of the control participants (no hypnosis, imagery, and no leading questions) selected the mug shot during interrogation, and that individual did not maintain the selection during cross-examination. Spanos did not report an analysis from this comparison, but it appears as if the rates of misidentification differed between hypnotic and control participants. The general finding that hypnotized participants did not differ from participants who received guided imagery and leading questions is not at all surprising in that both guided imagery and leading questions have been shown to produce false memories.

In a second study, Spanos, Quigley, Gwynn, Glatt, and Perlini [21] reported that hypnosis inflated participant confidence in mug shot identifications. Moreover, participants who underwent a hypnosis interrogation, and were then prepared for cross-examination (i.e., told to tell the truth but to be polite but firm and not let the prosecutor plant doubts in their mind), expressed higher certainty in their mug shot identifications than participants who were not prepared. In contrast, preparation did not influence nonhypnotized participants' certainty of identifications: prepared and nonprepared individuals responded equivalently.

All participants were subjected to cross-examinations of their earlier testimony. Hypnotized participants were neither more nor less likely to break down under cross-examination than nonhypnotized participants. However, across conditions, 37 out of the 78 participants evaluated (47%) did not break down in the face of both direct examination and cross-examination. In fact, a total of 73% of people who were prepared for cross-examination did not break down under cross-examination. In contrast, 86% of individuals who were not prepared by lawyers

to withstand cross-examination did break down, and reversed their earlier testimony.

In short, the combination of hypnosis and routine trial preparation to withstand cross-examination resulted in a high rate of certainty and resistance to changing mug shot identifications while testifying. Combined, research on confidence and memory accuracy provides substantial justification for *per se* inadmissibility rulings.

Per se Admissible

Despite research showing that there are perils associated with admitting hypnotically elicited recall, there are still four states that allow all hypnotically elicited testimony to be admitted.[b] Courts in these states view the issue of such testimony as one of credibility rather than admissibility. Instead of determining admissibility based on rules and case law concerning expert testimony and the admissibility of scientific evidence, they invoke evidentiary rules concerning witness competence and relevance [8]. For example, in *State v. Brown* 337 N.W. 2d 138, 151 (N.D. 1983), the court noted that if they were to rule out any testimony in which there were concerns about suggestibility or difficulty due to confidence at cross-examination, "we would not allow a lawyer to talk to his witness before trial, we would exclude most identification testimony, and relatives and friends of a party could be excluded as witnesses." The research reviewed above suggests that hypnosis increases the risk of inaccurate remembrances; however, the psychological principles that account for the cognitive errors that occur during hypnosis are universal.

Courts that have supported *per se* admissible rulings generally believe that an attack on credibility is the proper method of determining the value of hypnotically induced testimony. Given the available evidence, an expert witness should be able to establish doubts about the reliability of hypnotically elicited testimony; however, this might not always be the case because of the continued belief of the public in hypnotically elicited testimony and differences in expert opinions before the courts.

Case-by-Case Basis

Federal courts and the minority of states[c] determine the admissibility of hypnotically elicited recall on a case-by-case basis because courts in these jurisdictions have ruled that hypnotically enhanced testimony is reliable under certain conditions, or hypnosis is not unduly biasing [7, 8]. Experts who support this position can cite research showing the following:

1. Hypnosis increases the sheer volume of recall, and hypnotic recollections are not necessarily inaccurate. Accordingly, it can be argued that hypnosis has a place in criminal investigations, especially in situations where few leads are available, and when accurate recall can be potentially corroborated.
2. Hypnosis should not be "scapegoated". Many procedures, including leading questions, carry the risk of false memories that equals or surpasses that of hypnosis, and are not excluded from the bar [22, 23].
3. Limited research suggests that hypnosis does not necessarily preclude cross-examination. Moreover, the effects of hypnosis on confidence are variable.
4. Many false memory studies are conducted under sterile laboratory conditions, and hypnosis is reasonably reliable for enhancing memories that have been blocked due to emotional trauma [7].

However, experts whose opinions support *per se* inadmissibility can counter with the following claims:

1. The risk of false memories is a trade-off for any increase in memory accuracy. In many criminal investigations, hypnotically elicited inaccuracies in recollections are sufficiently biasing so as to outweigh advantages in securing "leads" that can be better obtained using nonhypnotic recall enhancement procedures. Although hypnosis has assisted investigations in a number of instances, it is unclear whether repeated recall attempts were responsible independent of hypnosis [24], or whether other recall techniques would have been equally, if not more, effective.
2. Many procedures are suggestive (e.g., misleading questions and guided imagery). Nevertheless, this fact provides no warrant for the use of hypnosis in criminal investigations, or for the admissibility of hypnotically elicited testimony.
3. The consilience of evidence indicates that hypnosis can inflate recall confidence in inaccurate memories, and that many individuals who are hypnotized do not reverse their testimony following cross-examination [25].

4. At least eight studies (see [19]) have compared hypnotic *versus* nonhypnotic memory in the face of relatively emotionally arousing stimuli (e.g., depictions of fatal stabbings, recollections of the death of Princess Diana, shop accidents and mutilations, and sexual intercourse). All studies have failed to confirm the hypothesis that hypnosis is especially effective in enhancing memory in emotionally arousing situations. An applied study with the Los Angeles Police Department concluded that hypnosis did not augment recognition memory for victims of violent crimes such as rapes and shootings [26].

Advocates of admitting hypnosis on a case-by-case basis are likely to contend that hypnosis, when conducted appropriately, without leading questions, and when other procedural safeguards are followed is no more biasing than routine interrogation.

The seminal case, *State v. Hurd*, 86 N.J. 525, 432 A.2d 86 (1981), delineated a set of "procedural safeguards" that must be followed in order for hypnotically elicited testimony to be admissible in New Jersey. Other states either require compliance with the Hurd rules or have developed their own set of guidelines [e.g., South Carolina in *State v. Chesseboro*, 552 S.E. 2d 300, 308 (2001)]. The Federal Bureau of Investigation (FBI) and the American Society of Clinical Hypnosis (ASCH) have also developed procedural guidelines, which range from making videotapes or other recordings of the procedures to evaluate suggestive influences, to discussing the imperfections of memory "in and out of hypnosis".

Unfortunately, no set of applied guidelines has been systematically evaluated. Previous guidelines probably do not go far enough in that they do not recommend assessing participants' confidence in their initial and posthypnotic recall. Without this information, it is impossible to fully evaluate the potentially biasing effect of hypnosis and hypnotic testimony on the witness. However, limited evidence regarding prehypnotic "warnings" about the imperfections of hypnotically elicited recall do not inspire confidence in their protective effect. In one study [27], warnings mitigated some of the memory distortions associated with hypnosis but did not improve recall above and beyond a nonhypnotic condition; in another study, warnings minimized memory distortions during but not after hypnosis [25]; and in a third study [28], repeated warnings that hypnotic and nonhypnotic memories are not necessarily accurate did not improve recall relative to a nonhypnotic condition. In short, the value of procedural guidelines has not been adequately investigated or well established.

Critics of guidelines argue that hypnotically elicited testimony is so unreliable that no array of procedural measures can outweigh the inherently biasing effect of hypnosis. As stated in *People v. Gonzales* [108 Mich. App. 145, 310 N.W. 2d 306 (1981), aff'd, 415 Mich. 615, 329 N.W. 2d 743 (1982)], the Hurd guidelines have the potential to confer an unwarranted "aura of reliability" on hypnotic testimony, thereby raising the specter of a miscarriage of justice. The Supreme Court of New Jersey found this and other arguments compelling, and recently reversed their earlier decision regarding the Hurd Guidelines. However, research on the biasing effects of hypnosis alone, absent suggestive questions, is mixed [22, 23].

Finally, advocates of a case-by-case position can argue that it is unduly restrictive to prohibit a person from testifying even about "nonhypnotic" recollections. This point may be valid. However, previous research has not examined the extent to which a previously hypnotized person's nonhypnotically refreshed memories are tainted, so the court can expect little evidence-based guidance on this issue.

Conclusion

Courts must contend with difficult issues that stretch the ability of triers of fact to fully comprehend. Accordingly, it is not surprising that many jurisdictions have opted for *per se* exclusion to avoid proverbial battles of experts, reduce costs and time associated with dealing with complex issues and discourages the use of a technique that can be misapplied or is unreliable. Indeed, a strong case can be made for *per se* inadmissibility of hypnotically augmented testimony. Nevertheless, it is also true that certain nonhypnotic recall enhancement and suggestive procedures may be equally or more prejudicial. Given that the majority of courts adopt a *per se* exclusion rule regarding hypnosis, it is appropriate to take steps to ensure that a variety of potentially biasing procedures are better understood and more carefully scrutinized in the judicial process.

End Notes

[a.] Alaska, Arizona, Arkansas, California, Connecticut, Delaware, Florida, Georgia, Hawaii, Illinois, Indiana, Iowa, Kansas, Maryland, Massachusetts, Michigan, Minnesota, Missouri, Nebraska, New Jersey, New York, North Carolina, Oklahoma, Utah, Virginia, Washington, and West Virginia.
[b.] North Dakota, Oregon, Tennessee, and Wyoming.
[c.] Alabama, Colorado, Idaho, Kentucky, Mississippi, New Mexico, Nevada, Ohio, Pennsylvania, South Carolina, South Dakota, Texas, and Wisconsin.

References

[1] Lynn, S.J., Neuschatz, J. & Fite, R. (2001). Hypnosis and memory: implications for the courtroom and psychotherapy, in *Memory, Suggestion, and the Forensic Interview*, M. Eisen & G. Goodman, eds, Guilford Press, New York.

[2] Perry, C., Orne, M.T., London, R.W. & Orne, E.C. (1996). Rethinking per se exclusions of hypnotically elicited recall as legal testimony, *International Journal of Clinical and Experimental Hypnosis* **44**, 66–81.

[3] Newman, A.W. & Thompson, W. (2001). The rise and fall of forensic hypnosis in criminal investigation, *Journal of the American Academy of Psychiatry and the Law* **29**, 75–84.

[4] American Psychological Association (1995). Division 17 Committee on Women, Division 42 Trauma and Gender Issues Committee. *Psychotherapy Guidelines for Working with Clients who May have an Abuse or Trauma History*.

[5] American Psychological Association (1996). *Working Group on Investigation of Memories of Childhood Abuse: Final report*.

[6] American Medical Association (1994). *Council on Scientific Affairs. Memories of Childhood Abuse*, CSA Report 5-A.

[7] Brown, D., Scheflin, A.W. & Hammond, D.C. (1998). *Memory, Trauma Treatment, and the Law*, WW Norton & Co.

[8] Webert, D.R. (2003). Are the courts in a trance? Approaches to the admissibility of hypnotically enhanced witness testimony in light of empirical evidence, *American Criminal Law Review* **40**, 1301.

[9] Martin, E.F. (2003). A daubert test of hypnotically refreshed testimony in the criminal courts, *Texas Wesleyan Law Review* **9**, 151.

[10] Erdelyi, M. (1994). Hypnotic hypermnesia: the empty set of hypermnesia, *International Journal of Clinical and Experimental Hypnosis* **42**, 379–390.

[11] Steblay, N.M. & Bothwell, R.K. (1994). Evidence for hypnotically refreshed testimony? The view from the laboratory, *Law and Human Behavior* **18**, 635–651.

[12] Lynn, S.J., Neuschatz, J., Fite, R. & Rhue, J.R. (2001). Hypnosis and memory: implications for the courtroom and psychotherapy, in *Memory, Suggestion, and the Forensic Interview*, M. Eisen & G. Goodman, eds, Guilford Press, New York.

[13] Kebble, M.R. & Wagstaff, G. (1997). Hypnotic interviewing: the best way to interview eyewitnesses?, *Behavioral Sciences and the Law* **16**, 115–129.

[14] Orne, E.C., Whitehouse, W.G., Dinges, D.F. & Orne, M.T. (1996). Memory liabilities associated with hypnosis: Does low hypnotizability confer immunity?, *International Journal of Clinical and Experimental Hypnosis* **44**, 354–369.

[15] Wagstaff, G. (2008). Hypnosis and the law: Examining the stereotypes, *Criminal Justice and Behavior* **35**, 1277–1294.

[16] Spanos, N.P. (1986). Hypnotic behavior: A social-psychological interpretation of amnesia, analgesia, and "trance logic", *Behavioral and Brain Sciences* **9**, 449–467.

[17] Lynn, S.J., Meyers, B. & Malinoski, P. (1997). Hypnosis, pseudomemories, & clinical guidelines: a sociocognitive perspective, in *Recollections of Trauma: Scientific Studies and Clinical Practice*, D. Read & S. Lindsay, eds, Plenum Press.

[18] Wells, G. & Bradfield, A.L. (1998). "Good, you identified the suspect:" Feedback to eyewitnesses distorts their reports of the witnessing experience, *Journal of Applied Psychology* **83**, 360–376.

[19] Lynn, S.J., Barnes, S.M. & Matthews, A. Hypnosis and memory: from Bernheim to the present, in K. Markman, W. Klein & J.A. Suhr, eds, Psychology Press, New York.

[20] Spanos, N.P., Gwynn, M.I., Comer, S.L., Baltruweit, W.J. & de Groh, M. (1989). Are hypnotically induced pseudomemories resistant to cross-examination? *Law and Human Behavior* **13**, 271–289.

[21] Spanos, N.P., Quigley, C.A., Gwynn, R.I., Glatt, R.L. & Perlini, A.H. (1991). Hypnotic interrogation, pretrial preparation, and witness testimony during direct and cross-examination, *Law and Human Behavior* **15**, 639–653.

[22] Scoboria, A., Mazzoni, G., Kirsch, I. & Milling, L.S. (2002). Immediate and persistent effects of misleading questions and hypnosis on memory reports, *Journal of Experimental Psychology: Applied* **8**, 26–32.

[23] Scoboria, A., Mazzoni, G. & Kirsch, I. (2006). Effects of misleading questions and hypnotic memory refreshment on memory reports: a signal detection analysis, *International Journal of Clinical and Experimental Hypnosis* **54**, 340–359.

[24] Payne, D.G. (1987). Hypermnesia and reminiscence in recall: historical and empirical review, *Psychological Bulletin* **101**, 5–27.

[25] Green, J.P., Lynn, S.J. & Malinoski, P. (1998). Hypnotic pseudomemories, prehypnotic warnings, and the malleability of suggested memories, *Applied Cognitive Psychology* **12**, 431–444.

[26] Sloane, M.C. (1981). *A Comparison of Hypnosis vs. Waking State and Visual vs. Non-visual Recall Instructions for Witness/victim Memory Retrieval in Actual*

Crimes, Doctoral dissertation, Florida State University. Dissertation Abstracts International; University Microfilms 81–25, 873.

[27] Burgess, C. & Kirsch, I. (1999). Expectancy information as a moderator of the effects of hypnosis on memory, *Contemporary Hypnosis* **16**, 22–31.

[28] Neuschatz, J., Lynn, S.J., Benoit, G. & Fite, R. (2003). Hypnosis and memory illusions: an investigation using the Deese/Roediger paradigm, *Imagination, Cognition, and Personality* **22**, 3–12.

STEVEN JAY LYNN, SEAN BARNES AND ABIGAIL MATTHEWS

Hypothetical Question

In adversary system jurisdictions, when a forensic expert testifies to facts that are within his personal knowledge, it is relatively simple to lay an evidentiary foundation. He acts just like an ordinary fact witness in stating the circumstances under which he became aware of facts and the observations he made. When an expert has done a laboratory examination of evidence and has arrived at an opinion, the witness will also be permitted to state that opinion if he has been qualified, preliminarily, to give opinion evidence as an expert.[a]

However, if the expert has no personal knowledge of the facts of the case, and has made no investigation of them, it is still possible to draw upon his scientific skill and expertise when called to court. He can be asked to assume certain facts disclosed to the court by other witnesses, and offer his opinion based upon the assumed facts. This has been a time-honored manner of eliciting expert testimony in common law jurisdictions.[b] The questioning process is referred to as eliciting *hypothetical question* opinion evidence. It is a method often used as a way of drawing upon prominent authorities in forensic fields involving science or medicine when these authorities lack personal knowledge of the case-specific facts.

The hypothetical question may be asked of expert witnesses during direct examination. On cross-examination the expert may thereafter be asked to assume supplemental facts and base his opinion taking into account these additional facts. Another hypothetical question approach on cross-examination simply supply the witness with a different hypothesis of the facts based on the cross-examiner's theory of the case.

It should be evident that there is a significant weakness in this artificial process of eliciting opinion evidence. In practice, while the witness must be familiarized, in the hypothetical question, with the undisputed relevant facts, these facts rarely portray the whole panorama of investigative evidence. Despite these misgivings, hypothetical question evidence has achieved a time-honored place in common law jurisdictions as a means of questioning an expert witness.

End Notes

[a.] This is the traditional way, in common law jurisdictions, in which expert opinion has been elicited.

[b.] In addition to common law principles, Federal Rule of Evidence 703, in the United States, also specifically permits an expert to offer opinions on facts of which the expert may not have had personal knowledge, but of which the expert was apprised either before trial or at the hearing or trial. The rule provides that if the facts of which the witness was apprised are "of a type reasonably relied upon by experts in the particular field in forming opinions or inferences upon the subject, the facts or data need not be admissible in evidence in order for the opinion or inference to be admitted. . . ." In that fashion, medical doctors are permitted to answer questions based on reports of X-ray technicians, nurses, other medical personnel, and hospital records.

Related Articles

Adversary Systems of Justice

Cross-Examination of Experts

Direct Examination of Experts

Expert Opinion: United States

Foundation Testimony

ANDRE MOENSSENS

Identification: Photographic *see* Visual Recognition Systems in Identification

Identification: Sex of Deceased *see* Sex Determination of Remains

Identification: Video *see* Visual Recognition Systems in Identification

Identification and Individualization

Identity, Identification, and Individualization

Prof Paul Kirk [1] introduced these terms as they relate to forensic science in a very elegant manner:

> *Identity* is defined by all philosophical authorities as uniqueness. A thing can be identical to only with itself, never with any other object, since all objects in the universe are unique. [...] Bowing to general scientific usage, we must however accept the term *identification* in a broader context, referring only to placing the object in a restricted class. [...] The real aim of all forensic science is to establish individuality, or to approach it as closely as the present state of science allows. *Criminalistics is the science of individualisation*.

An individualization can be viewed as a special case of identification, where the restricted class is populated by one object only.

Definitions of individualization in the forensic literature (e.g., fingerprint, footwear marks, or toolmarks) systematically refer to the capability of pointing to the right source *to the exclusion of all others* (objects or persons). Hence, by default, the size of the population of relevant sources considered at the outset of the examination is systematically set to its maximum, regardless of the specific circumstances of the case. We call this the *Earth population* paradigm. In that paradigm, the individualization conclusion cannot be reached in a deductive manner, but is *de facto* probabilistic in nature [2–4]. This is well spelled out by Tuthill and George [5]:

> The individualisation of an impression is established by finding agreement of corresponding individual characteristics of such number and significance as to preclude the possibility (or probability) of their having occurred by mere coincidence, and establishing that there are no differences that cannot be accounted for.

At the end of the examination process, the quantity of features observed in agreement between two objects (without discrepancies) is such that the examiner is prepared to rule out the possibility of a coincidental match, whatever the initial population of sources involved. That has been described by Stoney as a "leap of faith" [6].

If we ask ourselves the order of magnitude of the probability of adventitious matches in the mind of an examiner at the moment of its decision, we can derive it simply using the odds form of the Bayes theorem (*see* **Evidence Interpretation: a Logical Approach**, another formal analysis can also be found elsewhere [7]):

$$\frac{\Pr(H_p|E,I)}{\Pr(H_d|E,I)} = \mathrm{LR} \times \frac{\Pr(H_p|I)}{\Pr(H_d|I)} \qquad (1)$$

where E stands for the evidence, e.g., a correspondence between a mark and a print, H_p stands for the proposition that the mark has been left by the same source as the print, H_d stands for the proposition that the mark has been left by another source (here $H_d = \overline{H_p}$), and I stands for the background information; here, the adoption of the Earth population paradigm dominates all other information.

If the prior probability of H_p is set as the inverse of the size of the Earth population of sources (e.g., 1 in 7 billion), then the likelihood ratio (LR) required to achieve a posterior probability for H_p that is close

to 1 (e.g., 0.9999) can be derived from the following:

$$\frac{0.9999}{1-0.9999} = \mathrm{LR} \times \frac{1/(7\times 10^9)}{1-1/(7\times 10^9)} \qquad (2)$$

This can be re-arranged to give:

$$\mathrm{LR} = \frac{0.9999}{1-0.9999} \times \frac{1-1/(7\times 10^9)}{1/(7\times 10^9)} \qquad (3)$$

For a posterior probability of 0.9999, the LR should at least be above 7×10^{13} or a match probability (MP) (if we assume for the sake of argument that $\mathrm{LR} = 1/MP$) below 1.43×10^{-14}.

Leaving aside the consideration that the assignment of a given posterior probability is outside the duty of the scientist, we face, in the Earth population paradigm, the prospect of having to articulate probabilities out of reach of the current systematic research. Taking fingerprints as an example, the recent published statistical research [8] would allow us to quote MPs on the order of one in a billion confidently. Articulating any smaller number (down to the probability of zero) is nothing more than an unsupported leap of faith. In the above context, we agree with Saks and Koehler [9] that: "The concept of 'individualisation', which lies at the core of numerous forensic science subfields, exists only in a metaphysical or rhetorical sense."

DNA Evidence and Individualization

Current DNA short tandem repeat (STR) profiling techniques are typically based on 10–15 loci, so that the MPs for individuals unrelated to the defendant are extremely small, often less than one in a billion (*see* **Short Tandem Repeats: Interpretation**). This is enough for some commentators to suggest that the profile can be considered as unique "with reasonable scientific certainty" [10]. Apart from the technical issues for computing such posterior probabilities [11, 12], there are four arguments that call for a more humble approach here: (i) the threshold (e.g., above 99% as proposed by the FBI in [10]) amounts to quantifying the concept of reasonable doubt and is outside the realm of the scientist; (ii) invoking a very large population (such as 260 million for the FBI policy) or the population of the Earth is hardly meaningful in a forensic context; (iii) the levels of relatedness within this population with regards to the defendant are generally not available; and, (iv) more importantly, Foreman & Evett [13] argue that an order of magnitude of 1 in a billion is a fair and reasonable approach, given the current state of knowledge of dependence effects. Hence, we should discourage anyone from believing that a matching DNA profile amounts to definitive attribution of sources – committing the *uniqueness* fallacy according to Balding [14] – but to encourage the interpretation of DNA evidence within the specific context of the case. The fact that absolute certainty is not achievable does not prevent the findings from being useful in court.

Uniqueness

Considerable confusion exists among laymen, indeed amongst forensic scientists about the use of a word such as *unique* also. The phrase "all portions of friction ridge skin are unique" is no more than a statement of the obvious – every entity is unique and identical to itself only.

What matters is how two objects left in forensic contexts (as an inked print and as a latent mark for example) can be distinguished from each other; either in the case where they have come from the same source or in the case where they have come from different sources. This distinguishability or variability of prints/marks not only depends crucially on the examination method but also on the intrinsic qualities of the prints/marks to display selective features (extensiveness, clarity, etc.). One striking feature in most forensic identification fields is the constant confusion between the between-*persons/objects variability* (which may be infinite) with the *variability* that is expressed in *marks* left by these persons or objects. The loss of information from a complete three-dimensional organ or object to a two-dimensional mark detected on a surface needs to be investigated specifically. The crux of the matter is not really the uniqueness of the sources involved, but the ability of an examiner to distinguish reliably marks and prints left by these sources.

Hence, invoking statements such as "Nature never repeats itself" [15] while referring to the uniqueness of the organs under examination and by consequences to the strength of the conclusion that can be derived from their examination is not legitimate. Cole named this tendency the *fingerprint examiner's fallacy* [16].

Reporting Identification Evidence in Court

The above analysis suggests that scientists should refrain from reporting categorical opinions of individualization, but should report the strength of the findings and leave the integration of that element into the overall context of the case to the jury.

For fingerprint evidence, a systematic analysis of the standards for conclusions (either in terms of criteria for sufficiency or in terms of a range of authorized conclusions by the profession) shows how the evaluation phase has been dominated by policy and pragmatic decisions instead of the application of a scientific approach [17]. The practice of reporting only individualizations and exclusions, leaving all other corroborative evidence under the heading of "inconclusive", cannot be justified on scientific grounds.

Other identification fields have suggested the use of scales of conclusion that are focused on posterior probabilities by either following the Earth population paradigm [18–22] or by applying the Bayes theorem incorporating a highly debatable 0.5 prior probability [23, 24]. All these proposals fall short of providing a logical framework for the reporting task. For a discussion on the scales of conclusions, the reader can refer to the following papers [25–27].

All identification evidence types (DNA included) should be governed by the same principles when it comes to their interpretation.

There is an urgent need to find a common, coherent, and harmonized framework to report such evidence in court. The one proposed here is not new and is already embraced in other forensic fields. Broadly speaking, in all forensic areas seeking to help address issues of identity of sources, there are three inter-related factors that determine the nature of the inference that may be made: the quantity and quality of relevant detail, within-source variation, and between-source variation. These three factors impact on the LR defined as follows:

$$\mathrm{LR} = \frac{\Pr(E|H_p, I)}{\Pr(E|H_d, I)} \qquad (4)$$

The weight of forensic findings is essentially a relative and conditional measure that helps move a case in one direction or the other, depending on the magnitude of the LR. The numerator of the LR considers the within-source variability of the features defined by E and the denominator of the LR deals with their between-sources variability.

Reporting identification evidence in court could be viewed as simply assigning and reporting an LR without the need to claim individualization and set any operational threshold to reach it. All items of evidence for which the LR differs from one then becomes relevant [28].

It is a matter of principle that the scientist may not express an opinion with regard to the hypotheses themselves – that is a matter for the jury. That principle no doubt influenced the English Court of Appeal in the judgment in *R.v. Doheny. G. Adams* [29] which states "The scientist should not be asked his opinion on the likelihood that it was the Defendant who left the crime stain [. . .]."

Conclusions that use terms such as *very likely* or *almost certain* in relation to a proposition are only logically possible when the nonscientific evidence is taken into account (*see* **Evidence Interpretation: a Logical Approach**). Then, the statement follows from the combined effect of the scientific observation and an assessment of the prior probability that encapsulates all the evidence available to the court. The inferential process used here goes far beyond forensic science. Moreover, most of the time, the forensic scientist is not aware of the other pieces of evidence available in a case, and clearly does not have the information at hand to make inferences such as these. Forensic scientists should not make any assessment of the prior information because that ought to remain a matter for the court. To avoid such situations, we need to adopt a reporting convention that allows the scientific statement to remain consistent within a given framework irrespective of other evidence at hand. The LR provides the means for achieving this objective.

References

[1] Kirk, P.L. (1963). The ontogeny of criminalistics, *The Journal of Criminal Law, Criminology, and Police Science* **54**, 235–238.

[2] Meuwly, D. (2007). Forensic individualisation from biometric data, *Science & Justice* **46**(4), 205–213.

[3] Kwan, Q.Y. (1977). *Inference of Identity of Source [D. Crim. Dissertation]*, University of California, Berkeley.

[4] Champod, C. & Evett, I.W. (2001). A probabilistic approach to fingerprint evidence, *Journal of Forensic Identification* **51**(2), 101–122.

[5] Tuthill, H. & George, G. (2002). *Individualization – Principles and Procedures in Criminalistics*, 2nd Edition, Lightning Powder Company, Jacksonville.

[6] Stoney, D.A. (1991). What made us ever think we could individualize using statistics, *Journal of the Forensic Science Society* **31**(2), 197–199.
[7] Champod, C. (2000). Identification/individualization: overview and meaning of ID, in *Encyclopedia of Forensic Sciences*, J. Siegel, P. Saukko & G. Knupfer, eds, Academic Press, London, pp. 1077–1083.
[8] Neumann, C., Champod, C., Puch-Solis, R., Egli, N., Anthonioz, A. & Bromage-Griffiths, A. (2007). Computation of likelihood ratios in fingerprint identification for configurations of any number of minutiae, *Journal of Forensic Sciences* **52**(1), 54–64.
[9] Saks, M.J. & Koehler, J.J. (2008). The individualization fallacy in forensic science evidence, *Vanderbilt Law Review* **61**, 199–219.
[10] Budowle, B., Chakraborty, R., Carmody, G. & Monson, K.L. (2000). Source attribution of a forensic DNA profile, *Forensic Science Communications* **2**(3), http://www.fbi.gov/hq/lab/fsc/backissu/july2000/source.htm.
[11] Weir, B. (2001). DNA match and profile probabilities: comment on Budowle *et al.* (2000) and Fung and Hu (2000), *Forensic Science Communications* **3**(1), http://www.fbi.gov/programs/lab/fsc/current/weir.htm.
[12] Balding, D.J. (1999). When can a DNA profile be regarded as unique, *Science and Justice* **39**(4), 257–260.
[13] Foreman, L.A. & Evett, I.W. (2001). Statistical Analysis to support forensic interpretation of a new ten-locus STR profiling system, *International Journal of Legal Medicine* **114**, 147–155.
[14] Balding, D.J. (2005). *Weight of Evidence for Forensic DNA Profiles*, John Wiley & Sons, Chichester.
[15] McRoberts, A.L. (1996). Nature never repeats itself, *The Print* **12**(5), 1–3. http://www.iinet.com/market/scafo/library/120501.html.
[16] Cole, S.A. (2004). Grandfathering evidence: fingerprint admissibility rulings from *Jennings* to *Llera Plaza* and back again, *The American Criminal Law Review* **41**(3), 1189–1276.
[17] Champod, C. (2000). Fingerprints (dactyloscopy): standard of proof, in *Encyclopedia of Forensic Sciences*, J. Siegel, P. Saukko & G. Knupfer, eds, Academic Press, London, pp. 884–890.
[18] AFTE Criteria for Identification Committee (1992). Theory of identification, range of striae comparison reports and modified glossary of definitions, *AFTE Journal* **24**(2), 336–340.
[19] SWGTREAD (2006). Standard terminology for expressing conclusions of forensic footwear and tire impression examinations, *Journal of Forensic Identification* **56**(5), 806–808.
[20] Broeders, A.P.A. (1999). Some observations on the use of probability scales in forensic identification, *Forensic Linguistics* **6**(2), 228–241.
[21] Totty, R.N. (1991). Recent developments in handwriting examination, *Forensic Science Progress* **5**, 91–128.
[22] ASTM (2004). *Standard Terminology for Expressing Conclusions of Forensic Document Examiners*, Standard E 1658-04, West Conshohocken.
[23] ENFSI Expert Working Group Marks Conclusion Scale Committee (2006). Conclusion scale for shoeprint and toolmarks examination, *Journal of Forensic Identification* **56**(2), 255–280.
[24] Deinet, W. & Katterwe, H. (2007). Comments on the application of theoretical probability models including Bayes theorem in forensic science relating to firearm and tool marks, *AFTE Journal* **39**(1), 4–7.
[25] Champod, C. & Evett, I.W. (2000). Commentary on: Broeders, A.P.A. (1999) some observations on the use of probability scales in forensic identification, *Forensic Linguistics* **6**(2), 228–241; *Forensic Linguistics* **7**(2), 238–243.
[26] Champod, C., Evett, I.W., Jackson, G. & Birkett, J. (2000). Comments on the scale of conclusions proposed by the ad hoc committee of the ENFSI marks working group, *The Information Bulletin for Shoeprint/Toolmark Examiners* **6**(3), 11–18.
[27] Biedermann, A., Taroni, F. & Garbolino, P. (2007). Equal prior probabilities: can one do any better? *Forensic Science International* **172**(2–3), 85–93.
[28] Lempert, R.O. (1977). Modeling relevance, *Michigan Law Review* **75**, 1021–1057.
[29] *R. v Alan James Doheny, R. v Gary Adams Court of Appeal - Criminal Division* [1996]. E.W.C.A. Crim 728, (31st July, 1996).

CHRISTOPHE CHAMPOD

Identification of Human Remains

The previous articles have focused on the most important steps in achieving a biological profile from the skeleton, i.e., aging, sexing, determining stature, and ancestry, recording every single detail of the remains that may provide a thorough “identikit” of the person. This procedure can be completed with facial reconstruction from the cranium in cases of badly decomposed remains. In this way, such a complete identikit can be given to the investigating authorities and the media (newspapers and television) in order to achieve a suspicion of identity, in case somebody “recognizes” the biological profile. In the better organized countries, this kind of information

can be inserted into a postmortem database and will then be cross-matched with an "antemortem" database containing data from missing persons [1–4], in order to reach a suspicion of identity or at least possible matches for those specific human remains. At times, circumstantial evidence can help, in the sense that when the remains arrive the pathologist already has knowledge of a suspected identity (e.g., a burnt body in a car whose owner is known).

Whatever the case may be, the forensic anthropologist or pathologist is faced with the need to compare antemortem with postmortem data, in order to finalize the identification process. The most popular disciplines for positive identification are genetics and odontology. However, sometimes it is not possible to perform DNA or dental matching because of the lack of antemortem data (or relatives, in the case of DNA) or because of the bad state of preservation (e.g., dry bone, charred bone, and missing teeth). In these cases, bone morphology may be useful.

Generally speaking, morphological methods (forensic anthropology and odontology techniques), which simply compare shape, size, and/or peculiar anomalies of bone and teeth, are valid methods. They are advantageous because they are quicker to perform and much cheaper when compared with the costs of genetic reagents; however, they may suffer, in the view of some experts and judges, because of the qualitative (and not quantitative) responses they give. Genetic and fingerprinting methods give a quantitative result, because the specific traits are analyzed and compared and their relative frequencies among the population quantified using statistical methods. This allows one to answer in a quantitative manner on the question of probability of two individuals having similar characteristics. In the case of "morphological" methods (such as the odontological and anthropological ones), a final judgment can be given in terms of exclusion, certainty (very few times), or in less-defined terms of probability, possibility, consistency, etc. In other words, it is impossible to quantify the chances of two individuals sharing the exact same dental or osseous asset. Furthermore, a unanimous and clear agreement on the quality and quantity of the characters needed to achieve personal identification does not exist.

The recurrence of discordant characters settles the case, by excluding the identity; some or many concordant characters, particularly if not uncommon within the population, can permit one to express a judgment of compatibility or possibility; few characters or a combination of characters, rare among the population, will allow one to express a judgment of high probability or certainty [5]. Though it may at times seem a semantic matter, the importance of how the expert anthropologist may express him or herself in a report or in court is crucial. In certain parts of the anthropologist's work (such as sexing, aging, and estimation of stature.), a standard error, a confidence interval, or any instrument for expressing the error of a method may be given. As far as "personal identification", i.e., the significance of shared osseous traits, is concerned this is not the case for the reasons stated above. And the difficulties met (along with the limits one should admit) have recently been dealt with by forensic anthropologists which have started paying attention to the effects of court rulings (e.g., Daubert and Kumho) on the significance of anthropological testimony [5, 6].

It is therefore reasonable to admit that the aim of morphological techniques is more frequently that of excluding, reducing the candidates among a population to one (hence the use of the term individualization), more rarely, in the case of anthropology, establishing identity with certainty. The aim of this article is therefore to illustrate how anthropology may help identify (forensic genetics and odontology are treated in separate articles) the rationale behind this assumption, being that each bone of every single individual has its own peculiar shape and dimension. In order to identify a person, it is therefore possible to compare the shape of the bones retrieved with those on an antemortem X ray of a missing person, just as though each bone were a fingerprint [7–47].

Before speaking of bones, however, a brief comment should be made concerning fingerprints. One should keep in mind that the anthropologist may have to deal with remains with some soft tissue still present (Figure 1) and should know that the fingerprints can and must be preserved.

Qualified personnel exists for fingerprint analysis, however, pathologists and anthropologists frequently deal with very decomposed human remains and need to be aware of the fact that within an apparently useless piece of soft tissue lying over the phalanges of a skeleton there may still be a partial print. It is therefore useful for the anthropologist to give priority to such potential prints and aid in the recovery. Once fingerprints are taken, if the person has a criminal record, the automated fingerprint



Figure 1 Right hand of an almost completely skeletonized body with conservation of soft tissue on fingers and palms. It was still possible to "extract" fingerprints from these tissues

identification system (AFIS) system will find the appropriate identity. In some countries, it is in fact the pathologist or anthropologist who tries to enhance fingerprints on the cadaver, or at least guide the investigating authorities in doing so. Thus, the following lines provide very general indications for the treatment of decomposed fingers and the recovery of prints [5–7].

Bodies that have reached the stage of quasi-skeletonization because of putrefaction processes or postmortem factors (such as fire) may still show fingertip and hand soft tissue residues. Depending on the case, the papillary crests can present deformations consisting of folds of the epidermal or dermal layer (as in mummification); they can be flattened/thinned because of erosive effects of environmental phenomena, or dehydrated, and therefore retracted and fragile, as in the case of exposure to high temperatures and flames. In these cases specific methods should be used in order to enhance the papillary design. Different techniques have been developed for this purpose: chemical techniques (methanol solutions, sodium hydroxide, and ethylene diammine tetracetic acid (EDTA)) can be applied for the purpose of softening and rehydrating the skin. Once the finger is softened, subcutaneous injections of glycerin or saline solution can be performed to reinflate the finger. Chemical techniques are, however, potentially destructive, complex, and require constant monitoring. The choice of which technique to use depends mainly in the kind of degradation process the fingertips have been through. For example, if the body has been soaked in water for a long time or is in the wet stages of decomposition, this leads to peeling off of the skin with a consequent progressive loss of the papillary design and the first step should be to block the putrefactive process by hardening the skin via immersion in ethanol for a time ranging from a few minutes to 1–2 h. Dehydration due to the immersion in alcohol, though necessary, introduces serious problems, such as the thinning of the papillary crests, which are already compromised. At this point, however, it should be sufficient, once the epidermal glove, i.e., the peeled off skin, has been worn by the examiner, to ink the fingertip and transfer the fingerprint on paper.

Mummification and carbonization, on the other hand, provide a very dry substrate. The fingertips show more or less pronounced folding and hardening, which prevents the application of the normal techniques of fingerprinting. A technique frequently used in these cases consists in softening and reinflating the finger. This is obtained by carrying out alternated incubation in 90% methylalcohol and 5% sodium hydroxide solutions. However, the sodium hydroxide treatment is destructive for the skin and it is not advisable to repeat it. Once the fingertip is softened, it is possible to "reinflate" it by means of saline injections. When the folds have flattened and the papillary design becomes sufficiently readable, it is important to try to "ink" it or, if this is impossible, to take a photograph with tangential lighting. However, both for mummified fingers and burnt bodies a new versatile technique can be used. This involves applying a thin film latex (after cleaning the tip with ether) on the finger and obtaining an enduring negative copy of the print, which can then be inked and photographed [6]. The fingerprint obtained from the latex film will have to be inverted in place and color before fingerprint comparative analysis.

Regardless of the case, the anthropologist should always be in touch with the fingerprint expert who carries out the individualization, for the purpose of explaining what he or she is looking at.

Going back to bones, it is clear that the comparison, as always, must be carried out between antemortem and postmortem radiographs of different skeletal districts. Consequently, if radiographs of the various osseous regions (e.g., head, thorax, limbs, and abdomen.) are available, it will be possible to compare the morphology of the skeletal elements visible

in the radiograph with the same osseous elements belonging to the human remains (both working on the actual osseous elements, and on their radiographs) [8–49].

Exceptional morphological peculiarities (osteophytes, bone calluses, prosthetic or surgical devices, etc.) will make identification relatively simple; however, most of the time one has to deal with general bone morphology, for example, the observation of a transverse process, which may be particularly rounded, a clavicle with an unusual foramen, or a humeral head with a specific shape: given that an antemortem and a postmortem comparison yield similar characteristics, how many morphological elements are sufficient for individualization? There is no answer to this specific question; in other words, a minimum number of points does not exist – as it exists instead for fingerprints – in order to carry out an individualization. Some authors claim that one to four "significant" characters, without evident discrepancies, are considered sufficient for an individualization. The problem is that there is no clear description of what exactly is meant by a significant character: they must be visible in the same orientation as in the antemortem X ray, for example, and must be scarcely modifiable in time. However, the peculiarity of the characters and the consistency between the antemortem and postmortem shape is left, in part, to a subjective evaluation and to the experience of the operator.

Personal identification has been described in the literature for almost all the skeletal regions: frontal sinuses, cranial sutures, vertebrae, hip bones, osseous traits of the hands and feet, and even trabecular bone pattern (this last one only on high-quality hand and wrist X rays) [8–49]. In most instances, the examination is performed by a simple morphological comparison, at times with image superimposition, either of the bones with the antemortem X ray, or of the antemortem X ray with a postmortem X ray of the same skeletal region.

For some districts, the search for an "identification algorithm" has been attempted. The frontal sinus is perhaps the most studied one [12–17]. Apparently, the shape (Figure 2) of the frontal sinuses is so peculiar that it is even different in homozygous twins (like fingerprints). Different methods of classification of frontal sinus shape exist. One method (Yoshino's) is based on the attribution of a score for each of the following parameters: asymmetry, left sinus larger than the right one or vice versa, presence and number of indentations (and therefore arches) per septum, partial septa, presence of supraorbital cells, and so on. The entire procedure leads to a score, which is then translated in an "identification index". For the other skeletal districts such precise indications do not exist.

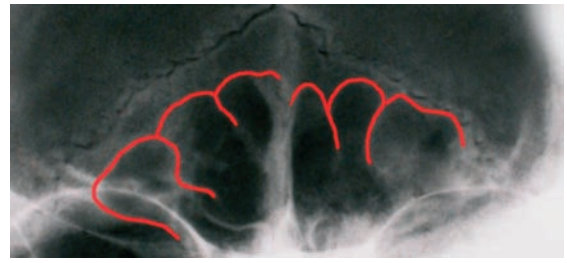


Figure 2 The red line delineates the outline of the frontal sinus in this X ray of the head. Comparison of the morphology of the antemortem and postmortem X ray can lead to positive individualization

Nonetheless, it is still possible to compare the forms of such other districts (Figure 3), such as the vertebrae and articular extremities. Of these, the transverse processes and the spinal process, which present very different shapes from one another, for example, are particularly useful. But there are no algorithms yet that allow a quantitative or semiquantitative estimate of the comparison. The answer is left to the morphological evaluation of the observer.

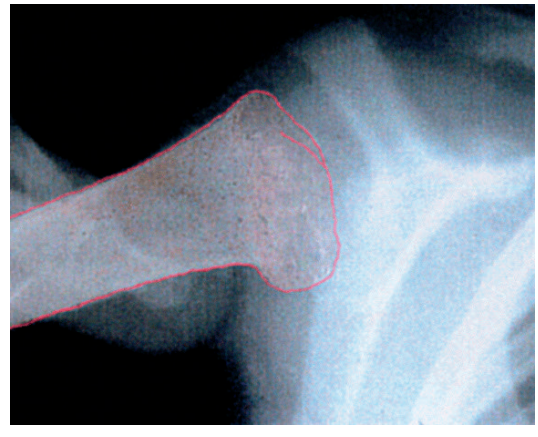


Figure 3 The radiograph represents a detail (humerus) of the antemortem thoracic radiograph of a skeletonized murder victim (a subadult). The thin line is the profile, once a similar orientation has been achieved, of the humerus from the skeleton

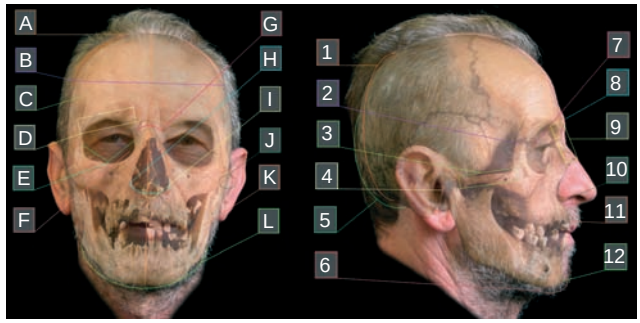


Figure 4 After having selected images of the missing person, proper spatial orientation of the skull and enlargement of the picture (trial and error) are achieved. Then superimpositions are usually performed in the frontal and lateral views. The figure shows an example of craniofacial superimposition in the frontal and lateral position. The letters (frontal) and the numbers (lateral) indicate the landmark for which the correspondence between face and skull is evaluated [36]. The observed criteria for the comparison in norma frontalis, marked with letters A to N are the following:

A: The length of the skull from bregma to mention must be included in the face. The bregma is generally covered with hair.
B: The width of the skull must match the forehead area.
C: The temporal line, if visible on the face, should correspond to the temporal cranial line.
D: The eyebrows generally follow the superior edge of the orbit to the medial and central third; they continue superiorly to the lateral third while the edge of the orbit deviates inferiorly.
E: The orbits contain the eye entirely.;
F: The lachrymal groove, if distinguishable on the photograph, lines up with the osseous groove.
G: The width of the nasal bridge must correspond in the two images.
H: The width of the nasal aperture falls within the external margins of the nose.
I: The nasal spine is situated above the inferior edge of the medial nasal crus.
J: The external auditory meatus is medial to the traghus. This can be adequately evaluated by a marker inserted in the ear.
K: The oblique line of the mandible, if visible on the face, corresponds to the same line on the skull.
L: The mandibular curve is similar to the curve visible on the face.

In norma lateralis:
1: The skullcap must coincide with the height of the head.
2: Sometimes it is possible to notice the margin of the frontal process of the zygomatic bone which should match the one on the cranium.
3: The margin of the zygomatic arch of the skull is also superimposable to that of the face.
4: The portion is slightly behind the traghus, and below the helix crus.
5: The occipital curve is placed inside the margin of the nape.
6: The anterior protuberance of the mandible is behind the chin. The chin form corresponds to the mandible form.
7: The lateral margin of the eye is situated within the orbit.
8: The profile of the glabella both of the skull and of the face must be similar.
9: The glabella, the nasal bridge, and the region of the nasal bones are the most significant. The prominence of the glabella and the depth of the nasal bridge closely follow the contour of the thin layer of overlying skin. The nasal bones fall within the margins of the nose.
10: The front nasal spine is situated posteriorly with respect to the base of the nose, close to the more posterior portion of the lateral septum cartilage.
11: The prosthion is posterior to the anterior margin of the superior lip.
12: The pogonion is posterior to the indentation noticeable in the chin where the orbicularis oris crosses the chin muscle

Since the antemortem material consists of radiological material, one should always be extremely careful in checking for proper orientation of the postmortem image, so that the two images can be comparable.

When there are no antemortem X rays with which to compare bone morphology, the last method the anthropologist can resort to is craniofacial superimposition (Figure 4). It consists of the

superimposition of a photograph of the skull onto a photograph, similarly oriented, of the face of the living person [36–43]. The correspondence between two subjects is evaluated on the basis of the correspondence of several anatomical landmarks that can be found on the skull and on the face. Although this procedure can be extremely useful and reliable for the dental area (in dental superimposition the same elements – teeth antemortem and teeth postmortem – are superimposed), it has far less credibility as a precise and accurate method when the soft tissues of the face are compared and matched to landmarks on the cranial skeleton. Figure 4 illustrates a general example that summarizes the method and whose aim is to demonstrate the difficulty in matching specific landmarks.

The supporters of this method report a reliability of 96%, whenever the possibility exists of both the frontal and lateral superimposition, but the discrepancy among numerous anthropologists concerning the degree of certitude supplied by this type of investigation is large.

Some authors have also tried to ameliorate the method by using photos in the oblique position as well, and checking the match of specific landmarks on three projections (frontal, lateral, and oblique). Others have even automated the system, and measured semiautomatically anatomical consistency between the digitized skull and face, as well as the distance between the landmarks and the thickness of soft tissues on the anthropometrical points; evaluation of matches has even been attempted by the use of polynomial functions and Fourier harmonic analysis for the forehead and mandibular line areas.

Regardless of the elaborate methodology one can apply, in our opinion, this method should never be used alone for individualization because there are several problems which are almost impossible to overcome in the comparison between a structure with soft tissues (face) and a structure made of hard tissue (skull). The method should be used only for excluding identity if gross incompatibilities are present or establishing consistent traits, which, however, cannot lead to a judgment of certainty.

In conclusion, results always need to be carefully examined by an experienced observer. Personal individualization has to be carried out with a set of data, after having carefully evaluated the limits and the possible sources of error of each method.

One last mention should be made concerning the role of the forensic anthropologist in mass disasters [44–49]. Once again, usually forensic pathologists, odontologists, and geneticists are the main figures called in for mass disasters. However, forensic anthropology is increasingly proving to be an indispensable discipline on these occasions, particularly where mass graves (especially in the case of war crimes) are involved, where extremely fragmented or charred remains need to be identified, or in the case of open mass disasters, where the biological profile of each victim may be crucial. The presence of the anthropologist is essential from the very beginning, as has been previously mentioned: disasters such as that of the Mont Blanc tunnel, for example, have proven how important recognition of minute charred remains is at the scene of the disaster. However, the role of the anthropologist can be fundamental also in individualization. Recent articles [48, 49] comment on the role of anthropologists in the identification of victims of war crimes in mass graves, when virtually no medical or dental antemortem data is available and where genetic antemortem samples may be difficult to obtain. Once again, most authors comment that great caution is advised and that frequently individualization is achieved through the combination of history, personal effects, biological profile (sex, age, stature, etc.) and, at times, peculiar pathological traits.

References

[1] Hanzlick, R. & Clark, S. (2008). The unidentified decedent reporting system – a model mational website registry for the unidentified deceased, *The American Journal of Forensic Medicine and Pathology* **29**, 106–113.

[2] Hanzlick, R. & Smith, G.P. (2006). Identification of unidentified deceased, *The American Journal of Forensic Medicine and Pathology* **27**, 79–84.

[3] Paulozzi, L.J., Cox, C.S., Williams, D.D. & Nolte, K.B. (2008). John and Jane Doe: the epidemiology of unidentified decedents, *Journal of Forensic Sciences* **53**(4), 1–6.

[4] Cattaneo, C., Ritz-Timme, S., Schutz, H.W., Collins, M., Waite, E., Boormann, H., Grandi, M.A. & Kaatsch, H.J. (2000). Unidentified cadavers and human remains in the UE: an unknown issue, *International Journal of Legal Medicine* **113**(3), 2–3.

[5] Kahana, Y., Grande, A., Tancredi, D., Penalver, J. & Hiss, J. (2001). Fingerprinting the deceased: traditional

and new techniques, *Journal of Forensic Sciences* **46**, 908–912.

[6] Porta, D., Maldarella, M., Grandi, M. & Cattaneo, C. (2007). A new method of reproduction of fingerprints from corpses in a bad state of preservation, *Journal of Forensic Sciences* **52**(6), 1319–1321.

[7] Cattaneo, C., De Angelis, D., Porta, D. & Grandi, M. (2006). Personal identification of cadavers and human remains, in *Forensic Anthropology and Medicine*, A. Schmitt, E. Cunha & J Pinheiro, eds, Human Press, pp. 359–379.

[8] Angyal, M. & Derczy, K. (1998). Personal identification on the basis of antemortem and postmortem radiographs, *Journal of Forensic Sciences* **43**(5), 1089–1093.

[9] Brogdon, B.G. (1998). Radiological identification of individual remains, in B.G. Brogdon, ed, *Forensic Radiology*, CRC Press.

[10] Christensen, A.M. (2004). The impact of Daubert: implications for testimony and research in forensic anthropology and the use of frontal sinuses in personal identification, *Journal of Forensic Sciences* **49**(3), 1–4.

[11] Grivas, C.R. & Kumar, D.A. (2008). Kumho Daubert and the nature of scientific inquiry: implications for forensic anthropology, *Journal of Forensic Sciences* **53**(4), 771–776.

[12] Yoshino, M., Miyasaka, S., Sato, H. & Seta, S. (1987). Classification system of frontal sinus patterns by radiography: its application to identification of unknown skeletal remains, *Forensic Science International* **34**(4), 289–299.

[13] Kirk, N.J., Wood, R.E. & Goldstein, M. (2002). Skeletal identification using the frontal sinus region: a retrospective study of 39 cases, *Journal of Forensic Sciences* **47**(2), 318–323.

[14] Yoshino, M., Miyasaka, S., Sato, H. & Seta, S. (1998). Classification system of frontal sinus patterns, *Canadian Society of Forensic Sciences Journal* **22**, 135–146.

[15] Cameriere, R., Ferrante, L., Mirtella, D., Rollo, F.U. & Cingolani, M. (2005). Frontal sinus for identification: qualità of classifications, possibile error and potential corrections, *Journal of Forensic Sciences* **50**(4), 770–773.

[16] Christensen, C. (2005). Testing the reliability of frontal sinuses in Positive Identification, *Journal of Forensic Sciences* **50**(1), 18–22.

[17] Dean, D.E., Tatarek, N.E., Rich, J., Brogdon, B.G. & Powers, R.H. (2005). Human identification from the ankle with pre and postsurgical radiographs, *Journal of Clinical Forensic Medicine* **12**(1), 5–9.

[18] Goodman, N.R. & Himmelberger, L.K. (2002). Identifying skeletal remains found in a sewer, *The Journal of the American Dental Association* **133**(11), 1508–1513.

[19] Kahana, T., Freund, M. & Hiss, J. (1997). Suicidal terrorist bombings in Israel: identification of human remains, *Journal of Forensic Sciences* **42**(2), 259–264.

[20] Kahana, T., Goldin, L. & Hiss, J. (2002). Personal identification based on radiographic vertebral features, *The American Journal of Forensic Medicine and Pathology* **23**(1), 36–41.

[21] Kahana, T., Ravioli, J.A., Urroz, C.L. & Hiss, J. (1997). Radiographic identification of fragmentary human remains from a mass disaster, *The American Journal of Forensic Medicine and Pathology* **18**(1), 40–44.

[22] Koot, M.G., Sauer, N.J. & Fenton, T.W. (2005). Radiographic human identification using bones of the hand: a validation study, *Journal of Forensic Sciences* **50**(2), 263–268.

[23] Kuehn, C.M., Taylor, K.M., Mann, F.A., Wilson, A.J. & Harruff, R.C. (2002). Validation of chest x-rays comparisons for unknown decedent identification, *Journal of Forensic Sciences* **47**(4), 725–729.

[24] Mann, R.W. (1998). Use of bone trabeculae to establish positive identification, *Forensic Science International* **98**(1–2), 91–99.

[25] Owsley, D.W. & Mann, R.W. (1992). Positive personal identity of skeletonized remains using abdominal and pelvic radiographs, *Journal of Forensic Sciences* **37**(1), 332–336.

[26] Rathbun, T.A. & Buikstra, J.E. (1984). *Human Identification: Case Studies in Forensic Anthropology*, Charles C. Thomas Publisher, Springfield.

[27] Rogers, T. & Allard, T.T. (2004). Expert testimony and positive identification of human remains through cranial suture patterns, *Journal of Forensic Sciences* **49**(2), 203–207.

[28] Smith, D.R., Limbird, K.G. & Hoffman, J.M. (2002). Identification of human skeletal remains by comparison of bony details of the cranium using computerized tomographic (CT) scans, *Journal of Forensic Sciences* **47**(5), 937–939.

[29] Sudimack, J.R., Lewis, B.J., Rich, J., Dean, D.E. & Fardal, P.M. (2002). Identification of decomposed human remains from radiographic comparisons of an unusual foot deformity, *Journal of Forensic Sciences* **47**(1), 218–220.

[30] Valenzuela, A. (1997). Radiographic comparison of the lumbar spine for positive identification of human remains. A case report, *The American Journal of Forensic Medicine and Pathology* **18**(2), 215–217.

[31] Campobasso, C.P., Dell'Erba, A.S., Belviso, M. & Di Vella, G. (2007). Craniofacial identification by comparison of antemortem and postmortem radiograhs: two case reports dealing with burnt bodies, *The American Journal of Forensic Medicine and Pathology* **28**(2), 182–186.

[32] Kahana, T., Hiss, J. & Smith, P. (1998). Quantitative assessment of trabecular bone pattern identification, *Journal of Forensic Sciences* **43**, 1144–1147.

[33] Kahana, T. & Hiss, J. (1994). Positive identification by means of trabecular bone pattern comparison, *Journal of Forensic Sciences* **39**, 1325–1330.

[34] Jayaprakash, P.T., Srinivasan, G.J. & Amravaneswaran, M.G. (2001). Craniofacial morphoanalysis: a new method for enhancing reliability while identifying skulls

by photosuperimposition, *Forensic Science International* **117**(1–2), 121–143.
[35] Aulsebrook, W.A., I§ can, M.Y., Slabbert, J.H. & Becker, P. (1995). Superimposition and reconstruction in forensic facial identification: a survey, *Forensic Science International* **75**(2–3), 101–120.
[36] Cattaneo, C. & Grandi, M. (2004). *Antropologia ed Odontologia Forense, Guida allo studio dei resti umani*, Monduzzi Editore, Bologna.
[37] Austin-Smith, M.A. & Maples, W.R. (1994). The reliability of skull/photograph superimposition in individual identification, *Journal of Forensic Sciences* **39**(2), 446–455.
[38] Yoshino, M., Imaizumi, K., Miyasaka, S. & Seta, S. (1995). Evaluation of anatomical cosistency in craniofacial superimposition images, *Forensic Science International* **74**(1–2), 125–134.
[39] Yoshino, M., Matsuda, H., Kubota, S., Imaizumi, K., Miyasaka, S. & Seta, S. (1997). Computer-assisted skull identification system using video superimposition, *Forensic Science International* **90**(3), 231–244.
[40] Ghosh, A.K. & Sinha, P. (2001). An economised craniofacial identification system, *Forensic Science International* **117**(1–2), 109–119.
[41] Al-Amad, S., McCullough, M., Graham, J., Clement, J. & Hill, A. (2006). Carniofacial identification by computer-mediated superimposition, *The Journal of Forensic Odonto-Stomatology* **24**(2), 47–52.
[42] De Angelis, D., Cattaneo, C. & Grandi, M. (2007). Dental superimposition: a pilot study forstandardising the method, *International Journal of Legal Medicine* **121**(6), 501–506.
[43] Fenton, T.W., Heard, A.N. & Sauer, N.J. (2008). Skull-photo superimposition and border deaths: identification through exclusion and the falure to exclude, *Journal of Forensic Sciences* **53**(1), 34–40.
[44] De Winne, J. (2001). Disaster victim identification at international level. The role of INTERPOL- now and in the future, *The Journal of Forensic Odonto-Stomatology* **19**(2), 40–42.
[45] Komar, D. (2003). Lessons from Srebrenica: the contributions and limitations of physical anthropology in identifying victims of war crimes, *Journal of Forensic Sciences* **48**(4), 713–716.
[46] Soomer, H., Ranta, H. & Penttila, A. (2001). Identification of victims from the M/S Estonia, *International Journal of Legal Medicine* **114**(4–5), 259–262.
[47] Primorac, D., Andelinovic, S., Definis-Gojanovic, M., Drmic, I., Rezic, B., Baden, M.M., Kennedy, M.A., Schanfield, M.S., Skakel, S.B. & Lee, H.C. (1996). Identification of war victims from mass graves in Croatia, Bosnia, and Herzegovina by use of standard forensic methods and DNA typing, *Journal of Forensic Sciences* **41**(5), 891–894.
[48] Komar, D. (2003). Lessons from Srebrenica: the contribution and limitations of physiscal anthropology in identifying victims of war crimes, *Journal of Forensic Sciences* **48**(4), 713–716.
[49] Baraybar, J.P. (2008). When DNA is not available, can we still identify people? Recommendations for best practice, *Journal of Forensic Sciences* **53**(3), 533–540.

Related Articles

Anthropology

DNA

Friction Ridge Skin: Fingerprint Detection and Recovery Techniques

Odontology

CRISTINA CATTANEO, DAVIDE PORTA AND DANILO DE ANGELIS

Identifiler

When the 13-core short tandem repeat (STR) loci for Combined Offender DNA Index System (CODIS) entry in the United States were adopted in 1997, scientists were required to use a minimum of two tests to generate this number of results. By 2001 the sophistication of multiplex technology had improved to the point where STR multiplex systems that encompassed the full set of core loci were produced. The 15-locus AmpF*l*STR® Identifiler™ STR multiplex system (Applied Biosystems, Foster City, CA) is one such system and required the addition of a fifth fluorescent dye (PET®) and mobility modifying nonnucleotide linkers to achieve this enhanced level of performance. The enlargement of forensic STR multiplexes to coamplify up to 15 loci provides extremely high discriminating power for unmixed full profile matches. This additional discriminating power is of great practical benefit for resolving complex kinship cases such as those that occur in mass disaster identifications. It also assists to minimize the likelihood of finding adventitious matching profiles on large forensic DNA database systems. The loci typed in the Identifiler™ system are displayed in the image of the allelic ladder in Figure 1.

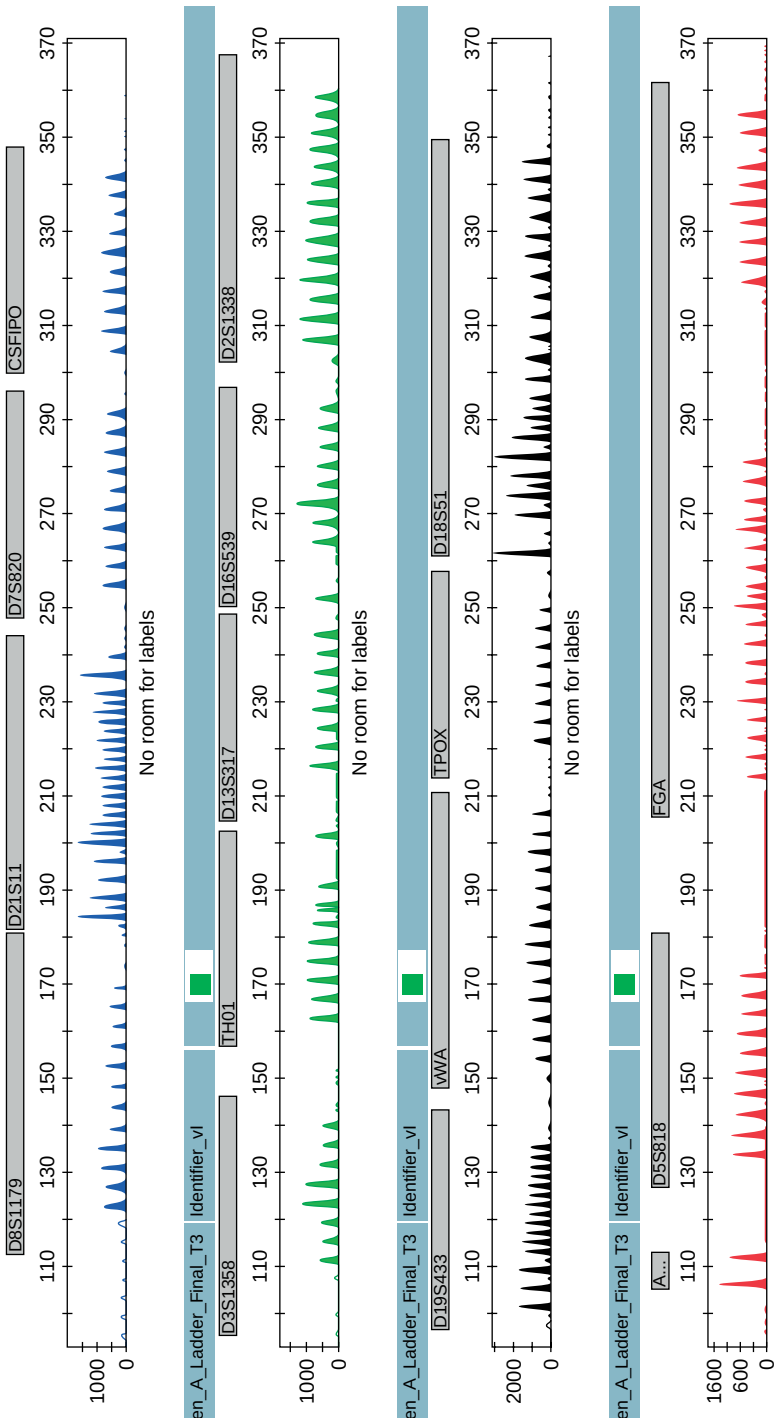


Figure 1 The 16 loci amplified using the AmpF/STR® Identifiler™ Multiplex System. (Reproduced from AmpF/STR® Identifiler™ PCR Amplification Kit Product Bulletin, Applied Biosystems, Foster City, CA)

Reference

[1] Butler, J.M. (2001). *Forensic DNA Typing: Biology and Technology Behind STR Markers*, Academic Press: San Diego, Ca.

Related Articles

Microsatellites

SIMON J. WALSH

Identity Disorder *see* Dissociative Disorders

Illusion of "Evil" *see* Evil: Illusion of

Image Processing and Analysis

History

From the late 1980s [1, 2], digital forensic image processing was becoming more widely used in court cases to enhance images. For still photographs, it has been used operationally in the United Kingdom at the Home Office Scientific Research and Development Branch since 1984 [3]. Fields where digital image processing is commonly used [2] are CCTV systems, document examination, fingerprint recognition, and photography.

Societies

There exist different societies that encompass forensic image processing. The SPIE [4] working group Investigative Image Processing has existed since 1999, and several conferences and proceedings have been conducted in this field. The American Academy of Forensic Sciences (AAFS) [5] has the Engineering Sciences group and there are ideas to set up a separate Digital Evidence and Multimedia group. For scientific working groups we see the scientific working group IT (SWGIT) [6] which has a liaison with the International Association for Identification (IAI). In Europe, the European Network of Forensic Science Institutes [7] has the Digital Imaging Working group and the Forensic IT Working Group.

CCTV

In CCTV we have seen developments from purely analog systems to the digital systems. For analog systems, several commercial image processing packages were developed for image processing of the video streams.

In CCTV systems, which are similar to the television systems, several standards exist: NTSC, PAL, and SECAM. For example PAL has 25 frames per second, and 50 fields per second. To have the best visual effect, the fields are composed of even and odd fields, such that the movie is more fluent. An example of odd fields and even fields is shown in Figure 2.

Analog CCTV images have to be digitized prior to processing them on a computer. When digitizing the image itself, the image is represented in the computer by a grid of numbers. The number at each grid point indicates the brightness at that point. Often a value of



Figure 1 Example of a frame that is composed of odd and even fields



Figure 2 The odd field of the frame in Figure 1

0 represents black and 255 represents white. In a color image three numbers are stored for each grid point (pixel) to indicate the red, green, and blue values at that point.

In Figure 1 an image is shown with one frame. Here we can see the odd and even lines. In Figure 2 one field of the image is shown. The license plate that is not moving is clearly readable. Here we see a time shift between the odd and even field.

In most packages image processing the next operations are found:

- contrast enhancement;
- edge filters;
- color image processing;
- deblurring; and
- measurements (such as height measurements, which are handled in the chapter).

With contrast enhancement, the information that is already there can be visualized. It is important to understand that there is no more detailed information available in an image than the grid point (pixel) that is visible in the image. For that reason zooming in does not help with a poor resolution of the image itself. This is visualized in Figure 3 with a license plate of a car. The license plate is magnified, and the limitations of the resolution and blur in the image are clearly visible.

There are possibilities to have more information if several images are available. If the target is not moving and there is noise [8] in the image, it is possible to reduce the noise and make the image itself more clearly visible. This is a possibility with a surveillance camera that stands still and captures multiple frames of an object that does not move.

Another possibility is to integrate the images of an object or a camera that moves, the super resolution method. In Figure 4 we have an example of a moving camera with some known number plates in lab environment. The images are registered first and then averaged.

There exist more sophisticated methods for deblurring, such as Wiener restoration, which is commonly used in image processing [9]. The movement and blurring filter of the movement should be known. In Figure 5 we have a known movement of the persons on their bicycles. With the known blurring function,

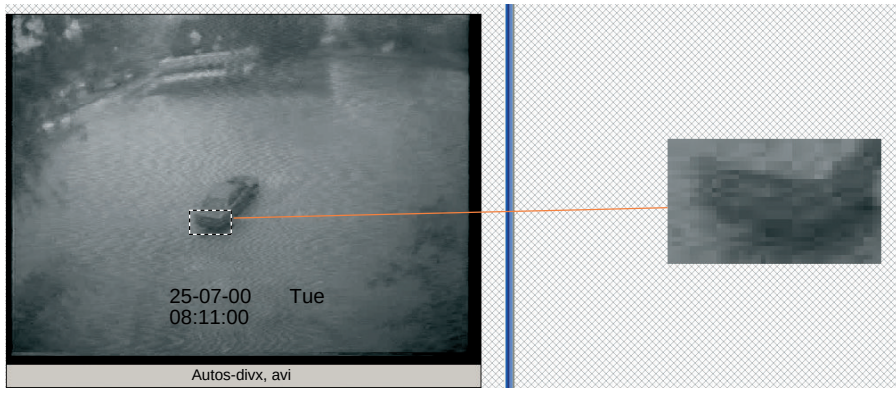


Figure 3 Magnification of a number plate of a car. On the right the selected part is magnified. It is not possible to reconstruct the number plate with this single image



Figure 4 Laboratory experiment: (a) one field, (b) super resolution image, (c) original



Figure 5 Wiener restoration of the field shown in Figure 2. The blurring function is estimated from the velocity of the bicycle

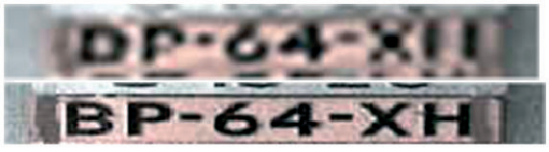


Figure 6 Laboratory experiment: super resolution with number plate in lab

the number plate on the bicycle becomes more clearly visible.

The methods should be interpreted and used with care, especially when there is digital compression in the images, which might alter the images. It is always necessary to validate the results before drawing a conclusion. For validation of test results it is important to have a large enough test set where the ground truth is known. To make a good validation, it is advisable to consult a forensic statistician. In Figure 6 we have an example of a wrong conclusion that could be drawn from these images when super resolution is used improperly.

Digital Compression

In digital systems mostly digital compression is used. For digitizing analog tapes in a forensic setting, it is always advisable to use lossless compression. In real world CCTVsystems band width and hard disk capacity cost money. This means that with compression a file with a size of 1 MB can be compressed (for example to JPEG) to 15 KB. The draw back is that we lose information. An example of images with various levels of compression is shown in Figure 7.

With video streams such as MPEG-2, MPEG-4/3GP, and many others, mostly a "lossy" compression scheme is used which compresses the video also by predicting the next frame. Also with these compression schemes one should be aware that we lose information, and alter the image. Mostly these compression schemes are optimized for the human eye.

Integrity

Questions of integrity are as follows:

- Is a specific image captured with a specific camera?



Figure 7 Different levels of compression. The size is given in the middle with a magnification of the eye under various compression ratios of JPEG



Figure 8 Example of an easy to recognize forged photograph

- With which brand/model has this image been captured?
- Has the image been altered?
- Has the digital compression altered the image?

An example of an easy to recognize forged photograph is shown in Figure 8. How can we see that the image has been altered? We might look at the next properties of the image:

- shadows
- edges
- compression artifacts
- pixel artifacts
- cut and paste artifacts
- light conditions
- noise
- original images if they are available.

When someone manipulates an image in a professional manner (and takes into account the properties above) it becomes nearly impossible to determine if the image has been manipulated. For this reason the chain of evidence should always be considered, in a similar way to other forensic evidence, before drawing conclusion.

Pixel Artifacts

Pixel artifacts can be used to examine

- if a picture has been taken with a certain camera; and
- whether two series of images taken with the same camera can be used to determine pixel artifacts.

Pixel artifacts are defects in the charge coupled device (CCD). Sometimes they are visible in the image itself, and can be used to draw a conclusion as to whether a certain image has been made with a certain camera. In coming to that opinion, it is important to have knowledge of the randomness of the pixel artifacts in a given batch of cameras and their CCDs. Since the manufacturing process of

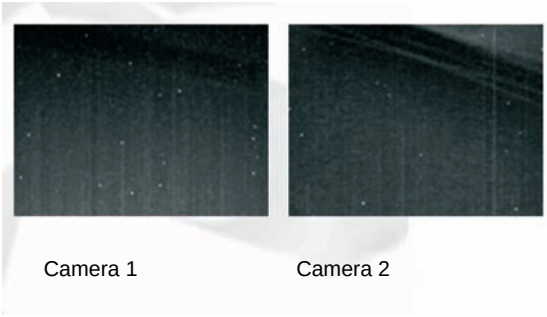


Figure 9 pixel artifacts visualized in images of two different cameras

CCDs is getting more precise, and less pixel artifacts are available, other methods such as measuring noise in the image should be considered. In Figure 9 an example of pixel artifacts is shown in an image with two different cameras. pixel artifacts are not always visible in a given image, depending on the gray level of the pixel and the surrounding pixels in a certain picture. There are methods such as averaging the images to make them clearer.

Determining the photo response nonuniformity (PRNU) [10] is an alternative approach for pixel artifacts. It is one source of pattern noise in digital cameras. An approach for JPEG-compressed images is available for download from http://sourceforge.net/projects/prnucompare/.

Background

A digital image consists of a matrix of image elements known as *pixels*. A digital image is made by means of CCDs or a complementary metal-oxide semiconductor (CMOS). The CCD of a digital camera is a matrix of light-sensitive elements. The light-sensitive elements on the CCD correspond with the pixels in a digital image.

Contamination and disruption of the light-sensitive elements can occur during the manufacture of the CCD. The result of this is that the light-sensitive elements work better or worse than the surrounding elements so that the intensity of the corresponding pixels in the image is always higher or lower than that of the surrounding pixels. These phenomena are known as *pixel artifacts*. Depending on the extent of the contamination this can occur in a cluster form whereby several connected light-sensitive elements become contaminated.

There are several studies [11–14] with various cameras of different manufacturers and with a series of cameras with successive serial numbers of the same manufacturer. Up to now show that the location of (clusters of) defects on the CCD surface are randomly distributed between CCDs. In general the pixel artifacts are clear and can be demonstrated in reproducible tests. As a result of these properties the pattern of pixel artifacts provides a kind of digital fingerprint of a camera.

Header and Footer Information. In a comparative study of digital images it is first checked whether the properties such as the format and header or footer information of the images are similar. Header or footer information of a digital image file is the information that the camera adds to an image file. This can contain information on the settings, the make and model of the camera, etc. This header or footer information is not visible in the images but can be made visible by examining the file itself. It is possible to alter the header and footer information by means of software.

The header information is a block of information that stands in the image file and precedes the block with the pixel values. Footer information stands in the block after the block with the pixel values of the image. This is shown in the graphic in Figure 10. Not all files contain header and footer information however. Files of the type JPEG usually contain header information.

Pixel Artifacts. Each digital image is formed by the actual image, noise, and any pixel artifacts. The

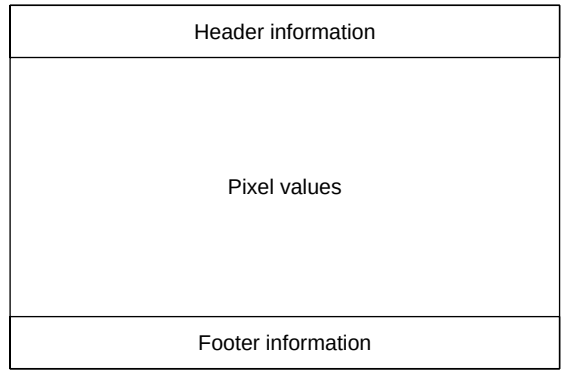


Figure 10 Schematic overview of a file

actual image and the noise differ between the images mutually made by the same digital camera at the same temperature, but the pixel artifacts are present to the same extent.

There are two investigation options:

1. Linking a digital camera to a serie(s) of image(s).
2. Comparing series of images mutually (e.g., by comparing video streams tapped from Internet traffic if the camera is not available).

In the first case, a series of test recordings is made with the camera in question, e.g., of a smooth white, gray, and/or black surface after which this set of images is used as the reference set.

In the second method, a series of images is used as the reference set. In this case, the camera is not available so that it is not possible to perform any test with camera itself.

If sufficient images are available, it is possible to find the pixel artifacts by adding a series of images (pixel by pixel) and to calculate an average. In this way changes parts of the image, the actual image and the noise, are averaged out and the consistent elements of the image, the pixel artifacts, remain behind. A filter such as a median filter, can also be used to increase the visibility of the pixel artifacts.

In some cases pixel artifacts can be seen in an image without any further processing so that only one image can be used instead of a series. It does not, however, follow that pixel artifacts are visible in every image made by the camera. Visibility depends on the object on the position of the pixel, temperature and the lighting of the image, and/or the way in which the image has been compacted.

The positions of the pixel artifacts in the reference set are compared with the positions of the pixel artifacts in the disputed images. Similar positions of pixel artifacts can be strong indication that the images were made by the same camera. Dissimilar positions can be a strong indication that the images were not made by the same camera.

Conclusions. Conclusions about the results of the investigation into pixel artifacts are reported in terms of support for the hypothesis that an image was or was not made by a certain camera. To express the extent of the support for a hypothesis some investigators make a choice from a verbal scale of "no support", "some support", "reasonable support", "much support," and "strong support" in a Bayesian framework.

The conclusions cannot be formulated quantitatively if insufficient statistical information is available to determine to what extent pixel artifacts are actually randomly distributed in the CCD or CMOS.

Comments.

- It is possible to add or remove pixel artifacts artificially by means of image processing software. The conclusions are based on the assumption that the integrity of the images is guaranteed.
- The investigation is based on the assumption that pixel artifacts are randomly distributed. To check this in the case of a specific CCD it is necessary to compare several CCDs of the same production process with each other.

Fingerprints

Several image processing methods are used for processing fingerprints [15]:

- contrast stretching;
- convolution filtering;
- separation of colors;
- dilation and erosion; and
- fast fourier transform (FFT) techniques for filtering regular patterns.

These methods are also used in combination and in local areas of the image. A question that arises when using these methods is if the methods have been validated (so that no information or minutiae are added that does not exist). Furthermore there is interest in new techniques that can be used for fingerprint enhancement with image processing (e.g., wavelet filtering).

The most common image processing methods as contrast enhancements are common knowledge, and do not have the risk of altering an image that minutiae are shown which are actually not there.

It is important to have new image processing methods validated and know what the risks are of these methods. Another issue is that the software that is used should be tested if it really does the image processing function that is requested.

More complicated methods such as FFT can degrade the image in such a way that the data can

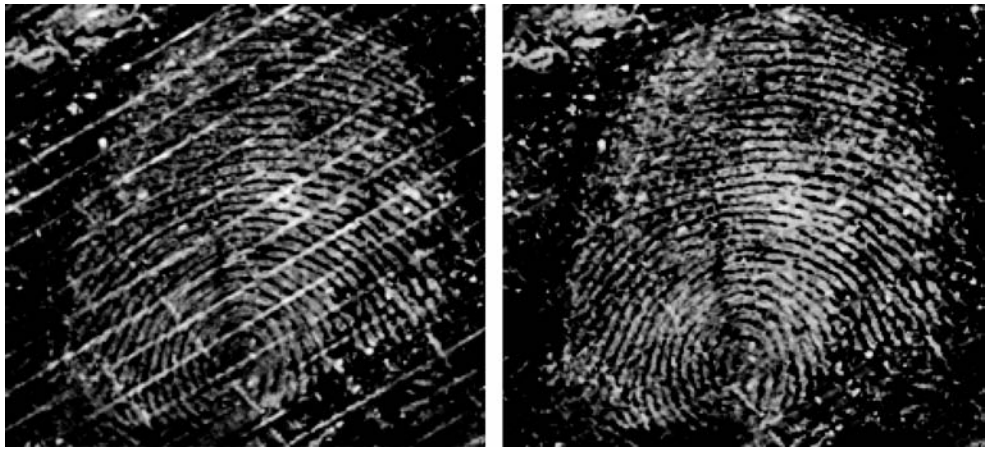


Figure 11 FFT filtering of image: (a) image with regular pattern and (b) filtered image

be altered. In 1994 this was mentioned in literature in 1993 by Watling [16] in 1995 by S. Bramble [17]. Both describe a warning on using the FFT, for instance Watling: "However one must be extremely careful when using the FFT spike boost so as not to cross the line between enhancement and restoration."

In a Daubert hearing in 1998 [18] concerning image processing of fingerprints the methods were accepted in court without much discussion. It even stated:

"The evidence in the record supports the trial court's unchallenged findings that the technique utilized by Berg has a reliability factor of 100% and a 0% margin of error and that the results are visually verifiable and could be easily duplicated by another expert using his or her own digital camera and appropriate computer software."

Which is not in agreement with the earlier publications.

In Figure 11 an example of image processing with FFT is shown. The regular pattern can be filtered out since it has a known frequency.

Documents

With documents color image processing is sometimes feasible. Examples are available from www.4n6site.com. If different colors (for example from different inks) have been used, it is possible to separate the colors and make the handwriting more clearly visible. Also on document image processing early publications suggest the limitations of image processing [19]: "However, the Laplacian filter also produced misleading answers on the other sequence problems."

Concluding Remarks

As mentioned, image processing can be used to enhance an image. This means that the limitations of the final image are limited by:

- the camera with lens system;
- the cabling between camera and recording system;
- the recording device and its settings and properties; and
- the playback device and its settings and properties.

In digital systems it is important to have the same playback software and hardware as where the images have been captured with. Since there exist many different digital CCTVsystems, a collection of players and hardware can help to speed up the examination. Also software can give a different playback results as is shown in Figure 12.

There has been much research in the field of image restoration, however in practice the usage is limited since the blurring functions are often more complicated and not known. Also the human eye compensates for many kinds of blurring functions. Measurement in images should be done with care and with enough validation as is shown in the chapter "Length measurement".



Figure 12 Two test recordings by two different cameras ((a) left camera, (b) right camera), where ten images were averaged. The positions of the pixel artefacts (white dots) are in different places

References

[1] Cannon, T.M. & Trussel, H.J. (1980). Applications of digital image restoration to photographic evidence. *Proceedings of the 1980 Carnahan Conference on Crime Countermeasures*, pp. 103–107.

[2] FBI Academy (1986). *Proceedings of the International Symposium on the Forensic Applications of digital Image Processing*, June 16–20 1986, Quantico.

[3] Lehar, A.F. & Stevens, R.J. Image processing system for enhancement and deblurring of photographs, *SPIE, Opt. Eng.* **23**, 303–308.

[4] www.spie.org (accessed on, 2008).

[5] www.aafs.org (accessed on, 2008).

[6] http://www.theiai.org/guidelines/swgit/index.php (accessed on, 2008).

[7] www.enfsi.eu (accessed on, 2008).

[8] Heerich, V. (2002). Urkundenlabor "Identification of a digital image by means of CCD Noise", *Presentation at the 3rd ENFSI Digital Imaging Workshop Meeting (DIWG) in Zurich*, 23–25 October 2002, Zurich.

[9] Tekalp, A.M. & Kaufman, A. (1994). Estimation techniques in image restoration, in *Digital Image Processing: Techniques and Applications*, C.T. Leondes, ed, Academic Press, pp. 1–34.

[10] Lukáš, J., Fridrich, J. & Goljan, M. (2006). Digital camera identification from sensor pattern noise, submitted to *IEEE Transactions on Information Forensics and Security* **1**(2), 205–214.

[11] Kursawa, K., Kuroki, K., Saitoh, N. (1999). CCD Fingerprint method–identification of video camera from videotapes images. *International Conference on Image Processing 1999*, IEEE Computer Society, Kobe, Vol. III, pp. 537–540.

[12] Geradts, Z., Bijhold, J., Kieft, M., Kurosawa, K., Kuroki, K. & Saitoh, N. (2001). Methods for identification of images acquired with digital cameras, in *Enabling Technologies for Law Enforcement and Security*, Vol. 4232, SPIE, Boston, pp. 505–512.

[13] Kursawa, K., Kuroki, K. & Saitoh, N. (2002). An approach to individual video camera identification, *Journal of Forensic Sciences* **47**(1), 97–102.

[14] Geradts, Z., Bijhold, J., Kieft, M., Kurusawa, K., Kuroki, K. & Saitoh, N. (2002). Digital camera identification, *Journal of Forensic Identification* **52**(2), 621–632.

[15] McRoberts, A.L. (1987). Digital image processing as a mean of enhancing latent fingerprints, *Proceedings of the International Forensic Symposium on Latent Prints, FBI*, Washington DC, July 7–10, 1987, pp. 165–166.

[16] Watling, W. (1993). Using the FFT in forensic digital image enhancement, *Journal of Forensic Identification* **43**(6), 573–584.

[17] Bramble, S. (1994). Operational experience of finger mark enhancement by frequency domain filtering, *Journal of Forensic Science* **39**(4), 920–932.

[18] STATE v. HAYDEN 109 90 Wn. App. 100, 950 P.2d 1024 (1998).

[19] Schuetzner, E.M. (1988). Examination of sequences of strokes with an image enhancement system, *Journal of Forensic Sciences* **33**(1), 244–248.

ZENO GERADTS

Implanted Memories *see* Deception: Truth Serum

Impressions: Manufactured Items ***see*** Marks or Impressions of Manufactured Items

Impressions: Tire ***see*** Tire Impressions

Impressions: Toolmarks ***see*** Toolmarks

Impulses ***see*** Compulsion

Impulsiveness ***see*** Aggression

In Limine Motions and Hearings

The term *in limine* is from Latin and means "at the outset" or "at the threshold".

In US courts – and also in other countries that follow the adversary system of jurisprudence – if, either prior to trial or during a trial, an attorney wishes to challenge an opponent's expert testimony, the attorney can file a motion *in limine* asking the court to exclude the anticipated expert evidence. The predicate for such a motion is that the proponent contends the proposed evidence is inadmissible because it is either of questionable reliability or does not meet either the *Frye*, the *Daubert*, or other proof-of-reliability tests for admissibility. The purpose of the motion is to avoid inserting irrelevant and unreliable matter into a proceeding and thereby causing a mistrial.

The court, if it believes the motion may have merit, will permit the attorneys to present evidence for and against the motion. This proffer of evidence assists the judge in arriving at a reasoned admissibility decision. Hearings on motions *in limine* are always conducted outside the presence of a jury; their purpose is merely to enlighten the judge on the proposed subject matter's admissibility, or, if an advocate believes a judge's decision on the motion is in error, to create a record upon which an appeal can be based.

The moving party – the side filing the motion to exclude evidence – has the burden of presenting supporting evidence. The court will, if the motion is granted, either exclude the forensic expert's testimony or limit its scope.

The practice of filing a motion *in limine*, requesting a special judge ruling on the admissibility of certain evidence, may also be filed by an attorney who proposes to use a certain novel expertise, if that attorney anticipates strong objection to the evidence from the opposing side, and seeks a favorable ruling on the admissibility issue.

Motions *in limine* are most often filed prior to trial, so that when the trial commences, the admissibility issue will have been settled "at the outset" and there will not be any delay of the proceedings while the jury is removed from the courtroom to permit attorneys to argue the evidentiary objections. In some cases, such motions may also be filed and entertained when a trial has already commenced. In such a case, the jury will be removed from the courtroom while the judge hears the evidence on the motion.

Related Articles

Daubert v. Merrell Dow Pharmaceuticals
Discovery of Expert Findings
Discovery in the United States: Civil Cases
Discovery in the United States: Criminal Cases
Discovery: Depositions
Discovery: Discovery Motions

Expert Opinion in Court: a Comparison of Approaches

Expert Opinion in Court: Civil Law Jurisdictions (France, Germany, Italy, and Spain)

Expert Opinion: United Kingdom, Canada, and Australia

Expert Opinion: United States

Frye v. United States

ANDRE MOENSSENS

Incest *see* Child Sexual Abuse Accommodation

Independent Living: Capacity for *see* Capacity for Independent Living

Industrial Incidents: Human Factors *see* Human Factors: Industrial Incidents

Infanticide *see* Postpartum Psychosis

Informed Consent *see* Therapeutic Jurisprudence

Injury: Burns, Scalds, and Chemical

Burns

Introduction

The possible findings in burned bodies cover a broad spectrum, ranging from minor, local, superficial burns of the skin to calcined skeletal remains without any soft tissue left and total incineration. The extent of the tissue changes depends on the temperature actually applied to the body, the time for which it is applied, the kind of transmission of the heat to the body, and other prevailing conditions. In most cases, the heat acts on the body beyond death. Consequently, the changes found are largely of postmortem origin. The effects of the heat on the body are burns of the exposed tissue, changes in the content and distribution of tissue fluids, fixation of the tissue, and shrinking processes (Table 1).

The forensic investigation of deaths related to a fire is important to determine the manner and cause of death, the vitality of the findings, and the identity of the victim. The basis of the assessment is a careful evaluation of the autopsy findings. Additional investigations, such as toxicology (e.g., determination of carboxyhemoglobin concentration, cyanide (CN) concentration, and blood alcohol concentration) or histology (particularly of the airways), may help to complete the assessment of the case.

External Findings

Among the externally discernible changes, the various stages of skin burns, the results of tissue shrinkage, and the consumption by the fire are the dominant features. In most fire deaths, the body was exposed after death to temperatures of several hundred degrees Celsius for at least several minutes, often by direct contact with the flames. Destruction may be so extensive that the less experienced tend to consider an autopsy pointless, because in their opinion it will not produce any useful findings. But this is definitely wrong: even charred torsos, with exposure of the body cavities and partial amputation of the extremities, the organs of the thorax, and abdomen, can usually still be assessed quite well, and body fluids

Table 1 Effects of heat on human bodies

Heat effect	External findings	Internal findings
Burns	Skin burns (first- to fourth-degree);	Burn injuries and destruction of internal organs and the skeletal system;
	pseudo-washerwoman's skin on palms and soles;	Edema, hemorrhages and vesicular detachment of the respiratory mucosa
	Singeing of the hair; Consumption by the fire	
Changes of the content and distribution of tissue fluid	Skin blisters	Evaporation of body fluids;
		Rupture of the abdominal wall with prolapse of intestinal loops;
		Exit of blood-tinged, foamy liquid from mouth and nose;
		Heat hematoma;
		Pseudohemorrhages in organs or cavities;
		Accumulation of fat in body cavities, vessels, or the heart
Heat fixation	Leathery, desiccated skin with brown discoloration	Hardening of internal organs and muscles;
		Fragmentation of erythrocytes
Shrinking of tissue	Tightening of the skin;	Shrinking of the internal organs up to the so-called "puppet organs"
	Heat-related tears in the skin;	
	Protrusion of the tongue;	
	Petechial hemorrhages on neck and head;	
	Pugilistic attitude	

as well as tissue samples can be obtained for further investigations.

Burns and Consumption by Fire. Skin burns are categorized into four degrees, with each degree characterizing a certain depth of the skin lesion. The categories are degree 1 (superficial burns), degree 2a (superficial partial-thickness burns associated with necrosis of the upper layers of the epidermis), degree 2b (deep partial-thickness burns associated with necrosis of the entire thickness of the epidermis), degree 3 (full-thickness burns with necrosis involving the dermis as well), and degree 4 (charring in which the heat lesion reaches deeper soft tissue layers). First-degree burns are characterized by a reddening of the skin. As a postmortem residue, a red margin may occasionally be observed. The leading sign of a second-degree burn is a fluid-filled skin blister. Second-degree burns of the palms of the hands and the soles of the feet appear as whitish discolorations of the epidermis associated with swelling, wrinkling, and vesicular detachment up to glovelike peeling. The findings resemble the so-called washerwoman's skin, as seen after prolonged exposure to a moist environment [1]. The skin in third-degree burns is firm and discolored brown.

Heat changes of the hair occur at temperatures above about 150 °C. This can be used to differentiate between burns and scalds or to indicate the approximate temperature on the skin reached in smoldering fires. The hair gets frizzy and brittle and assumes a fox-red or dark brown to black color (Figure 1). Temperatures of about 200 °C lead to the formation of gas bubbles in the shaft, at 240 °C the hair becomes frizzy due to the melting of the hair keratins, and above 300 °C charring occurs. Singeing of the head hair is usually not associated with high flames, but with a characteristic smell. In contrast to this, frizzy hair burns with high, open, and sustained flames, causing



Figure 1 Singening of the hair: the hair is frizzy and brittle and assumes a fox-red color

severe damage to the neighboring skin or mucosa [2]. The explanation for this phenomenon is the larger distance between the individual hairs, which allows better access to oxygen.

The distribution of burn injuries is seldom even and depends on various conditions. Especially, tight-fitting clothes can protect the underlying skin from burns for a long time. In the same way, it may be possible to prove a homicide by manual strangulation in a fire victim with the ligature still in place [3]. In cases of suicidal self-incineration using fire accelerants, burns may be absent from the feet and lower legs if the incineration took place while the body was in an upright position [4]. In burns caused by low heat, deep, anatomically circumscribed signs of consumption by the fire may occur (Figure 2). These are formed when the fire is maintained according to the wick principle: in those parts of the body where the skin has burned away, liquefied subcutaneous fatty tissue leaks out and maintains the fire [5, 6]. The bizarre distribution of the burn lesions in such cases has given rise to the myth of spontaneous human combustion [7, 8].

The question of to what extent the level of charring in a burned body allows drawing conclusions as to the duration of the fire is rarely asked [9]. In the literature there are a few reports on this topic, most of which refer to observations made during cremations



Figure 2 Circumscribed charring and consumption of the soft tissue of the left lower leg by a smoldering fire in a 73-year-old woman

[10–12]. These studies have shown that the course of events follows a fixed chronological order.

Some authors have proposed classifications as a tool to describe the extent and distribution of consumption by a fire [13–16]. According to our experience, the Crow–Glassman scale (CGS, Table 2) [15] is the best usable classification system. There is a clear relation to the scene of the fire. Lower degrees of destruction are found more often in fire victims recovered from buildings, whereas in car fires the extent of consumption by the fire is usually higher owing to the higher temperatures produced by these fires. Massive destruction does not only require higher temperatures but also an especially a longer duration of exposure [12, 17].

Shrinkage of Tissue. The reason why the tissue shrinks is the loss of fluid caused by the heat. Externally, it is characterized by tightening of the skin, splitting of the skin, protrusion of the tongue from the open mouth, petechial hemorrhages in the region of the neck and head, and the so-called pugilistic attitude.

The typical posture of charred bodies is called pugilistic or boxer's attitude with the arms being abducted in the shoulder joint and flexed in the elbow joint and the legs being abducted in the hip joint and flexed in the hip and knee joints (Figure 3). The reason for this phenomenon is the shrinkage of muscles and tendons caused by the heat. The flexion of the joints of the extremities is due to the predominance of the flexor muscles. This may even result in dislocation, most often recognizable at the wrists. In the same way, contracted feet may be observed, especially after advanced consumption of the legs.

Table 2 Crow–Glassman scale (CGS) of burn-related destruction of corpses [15]

Level 1	Second-degree burns, sometimes singeing of the hair; visual identification possible
Level 2	Burns of varying severity, sometimes with thermal destruction/amputation of ears, genitals, hands, or feet; visual identification may still be possible
Level 3	Consumption by the fire with partial amputation of arms and/or legs; cerebral cranium intact
Level 4	Bony lesions of the cerebral cranium; residual extremities still present
Level 5	Fragmented skeletal remains without soft tissue



Figure 3 Pugilistic attitude of a charred body as a result of the heat-mediated shrinking of the muscles and tendons

Splitting of the skin is a frequently observed phenomenon in charred bodies. It is very rare in burns of minor severity. If it appears in such cases, prolonged exposure to the heat has to be assumed. The splits have sharp edges that can be approximated, are often linear, but are occasionally also angled (Figure 4). In most cases they reach the subcutaneous fatty tissue, sometimes also the outer muscle layers. In this context, little attention has been paid to the fact that the tissue exposed in the depth of the splits is usually unburned and often not even sooted. Most likely the splits form during the cooling of the body.

Protrusion of the tongue from the open mouth is due to the heat-related shrinkage of the soft tissue of the neck (Figure 4). In the presence of severe burns on the neck and/or thorax, petechial hemorrhages may



Figure 4 Heat splits of the skin and protrusion of the tongue

occasionally be found in the lids and conjunctivae. For their formation, congestion due to the shrinkage of the soft tissue of the neck by heat or heat rigidity of the thorax while the circulation is still intact is discussed [18, 19]. So petechial hemorrhages in the region of the neck and head would have to be regarded as a vital sign.

Internal Findings

The internal findings in fire deaths are the result of a fixation of the tissue by the heat, processes of shrinking, thermal changes of the content and distribution of tissue fluids, and a rising gas pressure in hollow spaces. After the fire has opened the body cavities, direct burns occur also on the internal surfaces.

The loss of fluid and, after exposure of the body cavities, also the direct effect of the heat cause shrinking of the internal organs, which become firm, hardened, and cooked by the heat (so-called "puppet organs"). The surface of the organs becomes increasingly bosselated (formed into rounded, raised areas or knobs) and is reduced to a spongelike residual structure in the end. The tissue is meanwhile completely desiccated and disintegrates into ash at the slightest touch.

Respiratory Tract. The respiratory tract is the most important organ system for the diagnosis of vitality. Where fire fumes are inhaled, deposits of soot particles will be found. Edema, mucosal bleeding, and patchy or vesicular detachment of the mucosa in the nose, mouth, pharynx, larynx, trachea, and bronchi may be indicative of inhalation of hot gases. Often, increased secretion of mucus is observed in the air passages. This may be interpreted as an attempt to cool the surfaces of the air passages and thus as a sign of vitality, if other causes for the secretion of mucus (bronchial asthma, catarrhal bronchitis) have been ruled out [20, 21]. Damage caused to the respiratory tract by dry heat is limited more to the upper portions [22]. The inhaled hot air is sufficiently cooled down by the mucosa of the airways, so that after exposure to a "normal fire" hardly any changes are found in the medium and small bronchi [20–22].

Vessels. Intensive red discoloration of the intima of the vessels can be regularly observed in fire victims. This finding is the result of hemolysis [23] occurring at temperatures above 52 °C. Already at 48 °C erythrocytes begin to dissolve. If the circulation is still intact, the erythrocyte fragments ("fragmentocytes") can be demonstrated microscopically also in other organs, which has to be interpreted as a sign of vitality [24].

Gastrointestinal Tract. The abdominal wall protects the abdominal organs from direct damage by the

flames for a relatively long period of time. But as a result of the heating of the body, the tissue fluids boil away and the pressure inside the hollow organs and the abdominal cavity builds up, which often leads to the rupture of the abdominal wall and the prolapse of intestinal loops. Consumption of the abdominal wall by the fire further promotes the rupture. In rare cases, heat-related ruptures of the gastrointestinal organs are found, before they were directly exposed to the fire [25].

Bones. The skeleton is damaged by the fire, but it is not consumed completely. Even if it is exposed to a fire with high temperatures over a long period of time, there will usually still be remains to allow macroscopic assessment and successful determination of the species, the body measurements, and the sex as well as to identify skeletal anomalies and the presence of possible injuries [13, 26–30].

At temperatures above 700 °C, complete combustion of the organic substances with incineration and recrystallization of the inorganic matter occurs, which is called *calcination* [31, 32]. The bones are grayish-white and desiccated and disintegrate easily. The surface shows characteristic tears partly reflecting the course of the trabeculae, but often also being irregular in structure.

In charred or calcined bones, a minor mechanical strain may be enough to cause fractures. Especially in fractures localized within charred bone areas, the possibility of artifacts should be considered. But several authors have stressed that injuries sustained during life can be demonstrated even on charred or calcined bones [13, 29, 33].

In the assessment, special attention must be paid to the bony skullcap (Figure 5). In about 33% of all fire deaths, the skullcap is partially destroyed and the interior of the skull is exposed [17], which makes assessment even more difficult. Isolated fractures of the external table are seen especially in those cases where a defined area of the skullcap was in direct contact with the flames [11, 34]. Prolonged exposure to heat causes fractures of the entire thickness of the skullcap with occasional bursting of the sutures of the skull [11]. The tears in the skullcap caused by heat may radiate from a center, but can sometimes also be elliptic or circular in shape or resemble a spider's web fracture [34]. In rare cases, round or oval bone fragments may burst outward. Distinction of this finding from a gunshot injury may be difficult [35].



Figure 5 Skull cap with heat-related fractures and calcination of the tabula externa

Cranial Cavity and Brain. A frequent finding is the epidural heat hematoma. It is not a vital sign, but a postmortem effect due to the shift of fluid from the diploe and the venous sinuses when the skullcap is in direct contact with the flames. Accordingly, charring of the bony skull is usually found above the site of the heat hematoma. It is dry, crumbly, and of brick-red color. Occasionally, it may be surrounded by fat, and in rare cases accumulations of fat without extravasations of blood can also be found in the epidural space. Apart from that, the hematoma is sometimes found to be interspersed with brain tissue when the dura mater is torn as a result of shrinkage by heat [36].

Postmortem extravasates of blood may occur in all cavities of the skull including the ventricles of the brain. But, also in the brain tissue itself hemorrhages can be found as a result of the shrinkage

of tissue with laceration of small blood vessels. In these cases, distinction between postmortem and vital hemorrhages can be particularly difficult [37–39].

When the skullcap is intact, the brain of fire victims is often shrunken, and of a hardened consistency with filled sulci (grooves) on the surface. Dotzauer and Jacob described the finding as "a reduction of the brain volume associated with swelling of the internal matter" [40]. Histologically, condensation of the vessels and widening of the Virchow–Robinson spaces have been described [20, 40]. Again, this finding may be explained as the result of a loss of fluid and does not prove a vital thermal damage to the tissue.

Toxicology

In fire victims, the concentrations of carbon monoxide hemoglobin (CO-Hb), methemoglobin (met-Hb), and CN should be routinely determined in the corpse blood as part of the toxicological investigations [41–44]. Concentrations of more than 10% CO-Hb and more than $0.2\,mg\,l^{-1}$ CN in the blood suggest that fire fumes were inhaled and are to be considered as signs of vitality. A positive result for met-Hb indicates that nitrous gases were formed during the fire and were inhaled. As sole cause of death, elevated concentrations of CO-Hb and CN have to be considered especially in victims showing no or only minor heat changes on the body.

In deaths following flash fires, pulmonary tissue should also be kept as evidence to demonstrate the use of fire accelerants [45, 46]. As these substances are volatile, it is advisable to store the samples in airtight containers and to investigate them rapidly by means of headspace capillary gas chromatography mass spectrometry (GC-MS). A positive result for fire accelerants in the pulmonary tissue is also regarded as a sign of vitality.

Diagnosis of Vitality

Autopsy and Toxicological Findings. The signs of a vital exposure to the fire are summarized in Table 3. External signs that the body was exposed to heat when the victim was still alive may be the absence of burns and/or soot deposits in the corners of the eyes (so-called crow's feet), incompletely singed eyelashes, which result from squinting the eyes and, under certain conditions, petechial hemorrhages in the conjunctivae [18]. However, these findings are usually discernible only if the body was not, or was only slightly, consumed by the fire. As this is rare in deaths due to heat exposure, the internal findings are much more important than the external findings.

The most important constellation of findings to prove that the victim was alive during a fire is the combination of a CO-Hb concentration above 10% and soot deposits in the respiratory tract, the

Table 3 Signs of vitality after exposure to fire fumes versus heat

Exposure to fire fumes	Exposure to heat
Macroscopic findings	
Soot deposits in the airways	Crow's feet, incompletely singed eyelashes
Soot deposits in the esophagus	Skin blisters
Soot deposits in the stomach	Vesicular/patchy detachment of the pharyngeal mucosa or epiglottis; edematous swelling of the epiglottis; vesicular/patchy detachment of the upper esophageal mucosa
Histological findings	
Soot deposits in trachea and bronchi	Vesicular detachment of the tracheal and bronchial mucosa; pseudogoblet cells; massive secretion of mucus; nucleic elongation and palisade arrangement of the mucosal epithelium in trachea and bronchi; hyperemia and edema of the tracheal and bronchial mucosa
Toxicological findings	
CO-Hb concentration >10 %	
Cyanide concentration $>0.2\,ng\,ml^{-1}$	

esophagus, and the stomach [21] (Figures 6 and 7). In fact, these parameters show only whether the victim was exposed to *fire fumes* while alive; they are not indicative of vital *heat* exposure. Consequences of an inhalation of hot gases may be edematous swelling and vesicular or patchy detachment of the mucosa in the pharynx, the larynx, and/or the upper section of the esophagus [21, 24].

Although there is a certain relation between the vitality parameters of soot aspiration and CO-Hb, there is no reliable correlation [47, 48]. Deposits of soot in the airways should never lead to the premature conclusion that the victim died of intoxication by fire fumes. Soot particles may, under certain circumstances, occur in the respiratory tract also if burning was exclusively after death. On the other hand, the absence of soot in the respiratory tract does not allow drawing the conclusion that exposure to the fire occurred after death. Authors have pointed out repeatedly that the signs of vitality may be partly absent in cases of undoubtedly vital burning [14, 16, 21, 43, 49–52]. The significance of one vitality parameter alone is therefore limited.



Figure 6 Soot aspiration as a vital sign [Reproduced from Ref. 21. © Elsevier, 2003.]



Figure 7 Soot deposits in the esophagus as a vital sign [Reproduced from Ref. 21. © Elsevier, 2004.]

Histology. The histological signs of vitality are also summarized in Table 3. Histological examination of the respiratory tract is especially valuable to demonstrate changes due to the inhalation of hot gases. However, it should be remembered in this context that certain findings, e.g., nucleic elongation and palisade arrangement in the epithelium of the respiratory tract, may also be due to postmortem effects of heat on the body. On the other hand, other findings such as hyperemia and edema of the mucosa of the respiratory passages as well as

interstitial and intra-alveolar pulmonary edema are unspecific. Detachment of the mucosa from the epiglottis, the trachea, and the bronchi may also be due to autolysis and decomposition, although rarely in the vesicular form characteristic of thermal effects [20, 24]. The histological parameters should also be assessed only in context. According to our experience, the combination of pseudogoblet cells, increased secretion of mucus, and vesicular detachment of the epithelium is highly indicative of vital effects of heat on the respiratory system [21]. Histological findings of the pulmonary tissue in cases of rapid death may be acute congestion, interstitial and sectional intra-alveolar pulmonary edema, acute emphysema, as well as microthrombi.

Determination of the Cause of Death

The issues of vitality and cause of death are closely linked. For example, a high CO-Hb concentration in corpse blood supplies information both regarding the cause of death and the vitality. Apart from the signs of vitality, the absence of other causes of death is an essential condition for the conclusion that death was due to the fire. The diagnosis of a fire to cover up a homicide requires the presence of fatal injuries inflicted by another person *and* the absence of the classical parameters of vitality [33, 53–55].

The presence of mechanical lesions or severe pathological findings together with (seemingly) absent vital parameters may cause diagnostic problems. While in the presence of potentially fatal injuries (e.g., polytrauma following car accidents) perimortal burning has to be discussed, differentiation between peracute death due to the effects of heat and purely postmortem burning may be very difficult. In cases of this type, the CO-Hb concentrations in the corpse blood may be below 10%; the oral, pharyngeal, and laryngeal mucosa may show no heat damage; and there may be no soot aspiration. The most likely pathophysiological mechanism is a fulminant shock [24, 56–60]. But in some cases, the questions regarding the cause of death and the vitality cannot be answered definitely [14, 21, 47, 48].

Scalds

Scalds are the result of the effect of moist heat, usually hot water and/or steam, while scalding by other hot liquids, especially hot oil, is a comparatively rare picture to be examined in forensic practice. As the thermal conductivity of water and steam is much higher, they produce worse injuries than dry heat even after short exposure, as from the very beginning the actual skin temperature is considerably higher with moist heat and moist heat penetrates to deeper layers of the tissue much better than dry heat [61–64].

An actual skin temperature of 44 °C is regarded as the lowest temperature needed to cause damage to the skin; at this temperature, a second- to third-degree scald would be reached after an exposure time of 6 h [61–63]. Between 44 and 51 °C, a temperature increase of 1 °C reduces the exposure time needed for a certain degree of skin injury by half. Above 51 °C, there is no convective heat transport via the skin capillaries any more and the heat penetrates into the deeper tissue layers.

External Findings

Typical for the effect of moist heat is a uniform injury pattern of all the affected skin areas with sharply demarcated edges. Pouring a hot liquid over the skin may produce a streaklike trickle pattern. Scalds due to immersion are characterized by a straight, horizontal burn pattern corresponding to the level of the liquid. With tight-fitting clothes, the underlying skin areas may often be spared from scalding. As the temperatures applied by moist heat are usually lower than those of dry heat (water boils at 100 °C), moist heat does not produce fourth-degree skin burns (charring), heat-related exposures of body cavities, or amputations.

Internal Findings

The internal findings are unspecific. They are consequences of the shock associated with scalding [24], which can also be regarded as the cause of death. The phenomena seen after the inhalation of hot steam have been extensively described in the literature for deaths following a boiler explosion [24].

If hot steam is inhaled, the temperature hardly declines along the air passages, so that a direct thermal damage may occur even in the peripheral parts of the respiratory tract [22]. For differentiation between hot steam and dry air, the so-called “pleura sign” can be used at autopsy. If hot steam was inhaled, the parietal pleura is reddened, whereas the



Figure 8 Fourth-degree burns in a 52-year-old worker after explosion of an acid tank

costodiaphragmatic angles are pale [24]. In deaths caused by the effects of dry heat, this sign is absent even if the circumstances suggest a fire with high temperatures.

Child Abuse

Scald injuries in infants and small children are no rare events. Between 5 and 14% of the scalds occurring in children are claimed to be due to physical abuse. In scalds due to immersion, this percentage ranges probably even between 12 and 55% [65]. In all cases, it is therefore essential to clarify whether the injuries were caused by another person or could be the result of an accident.

If there are discrepancies between the localization and shape of the scalds on one hand and the given history of their origin on the other, if it is denied that they were caused by heat, if it is claimed that they were caused by other persons below the age of criminal responsibility, if pain during or after the injury is played down or denied, or if no doctor is consulted or is consulted only after a delay, all this may be indicative of child abuse.

Chemical Injury

Skin burns by chemical substances are rarely mentioned in the forensic literature. They are more frequently described under clinical aspects and occur mainly as household and industrial accidents. Chemical burns of the respiratory and gastrointestinal tract are seen both in accidents and suicides involving the intake of acid and alkaline solutions. Occasionally, child abuse is committed by forcing children to drink such substances.

The extent of the damage depends on the type of the substance taken, its concentration, amount, and time of exposure. The effect continues until the agent is removed or neutralized. Essentially, the damage is caused by coagulation (acid) or colliquation (alkaline solution) of the tissue. Alkaline solutions (most of which with a pH > 11) often produce worse injuries than acids (most of which with a pH < 2), as they liquefy the tissue and this process is not limited by protein precipitation.

Chemical injuries can manifest themselves as first- to third-degree burns. In exceptional cases they assume the appearance of fourth-degree burns (Figure 8).

As with scalding injuries, the internal findings away from the affected regions are due to shock reactions and thus unspecific.

References

[1] Bohnert, M. & Pollak, S. (2003). Heat-mediated changes to the hands and feet mimicking washerwoman's skin, *International Journal of Legal Medicine* **117**, 102–105.

[2] Bohnert, M., Faller-Marquardt, M. & Pollak, S. (2001). Zum unterschiedlichen Brandverhalten von Kopf- und Schamhaaren, *Archiv für Kriminologie* **207**, 42–48.

[3] Suarez-Penaranda, J.M., Munoz, J.I., Lopez de Abajo, B., Vieira, D.N., Rico, R., Alvarez, T. & Concheiro, L. (1999). Concealed homicidal strangulation by burning, *American Journal of Forensic Medicine and Pathology* **20**, 141–144.

[4] DeHaan, J.D. (1996). The dynamics of flash fires involving flammable hydrocarbon liquids, *American Journal of Forensic Medicine and Pathology* **17**, 24–31.

[5] DeHaan, J.D., Campbell, S.J. & Nurbakhsh, S. (1999). Combustion of animal fat and its implications for the consumption of human bodies in fires, *Science & Justice* **39**, 27–38.

[6] DeHaan, J.D. & Nurbakhsh, S. (2001). Sustained combustion of an animal carcass and its implications for the consumption of human bodies in fires, *Journal of Forensic Sciences* **46**, 1076–1081.

[7] Benecke, M. (1998). Spontaneous human combustion, *Skeptical Inquirer* **22**, 47–51.

[8] Gromb, S., Lavigne, X., Kerautret, G., Grosleron-Gros, N. & Dabadie, P. (2000). Spontaneous human combustion: a sometimes incomprehensible phenomenon, *Journal of Clinical Forensic Medicine* **7**, 29–31.

[9] Madea, B. (1992). Branddauer und verkohlungsgrad einer brandleiche, *Archiv für Kriminologie* **189**, 39–47.

[10] Günther, H. & Schmidt, O. (1953). Die Zerstörung des menschlichen Gebisses im Verlauf der Einwirkung hoher Temperaturen, *Deutsche Zeitschrift für die gesamte Gerichtliche Medizin* **42**, 180–188.

[11] Bohnert, M., Rost, T., Faller-Marquardt, M., Ropohl, D. & Pollak, S. (1997). Fractures of the base of the skull in charred bodies–post-mortem heat injuries or signs of mechanical traumatisation? *Forensic Science International* **87**, 55–62.

[12] Bohnert, M., Rost, T. & Pollak, S. (1998). The degree of destruction of human bodies in relation to the duration of the fire, *Forensic Science International* **95**, 11–21.

[13] Eckert, W.G., James, S. & Katchis, S. (1988). Investigation of cremations and severely burned bodies, *American Journal of Forensic Medicine and Pathology* **9**, 188–200.

[14] Gerling, I., Meissner, C., Reiter, A. & Oehmichen, M. (2000). Death from thermal effects and burns, *Forensic Science International* **115**, 33–41.

[15] Glassman, D.N. & Crow, R.M. (1996). Standardization model for describing the extent of burn injury to human remains, *Journal of Forensic Sciences* **41**, 152–154.

[16] Maxeiner, H. (1988). Umstände und befunde bei 202 brandtodesfällen, *Beiträge zur Gerichtlichen Medizin* **46**, 313–325.

[17] Bohnert, M., Schmidt, U., Große Perdekamp, M. & Pollak, S. (2001). Zum ausmass der brandzehrung–eine analyse von 68 brandleichen, *Archiv für Kriminologie* **207**, 104–113.

[18] Scharschmidt, A. & Bratzke, H. (1988). "Stauungsblutungen" als Brandfolge? *Archiv für Kriminologie* **182**, 94–100.

[19] Maxeiner, H. (1988). Blutaustritte im Kopf- und Halsbereich beim Verbrennungstod, *Zeitschrift für Rechtsmedizin* **101**, 61–80.

[20] Janssen, W. (1984). Injuries caused by heat and cold, in *Forensic histopathology*, Janssen, W, ed. Springer, Berlin, pp. 234–260.

[21] Bohnert, M., Werner, C.R. & Pollak, S. (2003). Problems associated with the diagnosis of vitality in burned bodies, *Forensic Science International* **135**, 197–205.

[22] Moritz, A.R., Henriques, F.C. & McLean, R. (1945). The effects of inhaled heat on the air passages and lungs, *The American Journal of Pathology* **21**, 311–331.

[23] Hagedorn, M., Pfrieme, B., Mittermayer, C. & Sandritter, W. (1975). Intravitale und pathologisch-anatomische Beobachtungen beim Verbrennungsschock des Kaninchens, *Beiträge zur Pathologie* **155**, 398–409.

[24] Brinkmann, B., Kleiber, M., Koops, E. & Püschel, K. (1979). Vitale Reaktionen bei akutem Verbrühungstod, *Zeitschrift für Rechtsmedizin* **83**, 1–16.
[25] Schneider, V., Pietrzak, T. & Klöppel, I. (1986). Postmortale Magen-Darm-Rupturen bei Brandleichen, *Archiv für Kriminologie* **177**, 29–33.
[26] Murray, K.A. & Rose, J.C. (1993). The analysis of cremains: a case study involving the inappropriate disposal of mortuary remains, *Journal of Forensic Sciences* **38**, 98–103.
[27] Grevin, G., Bailet, P., Quatrehomme, G. & Ollier, A. (1998). Anatomical reconstruction of fragments of burned human bones: a necessary means for forensic identification, *Forensic Science International* **96**, 129–134.
[28] Cattaneo, C., DiMartino, S., Scali, S., Craig, O.E., Grandi, M. & Sokol, R.J. (1999). Determining the human origin of fragments of burnt bone: a comparative study of histological, immunological and DNA techniques, *Forensic Science International* **102**, 181–191.
[29] Bohnert, M., Schmidt, U., Große Perdekamp, M. & Pollak, S. (2002). Diagnosis of a captive-bolt injury in a skull extremely destroyed by fire, *Forensic Science International* **127**, 192–197.
[30] de Gruchy, S. & Rogers, T.L. (2002). Identifying chop marks on cremated bone: a preliminary study, *Journal of Forensic Sciences* **47**, 933–936.
[31] Herrmann, B. (1977). On histological investigations of cremated human remains, *Journal of Human Evolution* **6**, 101–103.
[32] Bradtmiller, B. & Buikstra, J.E. (1984). Effects of burning on human bone microstructure: a preliminary study, *Journal of Forensic Sciences* **29**, 535–540.
[33] Iwase, H., Yamada, Y., Ootani, S., Sasaki, Y., Nagao, M., Iwadate, K. & Takatori, T. (1998). Evidence for an antemortem injury of a burned head dissected from a burned body, *Forensic Science International* **94**, 9–14.
[34] Spitz, W.U. (2006). Thermal injuries, in *Spitz and Fisher's medicolegal investigation of death*, 4th Edition, W.U. Spitz & D.J. Spitz, eds, Charles C Thomas, Springfield, Illinois, pp. 747–782.
[35] Hausmann, R. & Betz, P. (2002). Thermally induced entrance wound-like defect of the skull, *Forensic Science International* **128**, 159–161.
[36] Kondo, T. & Ohshima, T. (1994). Epidural herniation of the cerebral tissue in a burned body: a case report, *Forensic Science International* **66**, 197–202.
[37] Dirnhofer, R. & Ranner, G. (1982). Intracerebrale blutungen bei einer brandleiche – brandhämatom, bergungsverletzung oder intravitale entstehung, *Archiv für Kriminologie* **170**, 165–172.
[38] Schneider, V. (1982). Bemerkenswerte intracranielle befunde bei einer brandleiche, *Archiv für Kriminologie* **169**, 129–139.
[39] Schulz, F., Petri, S., Koops, E. & Matschke, J. (1996). Zur Frage der Vitalität und möglichen Ursachen von punktförmigen Blutungen im unteren Hirnstamm bei einer Brandleiche, *Archiv für Kriminologie* **198**, 160–166.
[40] Dotzauer, G. & Jacob, H. (1952). Über hirnschäden unter akutem verbrennungstod, *Deutsche Zeitschrift für gerichtliche Medizin* **41**, 129–146.
[41] Anderson, R.A., Watson, A.A. & Harland, W.A. (1981). Fire deaths in the Glasgow area: II. The role of carbon monoxide, *Medicine, Science and Law* **21**, 288–298.
[42] Anderson, R.A. & Harland, W.A. (1982). Fire deaths in the Glasgow area: III. The role of hydrogen cyanide, *Medicine, Science and Law* **22**, 35–44.
[43] Schwerd, W. & Schulz, E. (1978). Carboxyhaemoglobin and methaemoglobin findings in burnt bodies, *Forensic Science International* **12**, 233–235.
[44] Saukko, P.J. & Knight, B. (eds) (2004). *Knight's Forensic Pathology*, 3rd Edition, Arnold, London.
[45] Schuberth, J. (1994). Post-mortem test for low-boiling arson residues of gasoline by gas chromatography-ion-trap mass spectrometry, *Journal of Chromatography B: Biomedical Applications* **662**, 113–117.
[46] Schuberth, J. (1997). Gas residues of engine starting fluid in postmortem sample from an arsonist, *Journal of Forensic Sciences* **42**, 144–147.
[47] Rogde, S. & Olving, J.H. (1996). Characteristics of fire victims in different sorts of fire, *Forensic Science International* **77**, 93–99.
[48] Gormsen, H., Jeppesen, N. & Lund, A. (1984). The causes of death in fire victims, *Forensic Science International* **24**, 107–111.
[49] Betz, P., Roider, G., Meyer, Lv., Drasch, G. & Eisenmenger, W. (1996). Carboxyhemoglobin blood concentrations in suicides by fires, *Medicine, Science and Law* **36**, 313–316.
[50] Hirsch, C.S. & Adelson, L. (1969). Absence of carboxyhemoglobin in flash fire victims, *Journal of the American Medical Association* **12**, 2279.
[51] Wilk, E., Lindner, D. & Vock, R. (1999). Untersuchungen zum Inhalations-Hitzeschock, *Archiv für Kriminologie* **203**, 159–169.
[52] Wirthwein, D.P. & Pless, J.E. (1996). Carboxyhemoglobin levels in a series of automobile fires. Death to crash or fire? *American Journal of Forensic Medicine and Pathology* **17**, 117–123.
[53] Tsaroom, S. (1996). Investigation of a murder case involving arson, *Journal of Forensic Sciences* **41**, 1064–1067.
[54] Copeland, A.R. (1985). Homicide by fire, *Zeitschrift für Rechtsmedizin* **95**, 59–65.
[55] Fanton, L., Jdeed, K., Tilhet-Coartet, S. & Malicier, D. (2006). Criminal burning, *Forensic Science International* **158**, 87–93.
[56] Soejima, K., Schmalstieg, F.C., Sakurai, H. & Traber, D.L. (2001). Pathophysiological analysis of combined burn and smoke inhalation injuries in sheep, *American Journal of Physiology and Lung Cell Molecular Physiology* **280**, 1233–1241.
[57] Westphal, M., Morita, N., Enkhbaatar, P., Murakami, K., Traber, L. & Traber, D.L. (2004). Acute effects of

combined burn and smoke inhalation injury on carboxyhemoglobin formation, tissue oxygenation, and cardiac performance, *Biochemical & Biophysical Research Communications* **317**, 945–949.
[58] Cox, R.A., Burke, A.S., Oliveras, G., Enkhbaatar, P., Traber, L.D., Zwischenberger, J.B., Jeschke, M.G., Schmalstieg, F.C., Herndon, D.N., Traber, D.L. & Hawkins, H.K. (2005). Acute bronchial obstruction in sheep: histopathology and gland cytokine expression, *Experimental Lung Research* **31**, 819–837.
[59] Marschall, S., Rothschild, M.A. & Bohnert, M. (2006). Expression of heat-shock protein 70 (Hsp70) in the respiratory tract and lungs of fire victims, *International Journal of Legal Medicine* **120**, 355–359.
[60] Weis, A. & Bohnert, M. (2008). Expression patterns of adhesion molecules P-selectin, von Willebrand factor and PECAM-1 in lungs. A comparative study in cases of burn shock and hemorrhagic shock, *Forensic Science International* **175**, 102–106.
[61] Moritz, A.R. (1947). Studies of thermal injury III. The pathology and pathogenesis of cutaneous burns. An experimental study, *The American Journal of Pathology* **23**, 915–941.
[62] Moritz, A.R. & Henriques, F.C. (1947). Studies of thermal injury II. The relative importance of time and surface temperature in the causation of cutaneous burns, *The American Journal of Pathology* **23**, 695–720.
[63] Moritz, A.R., Henriques, F.C., Dutra, F.R. & Weisiger, J.R. (1947). Studies of thermal injury IV. An exploration of the casualty-producing attributes of conflagrations; local and systemic effects of general cutaneous exposure to excessive circumambient (air) and circumradiant heat of varying duration and intensity, *The American Journal of Pathology* **43**, 466–488.
[64] Price, P.H., Call, D.E., Hansen, F.L. & Zerwick, C.J. (1953). Penetration of heat in thermal burns, *Surgical Forum* **4**, 433–438.
[65] Renz, B. & Sherman, R. (1993). Abusive scaled burns in infants and children, *American Surgery* **59**, 329–334.

MICHAEL BOHNERT

Ink Analysis

The Different Classes of Inks

The inks most commonly encountered on documents today are writing inks and printing inks [1, 2]. Writing inks can subdivided into at least four classes – the thick (high viscosity) inks such as those used in ballpoint pens; the fluid (less viscous) inks such as those used in rolling ball, porous plastic nib, extruded plastic tip, and felt/fiber tip pens; the gel inks; and other inks such as erasable, "dry erase", and crayon inks.

Printing inks can be divided into at least two classes – those used in impact printers such as printing presses and typewriters and those used in nonimpact printers such as desktop printers, photocopiers, and thermal printers. Those used in printing presses include offset, intaglio, letterpress (typographic), gravure, and flexographic inks while those used in desktop printers include ink-jet inks and laser jet "inks" (toners). The toners used in laser jet printers are also used in photocopiers. Black and white desktop printers use black toners while color desktop printers use color toners (usually cyan, yellow, magenta, and black, also referred to as *CYMK*).

The General Composition of Inks

All inks, including those that date back hundreds of years, consists of a colorant or mixture of colorants in a vehicle (sometimes called *a carrier, a base*, or, *in printing inks*, *a varnish*). The colorants in fluid writing inks are dyes (dyes, as opposed to pigments, are soluble in the vehicle). The colorants in viscous writing inks are mostly dyes but may include some pigments, particularly carbon black (pigments are dispersed in the vehicle). The colorants in gel ink writing were originally colored pigments but now are a combination of colored pigments and dyes.

The colorants in printing inks are pigments dispersed in a vehicle except in at least the following cases: some color ink-jet printing inks contain both dissolved dyes and dispersed pigments; some typewriter ribbon inks contain dyes mixed with the pigment carbon black in a thick slurry; and thermal printers use ribbons containing thermal dye deposited on a film (the printing process is referred to as *thermal dye transfer*).

The vehicle or carrier of any ink consists mostly of a solvent or mixture of solvents and other components. For example, ballpoint pen inks contain solvents like 2-phenoxyethanol or benzyl alcohol or a mixture of these along with resins that thicken the ink (resins give "body" to the ink and assist in making the colorant adhere to the surface). They may also contain other components, albeit at a much lower level, such as humectants or corrosion inhibitors. Ballpoint pen inks consist of approximately 25% dyes, 50% solvents, and 25% resins.

In printing ink technology, the resin is sometimes called the *binder* and one of its functions is to get the colorants to adhere to the substrate (e.g., paper). Some binders are film-forming resins such as ethylcellulose and nitrocellulose (used in screen printing inks) [3]. Some minor (i.e., low concentration), but critical components of printing inks are the drying agents.

The Reasons for Performing Ink Analysis

The forensic scientist analyzes inks for at least three reasons:

1. to determine if two inks are the same or different (and if they are the same, how rare or common the ink is);
2. to determine the origin, source, or manufacturer of an ink; and
3. to determine the date of an ink (either when that ink was first produced or, ideally and much more difficult, when that ink was placed on a document).

All these determinations, except one, require the use of reference standards. The one exception is when inks being compared are found to be different.

Of the three determinations, the most commonly performed is the comparison of inks; however, the one has received the most attention due to its difficulty is the dating of inks, particularly writing inks. There are three progressively difficult levels of doing this.

The first is to determine the "date" of the ink by following when major changes in ink formulations first occurred (e.g., the introduction of glycol-based ballpoint pen inks, of copper phthalocyanine dyes in ballpoint pen inks, of water-based inks, and of gel inks). For this, the ink is compared with reference standards representing historical changes in ink production. One often consults a table of such historical changes [1].

The second is to determine the first manufacturing date of the ink or when certain components were first used in ink. For this, the ink is compared with a comprehensive collection of reference standards (discussed below).

The third, and most challenging level, is to determine when the ink was placed on a document. One approach to estimate the age of a single ink entry is to compare the ink with itself before and after inducing age (e.g., by heating). Another way is for the questioned ink to meet a stringent, but important requirement; namely, that it be on a document bearing ink entries of known dates and made with the same ink formula as that of the questioned ink. If this is met, then one can determine the relative age of these inks and, depending on the known dates of the other entries, find date bounds that bracket the age of the questioned ink. Both the *estimated age of a single ink entry* and the *relative age of inks entries* (of the same formula and on the same paper) are discussed below.

The Need for Reference Standards

On the basis of the discussion above, not much can be done (other than to say that two inks are different) without the use of reference standards. For writing inks, dried samples of different formulae can usually be distinguished by a set of established analytical procedures [4, 5]. Samples of different batches of the same writing ink are normally indistinguishable by this set of procedures. This is because the manufacturing of writing ink requires strict adherence to its formula (recipe); any deviation can cause the ink to fail in its performance (e.g., in the way the pen writes). Writing ink formulas are also meant to be stable over the average lifetime of their use, particularly if they are in a container such as a cartridge. However, such inks tend to "age" (fade, decompose, etc.) after they are placed on paper, exposed to the elements. Andrasko and Kunicki [6] also observed that some aging (ink drying and dye decomposition) occurs near the tip of a ballpoint pen. For nonwriting inks, particularly for ink-jet inks, thermal printing inks, and toners, the same approach of collecting different formulations is normally taken for building reference collections.

There are two types of reference standards: requested standards and acquired standards. Requested standards are those that are obtained from ink manufacturers through formal requests and acquired standards are those obtained from pens, ink samples, etc., acquired, for example, from stationery stores, businesses with promotional pens, or personal collections of pens or pen writings. When requesting samples of different ink formulas from the ink industry, one always asks about the existence of formula variations via component substitution and, if

any exist, requests to see these samples if differences can be found at the forensic laboratory level (using the established set of analytical procedures).

The forensic laboratory of the United States Secret Service has the largest collection of reference samples of writing ink formulas in the world. It is known as the *International Ink Library*. Yearly, this laboratory solicits samples of new, improved, or changed ink formulas from the ink industry in the United States and most of those abroad. As of 2007, the collection consisted of over 8500 inks. It consists of four parts: the *company files* which contain manufacturing information such as the first date of production of inks; the *ink samples* which may be in bulk, in cartridges, or in writing; the *scribble sheets* which contain the dried ink on paper; and the *thin-layer chromatograms* (TLCs) which separate their dye content.

It is critical that this collection be maintained up to date in order to make forensically significant statements about what a match between a questioned ink and a standard reference ink means [4, 5]. If a questioned sample, for example, is matched to a reference standard, then one cannot eliminate the possibility that the questioned sample is the same as the reference standard. By a match between A and B is meant that their analytical profiles (determined using a given set of analytical procedures) are indistinguishable.

Features That Do Not Tend to Change with Age

Ink formulas in closed systems (e.g., cartridges) tend to be stable over the lifetime of the ink's use. Once they are exposed to air and allowed to dry, certain features change with age (covered below), but others such as their elemental profiles, optical properties, and certain colorants are more resistant. The dried inks on the scribble sheets are sufficiently stable to be used for comparisons (their stability is nevertheless periodically checked). The established set of analytical procedures used to distinguish these inks [4, 5] provides the *analytical profile of these inks*. These profiles consist of three parts: the *physical*, the *optical*, and the *chemical features* of the inks. The physical features include the type of ink (i.e., one determines whether the ink comes from a ballpoint pen or a nonballpoint pen), a feature readily determined by observing the appearance of a written line under magnification. The optical features include the ink's color, near infrared absorbance, near infrared luminescence, and ultraviolet fluorescence. The chemical features include the ink's thin-layer chromatogram viewed under visible light, ultraviolet radiation, laser illumination (with filtered viewing), and in the near infrared luminescence mode.

There are two levels of TLC examinations: the library search level (level 1) and the higher resolution level (level 2). The first involves TLC plates that are highly reproducible (due to their low sensitivity to changes in ambient or developing chamber conditions), but, as a trade of, have low resolution. These are used for developing the fourth part of library (the TLCs) and for searching the library for matches via plate-to-plate comparisons. The second involves TLC plates that are highly discriminating (high resolution), but are also sensitive to external changes. These are used to further discriminate inks that were indistinguishable at the library search level; this is done via within-plate comparisons.

To address the comparison, sourcing, and dating of an ink using a collection of reference standards, one first obtains a (level 1) analytical profile of the ink (this requires removing inks for TLC examination). This profile is compared with those from inks in the collection. All those that match (i.e., are indistinguishable) are further sorted by removing additional sample from the questioned ink and from the scribble sheets corresponding to the matching standards and performing the level 2 examination in a single plate. If a single match is found, the questioned ink is said to bear the same characteristics (at the given level of examination) as the matching standard and, therefore, cannot be eliminated as being that ink. The match approaches an identity as the collection becomes more complete and the set of analytical procedures become more discriminating. If there is information from the manufacturer of a unique component (that no other ink has), then the match is an identity. The unique tagging of inks assists in this endeavor [1].

Features That Tend to Change with Age

Once ink is exposed to the elements, then, like paints, they begin to dry [7, 8]. This drying process primarily involves the evaporation of solvents, but it can also involve the oxidation of resins, the penetration of the ink into the paper, and possibly other interactions

of the ink with the paper's chemistry and physics. Overtime, there could also be natural degradation of ink dyes or degradation induced by heat or light.

Aginsky [9] found that inks that contain high boiling solvents and a binder such as a resin (even at low levels) have the property that their solvents do not readily evaporate. In essence, as such an ink dries via solvent evaporation, it thickens and its solvents evaporate more slowly. Thus, solvent evaporation along with ink setting is the major aging process in ballpoint pen inks that is currently extensively studied. This drying process affects several measurable features (aging parameters) such as how much ink solvent remains in the ink, how much ink solvent is extractable into organic solvents of varying strength, and how fast and to what extent the ink dyes are extractable into weak extracting solvents. The section below summarizes some of these and those due to dye degradation. It should be noted that these can be used for determining the *relative age of inks* (where several inks of the same formula and on the same document are compared for their age) and the *estimated age of a single ink entry*; the latter usually involves the accelerated aging approach (where two samples of the same ink, one of which is heated, are compared).

To be useful (e.g., for comparing inks of different age), these measurements must have no dependence on the amount of ink sampled. One way of achieving this *mass invariance* is to take ratios of measurements (e.g., measurements of the same sample at different wavelengths, retention times, temperatures, etc.) each of which depends linearly on the same amount sampled [2]. Furthermore, to determine if two inks are of the same or different age, one needs to make multiple measurements, average these, and use *statistical methods* such as *significance tests* to determine if there are significant differences between the two means (averages) or not [1, 2].

Sketch of Some Ink Aging Parameters

Aging Process: Solvent Evaporation plus Ink Setting

Extraction of ink volatile components into a solvent

- Vehicle-to-dye and vehicle-to-vehicle ratios [10].
- Percent extraction of solvents into a slow extracting solvent (compute difference between an unheated and heated ink specimen, each followed by a fast extracting solvent and the results are ratioed in such a way that a percent extraction into the slow extracting solvent is obtained) [9, 10]. Comment: This mass invariant method is the one most studied and used by Aginsky. To estimate the age of an ink, he implored the induced aging approach using heat.
- Amount of solvents extracted into a fast extracting solvent (compute percentage of solvent loss when comparing an equally sampled unheated and heated sample) [9–12]. Comment: Gaudreau and Brazeau [12] modified this Aginsky method by removing the ink sample from the cuvette before sampling the extract (i.e., the cuvette is free of the sample). This method is not mass invariant so equi-sampling is imperative.

Direct measurement of ink volatile components

- Thermal desorption of ink solvents (using solid phase microextraction SPME) [13, 14]. Comment: Andrasko [14] used a SPME holder as his thermal desorption device and used it to characterize inks. This simple device can also be used to make mass invariant measurements for example, by measuring the same sample at two different temperatures and taking a ratio of these.

Extraction of ink (or of individual dyes) into a weak solvent: [1, 2, 7, 8, 15–17]

- Rate of extraction
- Extent of extraction
- Time to extract X% (e.g., 90%) of ink (or a dye) Comment: The changes that this approach measures cease after three to four years while those of the mentioned above cease after one to two years.

Extraction of nonvolatile and noncolored vehicle components

- Solvent extraction [10]
- Thermal desorption using SPME [18]. Comments: Bügler *et al.* [18] characterized inks via their resin (which volatilize at high temperature) and solvent components; however, it can also be used to make mass invariant measurements (see comment above regarding the use of SPME by Andrasko [14]).

Surface reflectance vs. subsurface reflectance using polarizing light microscopy [11]

Comment: This surface vs. subsurface reflectance phenomenon is based on the ink drying (solvent evaporation).

Discoloration of ink when exposed to gaseous organic bases [10]

Comment: In this reversible phenomenon, as the ink ages (solvent evaporates) the ink becomes more resistant to discoloration.

Ratio of infrared peaks [19]

Comment: The changing infrared peaks involve the OH, CH, and CO groups and are believed to be from the evaporating solvent. For the ink studied, the ratio of these changes seems to cease after 10 to 20 years.

Aging Process: Dye Decomposition

Follow the demethylation process

Crystal Violet → Methyl Violet → Tetramethyl Para Rosaniline → others. Comment: Using high performance liquid chromatography (HPLC), Andrasko [20, 21] observed dye degradation for inks exposed to light, but also for inks stored in the dark (albeit the rate is slower). Using desorption ionization mass spectrometry (MS) Grim *et al.* [22–24] explored similar changes. Other who have applied desorption ionization MS include Weyermann *et al.* [25] and Ifa *et al.* [26] who basically used it to characterize inks.

References

[1] Cantu, A.A. (1995). A sketch of analytical methods for document dating, part I. The static approach: determining age independent analytical profiles, *International Journal of Forensic Document Examiners* **1**, 40–51.

[2] Cantu, A.A. (2009). The chemistry of fingerprint science and document examination, *Forensic Chemistry*, J. Almirall, ed, John Wiley and Sons, NY (in press).

[3] Dalwadi, D.H., Canet, C., Roye, N. & Hedman, K. (2005). Rheology: an important tool in ink development, *American Laboratory* November, 18–22.

[4] The American Society for Testing and Materials (ASTM) (1996). *Designation: E 1789-96: Standard Guide for Writing Ink Identification*, pp. 722–726.

[5] The American Society for Testing and Materials (ASTM) (1998). *Designation: E 1422-98. Standard Guide for Test Methods for Forensic Writing Ink Comparison*, pp. 530–537.

[6] Andrasko, J. & Kunicki, M. (2005). Inhomogeneity and aging of ballpoint pen inks inside of pen cartridges, *Journal of Forensic Sciences* **50**, 542–547.

[7] Cantu, A.A. & Prough, R.S. (1987). On the relative aging of ink – the solvent extraction technique, *Journal of Forensic Sciences* **32**, 1151–1174.

[8] Cantu, A.A. (1996). A sketch of analytical methods for document dating. part II. The dynamic approach: determining age dependent analytical profiles, *International Journal of Forensic Document Examiners* **2**, 192–208 (An erratum to this article is found in, *International Journal of Forensic Document Examiners*, 1996, **2**, 370–372).

[9] Aginsky, V.N. (1996). Dating and characterizing writing, stamp pad and jet printer inks by gas chromatography/mass spectrometry, *International Journal of Forensic Document Examiners* **2**, 103–116.

[10] Aginsky, V.N. (1993). Some new ideas for dating ballpoint inks – a feasibility study, *Journal of Forensic Sciences* **38**, 1134–1150.

[11] Aginsky, V.N. (1995). A microspectrophotometric method for dating ballpoint inks – a feasibility study, *Journal of Forensic Sciences* **40**, 475–478.

[12] Gaudreau, M. & Brazeau, L. (2002). Ink dating using a solvent loss ratio method, *Proceedings of the 60th Annual Conference of the American Society of Questioned Document Examiners*, San Diego.

[13] Brazeau, L. & Gaudreau, M. (2007). Ballpoint pen inks: the quantitative analysis of ink solvents on paper by solid-phase microextraction, *Journal of Forensic Sciences* **52**, 209–215.

[14] Andrasko, J. (2006). A simple microthermal desorption device, *Journal of Forensic Sciences* **51**, 925–928.

[15] Brunelle, R.L. & Read, R.W. (1984). *Forensic Examination of Ink and Paper*, Charles C. Thomas, Springfield.

[16] Brunelle, R.L. & Speckin, E. (1998). Technical report with case studies of the accelerated aging of ballpoint inks, *International Journal of Forensic Document Examiners* **4**, 240–254.

[17] Brunelle, R.L. & Crawford, K.R. (2003). *Advances in the Forensic Analysis and Dating of Writing Ink*, Charles C. Thomas, Springfield.

[18] Bügler, J.H., Buchner, H. & Dallmayer, A. (2005). Characterization of ballpoint pen inks by thermal desorption and gas chromatography-mass spectrometry, *Journal of Forensic Sciences* **50**, 1209–1214.

[19] Humecki, H. (1985). Experiments in ballpoint ink aging using infrared spectroscopy, in *Proceedings of the International Symposium on Questioned Documents, FBI Academy, Quantico, VA*, U.S. Government Printing Office, Washington, DC, pp. 131–135.

[20] Andrasko, J. (2001). HPLC analysis of ballpoint pen inks stored at different light conditions, *Journal of Forensic Sciences* **46**, 21–30.

[21] Andrasko, J. (2001). Changes in composition of ballpoint pen inks on aging in darkness, *Journal of Forensic Sciences* **47**, 324–327.
[22] Grim, D.M., Siegel, J. & Allison, J. (2001). Evaluation of desorption/ionization mass spectrometric methods in the forensic applications of the analysis of inks on paper, *Journal of Forensic Sciences* **46**, 1411–1420.
[23] Grim, D.M., Siegel, J. & Allison, J. (2002). Evaluation of laser desorption mass spectrometry and UV accelerated aging of dyes on paper as tools for the evaluation of a questioned document, *Journal of Forensic Sciences* **47**, 1265–1273.
[24] Grim, D.M., Siegel, J. & Allison, J. (2002). Does ink age inside a pen cartridge? *Journal of Forensic Sciences* **47**, 1294–1297.
[25] Weyermann, C., Marquis, R., Mazzella, W. & Spengler, B. (2007). Differentiation of blue ballpoint pen inks by laser desorption ionization mass spectrometry and high-performance thin-layer chromatography, *Journal of Forensic Sciences* **52**, 216–220.
[26] Ifa, D.R., Gumaelius, L.M., Eberlin, L.S., Manicke, N.E. & Cooks, R.G. (2007). Forensic analysis of inks by imaging desorption electrospray ionization (DESI) mass spectrometry, *Analyst* **132**, 461–467.

Related Articles

Forged and Counterfeit Documents
Ultimate Issue Evidence by Experts

ANTONIO A. CANTU

Ink Comparison and Interpretation

Introduction

The forensic examination of ink is usually performed for identification, comparison, or dating purposes [1–6]. The identification and the comparison processes aim at inferring the identity of the source [7] of an ink specimen, while the dating exercise focuses on the determination of the time of a particular entry on a document. The dating process is not a straightforward one and was discussed by many [8–14] and is still the subject of research. It is not the topic of this article, which focuses on comparison and identification of ink in forensic science.

The distinction between the identification and the comparison processes lies in the number of sources that are considered during the examination:

- The identification process aims to determine the source of a questioned specimen from a reference library [15]. In this process, a large number of potential sources are considered in a first step. On the basis of chosen characteristics of the questioned specimen, one (or several) control samples, similar to the questioned specimen, is (are) selected from the library. Information, which was collected on this (these) control sample(s), is then used to provide intelligence information in the course of the investigation.
- The comparison process endeavors to assess if a questioned specimen ink originates from one or a restricted number of particular sources [16]. In this process, a few sources are considered and the weight of the evidence represented by the (non-)differentiation between the questioned specimen and the control ink is evaluated and reported, usually for Court purposes.

Most of the principal techniques developed in the field of analytical chemistry were or are used to analyze inks in the forensic context. Recommended techniques focus on the detection and comparison of dye components and are described in [15, 16]. Other techniques focusing on the various components of inks, such as dyes, pigments, solvents, resins, preservatives, binders, surfactants, etc., were reported [17–19]. Neither the identification, nor the comparison process necessarily favors any of the various analytical techniques.

Definition of the Source of an Ink in the Forensic Context

There are several ways to define the source of an ink. They all depend to a great extent on the structure of the ink commercial market. The market is influenced by two phenomena: (i) emergence of new suppliers and new products and (ii) cost savings:

1. In the past decades, demand for writing instruments has generated an explosion of ink formulations, each of which may contain dozens of chemical components. In addition, the appearance of products manufactured in emerging

economies, in traditional Western markets, certainly contributes to the diversity of available ink formulations.

2. The very same increasing demand and globalization effect have pushed ink producers and writing instrument manufacturers to adopt highly efficient manufacturing processes to minimize production costs and maximize profits. In the last decade or so, many writing instrument manufacturers have merged or were bought by other companies. Notwithstanding, it is now very common for the inks and for the different components of writing instruments to be manufactured by different, unrelated and independent producers, which are also driven by economic factors. Ink producers are tempted (i) to minimize the development costs of an ink formula by selling batches of ink from the same formula to different writing instrument manufacturers and (ii) to reduce the production costs of a given ink as much as possible by replacing components of this ink formula with similar but cheaper ones if available. This is done providing that this manipulation does not affect the characteristics of the ink that are important to writing instrument manufacturers.
   Conversely, instrument manufacturers, influenced by market prices, might be tempted to put different inks in a given brand/model of instruments during the production life of a model as new, but cheaper, inks with the required characteristics become available.

Forensic ink examiners are hence faced with a considerable diversity of combinations between ink formulations and writing instrument brands/models when attempting to identify the source of questioned ink specimens. This variety is certainly an asset from the point of view of a forensic investigation; however, it renders the maintenance of ink databases and the interpretation of forensic ink analyses very challenging. Attempts have been made by some manufactures to introduce chemical tags in the composition of their ink, in order to identify them [20, 21]. However, the process was never really adopted by manufacturers because of costs and, in return, was never very successful because of the low adoption level among manufacturers.

From the description of the ink market, it can be seen that the source of an ink can be defined in various ways. The definition of what "source" is required (e.g., manufacturer, instrument) in a particular circumstances, impacts on both the comparison and the identification processes. In both processes, the meaning intended by the "source" defines the degree of detail of the answer provided by the examiner at the end of his examination. In an identification process, the significance given to the source of an ink constrains the construction of databases by defining precisely which control sources need to be referenced. In the comparison process, it defines the level of information on the ink market that needs to be available to the examiner (e.g., the number of manufacturers producing ink with the same formula, the number of production batches indistinguishable from each other, and the number of instruments carrying ink indistinguishable from each other).

Traditionally, the following meanings are attributed to the concept of source in forensic ink examination:

1. The source can be defined as an ink formulation. The demonstration of the identity of source of ink samples in this context is purely a matter of qualitative identity (i.e., the similarity of the intrinsic characteristics of the samples [7]) and is demonstrated by the nondifferentiation of the characteristics of their composition. In the identification process, samples from the different ink formulations need to be referenced in ink collections.
2. The source can be defined as an ink manufacturer. The demonstration of the identity of source of ink samples in this context is performed in two stages; at first the qualitative identity between the samples is confirmed, then the identity of the manufacturer is inferred from information such as the number of ink manufacturers using the same formulation, the size of their respective production, the geographical area, or the market shares of instrument manufacturers supplied by them, etc. In the identification process, samples from all the inks produced by the ink manufacturers need to be referenced. All changes of any ink component by any of the manufacturers need to be recorded in ink collections.
3. Since claims of the possibility to measure batch-to-batch variations have been reported [22–24], the source can also be defined as a particular production batch. The definition of the source of an ink sample is, in this context, a refined

situation of the definition in (2) above. The identity of source is inferred following the same logical path as before. At first the qualitative identity of two samples is demonstrated. In a second stage, the identity of the production batch is inferred. This inference takes into account the number of batches that remains identical to each other when using the considered analytical technique to detect batch-to-batch variations. In this situation, samples from every batch from all manufacturers need to be referenced in ink collections for identification purposes.

4. The source can be defined as a brand/model of writing instrument. As in (2) and (3) above, the identity of source of two samples is inferred from the qualitative identity between the two samples and from the information available on the number of instruments of this particular model/brand that contain an identical ink. The constitution of ink collections in this situation is far more complex than for the three aforementioned cases. Samples from all brands/models need to be collected. This needs to be performed only once in most situations. However, should the ink composition of any of the brand/model change, a new sample for that brand/model needs to be referenced.
5. Finally, the source can represent a particular instrument. As for (2), (3), and (4) mentioned earlier, the identity of source is inferred by demonstrating the qualitative identity between the two samples and from the number of writing instruments that have an ink similar to the samples and are available at the time when the questioned document was created. The constitution of reference collection, in such situation, would obviously prove to be impossible in an identification context, since one would have to sample every writing instrument during the manufacturing process.

Ink Identification

More practically, in ink identification, the source of a specimen is normally defined as the brand/model of an instrument as in (4) mentioned earlier, although information regarding formulation (as in 1) or the manufacturer (as in 2) is also often reported. The number of ink available to consumers has a direct consequence on the establishment and maintenance of ink collections. Furthermore, the possibility, either of writing instruments of different brands to have the same ink or of writing instruments from the same brand/model to have different inks, is decreasing the efficiency of the ink identification process. Most of the time, it appears impossible to obtain, retain, classify, and search exhaustive information on all of the brands/models of writing instruments and on their corresponding ink formulations, at any given time of their production lifetime. The very few reference ink collections, which have been set up over the past decades, prove to be far from exhaustive, although the size of some of them can be quite impressive [8, 25, 26]. Nonetheless, despite these difficulties, ink collections are still maintained and their technology upgraded [27]. Recent search simulations have shown a high identification rate for randomly selected specimen inks [28] and demonstrated that these collections are very useful tools to forensic investigations.

Ink Comparison

In a first step, the ink comparison process aims at verifying the qualitative identity of two ink sample. Such qualitative identity is based on the level of concordance of the characteristics of the samples and the possibility to explain the discordance between them. Modifications of inks on paper due to environmental conditions have been reported [13, 15, 16, 29–35] and needs to be taken into account when assessing the qualitative identity of two ink samples. This can be achieved empirically, or statistically by assessing the likelihood of the qualitative identity for a set of environmental conditions.

In a second step, the comparison process aims at weighting the evidence represented by the qualitative identity of two samples. It should be noted that in most cases, the absence of qualitative identity leads the simple rejection of the hypothesis that the two samples originate from the same source. The comparison process usually considers the source of an ink as being a particular instrument, as in (5) above. Obviously, the ideal result of the comparison process would be to link the questioned ink to the sole writing instrument under consideration to the exclusion of all others. Unfortunately, such a goal is unrealistic. While, in most cases, demonstrating the qualitative identity between ink from a questioned entry and an alleged source is a rather straightforward process,

linking writings on a document to a single given writing instrument, to the exclusion of all others, can only be achieved using some sort of statistical inference. Indeed, with respect to context in (5), modern mass-production processes result in the production of hundreds of thousands of writing instruments containing inks with the same properties, which, at first, cannot be ruled out as being the source of the questioned writings.

For example, a particular batch of ink may fill 100 000 pens. Every one of those pens would contain exactly the same ink, and thus it would be impossible to say that one of them was used to write the questioned document to the exclusion of all 99 999 others on the basis of the sole analysis of ink evidence. The evidential value notably depends on the total number of writing instruments produced containing similar inks. The weight of the evidence represented by a link established between a questioned ink specimen and a control ink, which was produced in a very small amount and used to fill only 100 000 pens, is more informative than a similar link, established between a questioned specimen and a blue ball-point pen that was manufactured at several million exemplars.

No frequency study has ever been reported on the availability of the various inks. From the structure of the ink market, it appears that data on the specificity of given inks is difficult to obtain. Indeed, a frequency study performed on a given market would not necessarily be representative of the structure of another geographical market and the measured frequencies are likely to vary significantly in time. Furthermore, the need to be able to measure, store, retrieve, and compare ink data at different times and locations requires a quality assurance system that is not available at the moment [27]. Finally, a frequency study would only provide a broad estimation of the specificity of a considered ink in a given case, since the evidential value really depends on the availability of the ink to the potential writers, at the time of the creation of the questioned document.

Since the specificity of the considered ink would need to be determined in every case, the last evaluative stage is usually not performed at all [16].

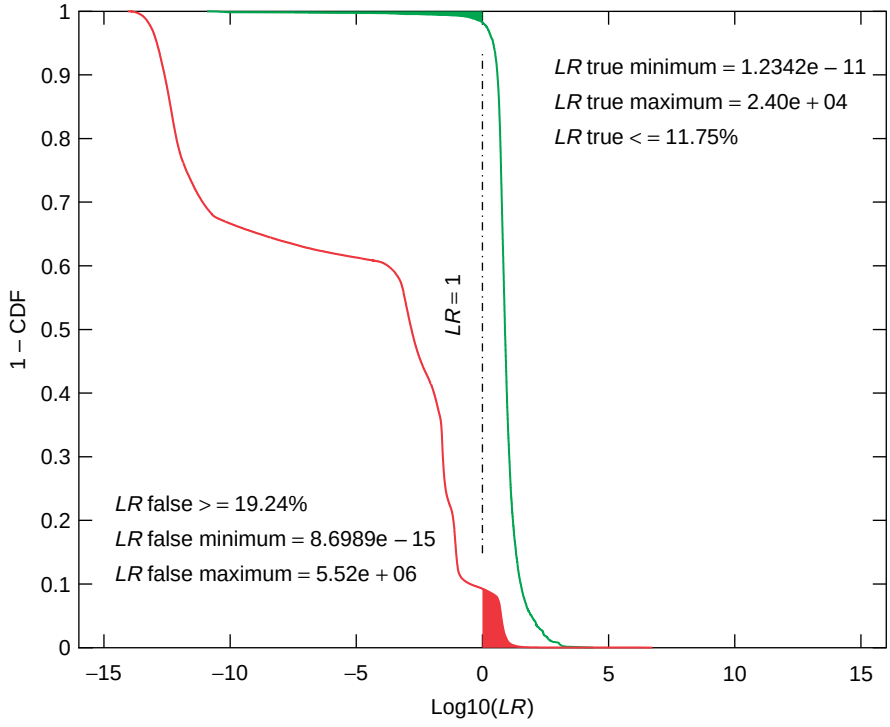


Figure 1 Simulations of the evidential value of a BIC® ink under various environmental conditions. It is possible to observe that the green curve is close to the separation line likelihood ratio ($LR = 1$), indicating a weak evidential value

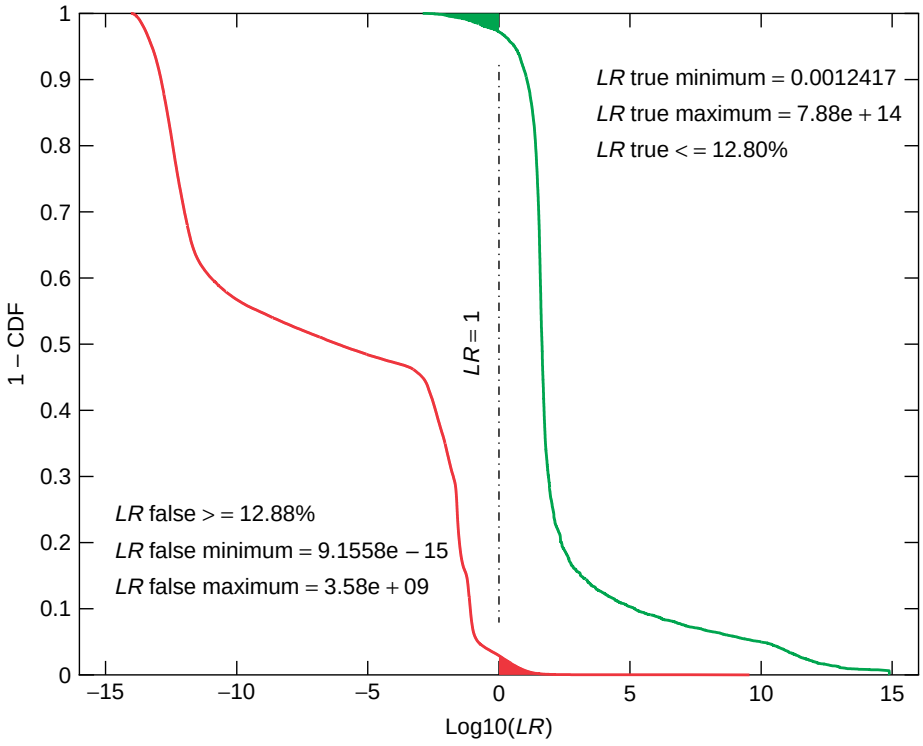


Figure 2 Simulations of the evidential value of a Pentel® ink under various environmental conditions. It is possible to observe that the green curve is located further away from the separation line ($LR = 1$), indicating, as a result, a generally higher evidential value

Thus, the result of a comparison between a questioned ink and a given writing instrument is reported as the (non-)differentiation of their respective chemical profile.

In very few cases, is it useful to define the source of an ink as in (1), (2), (3) or (4). However, in these contexts, the value of the information provided by the ink examination is limited given the number of corresponding instruments that can be considered.

Conclusion

Forensic ink analyses have been performed routinely for the last hundred years. Nevertheless, little has been written on the interpretation of the outcome of forensic ink analysis. The complexity of the relationships between ink producers and instrument manufacturers, reinforced by the rapid evolution of these relationships, has prevented the collection of exhaustive data. While such data could be used to inform probabilistic framework for the interpretation of ink evidence, such as the widely accepted likelihood ratio framework [36], the current absence of these data let forensic ink examiner with no other choice than to report the (non-)differentiation of questioned and control ink samples. Nonetheless, recent technological improvements, such as the digital acquisition and automatic comparison of ink data, render the collection and comparison easier, at different times and locations of these data [27]. This technology allows identifying ink specimen more efficiently and also the constitution of specific frequency datasets for the evaluation of ink evidence in comparison processes. The use of such technology already shows promising results. Indeed, simulations demonstrate that it is possible to measure the individual specificity of different inks, while taking into account environmental conditions (Figures 1 and 2).

References

[1] Ellen, D.M. (1997). *The Scientific Examination of Documents*, 2nd Edition, Taylor & Francis, London.

[2] Brunelle, R.L. & Reed, R.W. (1984). *Forensic Examination of Ink and Paper*, C.C. Thomas, Springfield.

[3] Breedlove, C.H. (1989). The analysis of ball-point inks for forensic purposes, *Journal of Chemical Education* **66**, 170–171.

[4] Brunelle, R.L. (2000). Ink analysis, in *Encyclopedia of Forensic Sciences*, J. Siegel, G. Knupfer & P. Saukko, eds, Academic Press, San Diego, pp. 591–597.

[5] Neumann, C. & Mazzella, W.D. (2004). Questioned document, in *Encyclopedia of Analytical Science*, P. Worsfold, A. Townshend & C. Poole, eds, Academic Press, pp. 465–471.

[6] ASTM E444-07 (2007). *Standard Guide for Scope of Work of Forensic Document Examiners*, ASTM International, West Conshohocken.

[7] Kwan, Q.Y. (1977). *Inference of Identity of Source*. PhD Thesis. Department of Forensic Science, University of California, Berkeley.

[8] Cantu, A.A. (1995). A sketch analytical methods for document dating. Part I. The static approach: determining age independent analytical profiles, *International Journal of Forensic Document Examiners* **1**, 40–51.

[9] Cantu, A.A. (1996). A sketch analytical methods for document dating. Part II. The dynamic approach: determining age dependent analytical profiles, *International Journal of Forensic Document Examiners* **2**, 192–208.

[10] Aginsky, V.N. (1993). Some new ideas for dating ballpoint inks – a feasibility study, *Journal of Forensic Sciences* **38**, 1134–1150.

[11] Aginsky, V.N. (1996). Accelerated aging – its use in methods for dating ink, *International Journal of Forensic Document Examiners* **2**, 179–181.

[12] Aginsky, V.N. (1998). Measuring ink extractability as a function of age – why the relative aging approach is unreliable and why it is more correct to measure ink volatile components than dyes, *International Journal of Forensic Document Examiners* **4**, 214–230.

[13] Weyermann, C., Kirsh, D., Vera, C. & Spengler, B. (2006). A GC/MS study of the drying of ballpoint pen ink on paper, *Forensic Science International* **168**, 119–127.

[14] Brazeau, L. & Gaudreau, M. (2007). Ballpoint pen inks: the quantitative analysis of ink solvents on paper by solid-phase microextraction, *Journal of Forensic Sciences* **52**, 209–215.

[15] ASTM E1789-04 (2004). *Standard Guide for Writing Ink Identification*, ASTM International, West Conshohocken.

[16] ASTM E1422-05 (2005). *Standard Guide for Test Methods for Forensic Writing Ink Comparison*, ASTM International, West Conshohocken.

[17] Tebbett, L.R. (1991). Chromatographic analysis of inks for forensic science applications, *Forensic Science Review* **3**, 71–82.

[18] Zlotnick, J.A. & Smith, F.P. (1999). Chromatographic and electrophoretic approaches in ink analysis [Review], *Journal of Chromatography B* **733**, 265–272.

[19] Neumann, C. & Margot, P. (2003). Forensic ink analysis: a review of the techniques, *Revue Internationale de Criminologie et de Police Technique* **56**, 341–360.

[20] Brunelle, R.L., Reed, R.W. & Charles, C. (eds) (1984). *Forensic Examination of Ink and Papers*, Thomas publishers, Springfield, pp. 124–133.

[21] Cantu, A.A. (1991). Analytical methods for detecting fraudulent documents, *Analytical Chemistry* **63**, 847A–854A.

[22] Blackledge, R.D. & Iwan, M. (1983). Differentiation between inks of the same brand by infrared luminescence photography of their thin-layer chromatograms, *Forensic Science International* **21**, 165–173.

[23] Roux, C., Novotny, M., Evans, I. & Lennard, C. (1999). A study to investigate the evidential value of blue and black ballpoint pen inks in Australia, *Forensic Science International* **101**, 167–176.

[24] Aginsky, V.N. (2006). Using TLC and GC-MS to determine whether inks came from the same manufacturing batch, *Journal of the American Society of Questioned Document Examiners* **9**, 19–27.

[25] Harget, J.W. (1990). The International Ink Library, *International Criminal Police Review* **45**, 33–34.

[26] Ramotowski, R.S. & Regen, E.M. (2007). Effect of electron beam irradiation on forensic evidence. 2. Analysis of writing inks on porous surfaces, *Journal of Forensic Sciences* **52**, 604–609.

[27] Neumann, C., Champod, C. & CAMAG Inc (2006). *The International Ink Library of the United States Secret Service: A New and Efficient Way of Managing the Data*, DHSARPA, Contract HSHQDC-06-R-00066.

[28] LaPorte, G.M., Arredondo, M.D., McConnell, T.S., Stephens, J.C., Cantu, A.A. & Shaffer, D.K. (2006). An evaluation of matching unknown writing inks with the United States International Ink Library, *Journal of Forensic Sciences* **51**, 689–692.

[29] Tappolet, J.A. (1984). *Etude de l'application en Criminalistique de la chromatographie en couche mince à haute performance (CCMHP) à l'examen de traits d'encres liquides noires, bleues et bleu-royal*, Institut de Police Scientifique, Université de Lausanne.

[30] Lewis, J.A. (1996). Thin-layer chromatography of writing inks – quality control considerations, *Journal of Forensic Sciences* **41**, 874–877.

[31] Hamed, H.R., Safey El-Din, N.M., El-Laithy, S.A., Mansour, O.Y. & Sabaa, M.W. (1997). Effect of accelerated fading on the stability of inks marked on different types of papers, *International Journal of Forensic Document Examiners* **3**, 229–236.

[32] Andrasko, J. (2001). HPLC analysis of ballpoint pen inks stored at different light conditions, *Journal of Forensic Sciences* **46**, 21–30.

[33] Andrasko, J. (2002). Changes in composition of ballpoint pen inks on aging in darkness, *Journal of Forensic Sciences* **47**, 324–327.
[34] Andrasko, J. & Kunicki, M. (2005). Inhomogeneity and aging of ballpoint pen inks inside of pen cartridges, *Journal of Forensic Sciences* **50**, 542–547.
[35] Weyermann, C., Kirsh, D., Costa-Vera, C. & Spengler, B. (2006). Photofading of ballpoint dyes studied on paper by LDI and MALDI MS, *Journal of the American Society of Mass Spectrometry* **17**, 297–306.
[36] Evett, I.W. (1986). A Bayesian approach to the problem of interpreting glass evidence in forensic science casework, *Journal of Forensic Science Society* **26**, 3–18.

Related Articles

Identification and Individualization
Ink Analysis
Interpretation: Document Evidence

CEDRIC NEUMANN

Insanity *see* Temporary Insanity

Insanity: Defense

Criminal laws in Western society are designed to identify those persons who should be held responsible for actions contrary to societal rules and subject to appropriate punishment. This concept is well entrenched in our moral codes and appears in early biblical and philosophical writings. From a legal perspective, how to codify those factors that exculpate one's responsibility for criminal conduct raises important issues. For example, a person who unintentionally hits another person in the midst of a violent seizure does not seem to warrant criminal sanction. Similarly, a two-year-old girl who scratches her 14-year-old brother should not be punished in the same way as when her older brother commits a similar act. In these scenarios, the legal system recognizes that not all actions call for legal determinations of guilt or identical subsequent punishments. The question of how the law might draw the line between those who are found responsible for their actions and those who are not the subject of great debate and central to the concept of the insanity defense [1].

Under United States law (which has its origins in British common law), to be found guilty of criminal conduct, one must have committed an illegal act, also referred to as an *actus reus* combined with the *mens rea*, or guilty mind, that accompanies the act [2]. Without a finding that both elements were present at the time of the crime, the defendant should not be held legally responsible. The Not Guilty by Reason of Insanity (NGRI) defense stands for the proposition that although there are some individuals who have committed an *actus reus*, they are not blameworthy because their mental illness prevents them from forming the requisite *mens rea*. Therefore, the appropriate response for such persons is treatment rather than punishment.

Although the public may believe that the insanity plea is an "easy way out" of the criminal justice system, the defense is infrequently raised and rarely successful. For example, the insanity defense is raised in less than 1% of all felony cases [3] and in the United States, on average, defendants are found NGRI only one out of every four times the defense is raised [4]. Furthermore, individuals who are found criminally insane are typically involuntarily committed to a forensic psychiatric facility for an indefinite period of time. Although the criteria for release from the hospital into the community are defined by each jurisdiction, the person's risk of future dangerousness is generally a significant factor in deciding whether or not he or she is appropriate for release.

Insanity Tests

Different legal standards of insanity have been proposed in an attempt to define when and how a person's mental illness should legally excuse their criminal acts. One of the earliest standards arose from the case of *Rex v. Arnold*,, in which Justice Tracy held that for a man to be found insane he must be "totally deprived of his understanding and memory, and doth not know what he is doing, no more than an infant, . . . a brute, or a wild beast . . ." [5]. Under this stringent test, the individual has to be so mentally disturbed that they do not even know what they are actually doing.

Subsequent cases decided by the English courts proposed other operational definitions for insanity.

One such case involved Mr. James Hadfield, an Englishman charged with treason after he attempted to shoot King George III while the King was entering the theater. Mr. Hadfield had the delusion (*see* **Delusions**) that God was going to destroy the world unless Mr. Hadfield killed the King and then was sacrificed through a legally sanctioned execution. At his trial, Mr. Hadfield's attorney, Thomas Erskine, argued for a finding of insanity. Lord Erskine proposed that the test of insanity should take into account whether the act was the "offspring of a delusion", rather than the act of a person who was like a "wild beast" and had lost total control [6]. This successful defense resulted in Hadfield's acquittal by reason of insanity and the subsequent passage of the Criminal Lunatics Act of 1800; one of the earliest legislative efforts to articulate aspects of the insanity defense [7].

In the aftermath of the Hadfield case, a new test for insanity arose from the trial of Edward Oxford. In 1840, Mr. Oxford was acquitted on grounds of insanity after he had attempted to assassinate Queen Victoria. The instructions to the jury indicated that Mr. Oxford should not be held responsible if a "controlling disease" made him unable to resist his actions. This standard became known as the "*Irresistible Impulse" insanity standard*, though its consideration was not limited to "impulsive" or sudden acts. Under this standard, a person could have the cognitive faculties to know what they were doing and the wrongfulness of their actions, but nevertheless be found "insane" because their mental disorder resulted in their inability to resist the impulse to commit the crime [8].

The United States briefly adopted the "Irresistible Impulse" standard in the mid-twentieth century in the Federal housebreaking case of Monty Durham [*Durham v. United States*, 214 F.2d 862 (1954); see also as amended in *United States v. Brawner*, 471 F.2d 969 (1972)]. The Durham standard proved controversial, and critics pointed out that there seemed to be no way to distinguish between an "Irresistible" impulse that was or was not resisted; thus, it could be argued that any action actually taken was, by definition, irresistible. Many jurisdictions still have a component of this standard, and mental health practitioners apply clinical data to help the court in its determinations on the issue of one's capacity to conform conduct at the time of an alleged criminal act based on symptoms of mental disease or defect.

One of the most famous and widely adopted tests of insanity is known as the *M'Naghten rule*. Daniel M'Naghten was a Scottish wood turner who attempted to murder Sir Robert Peel, a prominent leader of the Tory Party whom Mr. M'Naghten believed was persecuting him. By mistake, Mr. M'Naghten shot and killed Peel's secretary. At his trial, Mr. M'Naghten was found NGRI. Owing to the resulting public outrage, Queen Victoria ordered the 15 Law Lords of the House of Lords to compose a new, and presumably, more stringent standard. The result of their efforts became known as the *M'Naghten standard*, which stated the following:

> "To establish a defense on the grounds of insanity, it must be proved that, at the time of the committing of the act, the party accused was laboring under such a defect of reason, from disease of the mind, as not to know the nature and quality of the act he was doing; or if he did know it, that he did not know he was doing what was wrong" [9].

The M'Naghten test is often referred to as a "*cognitive" test of insanity* because it specifies that the person will not be found insane if they know the nature and quality and wrongfulness of their acts. Some jurisdictions have modified this test by substituting the words appreciate, understand, or distinguish the word "know".

In contrast to the more stringent M'Naghten standard, the New Hampshire case of *State v. Pike* (1869) formulated an insanity standard known as the "*product test*". This test is significantly more lenient than the M'Naghten standard and reads, "No man shall be held accountable, criminally, for an act which was the offspring and product of mental disease" [10]. The product test allows a defendant to be found criminally insane if their behavior results from their mental disease, even if they knew their actions were wrongful. The product test has not received broad acceptance and is the insanity standard only in New Hampshire and the Virgin Islands [11].

In 1955, the American Law Institute (ALI) produced the Model Penal Code, which also included a proposed insanity test. This ALI insanity test includes both a cognitive and volitional prong and reads as follows:

> "A person is not responsible for criminal conduct if at the time of such conduct as a result of mental disease or defect he lacks substantial capacity either to appreciate the criminal conduct or to conform his conduct to the requirements of the law" [12].

Although the ALI test of insanity was utilized by federal courts and approximately half of the states during the 1980s, the acquittal by reason of insanity of John Hinkley for his attempted assassination of President Ronald Reagan resulted in sweeping changes in both the federal and state statutes. As a result, many states eliminated the volitional prong from their insanity defense and returned to the stricter M'Naghten standard. In 1984, Congress passed the Insanity Defense Reform Act, which resulted in a new federal test of insanity that closely resembles the original M'Naghten test. Under this test, a person is not criminally responsible, if "As a result of a severe mental disease or defect, he was unable to appreciate the nature and quality or the criminality of his acts" [13].

Assessing Criminal Responsibility – The Forensic Interview

When preparing for an evaluation of a defendant's criminal responsibility, the expert should first clarify if he or she is appointed by the court or retained by the defense or prosecution. Although the examiner should always strive for honesty and objectivity regardless of whom has retained him or her, opinions rendered by a psychiatrist hired by the defense are not always disclosed to other parties. Before proceeding, the evaluator should carefully review the relevant statute as well as any case law that interprets the statutory language. The expert should also request collateral records to include police reports, witness and victim statements, booking and jail records, past psychiatric and medical records, police records of prior crimes, educational and work records, as well as audio or videotapes of confessions and 911 telephone tapes. Other evaluator's opinions may also assist in reviewing the consistency of the defendant's presentation. However, the examiner should first determine if any prior psychological examinations are prohibited from their review.

An evaluation of criminal responsibility involves an assessment of the defendant's mental state at the time of their alleged offense. Therefore, the evaluator should schedule their examination as soon as possible in order to observe the defendant's mental status in close proximity to the crime. Prior to conducting the evaluation, the defense attorney should be notified of the impending interview. In some situations, the defense attorney may request to be present during the assessment and may obtain a court order allowing them to do so. If this situation occurs, the evaluator should request that the defense counsel should not interrupt the examination or instruct the defendant how to respond to questions.

Prior to beginning the interview, the evaluator should inform the defendant who requested the evaluation, that the assessment is not confidential, that they are not acting as the defendant's treating psychiatrist, and that the defendant's statements may appear either in a written report or in deposition or court testimony. The evaluations may be audio or videotaped but such recording is not mandatory unless required by the court.

The examiner typically takes a broad bio-psychosocial history, which may be important to understand the defendant's background leading up to the instant offense. Key areas to review include past psychiatric history and prior hospitalizations, family psychiatric history, educational history, any history of learning disabilities or mental retardation, and the defendant's social and relationship history, particularly as related to any of their crime victims. The defendant's drug and alcohol history is also critically important, particularly during the months and days leading up to the instant offense. Finally, understanding the defendant's prior legal history can assist in determining if the defendant has a pattern of criminal behavior that may be unrelated to the presence, if any, of a psychiatric disorder.

Taking the Defendant's Account of the Offense

In assessing criminal responsibility, an evaluator generally begins by asking open-ended questions related to the alleged offense(s). Sample questions may include statements such as "Tell me what happened" or "What were you thinking and feeling before, during, and after your alleged crimes?" Through this technique, the examiner encourages the defendant to communicate their own thought processes regarding how the events unfolded. The examiner follows with more detailed questions that review the events in the weeks, days, hours and minutes leading up to and including the offense.

Rather than simply obtaining an account of "what happened", however, the evaluator must also determine if the defendant had a mental disorder at the

time of the offense and the relationship, if any, of the mental disorder to their criminal acts.

With regard to specific lines of inquiry, the evaluator typically asks questions to assess whether the defendant understood the nature and quality of his or her behavior at the time of the crime. For example, the evaluator might ask the defendant to explain what he was specifically thinking with regard to each action that comprised their criminal behavior. When evaluating a defendant's knowledge of wrongfulness, the examiner should review steps that the defendant took to plan and prepare for the crime as well any evidence to suggest that the defendant attempted to escape or avoid detection. Specific efforts to avoid detection that suggests knowledge of wrongfulness includes wearing gloves or a disguise during a crime, concealment of a weapon, waiting until the cloak of darkness, threatening to kill the witness if they go to the police, and giving a false name or alibi when questioned by the police. Examples of hiding evidence that indicate a knowledge of wrongfulness include wiping off fingerprints, washing away blood from the crime scene, destroying documents, or hiding a body [14] [15].

Although the above actions may suggest the defendant understood that their actions were wrongful, the evaluator should specifically query the defendant as to his or her understanding of both the legal and moral wrongfulness of their behavior. For example, this inquiry might include questioning a defendant as to why he threw a gun down a gutter or disposed of gloves that he was wearing. If the defendant responds that he threw these items away because he knew the police might arrest him, this answer might signify that he understood that his actions were legally wrongful. If, however, the defendant also describes a delusional belief that he needed to kill his neighbor in order to save the world from evil forces and was attempting to avoid arrest to continue this moral crusade, his psychotic beliefs may also render him unable to know the moral wrongfulness of his acts, despite his understanding that his behavior might be misinterpreted as wrongful by others.

Specific evidence that the defendant understood his actions were wrongful include statements by the defendant to the examiner that he knew his actions were wrongful at the time of the crime, notifying the police that a criminal act had occurred, and expressions of remorse or guilt immediately following the crime. At some point during the examination, the evaluator should ask questions such as "At the time of the crime, did you know that what you were doing was against the law?" or "Was there anything that made you think that what you were doing was a right thing to do?"

In those insanity tests that include a volitional prong, the assessment of the defendant's ability to conform their conduct or to resist a criminal impulse may be more challenging. First, if a defendant's mental disorder renders them unable to know their actions are wrongful, he or she may have no reason to actually conform or alter their conduct. In assessing a defendant's ability to refrain from criminal behavior, Rogers [16] has suggested examining the defendant's capacity to delay their actions, plan, organize, to consider alternatives, and to act differently in the face of potential negative consequences. Another method to assess a defendant's ability to refrain is sometimes referred to as the "*policeman at the elbow" test.* Under this line of questioning, the defendant is asked if they would have changed or stopped their behavior if a policeman was nearby, i.e., at their elbow. In general, individuals, who would stop their behavior if a policeman were to come on the scene, demonstrate the capacity to conform their conduct to the law should they choose to do so [3].

A defendant's account of an alleged offense is a retrospective account, and one that may be obtained at any point after an alleged offense. Therefore, the information the defendant provides may be a distorted or a self-serving account of what actually occurred. For example, a defendant who acted while floridly psychotic may have some confusion regarding their recollection of what he or she was thinking and feeling at the time. Even if a defendant describes events in their entirety and with a solid understanding of their symptoms, the evaluator must attempt to incorporate collateral data to help gain a fuller picture of what the defendant may have actually been experiencing.

Forming the Sanity Opinion

There are three important areas to review when rendering an opinion on a defendant's criminal responsibility. First, the evaluator must establish if the individual had a mental disease or defect at the time of the crime. The expert should determine what mental disorders qualify for consideration of insanity

through reviewing the governing statute and relevant case law. Psychiatric illnesses characterized by psychotic symptoms or significant cognitive impairment are the most common types of mental disorders that reach the threshold for consideration of the insanity defense. In some jurisdictions, certain disorders are specifically excluded and may include all personality disorders or antisocial personality disorder. In addition, voluntary intoxication alone may also be excluded as a qualifying mental illness. The examiner should become very familiar with detecting signs of malingered mental illness, particularly faked hallucinations or delusions [11].

Secondly, the expert evaluates whether or not there is a relationship between the defendant's mental symptoms and their alleged criminal actions. It is important to specifically review the motive for the defendant's behavior as well as their thoughts and feelings before, during, and after the crime. The evaluator should also screen for the presence of a rational alternative motive, such as stealing money for drugs or killing someone in a jealous rage. Such motives are often unrelated to the person's mental illness and generally exclude the person from further consideration of an insanity defense if they are the sole driving force for the criminal behavior [11].

Finally, the examiner must apply the relevant insanity test when evaluating the relationship between the person's mental disorder and their alleged acts. Under a M'Naghten test of insanity (i.e., cognitive standard), the evaluator reviews if the defendant knew what they were doing or understood that their actions were wrong, even if they had a qualifying mental disorder. Consider the case of a paranoid schizophrenic who burns down their neighbor's home because they believe the neighbor is demon possessed and is casting spells that would soon destroy the world. Despite their mental illness, this defendant may have understood that they were starting a fire to burn an inhabited dwelling (i.e., the nature and qualify of their acts) while at the same time believed that their actions were right because they were saving many lives from demonic destruction. Some jurisdictions allow consideration of insanity if the person's mental disorder resulted in the belief that their actions were morally justified, despite knowing that their actions were wrong in the eyes of the law.

When applying volitional tests of insanity, the examiner must carefully review if the person's mental disorder rendered them unable to control their behavior, even if they knew their actions were wrongful. Evidence that may indicate that the defendant had the ability to refrain includes stopping their behaviors when detected by someone during the course of the crime or deferring their actions until a more advantageous opportunity arises [11].

Summary

The insanity defense is based on the premise that some mentally ill individuals are not legally culpable due to their underlying illness and therefore should not be criminally punished. Tests of insanity have evolved over time and vary significantly by jurisdiction. Evaluators must conduct a careful retrospective analysis of the defendant's mental state at the time of the offense and use available collateral data to determine whether or not the defendant meets the specific test of insanity.

References

[1] Bonnie, R. (1973). The moral basis of the insanity defense, *American Bar Association Journal* **69**, 194–197.

[2] Walker, N. (1985). The insanity defense before 1800, *The Annals of the American Academy of Political and Social Sciences* **477**, 25–30.

[3] Melton, G.B., Petrila, J., Poythress, N.G. & Slobogin, C. (2007). *Psychological Evaluations for the Court: A Handbook for Mental Health Professionals and Lawyers*, 3rd Edition, The Guilford Press, New York.

[4] Callahan, L.A., Steadman, H.J., McGreevy, M.A. & Robbins, P.C. (1991). The volume and characteristics of insanity defense pleas: an eight-state study, *The Bulletin of the American Academy of Psychiatry and the Law* **19**(4), 331–338.

[5] Arnold's Case, 16 Howell St. Tr. 695, 764 (1724).

[6] Quen, J.M. (1974). Anglo-American insanity: an historical perspective, *The Bulletin of the American Academy of Psychiatry and the Law* **2**, 115–123.

[7] Appelbaum, P.S. (1994). *Almost a Revolution: Mental Health Law and the Limits of Change*, Oxford University Press, Inc, New York.

[8] Eigen, J.P. (1999). Lesion of the will: medical resolve and criminal responsibility in Victorian insanity trials, *Law and Society Review* **33**(2), 425–459.

[9] *M'Naghten's Case*, 10 Clark & Finelly 200, 8 Eng. Rep. 718 (1843).

[10] State v. Pike, 49 N.H. 399 (1869).

[11] Giorgi-Guarnieri, D., Janofsky, J., Keram, E., Lawsky, S., Merideth, P., Mossman, D., Schwart-Watts, D.,

Scott, C., Thompson, J.Jr. & Zonana, H (2002). AAPL practice guideline for forensic psychiatric evaluation of defendants raising the insanity defense. American Academy of Psychiatry and the Law, *The Journal of the American Academy of Psychiatry and the Law* **30**(2 Suppl), S3–S40.

[12] American Law Institute *Model Penal Code* § 4.01(2) 985.

[13] Insanity Defense Reform Act of 1984, U.S.C. Title 18 § 17, 3006A, 4241.

[14] Resnick, P.J. & Noffsinger, S. (2004). Competency to stand trial and the insanity defense, in *Forensic Psychiatry for the Clinician: Guidelines for Assessment*, L. Gold & R. Simon, eds, American Psychiatric Publishing, Inc., Washington, D.C., pp. 329–347.

[15] Knoll, J.L. & Resnick, P.J. (2007). Insanity defense evaluations: toward a model for evidence-based practice, *Brief Treatment and Crisis Intervention*, Available at: http://brief-treatment.oxfordjournals.org/cgi/reprint/mhm024v1 (accessed Jan 2008).

[16] Rogers, R. (1987). APA's position on the insanity defense. Empiricism versus emotionalism, *The American Psychologist* **42**(9), 840–848.

Related Articles

Dissociative Disorders

CHARLES L. SCOTT AND DEBRA A. PINALS

Insects *see* Entomology

Intellectual Disability *see* Mental Retardation

Intelligence *see* Mental Retardation

Intelligence: Footwear Impression as *see* Footwear and Foot Impressions: Intelligence

Intentional Harm *see* Aggression

International Classification of Disease *see* Psychopathology: Terms and Trends

Internet Resources for Forensic Scientists *see* Web Resources

Interpretation: DNA Low Amounts *see* Interpretation: Low Template DNA

Interpretation: DNA Mixtures *see* Mixture Interpretation: DNA

Interpretation: Document Evidence

Introduction

The literature dealing with the interpretation of "printed" documents is very limited. An article published by Souder [1] first highlights a theory of probability applied to two imaginary typewritten documents.

In general the absence of published literature reflects the absence of formal guidelines to help a scientist in assessing the relationship between a questioned document and a printing technology. Everything is left to the expert's judgment based on his experience or by invoking the random nature of a set of observations. This lack of formal guidelines is well exemplified by the following quote:

> Therefore, although all such marks on two photocopies may not much, a reasonable number that can be easily superimposed, by the use of either photographic transparencies or a comparison projector, is clear proof that the same machine has been used in the production of both. Their random nature indicates that chance match is extremely unlikely [2], p. 160.

With the massive availability of different printed documents, the forensic document examiner is more than even faced with the obligation of identifying different printing technologies. Detailed information about the identification of the printing technologies is beyond the scope of this article and the reader should refer to the specialized literature.

Generally every printout, independently from the printing technology, may present manufactured characteristics and acquired characteristics (individual characteristics). The manufactured characteristics may be defined by the type of ink (or printing material) used and the morphology aspect of the printout under the stereomicroscope. The manufactured characteristics refer to *class characteristics*, in the sense that all printed documents issued from a specific technology will share the same characteristics. For example, all the electronic typewriters equipped with a Brougham 10 daisy wheel will share the same class characteristics such as font design, size, and spacing.

Individualization intervenes when we can demonstrate that every object belonging to a defined class can be distinguished from the others on the basis of individual characteristics, provided that these exist. Individual characteristics are also described as acquired characteristics. This second definition is more elegant because brings in the notion of time. Indeed, time and usage of a certain object can provoke unique accidental deformations allowing its individualization.

It is obvious that the origin or the source of every manufactured or acquired characteristic should be carefully identified and explained. To simplify the reading, the author will focus the interpretation of printed documents on two technologies: typewriting and electrophotography. Table 1 lists a number of manufactured and acquired characteristics that may be encountered within the above-mentioned printing technologies.

A logical approach to help the scientist in the formal interpretation of such evidences is to assess the results in a likelihood ratio (LR) format. The main advantage of this approach is to invite the scientist to consider his results in the light of at least two

Table 1 List (not exhaustive) of manufactured and acquired defects encountered within typewriting and electrophotography

Printing technology	Manufactured characteristics	Acquired characteristics
Typewriting	Types of printing	Misalignments
	Typebar	Horizontal
	TypeBall	Vertical
	Typewheel	Rotational
	Design, size, and spacing of typeface	Typeface defects
Electrophotography	Roller marks	Random defect from the glass platen
	Types of toners	
	Single or bicomponent	Repetitive defect from
	Organic composition	OPC drum
	Inorganic composition	Fuser rollers
		Grabbing rollers
		White divots [3]

propositions. The following three examples aim to show the application of this approach.

Typewriting Example

An anonymous letter was typewritten with a mechanical typewriter having a 254 spacing and a cubic type font design. According to the limited questioned material, the exact cubic font type could not be identified. The letter "p" has a broken bulb defect that has been identified by the expert as an acquired defect. Indeed, in order to interpret these findings properly, the examiner should exactly understand the mechanical causes of the "p" defect. A mechanical typewriter showing the same manufactured and acquired characteristic is seized. The question is what the evidential value of this correspondence? To approach this question, there is the need to define the results, the propositions at hand and assign the LR.

- Define the evidence results $E = (E_1, E_2, E_3)$ as
  E_1: mechanical typeface with 254 spacing,
  E_2: cubic typeface, and
  E_3: acquired defect on letter "p".
- Define the propositions H_p: hypothesis for the prosecution and H_d: hypothesis for the defense:
  H_p: the seized typewriter is the origin of the anonymous letter and
  H_d: another typewriter is the origin of the anonymous letter.

The population defined by the generic term "another typewriter" is dictated by the case circumstances. In this case, it could be any other typewriter including all kind of type font and type printing mechanism. The LR becomes

$$LR = \frac{\Pr(E_1, E_2, E_3|Hp)}{\Pr(E_1, E_2, E_3|H_d)} = \frac{\Pr(E_1|H_p)}{\Pr(E_1|H_d)} \times \frac{\Pr(E_2|E_1, H_p)}{\Pr(E_2|E_1, H_d)} \cdot \frac{\Pr(E_3|E_1, E_2, H_p)}{\Pr(E_3|E_1, E_2, H_d)} \quad (1)$$

$$LR = \frac{\Pr(E_1, E_2, E_3|H_p)}{\Pr(E_1, E_2, E_3|H_d)} = \frac{1}{f_{Mechanical\ \&\ 254}} \times \frac{1}{f_{Cubic|Mechanical\ \&\ 254}} \cdot \frac{a}{b} \quad (2)$$

A search into the "TYPE" database [4] allows knowing the frequency of mechanical typeface having a 254 spacing (= 13%, there are 622 mechanical typefaces out of a total of 4683 typefaces) and using a cubic typeface style (= 3%, there are 22 mechanical cubic typefaces).

The probability a must be evaluated considering the reproducibility of the acquired defect among the reference known samples. Let us say that the defect is always reproducible, so $a = 1$.

The probability b must be evaluated considering the frequency of the acquired defect. Because we do not have the appropriate database, we may subjectively evaluate or use a reference data such as the study by Gayet [5].

Equation 2 becomes

$$LR = \frac{\Pr(E_1, E_2, E_3|H_p)}{\Pr(E_1, E_2, E_3|H_d)} = \frac{1}{0.13} \cdot \frac{1}{0.03} \cdot \frac{1}{0.44} \approx 582 \quad (3)$$

The findings are about 582 times more likely if the seized typewriter is the source of the questioned letter than if originated from another typewriter.

Photocopier Example

A blackmail message was sent to the police department menacing the building's total destruction. The message was determined as being a photocopy. A further examination of the recto of the photocopy showed two acquired marks of different size and shape (suggesting scratch glass platen marks). Investigators were able to rapidly locate a suspect and at his home a photocopier producing the same two acquired marks was found.

- Define the results of your evidence $E = (E_1)$ as
  E_1: two acquired defects.
- Define the propositions:
  H_p: the seized photocopier, at certain time, is the origin of the blackmail message; and
  H_d: another photocopier is the origin of the blackmail message.

The fact that the questioned document was printed using a photocopier (as opposed to any other printing device) is not considered important in the assessment of the case hence the device *photocopier* appears in both the propositions. The same approach had been adopted in the previous example with regards to the *typewriter*.

The likelihood becomes

$$LR = \frac{\Pr(E_1|H_p)}{\Pr(E_1|H_d)} = \frac{1}{f} \quad (4)$$

The numerator has been set to 1 because of the persistency and reproducibility of the defects under the case circumstances (i.e., no replacement of the glass platen).

f is the probability of finding two marks at the same position (X, Y) on the glass platen. We can assume that we may distinguish only marks separated by 2 mm (that is only an assumption, in reality we can distinguish marks with relative distances smaller than 1 mm), so on an A4 document (210 mm × 297 mm) we have 105 × 148.5 squares of size 2 mm. So, if we assume that the characteristics are uniformly acquired on the glass platen surface, the probability to find the same 2-mm mark on another copier is $1/105 \times 148.5 = 0.000064$. To be more conservative this probability should be multiplied by 4. In fact, depending on the position of the document to be copied on the glass platen, the same defect on the glass platen can generate four different positions of the marks on the resulting photocopied document. So our probability will be $0.000064 \times 4 = 0.000256$.

If we assume that each mark could be considered as independent, we can square the probability so that in our scenario of two marks we obtain $0.000256 \times 0.000256 = 6.55 \times 10^{-8}$.

$$LR = \frac{\Pr(E_1|H_p)}{\Pr(E_1|H_d)} = \frac{1}{f} = \frac{1}{0.000256 \times 0.000256}$$
$$= \frac{1}{6.55 \times 10_{-8}} = 15'258'789 \quad (5)$$

In this simple case, we obtain a very high LR. The findings are about 15 million times more likely if the seized photocopier is the source of the blackmail letter than if originated from another photocopier. It should also be noted that in our case we do not take into account the chemical toner composition, neither the shape nor the orientation of the marks that would surely lower the probability.

Laser Printer Example

Consider a laser printed A4 contract showing two repetitive defects (1 mm) along the page. A repetitive artifact is a printing quality defect that has a pattern which reappears on the printed page at regular intervals. It results from physical damage to or contamination of a rotating unit such as a charging roller, an organic photoconductor (OPC) drum, or a fuser. Measuring the distance between the repetitive defects allows inferring the rotating unit circumference. The vertical distance between each repetitive defect was 94-mm. An HP LaserJet 4000 showing the same printing defect is seized. This laser printer has an OPC drum circumference of 94 mm.

- Define the results of your evidence $E = (E_1, E_2)$ as
  E_1: a OPC drum with a 94-mm diameter and
  E_2: two acquired repetitive defects
- Define the propositions:
  H_p: the seized HP LaserJet 4000 printer has printed the contract; and
  H_d: another Laser printer has printed the contract.

The likelihood becomes

$$LR = \frac{\Pr(E_1|H_p)}{\Pr(E_1|H_d)} \frac{\Pr(E_2|E_1, H_p)}{\Pr(E_2|E_1, H_d)} =$$
$$\frac{1}{f_{94\text{-mm OPC}}} \cdot \frac{1}{f_{\text{defects}}} \quad (6)$$

where f_{94}-mm OPC is the probability of finding a laser printer equipped with a 94-mm OPC drum. This probability was assigned to 0.22 in reference to a database [6].

Where f_{defects} is the probability of finding two repetitive defects at the same position X on a 94-mm OPC drum. Consider a defect that is present on a drum with a circumference of 94-mm and a drum width of 210 mm (the width of an A4 page). The probability of this defect occurring in the same horizontal position and any radial position on another drum of this size is 0.0047 (or 1 in 210).

Now considering we have two defects of this size occurring on two different horizontal X positions on the drum, then the probability is squared. Again, we assume that the defects are independent authorizing the frequency multiplication.

$$LR = \frac{1}{0.22} \cdot \frac{1}{0.0047 \times 0.0047} = 205'769 \quad (7)$$

The findings are 205'769 times more likely if the seized laser printer is the source of the questioned letter than if originated from another laser printer. Again, the chemical toner composition, neither the

shape nor the orientation of the marks were taken into account that surely increase the LR value.

Conclusion

In the above three examples, the subjective "identification" approach was formalized adopting a logical framework assigning the results in a LR format. This interpretation methodology invites the scientist to consider his results in the light of at least two alternative propositions that are mutually exclusive. It mainly provides a support to properly evaluate the evidence in a particular case with impartiality, but it can also provide a help to the decision through the probabilities quantification involved in the LR.

References

[1] Souder, W. (1934). The merits of scientific evidence, *Journal of the American Institute of Criminal Law and Criminology* **25**, 683–684.
[2] Ellen, D. (2006). *Scientific Examination of Documents Methods and Techniques*, 3rd Edition, CRC Taylor & Francis Group.
[3] Winter, R. (2000). The detection of laser printer defects for printer identification. *58th ASQDE Meeting*, Ottawa.
[4] Kelly, M. (2004). The significant contributions of Dr. Philip D. Bouffard to the examination and classification of typewriting, *Journal of the American Society of Questioned Document Examiners* **7**(2), pp. 91–96.
[5] Gayet, J. (1961). *Manuel de Police Scientifique*, Payot, Paris, pp. 238–239.
[6] Haas, B. (1998). Identification of laser printers. *Joint Meeting of the European Conferences for Police and Government Handwriting and Documents Experts*, Clackmannanshire.

W.D. MAZZELLA

Interpretation: Earprints *see* Earprints: Interpretation of

Interpretation: Friction Ridge and Fingerprints *see* Friction Ridge Examination (Fingerprints): Interpretation of

Interpretation: Glass *see* Glass Evidence: Bayesian Approach to

Interpretation: Ink *see* Ink Comparison and Interpretation

Interpretation: Legal Perspective

Introduction

In comparing fibers, hairs, paint chips, glass fragments, toolmarks, soil samples, bloodstains, and other forms of trace evidence, it is common to speak of "matches" and "exclusions". If two samples do *not* match (and if the match criteria cannot produce false negatives, and if there has been no police or laboratory misconduct or error), then the samples cannot have come from the same source. But testimony that the samples match has an additional wrinkle. For instance, even if two DNA samples are identical, it is logically possible that someone else has the same DNA profile and is the source. Likewise, the fact that a bullet fragment from a crime scene and a bullet found in a defendant's residence have similar elemental compositions does not necessarily mean that they came from the same box of ammunition, the same melt of lead, or some other unit of production [1, 2]. Because such matches are not within the experience of ordinary judges and jurors, the legal system normally expects expert witnesses to assist the factfinder by providing further information or conclusions about how a match should be evaluated. This article describes the law concerning the manner in which the significance of a laboratory finding of a match may be described.

Probabilistic Statements

Frequencies and Random Match Probabilities

Courts routinely admit estimates of how infrequent the matching features are in the general population [3, § 12.4.1]. Such quantities indicate the probability that a randomly selected sample in a defined population would have the matching type if that sample were *not* the source. The underlying logic is that if the match is unlikely to arise under this hypothesis, then the match is unlikely to be coincidental. Thus, the random match probability, as forensic scientists usually call it, is essentially the same as the statistician's p-value – the probability of finding an outcome at least as extreme as the one observed when the null hypothesis is true.

It has been argued that very small random match probabilitites are unfairly prejudicial or misleading. One potential danger is that the jury will think that it gives the probability that the match is random. A random match probability of 1%, for example, might be misreported or misconstrued as a 99% chance that the trace originated from the defendant. This is an example of the *fallacy of the transposed conditional* (also known as the *prosecutor's fallacy*) [3, § 12.4.1(b)(1)].

Some American courts have intimated that transposition by the expert or prosecutor could constitute reversible error, and one federal court of appeals granted postconviction relief partly on this ground [3, § 12.4.1(b)(1)]. Likewise, courts in England and Australia have reversed convictions when the judge's summation contained the fallacy. In reversing rape convictions in *R. v. Doheny*, [1997] 1 Cr. App. R. 369, 374, the English court of appeals suggested that to avoid transposition, an expert presenting DNA evidence should give the jury the frequency and then "say how many people with the matching characteristics are likely to be found in the United Kingdom – or perhaps in a more limited subgroup, such as, for instance, the Caucasian sexually active males in the Manchester area". This proposal has not been well received [4, 5], but much of the criticism seems overdrawn [3, § 12.4.1(b)(2)].

Likelihood Ratios

Some statisticians do not regard random match probabilities as a suitable measure of probative value. One alternative is to use a likelihood ratio to express how much the data support one hypothesis compared to another [6–8]. Indeed, the likelihood ratio can be generalized to describe the degree of support that measurements of a continuous variable, such as the refractive index of a sheet of glass, gives to competing hypotheses. Framing the issue in terms of likelihoods eliminates the need for an arbitrary cutoff point for declaring a match, and it is particularly useful when the numerator (the probability of the match when the suspect is the source) is not assumed to be 1 [2, 9].

Likelihood ratios have been introduced in a growing number of criminal cases [3, § 12.4.2(b)]. They are routinely quoted in civil and criminal cases in which the parentage of a child is relevant. In these contexts, the likelihood ratio is called the *paternity index* [10].

Posterior Probabilities

Neither the likelihood ratio nor a frequency gives a judge or jury the quantity that it ultimately seeks – the probability that a given hypothesis is true. That probability can be obtained according to Bayes' rule by combining a likelihood function with a prior probability function. Efforts to do so, however, have met with a mixed reception in the courts. Articulating the prior probability is the major sticking point. Several approaches have been discussed in the case law and legal literature: (i) applying a uniform prior distribution across the suspect population; (ii) assuming that prior probability is one-half; and (iii) producing a list of posterior probabilities for a broad range of priors.

Assuming a Uniform Prior in an Enumerated Population. In rare cases, it is possible to enumerate a suspect population. For example, in *State v. Klindt*, 389 N.W.2d 670 (Iowa 1986), the prosecution sought to prove that a torso found in the Mississippi River was the remains of a woman named Joyce Klindt. Investigators narrowed the possibilities among a list of missing persons in four nearby states to four women. A statistician computed likelihoods based on immunogenetic markers and other information about these women and applied these to a uniform distribution of prior probabilities for the four missing women. At the murder trial of Joyce Klindt's husband, the statistician testified that "[b]ased upon the likelihood of the concurrence of these factors among the missing

women, . . . the probabilities were over ninety-nine percent that the torso was Joyce Klindt's rather than any of the other three." *Id*. at 671.

The defendant objected to this testimony, but the objection was "basically that the underlying data was [sic] not sufficiently established." *Id*. The Supreme Court of Iowa quickly disposed of this claim upon determining that the data used to derive likelihood ratios were reliable. However, the court did not discuss the statistician's assumption that the prior probability with respect to each woman was 1/4.

Assuming a Prior of One-Half. In thousands of paternity cases, a posterior "probability of paternity" is calculated using a likelihood ratio (involving genetic information on the mother, child, and alleged father) and an arbitrarily chosen prior probability of 1/2. In these cases, the expert normally makes no effort to examine other possible fathers and to ascertain the prior odds for each such man. Rather, the alternative hypothesis usually is taken to be that a single "random man" is the father and that the prior probability of this hypothesis is 1/2. Testimony as to the resulting posterior probability, with or without disclosure of the prior odds, usually has been admitted with little appreciation of the nature of the computation [11]. The Supreme Court of Oregon, however, disapproved of the practice of using fixed prior odds. It announced in *Plemel v. Walter*, 735 P.2d 1209 (Or. 1987), that experts who desired to testify to a posterior probability could do so only by presenting a comprehensive table of prior and posterior probabilities, leaving it to the jury to decide which, if any, prior odds to use [12].

Bayesian computations also are used in criminal cases in which the defendant is alleged to be the father of a child or fetus resulting from rape or incest as well as in homicide cases involving missing bodies, and a few other offenses [13]. Although here too the use of the arbitrary prior probability of 1/2 often is not challenged, several courts have reversed convictions based on such testimony. The leading cases are *State v. Spann*, 617 A.2d 247 (N.J. 1993), and *State v. Skipper*, 637 A.2d 1101 (Conn. 1994). In *Spann*, a guard at a county jail was convicted of sexually assaulting a woman incarcerated there. human leukocyte antigen (HLA) tests involving the woman's child showed an exclusion probability of 99% and a paternity probability of 96.55%. On cross-examination, defense counsel established that the 96.55% figure presupposed a prior probability of 50%, but the expert characterized this premise as "neutral", "purely objective", and "one of the beauties of the test". 617 A.2d at 252. The prosecutor argued in summation that "guilt. . . . is proved to a mathematical certainty by carefully applying an objective scientific technique to the hard facts of this case." *Id*. at 249.

The Supreme Court of New Jersey held that the prosecution's portrayal of the posterior probability was unfair and that the failure to explain what effect other prior probabilities would have had on it was misleading. In dictum, the court stated that "the expert's testimony should be required to include an explanation to the jury of what the probability of paternity would be for a varying range of such prior probabilities, running, for example, from. 1 to. 9." *Id*. at 254.

In *State v. Skipper*, 637 A.2d 1101 (Conn. 1994), defendant was charged with various offenses arising from a sexually abusive relationship that he initiated with the young daughter of a neighbor. DNA testing performed on an aborted fetus, the girl, and the defendant resulted in a likelihood ratio of 3496 and a corresponding paternity probability of 99.97%. Unlike *Spann*, the prior probability of 1/2 was never disclosed, even on cross-examination. Instead, an expert "explained" that "[t]he calculation is . . . simply a conversion of the paternity index to the percentage. So, it's 99.97 is basically 3500 to 1, which is odds . . . expressed as a percentage." *Id*. at 1105.

The court held that this expert evidence was inadmissible and reversed the defendant's conviction. It stated that "[t]he assumption that there is a substantial possibility that the defendant had intercourse with the victim . . . raises serious concerns in sexual assault cases. It is antithetical to our criminal justice system to presume anything but innocence at the outset of a trial." Indeed, the opinion suggests that any nonzero prior probability – no matter how small – undermines the presumption of innocence:

> Whether a prior probability of 50 percent is automatically used or whether the jury is instructed to adopt its own prior probability, when the probability of paternity statistic is introduced, an assumption is required to be made by the jury before it has heard all of the evidence – that there is a quantifiable probability that the defendant committed the crime.

Id. at 1107–08 (footnote omitted).

The *Skipper* opinion, however, fails to explain why or how the presumption of innocence means

that the prior probability is zero. The conventional understanding of the "presumption" is that the mere fact of a criminal charge may not be considered supportive of guilt; hence, if the government adduces insufficient evidence at trial to permit a reasonable jury to conclude beyond a reasonable doubt that defendant is guilty as charged, the "presumption" is not overcome, and defendant is entitled to a directed verdict of acquittal [14]. Another interpretation is that *ab initio*, "the defendant should be thought no more likely than anyone else to be guilty" [15–17].

Varying the Prior Odds. Although American courts are divided as to the admissibility of posterior probabilities derived via Bayes' rule with undisclosed prior probabilities – particularly in criminal cases [13] – no appellate opinion holds that it is impermissible to permit testimony about the impact of the trace evidence on the prior probability. This testimony could take two forms. In the first, the jurors are invited to select the prior probability on the basis of the other evidence in the case, and to use the trace evidence according to Bayes' rule to deduce a posterior probability. This approach was employed by the defendant in *R. v. Adams* [1996] 2 Cr. App. R. 467, and, on retrial, [1998] 1 Cr. App. R. 377. The Court of Appeal denigrated such te stimony as "a recipe for confusion, misunderstanding and misjudgment." 1 Cr. App. R. at 384. Similar remarks can be found in American cases, although these are dicta since no such presentation was used. For example, People v. Nelson, 48 Cal.Rptr.3d 399 (Ct. App. 2006), *aff'd*, 185 P.3d 49 (Cal. 2008). This condemnation seems overstated, but the approach is criticized in more detail in [5], pp. 67–71.

In the second approach, no juror is asked to settle on a single figure for the prior probability, but only to inspect a table of prior and posterior probabilities to appreciate the power of the trace evidence [11]. This variable-prior-probability method avoids certain logistical, psychological, and jurisprudential difficulties that confront the juror-choice-of-prior-probability method. Jurors may not be very skilled at translating their impressions of the nonstatistical evidence into a single number; unless they actually articulate this number before learning of the trace evidence, they may be tempted to double-count the trace evidence by including it to some extent in the prior probability estimate; and arriving at a single number for the posterior probability would invite a comparison to a threshold probability required for conviction – a quantity that is difficult and perhaps undesirable to articulate [18].

Categorical Opinions

Experts often are asked for opinions, not numbers. The basic question then becomes: When should opinions about the impact of the trace evidence be admissible? These characterizations might range from the relatively uninformative statement that two samples are "consistent" with each other (or that they "match," or that the suspect "cannot be excluded") to pointed descriptions of the value of the match (e.g., "strong evidence of a common source") or of the infrequency of the matching features (e.g., a "very rare" type, or a "unique" match).

"Consistent-with" Testimony

For some types of trace evidence, such as hair fibers, toolmarks, and handwriting, little or no frequency data have been compiled. Some experts therefore describe the striking similarities between the samples, report that they match, and leave it at that. Indeed, in response to doubts about the ability of analysts to ascertain authorship on the basis of handwriting samples or to determine that a "rolled" fingerprint and a "latent" print come from the same person, a few courts have confined experts to testimony about the existence of the match [3, § 12.5.1].

The difficulty with this approach is obvious – the jury has no experience with which to judge how probative the esoteric characteristics are. Intuition and the experience of forensic examiners suggest that the failure to exclude the defendant surely is relevant – hair fibers and toolmarks, for example, are variable – but this variation has not been quantified. Consequently, the juror presented merely with testimony that the defendant or the material in his possession is consistent with the trace is hard-pressed to know what to make of the match.

As such, testimony that stops after reporting a "match" arguably is unfairly prejudicial. However, opposing counsel should be able to put the evidence in perspective by pointing out that for all that the expert has said, the trace might have come from anyone in the courtroom. If the evidence is indeed relevant and if the risk that the jury will be unduly impressed can be readily countered, then testimony

that the sample from the defendant is "consistent with" the trace should be admissible despite the difficulty in ascertaining just how incriminating the evidence is.

It also could be argued that the testimony of a match without substantial data on the frequency of the matching characteristics is not sufficiently scientific to be admitted under the standard of scientific validity adopted by the US Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals, Inc*, 509 U.S. 579 (1993). However, the analyst may be able to declare matches and nonmatches accurately, and even without systematic study of population data, it may be clear to scientists that the outcomes are at least somewhat probative. The question for the law is then whether the trace evidence is of sufficient assistance to justify its admission despite the risk that it will be overvalued. Under American and English law, the finding of a match should be admitted unless the jury will give so much weight to the evidence that its prejudicial impact substantially outweighs its probative value.

Qualitative Expressions for Probability and Likelihood Ratios

The conclusion that testimony about the matching nature of two samples often should be admissible notwithstanding the lack of systematic data on the population frequencies conflicts with a line of cases that rely on the general-acceptance standard of *Frye v. United States*, 293 F 1013 (D.C. Cir. 1923), to hold that a DNA match is inadmissible unless the expert gives a numerical frequency [3, § 12.5.2]. But these cases confuse a question of science with an issue of law. What should be well established in science is that the methodology has some reasonable individualizing power. How the findings made with a generally accepted technology should be presented to achieve a satisfactory balance between probative value and prejudice is a quintessentially legal question, not a scientific one. Where empirical knowledge justifies characterizing a matching type as "rare," "extremely rare," or the like, the expert should be permitted to say as much. See, e.g., State v. Bloom, 516 N.W.2d 159, 166–67 (Minn. 1994); [19].

Somewhat more problematic are statements from the expert that the evidence is "weak", "good", "very strong", etc., intended as a description of the likelihood ratio for the evidence. If the expert makes the actual meaning of these phrases clear by providing the underlying likelihood ratio, then it is not obvious that they add much to that more precise numerical statement, although neither is any great prejudice apparent [9], p. 57. Yet, if it is not made clear that the expressions are merely descriptions of the likelihood ratio, whether the jury will make reasonable use of them is questionable. In that situation, the balance of probative value and prejudicial effect could favor exclusion.

Uniqueness

The most extreme case of a purely verbal description of the infrequency of the matching features arises when those features are said to be unique (or the defendant is said to be the source). Uniqueness, the condition where there is one and only one occurrence of the particular features in a population, is impossible to prove directly for an infinite population [20], but scientific research and common experience have led to judgments of extreme rarity for a few types of trace evidence. For example, courts have taken "judicial notice" of the uniqueness of fingerprints.[a]

Sometimes an expert's claim of "uniqueness" is not generally accepted or adequately established by scientific research. In these situations, the opinions should not be admitted under the *Frye* or *Daubert* standards. However, "[t]here is no 'bright-line' standard in law or science that can pick out exactly how small the probability of the existence of a given profile in more than one member of a population must be before assertions of uniqueness are justified . . . " [21]. When a great number of samples have been compared to one another and shown to be distinct, and when the process that gives rise to the population is known to involve an infinitesimal probability of generating a duplicate type for each unit or individual in the population, courts should allow an opinion of probable uniqueness along with, or instead of testimony about the vanishingly small random match probability [3, § 12.5.3].

End Notes

[a.] The doctrine of judicial notice permits a court to accept as true a fact that is not subject to reasonable dispute. The application of the doctrine to fingerprinting, however, was successfully contested in *United States v. Mitchell*, 365 F.3d 215, 252 (3d Cir. 2004).

References

[1] Committee on Scientific Assessment of Bullet Lead Elemental Composition Comparison, National Research Council (2004). *Weighing Bullet Lead Evidence*, National Academy Press, Washington, DC.

[2] Kaye, D.H. (2006). The current state of bullet-lead evidence, *Jurimetrics: The Journal of Law, Science, and Technology* **46**, 99–114.

[3] Kaye, D.H., Bernstein, D. & Mnookin, J. (2004). *The New Wigmore, A Treatise on Evidence: Expert Evidence*, Aspen, New York.

[4] Evett, I.W. & Weir, B.S. (1998). *Interpreting DNA Evidence, Statistical Genetics for Forensic Scientists*, Sinaur Associates, Sunderland.

[5] Redmayne, M. (2001). *Expert Evidence and Criminal Justice*, Oxford University Press, Oxford.

[6] Edwards, A.W.F. (1972). *Likelihood: An Account of the Statistical Concept of Likelihood and Its Application to Scientific Inference*, Cambridge University Press, Cambridge.

[7] Royall, R.M. (1997). *Statistical Evidence: A Likelihood Paradigm*, Chapman & Hall, London.

[8] Kaye, D.H. & Koehler, J.J. (2003). Misquantifying probative value, *Law and Human Behavior* **27**(6), 645–659.

[9] Robertson, B. & Vignaux, G.A. (1995). *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*, John Wiley & Sons, Chichester.

[10] Kaye, D.H. (1989). The probability of an ultimate issue: the strange cases of paternity testing, *Iowa Law Review* **75**, 175–109.

[11] Ellman, I.M. & Kaye, D.H. (1979). Probabilities and proof: can HLA and blood test evidence prove paternity? *New York University Law Review* **55**(6), 1131–1162.

[12] Kaye, D.H. (1988). Plemel as a primer on proving paternity, *Willamette Law Journal* **24**(4), 867–883.

[13] Faigman, D., Kaye, D.H., Saks, M. & Sanders, J. (2006–2007). *Modern Scientific Evidence: The Law and Science of Expert Testimony*, West Group, St. Paul, Minnesota, Vol. 4, p. 7.

[14] Wigmore, J.H. (1981). *Evidence in Trials at Common Law (Chadbourn Rev.)*, Little, Brown, and Company, Boston, Vol. 9, p. 2511.

[15] Allen, R.J., Balding, D.J. Donnelly, P. Friedman, R., Kaye D.H., LaRue, L.H., Park R.C., Robertson, B. & Stein, A. Probability and proof in *State v. Skipper*: An Internet exchange, *Jurimetrics: The Journal of Law, Science and Technology* **35**(3), 292–310.

[16] Dawid, A.P. (1994). The island problem: coherent use of identification evidence, in *Aspects of Uncertainty: A Tribute to D. V. Lindley*, P.R. Freedman & A.F.M. Smith, eds, John Wiley & Sons, New York, pp. 159–170.

[17] Dawid, A.P. (2002). Bayes's theorem and juries, in *Bayes's Theorem*, R. Swinburne, ed, Oxford University Press, New York, pp. 71–90.

[18] Tribe, L. (1971). Trial by mathematics: precision and ritual in the legal process, *Harvard Law Review* **84**(8), 1329–1391.

[19] Kaye, D. (2009). *The Double Helix and the Law of Evidence: Controversies over the Admissibility of Genetic Evidence of Identity*, Harvard University Press, Cambridge.

[20] Saks, M. & Koehler, J.J. (2008). The individualization fallacy in forensic science, *Vanderbilt Law Review* **61**, 199–219.

[21] Committee on Forensic DNA Science, National Research Council (1996). *The Evaluation of Forensic DNA Evidence*, National Academy Press, Washington, DC, p. 194.

DAVID H. KAYE

Interpretation: Low Template DNA

Introduction

The term *low template DNA* (LTDNA) has become the preferred term to describe any sample where there are a few DNA molecules present [1]. The term has been coined to avoid confusion with low copy number (LCN), which rather describes one particular enhancement technique. LCN has become synonymous with elevated PCR cycle technology [2–5]. In modern forensic laboratories, an analyst using the term *LCN* usually means amplification using 34 cycles probably with a duplication and consensus strategy for LTDNA samples The connection between LTDNA and 34 cycles is valid in part, but not comprehensively useful. It may be possible to obtain profiles from trace DNA amplified at 28 cycles (or indeed any other cycle number) and equally not all 34-cycle work is done on truly trace DNA. Equally there are other ways to increase the sensitivity of the analysis than elevation of the number of short tandem repeat (STR) amplification cycles, for example nested PCR [6, 7], MiniSTR [8] and whole genome amplification (WGA) [9, 10], post-PCR clean-up, and enhanced injection volume [11].

Despite being developed in the LTDNA context at least some of the concepts discussed in this chapter have application to all profiles, regardless of

the amplification strategy, wherever profile quality is compromised because of presumed low template number. Equally, 34-cycle amplification may produce profiles that appear to suffer none of the compromising aspects.

Increasing the number of PCR amplification cycles from 28 allows the amplification of less than 100 pg of total genomic DNA. It is expected that 3.5 pg contains one haploid DNA template, therefore 100 pg should equate to approximately 30 haploid templates or 15 diploid templates. With LTDNA analysis, results may be less reproducible compared with analysis involving more than 100 pg of DNA, and analysis may be more prone to contamination. However, the quantification step in DNA amplification is not completely accurate at low levels. The consequence of this is that quite often the LCN technology is applied to samples that give a low quantification result but which produce reproducible profiles with large peaks and no apparent signs of any compromise to quality.

Profile Quality

The low template number and increased amplification cycles may affect the morphology of LTDNA profiles [12]. The most noticeable effects are observed in/on dropout, heterozygote balance ratios, and stutter size. There are two postulated causes for this: sampling from low-level templates and stochastic effects in the early amplification rounds. When the template number is small, it is possible that each allele is unequally represented in an aliquot from the sample. Equally, if the number of templates is small, a stutter event or nonamplification in the early rounds of amplification can have a marked effect on the amount of final allelic and stutter products produced. However, as the heights of the peaks in the profile increase, the performance of 34-cycle amplification approaches that of 28-cycle amplification with normal template levels. For ultra trace work, it is impossible to completely remove the possibility contaminant alleles.

Heterozygote Balance and Dropout

If the sampling and amplification of template were perfect for each strand of haploid DNA template, then the two peaks of a heterozygote should be of equal height or area (hereafter height). However, in all PCR work some variation is induced in the resulting peaks' heights. For 34-cycle low template work, the induced variation is larger. There are two terms in common usage to describe heterozygote balance. These are

$$Hb = \frac{\text{height of the smaller peak}}{\text{height of the larger peak}} \quad \text{or} \qquad (1)$$

$$Hb = \frac{\text{height of the higher molecular weight peak}}{\text{height of the lower molecular weight peak}} \qquad (2)$$

The latter contains more information (the order of the alleles is kept) whereas the former is in more common usage. For 28-cycle full template work it is expected that the heterozygote balance (defined the second way) is expected to lie approximately between 0.5 and 2.0. For LTDNA work, it can be anywhere between 0 and infinity. A heterozygote balance of 0 or infinity is equivalent to the event of allelic dropout.

The distribution of *Hb* for LTDNA templates amplified at 34 cycles shows the expected peak slightly below 1, which arises from the slight tendency for lower molecular weight alleles to amplify better (Figure 1).

It is usually possible to examine empirical data from known heterozygotes to determine a peak height above which dropout has not been observed. This is often termed the *homozygote peak height threshold guideline* and may be of the order of 3000 rfu, although each laboratory would need to determine their homozygote peak height threshold as part of their validation. Therefore, if a single peak at a locus is present at a height greater than the threshold it is potentially possible to rule out dropout and designate the locus as homozygous.

Stutter Size

Stutter peak heights (relative to the size of the associated allele peak) can be bigger for low template 34-cycle work. This is thought to occur because a stutter in an early round of amplification has a proportionally larger contribution to the final product if there are few templates (Figure 2).

A stutter guideline is often defined below which a peak may be composed completely of stutter product. This does not mean that it must all be stutter, it is still possible that part of the peak is allelic, but that there is no need to invoke the presence of an allele to explain

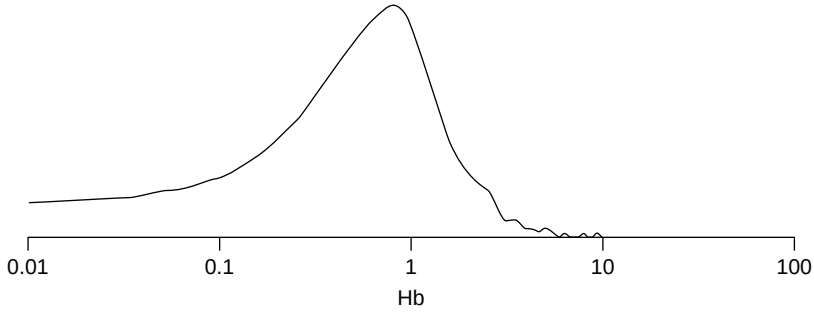


Figure 1 Distribution of the frequencies of *Hb* values for a set of $N = 1688$ heterozygote profiles obtained after 34 cycles [13]

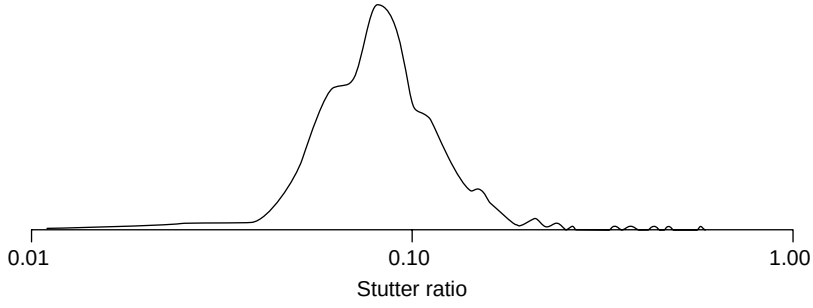


Figure 2 Distribution of the frequencies of *stutter ratio* values for a set of $N = 844$ loci heterozygote profiles obtained after 34 cycles [13]

the peak. Just because a peak could be stutter does not mean that it is stutter. Given the greater variance in stutter ratio at LTDNA levels, broader criteria are often used than for 28-cycle work with good template levels. Each laboratory should determine their own stutter guidelines as part of their validation. Application of these guidelines should always be undertaken in conjunction with duplication to report consensus profiles and should be done before looking at the reference profiles.

Contamination/Drop-in

It has been proved impractical to completely eliminate the possibility of laboratory-introduced contamination in LTDNA work, even when strict conditions of cleanliness are in place. Many readers would respond to this statement with deep concern. However, it is also impractical to eliminate the possibility of contamination of samples at the scene or at other stages of the evidence chain. It is of utmost importance for every forensic laboratory to minimise the possibility of gross contamination of the crime sample with the suspect's sample. It is also important that the laboratory adopts protocols that minimize the opportunity for the introduction of DNA from analysts involved in the collection or analysis of samples and the potential contamination from laboratory based sources, such as nonDNA free plasticware [14]. This can be achieved through appropriate training and good laboratory practices together with the establishment of elimination databases for staff and other relevant personnel. Further, confidence in the "health" of the laboratory can be gauged from appropriately targeted environmental monitoring. Regular environmental testing of surfaces assists in determining when a more through deep clean of these surfaces is required.

Most contamination events observed in LTDNA profiles appear to be a single allele or a few alleles, often of small peak height. The term *drop-in* has

been coined to describe this phenomenon. Drop-in rates for an operational LTDNA laboratory should be constantly monitored and the laboratory should be able to demonstrate that their drop-in rate is low. This may typically be within a range of 1–3% per locus or even less. By regular monitoring of the drop-in rates, any unacceptable rise can be easily observed. Appropriate investigations into the cause of the increase and appropriate cleaning can be initiated. The interpretation procedures recommended by Gill and Buckleton [15] are very robust to contamination events of this type, as long as the possibility of introducing the suspect's DNA is eliminated and the drop-in rate does not exceed 30% per locus.

The Interpretation of LTDNA Profiles

LTDNA profiles, single stain or mixtures, may be interpreted using a frequentist or likelihood-ratiobased approach [16]. The usual strengths of the likelihood-ratio approach also apply to LTDNA work and are not repeated here. However, for those people philosophically opposed to, or constrained not to use such an approach, there are still valid ways in which the profiles may be interpreted.

It is often impossible to assign the source of a trace DNA sample to a particular body fluid or cell type. Nor is it likely to be straightforward to state when or how the DNA was deposited. Difficulties include demonstrated secondary transfer [5, 17–19].

Any case that involves prior contact between persons of interest in an investigation is likely to be unsuitable for LTDNA analysis. The possibility of prior contact is to be considered along with other information relevant to the case, before LTDNA analysis was attempted [20].

As the reporting of the number of DNA contributors to LTDNA profiles cannot always be undertaken with the same degree of confidence as the reporting of 28-cycle profiles, a caveat regarding the assumption of contributor number is usefully included as part of the reporting regime.

Typically scientists interpreting LTDNA work operate at what has been termed the *sublevel 1 level of propositions* [21–24].

Using the Bayesian approach two propositions are formed that align with the prosecution and defense viewpoints.

H_p: the DNA in the crime stain is from the suspect and

H_d: the DNA in the crime stain is from someone else.

The use of the term DNA rather than blood, semen or saliva, is what makes this a sublevel 1 set of propositions.

The current interpretation strategies are based upon replicate amplifications. This requires that a LTDNA sample is amplified at least twice. This approach improves the reliability of the technique considerably and ameliorates most of the perceived negative aspects. Interpretation may proceed via what is termed the *biological* or the *statistical* model.

The Biological Model

The biological model uses a "consensus" approach. A "consensus profile" is developed by "scoring those alleles that appear in two or more of the replicates but not at similar heights in the negative controls". This consensus should be developed before the reference samples are examined. The consensus profile many contain zero, one, two, or more alleles per locus. In some instances, it can be difficult to determine whether the resulting consensus profile arises from one or more individuals. Peak height imbalance is less reliable in this determination than in normal template work. The decision as to how many individuals are present in the consensus profile is one of the more problematic decisions in LTDNA casework.

If there are between zero and two alleles per locus in the consensus profile, it is normal to treat the profile as unmixed.

For example, Table 1 shows two amplification results for a sample; replicate 1 and replicate 2, and the resulting consensus profile generated prior to consideration of the results from the corresponding extraction negatives.

The next step is to consider the DNA profiles from the extraction negative associated with the sample. If duplicated alleles are sighted in the extraction negatives that correspond in height to those in the consensus profile generated for the sample, then these alleles are removed from the profile. These duplicated alleles could indicate a contamination event rather than a random spurious drop-in event.

In the example in Table 2, the alleles in the extraction negative profiles correspond in height to

Table 1 Amplification results from two replicates of the same evidential sample

	D3	vWA	D16	D2	D8	D21	D18	D19	TH0	FGA	Amel
Rep 1	15,16	16	11,12,13	18,19	11	30,31	18	14	6,7,8	20,23	X,Y
Rep 2	15	NR	11,13	18	10	30	19	14	6,9	20,23	X
Consensus	15,0	NR	11,13	18,0	NR	30,0	NR	14,0	6,0	20,23	X,0

Where "NR" means no result, 0 is a code that signifies that an unseen allele may be present (F in the United Kingdom) and "Rep" means one replicate amplification of sample

Table 2 Removal of alleles contained in the extraction negatives

	D3	vWA	D16	D2	D8	D21	D18	D19	TH0	FGA	Amel
Sample consensus	15,0	NR	11,13	18,0	NR	30,0	NR	14,0	6,0	20,23	X,0
E Neg 1	16	–	13	–	–	–	–	–	–	23	–
E Neg 2	–	15,17	–	–	–	27	–	–	9	–	–
Report	15,0	NR	11,0	18,0	NR	30,0	NR	14,0	6,0	20,0	X,0

those in the consensus profile. The presence of the D16 13 allele and the FGA 23 allele results in these two alleles being removed from the consensus profile.

If the peak heights of alleles in an extraction negative profile are markedly smaller in height than the peak heights of corresponding alleles in a consensus profile, then the reporting scientist may conclude that the alleles detected in the extraction negative could not have contaminated the sample. Therefore, in these circumstances it may be appropriate for the reporting scientist not to delete alleles from the consensus profile of the sample.

An assumed single source stain is interpreted/assessed statistically using either the match probability or the likelihood-ratio approach. In many cases, those loci with one allele present, say allele a, cannot be assumed to be homozygous. In such cases, the "2p" rule is typically applied. This rule assigns the match probability as either $2p_a$ using the product rule where p_a is the allele probability for allele a, or preferably one of a number of variants that include the subpopulation correction [25]. These are

$$\frac{[2\theta + (1-\theta)p_a]}{(1+\theta)}\left(2 - \frac{[3\theta + (1-\theta)p_a]}{(1+2\theta)}\right)$$

One of these variants, for a homozygous defendant of type *aa* is

$$\frac{2[2\theta + (1-\theta)p_a]}{(1+\theta)}$$

and

$$\frac{[\theta + (1-\theta)p_a]}{(1+\theta)}\left(2 - \frac{[2\theta + (1-\theta)p_a]}{(1+2\theta)}\right)$$

for a heterozygous defendant of type *ab*. This is often approximated as

$$\frac{2[\theta + (1-\theta)p_a]}{(1+\theta)}$$

The biological model reduces the replicates to a consensus profile and in doing so is wasteful of evidence. This tends to be conservative if H_p is true but is not conservative if H_d is true (this statement is universally true of all transformations or reductions that waste evidence). Of most concern are the situations where the suspect is an *ab* heterozygote and

- one replicate is *ab* and the second replicate is *ac* and hence the consensus profile shows the *a* allele. There is an additional nonreplicated *c* allele.
- each replicate is *a* and hence the consensus profile shows the *a* allele and no others. The situation is at its most extreme if the *a* peaks are near, but just below, the threshold.

If caseworkers are vigilant for these situations the biological model is expected to operate in a conservative manner under most realistic circumstances. However, the perceived deficiencies led to searches for methods that utilize the evidence more thoroughly.

The Statistical Model

The only model currently available that attempts to use information from the LTDNA profile more thoroughly is termed the *statistical model*. This treats the alleles as either present or absent, and also models dropout, drop-in, and stutter as binary events. This model may be used to explore the robustness of the biological model or substituted for it in casework. The only forensic service currently using this model in casework is the FSS in the United Kingdom. In that service it is embodied in the software LoComatioN [26].

While this model does make more comprehensive use of the information, it still wastes some. Specifically the treatment of the alleles as present or absent wastes the information inherent in peak height. It is still unclear how large this loss is when LTDNA profiles are considered, as peak height is a much less reliable indicator.

If there are more than two alleles at some loci then it is likely that the profile will be treated as a mixture. DNA profile morphology including relevant peak heights in the replicates of a sample and the number of alleles observed at each locus across the multiple amplifications should always be considered when determining the likely number of contributors to the DNA profile. When peak heights are small, the relative heights of peaks are virtually uninformative in mixture analysis. Such mixed profiles should be classified as uninterpretable.

If there are multiple contributors, it may be possible to separate the contributions into major and minor components. In such a sample, the peak heights of one DNA component is markedly greater than the peak heights of other DNA component(s) in the sample and the profile of the major DNA component is consistent between two replicate profiles.

For some mixed profiles, it may be appropriate to condition on a known DNA contributor to the sample, for example, if the sample is from an item of worn clothing. By conditioning on the wearer's DNA profile it may be possible to determine the profile of any additional DNA contributor, provided this profile is consistent between two replicates and well above any background DNA that may also be present.

The Statistical Model Theory

As noted, the statistical model, first proposed by Gill *et al.* [27] considers three separate phenomena as follows:

C: the event that an allele appears at a given locus in the crime sample as a result of contamination/drop-in. The complement to this event is denoted by $\bar{C}$ with $\Pr(C) = 1 - \Pr(\bar{C})$.
D: the event that a given allele "drops out" because of stochastic variation in the amplification. The complement to this event is denoted by $\bar{D}$ with $\Pr(D) = 1 - \Pr(\bar{D})$.
St: the event that a band has been caused by stuttering of an adjacent allelic peak.

In general, we have replicates $R_1, R_2, \ldots, R_n$ although in practise n is generally 2. We proceed using the notation given in Curran *et al.* [28]. Under any given proposition, H, we have three sets to consider:

T – the alleles of those people who are known (i.e., they have been identified and have contributed DNA to a sample that has been typed) and who are assumed to have contributed to the mixture.
V – the alleles of those people who are known but who are assumed to have not contributed to the mixture.
U – the alleles of the unknown contributors.

The contents of the sets T, V, and U are dependent on the proposition under consideration (H). We wish to calculate

$$\begin{aligned}&\Pr(R_1, R_2, \ldots, R_n, T, U, V \mid H)\\ &= \Pr(R_1, R_2, \ldots, R_n \mid T, U, V, H)\\ &\quad \times \Pr(T, U, V \mid H)\end{aligned} \tag{2}$$

When the number of unknown contributors to the stain is nonzero, the set U may contain more than one mutually independent genotype, therefore this becomes

$$\begin{aligned}&\Pr(R_1, R_2, \ldots, R_n\ T, U, V \mid H)\\ &= \sum_j \Pr(R_1, R_2, \ldots, R_n \mid T, U_j, V, H)\\ &\quad \times \Pr(T, U_j, V \mid H)\end{aligned} \tag{3}$$

where U_j is the jth possibility for genotypes (or more specifically alleles) of the unknown contributors. The net effect of the hypothesis H is to dictate the elements of T, V, and U, therefore in the following steps we assume that H has been accounted for. Furthermore, if the set U is nonempty (and contains m elements) then it is possible to simplify (1) for computation to

$$\Pr(T, V) \sum_{j=1}^{m} \Pr(R_1, R_2, \ldots, R_n \mid T, U_j, V) \times \Pr(U_j \mid T, V) \tag{4}$$

In most circumstances the union of the sets V and T remains the same in the numerator and denominator in which case we can cancel the term $\Pr(T, V \mid H)$. Using the assumption that the replicates are independent of each other, this expression becomes

$$\sum_{j=1}^{m} \left[\prod_{i=1}^{n} \Pr(R_i \mid T, U_j, V) \right] \Pr(T, U_j, V) \tag{5}$$

Furthermore, the replicate information does not depend on people who have not contributed so that $\Pr(R_i \mid T, U_j, V) = \Pr(R_i \mid T, U_j)$

$$\sum_{j=1}^{m} \left[\prod_{i=1}^{n} \Pr(R_i \mid T, U_j) \right] \Pr(T, U_j, V) \tag{6}$$

Expression (4) allows us to reduce the problem into two nested but discrete steps: evaluation of the genotype probabilities and evaluation of the probabilities of the replicates with respect to genotype information.

Assessing Genotype Probabilities

The evaluation of the genotype probabilities, with the inclusion of population substructure, is given in Curran *et al.* [28]. There is one minor difference in that the possible values for the unknown contributors are not constrained in the same way. In general mixture interpretation genotypes of unknown contributors are restricted to alleles that have appeared in the crime scene sample. Furthermore, the set of unique alleles observed in the known contributors, T, and any unknown contributor combination must be equal to the set of unique alleles observed in the crime scene sample. For example, if the scene has alleles $C_g = \{a, b, c\}$ then the combination of $T = \{a, b\}$, $U = \{c, c\}$ is allowable whereas $T = \{a, b\}$, $U = \{b, b\}$ is not because it does not explain the presence of the c allele. With the possibility of drop-in, such a restriction must obviously be removed. Furthermore, allelic dropout allows the proposal of alleles that have not been seen in the crime scene sample. From a computational and practical point of view, it is useful to have some mechanism for generating candidate genotypes for the unknown contributors. One possibility is to allow the choice of drawing $2x$ alleles from the set of unique alleles observed in all of the replicates, where x is the number of unknown offenders.

Finally, there can be good reasons for not using all possible genotypes such as particular combinations being inconsistent with the estimated mixing proportion of the sample, or inconsistent with preferential amplification rules. See Gill *et al.* [29] for a deeper discussion of these issues.

Assessing Replicate Probabilities

In a manner similar to mixture interpretation, assessing replicate probabilities can be made easier by partitioning replicate information into a number of disjoint sets.

Firstly, we collapse T, U_j into a set of unique alleles G_g, For the time being let us regard G_g as a set of unique alleles for a particular combination and let R be a particular replicate, then $\Pr(R_i \mid T, U_j) = \Pr(R \mid G_g)$. We can evaluate this probability by constructing three sets:

1. The alleles that have dropped out Δ (delta).
2. The alleles that have not dropped out ρ (rho).
3. The potential contaminants/drop-in peaks or stutter peaks/alleles χ (chi).

Using this notation we can now define our three sets

$$\begin{aligned} \Delta &= G_g \ \{R|G_g\} \\ \mathrm{X} &= R \ \{R|G_g\} \\ \mathrm{P} &= R \ \{\mathrm{X} \cup \Delta\} \end{aligned} \tag{7}$$

where "\" is the set exclusion operator. The set $\{\mathrm{X} \cup \Delta\}$ is a conventional set in that it doesn't contain duplicated elements. From these sets we can write down the probability of any replicate, given any genetic combination using the following rules:

1. For every element of P write $\Pr(\bar{D})$.
2. For every element of X write $\Pr(C)p_i$ where i is the ith element of X and p_i is the probability of the ith allele in the population. If $X = \phi$ then write $\Pr(\bar{C})$.
3. For every element of Δ write $\Pr(D)$.

There are two additional factors that should be considered in our rule regarding the elements of the set of contaminants, X. The first is the concept of ordering. Consider a situation where X contains more than one element, e.g., where the replicate contains alleles 13, 14, 15, and the putative genotype for the one unknown contributor is 13,13. In this case X would contain alleles 14 and 15. There is no particular reason to think that the order in which these contaminants were "added" has any relevance, therefore we must account for the possible orderings of these alleles – in this case we should multiply the probability by a factor of two. In general, if X contains k alleles, then there are $k!$ possible arrangements. The second factor that we might consider is the effect of population substructure. However, we argue that the allele probabilities for contaminants are more likely to be related to sources within the laboratory rather than the population of the offender. In arguing so, we admit then, that using the population frequencies for the contaminant alleles are probably incorrect and that treating contamination events as independent on an allele by allele basis is also incorrect. However, they suffice as an approximation.

Conclusion

LTDNA interpretation is considered the hardest DNA evidence type to interpret and is undertaken by the most experienced analysts after extensive training. Analysts would have spent some time interpreting simple stains before moving on to mixtures and LTDNA cases [30].

It is acknowledged that in many cases it is difficult to associate a LTDNA DNA profile to its DNA source (e.g., semen, blood, or saliva), because of the small amounts of cellular material present in the sample. In addition, owing to the extreme sensitivity of the analysis, we can expect to obtain DNA profiles from trace biological material left for reasons unconnected with the facts under investigation. In other words, the relevance of such profiles is usually more difficult to establish compared with the relevance of DNA profiles obtained from rich samples extracted from visible biological stains. Research has shown that trace DNA can be obtained following primary, secondary, or tertiary transfer.

There have been expressions of concern in the literature and courts about the reliability of LCN work [31]. LTDNA technology had been offered originally as an investigative tool to the police, typically in cases where no names were offered for investigation or in cases reopened by the UK Criminal Case Review Commission. The transition from an investigation tool (aiming at generating leads for further investigation through the national DNA database) to evidence presented in court has not been straightforward. The recent ruling and discussion following the Omagh trial highlighted issues of contamination and the reliability of the technique [32–35]. In the aftermath of the Omagh bombing decision, a report commissioned by the Home Office has supported the validity of the technique [1]. The Forensic regulator has recommended some very sensible ways forward [36].

Since LTDNA analysis began in the late 1990s there has been considerable improvement in both profile quality and in the control of, and understanding, of the different sources of contamination that may occur. LTDNA analysis is costly, when compared with other examinations, and is usually reserved for targeted cases.

If careful attention is given to training, good scene, storage, and laboratory practice, robust interpretation guidelines and care in presenting appropriate caveats then LTDNA analysis can take its place as a forensic tool.

References

[1] Caddy, B., Taylor, G.R. & Linacre, A.M.T. (2008). *A Review of the Science of Low Template DNA Analysis*, Home Office, London.

[2] Van Oorschot, R.A. & Jones, M. (1997). DNA fingerprints from fingerprints, *Nature* **387**(6635), 767.

[3] Van Oorschot, R.A. & Jones, M. (1998). Retrieval of DNA from touched objects, *The 14th International Australia and New Zealand Forensic Science Society Symposium for Forensic Sciences*, Adelaide, October 1998.

[4] van Oorschot, R.A.H., Phelan, D.G., Furlong, S., Scarfo, G.M., Holding, N.L. & Cummins, M.J. (2003). Are you collecting all the available DNA from touched objects? *International Congress Series* **1239**, 803–807.

[5] Murray, C., Lowe, A., Richardson, P., Wivell, R., Gill, P. & Tully, G., *et al.* (2001). Use of low copy number (LCN) DNA in forensic inference, *Twelfth International Symposium on Human Identification*, Biloxi.

[6] Strom, C.M., Rechitsky, S., Wolf, G. & Verlinsky, Y. (1994). Reliability of polymerase chain reaction (PCR) analysis of single cells for preimplantation genetic analysis, *Journal of Assisted Reproduction and Genetics* **11**, 55–62.

[7] Strom, C.M. & Rechitsky, S. (1998). Use of nested PCR to identify charred human remains and minute amounts of blood, *Journal of Forensic Sciences* **43**(3), 696–700.

[8] Coble, M.D. & Butler, J.M. (2005). Characterization of new miniSTR loci to aid analysis of degraded DNA, *Journal of Forensic Sciences* **50**, 1–11.

[9] Hanson, E. & Ballantyne, J. (2005). Whole genome amplification strategy for forensic genetic analysis using single or few cell equivalents of genomic DNA, *Analytical Biochemistry* **346**, 246–257.

[10] Ballantyne, K.N., van Oorchot, R.A.H. & Mitchell, R.J. (2007). Comparison of two whole genome amplification methods for SYR genotyping of LCN and degraded DNA samples, *Forensic Science International* **166**(1), 35–41.

[11] Forster, L., Thomson, J. & Kutranov, S. (2008). Direct comparison of post-28-cycle PCR purification and modified capillary electrophoresis methods with the 34-cycle "low copy number" (LCN) method for analysis of trace forensic DNA samples, *Forensic Science International: Genetics* **2**(4), 318–328.

[12] Whitaker, J.P., Cotton, E.A. & Gill, P. (2001). A comparison of the characteristics of profiles produced with the AMPFlSTR® SGM Plus™ multiplex system for both standard and low copy number (LCN) STR DNA analysis, *Forensic Science International* **123**, 215–223.

[13] Petricevic, S., Whitaker, J., Buckleton, J.S., Patel, J., Simon, P., Vintiner, S., *et al.* Low Copy Number DNA profiling a valid forensic technique? *Forensic Science International: Genetics in draft.*

[14] Gill, P. & Kirkham, A. (2004). Development of a simulation model to assess the impact of contamination in casework using STRs, *Journal of Forensic Sciences* **49**(3), 485–491.

[15] Gill, P., Whitaker, J.P., Flaxman, C., Brown, N. & Buckleton, J.S. (2000). An investigation of the rigor of interpretation rules for STR's derived from less that 100 pg of DNA, *Forensic Science International* **112**(1), 17–40.

[16] Buckleton, J.S., Triggs, C.M. & Walsh, S.J. (2004). *DNA Evidence*, CRC Press, Boca Raton.

[17] Lowe, A., Murray, C., Whitaker, J., Tully, G. & Gill, P. (2002). The propensity of individuals to deposit DNA and secondary transfer of low level DNA from individuals to inert surfaces, *Forensic Science International* **129**, 25–34.

[18] Lowe, A.L., Murray, C., Whitaker, J.P., Tully, G. & Gill, P. (2002). The propensity of individuals to deposit DNA and secondary transfer of low level DNA from individuals to inert surfaces, *Forensic Science International* **129**(1), 25–34.

[19] Farmen, R.K., Jagh, R., Cortez, P. & Fryland, E.S. (2008). Assessment of individual shedder status and implication for secondary DNA transfer, *Forensic Science International: Genetics Supplement Series* **1**(1), 415–417.

[20] Evett, I.W., Gill, P.D., Jackson, G., Whitaker, J. & Champod, C. (2002). Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks, *Journal of Forensic Sciences* **47**(3), 520–530.

[21] Evett, I.W., Jackson, G. & Lambert, J.A. (2000). More on the hierarchy of propositions: exploring the distinction between explanations and propositions, *Science and Justice* **40**(1), 3–10.

[22] Cook, R., Evett, I.W., Jackson, G., Jones, P.J. & Lambert, J.A. (1998). A hierarchy of propositions: deciding which level to address in casework, *Science and Justice* **38**(4), 231–240.

[23] Cook, R., Evett, I.W., Jackson, G., Jones, P.J. & Lambert, J.A. (1999). Case pre-assessment and review in a two-way transfer case, *Science and Justice* **39**(2), 103–111.

[24] Cook, R., Evett, I.W., Jackson, G., Jones, P.P. & Lambert, J.A. (1998). A model for case assessment and interpretation, *Science and Justice* **38**(3), 151–156.

[25] Buckleton, J. & Triggs, C.M. (2006). Is the 2p rule always conservative? *Forensic Science International* **159**, 206–209.

[26] Gill, P., Kirkham, A. & Curran, J. (2007). LoComatioN: a software tool for the analysis of low copy number DNA profiles, *Forensic Science International* **166**(2–3), 128–138.

[27] Gill, P., Whitaker, J.P., Flaxman, C., Brown, N. & Buckleton, J. (2000). An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA, *Forensic Science International* **112**, 17–40.

[28] Curran, J.M., Triggs, C.M., Buckleton, J. & Weir, B.S. (1999). Interpreting mixtures in structured populations, *Journal of Forensic Sciences* **44**(5), 987–995.

[29] Gill, P., Sparkes, R., Pinchin, R., Clayton, T.M., Whitaker, J.P. & Buckleton, J. (1998). Interpreting simple STR mixtures using allele peak area, *Forensic Science International* **91**(1), 41–53.

[30] Gill, P. (2001). *The Development of An Expert System that Deskills LCN Reporting. Project Plan*, UK Forensic Science Service, Birmingham.

[31] Budowle, B., Hobson, D.L., Smerick, J.B., Smith, J.A.L. (2001). Low copy number – consideration and caution, *Twelfth International Symposium on Human Identification*, Biloxi.

[32] (2007). *The Queen v. Sean Hoey. In: Crown Court sitting in Northern Ireland*. Bill No: 341/05. Neutral citation no. [2007] NICC 49. Ref: WEI7021. December 20, 2007.

[33] McCartney, C. (2008). LCN DNA: proof beyond reasonable doubt? *Nature Reviews Genetics* **9**(5), 325.

[34] McCartney, C. (2008). Reply: LCN DNA: proof beyond reasonable doubt? – a response, *Nature Reviews Genetics* **9**(9), 726.
[35] Gill, P. (2008). Reply: LCN DNA: proof beyond reasonable doubt? – a response, *Nature Reviews Genetics* **9**(9), 726.
[36] Regulator TFS (2008). *Response to Professor Brian Caddy's Review of the Science of Low Template DNA Analysis*, 7 May 2008. (Regulator TFS is The Forensic Science Regulator).

Related Articles

Bayesian Networks
DNA: Degraded Samples
Evidence Interpretation: a Logical Approach
Low Copy Number DNA
Mixture Interpretation: DNA
Short Tandem Repeats: Interpretation

JOHN S. BUCKLETON

Interpretation: Observer Effects

Introduction

The term *observer effect* generally describes circumstances in which the results of an observation are affected by the observer. In physics, the term has been used to describe circumstances in which the act of observation changes the phenomenon being observed, as where measurement of electrical current in a circuit changes the current flow. By contrast, in the social sciences and in forensic science, the term is used to describe circumstances in which the observer's preconceptions or motives influence conclusions drawn from data [1]. The preconceptions and motives are thought to influence the *perception and interpretation* of the evidence rather than changing the evidence itself. Observer effects are sometimes also called *examiner bias* [2, 3] although it important to note that the "bias" entailed in the phenomenon may occur without the observer intending or even being aware of it [4, 5].

Observer effects are closely related to *context effects* and the two terms are sometimes used synonymously [4]. The term *context effect* originated in psychology and has been used to describe circumstances in which the perception of a stimulus is affected by the surrounding context, as where a gray object looks lighter against a dark background than against a light background (see, e.g., http://web.mit.edu/persci/people/adelson/checkershadow_illusion.html). In forensic science, the term *context effect* has been used more broadly to describe situations in which the results of a forensic analysis are affected by the circumstances in which it is performed, and particularly by the information available to the analyst, as when an analyst becomes more likely to identify a latent print as that of a suspect when told that another analyst has already made the identification or when told that other evidence indicates the suspect made the print. The "other evidence" might be said to provide a "context" that changes the analyst's interpretation of the data contained in the prints. Alternatively, the "other information" might be said to have changed the analyst's expectations about the data and hence to have induced an *observer effect*.

The same phenomenon is sometimes described as *confirmation bias*. In psychology and cognitive science, the term *confirmation bias* refers to a human tendency to evaluate evidence in a manner that supports or confirms one's preconceptions, such as a tendency to search out and give more weight to evidence that supports a favored hypothesis than to evidence that contradicts it [5]. In forensic science, the term *confirmation bias* has been used in a manner that is roughly interchangeable with the terms *observer effect* and *context effect* [6, 7].

Underlying Psychological Phenomenon

Underlying all of these terms is a basic phenomenon of human psychology – the tendency of observers to interpret data in a manner consistent with their expectations and desires. The existence of this phenomenon is well established – it has been called *one of the most venerable ideas of traditional epistemology* as well as *one of the better demonstrated findings of twentieth century psychology* [8]. An early discussion of the

phenomenon can be found in the writings of Francis Bacon, who commented in 1620 that

> The human understanding when it has once adopted an opinion draws all things else to support and agree with it. And though there be a greater number and weight of instances to be found on the other side, yet these it either neglects and despises, or else by some distinction sets aside and rejects, in order that by this great and pernicious predetermination the authority of its former conclusions may remain inviolate [9].

Even earlier, Julius Caesar famously noted men's tendency to "believe quite readily that which they wish to be true" [10]. In the twentieth century, psychologists confirmed the existence and strength of these phenomena with innumerable experiments in which people's interpretation of a variety of stimuli, data, and forms of evidence were shown to be influenced by preconceptions and desires. The effects are greatest when the underlying data are somewhat ambiguous and when observers are influenced by strongly held expectations and motives. Extensive reviews of this literature are found in [1, 4, 5, 8, 11, 12].

The most relevant studies relate to the interpretation of scientific data. Like everyone else, scientists have a tendency to interpret data in a manner that supports theories that they favor. The tendency of academic scientists to cling to pet theories, well past the point at which such adherence could be justified by the evidence, has been widely noted [13–15]. In his iconic book *Galileo's Revenge: Junk Science in the Courtroom* [16], Peter Huber traced several false scientific theories to misinterpretation of data arising from uncontrolled observer effects. A scientist who is committed to a pet theory inevitably (and unconsciously) interprets data in a manner consistent with that theory. By Huber's account, uncontrolled observer effects are one of the hallmarks of junk science (*see* **Interpretation: Legal Perspective**).

Minimizing Observer Effects

Allowing preconceptions to influence the interpretation of data is said to be the "cardinal sin for the formal, pure scientist" [8]. Accordingly, when scientists must rely on subjective judgment to interpret the results of an experiment, they routinely take careful steps to mask or shield the person interpreting data from extraneous information that might improperly influence the interpretation. For example, scientists in most fields use "blind" or "double-blind" procedures when relying on subjective judgment to interpret data. Blind procedures are also widely used for peer-review of scientific articles, for grading of written examinations, and for other functions for which it is important to minimize observer effects.

The field of forensic science has been criticized for failing to take adequate steps to minimize observer effects [1, 4]. Through communications with police, lawyers, and other experts, forensic scientists often are exposed to information that may influence their expectations and perhaps even their hopes and desires about what a particular examination might reveal [1]. Moreover, when performing examinations forensic scientists often rely in part on subjective judgment to evaluate potentially ambiguous data [17]. Even seemingly objective procedures like DNA testing sometimes require analysts to use subjective judgment to resolve crucial ambiguities and hence could potentially be influenced by observer effects [3, 18–21]. Yet forensic scientists rarely take steps to shield themselves from extraneous information – i.e., information unnecessary for making a scientific assessment – when making comparisons or interpreting test results [1] (*see also*, **Threat Assessment: School**; **DNA: Degraded Samples**).

The difference between forensic science and other scientific fields may stem in part from uncertainty about what types of information are necessary and relevant to a forensic science assessment. In some forensic disciplines consideration of nonscientific evidence is not only accepted but required by professional norms. For example, fire investigators are trained to consider whether a suspect had a motive for setting a fire when deciding whether to classify a fire as arson [22] (*see also* **Fire and Explosion Investigations: Overview**; **Fire: Scene Investigation**). There have been reports of forensic scientists relying on nonscientific evidence in other disciplines as well, such as a DNA analyst who defended a "match" that implicated the defendant in a rape case by saying: "I know I am right. They found the victim's purse in [the defendant's] apartment" [21]. In 1997 the US Justice Department's Office of Inspector General reported that FBI explosives experts had relied on extraneous evidence when making key scientific determinations in a number of important cases [23]. In the first World Trade Center bombing in 1993, for example, FBI examiners had relied on the

fact that the suspects had access to urea nitrate to reach the conclusion that urea nitrate had been used to make the bomb. The Inspector General condemned this reasoning as circular and biased. The FBI laboratory management agreed and pledged to take steps to ensure that the problem never happened again. Academic commentators have been unanimous in condemning the use of such nonscientific evidence as a basis for scientific conclusions [1, 6, 7, 17, 18]. They argue that the role of the forensic scientist is to offer conclusions derived from a scientific discipline, not to offer conclusions based in part on analysis of other evidence in the case (*see* **Fire and Explosion Investigations: Overview**).

Forensic scientists have become more aware of the importance of addressing observer effects as a result of a high-profile error by the FBI's fingerprint identification unit in the Madrid train bombing case [6, 7, 24]. Three highly trained FBI latent print examiners, and a well regarded independent examiner, all identified a latent print associated with the Madrid bombing as having come from Brandon Mayfield, a resident of Portland, Oregon. The error came to light when Spanish authorities determined the latent print actually matched an Algerian suspected terrorist. The FBI acknowledged the error and apologized to Mr Mayfield. In a subsequent analysis, the agency attributed the error in part to *confirmation bias*. After an initial erroneous identification, each subsequent examiner was fully aware that the latent print had previously been matched to Mayfield. This knowledge may have led them to place too much weight on similarities between the two prints while ignoring or discounting discrepancies.

Empirical Studies of Observer Effects in Forensic Science

Empirical studies have confirmed that observer effects can influence the interpretation of latent prints, leading to errors [25–28]. For example, in one clever study, psychologists asked five highly experienced latent print examiners to compare pairs of prints [26]. The examiners were not told that they had previously individualized these pairs (i.e., found them to match) during casework. Each examiner was instead told (incorrectly) that the prints were those that the FBI had erroneously matched in the Madrid train bombing case. Although they were instructed to "ignore all the contextual information and focus solely on the actual prints", three of the five examiners changed their previous judgment of "match" to "no match" and a fourth changed from "match" to "cannot tell". Only one of the five examiners consistently maintained that the prints were a "match". In other words, after exposure to "extraneous information" suggesting that the prints should *not* match, four of five examiners reached a different conclusion than they had reached previously when comparing the same prints.

Studies in other forensic science domains have also found evidence of observer effects on examiners' judgments [29]. Research on the precise mechanisms by which observer effects influence interpretation are just beginning to emerge [30]. Additional research in this important area is clearly needed.

Case Managers and Sequential Unmasking

There are a number of procedures for minimizing observer effects [1, 4, 31]. One method is to separate various laboratory functions and assign them to different people. A *case manager* who is fully informed of the facts of the case decides what to test and how to test it; a *laboratory analyst* who is "blind" to extraneous information analyzes and interprets the test results. To the extent possible, the analyst remains "blind" to extraneous information, such as information about the legal consequences of the judgment and nonscientific evidence in the case, when deciding whether samples "match". Once the analyst interprets and records the test results, the *case manager* (who is aware of the broader facts of the case) is then responsible for placing the test results in context and assessing the compatibility of forensic observations with various theories of what occurred.

Another approach is simply to perform the analysis and evaluation of various samples in a sequence so that analysis and interpretation of early samples cannot be influenced by knowledge of the later samples. Evidentiary samples, which are generally more difficult evaluate, are analyzed first, before the analyst knows the features of the reference samples. For example, fingerprint examiners can decide whether a latent print is interpretable and which features of the latent print constitute reliable data before knowing whether those features are consistent or inconsistent with a reference print. Similarly, DNA analysts

can decide which "alleles" in an evidentiary sample profile are real and which are spurious before knowing whether those alleles match up with a suspect's. Information required to draw an ultimate conclusion is "unmasked" when needed but the analyst performs as much work as possible while "blind" to unnecessary facts [31].

References

[1] Risinger, D.M., Saks, M.J., Thompson, W.C. & Rosenthal, R. (2002). The Daubert/Kumho implications of observer effects in forensic science: hidden problems of expectation and suggestion, *California Law Review* **90**(1), 1–56.

[2] Beckham, J.C., Annis, L.V. & Gustafson, D.J. (1989). Decision making and examiner bias in forensic expert recommendations for not guilty by reason of insanity, *Law and Human Behavior* **13**(1), 79–87.

[3] Thompson, W.C. (1998). Examiner bias in forensic RFLP analysis, *Scientific Testimony: An Online Journal*, http://www.scientific.org/case-in-point/articles/thompson/thompson.html.

[4] Saks, M.J., Risinger, D.M., Rosenthal, R. & Thompson, W.C. (2003). Context effects in forensic science, *Science and Justice* **43**(2), 77–90.

[5] Plous, S. (1993). *The Psychology of Judgment and Decision Making*, McGraw-Hill, New York.

[6] Stacey, R.B. (2004). Report on the erroneous fingerprint individualization in the Madrid train bombing case, *Journal of Forensic Identification* **54**(6), 706–718.

[7] Office of the Inspector General U.S. Department of Justice (2006). *A Review of the FBI's Handling of the Brandon Mayfield Case*, Office of the Inspector General U.S. Department of Justice, Washington, DC, pp. 1–330.

[8] Nisbett, R. & Ross, L. (1980). *Human Inference*, Prentice Hall, Englewood Cliffs, p. 67.

[9] Bacon, F. (1620). *Novum Organum, book I*, 109, point 46, reprinted, in Hutchins, R.M. (ed) (1952) *Great Books of the Western World*, Vol. 30, Britannica Publishing, New York.

[10] Caesar, G.J. (1873). *Commentaries on the Gallic war 155 (51 BCE)* John Wiley & Sons.

[11] Schneider, D.J., Hastorf, A.H. & Ellsworth, P.C. (1979). *Person Perception*, 2nd Edition, Addison-Wesley Publishing, Reading.

[12] Gilovich, T. (1991). *How We Know What isn't So: The Fallibility of Human Reason in Everyday Life*, The Free Press, New York.

[13] Barber, B. (1952). *Science and the Social Order*, Collier, New York.

[14] Kuhn, T.S. (1962). *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago.

[15] Mahoney, M.J. (1976). *Scientist as Subject: The Psychological Imperative*, Ballinger, Cambridge.

[16] Huber, P.W. (1991). *Galileo's Revenge: Junk Science in the Courtroom*, Basic Books, New York.

[17] Thompson, W.C., Cole, S.A. (2007). Psychological aspects of forensic identification evidence, in *Expert Psychological Testimony for the Courts*, M. Costanzo, D. Krauss & K. Pezdek, eds, Lawrence Erlbaum and Associates, New York.

[18] Thompson, W.C. (1995). Subjective interpretation, laboratory error and the value of forensic DNA evidence: three case studies, *Genetica* **96**, 153–168.

[19] Thompson, W.C. (1997). A sociological perspective on the science of forensic DNA testing, *UC Davis Law Review* **30**(4), 1113–1136.

[20] Thompson, W.C. (1997). Accepting lower standards: The National Research Council's second report on forensic DNA evidence, *Jurimetrics Journal* **37**(4), 405–424.

[21] Thompson, W.C., Ford, S., Doom, T., Raymer, M. & Krane, D. (2003). Evaluating forensic DNA evidence: essential elements of a competent defense review: part 1, *The Champion* **27**(3), 16–25.

[22] Lentini, J. (2006). *Scientific Protocols for Fire Investigation*, CRC Press, Boca Raton.

[23] Office of the Inspector General, U.S. Department of Justice (1997). *The FBI Laboratory: An Investigation Into Laboratory Practices and Alleged Misconduct In Explosives-Related and Other Cases*, Office of the Inspector General, U.S. Department of Justice, Washington, DC.

[24] Thompson, W.C. & Cole, S.A. (2005). Lessons from the Brandon Mayfield case, *The Champion* **29**, 32–34.

[25] Dror, I.E., Peron, A., Hind, S.L. & Charlton, D. (2005). When emotions get the better of us: the effect of contextual top-down processing on matching fingerprints, *Applied Cognitive Psychology* **19**(6), 799–809.

[26] Dror, I.E., Charlton, D. & Peron, A. (2006). Contextual information renders experts vulnerable to making erroneous identifications, *Forensic Science International* **156**, 74–78.

[27] Dror, I.E. & Charlton, D. (2006). Why experts make errors, *Journal of Forensic Identification* **56**(4), 600–616.

[28] Dror, I.E. & Rosenthal, R. (2008). Meta-analytically quantifying the reliability and biasability of forensic experts, *Journal of Forensic Sciences* **53**(4), 900–903.

[29] Miller, L.S. (1987). Procedural bias in forensic examination of hair, *Law and Human Behavior* **11**(2), 157–163.

[30] Schiffer, B. & Champod, C. (2007). The potential (negative) influence of observational biases at the analysis stage of fingermark individualization, *Forensic Science International* **167**, 116–120.

[31] Krane, D.E., Ford, S., Gilder, J.R., Inman, K., Jamieson, A., Koppl, R., Kornfield, I.L., Risinger, D.M., Rudin, N., Taylor, M.S. & Thompson, W.C. (2008). Sequential unmasking: a means of minimizing observer effects in forensic DNA interpretation (letter), *Journal of Forensic Sciences* **53**(4), 1006–1007.

Related Articles

Case Assessment and Interpretation
Identification and Individualization
Interpreting Expert Opinions: History of

WILLIAM C. THOMPSON

Interpretation: Paint Evidence ***see*** Paint: Interpretation

Interpretation: Principles ***see*** Evidence Interpretation: a Logical Approach

Interpretation: Sampling ***see*** Sampling and Estimation of Quantities

Interpretation: Signatures and Handwriting ***see*** Handwriting and Signatures, Interpretation of Comparison Results

Interpretation: STR Profiles ***see*** Short Tandem Repeats: Interpretation

Interpretation: Toxicology ***see*** Postmortem Toxicology: Interpretation

Interpreting Expert Opinions: History of

Introduction

The underlying principle of interpretation is the achievement of the best assessment of the uncertainty associated with the comparison of evidence found at the scene of the crime and evidence found in association with a suspect.

Interpretation in forensic science has a long history, dating back to comments made by classical scholars. See Aitken and Taroni [1], Robertson and Vignaux [2] and Sheynin [3] for appropriate references.

Records were scanty until the mid-nineteenth century. In his book on the emergence of probability, Hacking [4] describes how Leibniz in the seventeenth century wrote about the relationship between probability and the law. A probabilistic analysis of the cumulative force of circumstantial evidence was given by Bernouilli [5]. The calculus of probabilities was applied to problems including the presumption of death, the value of annuities, marine insurance, the veracity of testimony, and the probability of innocence by his nephew, Bernouilli [6].

From the mid-nineteenth century till the 1940s, interpretation in forensic science was concerned mainly with the determination of a relative frequency of a characteristic of interest. In the war years of the 1940s, the logarithm of the likelihood ratio, known as the *weight of evidence*, was used in Bletchley Park by code breakers. The role of the likelihood ratio in interpretation in forensic science did not emerge until the publication of a seminal paper by Lindley [7], though the philosophy had been expressed in a report for the Dreyfus case by Darboux, Appel, and Poincaré [8].

Early forensic scientists recognized that adequate data and consideration of the case as a whole should be used to reach a decision regarding the value of forensic scientific evidence. Their points of view were generally compatible with a Bayesian framework [9]. Bertillon wrote that experts must be prepared to present evidence in a numerical form and not just as an opinion. He also considered the presentation of results without other information pertaining to the inquiry as a methodological error, an early consideration of the differing roles of the likelihood ratio and the prior odds in the assessment of evidence.

In the 1960s in a series of papers, Kingston and Kirk [10, 11] and Kingston [12–14], discussed the role of statistics in forensic science (or criminalistics). These were predated by Kirk [15] in which Kirk stated that "(p)robability is the very keynote of interpretation of all physical evidence". A distinction has been made between *individualization* and *identification*. The process of addressing the issue of whether or not a particular item came from a particular source is termed "individualization". "Criminalistics is the science of individualization" [16]. An "identification" is, more correctly, defined as "the determination of some set to which an object belongs or the determination as to whether an object belongs to a given set" [12]. However, established forensic and judicial practices have led to "individualization" being termed "identification". The establishment of an identification of a suspect or defendant with a criminal is an opinion on the issue of interest to the court and outside the remit of the forensic scientist. The step from probabilities relating to evidence to probability relating to identity has been termed a "leap of faith" [17].

The problems that may arise when numbers are introduced into a court of law are discussed at length in Tribe [18], in response to an article by Finkelstein and Fairley [19] with a response by these authors to the article by Tribe in Finkelstein and Fairley [20]. For example, in an issue which is still a matter of debate, Tribe raises the concern that a very low relative frequency for a characteristic may "dwarf all efforts to put it into perspective". This concern "pervades all cases in which the trial use of mathematics is proposed".

Current research and development on interpretation concentrates on ideas known as *Bayesian*, named after the nonconformist minister Reverend Thomas Bayes, FRS (1702–1761). The paper in which the principles of the reasoning that has come to be known as *Bayesian inference* was espoused was published posthumously in 1763 [21–23].

The history of interpretation may be considered, roughly in chronological order, in the four stages of relative frequencies, discriminating power, significance probabilities, and likelihood ratios. A brief description of errors in interpretation follows. The article finishes with a short comment on the role of Bayesian networks in interpretation in forensic science.

Relative Frequencies

The relative frequency of a characteristic has been used as a measure of the strength of evidence for many years, since at least the mid-nineteenth century. The relative frequency in a sample has been used as an estimate of the probability of the characteristic being present in a member of the population of which the sample is representative.

An early example of the use of a relative frequency in law was to query the authenticity of signatures on wills [24, 25]. The use of probabilistic and statistical evidence in the Howland will case is probably the earliest instance of their use in American law. The evidence related to the agreement of 30 downstrokes in a contested signature with those of a genuine signature. It was argued that the probability of this agreement if the contested signature were genuine was extremely small; the probability of observing two spontaneous signatures with the number of overlaid strokes observed in those two signatures was estimated as $(1/5)^{30}$. Hence, the contested signature was a forgery, an example of a transposed conditional and a leap of faith.

Later examples in which the relative frequency of a characteristic in some background or relevant population was used as evidence include the Dreyfus case, the *Collins* case [26] and, in the present day, cases involving DNA profiling. In all the examples, the smaller the frequency, the stronger is the support provided by the evidence for the prosecution proposition.

Propositions may be at one of three levels known as *source, activity, and crime* [27]. Examples of prosecution propositions at each level include the following:

Source: the defendant's DNA was at the scene of the crime;

Activity: the defendant assaulted the victim;
Crime: this is close to the activity level but includes nonscientific evidence as to whether a crime occurred.

The decision-making processes in forensic science that were used in the early 1960s were contrasted with a statistical approach to evidence evaluation by Kingston and Kirk [10] who advocated the use of statistics in forensic science. They commented that the emphasis for a high quality of interpretation of evidence should be upon a formal system rather than on the subjective opinion of an individual (such as an expert witness); a comment that still has considerable relevance.

It was recognized in the 1960s that there are four aspects to the accurate collection and interpretation of data: planning, randomness, independence, and uncertainty [10]. Some updated comments are appropriate.

- *Planning:* Data have to be collected properly to ensure that any analysis based on them may be meaningful. Meaningful results do not arise from the application of statistical analyses to data that have been collected poorly or in an inappropriate manner.
- *Randomness:* It is necessary to make a careful study of the type of evidence that is likely to appear so that the population selected from which a sample is taken and from which background information is derived is one which is equivalent to a random sample. The experimenter is under reasonable control of the choice of the sample but is not under control of the evidence. A true random sample is rarely achievable in the context of forensic science and, in such a situation, inferences have to be made from what is known as a *convenience sample*. This approach is defended by Evett and Weir [28]: "(i)n the last analysis, the scientist must also convince a court of the reasonableness of his or her inference within the circumstances as they are presented as evidence".
- *Independence:* Independence is not necessary before a statistical analysis may be used. In practice, dependence should be assumed and any assumption of independence justified.
- *Uncertainty:* The existence of uncertainty means a statistical analysis must be used. The existence of uncertainty places a limitation on the appropriateness of any decisions made and validates the contention that a statistical approach offers the best solution to the problem.

Discriminating Power

Discriminating power was developed as an answer to the problem of the determination of the quality of a method for distinguishing between two samples from different sources. If an analysis of characteristics of a sample from a crime scene and of a sample recovered from a suspect failed to distinguish the two samples as originating from different sources, it was necessary to know how strong this outcome was as evidence that the samples came from the same source. These problems were of considerable interest to forensic scientists in the late 1960s and in the 1970s; see, for example, theoretical work in Parker [29, 30], Jones [31] and Smalldon and Moffat [32]. Early experimental attempts to answer these questions have been described in Tippett *et al.* [33] for fragments of paint, and in Gaudette and Keeping [34] for human head hairs.

Two individuals are selected at random from some population. The probability that they are found to match with respect to some characteristic (*e.g.*, blood profile or paint fragments on clothing or head hairs) is known as the *probability of nondiscrimination* or the *probability of a match*, PM. The complementary probability, that is, the probability that they are found not to match with respect to this characteristic is known as the *probability of discrimination* [31] or *discriminating power*, DP [32] such that $DP = 1 - PM$. Gaudette and Keeping [34] provided an estimate of the probability that hairs "selected at random" from two individuals are indistinguishable. This probability can be thought of as an average probability. Its value can be used as a broad guideline to indicate the effectiveness of the evidential type in general, but its use as the value of the evidence in a particular case could be very misleading.

Significance Probabilities

The role of significance probabilities in evidence evaluation is best discussed by means of an example. The example chosen is that of a broken window and the characteristic of interest is the refractive index. A fragment F of glass is found on the clothing of a suspect. Let θ_0 be the value of the parameter,

representing the mean refractive index of the broken window. This is the assumed constant.

Let X be the refractive index for F. It is assumed that if F came from a window of refractive index θ, then X is such that

$$X \sim N(\theta, \sigma^2) \tag{1}$$

there being variation in the refractive index within a particular window. The question is whether F came from the window at the crime scene. If this is so, then θ will be equal to θ_0. The converse may not be true.

An argument based on significance probabilities is as follows. Suppose $\theta = \theta_0$. The deviation of an observation x from the mean θ of the distribution is measured in terms of the probability of observing a value for the random variable X as extreme as x, in relation to θ_0.

The supposition that $\theta = \theta_0$ may be referred to as the *null hypothesis*, conventionally denoted by H_0. If H_0 is true,

$$X \sim N(\theta_0, \sigma^2) \tag{2}$$

The probability $\Pr(|\ X\ | > x)$ is the probability of what has been observed (x) or anything more extreme if H_0 ($\theta = \theta_0$) is true (and hence, as discussed above, that F came from the window at the crime scene). The phrase "anything more extreme" is taken to mean anything more extreme in relationship to an implicit alternative hypothesis that, if H_0 is not true, then $\theta \neq \theta_0$.

The probability of what is observed or anything more extreme, calculated assuming the null hypothesis is true, is known as the *significance probability*. It may be thought to provide a measure of the compatibility of the data with the null hypothesis. It is conventionally denoted by P. A small value of P casts a doubt on the null hypothesis. Certain values of P have been used to provide values known as *significance levels* at which a scientist decides to act as if the null hypothesis is false. Typical values are 0.10, 0.05, and 0.01. Thus, for example, a value x of the refractive index for which $P < 0.05$ would be said to be *significant at the 5% level*. The expression for P may be written more explicitly in the notation of conditional probability as

$$P = \Pr(|\ X\ | > x \mid \theta = \theta_0, \sigma) = 0.01 \tag{3}$$

A small value for P would seem indicative of the falsity of the null hypothesis, but it is not the probability that the null hypothesis is true. It is incorrect to use the value for P as the probability that the suspect was at the crime scene. The transposed probability

$$\Pr(\theta = \theta_0 \mid \ \ |\ X\ | > x, \sigma) \tag{4}$$

would be much more useful, but this is not what has been calculated. The relationship between equation (3) and equation (4) is similar to that between the probability of the evidence given the suspect is guilty and the probability that the suspect is guilty given the evidence. The interpretation of the first of the last two probabilities as the second is the fallacy of the transposed conditional.

Significance probabilities also *combine* in a different way from the probabilities of events. For characteristics which are dependent, it is possible that the significance probability of the joint observation may be greater than either of the individual significance probabilities. For example, it is possible that the individual significance probabilities are both <0.05, but the combined significance probability >0.05.

The significance probability is the output of the first part of a two-stage process [35] which models some processes used by forensic scientists from then until the present day. If the outcome of the significance test is significant, then the piece of evidence which is being compared with the crime scene is deemed not to have come from the crime scene. If the outcome is not significant, then the piece of evidence which is being compared with the crime scene is deemed to have perhaps come from the crime scene. This leads to an effect known as the *fall-off-the-cliff* effect (credited in [36], to Smalldon). A comparison which is significant at the 4.9% level leads to an exclusion of the evidence and one which is significant at the 5.1% level leads to the continuation of the evidence in the investigative process. The second stage is one of assessment of the rarity of the similarity. If the similarity is of a rare characteristic, then the evidence is stronger than if it is of a common characteristic.

Likelihood Ratio

The likelihood ratio is the ratio $\Pr(E \mid H_p)/\Pr(E \mid H_d)$, which converts prior odds $\Pr(H_p)/\Pr(H_d)$ in favor of the prosecution proposition (H_p) relative to the defense proposition (H_d) into posterior odds

in favor of the prosecution proposition, given evidence E, $\Pr(H_p \mid E)/\Pr(H_d \mid E)$. A brief historical review is given here. A detailed description of the role of the likelihood ratio is given in a separate entry.

An interesting and somewhat prophetic comment was made by Parker and Holford [35]. The two-stage approach follows, they said, the traditions of forensic scientists. Interestingly, they then went on to make the following remark.

> We could (therefore) set up an index R whose numerator is the likelihood that the crime hair comes from the suspect and whose denominator is the likelihood that it comes from the population at large. In ordinary language one would then assert that it was R times more likely for the hair to have come from the suspect than from someone drawn at random from the population. But a statement of this nature is of rather limited use to forensic scientists, by-passing as it does the similarity question altogether [35].

Note that these are posterior probabilities about the source of the evidence and not about the relative likelihood of the evidence under prosecution and defense propositions.

The logarithm of the likelihood ratio is known as the *weight of evidence* [37] as it is possible to write the odds form of Bayes Theorem in a log-odds form, which is additive and the logarithm of the likelihood ratio may be thought of as weights in the two pans of the scales of justice.

The use of the odds form of Bayes Theorem to evaluate evidence is extended to include the feature of missing evidence by Lindley and Eggleston [38]. The idea of considering a restricted population, as on an island, for theoretical developments of evidential evaluation, a context known as the *island problem*, was introduced by Eggleston [39], Following the publication of various papers by [7, 40, 41], the 1980s saw a flourishing of the development of related ideas, led by Evett. See, for example, Evett [42] which discussed the question "What is the probability that this blood came from that person?", Evett [43] which provided a "quantitative theory for interpreting transfer evidence in criminal cases", Evett [44] which considered the two-trace problem, and, Evett [45] which considered the evaluation of DNA profiles in the case where the defense is "It was my brother". Review papers by Evett include Evett [46] and Evett [47].

Relationship of Significance Probabilities to Likelihood Ratio

As an illustration of the relationship of significance probabilities and the likelihood ratio, consider a characteristic X of interest. The prosecution and defense propositions are H_p and H_d. The mean of X at the crime scene is θ_0 and the distribution of X at the crime scene is $N(\theta_0, \sigma^2)$. For X from an unknown source, of unknown θ, assume that the unknown θ has a normal distribution with mean θ_0 and variance τ^2, where, typically, $\tau^2 \gg \sigma^2$. Then,

$$f(x \mid H_d) = \int f(x \mid \theta)\ f(\theta)\ \mathrm{d}\theta \tag{5}$$

and so

$$(X \mid H_d) \sim N(\theta_0, \sigma^2 + \tau^2) \tag{6}$$

With $\tau^2 \gg \sigma^2$, the distribution of $(X \mid H_d)$ is approximately $N(\theta_0, \tau^2)$. Thus, the likelihood ratio is

$$\begin{aligned} V &= \frac{f(x \mid \theta_0, H_p)}{f(x \mid H_d)} \\ &= \frac{(2\pi\sigma^2)^{-1/2}\ \exp\left\{-(x-\theta_0)^2/2\sigma^2\right\}}{(2\pi\tau^2)^{-1/2}\ \exp\left\{-(x-\theta_0)^2/2\tau^2\right\}} \end{aligned} \tag{7}$$

Consider $\tau = 100\ \sigma$. Let $z^2 = (x-\theta_0)^2/\sigma^2$ be the square of the standardized distance between the observation x and the mean specified by the null hypothesis, θ_0. Then,

$$\begin{aligned} V &= 100\ \exp\left(-\frac{z^2}{2} + \frac{z^2}{2\times 10^4}\right) \\ &\simeq 100\ \exp(-z^2/2) \end{aligned} \tag{8}$$

For example, for measurements of refractive indexes, consider $x = 1.518540$, $\theta_0 = 1.518458$, $\sigma = 4\times 10^{-5}$, $\tau = 4\times 10^{-3}$. Then $z^2 = 2.05^2$ and $P = 0.04$, which, at the 5% level, would lead to rejection of the hypothesis that the fragment of glass came from the window at the scene of the crime. However,

$$V = 100\exp\left(\frac{-2.05^2}{2}\right) = 12.2 \tag{9}$$

a value for V that is supportive of H_p against H_d. Such an apparent contradiction between the two approaches is not a new idea and has been named Lindley's paradox (see, for example, [40, 48]).

Errors

As the procedures for the interpretation of scientific evidence developed, many examples of erroneous interpretations of scientific evidence in the law arose. Probably the most famous is that of the prosecutor's fallacy [49], also known as the *fallacy of the transposed conditional*. The fallacy is to equate the probability of finding evidence on an innocent person with the probability that the person on whom that evidence is found is innocent. In practice, it is not possible to derive the second probability from the first without knowing the probability of the person being innocent before the evidence was presented and without knowing the probability of finding the evidence on a guilty person.

A related fallacy is the defense fallacy [49]. Given a probability, p, of finding the evidence on an innocent person, the defender's fallacy then multiplies this probability, p, by the size of a relevant population, say N, to obtain the expected number of people Np with the evidence in that population. It is then argued that a large value of Np renders the evidence meaningless as the accused is only one amongst Np people. However, in rebuttal, before the evidence was presented, the accused was one of N people. The evidence has reduced the pool of potential criminals from N to Np and is therefore of some worth.

Another statistic, while not an error but difficult to interpret, is that which gives the probability of the existence of more than one person with the evidence under consideration, conditional on the existence of one person (the accused) with the evidence. As an example of its use, the court in *Collins* apparently concluded that, if a random item sampled from a large population belongs to a rare type, then the probability of at least one more item of that type in the population is about 0.41. While not an error, this is not a very helpful statistic in that it is not possible to combine this result with other evidence to assess the value of the combined evidence.

Another error is the numerical conversion error. Given a characteristic with probability p, the use of the numerical conversion error is to argue that one has to test $1/p$ people before observing another person with the characteristic. For example, if $p = 0.01$, $1/p = 100$. However, the probability of observing another person with the characteristic in 100 people is $1 - (1 - p)^{100}$, which equals 0.634. Further examples are given in Aitken and Taroni [1].

Bayesian Networks

The use of deterministic networks for legal analysis had been proposed in the 1930s [50]. It was not until the late 1980s that a statistical theory was developed and applied to forensic scientific examples.

Propositions and evidence are represented by nodes in a graph. Edges in the graph correspond to direct associations between the nodes joined by the edge. Nodes that are separated by other nodes are said to be independent conditionally on the intervening nodes. Conditional probability tables and probability density functions are needed to complete the construction of a network. Once evidence is known, the status of the corresponding node is switched from uncertain, with associated probabilities, to certain, where one and only one outcome has a probability of 1, while the others have a probability of zero. This process is known as *instantiation*. The probabilities associated with other nodes are adjusted accordingly. Initial work may be found in Aitken and Gammerman [51] and in Dawid and Evett [52]. A review of work up to 2005 is given by Taroni *et al.* [53].

References

[1] Aitken, C.G.G. & Taroni, F. (2004). *Statistics and the Evaluation of Evidence for Forensic Scientists*, John Wiley & Sons, Chichester.

[2] Robertson, B. & Vignaux, G.A. (1995). *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*, John Wiley & Sons, Chichester.

[3] Sheynin, O.B. (1974). On the prehistory of the theory of probability, *Archive for History of Exact Sciences* **12**, 97–141.

[4] Hacking, I. (1975). *The Emergence of Probability*, Cambridge University Press, Cambridge.

[5] Bernouilli, J. (1713). *Ars Conjectandi*, Basle, Switzerland.

[6] Bernouilli, N. (1709). *Specimina Artis Conjectandi ad Quaestiones Juris Applicatae*, Basle, Switzerland.

[7] Lindley, D.V. (1977a). A problem in forensic science, *Biometrika* **64**, 207–213.

[8] Darboux, J.G., Appell, P.E. & Poincaré, J.H. (1908). Examen critique des divers systèmes ou études graphologiques auxquels a donné lieu le bordereau, in *L'affaire Drefus – La révision du Procès de Rennes – Enqueîe de la Chambre Criminelle de la Cour de*

Cassation, igue française des droits de l'homme et du citoyen, Paris, pp. 499–600.

[9] Taroni, F., Champod, C. & Margot, P. (1998). Forerunners of Bayesianism in early forensic science, *Jurimetrics Journal* **38**, 183–200.

[10] Kingston, C.R. & Kirk, P.L. (1964a). The use of statistics in criminalistics, *The Journal of Criminal Law, Criminology, and Police Science* **55**, 514–521.

[11] Kingston, C.R. & Kirk, P.L. (1964b). Evidence evaluation and problems in general criminalistics, *Journal of Forensic Sciences* **9**, 434–444.

[12] Kingston, C.R. (1965a). Applications of probability theory in criminalistics, *Journal of the American Statistical Association* **60**, 70–80.

[13] Kingston, C.R. (1965b). Applications of probability theory in criminalistics – II, *Journal of the American Statistical Association* **60**, 1028–1034.

[14] Kingston, C.R. (1966). Probability and legal proceedings, *The Journal of Criminal Law, Criminology, and Police Science* **57**, 93–98.

[15] Kirk, P.L. (1953). *Crime Investigation*, Interscience Publishers, New York.

[16] Kirk, P.L. (1963). The ontogeny of criminalistics, *The Journal of Criminal Law, Criminology, and Police Science* **54**, 235–238.

[17] Stoney, D.A. (1991). What made us think we could individualise using statistics? *Journal of the Forensic Science Society* **31**, 197–199.

[18] Tribe, L. (1971). Trial by mathematics: precision and ritual in the legal process, *Harvard Law Review* **84**, 1329–1393.

[19] Finkelstein, M.O. & Fairley, W.B. (1970). A Bayesian approach to identification evidence, *Harvard Law Review* **83**, 489–517.

[20] Finkelstein, M.O. & Fairley, W.B. (1971). A comment on 'Trial by mathematics', *Harvard Law Review* **84**, 1801–1809.

[21] Bayes, T. (1764). An essay towards solving a problem in the doctrine of chances, *Philosophical Transactions of the Royal Society of London for 1763* **53**, 370–418, Reprinted with Barnard (1958) in Pearson and Kendall (1970), 131–153.

[22] Barnard, G.A. (1958). Thomas Bayes – a biographical note (together with a reprinting of Bayes, 1764), *Biometrika* **45**, 293–315, Reprinted in Pearson and Kendall (1970), 131–153.

[23] Pearson, E.S., Kendall, M.G. (eds) (1970). *Studies in the History of Statistics and Probability*, Charles Griffin, London.

[24] Mode, E.B. (1963). Probability and criminalistics, *Journal of the American Statistical Association* **58**, 628–640.

[25] Meier, P. & Zabell, S. (1980). Benjamin Peirce and the Howland will, *Journal of the American Statistical Association* **75**, 497–506.

[26] Fairley, W.B. & Mosteller, W. (1977). *Statistics and Public Policy*, Addison-Wesley, London, pp. 355–379.

[27] Cook, R., Evett, I.W., Jackson, G., Jones, P.J. & Lambert, J.A. (1998). A hierarchy of propositions: deciding which level to address in casework, *Science and Justice* **38**, 231–239.

[28] Evett, I.W. & Weir, B.S. (1998). *Interpreting DNA Evidence*, Sinauer Associates, Inc., Sunderland, MA.

[29] Parker, J.B. (1966). A statistical treatment of identification problems, *Journal of the Forensic Science Society* **6**, 33–39.

[30] Parker, J.B. (1967). The mathematical evaluation of numerical evidence, *Journal of the Forensic Science Society* **7**, 134–144.

[31] Jones, D.A. (1972). Blood samples: probability of discrimination, *Journal of the Forensic Science Society* **12**, 355–359.

[32] Smalldon, K.W. & Moffat, A.C. (1973). The calculation of discriminating power for a series of correlated attributes, *Journal of the Forensic Science Society* **13**, 291–295.

[33] Tippett, C.F., Emerson, V.J., Fereday, M.J., Lawton, F. & Lampert, S.M. (1968). The evidential value of the comparison of paint flakes from sources other than vehicles, *Journal of the Forensic Science Society* **8**, 61–65.

[34] Gaudette, B.D. & Keeping, E.S. (1974). An attempt at determining probabilities in human scalp hair comparison, *Journal of Forensic Sciences* **19**, 599–606.

[35] Parker, J.B. & Holford, A. (1968). Optimum test statistics with particular reference to a forensic science problem, *Applied Statistics* **17**, 237–251.

[36] Evett, I.W. (1991). Interpretation: a personal odyssey, in *The use of Statistics in Forensic Science*, Aitken, C.G.G. & Stoney, D.A., eds, Ellis Horwood, Chichester.

[37] Good, I.J. (1985). Weight of evidence: a brief survey, in *Bayesian Statistics, 2*, Bernardo, J.M., De.Groot, M.H., Lindley, D.V. & Smith, A.F.M., eds, Elsevier Science North Holland.

[38] Lindley, D.V. & Eggleston, R. (1983). The problem of missing evidence, *The Law Quarterly Review* **99**, 86–99.

[39] Eggleston, R. (1978, 1983). *Evidence, Proof and Probability*, 1st and 2nd Edition, Weidenfeld & Nicolson, London.

[40] Lindley, D.V. (1957). A statistical paradox, *Biometrika* **44**, 187–192.

[41] Lindley, D.V. (1977b). Probability and the law, *The Statistician* **26**, 203–220.

[42] Evett, I.W. (1983). What is the probability that this blood came from that person? A meaningful question? *Journal of the Forensic Science Society* **23**, 35–39.

[43] Evett, I.W. (1984). A quantitative theory for interpreting transfer evidence in criminal cases, *Applied Statistics* **33**, 25–32.

[44] Evett, I.W. (1987a). On meaningful questions: a two-trace transfer problem, *Journal of the Forensic Science Society* **27**, 375–381.

[45] Evett, I.W. (1992). Evaluating DNA profiles in the case where the defence is 'It was my brother', *Journal of the Forensic Science Society* **32**, 5–14.

[46] Evett, I.W. (1987b). Bayesian inference and forensic science: problems and perspectives, *The Statistician* **36**, 99–105.
[47] Evett, I.W. (1990). The theory of interpreting scientific transfer evidence. *Forensic Science Progress*, Springer-Verlag, Vol. 4, pp. 141–179.
[48] Good, I.J. (1956). Discussion of paper by G. Spencer Brown, in *Information Theory: Third London Symposium 1955*, Cherry, C., ed, Butterworths, London. 13–14.
[49] Thompson, W.C. & Schumann, E.L. (1987). Interpretation of statistical evidence in criminal trials. The prosecutor's fallacy and the defence attorney's fallacy, *Law and Human Behaviour* **11**, 167–187.
[50] Wigmore, J. (1937). *The Science of Proof: As given by Logic, Psychology and General Experience and Illustrated in Judicial Trials*, 3rd Edition, Little, Brown & Company, Boston, MA.
[51] Aitken, C.G.G. & Gammerman, A. (1989). Probabilistic reasoning in evidential assessment, *Journal of the Forensic Science Society* **29**, 303–316.
[52] Dawid, A.P. & Evett, I.W. (1997). Using a graphical model to assist the evaluation of complicated patterns of evidence, *Journal of Forensic Sciences* **42**, 226–231.
[53] Taroni, F., Aitken, C.G.G., Garbolino, P. & Biedermann, A. (2006). *Bayesian Networks and Probabilistic Inference in Forensic Science*, John Wiley & Sons, Chichester.

COLIN G. G. AITKEN

Interrogation

Scholars and police detectives often distinguish between interviews and interrogations. The goal of an interview is to gather information relevant to a criminal investigation, whereas the goal of an interrogation is to persuade a suspect to admit to committing or participating in a crime [1–2]. Once detectives have concluded (correctly or incorrectly) that a suspect is guilty of committing a crime, they deploy several psychologically powerful interrogation techniques to move the suspect from denial to confession.

Historically, physical and psychological torture have been the primary tools of interrogators and, in many countries, torture remains a standard interrogation technique [3]. In the United States, at least through the early 1930s, police used the "third degree" – routinely beating, threatening, and abusing criminal suspects in the pursuit of confessions [4]. As a result of the Wickersham Commission Report in 1931 [5] and a number of Supreme Court cases beginning with *Brown v. Mississippi* [6], physical abuse of suspects began to decline in the 1930s and by the 1960s, appeared to be rare in the United States [7]. As a substitute for physical and psychological abuse, police began to develop increasingly subtle and sophisticated psychological interrogation techniques.

In 1940, W.R. Kidd published *Police Interrogation*, the first police interrogation manual, instructing police on the new science of modern interrogation. In 1942, Fred Inbau published the first edition of his seminal interrogation manual, *Lie Detection and Criminal Interrogation.* With John Reid and others, Inbau revised and expanded this training manual through several revisions. The most recent edition – *Criminal Interrogation and Confessions* by Inbau, Reid, Buckley, and Jayne – was published in 2001. The Chicago-based firm, *Reid and Associates*, transformed these manuals into a training program and, in 1974, began offering seminars to police and security professionals across North America. More than a quarter million police have now been trained in the Reid Method of interrogation. Undoubtedly, many more police officers have either read the Inbau *et al.* manual, or have learned Reid tactics by observing or working with other interrogators. Research confirms that these tactics are frequently used by detectives in the United States [8–10].

If a suspect is in police custody, police are required to issue the well-known *Miranda* warnings prior to interrogation, informing the suspect that he has the right to remain silent, the right to appointed counsel, and that anything he says can be used against him in court [11]. Surprisingly, *Miranda* has proven to be only a minimal safeguard against coercion. The overwhelming majority of suspects (estimates range from 78 to 96%) relinquish their rights and thus appear to consent to interrogation [12].

In practice, police have devised several strategies to elicit waivers to the *Miranda* warnings. These include reading the *Miranda* warnings in a quick, confusing or perfunctory manner; de-emphasizing or minimizing the significance of the *Miranda* warnings; obscuring the adversarial nature of the police-suspect interaction; and telling the suspect that the interrogator can help him only if he first waives his *Miranda* rights [13].

Police interrogators persuade a suspect to confess by manipulating his perception of the nature and gravity of his immediate situation, the limited choices available to him, and the consequences that follow from each of these choices. The interrogator labors to convince the suspect that admitting culpability is the best and most sensible course of action. First, the suspect's resistance is broken by causing him to perceive that he is trapped and that his situation as hopeless [14]. The interrogator leads the suspect to believe that his guilt can be objectively demonstrated to the satisfaction of any reasonable person; that this fact is indisputable and cannot be changed, that there is no way out of his predicament, and that his fate is already determined.

Alone in the interrogation room with one or more police officers, the suspect is repeatedly accused of committing the crime. The interrogator is likely to express unwavering confidence in these assertions of guilt, ignoring or cutting off any of the suspect's denials or objections. This is likely to have the effect of causing the disoriented suspect to believe that he bears the burden of proving his innocence. If the suspect offers an alibi, the interrogator may attack it as inconsistent, implausible, contradicted by all the case evidence, or simply impossible – even if none of these characterizations is true. Perhaps the most powerful tool that police use to convince the suspect that his situation is hopeless is to confront him with objective and irrefutable evidence of his guilt – whether or not such evidence actually exists.

American law permits interrogators to pretend they have evidence when they do not, and police often confront suspects with fabricated evidence such as nonexistent eyewitnesses, false fingerprints, make-believe videotapes, false polygraph results, etc. [15]. Confronted with real or fabricated evidence, many suspects conclude that conviction and imprisonment are inevitable, and decide that there is no choice but to confess.

During the next phase of interrogation, the detective seeks to create the impression that confessing will actually improve the suspect's otherwise hopeless situation. Interrogators emphasize their discretionary power to favorably influence the suspect's case and describe how unfavorably his case is likely to be perceived by key players in the justice system: prosecutors, jurors, or a judge. The interrogator's message to the suspect is that he will be charged with a less serious crime and receive a shorter prison sentence if he accepts responsibility, admits guilt, and expresses remorse [16]. Conversely, continued denial will result in a more serious charge and a longer prison sentence. Detectives suggest that the suspect is dependent on the interrogator's help and can only be released from the pressures of interrogation by confessing [14].

To clear the path for an admission, interrogators suggest honorable or sympathetic reasons why the suspect might have committed the crime. For example, interrogators may suggest that instead of premeditated murder, the suspect was acting in self-defense, or that he accidentally killed the victim. In interrogation manuals, this technique is known as using "themes" [2]; [17]. These "themes" are actually hypothetical scenarios or rationalizations that morally, psychologically, and legally downplay, excuse, or justify the suspect's act. Interrogators use such scenarios to make a suspect comfortable with making an admission, and to persuade him that if he makes an admission, he (with the interrogators' help) can control how his admission is framed for relevant audiences (i.e., prosecutors, judges, juries, his friends and family, the victim's friends and family, and perhaps the media). Scenarios can be a subtle yet powerful inducement. They work by shifting blameworthiness from the suspect being interrogated to another person, or to the social circumstances that caused the act, or by redefining the act itself. Delivered against the backdrop of the techniques that have preceded them – social isolation, accusations of guilt, attacks on denials, evidence ploys, etc. – these scenarios are intended to motivate the suspect to admit the crime because he is being offered a chance to mitigate his punishment by choosing the less incriminating of two bad explanations.

The tactic of reframing the interrogation as an "opportunity" simultaneously misrepresents the role of police as allies instead of adversaries, and suggests that interrogators can and will help the suspect to minimize the consequences of his crime. Interrogators may explicitly tell a suspect that they can help him by how they write up their report following questioning, or by what they tell the prosecutor prior to charges being filed, or by how they testify before a judge or jury at trial. Interrogators may also offer to help a suspect in ways that extend beyond the criminal justice system, such as arranging for counseling or help from social services. Of course, the catch is

that these benefits can only be realized if the suspect complies with the interrogators' demands for an admission. Police often imply that this "opportunity" for the suspect to "tell his side of the story" is time-limited and will vanish once the interrogation ends.

Effective psychological interrogation is an incremental, cumulative process. Through tactics that rely on deception, manipulation, and coercion, the interrogator gradually restructures the psychological context within which the suspect makes the decision about whether or not to confess. A suspect may enter the interrogation room confident that he will never agree to make an admission. However, once the suspect has been convinced that he will almost certainly be arrested, convicted, and punished, his assessment of his immediate situation and his decision-making calculus are likely to change such that confessing may be seen as the only available means of improving his prospects [15].

The modern interrogation process is designed to break the resistance of a rational person who knows he is guilty and to persuade him that the benefits of confessing outweigh the costs [14]. The presumption is that the suspect under interrogation is guilty and will, at least initially, steadfastly deny his guilt. This presumption of guilt should be based on solid evidence. However, it is sometimes based on little more than a detective's intuition. Acting as a sort of human lie detector, interrogators may mistakenly evaluate the suspect's verbal and nonverbal behaviors as indicative of guilt and then subject an innocent person to coercive interrogation techniques [18]. Indeed, Inbau *et al.* caution that interrogation techniques should only be used when a detective is certain of the suspect's guilt, not as a means of establishing guilt without corroborating evidence. However, once a detective has misclassified an innocent person as guilty, he is likely to resort to coercive interrogation techniques [19]. Further, because no credible evidence is likely to exist against an innocent suspect, getting a confession becomes even more important than usual – police detectives desperately need a confession to successfully build a case, especially in high profile and capital prosecutions [20]. Unfortunately, most interrogation training manuals give little thought to how the methods they advocate sometimes lead innocent suspects to confess. Instead, police trainers and interrogators, in the face of abundant empirical evidence to the contrary [21], assert that their methods will only produce confessions from the guilty. The central research issue is no longer whether police-induced false confessions exist, but why they occur and what can be done to prevent them.

Several "best practice" reforms have been proposed to increase the number of true confessions while decreasing the number of false ones. Videotaping of interrogations may be the single most important policy reform available. Videotaping creates an objective, comprehensive, reviewable record of what occurred in the interrogation room. When a videotape is available, triers of fact do not need to guess who is telling the truth about what happened during a disputed interrogation. Videotaping opens police practices up to the possibility of external scrutiny, and because interrogators know they may be scrutinized, they are less likely to use the psychologically coercive and improper interrogation techniques that are the primary cause of false confessions in America [15]. Lawyers, judges, jurors, and experts should be able to evaluate not only what was said and how it was said, but also to see the rich array of nonverbal cues (e.g., facial expressions, gestures, posture, personal space, and gaze). At the time of this writing, seven states and District of Columbia have laws requiring police to record interrogations in their entirety in some or all criminal cases, and many police departments have begun to voluntarily record interrogations [22].

Ideally, the entire interrogation should be videotaped; both the suspect and the interrogator should be visible, and the video should carry a time/date stamp. Recording the entire interrogation ensures that we do not rely on partial, selective, or biased records of what transpired. Too often, police videotape only the final, damning confession. To evaluate whether a confession is coerced or contaminated, we need to be able to examine the interaction that culminated in a confession. The requirement that both the suspect and the interrogator be visible follows directly from research demonstrating that only such an inclusive camera focus minimizes biased interpretations about the coerciveness of the interrogation [23]. Finally, including a time/date stamp on the recording ensures that there is an accurate accounting of not only the length of the interrogation but also the length of breaks or gaps in the interrogation process.

A second important reform is the use of social science expert testimony in cases involving a disputed

interrogation or confession. There is now a substantial and well-accepted body of scientific research on this topic, and case law supports the admissibility of such expert testimony [24]. If a factually disputed confession is introduced at trial, the jury will want to know how an innocent person might possibly have been made to confess falsely, especially if to a heinous crime. Most people appear to believe that an innocent person will not falsely confess to police unless he is physically tortured or mentally ill [25–26]. The purpose of expert testimony at trial is to provide an overview of research on interrogation and confessions to assist the jury in making a fully informed decision about what weight to place on the defendant's confession. By educating the jury about the existence, psychology, causes, and indicia of police-induced false confession, social science expert witness testimony at trial should reduce the number of confession-based wrongful convictions.

Police interrogation is an essential investigative activity, but it requires careful monitoring and regulation if we are to trust the voluntariness and reliability of the admissions produced by this inherently manipulative process. In a democratic society that places a high value on individual rights, fair procedures, and just outcomes, custodial interrogation is likely to remain controversial. We must be careful to give police the tools they need to help convict the guilty, while simultaneously putting safeguards in place to protect the innocent.

References

[1] Costanzo, M. & Leo, R., (2007). Research findings and expert testimony on police interrogations and confessions to crimes in *Expert Psychological Testimony for the Courts*, M. Costanzo, D. Krauss, K. Pezdek, eds, Erlbaum, Mahwah.

[2] Inbau, F.E., Reid, J.E., Buckley, J.P. & Jayne, B.C. (2001). *Criminal Interrogation and Confessions*, 4th Edition, Aspen, Gaithersburg.

[3] Costanzo, M., Gerrity, E. & Lykes, M.B. (2007). Psychologists and the use of torture in interrogations. *Analyses of Social Issues and Public Policy*.

[4] Hopkins, E.J. (1931). *Our Lawless Police: A Study of the Unlawful Enforcement of the Law*, Viking Press, New York.

[5] National Commission on Law Observance and Enforcement (1931). *Report on Lawlessness, In Law-Enforcement*, Vol. 11, U.S. Government Printing Office, Washington, DC.

[6] *Brown v. Mississippi*, 297 U.S. 278 (1936).

[7] Leo, R.A. (1992). From coercion to deception: the changing nature of police interrogation in America, *Crime, Law and Social Change: An International Journal* **12**, 35–59.

[8] Simon, D. (1991). *Homicide: A Year on the Killing Streets*, Houghton Mifflin Company, Boston.

[9] Leo, R.A. (1996). Inside the interrogation room, *Journal of Criminal Law and Criminology* **86**, 266–303.

[10] Feld, B. (2006). Juveniles competence to assess Miranda rights: an empirical study of policy and practice, *Minnesota Law Review* **91**, 26–100.

[11] *Miranda v. Arizona*, 384 U.S. 436 (1966).

[12] Leo, R.A. (2001). Questioning the relevance of *Miranda* in the twenty-first century, *The Michigan Law Review* **99**, 1000–1029.

[13] Leo, R.A. & White, W.S. (1999). Adapting to Miranda: modern interrogators' strategies for dealing with the obstacles posed by Miranda, *Minnesota Law Review* **84**, 397–472.

[14] Ofshe, R.J. & Leo, R.A. (1997). The decision to confess falsely: rational choice and irrational action, *Denver University Law Review* **74**, 979–1122.

[15] Leo, R.A. (2008). *Police Interrogation and American Justice*, Harvard University Press, Cambridge.

[16] Leo, R., Costanzo, M. & Shaked, N. (2007). Psychological and cultural aspects of interrogation and false confessions: using research to inform legal decision-making, in *Psychology in the Courtroom*, J. Lieberman, & D. Krauss, eds, Ashgate, London.

[17] Ofshe, R.J. & Leo, R.A. (1997). The decision to confess falsely: rational choice and irrational action, *Denver University Law Review* **74**, 979–1122.

[18] Drizin, S.A. & Leo, R.A. (2004). The problem of false confessions in the post-DNA world, *North Carolina Law Review* **82**, 891–1007.

[19] Meissner, C.A. & Kassin, S.M. (2004). "You're guilty, so just confess!" Cognitive and confirmational biases in the interrogation room, in *Interrogations, Confessions, and Entrapment*, G.D. Lassiter, ed, Kluwer Academic, New York, pp. 85–106.

[20] Gross, S. (1996). The risks of death: why erroneous convictions are common in capital cases, *Buffalo Law Review* **44**, 469–500.

[21] Kassin, S. & Gudjonsson, G.H. (2004). The psychology of confessions: a review of the literature and issues, *Psychological Science in the Public Interest* **5**, 33–67.

[22] Sullivan, T.P. (2006). *Police Experience with Conducting Custodial Interrogations*, Northwestern University School of Law, Center on Wrongful Convictions.

[23] Lassiter, G.D. & Geers, A.L. (2004). Evaluation of confession evidence: effects of presentation format, in *Interrogations, Confessions, and Entrapment*, G.D. Lassiter, ed, Kluwer Press, New York, pp. 197–214.

[24] Costanzo, M. & Leo, R. (2007). Research findings and expert testimony on police interrogations and confessions to crimes, in *Expert Psychological Testimony for*

the Courts, M. Costanzo, D. Krauss & K. Pezdek, eds, Erlbaum, Mahwah, New Jersey.
[25] White, W.S. (1997). False confessions and the constitution: Safeguards Against untrustworthy confessions, *Harvard Civil Rights-Civil Liberties Law Review* **32**, 105–157.
[26] Ainsworth, P. (1995). *Psychology and Policing in a Changing World*, John Wiley & Sons, New York.

Related Articles

Confessions: Evidentiary Reliability of
Deception: Detection of and Brain Imaging
Interrogative Suggestibility

MARK COSTANZO AND RICHARD A. LEO

Interrogative Suggestibility

Law enforcement uses a variety of psychologically sophisticated techniques to extract confessions from those whom they believe are guilty of an offense. The Reid Technique [1] is the most widely taught procedure on how to extract confessions from suspects using legally sanctioned methods (*see* **Interrogation**). Unfortunately, certain individuals are particularly susceptible to police influence (*see also* **Eyewitness: Suggestibility of**). When these police techniques are used on vulnerable individuals, it becomes more difficult for the suspects to assert their Miranda [2] rights to remain silent, seek the consultation of an attorney before or during the interrogation, or to request one if indigent (*see* **Capacity to Waive Miranda Rights**).

If law enforcement should have known that a suspect was particularly vulnerable to not knowing or appreciating rights, such factors may be relevant to the voluntariness of the *Miranda* waiver or the voluntariness of the confession. Suspects, even if guilty, may be at high risk for being incapable of asserting their rights and resisting police demands. A confession may be suppressed, if based on the *totality of circumstances* [3], the court believes a defendant was unable to make a knowing, intelligent, or a voluntary waiver of his rights or that the confession itself was not voluntary.

Types of False Confessions

Additionally, there are some individuals who do not commit the offense or not to the level to which they are being accused. Some of these suspects are quite susceptible to police influence and to giving false or erroneous accounts of their participation in the offense. A false confession is defined here as one in which an individual admits to a crime or overemphasizes or mischaracterizes his or her involvement in the crime. It is more common for an individual to overstate his or her role in a crime than to confess to a crime they had no involvement in, but both of these do occur. Not only do suspects sometimes falsely confess to crimes that they did not commit, but also they do so quite regularly [4].

It is impossible to estimate the rate and frequency of false confessions because we do not know who is genuinely guilty of the act and who is not. Retractions happen in both true and false confessions, and include so-called confessions, which a defendant says the law enforcement falsely created. We know that innocent defendants are sometimes wrongly convicted. They are later exonerated by DNA evidence or when the truly guilty party is apprehended. Estimates of false confessions vary depending upon the research design and how the term *false confession* is operationalized. States do not keep a record of the number of suspects annually interrogated, how many of the interrogations result in a true confession, false confession, or no confession, and how many of those are charged with a crime and eventually are found guilty. False confession estimates range from a low of fewer than 30–60 false confessions per year [5] to reports as high as 600 [6].

Types of False Confessions

There are numerous models conceptualizing the categories of false confessions. Kassin and Wrightsman [7] describe three types of false confessions. *Voluntary false confessions* are given without pressure from police. An individual willingly goes to the police to falsely confess to a crime because of the morbid need for notoriety, to protect a friend or relative, or a

pathological need to be punished. *Coerced-compliant false confessions* result from the use of psychological, physical, or environmental pressures by police to extract a confession. In these situations, a suspect may falsely confess to achieve some immediate goal, such as to end the stressful interrogation, to go home, or to get less punishment.

Because the suspect knows he did not commit the crime, a retraction of the confession is likely once the suspect is out of the interrogative environment. *Coerced-internalized false confessions* are when the suspect, who has been subjected to intense police interrogative procedures, such as the use of the Reid Technique, comes to believe or internalize that he or she committed the crime for which they have been accused. This happens most frequently with individuals who have some memory lapse for their behavior at the time of the offense, often due to psychosis or drug intoxication. Law enforcement gets the suspect to believe that they have incriminating evidence (i.e., witnesses, fingerprints, polygraph results). The suspect, who does not really remember what transpires, internalizes the police version of the crime. Individuals who give this type of false confession are highly suggestible.

McCann [8] added a fourth type of false confession, the *coerced-reactive false confession*. The individual is pressured or induced to confess by some other person other than law enforcement. The pressure may come from an abusive husband who threatens to harm if the wife does not confess; or from a gang member who threatens the individual's family if he "does not take the rap".

There are other models of how false confessions can be conceptualized. Ofshe and Leo's model [4] emphasizes the levels of police coerciveness and makes a distinction with the temporal stability in a suspect continuing to falsely believe he or she committed a crime.

In reality, false confessions may not fall within discrete categories but rather overlap with each other. None of these models adequately account for false confessions due to the interaction of police with defendants who have psychotic thinking or the impact of the physical environment (i.e., prolonged isolation, and restraints) [9]. Also, some suspects internalize guilt without being coerced to do so.

Interrogative Suggestibility Construct

Gisli Gudjonsson coined the term *interrogative suggestibility* as the extent to which an individual comes to accept messages or information communicated during formal questioning, whereby that individual comes to believe that information given is essentially true [9, 10]. This construct is different from suggestibility related to hypnosis, which primarily deals with motor processes [11]. According to Gudjonsson, there are three prerequisites with interrogative suggestibility. The suspect must be *uncertain* as to the correct answer to a question. The subject must believe that the interviewer's intentions are genuine and constructive and not involving deception. This is referred to as *interpersonal trust.* Finally, there is an *expectation* component. The subject is reluctant to declare what he does not know because he thinks the interviewer believes that the subject knows the answer.

Interrogative suggestibility appears to be the main component of the coerced-internalized false confession, although there is some relationship to other types of false confessions. The suspect is *uncertain* of his participation in an alleged offense. Law enforcement tells the suspect they have proof that he was involved in the offense due to alleged incriminating data, such as failing a polygraph and/or fingerprint matches. When the police tell the suspect they know he committed the crime but are allowing him the chance to tell his side of the story, which is oftentimes suggested to the suspect by law enforcement as an alternative to the more morally reprehensible version already discussed by them, the suspect believes the officer's statement that it is in his own interest to confess. Thus, interpersonal *trust* is established. The suspect is also hesitant to continue to state he does not know whether he committed the offense because there is the *expectation* that he should know. The theory of interrogative suggestibility is also applicable to the formal interview of witnesses and victims [10].

Gudjonsson Suggestibility Scales

The Gudjonsson Suggestibility Scales (GSS) [9, 10] provide the mental health professional with an objective method to help assess interrogative suggestibility. There are two versions of the GSS, and both have the same format, scoring, and administration. The differences between the versions are the narrative

story presented to the subject and the standardized questions about the story. The GSS is presented as a memory test. The subject is asked to read a narrative story containing 40 facts. A subject is only given one version of the test. After the story has been read, the subject is asked to narrate everything he could remember about the story". After 50 min have elapsed (or immediately for someone with poor initial recall), 20 questions are asked, 15 of which are designed as leading or misleading. One obtains a Yield 1 score based on the number of leading questions given in to by the subject. The subject is then firmly told that a number of errors were made (regardless of actual performance), that the questions will be repeated, and he or she needs to be more accurate. After the questions are repeated, the examiner calculates the number of times the subject has given in to the second round of leading questions, for a Yield 2 score. Shift is also measured. This is the extent to which a subject has changed his response, right or wrong, to a different response. Yield 1 and Shift scores are then summed to produce a Total Suggestibility (Total GSS) score. Yield 1, Yield 2, Shift and Total GSS scores can then be compared to an appropriate normative group (e.g., adults or juveniles).

Interrogative Suggestibility Research

The GSS was developed using population groups from Iceland and Great Britain. This should not present a problem in the use of the test with populations from the United States. Although there were some ethnic differences in suggestibility scores, the differences were minor compared to other sources of variance [12]. Even so, the GSS has proved useful in providing relevant behavioral samples as how a defendant gives in to leading questions and shifts from one response to a different response under pressure [13]. There have been inconsistent findings concerning gender and suggestibility [9] as measured by the GSS, showing no simple correlations.

Age has also been shown to have an influence on an individual's suggestibility. Children below the age of 12 have been shown to be more suggestible than adolescents. They had higher Yield and Shift scores. Yet by age 12 and older, adolescents were no more likely than adults to give in to leading information. They were more likely to change to a different response when given negative feedback compared to adults. By age 17, there were no differences in either Yield or Shift types of suggestibility compared to older individuals [9].

Numerous research studies summarized by Gudjonsson [9] have shown an inverse relationship between intelligence and interrogative suggestibility. Individuals with mental retardation and others of low intelligence are more suggestible than those with higher intelligence. In particular, people with mental retardation are substantially more likely to change their answers in response to feedback that is even friendly, as opposed to feedback that is unfriendly or neutral [14, 15]. Memory is also correlated with suggestibility [9]. The faster the memory deteriorates, the more suggestible the person becomes. Individuals who have poor memory do not trust their own judgments and rely more on what others tell them. An individual is thus more prone to give in to misleading information. Shifting from one response to a different response is more associated with interpersonal and social–psychological factors.

Research has also shown that poor assertiveness, evaluative anxiety, state anxiety, and avoidance coping strategies are related to suggestibility. More chronic forms of anxiety may also be related [16].

Research has shown that prior criminal convictions decrease the individual's suggestibility [9]. It is assumed that the prerequisite of interpersonal trust is no longer as potent in those who have prior negative experiences with law enforcement. Sleep deprivation is also correlated with suggestibility, particularly after negative feedback [17–19]. The longer the sleep deprivation occurs, the higher the suggestibility becomes. The relationships of suggestibility with sleep deprivation combined with state anxiety are important components when considering police interrogation procedures. Mental illness *per se* is not correlated with suggestibility [9].

Related Constructs

Mental health professionals can provide relevant data to judges and juries regarding psychological factors which heighten the risk of a specific defendant producing an involuntary waiver of rights and/or a false or coerced confession. This assessment is case-specific and thus, it is important to integrate a

defendant's vulnerabilities with interrogation tactics and techniques actually used. Case law [20] allows the jury to hear such testimony to help them to decide how much weight to give to a defendant's confession.

In addition to a comprehensive clinical interview and review of third party data, psychological testing provides useful information regarding a defendant's reality testing, judgment, functioning under stress, impulse control, accommodation, assertiveness, and interrogative suggestibility.

Gudjonsson states that *compliance* is another psychological construct that is related to false confessions. Compliance differs from suggestibility in that compliance does not require a private or internal acceptance of the proposition or request, that is, the defendant may confess or give a statement about the crime, even though he does not believe or agree with what is said [9]. Compliance is more directly related to the coerced-compliant type of false confession. Although suggestibility and compliance are two different constructs, research has shown a modest correlation between the two [21, 22]. An evaluation of interrogative suggestibility is warranted even in those cases where there is no doubt the defendant knew whether or not he committed the offense. Since most police interrogations involve some misleading or leading questioning and some pressure for the suspect to change their answers, information regarding a defendant's susceptibility to those procedures is relevant.

A third construct specified by Gudjonsson that is related to confessions is *acquiescence*. It is the tendency to act in accordance with the wishes of others, to affirm information, regardless of the content [23]. Gudjonsson [9] reviewed the research which shows only weak correlations between acquiescence and suggestibility. There was little correlation between acquiescence and compliance.

Gudjonsson has shown [9] that those who falsely confess obtain higher GSS scores than those who do not. A high score on the GSS does not signify a false confession though. A highly suggestible individual may confess accurately about a crime. A low suggestible individual may falsely confess. In addition to the GSS, the use of personality tests such as the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) and 16 Personality Factor (16 PF) can help the clinician assess compliance and acquiescence and integrate those constructs with other aspects of an individual's psychological functioning at the time of the interrogation. As suggestibility is correlated with intelligence and memory, intelligence testing is usually warranted.

References

[1] Inbau, F., Reid, J., Buckley, J. & Jayne, B. (2001). *Criminal Interrogation and Confessions*, 4th Edition, Aspen Publishing Company, Gaithersburg.

[2] Miranda v. Arizona, 384 U.S. 436 (1966).

[3] Fare v. Michael, 442 U.S. 707 (1979).

[4] Ofshe, R. & Leo, R. (1997). The decision to falsely confess: rational choice and irrational and irrational action, *Denver University Law Review* **74**, 979–1122.

[5] Kassin, S. & McNall, K. (1991). Police interrogations and confessions: communicating promises and threats by pragmatic implication, *Law and Human Behavior* **15**, 233–251.

[6] Cassell, P.G. (1996). Miranda's social costs: an empirical reassessment, *Northwestern University Law Review* **90**, 387–499.

[7] Kassin, S. & Wrightsman, L. (1985). Confession evidence, in *The Psychology of Evidence and Trial Procedures*, S. Kassin & L. Wrightsman, eds, Sage Publications, London, pp. 67–94.

[8] McCann, J. (1998). A conceptual framework for identifying various types of false confessions, *Behavioral Sciences and the Law* **16**, 441–453.

[9] Gudjonsson, G. (2003). *The Psychology of Interrogation and Confessions: A Handbook*, John Wiley & Sons, West Sussex.

[10] Gudjonsson, G. (1997). *The Gudjonsson Suggestibility Manual*, Psychology Press, Hove UK.

[11] Gheorghiu, V.A. (1989). The difficulty in explaining suggestion: Some conceivable solutions, in *Suggestion and Suggestibility: Theory and Research*, V.A. Gheorghiu, P. Netter, H.J. Eysenck & R. Rosenthal, eds, Springer, London, pp. 99–112.

[12] Gudjonsson, G., Clare, I., Rutter, S. & Pearse, J. (1993). Persons at risk during interviews in police custody: the identification of vulnerabilities, *Royal Commission on Criminal Justice*, HMSO, London.

[13] Frumkin, I.B. & Garcia, A. (2003). Psychological evaluations and the competency to waive *Miranda* rights, *The Champion* **27**(9), 12–23.

[14] O'Connell, M., Garmoe, W. & Goldstein, N. (2005). *Miranda* comprehension in adults with mental retardation and the effects of feedback style on suggestibility, *Law and Human Behavior* **29**, 359–369.

[15] Everington, C. & Fulero, S. (1999). Competence to confess: Measuring understanding and suggestibility of defendants with mental retardation, *Mental Retardation* **37**(3), 212–220.

[16] Gudjonsson, G., Rutter, S. & Clair, I. (1995). The relationship between suggestibility and anxiety among suspects detained at police stations, *Psychological Medicine* **25**, 875–878.
[17] Blagrove, M. (1996). Effects of length of sleep deprivation on interrogative suggestibility, *Journal of Experimental Psychology. Applied* **2**(1), 48–59.
[18] Blagrove, M. & Akehurst, L. (2000). Effects of sleep loss on confidence–accuracy relationships for reasoning and eyewitness memory, *Journal of Experimental Psychology. Applied* **6**, 59–73.
[19] Blagrove, M., Cole-Morgan, D. & Lambe, H. (1994). Interrogative suggestibility: the effects of sleep deprivation and relationship with field dependence, *Applied Cognitive Psychology* **8**, 169–179.
[20] Crane v. Kentucky, 476 U.S. 683, 106 S.Ct. 2142 (1986).
[21] Gudjonsson, G. (1989). Compliance in an interrogation situation: a new scale, *Personality and Individual Differences* **10**, 535–540.
[22] Gudjonsson, G. (1990). The relationship of intellectual skills to suggestibility, compliance and acquiescence, *Personality and Individual Differences* **11**, 227–231.
[23] Cronbach, L. (1946). Response sets and test validity, *Educational and Psychological Measurement* **6**, 475–494.

I. Bruce Frumkin and Gretchen M. Lamendola

Intersecting Lines: Documents

Introduction

In the study of questioned documents, it may be important to determine the order of execution of certain elements of a document. One of the most common sequencing problems is the order of intersecting writing lines. In both criminal investigations and civil litigation, the order of intersecting lines can take on great importance. For example, one party to a contract may be accused of altering the document by wrongly adding a handwritten notation. If this handwritten entry intersects a signature (*see* **Handwriting and Signatures, Comparison of**) or other text, a forensic document examiner may be asked which came first, the signature/text or the handwritten notation. The answer to that question may determine whether the contract has been altered (*see* **Alterations: Erasures and Obliterations of Documents**).

Line intersection problems have concerned document examiners for more than 100 years. In 1901, Ames discussed an optical illusion that is familiar to every document examiner. That is, when a dark ink and a lighter ink cross each other, the darker ink will almost always appear to be on top regardless of the actual sequence of execution. He also discussed another phenomenon well known in his time – writing ink flowing into a previously written line as the first line crosses the second [1].

Since Ames' time, forensic document examiners have continually conducted research into the problem of line sequencing. Early writers were generally pessimistic regarding the potential success of determining the sequence of intersecting ink lines [2], and over the years many examination methods were proposed. Some of these methods are more reliable than the others; however, no method will allow every question to be answered [3]. It is imperative that sequencing problems be approached with caution. Most document examiners will agree that "substantial, repeated intersections of two writings offer a higher probability of success than a single, indifferent intersection . . ." [4].

There are no books dedicated to the forensic examination of line intersections. Texts on the general topic of forensic document examination dedicate only a small amount of space to the subject. Most of the information regarding the examination of line intersections is in forensic science journals and unpublished conference papers. Some of these papers provide conflicting information. Some of the methods found in these articles and papers are rarely, or perhaps never, used in actual forensic casework.

While line crossings often involve ink from writing pens, sequencing problems may also involve computer-printed text, typewritten entries, paper folds, and handwriting indentations. When line crossings involve the same or similar writing or printing mediums, they may be termed *homogeneous line intersections*. When different mediums intersect one another, the term *heterogeneous line intersection* is appropriate [5].

Microscopic Examination

Numerous methods have been proposed to determine the sequence of intersecting ink lines. In 1996, Poulin

published a review of more than 30 methods [6]. For forensic document examiners, nondestructive methods of examination are always preferable; however, some methods used to examine line intersections are destructive to the document. Perhaps the best-known nondestructive technique used to determine the order of writing strokes is an examination of the intersecting lines using a stereo binocular microscope. The separate optical paths of this microscope provide a three-dimensional image, which is essential in the microscopic study of line crossings.

Document examiners use both vertical and oblique illumination in conjunction with the microscopic examination of a line intersection. When vertical lighting is used, the presence or absence of a specular reflection is important [7, 8]. In some heterogeneous line crossings, such as when certain inks cross a carbon-film-typewritten or laser-printed character, a specular reflection may be observed if the ink is on top (Figure 1). Conversely, the lack of a specular reflection may be indicative of the ink being underneath [9, 10].

Oblique lighting, also referred to as *grazing or side lighting*, is useful in studying the groove caused by a writing stroke. Interruptions of the writing groove can sometimes be used to determine the order of the writing strokes. On occasion, a pen will skip when it hits the groove of the initial stroke. This causes a gap in the second line, which also provides evidence of the sequence of writing. An accumulation of excess ink can occur on the far edge of an underlying writing groove when crossed by a second pen. When different colored inks cross, particles of the first ink may be carried away by the second. Sometimes an ink line narrows when it crosses the groove of an existing writing line [11].

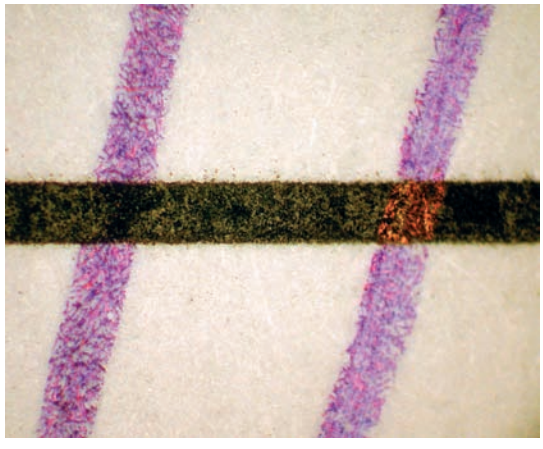


Figure 1 Microscopic examination of line intersections with vertical illumination showing, on the right, a specular reflection at the point of the intersection indicating that the ballpoint pen ink is on top of the toner. The intersection on the left does not display a specular reflection. For this intersection, the ink is underneath the toner

Typically, document examiners conduct microscopic examinations of line intersections at relatively low magnification (60× and below). Higher power magnification (125–500×) has also been used to determine sequence of execution when lines do not intersect in the traditional sense. For example, laser-printed and photocopied documents are covered with scattered toner particles that are invisible to the naked eye. Aginsky reported on a microscopic technique used to determine whether the stray toner particles are below or above ballpoint pen ink writing [12].

Another microscopic technique involves the use of the scanning electron microscope (SEM) [13–15]. This method can be effective with some types of line intersection problems, but can also be destructive to the document. Depending on the type of SEM used, sample preparation may require extraction of the line crossing from the document, as well as a coating of the line crossing with a conductive material. The latter can be avoided with the latest generation of SEMs (environmental SEM (ESEM)). The SEM can also be used in conjunction with the Kromekote paper technique discussed below [16, 17].

Lifting Method

One somewhat destructive technique is the use of Kromekote paper to lift an image of the line intersection [18]. Kromekote paper is a highly glossed coated paper. The technique is reportedly effective with ballpoint pen crossing when two or more line crossings are available for examination. The technique is not reliable with fluid inks and pencil strokes, but some researchers have reported success with other mediums (*see* **Writing Instruments and Printing Devices**). The technique requires placement of a small piece of Kromekote paper over the suspect line intersection. Pressure is applied to the back of the Kromekote paper using a heated tacking iron. In some variations of the technique, the application of cold pressure, pressure, and steam, as well as the use

of solvents, is involved. Photographic paper has been used in the place of Kromekote paper [19].

If performed correctly, a portion of the ink at the line intersection is transferred to the Kromekote paper. A careful study of the lifted intersection provides information regarding the sequence of the line crossing by observing the edges of the intersection. Other lifting methods, such as the wax lift method [20], have also been used.

The Kromekote method, as well as other lifting methods, should be used only after a careful microscopic examination of the suspect intersection. It may not be possible to conduct an effective microscopic examination after the Kromekote method is conducted.

Infrared Examination Methods

Forensic document examiners commonly use various nondestructive lighting techniques to differentiate inks (*see* **Ink Analysis**). One of these techniques relies on electronic imaging systems capable of "seeing" into the near-infrared region along with various bandpass and longpass filters, as well as a high intensity light source. Using this equipment, some inks can be observed to luminesce, while other inks do not. Utilizing another technique, some inks will transmit light, effectively becoming invisible.

The sequence of inks cannot be determined on the basis of the general appearance of a luminescing ink crossing a nonluminescing ink. In some cases, a luminescing ink below a nonluminescing ink will appear to be on top and *vice versa*. For this reason, infrared techniques have limited usefulness in line sequencing problems.

If a nonluminescing ink crosses a luminescing ink while the latter is still wet, it may be possible to see some of the luminescing ink that has been carried away along the nonluminescing ink line (Figure 2). This technique is useful in determining whether the nonluminescing ink was written last [21]. The examiner must ensure that the luminescing ink has not merely bled into the nonluminescing ink line. Of course, this technique would not be useful with a homogeneous ink crossing, and if the luminescing ink has dried, it is unlikely that this effect can be observed. Also, note that in the illustration there is an optical illusion that makes the luminescing ink appear to be uppermost. A variation of this technique involves the use of a laser to generate the luminescence in the ink [22]. If the ink from a porous tip pen transmits light, and the pen crosses a pencil writing, it may be possible to see particles of graphite carried away in the ink line [23].

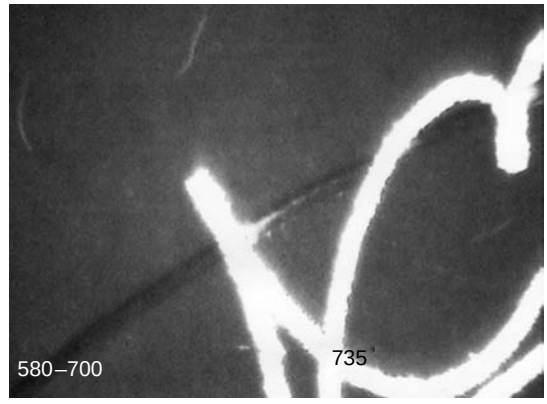


Figure 2 Image showing ink crossings that are observed using an infrared luminescence technique. The underlying freshly written and luminescing fluid ink has been pulled along the ink line of the uppermost and nonluminescing ink

Radley reported a "waisting effect" that sometimes occurs when a nonluminescing ballpoint ink crosses a luminescing ballpoint ink. This effect is the result of the second pen entering the groove of the first ink line, which can cause the crossing stroke to have an hourglass-like appearance at the point of the intersection [24].

Paper Folds

It is often possible to determine whether an ink line that intersects a paper fold was written before or after the paper was folded (*see* **Paper Analysis**). An ink line written prior to a paper fold will generally have a smooth and continuous appearance, although, in some cases, a break in the writing line occurs due to wear and tear at the fold.

When the writing line crosses a previously existing fold, the point of the pen may skip for a short distance prior to returning to the paper. In addition, an accumulation of excess ink may appear at the point where the pen point hit the fold (Figure 3). When a fluid ink is used, the ink may flow into the paper fold [25].

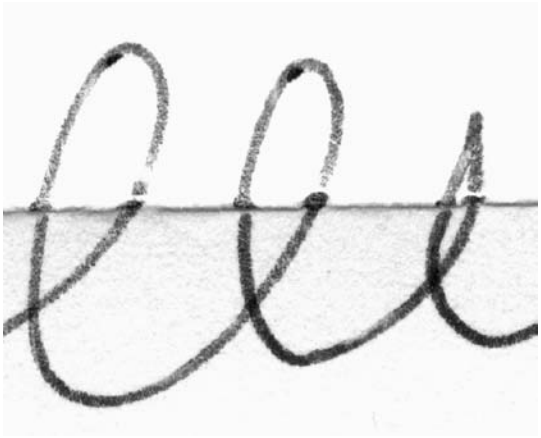


Figure 3 Paper folding prior to ballpoint pen writing. An accumulation of ink appears in the area where the point of the pen hit the fold. In addition, a break in the writing line appears where the point of the pen skipped prior to returning to the paper

Ink and Toner Intersections

Laser printers and electrostatic photocopiers are ubiquitous in our business culture. These machines make documents by using a toner to construct the text on the paper. The toner is a powder that is fused to the paper by heat and pressure. It stands to reason that questions regarding the sequence of ink and toner intersections are common.

Some aspects of ink and toner line intersections were discussed in the section of this article discussing microscopic examinations. Another frequent method of examining ink and toner intersections involves scraping away the toner at the point of the intersection. Obviously, this method is destructive to the document. One might expect that if ink is found under the toner, then the ink was placed on the paper first. If no ink is found, the writing occurred after the document was printed or copied. In actual forensic casework, it is not so simple.

Research has shown that fluid inks can soak through the toner and into the paper below [26]. An analogy would be pouring water over a pile of gravel. The water would run through the empty spaces in the gravel and soak through to the ground. At a microscopic level, laser-printed text is somewhat porous. Fluid inks may soak through to the bottom, giving the impression that the ink line was made first. This is less likely to occur when ballpoint pen inks, which are more viscous, are used. Because of this soaking-through phenomenon, caution must be used when interpreting the results of this method.

Ink and Typewritten Text Intersections

Most modern typewriters use correctable carbon film ribbons. The typewritten characters are created by the pressure of the typeface striking the film of the ribbon. It has been previously mentioned that a ballpoint pen ink over carbon film ribbon will cause a specular reflection. The specular reflection indicates that the writing ink is uppermost.

Document examiners have also used the scraping method discussed in the previous section regarding toner [27]. The same cautions apply. Fluid ink can soak through the carbon film to the paper below giving the false impression that the writing came first. Another drawback with the scraping method is the risk that ink below the carbon film might be removed during the scraping process. This is more likely to occur with ballpoint pen inks since these inks do not soak into the paper fibers like fluid inks.

Since carbon film ribbons are designed to be removable from the paper surface, another method of examining ink and carbon film intersections involves lifting the carbon from the paper with an adhesive tape or a correction ribbon.

Electrostatic Detection Apparatus Technique

The electrostatic detection apparatus (ESDA) is a device that is used to detect indented writing impressions on paper. When writing occurs on one sheet of paper with another sheet underneath, indented writing impressions are created on the sheet below. Mohammed demonstrated that the ESDA is useful for sequencing inkjet printing and indented writing impressions. He found that the ESDA is not useful in sequencing laser printing and writing impressions [28]. The ESDA has also been used to sequence writing impressions with ballpoint pen writing [29, 30]; fiber tip, porous tip, and rollerball pen writing [31]; and writing made with highlighter ink [32].

Other Techniques

A study of the embossing on the reverse side of paper may provide information as to which writing line

crossed another first. One study reported successful results on more than 90% of 300 line crossings [33].

Researchers in Taiwan have applied chromaticity measurements and clustering techniques [34] as well as the CIELAB uniform color system and sample duplication [35] to the problem of line intersections.

Laser profilometry, a technique that produces a 3D topographical representation of a line intersection, has a high degree of accuracy under testing conditions. It also has the advantage of being nondestructive [36, 37].

Attenuated total reflectance-Fourier transform infrared (ATR-FTIR) techniques have been successfully used on line sequencing problems involving laser printer toner from two different models of laser printers, as well as toner and writing ink sequencing problems [38, 39].

Time-of-flight secondary ion mass spectroscopy (ToF-SIMS) has been used to assess the sequence of heterogeneous ink crossings by measuring the secondary ion mass spectra of the inks [40].

Atomic force microscopy (AFM), a type of scanning probe microscopy, reportedly produces qualitative results comparable to those from scanning electron microscopy without the drawbacks associated with many versions of the SEM [41].

References

[1] Ames, D. (1901). *Ames on Forgery*, The Boston Book, Boston, p. 67.

[2] Villanova, P. (1969). Establishing the sequence of superimposed lines, *International Criminal Police Review* **24**, 214–220.

[3] Conway, J. (1959). *Evidential Documents*, Charles C. Thomas, Springfield, p. 164.

[4] Hilton, O. (1982). *Scientific Examination of Questioned Documents*, Revised Edition, Elsevier Science Publishing, New York, p. 110.

[5] Mathyer, J. (1980). The problem of establishing the sequence of superimposed lines, *International Criminal Police Review* **35**, 238–250, 271–281.

[6] Poulin, G. (1996). Establishing the sequence of strokes: the state of the art, *International Journal of Forensic Document Examiners* **2**, 16–32.

[7] Godown, L. (1982). Recent developments in writing sequence determination, *Forensic Science International* **20**, 237–232.

[8] Daniels, J. (2004). Thermal gradient mechanism of line crossing anomaly, *Journal of the American Society of Questioned Document Examiners* **7**, 77–82.

[9] Singla, A., Jasuja, O. & Kaur, J. (1994). Determining the sequence of intersecting ball-pen lines and correctable carbon ribbon strokes, *Forensic Science International* **64**, 141–145.

[10] Planty, M. (1997). Determining the relative chronology of intersecting ball-point pen lines and laser printed document marks – Linton Godown revisited, *International Journal of Forensic Document Examiners* **3**, 31–34.

[11] Strach, S. (1978). Establishing the sequence of intersecting ball-point pen strokes, *Forensic Science International* **11**, 67–74.

[12] Aginsky, V. (2002). Determining the sequence of non-intersecting media on documents: ballpoint pen ink and laser toner entries, *Journal of the American Society of Questioned Document Examiners* **5**, 1–4.

[13] Tollkamp-Schierjot, C. & Fackler, H. (1996). Use of low voltage SEM in the detection of forgeries, *International Journal of Forensic Document Examiners* **2**, 333–341.

[14] Waeschle, P. (1979). Examination of line crossings by scanning electron microscopy, *Journal of Forensic Sciences* **24**, 569–577.

[15] Baier, P. (1983). Technical improvements of scanning electron microscope methods in document examination, *Forensic Science International* **22**, 265–278.

[16] Mathyer, J. & Pfister, R. (1984). The determination of sequence of crossing strokes by the 'kromekote' paper lifting process and by the scanning electron microscopic method, *Forensic Science International* **24**, 113–124.

[17] Koons, R. (1985). Sequencing of intersecting lines by combined lifting process and scanning electron microscopy, *Forensic Science International* **27**, 261–276.

[18] Igoe, T. & Reynolds, B. (1982). A lifting process for determining the writing sequence of two intersecting ball-point pen strokes, *Forensic Science International* **20**, 201–205.

[19] Gupta, A., Gulshan, R. & Chugh, O. (1987). Determination of writing sequence of strokes of ball-point pen versus ball-point pen & other conventional writing instruments, *Forensic Science International* **34**, 217–223.

[20] Taylor, L. (1984). Intersecting lines as a means of fraud detection, *Journal of Forensic Sciences* **29**, 92–98.

[21] Kelly, J. & Lindblom, B. (2006). *Scientific Examination of Questioned Documents*, 2nd Edition, CRC Press, Boca Raton, pp. 327–328.

[22] Day, S. (1984). Evaluation of the application of the argon-ion laser to document examination: a review of casework and experimental data, *Journal of the Forensic Science Society* **25**, 285–296.

[23] Howes, D. (1985). *Sequence of Writing – Pencil and Porous Tip Pen. Proceedings of the International Symposium on Questioned Documents*, U.S. Government Printing Office, Washington, D.C, p. 129.

[24] Radley, R. (1982). Determination of sequence of ball point pen writing utilising infrared luminescence techniques, *Journal of the Forensic Science Society* **22**, 373–375.

[25] Osborn, P. (1964). Discussion of the sequence of fluid ink lines and intersecting paper folds, perforations, tears,

and cut edges, *Journal of Criminal Law and Criminology* **55**, 412–419.
[26] Novotny, M. & Westwood, P. (2005). Determining the sequence of original ink writing and toner printing, *Journal of the American Society of Questioned Document Examiners* **8**, 37–47.
[27] Hart, L. & Carney, J. (1989). Typewriting versus writing instrument: a line intersection problem, *Journal of Forensic Sciences* **34**, 1329–1335.
[28] Mohammed, L. (1998). Sequencing writing impressions and laser printing or inkjet printing using the ESDA, *Journal of the American Society of Questioned Document Examiners* **1**, 40–42.
[29] Radley, R. (1993). Determination of sequence of writing impressions and ball pen inkstrokes using the ESDA technique, *Journal of the Forensic Science Society* **33**, 69–72.
[30] Giles, A. (1993). Extending the ESDA's capability: the determination of the order of writing and impressions using the technique of electrostatic detection, *Forensic Science International* **59**, 163–168.
[31] Radley, R. (1995). Determination of sequence of intersecting ESDA impressions and porous tip, fiber tip, and rollerball pen inks, *Science & Justice* **35**, 267–272.
[32] Allen, M. (1997). Using ESDA to sequence highlighter ink and impressions, *International Journal of Forensic Document Examiners* **3**, 49–51.
[33] Jasuja, O., Singla, A. & Chattopadhyay, P. (1987). A simple method for determining the sequence of intersecting ball pen lines, *Journal of the Forensic Science Society* **27**, 227–230.
[34] Liu, K., Cheng, K., Shieh, T., Miou, C. & Jeng, B. (1997). The determination of the order of writing of cross strokes by chromaticity measurements and clustering techniques, *International Journal of Forensic Document Examiners* **3**, 138–145.
[35] Cheng, K., Liu, K., Lee, S. & Shieh, D. (1998). Determination of the writing sequence of crossing strokes by CIELAB color system and sample duplication, *International Journal of Forensic Document Examiners* **4**, 12–21.
[36] De Kinder, J. & Berx, V. (2005). The application of profilometry in the analysis of the lines crossing, *Journal of the American Society of Questioned Document Examiners* **8**, 1–8.
[37] Spagnolo, G. (2006). Potentiality of 3D laser profilometry to determine the sequence of homogenous crossing lines on questioned documents, *Forensic Science International* **164**, 102–109.
[38] Gal, T., Sandor, J. & Karoly, A. (2007). Determining the sequence of crossed lines by FT-IR-ATR-microscopy, *Global Forensic Science Today* **1**, 8–11.
[39] Bojko, K., Roux, C. & Reedy, B. An examination of the sequence of intersecting lines using ATR-FTIR spectral imaging. *Journal of Forensic Sciences*, in press.
[40] He, A., Karpuzov, D. & Xu, S. (2006). Ink identification by time-of-flight ion mass spectroscopy, *Surface and Interface Analysis* **38**, 854–858.
[41] Kasas, S., Khanmy-Vital, A. & Dietler, G. (2001). Examination of line crossings by atomic force microscopy, *Forensic Science International* **119**, 290–298.

FARRELL C. SHIVER

Interview: Police *see* Interrogation

Interviewer Influence upon Memory *see* Children: Suggestibility of

Intoxication: Alcohol *see* Alcohol: Behavioral and Medical Effects

Investigation: Scene *see* Evidence Collection and Preservation: Casting

Involuntary Commitment *see* Civil Commitment

Ipse Dixit Testimony

This is a Latin term, meaning "he said it himself". It refers to a statement wherein the speaker is the sole authority for its truth or assertion. When dealing with forensic evidence, the phrase may be used where an expert witness asserts that the methodology

used in examining evidence is sound, reliable, and replicable, but, upon further request, fails to provide evidence for this assertion. In jurisdictions that follow precedents such as those embodied in *Daubert v. Merrell Dow Pharmaceuticals (1993) (see* ***Daubert v. Merrell Dow Pharmaceuticals****), Kumho Tire v. Carmichael* (1999) (*see* ***Kumho Tire v. Carmichael***), or *General Electric v. Joiner* (1997) (*see* **General Acceptance Test for Novel Expert Evidence**), the courts will ordinarily not allow self-serving assertions of reliability to be received in evidence. The court precedents referred to provide, specifically, that *ipse dixit* evidence is insufficient to prove the reliability of a technique or methodology in the absence of proof of experimentation by other scientists, the existence of data bases, and a peer review literature compelling a belief in the conclusion of accuracy.

The Supreme Court of the United States, in *General Electric v. Joiner,* made the point specifically when discussing the difference between the reliability component required for admissibility of the evidence by *Daubert*, which may be affected by methodological error in a process or technique, and an expert witness or practitioner's erroneous conclusion, which would affect the qualifications of the witness but not necessarily the soundness of the underlying technique. The Court added, however:

> [C]onclusions and methodology are not entirely distinct from one another. Trained experts commonly extrapolate from existing data. But nothing in either *Daubert* or the Federal Rules of Evidence requires a district court to admit opinion evidence which is connected to existing data only by the *ipse dixit* of the expert. A court may conclude that there is simply too great an analytical gap between the data and the opinion proffered.

Related Articles

Expert Opinion in Court: a Comparison of Approaches

General Electric v. Joiner

***In Limine* Motions and Hearings**

ANDRE MOENSSENS

Irresistible Impulse *see* Behavioral Science Evidence

Irresistible Impulse Test of Insanity *see* Insanity: Defense

Isopropanol *see* Alcohol: Use, Abuse, Tolerance, and Dependency



Jaffee v. Redmond ***see*** Duty to Warn

Judicial Notice of Scientific Principles and Facts

Judicial notice is a legal principle applicable primarily, though not exclusively, in common law, adversary system jurisdictions (*see* **Adversary Systems of Justice**). It is a rule of judicial economy and convenience, designed to save time in litigation, whereby a court may "judicially note" the truth and accuracy of a fact, thereby relieving the party who has the burden of proof of that fact from having to present evidence of it. Judicial notice, then, is a substitute for evidence (*see* **Expert Opinion in Court: Civil Law Jurisdictions (France, Germany, Italy, and Spain)**).

When a court takes judicial notice of a fact, the opposing party is often also prohibited from presenting evidence of its untruth. It would indeed reduce the judicial process to a mockery if a litigant were permitted to offer proof that the Eiffel Tower is located in Berlin!

Most jurisdictions that sanction use of the rule place limitations on the type of facts that a judge can note judicially (*see* **Evidence: Rules of**). The common law has always permitted courts to take judicial notice of legislative facts and law, such as constitutions, statutes, decisional law, and regulations in effect within their jurisdictions, as long as these legal principles are published and capable of easy verification. A court is not required, however, to note the laws of foreign jurisdictions.

As to adjudicative facts, courts can take judicial notice of

(i) matters of common knowledge;
(ii) facts capable of certain verification; and
(iii) scientific principles.

Matters of common knowledge must be generally known to the public residing in the jurisdiction where the court sits. If the trial judge personally happens to be familiar with a fact that is relevant to the case, but that is not generally known in the jurisdiction where the court sits, the court cannot take judicial notice of that fact. Whatever party would like the fact to be considered as evidence in the case must present supporting evidence of it. Examples of commonly known matters are most geographical and historical data that are relevant to a particular case. Examples are: that certain streets within the city run in north–south direction; that traffic on a particular avenue runs one-way in westerly direction; that Washington DC is the capital of the United States; that Massachusetts was one of the original thirteen states of United States; that the Eiffel Tower is located in Paris, France.

The category of facts capable of easy verification applies to anything that can be readily found and

verified by reference to an almanac, a calendar, or a reliable compilation of data generally considered to report its data accurately. These facts need not be commonly known to the general public. Examples of such facts include the state of the moon on a particular hour of a day; the main telephone number of a company or institution in the city; the weather conditions on the day that an incident occurred; an electrical malfunction that caused the power to be interrupted at a particular time.

Of greatest interest to forensic scientists is the third category of judicially noted facts: well-established scientific facts the accuracy of which is generally beyond doubt. Some scientific facts are truly beyond doubt. Examples would be: that the earth is a sphere and not flat; that the sun rises in the East and sets in the West; that the "big bang" theory is the prevailing scientific view of the origin of the universe; that radar principles permit the reliable measurement of speed [1, 2].

On the accuracy of some scientific facts (*see* **Scientific Method Compared to Legal Method**), however, courts in different jurisdictions may have inconsistent views. In many older precedent-setting court decisions, it has been held that the digits of each finger or thumb bear unique friction skin ridge characteristics [3]. This reasoning was thereafter also extended by some courts to the uniqueness of small areas of a latent fingerprint. A more recent influential and well-reasoned court decision held that while the original part of the equation – ridge structure uniqueness – can be properly noted, courts should not take judicial notice of the fact that a particular partial or blurred latent finger mark of an unknown person can be positively matched to that of a known person [4].

When scientific principles underlying certain forensic techniques are truly beyond question, their widespread acceptance is seldom contested in court. The issue of judicial notice of the principle may never be presented to a judge. However, when it comes to the application of certain principles to the facts of a specific case, a court will rarely be in a position to take judicial notice of the specific result.

Examples are easy to find. Deoxyribonucleic acid (DNA) forensic applications rest on Nobel-prize winning scientific principles about the molecular nature of DNA, researched and articulated at Cambridge University in the United Kingdom by scientists James Watson and Francis Crick in 1953 [5]. Between these underlying principles and the ultimate opinion of an expert witness evaluating the "match" of a known and unknown biological specimen lie myriad steps, each of which may require subjective evaluation by the expert subject to error or misinterpretation. The results of these applications are not appropriate fact for judicial notice by courts. In all those and most other forensic applications, a qualified witness must present proof and opinion evidence about the results of an analysis specific to the case in point.

References

[1] State v. Tomanelli, 153 Conn. 365, 216 A.2d 625 (1966).
[2] People v. Magri, 3 N.Y.2d 562, 147 N.E.2d 728 (1958).
[3] Grice v. State, 142 Tex. Crim. 4, 151 S.W.2D 211 (1941).
[4] United States v. Mitchell, 365 F.3d 215 (3rd Cir. 2004), certiorari denied 543 U.S. 974 (2004). As evidence of this lack of certainty, the court cited the fact that in multiple days of expert testimony, different views were expressed by a number of expert witnesses.
[5] Moenssens, A.A., Henderson, C.E. & Portwood, S.G., *Scientific Evidence in Civil and Criminal Cases*, 5th Edition, Foundation Press, 2007.

ANDRE MOENSSENS

Juror Knowledge *see* Eyewitness Testimony

Jury Decision Making *see* Mitigation Testimony

Jury Dynamics

The right to a trial by jury, guaranteed by the Sixth Amendment to the Constitution for criminal cases and the Seventh Amendment for civil cases, entrusts juries to make some of the most complicated and profound decisions of civic life. How well laypeople are able to function in these roles is a question that

has vexed attorneys, judges, defendants, plaintiffs, victims, and indeed, the general public since the founding of our democracy (*see* **Jury Instructions on Expert Testimony**; **Hearsay Evidence**).

Of particular concern is how jurors make decisions in tough cases – criminal cases in which the evidence is neither so weak as to warrant outright dismissal nor so compelling as to induce a guilty plea, and civil cases where the evidence seems evenly balanced between the two sides.

The Verdict on Jury Verdicts

Over the past few decades, social scientists and legal scholars have begun to understand what goes on behind the closed door of a jury deliberation room. The empirical evidence – gathered by way of interviews of jurors and judges, archival research, experimental field and simulation studies, and most recently, by observing jury deliberations – suggests that juries function reasonably well. Several findings lead to that conclusion. First, judges tend to agree with the jury's verdict in the vast majority of cases [1, 2]; second, in criminal cases, the strength of the evidence is a stronger determinant of a jury's verdict than the extralegal factors such as personality or demographic characteristics of the jurors themselves [3, 4]; and finally, in civil cases, jurors' damage awards are related to the extent of documented injuries and losses, as they should be [5]. Several recent and comprehensive reviews of empirical research on juror and jury decision making [6, 7] detail these and other auspicious findings.

But despite the empirical research showing that jurors and juries make defensible (and usually, quite reasonable) decisions, there are lingering suspicions that, on occasion, they might not be up to the task. One concern focuses on jurors' attention to and reliance on information that is unrelated to case facts brought to light during the trial. Another concern has to do with biases that jurors bring with them into courtrooms and the prospect that those biases distort the ways that they process evidence.

The Effects of Extralegal Information

Jurors should base their verdicts only on the evidence deemed admissible in court, but do they? Information that is legally irrelevant to the decision required of them is termed as *extralegal information*. A criminal defendant's appearance, personal characteristics, and propensity for wrongdoing constitute extralegal information, yet research shows that such extralegal information can indeed influence jurors' judgments. For example, mock jurors who learned about an alleged sex offender's previous criminal acts perceived the crime victim as more credible and the defendant as more likely to be guilty than did jurors who were not so exposed [8]. Trial-related information in the media is another source of extralegal information to which jurors are exposed. Empirical studies based on juror interviews and experimental techniques that systematically vary the kind of extralegal information that jurors hear have shown that such pretrial publicity can bias jurors' in-court decisions [9, 10].

Frequently, extralegal information is conveyed by one side or the other in the courtroom. When this occurs, judges typically instruct jurors to disregard this so-called inadmissible evidence so that it does not distort their reasoning or mislead them, resulting in an incorrect verdict. On other occasions, judges instruct jurors that they may use certain evidence for one purpose (e.g., to assess a defendant's credibility) but not for another (e.g., to determine guilt). Although commentators have expressed skepticism about jurors' ability to follow these instructions [11] and US Circuit Court judge, Learned Hand, described the instructions as "recommendation[s] to the jury of a mental gymnastic which is beyond, not only their powers, but also anybody else['s]" [12], most judges assume that jurors can follow these directives.

A recently conduced meta-analysis (a quantitative technique that combined the results of 175 individual studies – primarily simulation studies – to identify reliable patterns in the data) examined the effects of inadmissible evidence on jurors' decision making and the efficacy of judicial instructions to disregard it [13]. The meta-analysis showed that jurors' verdicts were slanted in the direction of the inadmissible evidence; i.e., when the evidence favored the prosecution, guilty verdicts increased; when it favored the defense, guilty verdicts decreased. Further, judicial admonitions to disregard inadmissible evidence fell on slightly deaf ears: both proprosecution and the prodefense biases remained even after the judge instructed jurors to disregard the evidence, though the lingering effect of inadmissible evidence was reduced when the judge gave a rationale for his or her decision by stating, for example, that the evidence was

hearsay, not reliable, or without bearing on the case. The biasing effects of inadmissible evidence are also reduced when the judge issues a reminder at the conclusion of the trial to disregard any evidence so deemed. In general, though, jurors apparently have difficulty in ignoring information they believe to be reliable and probative, especially if they also think that it will lead to a just verdict [14].

Distorted Evaluations of Trial Evidence

Judges often remind jurors that their verdicts should be determined only with reference to the evidence and the legal instructions, and should not be influenced by bias, prejudice, or emotion. Appellate courts presume that jurors follow these mandates.

But is it realistic to expect any human to come into such a formal, institutional setting as a trial completely devoid of preexisting sentiments, inclinations, or expectations? Obviously not. Shari Diamond asks us to consider the differential reactions of two hypothetical jurors as they ponder the claims of a plaintiff in an automobile accident lawsuit. The first juror believes that many plaintiffs are excessively litigious and that they pursue frivolous claims in order to line their pockets. The second juror is decidedly more sympathetic; he has a brother who experiences excruciating pain as a result of a soft tissue injury. Diamond suggests that both of these jurors are simply allowing their experiences and commonsense knowledge of the world (and not any overwhelming prejudice or bigotry) to follow them into the courtroom. Psychologists and other social scientists who scrutinize the behavior of jurors are beginning to understand some of the subtle effects that these values and experiences have on jurors' reactions to trial evidence [15].

One way that subtle biases can influence fact-finding is by distorting how jurors interpret evidence as they hear it. In particular, when a juror already tends to favor one side in a case, new evidence tends to be distorted by this predecisional bias so that it is interpreted as supporting the currently favored verdict [16].

In one study, mock jurors who were given case facts, opening statements, witness affidavits, and instructions not to reach "hasty opinions or conclusions" then got successive pieces of evidence and indicated, for each item, whether it favored the plaintiff or defendant and which verdict they tended to favor. Finally, they were asked to choose a verdict. Regardless of the side that jurors favored, they tended to interpret new evidence as supportive of their preferred verdict, perhaps because doing so provided coherence and closure. Doing so meant that they did not have to grapple with inconsistencies and contradictions in the evidence.

Is there a way to counter the effects of these biases? Researchers suggest that an additional pretrial instruction from the judge reminding jurors to avoid even the slightest predecisional leaning might have some ameliorative consequences.

Effects of Jury Deliberations

Much of the empirical research on fact finders in trial settings focuses on individual jurors' cognitive and affective responses to trial-based and extralegal information. But verdicts are not determined by individuals; they result from a deliberative process during which jurors theoretically pool their knowledge and opinions to reach a verdict that reflects a consensus perspective. The extent to which jury decisions are influenced by the process of deliberating is unclear, however. Some data suggest that jury verdicts can be predicted simply by considering the distribution of predeliberation preferences [1] whereas other data show that verdicts can be significantly influenced by the process of deliberation [17]. For example, verdicts in criminal cases tend to show a leniency bias; i.e., the jury's verdict is often more lenient than would be predicted from individual, predeliberation sentiments, suggesting that the arguments of proacquittal jurors must have been persuasive in the jury room.

Laws prohibiting the observation of jury deliberations are in place in all states and the federal court system, though a recent study of jury reform included limited filming of actual deliberations [18]. This means that to understand the process by which individual jurors reach consensus, one must rely on mock jury methodology. This was the research paradigm that Samuel Sommers used to examine the effects that a jury's racial composition has on its decision making [19]. (Sommers and his colleague, Phoebe Ellsworth, had previously determined that individual white jurors tended *not* to be prejudiced in judging the case of a hypothetical black assault defendant so long as race was a salient issue in the case. However, when the "race card" was not played, white jurors were more likely to convict a black defendant

than a similarly described white defendant, suggesting that racial prejudice is still a factor in juror judgments [20].)

In particular, Sommers wondered whether diversity in the composition of the jury could offset the effects of individuals' prejudicial beliefs. He showed a condensed sexual assault trial to juries that consisted either of all-White members or of four White and two Black members. He filmed and analyzed the content of the deliberations and determined that all-White and diverse juries differed in some very important ways. All-White juries made more factual errors than diverse juries yet were less likely to correct inaccuracies than were diverse juries. They also considered fewer case facts and were less likely to discuss race-related issues than were diverse groups.

Perhaps more surprising was Sommers' finding that many of these effects were a result of differential behavior by White jurors, rather than the influence of Black jurors. White jurors on diverse juries were more likely than their counterparts on homogeneous juries to discuss controversial race-related issues. They also raised more case facts and made fewer factual errors. These data tell us something important about juries and the ways that social scientists study them. They tell us that representation of diverse members of society on a jury is not solely a moral and legal ideal, but that it can enhance the process of decision making. They also remind us of the importance of evaluating jurors both as individuals and as collective members.

Helping Jurors and Juries Perform Better

Empirical research on juror and jury decision making has revealed that in many cases, jurors competently perform their duties [17, 21]. However, in addition to concerns about pre-existing bias and prejudice improperly influencing juror decision making, there are also concerns about jurors' ability to handle evidence in complex civil litigation (e.g., antitrust litigation, and class action lawsuits which can involve hundreds of plaintiffs and multiple defendants) as well as complex scientific evidence in some criminal cases (e.g., DNA evidence involving statistical probabilities) and civil cases (e.g., scientific evidence involved in patent litigation). Rather than eliminating jury trials or replacing juries comprised of citizens with juries comprised of experts (sometimes referred to as *blue ribbon juries*), social scientists and legal scholars have found that revising court procedures can reduce the impact of pre-existing biases and assist jurors in evaluating complex evidence. The result has improved the performance and satisfaction of juror and jury.

Reducing Pre-Existing Juror Bias and Prejudice. It is natural for jurors to come to trial with pre-existing attitudes and knowledge about the world and, on occasion, knowledge about the present case. Identifying juror bias and prejudice and determining whether it is likely to interfere with a criminal defendant's right to trial by an impartial jury, as protected by the Sixth Amendment, is the duty of the attorneys representing the parties and the judge overseeing the case. The questioning of members of the jury pool prior to trial (i.e., the voir dire process) provides the opportunity to detect bias and prejudice and remove jurors who are deemed unable to be impartial.

The questioning can be conducted in a group setting or with individual jurors and can be carried out by the judge and/or the attorneys. Potential jurors tend to respond more candidly and completely when questioned by attorneys rather than the judge [22]. Individualized, as opposed to group, questioning of potential jurors also increases the likelihood of honest, candid, and complete answers from jurors, particularly when the questions involve matters of a sensitive or private nature [23]. Individual questioning also helps to avoid potential tainting of the jury pool (e.g., by means of a juror describing pretrial publicity or voicing particularly strong opinions about one of the parties) or of jurors conforming to the group by providing answers that seem to be "correct" or "appropriate" and that do not deviate from the norm [24, 25]. Having potential jurors complete pretrial, supplemental juror questionnaires is another way of attempting to gather candid information from them about potentially problematic biases and attitudes [26].

Assisting Jurors' Comprehension and Memory. Trial procedures that improve jurors' comprehension and memory of the evidence and the applicable law include providing preinstructions to jurors, allowing jurors to take notes during trial, and providing jury instructions in plain English.

In most jurisdictions, the trial judge has the discretion to include at least some information on

the law that will govern the case in the preliminary instructions provided to jurors at the beginning of the trial. These instructions include the basic principles of law that will govern the trial and an explanation of the role and responsibilities of the jury. Such instructions have been found to provide jurors with a framework that helps them to organize trial facts in ways that improve their understanding and memory of the case [27, 28].

Another procedure proposed to improve jurors' memory for trial evidence is allowing them to take notes during trial. Although judges in some states are required to inform jurors that they may take notes during trial, it is more often the case that juror notetaking may be permitted or prohibited at the judge's discretion. Initially courts were concerned that jurors who took notes would be distracted by notetaking and miss ongoing testimony and that, in addition, jurors who took notes might have some unfair advantage over non-notetakers during deliberations. Empirical studies of juror notetaking have shown no effect on jurors' comprehension or memory, but they have also found no evidence to support the concerns about juror distraction or unfair advantage during deliberations [29, 30].

However, compared to non-notetakers, jurors who took notes during trial have rated themselves as more attentive, more involved in the trial, better able to keep up with the proceedings, and more satisfied with the final verdict [30, 31].

Jurors are asked to combine the case facts with the relevant law to decide on a verdict, but understanding the jury instructions on the applicable law can be daunting for laypersons with no formal legal training. The language used in the instructions as well as the way in which the instructions are organized can make it difficult for jurors to comprehend or recall the law on which they have to rely to decide the case. Drafting jury instructions in ways that avoid abstract statements of legal principles and eliminate unfamiliar terminology, avoid awkward grammar, use case-specific language when referring to parties and pieces of evidence, and organize the instructions hierarchically in terms of legal tests and criteria can result in significant improvements in jurors' comprehension of the instructions [32, 33].

In general, then, the verdict on jurors and juries is cautiously optimistic. Although biases and predilections sometimes influence jurors' determination of the facts and application of the law, judges and other court personnel can institute structural and procedural changes to address these concerns and to render jurors' decision making more rational and predictable.

References

[1] Kalven, H. & Zeisel, H. (1966). *The American Jury*, Brown Little, Boston.

[2] Eisenberg, T., Hannaford-Agor, P., Hans, V., Waters, N., Munsterman, G.T., Schwab, S. & Wells, M. (2005). Judge-jury agreement in criminal cases: A partial replication of Kalven and Zeisel's The American Jury, *Journal of Empirical Legal Studies* **2**, 171–207.

[3] Garvey, S., Hannaford-Agor, P., Hans, V., Mott, N., Munsterman, G.T. & Wells, M. (2004). Juror first votes in criminal trials: Black power in the criminal justice system, *Journal of Empirical Legal Studies* **1**, 371–399.

[4] Devine, D., Clayton, L., Dunford, B., Seying, R. & Pryce, J. (2001). Jury decision making: 45 years of empirical research on deliberating groups, *Psychology, Public Policy, and Law* **7**, 622–727.

[5] Greene, E. & Bornstein, B. (2003). *Determining Damages: The Psychology of Jury Awards*, American Psychological Association, Washington, DC.

[6] Diamond, S. (2006). Beyond fantasy and nightmare: A portrait of the jury, *Buffalo Law Review* **54**, 717–763.

[7] Greene, E., Chopra, S., Kovera, M., Penrod, S., Rose, V.G., Schuller, R. & Studebaker, C. (2002). Jurors and juries: A review of the field, in *Taking Psychology and Law into the Twenty-first Century*, J. Ogloff, ed, Kluwer, New York, pp. 225–284.

[8] Bottoms, B. & Goodman, G. (1994). Perceptions of children's credibility in sexual assault cases, *Journal of Applied Social Psychology* **24**, 702–732.

[9] Studebaker, C.A. & Penrod, S.D. (1997). Pretrial publicity: The media, the law and common sense, *Psychology, Public Policy, and Law* **2/3**, 428–460.

[10] Vidmar, N. (2002). Case studies of pre-and midtrial prejudice in criminal and civil litigation, *Law and Human Behavior* **26**, 73–106.

[11] Eichhorn, L. (1989). Social science findings and the jury's ability to disregard evidence under the Federal Rules of Evidence, *Law and Contemporary Problems* **52**, 341–353.

[12] Nash v. U.S., 54 F.2d 1006 (2nd Cir., 1932),. p. 1007.

[13] Steblay, N., Hosch, H., Culhane, S. & McWethy, A. (2006). The impact on juror verdicts of judicial instructions to ignore inadmissible evidence: A meta-analysis, *Law and Human Behavior* **30**, 469–492.

[14] Kassin, S.M. & Sommers, S.R. (1997). Inadmissible testimony, instructions to disregard, and the jury: substantive versus procedural considerations, *Personality and Social Psychology Bulletin* **23**, 1046–1054.

[15] Smith, V. & Studebaker, C. (1997). What do you expect? The influence of people's prior knowledge of crime

categories on fact-finding, *Law and Human Behavior* **20**, 517–532.
[16] Carlson, K. & Russo, J. (2001). Biased interpretation of evidence by mock jurors, *Journal of Experimental Psychology: Applied* **7**, 91–103.
[17] Hastie, R., Penrod, S. & Pennington, N. (1987). *Inside the Jury*, Harvard University Press, Cambridge.
[18] Diamond, S. & Vidmar, N. (2001). Jury room ruminations on forbidden topics, *Virginia Law Review* **87**, 1857–1915.
[19] Sommers, S. (2006). On racial diversity and group decision making: Identifying multiple effects of racial composition on jury deliberations, *Journal of Personality and Social Psychology* **90**, 597–612.
[20] Sommers, S. & Ellsworth, P. (2001). White juror bias: An investigation of prejudice against black defendants in the American courtroom, *Psychology, Public Policy, and Law* **7**, 201–229.
[21] Ellsworth, P.C. (1989). Are twelve heads better than one?, *Law and Contemporary Problems* **52**, 205–224.
[22] Jones, S.E. (1987). Judge- versus attorney-conducted voir dire, *Law and Human Behavior* **11**, 131–146.
[23] Neitzel, M.T. & Dillehay, R.C. (1982). The effects of variation in voir dire procedures in capital murder trials, *Law and Human Behavior* **6**, 1–13.
[24] Suggs, D. & Sales, B.D. (1981). Juror self-disclosure in the voir dire: A social science analysis, *Indiana Law Journal* **56**, 245–271.
[25] Haney, C. (1984). On the selection of capital juries: The biasing effects of the death-qualification process, *Law and Human Behavior* **8**, 121–132.
[26] Bilecki, D. (1989). Efficient method of jury selection for lengthy trials, *Judicature* **73**, 43–47.
[27] ForsterLee, L. & Horowitz, I. (2003). The impact of procedural strategies on information processing in civil litgation, *Judicature* **86**, 184–190.
[28] Smith, V.L. (1991). Impact of pretrial instructions on jurors' information processing and decision making, *Journal of Applied Psychology* **76**, 220–228.
[29] Heuer, L. & Penrod, S.D. (1994). Juror notetaking and question asking during trials: A national field experiment, *Law and Human Behavior* **18**, 121–150.
[30] Dann, B.M., Hans, V.P. & Kaye, D.H. (2006). Can jury trial innovations improve juror understanding of DNA evidence?, *National Institute of Justice Journal* 2–7.
[31] Rosenhan, D.L., Eisner, S.L. & Robinson, R.J. (1994). Notetaking can aid juror recall, *Law and Human Behavior* **18**, 53–61.
[32] Elwork, A., Sales, B.D. & Alfini, J.J. (1977). Juridic decisions: In ignorance of the law or in light of it?, *Law and Human Behavior* **1**, 163–190.
[33] Severance, L.J., Greene, E. & Loftus, E.F. (1984). Toward criminal jury instructions that jurors can understand, *Journal of Criminal Law and Criminology* **75**, 198–233.

EDIE GREENE AND CHRISTINA STUDEBAKER

Jury Instructions on Expert Testimony

In trials conducted in adversary system jurisdictions (*see* **Adversary Systems of Justice**) where the fact finder is a jury composed of ordinary citizens, such as in the United States, Canada, the United Kingdom, and other commonwealth countries, the jurors will "deliberate" and then render their verdict after they have heard and seen all of the evidence that was presented by both sides in the case. Before the jurors commence their deliberation, however, the judge must "instruct" the jurors in the law that applies to the issues in the case. These instructions deal with burdens of proof and a number of other legal concepts. Among these concepts are the rules by which jurors are to judge evidence that may be in conflict, or which they have difficulty in accepting as true.

In most jurisdictions, the jurors will be told that they must determine what happened. In doing so, their role is to judge the credibility of all the evidence they have seen and heard. They will further be told that they may accept, in whole or in part, each item of evidence presented to them, or they may reject such evidence, in whole or in part, if they find it to be unconvincing or not worthy of belief. One of these instructions also deals with how jurors should evaluate expert witness testimony.

The instruction on the weight of the evidence is particularly important when opinion testimony is in conflict. While jurisdictions may differ on how such instructions are worded, a typical instruction on the evaluation of expert testimony will be as follows:

> You have heard the testimony of experts in this case. The credibility or worth of the testimony of an expert witness is to be considered by you just as it is your duty to judge the credibility or worth of the testimony of all other witnesses you have heard or evidence you have seen. You are not bound to accept expert testimony as true if you find it unconvincing, and you may weigh and credit testimony of expert witnesses the same as that of other witnesses, and give it the weight to which you think it is entitled.[a]

It is also permissible for the judge to supplement the standard expert witness jury instruction by a special provision more applicable to a particular case. In many jurisdictions, in charging the jury, the judge

may not refer to the testimony of any particular witness and may not single out certain testimony or evidence. In other jurisdictions, however, these restrictions do not apply and a judge may single out expert witnesses or evidence that needs to be specifically addressed by the jury.

Forensic experts will note that their opinion testimony is not entitled to a greater weight than that given to ordinary fact witnesses merely because they are experts. Jurors are permitted to ignore expert opinions that they do not believe or testimony that is unconvincingly presented. Jurors are likely to ignore opinion testimony they do not understand or where the witness does not explain scientific principles upon which an opinion is based in a credible manner. Other factors that influence jurors in how much weight to give to an expert's testimony are as follows: the qualifications of the witness; the appearance and presentation of the expert; possible bias or personal interest of the expert, his facility of expression and clarity of thought, as well as the comportment during cross-examination. Most jurors attach great persuasive weight to these factors in their credibility judgment of all testimony, including that of forensic experts.

While jurors are free to attach such weight to expert forensic testimony as they feel it deserves, they cannot go beyond the evidence seen and heard in the courtroom. Thus, it would be improper for jurors to perform their own investigations. The discovery, at any time during the trial, that jurors conducted experiments on their own, or, during a recess, went to the crime scene without being required to do so by the trial judge, is certain to result in a mistrial being declared. That is why judges typically instruct juries to use *only the evidence heard in court* to inform their deliberations.

End Notes

[a.] Adapted from Pattern Jury Instructions approved by several jurisdictions.

ANDRE MOENSSENS

Jury Nullification ***see*** Behavioral Science Evidence

Justice Systems: Adversary ***see*** Adversary Systems of Justice

Justice Systems: Civil Law Countries ***see*** Civil Law Systems of Justice

Juvenile: Neuropsychological Assessment ***see*** Neuropsychological Assessment: Child

Juvenile: Psychiatric Research ***see*** Northwest Juvenile Project

Juvenile Adjudicative Competence ***see*** Children: as Defendants

Juvenile Death Penalty ***see*** Death Penalty and Age

Juvenile Justice: Adolescent Development

In 2005, in a landmark decision, the US Supreme Court outlawed the death penalty for offenders who were younger than 18 when they committed their crimes (*see also* **Death Penalty and Age**). The ruling

centered on the issue of culpability or criminal blameworthiness. Unlike competence, which concerns an individual's ability to serve as a defendant during trial or adjudication, culpability turns on the offender's state of mind at the time of the offense, including factors that would mitigate, or lessen, the degree of responsibility.

The Court's ruling ran counter to a nationwide trend toward harsher sentences for juveniles. Over the preceding decade, as serious crime rose and public safety became a focus of concern, legislators in virtually every state had enacted laws that lower the age at which juveniles could be tried and punished as adults for a broad range of crimes. This and other changes have resulted in the trial of more than 200 000 youths in the adult criminal system each year [1].

Proponents of the tougher laws argue that youths who have committed violent crimes need more than a slap on the wrist from a juvenile court. It is naive, they say, to continue to rely on a juvenile system designed for a simpler era, when youths were getting into fistfights in the schoolyard; drugs, guns, and other serious crimes are adult offenses that demand adult punishment. Yet the premise of the juvenile justice system is that adolescents are different from adults, in ways that make them potentially less blameworthy than adults for their criminal acts.

The legal system has long held that criminal punishment should be based not only on the harm caused, but also on the blameworthiness of the offender. How blameworthy a person is for a crime depends on the circumstances of the crime and of the person committing it. Traditionally, the courts have considered several categories of mitigating factors when determining a defendant's culpability. These include the following:

- impaired decision-making capacity, usually due to mental illness or disability;
- the circumstances of the crime – for example, whether it was committed under duress; and
- the individual's personal character, which may suggest a low risk of continuing crime.

Such factors do not make a person exempt from punishment – rather, they indicate that the punishment should be less than it would be for others committing similar crimes, but under different circumstances.

Should developmental immaturity be added to the list of mitigating factors? Should juveniles, in general, be treated more leniently than adults? A major study by the MacArthur Foundation Research Network on Adolescent Development and Juvenile Justice now provides strong evidence that the answer is yes.

The MacArthur Study of Juvenile Culpability

The study of juvenile culpability was designed to provide scientific data on whether, in what ways, and at what ages adolescents differ from adults.

Many studies have shown that by the age of 16, adolescents' cognitive abilities – loosely, their intelligence or ability to reason – closely mirrors that of adults. But how people reason is only one influence on how they make decisions. In the real world, especially in high-pressure crime situations, judgments are made in the heat of the moment, often in the company of peers. In these situations, adolescents' other common traits – their shortsightedness, their impulsivity, their susceptibility to peer influence – can quickly undermine their decision-making capacity.

The investigators examined the age differences in a number of characteristics that are believed to undergird decision making and that are relevant to mitigation, such as impulsivity and risk processing, future orientation, sensation-seeking, and resistance to peer pressure. These characteristics are also thought to change over the course of adolescence and to be linked to brain maturation during this time. The subjects – close to 1000 individuals between the ages of 10 and 30 – were drawn from the general population in five regions. They were ethnically and socioeconomically diverse.

The study's findings showed several characteristics of adolescence that are relevant to determinations of criminal culpability. As the accompanying figure indicates, although intellectual abilities stop maturing around age 16, psychosocial capability continues to develop well into early adulthood (Figure 1).

Short-Sighted Decision Making

One important element of mature decision making is a sense of the future consequences of an act. A variety of studies in which adolescents and adults are asked

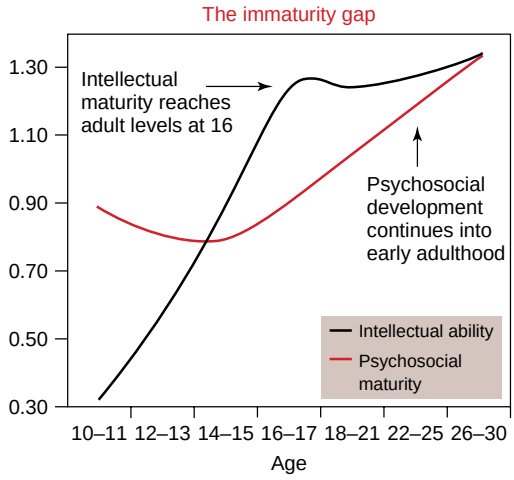


Figure 1 Developmental course of intellectual and psychosocial development

to envision themselves in the future have found that adults project their visions over a significantly longer time, suggesting much greater future orientation.

These findings are supported by data from the network's culpability study. Adolescents characterized themselves as less likely to consider the future consequences of their actions than did adults. And when subjects in the study were presented with various choices measuring their preference for smaller, immediate rewards versus larger, longer-term rewards (e.g., "Would you rather have $100 today or $1000 a year from now?"), adolescents had a lower "tipping point" – the amount of money they would take to get it immediately as opposed to waiting.

How might these characteristics carry over into the real world? When weighing the long-term consequences of a crime, adolescents may simply be unable to see far enough into the future to make a good decision. Their lack of foresight, along with their tendency to pay more attention to immediate gratification than to long-term consequences, is among the factors that may lead them to make bad decisions.

Poor Impulse Control

The network's study also found that as individual's age, they become less impulsive and less likely to seek thrills; in fact, gains in these aspects of self-control continue well into early adulthood. This was evident in individuals' descriptions of themselves and on tasks designed to measure impulse control. On the "Tower of London" task, for example – where the goal is to solve a puzzle in as few moves as possible, with a wrong move requiring extra moves to undo it – adolescents took less time to consider their first move, jumping the gun before planning ahead.

Network research also suggests that adolescents are less sensitive to risk and more sensitive to rewards – an attitude than can lead to greater risk taking. The new data confirm and expand on earlier studies gauging attitudes toward risk, which found that adults spontaneously mention more potential risks than teens. Juveniles' tendency to pay more attention to the potential benefits of a risky decision than to its likely costs may contribute to their impulsivity in crime situations.

Vulnerability to Peer Pressure

The law does not require exceptional bravery of citizens in the face of threats or other duress. A person who robs a bank with a gun in his back is not as blameworthy as another who willingly robs a bank; coercion and distress are mitigating factors. Adolescents, too, face coercion, but of a different sort.

Pressure from peers is keenly felt by teens. Peer influence can affect youths' decisions directly, as when adolescents are coerced to take risks they might otherwise avoid. More indirectly, youths' desire for peer approval or their fear of rejection may lead them to do things they might not otherwise do. In the network's culpability study, individuals' reports of their vulnerability to peer pressure declined over the course of adolescence and young adulthood. Other network research now underway is examining how adolescent risk taking is "activated" by the presence of peers or by emotional arousal. For example, an earlier network study, involving a computer car-driving task, showed that the mere presence of friends increased risk taking in adolescents and college undergraduates, though not adults [1].

Although not every teen succumbs to peer pressures, some youths face more coercive situations than others. Many of those in the juvenile justice system live in tough neighborhoods, where losing face can be not only humiliating but also dangerous. Capitulating in the face of a challenge can be a sign of weakness, inviting attack, and continued persecution. To the extent that coercion or duress is a mitigating

factor, the situations in which many juvenile crimes are committed should lessen their culpability.

Confirmation from Brain Studies

Recent findings from neuroscience line up well with the network's psychosocial research, showing that brain maturation is a process that continues through adolescence and into early adulthood. For example, there is good evidence that the brain systems that govern impulse control, planning, and thinking ahead are still developing well beyond age 18. There are also several studies indicating that the systems governing reward sensitivity are "amped up" at puberty, which would lead to an increase in sensation-seeking and in valuing benefits over risks. And there is emerging evidence that the brain systems that govern the processing of emotional and social information are affected by the hormonal changes of puberty in ways that make people more sensitive to the reactions of those around them – and thus more susceptible to the influence of peers [2].

Policy Implications: A Separate System for Young Offenders

The scientific arguments do not say that adolescents cannot distinguish right from wrong, or that they should be exempt from punishment. Rather, they point to the need to consider the developmental stage of adolescence as a mitigating factor when juveniles are facing criminal prosecution. The same factors that make youths ineligible to vote or to serve on a jury require us to treat them differently from adults when they commit crimes.

Some have argued that courts ought to assess defendants' maturity on a case-by-case basis, pointing to the fact that older adolescents, in particular, vary in their capacity for mature decision making. But the tools needed to measure psychosocial maturity on an individual basis are not well developed, or it is possible to distinguish reliably between mature and immature adolescents on the basis of brain images. Consequently, assessing maturity on an individual basis, as we do with other mitigating factors, is likely to produce many errors. However, the maturing process follows a similar pattern across virtually all teenagers.

Therefore, it is both logical and efficient to treat adolescents as a special legal category – and to refer the vast majority of offenders under the age of 18 to juvenile court, where they will be treated as responsible but less blameworthy, and where they will receive less punishment and more rehabilitation and treatment than typical adult offenders. The juvenile system does not excuse youths of their crimes; rather, it acknowledges the development stage and its role in the crimes committed, and punishes appropriately.

At the same time, any legal regime must pay attention to legitimate concerns about public safety. There will always be some youths – such as older, violent recidivists – who have exhausted the resources and patience of the juvenile justice system, and whose danger to the community warrants adjudication in criminal court. But these represent only a very small percentage of juvenile offenders. Trying and punishing youths as adults is an option that should be used sparingly.

Legislatures in several states have begun to reconsider the punitive laws enacted in recent decades. They have already recognized that prosecuting and punishing juveniles as adults carries high costs, for the youths and for their communities. Now we can offer lawmakers in all states a large body of research on which to build a more just and effective juvenile justice system.

Acknowledgment

This article was provided in cooperation with the MacArthur Foundation Research Network on Adolescent Development, and Juvenile Justice, an interdisciplinary, multi-institutional program focused on building a foundation of sound science and legal scholarship to support reform of the juvenile justice system. The network conducts research, disseminates the resulting knowledge to professionals and the public, and works to improve decision making and to prepare the way for the next generation of juvenile justice reform. For more information, contact the MacArthur Foundation Research Network on Adolescent Development and Juvenile Justice, Temple University, Department of Psychology, Philadelphia, PA 19122, USA, http://www.adjj.org.

References

[1] Allard, P. & Young, M. (2002). Prosecuting juveniles in adult court: perspectives for policymakers and practitioners, *Journal of Forensic Psychology Practice* **6**, 65–78.

[2] Gardner, M. & Steinberg, L. (2005). Peer influence on risk-taking, risk preference, and risky decision-making in adolescence and adulthood: an experimental study, *Developmental Psychology* **41**, 625–635.

Further Reading

Nelson, E., Leibenluft, E., McClure, E. & Pine, D. (2005). The social reorientation of adolescence: a neuroscience perspective on the process and its relation to psychopathology, *Psychological Medicine* **35**, 163–174.

Related Articles

Children: as Defendants

LAURENCE STEINBERG

Juvenile Justice: Mental Illness

see Children: as Defendants

Juvenile Justice: Parental Rights

see Parental Rights and Prerogatives

Juvenile Justice: Transfer to Adult

Through much of the twentieth century, the juvenile court was the primary legal forum to respond to children who broke the criminal laws. With the rise in youth crime beginning in the late 1970s, legislators and commentators spoke ominously of a nation under siege by a rising generation of violent young criminals. These fears led many Americans to blame the juvenile court and demand that legislators "get tough" with violent and chronic young offenders.

In response to recurring epidemics of youth violence over the past three decades, 46 states made significant changes in laws that lowered the age and broadened the circumstances under which young defendants could be prosecuted in the criminal courts. Prosecution in the criminal court was designed to punish young offenders more harshly and for longer periods of time, thereby deterring them and other youths from further crimes.

But have these efforts been effective? Does the prospect of harsher sentences and adult time deter youth from committing crimes? Although there are strong proponents on each side of the argument, new evidence has raised questions about the effectiveness of the new laws.

Several studies have examined whether the prosecution of adolescents as adults reduces crime and recidivism. Two studies capitalized on unique conditions in the New York Metropolitan area, where the laws of two states, New York and New Jersey, span the border of a single metropolitan area. On the New York side of the border, juveniles as young as 13 are charged in adult court, while on the New Jersey side, nearly all cases of juvenile offenders below the age of 18 are processed in juvenile court. By comparing similar offenders in the two settings who were arrested and charged with the same felony offenses during the same time period, the researchers were able to determine whether treating juveniles as adults in the legal system is an effective deterrent to crime.

They find that adolescents processed in the New York adult courts were more likely to be rearrested. They were rearrested more often and more quickly and for more serious offenses, and they were reincarcerated at higher rates than those in the New Jersey juvenile courts. The results suggest that harsher sentences and adult punishments are ineffective deterrents to crime among the juveniles in this sample.

Teens Prosecuted in Adult Courts at Greater Risk of Repeat Offenses

The study examined more than 2000 adolescents who committed one of three types of serious crimes (aggravated assault, armed robbery, burglary) during 1992 and 1993. The youth were tracked through 1999 to determine rearrest rates for several types

of crimes. By using the two groups from the same metropolitan area, with similar economic opportunity, access to weapons, drug use, gang influences, and other influences on crime, any differences in rearrest between the two groups can be assumed to be due to the different court systems. The rearrest rates were calculated after controlling for time on the street.

Table 1 shows that youth prosecuted in the adult courts in New York were 85% more likely to be rearrested for violent crimes than those prosecuted in the New Jersey juvenile courts, and 44% more likely to be rearrested for felony of property crimes. The odds of rearrest were greatest for those youths with no prior arrest record who were prosecuted and sentenced as adults. Only for one type of crime, drug offenses, were youths in the adult courts less likely to be rearrested. The chances of being reincarcerated were 26% greater for youths prosecuted as adults. When the researchers compared the number of each type of offense during the follow-up period, the results were nearly identical.

Youths who received lighter sanctions – those whose cases were either dismissed or who received lighter sentences – also were less likely to be rearrested; this was true in both states. In other words, teens whose cases are diverted from court or dismissed are less likely to be arrested again. More work is needed to determine whether this stems from the courts' ability to identify those youth at greater risk for reoffending and to give them a sanction, or if the mere fact of a sanction causes an adolescent to feel more like a criminal and then act more like one after release.

The research also showed that longer sentences did not reduce the likelihood of rearrest either in the juvenile or the adult court. But, the research did show that a history of prior arrests and rearrests is a reliable predictor of future rearrests. So too are several demographic factors: males are more likely to be rearrested than females, and African-Americans are more likely to be rearrested than other race-ethnicities. More study is needed to determine whether these differences stem from different behaviors of the individuals in the groups or from different arrest policies.

Table 1 Odds of rearrest and reincarceration for adolescents in adult court relative to the juvenile system[(a)]

Recidivism measure	Statistical significance	Odds ratio relative to juvenile system
Any rearrest	ns	–
Violence	0.00	0.85
Property	0.00	0.44
Weapon	Ns	–
Drug	0.00	0.65
Reincarceration	0.05	0.6

[(a)] Log odds of 0.0 indicates no difference

Teens in Adult Corrections Face Harsher Settings and Experience More Developmental Problems

In a related study, the same research team, network researchers compared the correctional experiences of 425 adolescents placed in juvenile versus adult correctional facilities in 2000–2001. This research sought clues that might explain why adolescents adjudicated and sentenced in the criminal courts often have higher rearrest rates and are more often returned to jail or prison. The research used a similar design in which youth in juvenile corrections were compared with matched samples of youths in nearby states where they were incarcerated as adults. The incarcerated youths were interviewed within three months of their scheduled release date and asked about their correctional experiences, the therapeutic and rehabilitative services they received, and their mental health and social outcomes. Four states were included in the study, each with varying programs and facilities where teenage offenders were incarcerated. The experiences of youth in juvenile correctional facilities in New Jersey and California were compared with those of similar groups of youth placed in adult correctional facilities in New York and Arizona.

The results suggest clear differences in the therapeutic and service contexts of each of these settings. Figure 1 shows that youths placed in adult correctional settings reported significantly weaker correctional climates along four critical dimensions: fairness, counseling and therapeutic services, educational and job training services, and program structure, compared with matched groups of youths placed in juvenile facilities. At the same time, the juvenile facilities were more chaotic. Adolescents in the juvenile programs reported higher rates of witnessing violence and violent victimization. They also reported higher rates of involvement in several types of crimes

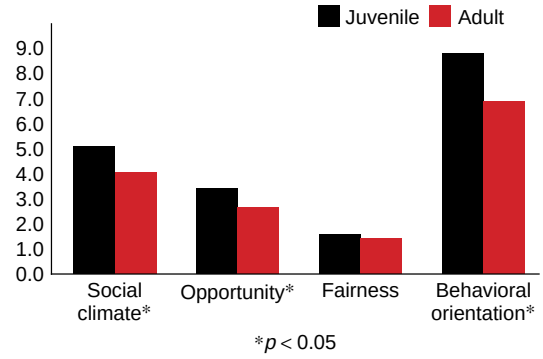


Figure 1 Institutional climate by correctional system

while incarcerated as well as more drug use. Despite these unruly settings, they reported greater feelings of safety compared with youths placed in adult settings. This paradox may reflect the social networks that were dominant in the two different types of placements, older criminal offenders in more organized prison gangs were the dominant social group in the adult facilities, compared to the loosely organized groups of peers that populated the juvenile facilities.

This greater sense of danger, then, perhaps explains the higher rates of mental health problems reported by youths in the adult facilities. Figure 2 shows that the current levels of mental health symptoms of youths in adults corrections were significantly worse on two dimensions of mental health functioning compared with rates reported by youths in juvenile facilities. The significant dimension includes the important Global Severity Index, a scale that spans all of the dimensions of mental health in this assessment tool.

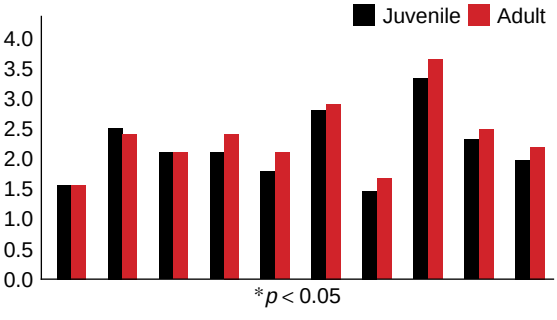


Figure 2 Current mental health functioning by correctional system

The same youths also reported higher rates of three dimensions of posttraumatic stress disorder. Figure 3 shows that youths in adult corrections had higher rates of the same types of mental health problems experienced by soldiers returning from war and survivors of natural disasters.

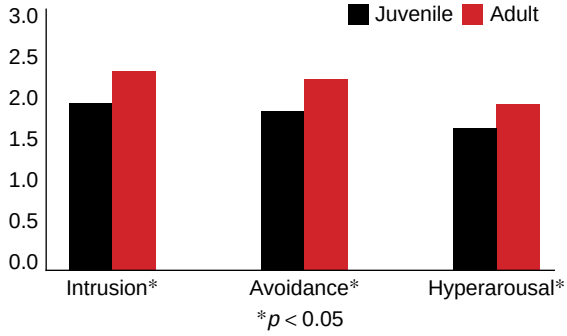


Figure 3 posttraumatic stress disorder (PTSD) symptoms by correctional system

There were also differences across the juvenile corrections facilities. Youths placed in larger juvenile justice facilities with a wider age range of inmates had outcomes similar to that of youths placed in the adult facilities. Youths sentenced as adults who spent some time in juvenile facilities before being administratively transferred to adult placements experienced fewer mental health problems and reported better service environments. These differences suggest that the size and diversity of populations in correctional placements are important dimensions of the correctional experience that interacts with the broader adult–juvenile legal categories to shape the correctional experiences of youths punished as adults.

Policy Implications

Policymakers and others advocating for harsher youth sentences argue that the threat of "adult time for adult crime" is a sound deterrent. The first study, however, shows the opposite: recidivism among 15- and 16-year olds prosecuted in adult courts in New York is actually more common and more serious than it is in New Jersey, which refers adolescents who have committed similar offenses to juvenile courts. Youths sentenced in the criminal courts in New York

were more likely to be rearrested, their rearrests were more frequent and their new offenses more serious, and they were more likely to be reincarcerated within a few years.

The study also suggests that the longer sentences in adult courts are not responsible for the differences in rearrest rates or in correctional outcomes. Rather, the second study finds that the adult courts may expose adolescents to harsher incarceration settings and less effective probation supervision in the criminal justice system. One reason may be that a felony conviction has a more harmful effect on subsequent employment, citizenship, or other positive adult roles, factors that otherwise could lessen the tendency to return to crime. Another possibility is that prosecution in an adult court communicates to the adolescent that he or she is unsalvageable, and hence repeat offenses become a self-fulfilling prophecy. A third reason is the stark differences in correctional experiences for those youths who are incarcerated as adults. Not only do they receive fewer and weaker services but also they are confined with adult offenders during the critical developmental period of the transition from adolescence to adulthood. This environment obviously has its effects on mental health. However, also, teens in adult corrections have limited exposure during this critical developmental stage to a broader set of social norms and a more diverse behavioral toolkit from the wider social networks of family, school or work, and community. Network researchers are not the only ones to reach these conclusions. Studies in Florida, for example, show similar elevated risks of rearrest for juveniles in adult court, and similar toxic environments for those youths placed in adult correctional facilities.

Authority for transfer decisions should be returned to juvenile court judges who can consider individualized criteria other than age and offense in determining how to prosecute an adolescent. Policies that result in a wholesale transfer of adolescents from juvenile to adult courts often fail to deter repeated instances of serious and violent crime. Although some of the most extreme cases may still need to be prosecuted in adult court, these should be the exception and not the rule. However, return of decision-making authority to judges must be accompanied by new models for decision making. In the past, judges have not been able to consistently identify the most serious offenders, and there has been a tendency toward harsher sentencing that often reflected racial discrimination. If new models are not offered to judges, the same problems recur. What is needed is a better method to distinguish between cases in which the community must be protected from predatory youth and those in which delinquent youth must be protected from negative effects of incarceration.

Acknowledgment

This research was supported by the MacArthur Research Network on Adolescent Development and Juvenile Justice, an interdisciplinary, multi-institutional program focused on building a foundation of sound science and legal scholarship to support reform of the juvenile justice system.

Further Reading

Alpert, G. MacDonald, J. & Dunham, R. (2005). Police suspicion and discretionary decision making during citizen stops, *Criminology* **43**, 407–434.

Banks, R.R. Eberhardt, J.L. & Ross, L. (2006). Discrimination and implicit bias in a racially unequal society, *California Law Review* **94**, 1169–1190.

Barnowski, R. (2003). *Changes in Washington State's Jurisdiction of Juvenile Offenders: Examining the Impact'*, available at www.wsipp.wa.gov/rptfiles/JuvJurisChange.pdf.

Barry, C.F. & Marcy, R.P. (1996). The end of the line: an empirical study of judicial waiver, *Journal of Criminal Law and Criminology* **86**(2), 449–492.

Barry, C.F. & Marcy, R.P. (2001). The back-door to prison: waiver reform, 'Blended Sentencing,' and the law of unintended consequences, *Journal of Criminal Law and Criminology* **91**(4), 997–1072.

Beckett, K. (1997). *Making Crime Pay: Law and Order in Contemporary American Politics*, Oxford University Press.

Bennett, W.J., DiIulio Jr, J. & Walters, J.P. (1996). *Body Count: Moral Poverty and How to Win America's War against Crime and Drugs*, Simon and Schuster, New York.

Bishop, D. *Juvenile Offenders in the Adult Criminal System* (see 9).

Bishop, D. (2000). Juvenile offenders in the adult criminal system, *Crime and Justice: A Review of Research* **27**, 81–167.

Bishop, D.M. (2005). The role of race and ethnicity in juvenile justice processing, in *Our Children, Their Children: Confronting Racial and Ethnic Differences in American Juvenile Justice*, D. Hawkins & K. Kempf-Leonard, eds, University of Chicago Press.

Bishop, D.M. and others (1996). The transfer of juveniles to criminal court: does it make a difference? *Crime and Delinquency* **42**, 171–191.

Bobo, L. & Johnson, D. (2004). A taste for punishment: black and white Americans' views on the death penalty and the war on drugs, *Du Bois Review* **1**, 151–180.

Brannen, D.N. and others *Transfer to Adult Court: A National Study of How Juvenile Court Judges Weigh Pertinent Kent Criteria, Psychology, Public Policy, and Law* (forthcoming).

Bridges, G. & Steen, S. (1998). Racial disparities in official assessments of juvenile offenders: attributional stereotypes as mediating mechanisms, *American Sociological Review* **63**, 554–570.

Brink, D.O. (2004). Immaturity, normative competence, and juvenile transfer: how (not) to punish minors for major crimes, *Texas Law Review* **82**, 1555–1585.

Butterfield, F. (1995). *All God's Children: The Bosket Family and the American Tradition of Violence*, Harper Perennial, New York.

Cauffman, E. & Steinberg, L.D. (2001). (Im) Maturity of judgment in adolescence: why adolescents may be less culpable than adults, *Behavioral Sciences and the Law* **18**, 741–760.

Champion, D.J. (1989). Teenage felons and waiver hearings: some recent trends, 1980–1988, *Crime and Delinquency* **35**, 577–585.

Correll, J. and others (2002). The police officer's dilemma: using ethnicity to disambiguate potentially threatening individuals, *Journal of Personality and Social Psychology* **83**, 1314–1329.

Cranor, C.F. (1993). *Regulating Toxic Substances: A Philosophy of Science and the Law*, Oxford University Press.

Eberhardt, J. and others (2006). Looking deathworthy: perceived stereotypicality of black defendants predicts capital-sentencing outcomes, *Psychological Science* **17**, 383–387.

Emens, E.F. (2005). Aggravating youth: *Roper v. Simmons* and age discrimination, *The Supreme Court Review* 51–102.

Fagan, J. (1995). Separating the men from the boys: the comparative impacts of juvenile and criminal court sanctions on recidivism of adolescent felony offenders, in *Sourcebook: Serious, Violent, and Chronic Juvenile Offenders*, J.C. Howell, and others Sage, Thousand Oaks, CA.

Fagan J. & Davies G. (2000). Street stops and broken windows: race, *terry* and disorder in New York city, *The Fordham Urban Law Journal* **28**, 457.

Fagan J., Kupchik, A. & Liberman A., "Be Careful What You Wish for: Legal Sanctions and Public Safety among Adolescent Offenders in Juvenile and Criminal Court," *Columbia Law School, Public Law Research Paper* no. 03-61 (available at SSRN: http://ssrn.com/abstract = 491202 [July, 2007]).

Fagan, J. & West, V. (2005). The decline of the juvenile death penalty: scientific evidence of evolving norms, *Journal of Criminal Law and Criminology* **95**, 427.

Fagan, J. & Zimring, F.E. (eds) (2000). *Changing Borders of Juvenile Justice: Transfer of Adolescents to the Criminal Court*, University of Chicago Press.

Feld, B. (1988). The juvenile court meets the principle of the offense: punishment, treatment, and the difference that it makes, *Boston University Law Review* **69**, 821–952.

Feld, B.C. (1995). Violent youth and public policy: a case study of juvenile justice law reform, *Minnesota Law Review* **79**, 965–1128.

Feld, B.C. (1997). Abolish the juvenile court: youthfulness, criminal responsibility, and sentencing policy, *Journal of Criminal Law and Criminology* **88**, 68.

Feld, B.C. (1999). *Bad Kids: Race and the Transformation of the Juvenile Court*, Oxford University Press.

Forst, M.A. Fagan, J. & Vivona, T.S. (1989). Some paradoxical effects of the treatment-custody dichotomy for adolescents in adult prisons, *Juvenile and Family Court Journal* **40**, 1–15.

Freeman, R. (1996). Why do so many young American men commit crimes and what might we do about it? *The Journal of Economic Perspectives* **10**(1), 25–42.

Garland, D. (1990). *Punishment and Welfare*, University of Chicago Press.

Gelman, A. Fagan, J. & Kiss, A. (2007). An analysis of the New York city police department's 'stop-and-frisk' policy in the context of claims of racial bias, *Journal of the American Statistical Association* **102**, 813–823.

Gillespie, L.K. & Norman, M.D. (1984). Does certification mean prison? Some preliminary findings from Utah, *Juvenile and Family Court Journal* **35**, 23–34.

Glassner, B. and others (1983). A note on the deterrent effect of juvenile vs. Adult jurisdiction, *Social Problems* **31**(2), 219–221.

Graham, S. & Lowery, B.S. (2004). Priming unconscious racial stereotypes about adolescent offenders, *Law and Human Behavior* **28**, 483–504.

Greenwald, A.G. and others (2003). Targets of discrimination: effects of race on responses to weapons holders, *Journal of Experimental Social Psychology* **39**, 399–405.

Greenwood, P.W. Abrahamse, A. & Zimring, F. (1984). *Factors Affecting Sentencing Severity for Young Adult Offenders*, Rand Corporation, Santa Monica, CA.

Gruber J. (eds) (2001). *Risky Behavior among Youth: An Economic Analysis*, University of Chicago Press.

Hagan, J.C. Bumiller, K. (1983). *Making* sense of sentencing: A review and critique of sentencing research, in *Research on Sentencing: The Search for Reform*, A. Blumstein, eds, Vol. 2 and others National Academy Press, Washington, DC, pp. 1–54.

Houghtalin, M. & Mays, G.L. (1991). Criminal dispositions of new Mexico juveniles transferred to adult court, *Crime and Delinquency* **37**, 393–407.

House Committee on the Judiciary, *Putting Consequences Back into Juvenile Justice at the Federal, State, and Local Levels: Hearings before the Subcommittee on Crime*, 106th Congress, 1999.

Jensen, E.L. & Metsger, L.K. (1994). A test of the deterrent effect of legislative waiver on violent juvenile crime, *Crime and Delinquency* **40**(1), 96.

Kupchik, A., Fagan, J. & Liberman, A. (2003). Punishment, proportionality, and jurisdictional transfer of adolescent offenders: a test of the leniency gap hypotheses, *Stanford Law and Policy Review* **14**, 57–83.

Langan P. & Levin D. U.S. Department of Justice, Bureau of Justice Statistics (2002). "Special Report," *Recidivism of*

Prisoners Released in 1994 (www.ojp.usdoj.gov/bjs/pub/pdf/rpr94.pdf [June, 2002]).

Lanza-Kaduce, L. and others Florida Department of Juvenile Justice (2002). *Juvenile Transfer to Criminal Court Study: Final Report* (www.prisonpolicy.org/scans/juveniletransfers.pdf [January, 2002]).

Laval, S. Miller-Wilson and Patricia Puritz, *Pennsylvania: An Assessment of Access to Counsel and Quality of Representation in Delinquency Proceedings* (www.jlc.org/File/publications/paassessment.pdf [October 2003]).

Lauritsen, J. Racial and ethnic differences in juvenile offending, in *Our Children, Their Children*, D. Hawkins & K. Kempf-Leonard, eds.

Lee, D. & McCrary, J. (2005). *Crime, Punishment, and Myopia*, Working Paper W11491 National Bureau of Economic Research, Cambridge, available at http://ssrn.com/abstract = 762770.

Levitt, S.D. (1998). Juvenile crime and punishment, *The Journal of Political Economy* **106**, 1156–1185.

Lyons, D. National Conference of State Legislatures, "State Legislature Report," *1995 Juvenile Crime and Justice State Enactments* 20, no.17 (November 1995).

McGowan, A. and others (2007). Effects on violence of laws and policies facilitating the transfer of juveniles from the juvenile justice system to the adult justice system: a report on recommendations of the task force on community preventive services. *Morbidity and Mortality Weekly Report* **56**(RR-9), 1–11.

Morse, S.J. (1998). Immaturity and responsibility, *Journal of Criminal Law and Criminology* **88**, 15.

Pizzi, W. Blair, I.V. & Judd, C.M. (2005). Discrimination in sentencing on the basis of afrocentric features, *Michigan Journal of Race and Law* **10**, 327–355.

Plant, E.A. & Peruche, B.M. (2005). The consequences of race for police officers' responses to criminal suspects, *Psychological Science* **16**, 180–183.

Plant, E.A. and others, (2005). Eliminating automatic racial bias: making race non-diagnostic for responses to criminal suspects, *Journal of Experimental Social Psychology* **41**, 141–156.

Platt, A. (1967). *Child Savers: The Invention of Delinquency*, University of Chicago Press.

Redding, R.E. (2003). The effects of adjudicating and sentencing juveniles as adults, *Youth Violence and Juvenile Justice* **1**(2), 128–155.

Reppucci, N.D. (1999). Adolescent development and juvenile justice, *American Journal of Community Psychology* **27**(3), 307–326.

Richard, E.R. & James, C.H., Blended sentencing in American juvenile courts, in *Changing Borders of Juvenile Justice*.

Robinson, P.H. & Darley, J.M. (2004). Does criminal law deter? A behavioural science investigation, *Oxford Journal of Criminal Law* **24**, 173–205.

Rosenbaum, P.R. & Rubin, D.B. (1984). Reducing bias in observational studies using subclassification on the propensity score, *Journal of the American Statistical Association* **79**, 516–524.

Roysher, M. & Edelman, P. *Treating Juveniles as Adults*, (see 11).

Roysher, M. & Edelman, P. (1981). "Treating juveniles as adults in New York: what does it mean and how is it working?" in *Major Issues in Juvenile Justice Information and Training*, J.C. Hall, eds, and others Academy for Contemporary Problems, Columbus, OH.

Rudman, C.S. Fagan, J. & Hartstone, E.C. (1986). Violent youth in adult court: process and punishment, *Crime and Delinquency* **32**, 75–96.

Salekin, R. and others (2002). Juvenile transfer to adult courts: a look at prototypes for dangerousness, sophistication-maturity, and amenability to treatment through a legal lens, *Psychology, Public Policy, and Law* **8**(4), 373–410.

Sampson, R.J. & Laub, J.H. (1993). *Crime in the Making: Pathways and Turning Points through Life*, Harvard University Press.

Sealander, J. (2003). *The Failed Century of the Child: Governing America's Young in the Twentieth Century*, Cambridge University Press.

Sickmund, M. Office of Juvenile Justice and Delinquency Prevention, "OJJDP Update on Statistics," *How Juveniles Get to Criminal Court* (www.ncjrs.gov/pdffiles/juvcr.pdf [October, 1994]).

Simon, J. (2007). *Governing through Crime*, Oxford University Press.

Singer, S. (1996). *Recriminalizing Delinquency: Violent Juvenile Crime and Juvenile Justice Reform*, Cambridge University Press.

Singer, S.I. & McDowall, D. (1988). Criminalizing delinquency: the deterrent effects of the New York juvenile offender law, *Law and Society Review* **22**(3), 521–536.

Snyder, H. & Sickmund, M. *Juvenile Offenders and Victims* (see 44), p. 236.

Sobie, M. *The Juvenile Offender Act*, (see 10).

Sobie, M. (1981). The juvenile offender act: effectiveness and impact on the New York juvenile justice system, *New York Law School Law Review* **27**, 677–691.

Spelman, W. (2000). The limited importance of prison expansion, in *The Crime Drop in America*, A. Blumstein & J. Wallman, eds, Cambridge University Press.

Steinberg, L. & Scott, E. (2003). Less guilt by reason of adolescence, *The American Psychologist* **58**(12), 1009–1018.

Strom, K. and others U.S. Department of Justice, Bureau of Justice Statistics (1998). "Special Report," *State Court Processing Statutes 1990-94, Juvenile Felony Defendants in Criminal Courts* (www.ojp.gov/bjs/pub/pdf/jfdcc.pdf [September, 1998]).

Strom, K. U.S. Department of Justice, Bureau of Justice Statistics (2000). "Special Report," *Profile of State Prisoners under Age 18, 1985–97* (http://ojp.usdoj.gov/bjs/pub/pdf/pspa1897.pdf [February, 2000]).

Tanenhaus, D. (2004). *Juvenile Justice in the Making*, Oxford University Press.

Thomas, C.W. & Bilchik, S. (1985). Prosecuting juveniles in criminal courts: a legal and empirical analysis *Journal of Criminal Law and Criminology* **76**, 439–479.

Tittle, C.R. (1980). *Sanctions and Social Deviance: The Question of Deterrence*, Praeger, New York.

Tittle, C.R. (1980). Evaluating the deterrent effects of criminal sanctions, in *Handbook of Criminal Justice Evaluation*, M. Klein & K. Teilmann, eds, Sage Publications, Beverly Hills.

Tonry, M. (1996). *Malign Neglect: Race, Crime, and Punishment in America*, University of Chicago Press.

Tonry, M. (2004). *Thinking about Crime: Sense and Sensibility in American Penal Culture*, Oxford University Press.

Torbet, P. and others Office of Juvenile Justice and Delinquency Prevention, *State Responses to Serious and Violent Juvenile Crime*, p. 3 (www.ncjrs.gov/pdffiles/statresp.pdf [July, 1996]).

Winner, L. and others (1997). The transfer of juveniles to criminal court: reexamining recidivism over the long term, *Crime and Delinquency* **43**, 548–563.

Zimring, F. (2000). Penal proportionality for the young offender: notes on immaturity, capacity, and diminished responsibility, in *Youth on Trial: A Developmental Perspective on Juvenile Justice*, R. Schwartz & T. Grisso, eds, University of Chicago Press.

JEFFREY FAGAN

Juveniles and Criminal Responsibility *see* Children: as Defendants



Kidnapping: Victim Behavior *see* Stockholm Syndrome

Kumho Tire v. Carmichael

Facts and Issues Summarized

Plaintiff Patrick Carmichael brought action in a federal court in Alabama against the manufacturer of a tire that blew out as he was driving his minivan on July 6, 1993. One passenger was killed and others were severely injured in the accident. Plaintiff's expert witness, an engineer referred to as *a tire failure analyst*, claimed the tire's defective manufacture resulted in the blowout that caused the accident. The analyst based his testimony on a "visual and tactile inspection".

The defendant, Kumho, moved to exclude the testimony on grounds that it did not satisfy the admissibility requirements of Federal Rule of Evidence 702 (*see* also **Expert Opinion: United States**) and the *Daubert* (*see* also ***Daubert v. Merrell Dow Pharmaceuticals***) criteria. The district court agreed with the defendant and granted summary judgment for the defendant.

An intermediate appellate tribunal – the Court of Appeals for the Eleventh Circuit – reversed the trial judge's decision and remanded the original action for trial, holding that the trial judge should not have applied the *Daubert* factors because these guidelines applied only to scientific evidence. The plaintiff's expert, by contrast, offered his opinion based on acquired experience in working for a tire manufacturing company [1].[a]

On further review, the US Supreme Court had to decide if *Daubert*'s judicial gatekeeping role pertained only to expert testimony based on science or to all forms of expert testimony. Also at issue was how the *Daubert* factors had to be applied in the evaluation of the reliability of an expert opinion, and what the judges were to do if these factors did not squarely fit the type of expertise involved.

In reversing the decision of the intermediate appeals court, the Supreme Court held that trial courts have an obligation to consider both relevancy and validity of all challenged expert testimony. The Court also reiterated, as it had done in the previous *Daubert* decision, that the *Daubert* criteria were to be used flexibly. If specific factors had little relevance to a particular occupation or expert discipline, a trial court was free to exercise its discretion in utilizing other criteria for determining whether the proffered evidence was sufficiently reliable to satisfy admissibility standards. Finally, the Court held that the trial court had not abused its discretion in excluding the evidence in this case.

Discussion and Its Importance to Forensic Scientists

Prior to the Court's *Kumho Tire* decision, several trial courts had held that *Daubert*'s reliability standard

and the factors guiding its application only applied to "scientific" expert testimony and did not apply to skill-based opinions. Other trial courts had held to the contrary and applied the full panoply of *Daubert* factors in evaluating expert opinions. This had resulted in some occupations waging what might be referred to as *science wars*.

Some occupational specialties that had been traditionally referred to as *offering scientific evidence* had vigorously maintained that they were *not* involved in "science" *per se* for the precise purpose of avoiding application of Daubert standards, when its practitioners feared they could not meet the Supreme Court standards. Other disciplines, by contrast, had strenuously maintained that they were engaged in "science" despite the tortured nature of some arguments they presented.

This, in turn, had led to disputes among experts and legal scholars about *Daubert's* meaning of the terms "science," "scientific knowledge," the "scientific method," "forensic science," and similar descriptive terms. By deciding that the *Daubert* reliability standard applied to all forms of expert testimony, the Supreme Court rendered all of these arguments irrelevant in the future[b] *Kumho Tire* also addressed the issue of whether trial judges were required to apply the *Daubert* factors rigorously to experience – or skill-based expertise. The Court's opinion clarified that the enumerated factors of "testing," "peer review," "error rates," and "acceptability in the relevant scientific community" were merely flexible guidelines for trial judges. If they applied to a specific occupational speciality and would be helpful in evaluating the reliability of opinions it generated, the factors certainly could be used. But the Court also recognized that "The *Daubert* factors do *not* constitute a definitive checklist, or test". Because the gatekeeping function must be tied to the particular facts of a case, trial judges have wide latitude of discretion and must use criteria that are appropriate to each case. The opinion of the Court stated:

> *Daubert* is not to the contrary. It made clear that its list of factors was meant to be helpful, not definitive. Indeed, those factors do not all necessarily apply in every instance in which the reliability of scientific testimony is challenged. It might not be surprising in a particular case, for example, that a claim made by a scientific witness has never been the subject of peer review, for the particular application may never previously have interested any scientist. Nor, on the other hand, does the presence of *Daubert*'s general acceptance factor help show that an expert's testimony is reliable where the discipline itself lacks reliability, as, for example, do theories grounded in any so-called generally accepted principles of astrology or necromancy [2].

How were these criteria applied to the facts in the *Kumho Tire* case? The trial court had concluded that the plaintiff's expert was not based on an approved methodology, which it found to be unreliable. The Supreme Court approved of this determination.

On the specific *Kumho Tire* issue of whether overdeflection had caused the tire's tread to separate from its steel-belted carcass, the relevant issue was whether this particular expert had sufficient specialized knowledge to assist the jurors in deciding that fact. The trial court did not dismiss out of hand the expert's visual and tactile inspection methodology, but found a lack of connection between his testimony and the particular matter at issue. While the expert claimed his method was accurate, the Court noted that nothing requires a trial court to admit opinion evidence that is connected to existing data only by the *ipse dixit* of the expert.[c] It did not find this necessary connection and thus, the trial judge's ruling was determined as not being an abuse of discretion.

The *Kumho Tire* decision is important to forensic scientists not only for its determination at all forms of expert testimony must be shown to lead to reliably valid opinions, whether science-based or obtained through experience or special skills, but also that specific disciplines can no longer rely on the fact that older court decisions long accepted expert conclusions in a particular field. Courts can relitigate the admissibility of a certain type of evidence if a credible case can be made by a litigant that there has never been an adequate scientific inquiry into the validity of the assumptions on which this field has rested its conclusions in the past.

Thus, despite the fact that decades of judicial precedent may exist in a particular forensic discipline, the underlying lack of validity on which expert opinions rest may be relitigated. This has in fact happened, in the United States and elsewhere, when dealing with such venerable expert opinions as those offered by handwriting experts, firearm and toolmark examiners, fingerprint experts, fire investigators, bitemark specialists, trace evidence analysts, and other disciplines. The results of these challenges are discussed within the topics on each discipline.

End Notes

[a.] The facts and procedural history are set out in the Court's opinion. For another commentary discussing the path in the courts as standards moved from *Frye*, through *Daubert* to *Kumho Tire*, see: Sanders, J., "*Kumho and How We Know*," (2001), Law & Contemp. Probs. [Spring/Summer] at 373. See also, Savage, D.G., Putting the Brakes on Junk Analysis," (1999), ABA J. [May] at 38 and Saks, M.J., "Banishing Ipse Dixit: The Impact of Kumho Tire on Forensic Identification Science," (2000), Wash. and Lee L.Rev. (Summer).

[b.] The Court stated, [I]t would prove difficult, if not impossible, for judges to administer evidentiary rules under which a gatekeeping obligation depended upon a distinction between "scientific" knowledge and "technical" or "other specialized" knowledge. There is no clear line that divides the one from the others. Disciplines such as engineering rest upon scientific knowledge. Pure scientific theory itself may depend for its development upon observation and properly engineered machinery. And conceptual efforts to distinguish the two are unlikely to produce clear lines capable of application in particular cases". *Kumho Tire*, 526 U.S. at 148, 119 S.Ct. At 1174.

[c.] This reference in the Court's opinion was to an earlier decision expounding on *Daubert*, namely: *General Electric v. Joiner*, 522 U.S. 136, 118 S.Ct. 512, 139 L.Ed.2d 508 (1997), in which the Court used the *ipse dixit* language to illustrate the analytical gap between method used and opinion reached. See, in this regard, the separate articles on *General Electric v. Joiner* (*see* ***General Electric v. Joiner***) and *Ipse Dixit* Testimony (*see* ***Ipse Dixit* Testimony**; ***General Electric v. Joiner***; ***Ipse Dixit* Testimony**).

References

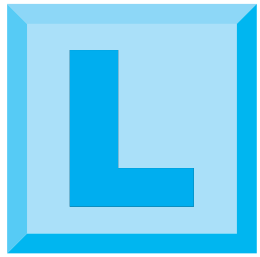
[1] *Kumho Tire v. Carmichael*, 526 U.S. 137, 119 S.Ct. 1167, 143 L.Ed. 2d 238 (1999).
[2] *Kumho Tire*, 526 U.S. at 151, 119 S.Ct. at 1175.

Related Articles

General Acceptance Test for Novel Expert Evidence

Weisgram v. Marley

ANN C. SMITH



Laboratory Accreditation ***see*** Accreditation: Organizational, Accreditation: Laboratory

Lactational Insanity ***see*** Postpartum Psychosis

Latent Fingerprint Detection ***see*** Friction Ridge Skin: Fingerprint Detection and Recovery Techniques

Laying the Foundation for Evidence ***see*** Foundation Testimony

LCN ***see*** Low Copy Number DNA

LCN: Interpretation ***see*** Interpretation: Low Template DNA

Learned Helplessness ***see*** Stockholm Syndrome

Learned Treatises as Evidence

Qualified experts who testify in court are expected to be familiar with published books, pamphlets, and periodicals that are part of the witness's recognized professional literature. A lack of familiarity with the literature is a reflection on the competence and believability of an expert witness. This is not to assert that everything published in a particular field must be known and remembered by a person testifying as an expert. However, in every profession there are certain standard publications that all experts are expected to have read and studied.

In adversary system jurisdictions, statements of an author of learned publications can be used by a cross-examining attorney if these statements conflict with opinions the expert offered on direct examination.

That opportunity of impeaching the expert is explored in a different article (*see* **Cross-Examination of Experts**). No "hearsay" impediment exists when a text is used to impeach the credibility of an expert on cross-examination. Cross-examination testimony is not offered for the truth of what it contains, but only to reflect on the credibility of the speaker, by the cross-examiner's attempt to show there are others in the expert's profession who hold contrary opinions **(***see* **Hearsay Evidence**). The prohibition against the use of hearsay does not pertain in civil law jurisdictions. It is a rule that is peculiar to evidentiary systems based upon the common law of England.

Rules against the use of hearsay evidence generally were originally conceived because of a special feature of adversary system jurisdictions. The feature is that a person against whom oral or written evidence is introduced must have a right to confront and cross-examine its maker. During the development of evidentiary law, it came to be recognized that some hearsay offers special guarantees of trustworthiness that compensate for the inability to produce the declarants as witnesses. These special circumstances gave rise to the creation of a number of exceptions to the rule against the use of hearsay.

In American jurisdictions that follow a rule similar to that contained in Federal Rule of Evidence 702, (*see* **Federal Rule of Evidence 702**), there exists an exception that permits learned treatises to be used, not only to impeach, but as substantive evidence of what the publication asserts. The theory for recognizing a learned publication exception is that the authors, while not in court, are persons who, by the recognition and status they have achieved in their disciplines, have staked their professional reputation on the accuracy of statements contained in the treatise or article.

As in the introduction of physical and documentary evidence, it is necessary to present oral evidence laying a foundation before a learned treatise or text can be used during a court proceeding. Laying the foundation requires testimony by the proponent's expert that the text is recognized as authoritative in his profession. In the alternative, authoritativeness may be elicited from an opponent's expert as well, by eliciting an admission on cross-examination as to its professional recognition. In those rare cases where a text is so well recognized that most persons know of it, a court may even take judicial notice of its authoritativeness (*see* **Judicial Notice of Scientific Principles and Facts**).

After the reliability of a publication is established, the relevant passage of the learned text can be read to the jurors, who may receive it as substantive evidence of the truth of its assertions. Jurors will, however, not be permitted to read the text themselves or take the publication to the jury room during deliberations.

It is always permissible for the party against whom a passage of an authoritative published text is offered to dispute the accuracy or currency by presenting expert testimony that the text read by the opponent is either not the current edition and is therefore outdated, that it no longer represents prevailing scientific opinion on the issue, or that other experts hold differing views.

Related Articles

Federal Rule of Evidence 702

Hearsay Evidence

Judicial Notice of Scientific Principles and Facts

ANDRE MOENSSENS

Legal Method Compared to Scientific Method *see* Scientific Method Compared to Legal Method

Length Measurement

Introduction

With the growing number of security cameras in the public and private domain, more and more incidents are recorded. In forensic practice, facial comparison of the perpetrator to the suspect is not always possible, the quality of the images being too poor, or the face of the perpetrator not being visible. In

these cases, it may, nonetheless, be possible to make an estimation of the body height of the perpetrator. Figure 1 shows an example of an image from a robbery of a cinema for which a height estimation of the perpetrator was requested. The results can be used to exclude or gather evidence against suspects and, as such, are interesting to the police, judges, and lawyers.

In this article, we first describe typical difficulties that arise when making height measurements in images. After this, we give an outline of four methods for height estimations, mentioned in literature. The section "Validation of the Measurements" addresses the question how results of the measuring process can be validated. How the evidential value of height estimations could be determined and reported is described in the section "Evidential Value (Likelihood Ratios)". The sections "Possible Future Research" and "Conclusions" describe the respective issues.

Typical Difficulties

Performance of height measurements in forensic casework is more involved than in a lab context. This is caused by different factors. First of all, the focal point of a surveillance camera is usually not known. The same applies to the position and orientation of the camera, which may have changed after the incident. The parameters can be retrieved by means of camera calibration, which will be described below.



Figure 1 Case example: questioned image. A height measurement was requested of the person in the oval

Secondly, surveillance cameras often have a lens distortion, the phenomenon in which straight lines appear as curved in the image. This results in deformation of people and objects in the image.

On the camera image in Figure 1, a few other problems can be seen, typical for closed-circuit television CCTV-images, which make it difficult to measure the height of the perpetrator. Circumstances such as lighting and the camera position may not be ideal. Also, resolution of an image may not be high because of limited storing capacity. A small number of pixels is a restriction on the accuracy of body height measurements.

Finally, the features of the perpetrator themselves influence the measurement. The suspect's pose, footwear, and headwear all affect the measured height. Heights are normally measured including footwear and headwear. More details are given below.

Camera Calibration

When a 3D object is projected on an image, the world and pixel coordinates of the object are related by a set of parameters such as the focal length of the lens, the size of the pixels, and the position and orientation of the camera. 3D measurements on a 2D image can be done after these parameters have been determined. This process is known as *camera calibration* and can be done in different ways.

A classical method is based on measurements of 3D coordinates of points on the scene that correspond to points on the image. In Ref. [1], a different technique is proposed, for which a planar pattern must be shown to the camera at different orientations. In the absence of a calibration object or knowledge about scene points, an estimation of the camera parameters can be found using the motion of a walking human as in Ref. [2].

Lens Distortion

Camera calibration is based on an ideal pinhole camera. The main problem with this approach in forensic applications is that surveillance cameras usually have significant lens distortion. Therefore, besides resolving the camera calibration parameters, the lens distortion problem must also be addressed.

A good method to correct for lens distortion is by placing a reference board with a raster in front of the camera. In Ref. [3], lines are selected that

should be straight in the image. The distance between the line through the end points and the middle point is computed to obtain a measure for the distortion. With software such as Mathematica, this distortion measure can be minimized and a corrected image can be created. Software packages such as PTLens correct for lens distortion automatically.

Questioned Person

Besides difficulties in body height estimation caused by the camera, there are those caused by the questioned person or perpetrator himself. As in Figure 1, perpetrators usually do not stand up straight in front of the camera, but will be walking, running, standing, or in other natural poses. This has to be taken into account, because stance influences the measured height [4]. Even standard standing measurements, for example, in doctor's surgeries, exhibit about 2-cm difference between natural and upright stance [4], and walking and running can cause even more differences in the height of the top of the head [5]. The amount of vertical displacement increases with speed.

Another factor that affects the measured height of a perpetrator is the thickness of shoe soles and headgear [4]. Here, a difference of up to 3 cm is reported between different shoe soles, excluding high heels.

Methods

In the literature, four different methods for making height measurements in camera images are given. These are described here, with their benefits and disadvantages. All methods are based on photogrammetry (literally: measuring in photographs), either on a single image or on different images taken from different sides and angles.

For every method, we attempt to report the uncertainty of the results. Unfortunately, this is seldom reported accurately. We reproduce the uncertainties as stated by the authors referred to. These values cannot be compared with each other, because they are all given in different measures. Additionally, the camera images and circumstances of the test scenes were diverse.

Reverse Projection Photogrammetry

An initial approach to performing height measurements in images that can be used when the original camera is available [6, 7] is the following. Using the same camera system, under identical circumstances as during the capture of the questioned image (same position, tilt, and focal length), an image is produced of an ordinary ruler of known height and proportions, placed at the same position as the questioned person (Figure 2). Position and orientation of the camera are checked using fixed reference points in the scene. The height distribution on the ruler is clear and with the same distortion as the questioned image, and hence the height of the questioned person may be read off the ruler. The method is referred to as *reverse projection photogrammetry* and used, e.g., by the Federal Bureau of Investigation (FBI).



(a) (b)

Figure 2 Security camera image of a perpetrator (on the left) and a reconstruction image of a ruler at the same location (on the right)

A problem connected to this method is [7] that it is not straightforward to identify whether the ruler is placed correctly at the spot of the questioned person, and results might be sensitive to deviations in location, especially when the image is distorted. In Ref. [6], height measurements of reference persons are mentioned, but the results are not included. The method is fast and easy, and therefore certainly useful for investigational purposes.

Single View Metrology (Vanishing Points)

The following method is, e.g., used in the United States of America (FBI) and in the United Kingdom (Home Office Scientific Development Branch).

When some knowledge of the scene is available, height measurements can be done on a single image, without the need of the original camera or camera calibration. Essential for a height measurement by single view metrology are

- a reference height of an object in the scene
- a set of vertical parallel lines
- two sets of horizontal parallel lines, in different directions.

As lines that are parallel in 3D will converge at a vanishing point when projected on an image, a vanishing point can be constructed by extrapolation of two parallel lines. All vanishing points of lines on a plane together form a vanishing line (or horizon). The construction of the plane vanishing line and a vertical vanishing point together with a reference height makes it possible to compute a height on the image [8]. In Figure 3, an example is shown of this method for making height measurements.

In Ref. [8], an uncertainty analysis can be found of the measurements. The measurement errors depend on the operator's uncertainty in locating the vanishing points, the top of the head, its projection on the ground plane, and a scale factor. In Ref. [4], it is shown that training the operators in locating the correct points on the top of the head and on the floor can reduce the uncertainty in measurements. The introduction of more reference heights may also reduce the uncertainty. In Ref. [8], the uncertainty in the height measurement of a man standing straight in an office is estimated on an image corrected for lens distortion. Here, a 99.7% confidence interval (3 standard deviations) was reported of ±3.9 cm around a mean, which reduced to ±3.5 and ±3.3 cm when two or three reference heights were used,

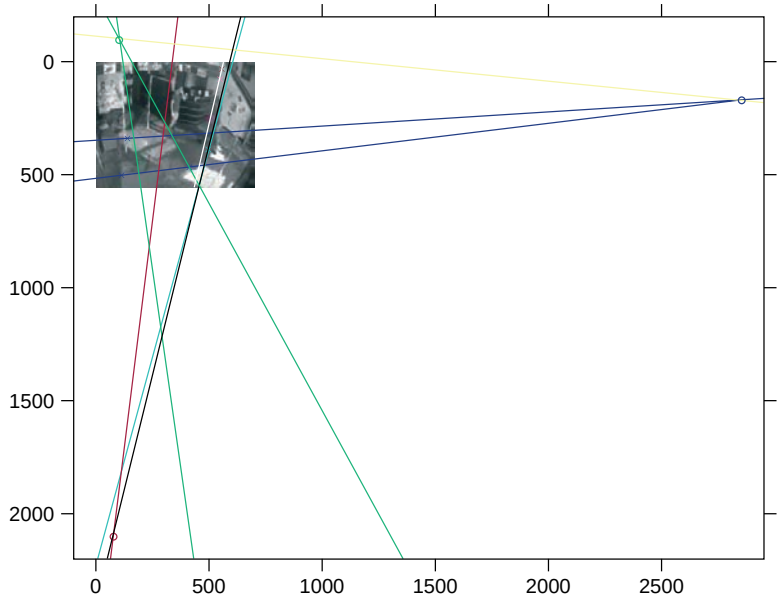


Figure 3 Illustration of modeling of the scene using vanishing points (reproduced with kind permission from J. Tighe, Home Office Scientific Development Branch, UK)

respectively. Measurements of the same test person on another image, recorded simultaneously from a different angle, resulted in a higher uncertainty of ±5.0 cm. This illustrates that the point of view may influence uncertainty.

An advantage of single view metrology is that it is fast, easy, and cheap. Applicability in casework is not always clear because enough parallel lines and a reference height must be visible on the image. Also, a correction for lens distortion may be needed, because lines are not always straight on the image. When radial distortion in an image is not minimized, it will introduce another uncertainty in the measurement.

3D Modeling of the Perpetrator

In the technical note [7], a description is given of how anthropometric measurements were performed on surveillance images from a bank robbery scene by making a 3D model of the perpetrator. The 3D model was consecutively held against images of a suspect.

For this method, in principle, different images of the perpetrator are needed, taken simultaneously by different cameras. The cameras are first calibrated by placing targets on the crime scene and taking photos with a digital camera. These measured points are imported in photogrammetric software such as Photomodeler Pro or RolleiMetric, and are used as control points for the calibration of the security cameras (including calculation of the location and orientation). With the information of the cameras, the software can calculate the 3D coordinates of concurrent points that are indicated in the different surveillance images. By indicating specific points, such as joints, on the different images of the perpetrator, three human models are created and some bodily measures calculated, among which is the total body height (including a helmet). Validation measurements were done on fixed objects, of which the error was less than 1%. Obviously, the errors will be greater for persons, because clothes can make it difficult to indicate the location of the joints. Validation measurements comparing a perpetrator with a suspect are given, but they seem inappropriate.

Applicability of the above in casework is troublesome, as images from different cameras are required. When there is footage from only one camera, there is a way to make use of information about the vertical plane the questioned person is standing in, as described in Ref. [7]. Of course, in casework this information may not be available as well.

3D Modeling of the Crime Scene

A fourth method of making height measurements in images is through the construction of a 3D model of the crime scene. This can be done in different ways. A technique that is often used is by means of photogrammetric software. As described previously, this software makes use of the fact that 3D coordinates of points can be measured by making photos of it from different positions. When a sufficient number of common points is identified on each image, a 3D model of the scene is made.

Instead of using photographs, a laser scanner can be used for the construction of a 3D model of the scene. In that case, a laser beam is sent to the scene from different angles. From all the points that reflect the laser beam, a point cloud will be obtained. Since the direction of the laser beam is known and the distance to every point is proportional to the time of flight of the laser beam, the 3D coordinates of the points can be known.

Next, a human operator links the scene points to the corresponding points in the questioned image. This makes it possible to determine the position, rotation, and focal length of the camera taking the images. This procedure of finding the right camera parameters is referred to as a *camera match*. Using the camera information, a virtual camera can then be placed in the 3D model of the room, looking at the model from the same perspective as the real camera at the real crime scene. Figure 4 shows a screen shot of the software 3dsMax, on which a 3D model of a simulated crime scene can be seen, together with its projection on a questioned image.

On the basis of the 3D model and camera parameters, in software such as 3dsMax it is possible to measure heights and distances on the image. The height of a person can be measured by placing a cylinder in the scene, from feet to the top of the head, or by creating human models (so-called *bipeds*) that can be adjusted to the stance of the person. The process of making measurements is depicted in Figure 5.

A description of the method using a high-resolution digital camera images is given in Ref. [9]. An uncertainty analysis of this method is given in section “Validation of the Measurements”. An advantage of this method is its wide applicability in forensic casework, since a 3D model of the crime scene can always be made, independent of the surveillance camera and the available images. A disadvantage is that



Figure 4 Example of a 3D computer model and its projection on the image (bottom right)



Figure 5 Example of performance of a height measurement using either a cylinder (on the left) or a biped (on the right)

taking the pictures or carrying out the laser scan is time consuming and expensive.

Validation of the Measurements

As for every measurement in real life, for each height estimation, an error is made; that is to say, measured heights are always estimations of actual values. Whatever method is used to measure body height, the key issue is the precision of the measurements. What is needed is a proper estimate of the accuracy of the height measurement on the person in the image.

The solid way to investigate this seems to be by doing validation measurements on test persons of known height, under more or less identical circumstances. This will make it possible to carry out a sound analysis on the measurement error.

We describe a standard operating procedure for the validation of height measurements used at the Netherlands Forensic Institute, based on the method of 3D modeling of the crime scene. The procedure consists of

1. carrying out a "reconstruction" at the crime scene, and
2. performing a statistically sound analysis of the results.

The end result will be a confidence interval for the perpetrator's height.

The validation procedure is based on the method of 3D modeling of the crime scene, but this is not essential. When using another method, the analysis can be performed comparably.

Reconstruction

The reconstruction consists of positioning test persons of known heights at the crime scene, in the same stance as the perpetrator, using the same camera. In the lab, the height of the test persons in the images is measured, which gives an idea about the accuracy of the measurement process.

The conditions under which control measurements take place must be similar to the original conditions. This means that the same camera must be used for the recording, so that the resolution and the lens distortion will be the same. It is also important that the camera is in the same position as before. This can be checked by placing an overlay of (characteristic lines of) the scene onto the new camera image. An example of this method is shown in Figure 6, in which the camera has been rotated after the crime scene was recorded.

By marking the location of the feet of the perpetrator on the overlay, the test persons are positioned such that the location and pose are more or less the same as the perpetrator's. The images of the test persons are captured, and their heights are measured in the same way as the perpetrator's.

Numerical Evaluation

Variation between actual and measured heights is introduced by the following factors:

1. creation of the 3D model;
2. finding of camera position, orientation, and focal length;
3. presence of lens distortion at the location of the perpetrator in the chosen image;
4. pose of the perpetrator in the chosen image;
5. presence and height of headwear and footwear; and
6. interpretation of head and feet in the images by the operators.

This variation may be decomposed into a systematic part and a random part. The systematic part is caused by variation in the modeling of the scene of crime (points 1, 2, 3) and difference in pose of test persons, as well as headwear and footwear (4, 5). Manual intervention by operators in the process (6) results in a random part of the variation.

From a statistical point of view, it makes sense to concentrate on the "total error" made in each particular height measurement, actual minus measured height, and assume it is normally distributed with certain mean (systematic error) and variation (random error). The mean and variation are estimated



Figure 6 Questioned image (on the left) next to reconstruction image (on the right). Following the overlay, it is clear that the camera has rotated after the incident

on the basis of outcomes for test persons. As the actual heights of the test persons are known, the average measurements on the test persons show whether there is a systematic error in the results. The measured height of the perpetrator must be adjusted with this error. The variation in the measurements is then used to determine a confidence interval for the height of the perpetrator.

Using 3D crime scene models, systematic errors found in casework go up to over 10 cm. However, because in these cases the perpetrator did not stand up straight, the systematic error may, e.g., be explained by the loss in height caused by his pose. Hence, this does not automatically mean that the measurement is not reliable. More problematic is a high variation in the measurements, which results in a large confidence interval. In casework, for 95% confidence intervals the band widths encountered reach from 5 up to 20 cm.

Evidential Value (Likelihood Ratios)

When height estimation is used as evidence in court, it is relevant to quantify its evidential value. This is not realized by a confidence interval. In fact, given a confidence interval and a suspect height, one cannot go much further than check whether the second is contained in the first. However, this is highly dependent on the chosen significance level, which is naturally somewhat arbitrary.

In literature [10], it has been described how evidential value can be quantified by means of so-called likelihood ratios (LRs). LR expresses numerically how rare the resemblance/difference is between a suspect's height and the confidence interval of the estimated height of the perpetrator.

In Ref. [11], a thought experiment is described about LR calculation for heights using witness statements. A theoretical foundation for the calculation of the LR based on population parameters and validation measurements on suspect and test persons has been described in Ref. [12].

There are different factors that influence the size of the LR. When the perpetrator's and suspect's heights do not match, the LR will be low and vice versa. Consequently, when the confidence interval is large, it will be more common to find a resemblance, so in that case the LR will also be low. Therefore the variation of the measurements must be low to retrieve a high evidential value. However, when the perpetrator is very tall or very small compared to the rest of the population, it is uncommon to equal the height, which may result in a high LR.

Possible Future Research

A question that arises is how much of the behavior of the estimation error is determined by the situation (camera, pose, lens distortion, resolution) and how much it behaves independently. This can be examined by testing methods on images from different cameras, and experiments investigating the latter will be valuable.

What is needed is a statistically sound comparison of the results in forensic casework of the four different methods. Points of interest are correctness and width of confidence intervals. Next to this, it is interesting to know in what percentage of the cases the individual methods are applicable.

It is furthermore interesting to look at the effect of radial distortion on measurements. Comparisons can be made between measurements on distorted images and measurements on the same images after correcting for lens distortion. The position of the measured people on the image might also influence the result.

Finally, further research is needed on the practical implementation of the calculation of numerical LRs for height measurements. This is important for the communication and translation of the evidential value of findings to judges, police officers, and the like. A typical problem in this respect is how to acquire proper databases of heights for fixed subpopulations.

Conclusion

Typical complicating factors in the process of body height estimation in digital images are, on one hand, camera parameters such as location, orientation and lens distortion; and, on the other, pose, footwear and headwear of the questioned person. Methods used in the literature include single view metrology (on the basis of vanishing points) and 3D modeling of either the perpetrator or the crime scene.

In all papers mentioned, except [8], validation of results has been limited. For any method used, results need to be validated by doing a reconstruction with persons of known heights posing in front of the same

camera, in the same pose as the person in the questioned image. The systematic and random error made in measurements is then modeled, and confidence intervals for the questioned person's height can be determined.

LRs quantifying evidential value on the basis of population parameters and validation measurements need to be studied in more detail. Moreover, dependency of results on the cameras needs to be studied, and the different methods need to be compared in actual casework.

Acknowledgments

The authors would like to thank John Tighe of HOSDB for his cooperation on the part on single view metrology.

References

[1] Zhang, Z. (2000). A flexible new technique for camera calibration, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(11), 1330–1334.
[2] Lv, F., Zhao, T. & Nevatia, R. (2006). Camera calibration from video of a walking human, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(9), 1513–1518.
[3] Bijhold, J. & Geradts, Z. (1999). Forensic photo/videogrammetry; monte Carlo simulation of pixel and measurement errors, in *Proceedings of SPIE*, Investigation and Forensic Science Technologies, Vol. 3576, pp. 239–246.
[4] Compton, D., Prance, C., Shears, M. & Champod, C. (2001). A systematic approach to height interpretation from images, in *Proceedings of SPIE*, Technologies for Law Enforcement and Security, Vol. 4232, pp. 521–532.
[5] Murray, M.P., Kory, R.C., Clarkson, B.H. & Sepic, S.B. (1966). Comparison of free and fast speed walking patterns of normal men, *American Journal of Physical Medicine* **45**(1), 8–24.
[6] Klasen, L. & Fahlander, O. (1996). Using videogrammetry and 3-D image reconstruction to identify crime suspects, in *Proceedings of SPIE, Investigative Image Processing*, Vol. 2942, pp, 162–169.
[7] Lynnerup, N. & Vedel, J. (2005). Person identification by gait analysis and photogrammetry, *Journal of Forensic Sciences* **50**(1), 112–118.
[8] Criminisi, A., Zisserman, A., Van Gool, L., Bramble, S. & Compton, D. (1998). A new approach to obtain height measurements from video, in *Proceedings of SPIE*, Investigation and Forensic Science Technologies, Vol. 3576, pp, 227–238.
[9] De Angelis, D., Sala, R., Cantatore, A., Poppa, P., Dufour, M., Grandi, M. & Cattaneo, C. (2007). New method for height estimation of subjects represented in photograms taken from video surveillance systems, *International Journal of Legal Medicine* **121**, 489–492.
[10] Aitken, C. & Taroni, F. (2005). *Statistics and the Evaluation of Evidence for Forensic Scientists*, 2nd Edition, John Wiley & Sons, Chichester.
[11] Lucy, D. (2005). *Introduction to Statistics For Forensic Scientists*, John Wiley & Sons, Chichester.
[12] Alberink, I. & Bolck, A. (2008). Obtaining confidence intervals and Likelihood Ratios for body height estimations in images, *Forensic Science International* **177**(2–3), 228–237.

Related Articles

Image Processing and Analysis
Reconstruction: Three Dimensional

GERDA EDELMAN AND IVO ALBERINK

Lie Detection *see* Deception: Detection of Deception: Truth Serum

Light Bulbs and Filaments: Examination of

Introduction

While night road traffic represents only about 20% of daytime traffic, one-fourth of all road accidents and more than one-third of all serious accidents (up to 50% in some instances) occur at night time [1, 2]. This imbalance is mainly due to a lack of visibility from drivers. Drivers assumed to be at fault sometimes claim that they did not see the other vehicle, alleging that the other vehicle's lights were not turned on. In many instances, the other driver, who claims that he/she was driving with the lights on, contests this. Therefore, during the investigation of a road traffic accident, law

enforcement personnel are sometimes required to determine whether the lights of a vehicle were on or off at the time of the accident [3]. While special units of uniformed police officers investigate most road accidents, criminalists at the forensic laboratory, who may not go to the scene, often perform the examination of light bulbs. However, a thorough light bulb examination must include examinations of the scene, the vehicle, and the different items of evidence at the laboratory. Thus, it is fundamental that both the field investigator and the criminalist communicate properly [4]. Only by including the circumstances of the case in the interpretation of the results, one can guarantee that the different observations made on the evidence are exploited to their full potential [5].

Light Bulb Types

A vehicle lighting system has to fulfill two main goals: to allow the driver to see the road ahead and to allow other motorists to see the vehicle. While all light bulbs are devices that produce light when electrically energized, it is possible to find three main types of light bulbs in modern vehicles: incandescent light, arc-discharge light, and light-emitting diodes (LEDs).

Incandescent Light

This is still the most widespread system found on vehicles. It is being slowly replaced by LED (mostly for rear and indication lights) and arc-discharge lights (for low and high beams). There are two main types of incandescent light bulbs: regular and halogen types, which are shown in Figure 1.

The heart of such light bulbs is the tungsten filament, which has a diameter of less than 0.1 mm and produces light by incandescence when energized. In the regular type of incandescent light bulb, the volume inside the glass globe is under vacuum to prevent the filament's oxidation. In time, tungsten slowly evaporates and deposits on the glass inner wall, thus the lifetime of the bulb is limited. The efficiency of such a bulb is about 2%. In the halogen type of incandescent light bulb, the globe is made of quartz, and the inner volume is filled with a halogen gas (usually iodine or bromine). As a result, tungsten does not deposit on the inner wall and the lifetime of the bulb is highly increased. Its efficiency usually reaches about 3.5%.

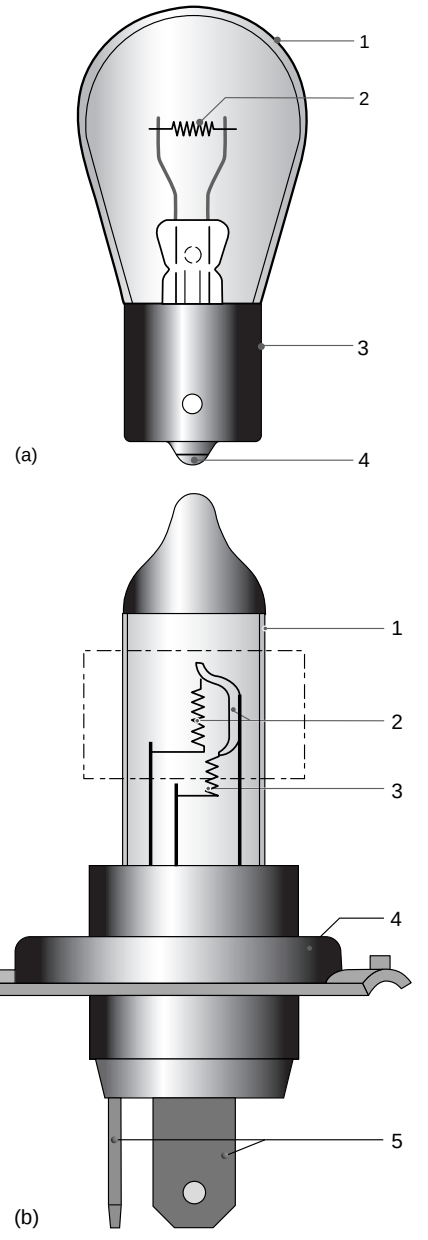


Figure 1 (a) Schematic of a regular incandescent light bulb: (1) glass globe, (2) tungsten filament, (3) metal socket base, (4) electrical contact. (b) Schematic of a halogen incandescent light bulb: (1) quartz globe, (2) low beam filament with cap, (3) high beam filament, (4) metal socket base, (5) electrical contacts. [Automotive lighting technology, windshield and rear-window cleaning, Robert Bosch GmbH, Plochingen, Germany, 2nd edition, 2003, Figures 3 and 4 [Reproduced with permission from Robert Bosch GmbH. © 2003.]

Halogen-type light bulbs are only used for high power (~40 W and more) headlights such as low/high beams and fog lights. While regular incandescent light bulbs are used for all other purposes, such as taillights and park lights. In some older vehicles and/or low-end vehicles, they can be found as low/high beam headlights.

Note that a light bulb may have more than one filament, as shown in Figure 1(b). In this case, it is used as low/high beam combination, but it can also be used for other purposes such as park/brake light.

Arc-Discharge Light

In arc-discharge or high-density discharge (HID) light bulbs, an arc passes through a gas (usually xenon with metal halides), which produces light. Thus, no filaments are present. This type of light bulb presents very little wear and, thus, a very long lifetime, and its efficiency is increased to about 22%. Nonetheless, it requires complex electronic control of the potential applied to the bulb. It remains extremely expensive and is found on high-end vehicles and as an option in many middle class vehicles. It is exclusively used for low/high beams.

Light-Emitting Diode (LED)

In a LED, a semiconductor diode emits light when a current passes through it. LEDs present no wear, an extremely long lifetime, a high resistance to mechanical shock, and an efficiency of up to 15%. LEDs are under constant development and are used on vehicles more frequently everyday, mostly to replace taillights and interior lights. Future developments will likely lead to the use of LEDs as headlights.

Failure Mechanisms

Recent research in forensic sciences has not been able to identify failure mechanisms for determining the condition of a light bulb at the time of its breakage when it is not equipped with a filament. Only the incandescent-type light bulb has a filament and, thus, can be examined to generate meaningful forensic results. Therefore, this discussion is limited to the incandescent-type.

The light produced by the incandescence of the filament is caused by heat generated by the flow of electrical current (Joule's effect). The filament can reach temperatures up to approximately 3000 °C [6]. The filament is placed in an inert atmosphere because it would completely burn if exposed to air at that temperature. As a result, the filament breaks differently depending on whether or not it is energized and whether or not the glass bulb is compromised.

From a failure mechanism perspective, it is important to consider whether the filament is energized or not. Then, it is clear that a mechanical shock has occurred; otherwise, there would be no mechanical stress applied to the light bulb and no failure would occur. When the filament is energized, several situations can occur. If the glass breaks, but the filament does not rupture immediately, the filament burns in the air and produces tungsten oxide until it opens [7]. Because of Joule's effect, the thinner the diameter of the filament, the more resistance it has and the more heat it generates; thus, the faster it will burn. As a result, one section of the filament usually disappears until it ruptures. The rupture pattern consists of two filament ends exhibiting decreasing diameters to the point of rupture and a globular appearance, as shown in Figure 2(a). This is accompanied by significant production of tungsten oxide crystals. If the glass breaks and the filament ruptures immediately upon glass breakage (due to the same shock), then the fracture is constituted of two ends of the same diameter, and the fracture surface will be globular in appearance (also called *melted fracture*) as shown in Figure 2(b) [8]. Usually, no tungsten oxide crystals are present on the filament, however there is usually some bluing or light oxidation at its surface. Nevertheless, when the filament is on at the time of the glass breakage, small glass particles deposit on the filament, melt on it, and solidify into spherical shapes (see Figure 2c).

If the globe does not break, then the filament does not come into contact with air. Thus, no oxidation of the tungsten occurs. However, the energized filament can rupture, in which case the fracture melts. If the filament is not energized, with or without glass breakage, the fracture will be brittle, as shown in Figure 2(d).

Forensic Approach

Limitations

It is important to understand that the forensic examination of a light bulb may reveal two things:

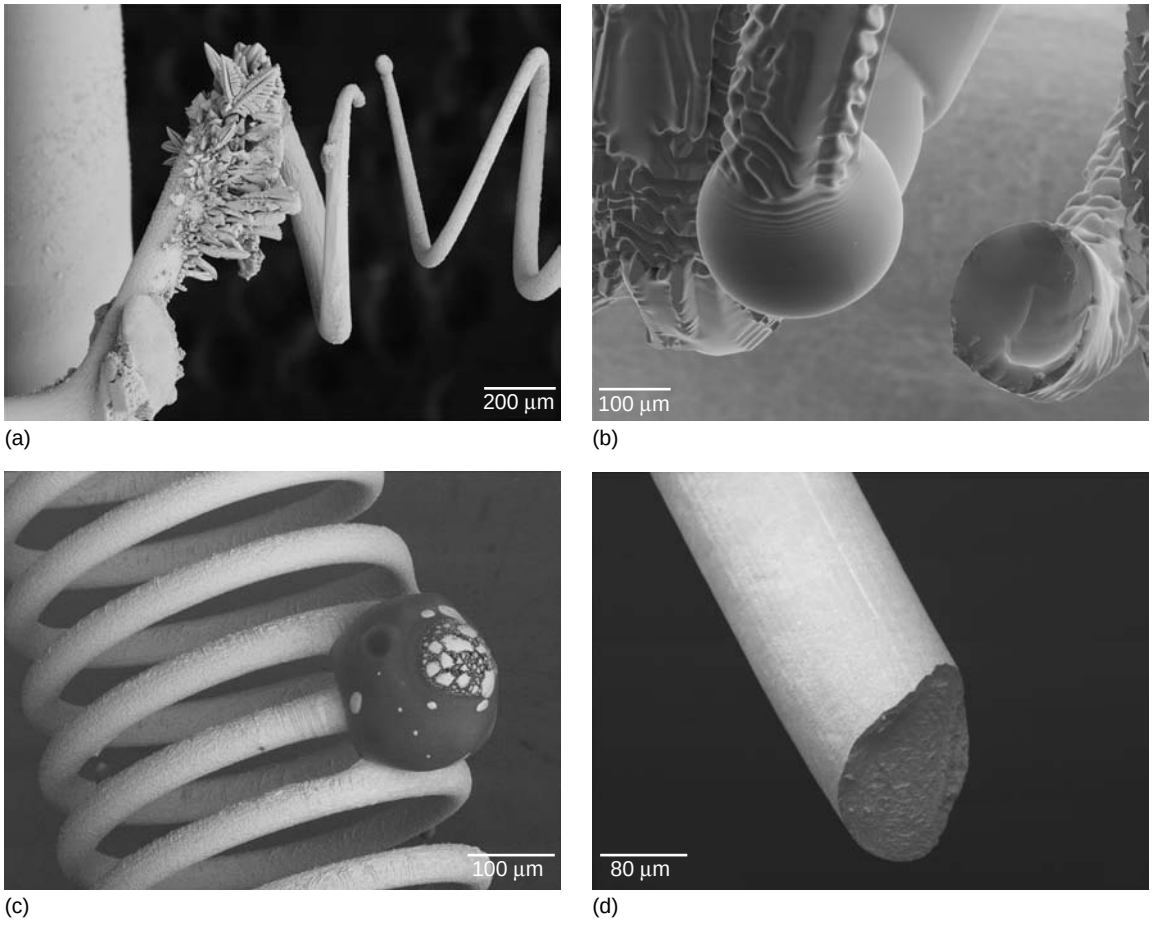


Figure 2 (a) A melted fracture due to burning of the filament in the air. Note the diminishing diameters before the fracture. (b) Melted fracture due to mechanical shock while the filament is energized. (c) Glass beads resulting from the globe breakage while the filament is incandescent. (d) Brittle fracture due to a mechanical shock while the filament is cold (not energized) [Reproduced with permission from SwissForensic.]

- Whether or not the filament was energized at the time of the glass globe breakage.
- Whether or not the filament was energized at the time of its rupture.

As a consequence, it is crucial not to confound the time of the glass breakage (or filament rupture) with the time of the accident [9]. While these two (or three) events could occur simultaneously, they could also be completely separate and independent events. It is not uncommon to observe vehicles driving on the road with either a broken taillight or headlight, or simply with one light not working. Therefore, the criminalist can generally not testify to the condition of the light bulb at the time of the accident, but rather at the time of the glass breakage or filament rupture. Although this limitation may affect the strength of the conclusion, it is often unavoidable. Of course, there are exceptions, notably when the scene and vehicle examinations have been thoroughly conducted and all evidence retrieved, and/or if the exact condition of the lighting system (including all light bulbs) was known just prior to the accident.

Overall Investigation

The forensic investigation involving the examination of light bulbs to determine whether they were on or

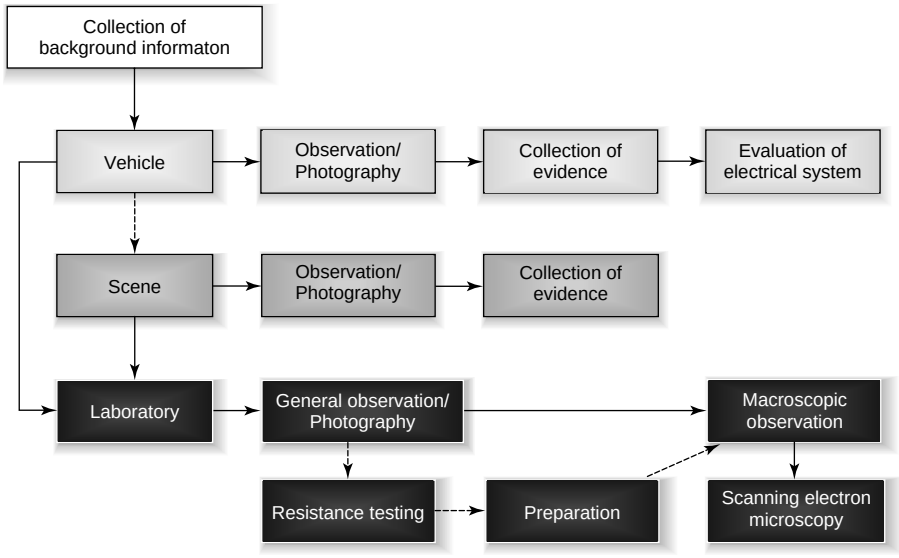


Figure 3 The different steps of the forensic examination of light bulbs

off at the time of an accident is conducted according to the steps shown in Figure 3.

First, it is crucial to obtain background information about the case and the testimony of the first responders at the scene, in particular, to understand who did what. For example, if a police officer tried to turn on the lights at the scene, this will greatly influence the outcome of the interpretation [10].

Scene Examination

The examination of the scene where the accident occurred should not be neglected. Even though the scene may have been "examined" by police officers in an attempt to reconstruct the accident, they may not have looked for light bulb remains. Often, if a light assembly has been damaged by mechanical shock, small pieces could have fallen and may remain on the ground. The search for these pieces is difficult and must be conducted meticulously. Photograph the scene along with any traces of impact. Note and photograph the position of any found items prior to removal. Be extremely gentle when moving them, because very tiny crucial pieces of evidence may be lost. Also remember that foreign materials, which have nothing to do with the vehicles involved, may be found on the accident site. This is particularly true in countries or places that are known not to be clean, where objects may remain on the side of the road for many months or years.

Vehicle Examination

When examining the vehicle, first ensure that the vehicle is properly identified [11]. Second, observe and note the position of the contact and (light) switches. Measure and record the remaining battery potential using a voltmeter. Then, thoroughly examine the fuse box. Check the integrity of each fuse. Record any abnormality or missing fuse. Be careful because many modern cars have two fuse boxes, usually one in the passenger compartment and one under the hood. Next, carefully observe and photographically record all damages to the vehicle. This is very important for interpreting the results. Sometimes the vehicle may have been damaged prior to the accident. If the damage is not consistent with the dynamics of the accident or if there are serious indications of aging, such as rust, then it is probably anterior to the accident in question. In any case, it is important to determine what damage resulted from the accident and what did not. Finally, examine and photograph all the light assemblies.

Then collect all damaged light assemblies, if possible, or, at least, the light bulbs. Additionally, it is good practice to collect all other light assemblies,

too, because they may bear some pertinence later in the investigation. For example, the rear brake light may reveal information about whether the vehicle was braking or not. When a filament is hot, it distorts more easily. As a result, if the shock was of sufficient force, the brake light filaments may be distorted, while the rear fog light filaments, if they were off, would not distort. This provides baseline information to the investigator regarding the force of the shock and the level of distortion expected. Also include the dashboard light bulbs, because these contain very tiny filaments, which distort very easily. As a result, it may be possible to determine which indicator light was on or off at the time of the shock.

Once all of the pertinent light bulbs have been collected, test the electrical lighting circuit using a multimeter. Use service manuals whenever necessary to identify the proper functioning of the lighting system. Then, verify light sockets to determine whether the circuit presents any modifications and/or abnormalities. At the end of the vehicle examination, the criminalist must have a clear documentation of the damages to the vehicle, the conditions of the light assemblies and light bulbs, and a full comprehension of the vehicle's lighting system.

Laboratory Examination

Before performing any manipulation on a light bulb, take a general photograph. One must always remember that filaments are extremely fragile and small titbits of important evidence can be lost from a bulb if not properly handled. Taking general photographs of light bulbs is never easy because of the reflection on the glass. The combined use of a polarization filter and diffuse lighting provides good results. Use a brightly colored background, either blue or pink, Post-It notes are extremely helpful.

Then, carefully observe the light bulb using a macroscope (magnification up to 40×). Note the condition of the glass and any cracks or abnormalities. Then, observe the filament(s), note their conditions and photograph them whenever possible. Carefully observe the contacts for any damage and irregularities. Finally, record all identifying marks. It is important to obtain the brand and model of the light bulb and its specifications. Further information may be obtained from the manufacturer.

Then, measure the resistance of the filament at the contacts using an ohmmeter. A normal resistance should range from 34 mΩ (for 5-W bulb) to 420 mΩ (for 60-W bulb) on a 12-V system. Often, it is possible to see whether the filament is open or not; however, sometimes, it is only possible to observe this through a resistance measurement.

It may also be necessary to remove the glass from the bulb. Proper macroscopy, or even electronic microscopy, requires the direct observation of the filament and its fracture surfaces. This cannot be achieved through the glass. To remove the glass from a regular incandescent-type light bulb, score the bottom part of the bulb with a file close to the metal base. Then, locally heat the glass at the score with a small flame. Finally, place a drop of water on the score. The rapid contraction of the glass due to cooling will break the bulb evenly along the score, thus allowing for a clean removal of the bulb. For the halogen type, wrap the bulb in paper towels and slowly squeeze it in a vice, until it breaks. The breaking is quite violent, so be very gentle when conducting this operation, and particularly when unwrapping because small pieces of filament can be lost.

Then, observe the filament in detail. If it is open, carefully locate and observe the fracture surfaces and photograph them. Check the filament for the presence of oxidation (bluing), tungsten oxide, and, most importantly, glass beads. Whenever possible, observe the filament with a scanning electron microscope. This allows for a significant improvement in the observation of the fracture surface. The presence of glass beads can also be easily pinpointed on the filament, and elemental analysis capabilities allow for the identification of glass and tungsten oxide.

Interpretation of the Results

Because the failure mechanisms of incandescent light bulbs are known, it is possible to determine the state of the light bulb at the time of the breakage as a function of its condition after the breakage. Figure 4 can aid in the interpretation of light bulb examination results in most instances. The first dichotomy depends on the condition of the glass. The second dichotomy is based on the filament's condition.

If the glass is intact and the filament is broken, the fracture surface is determinant. If the fracture is brittle (cold fracture), then the light bulb was off; if the fracture is melted (hot fracture), then the light

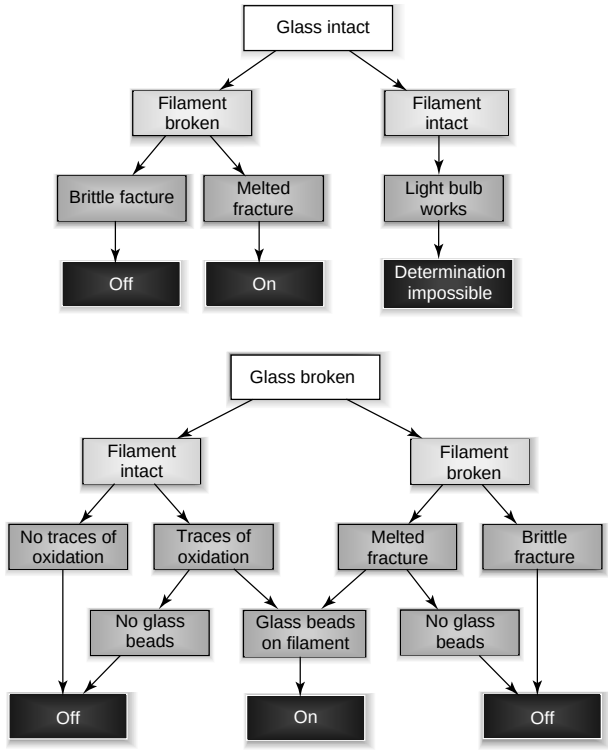


Figure 4 Scheme of the aid to the interpretation of light bulb examination results

bulb was on. When both the glass and the filament are intact and the light bulb still works, then the determination is impossible.

Things become slightly more complicated when the glass is broken, because if the light bulb is energized after glass breakage, further damage to the filament could occur. The criminalist must then take this possibility into account when interpreting the results. If the filament is intact, it is important to determine whether there are traces of oxidation. If the filament is clean and presents no traces of oxidation, this means that the light bulb was off at the time of glass breakage and was not energized afterward. If traces of oxidation are present, then it is important to determine whether glass beads are present on the filament. While there may be cases in which no glass beads would deposit on the filament when this one is on at time of glass breakage, only the presence of glass beads on a filament is a true indication that it was energized at the time of glass breakage. So, if no glass beads are present on a filament, which exhibits oxidation traces, it is likely that it was turned on after glass breakage. As a result, one cannot conclude that it was energized at the time of glass breakage. When the filament is broken, the fracture surface is determinant. If the fracture is brittle, the filament was not energized because it was cold. If the fracture is melted, then it means that the filament could have been energized at the time of breakage, or at some point after the breakage. Again, at that point, glass beads as well as the fracture surface, are decisive [12]. If they are present, the light bulb was on, if no glass beads are present, then the light bulb was likely off at the time of glass breakage.

Acknowledgment

The author would like to thank Dr Sarah D. Brown for her editorial review of this article.

References

[1] Economic Commission for Europe (2007). *Statistiques des Accidents de la Circulation Routière en Europe*

et en Amérique du Nord, United Nations, Geneva, Switzerland.
[2] Federal Highway Administration (2007). *Nighttime Visibility Facts and Statistics*, http://safety.fhwa.dot.gov/roadway_dept/retro/gen/night_facts.htm.
[3] Baker, J.S. & Aycock, T.L. & Lindquist, T. (1985). Lamp examination for on or off in traffic accidents, in *The Traffic-accident Investigation Manual: At-scene Investigation and Technical Follow-up*, J.S. Baker & L.B. Fricke, eds, Northwestern University Traffic Institute, Evanston, IL.
[4] Baudoin, P., Lavabre, R. & Vayne, F. (2002). An unusual oxidation type on bulb filament after a car crash dive, *Journal of Forensic Sciences* **47**(2), 377–380.
[5] Baudoin, P. & Lavabre, R. (1996). A particular case of oxidation colors on bulb filament after a car crash, *Journal of Forensic Sciences* **41**(2), 304–309.
[6] Bauer, H. (2003). *Automotive Lighting Technology, Windshield and Rear-window Cleaning*, Robert Bosch GmbH, Plochingen, Germany.
[7] Mathyer, J. (1975). Éléments de preuves fournis par l'examen des lampes électriques à incandescence, plus particulièrement de celles de véhicules impliqués dans des accidents de circulation, *Revue Internationale de Police Criminelle* **284**, 2–18.
[8] Thorsen, K.A. (1986). Forensic examination of bulb filaments by the SEM, in *Electron Microscopy in Forensic, Occupational, and Environmental Sciences*, eds, S. Basu & J.R. Millette, eds, Plenum Press, New York.
[9] Mathyer, J. (1975). Examen des lampes électriques à incandescence sur les véhicules impliqués dans des accidents de circulation (2ème partie et fin), *Revue Internationale de Police Criminelle* **285**, 34–41.
[10] Stauffer, E. (2007). Interpretation of automotive light bulb examination results: an intriguing case, *Journal of Forensic Sciences* **52**(1), 119–124.
[11] Smylie, W.T. (2006). Vehicle identification, in *Forensic Investigation of Stolen-recovered and other Crime-related Vehicles*, E. Stauffer & M. Bonfanti, eds, Elsevier Academic Press, Burlington, MA.
[12] Thorsen, K.A. (1981). Examination of bulb filaments by the scanning electron microscope, *Canadian Society for Forensic Science Journal* **14**(2), 55–69.

ERIC STAUFFER

Lineups: Eyewitness Identification ***see*** Eyewitness Lineups: Identification from

Lineups: Eyewitness Testimony ***see*** Eyewitness Lineups: Identification from

Live Independently: Capacity to ***see*** Capacity for Independent Living

Livor Mortis ***see*** Death: Time of

Longitudinal Juvenile Research ***see*** Northwest Juvenile Project

Low Copy Number DNA

Introduction

The late 1990s saw the advent of short tandem repeat (STR) DNA testing (*see* **DNA**). STR tests combine the sensitivity of a polymerase chain reaction (PCR)- based test with great discriminating power (quadrillions or quintillions of randomly chosen, unrelated individuals may need to be compared before two are found to have the same DNA profile) (*see* **Short Tandem Repeats**). PCR is a procedure that allows a small amount of DNA (which by itself would not be enough to type) to be amplified into an amount large enough for typing. It does this by making billions of copies of DNA fragments from a polymorphic area (or areas) of the genome. PCR is

not a genetic test itself, but merely a tool to increase the amount of genetic material that is being tested.

The amplification of DNA takes place in a test tube. The DNA that is extracted from each sample is placed in a separate tube, along with a mixture of primers, enzymes, and other reagents. The tubes are then placed in a machine known as a *thermal cycler*, which can control their temperature precisely while going through a series of heating and cooling cycles. Each cycle has three steps that are repeated 28 times (in conventional DNA profiling as recommended by the widely used Profiler Plus and SGM+ test kits), doubling the number of copies of the target DNA each time, and ultimately producing literally billions of copies. The amplified DNA, which was initially like a needle in a haystack of other DNA, is amplified to the point that there are far more needles than hay and those needles can then be characterized relatively easily.

Applied Biosystems, the manufacturer of the popular Profiler Plus and SGM+ test kits, recommends that a starting quantity of DNA should be between 1.0 and 2.5 ng (1 ng = 1×10^{-9} gm) in order for their kits to produce accurate and reliable results. A diploid human cell contains roughly 6.6 pg (1 pg = 1×10^{-12} g) of genomic DNA such that 1 ng of human DNA comes from approximately 152 diploid cells. Use of more than 2.5 ng of DNA template is known to give rise to a variety of issues that can complicate the interpretation of DNA test results including: the presence of off-scale peaks; split peaks (such as those that arise due to a technical artifact known as +A; *see* Figure 1); and imbalance in the height of peaks between loci. Use of less than 1.0 ng of template DNA is also known to give rise to a set of issues that can complicate the interpretation of DNA test results including: imbalance of the height of peaks within loci; imbalance of the height of peaks between loci; and allelic drop-out (Figure 1) and drop-in.

Stochastic Effects

Walsh *et al.* [1] recognized even before PCR-based DNA tests became popular that use of quantities of DNA below the recommended 1.0 ng starting amount could be problematic. "Stochastic" effects (sometimes also referred to as "preferential amplification") are at the heart of the problems associated with small quantities of template. These stochastic effects essentially arise from sampling errors that can occur when very few samples are made (much like those that might happen when blindly drawing black and white beans from a bag – a small number of draws, but not a large number, might suggest that all the beans in the bag are black even though they account for only 50% of a large number of beans in the bag)

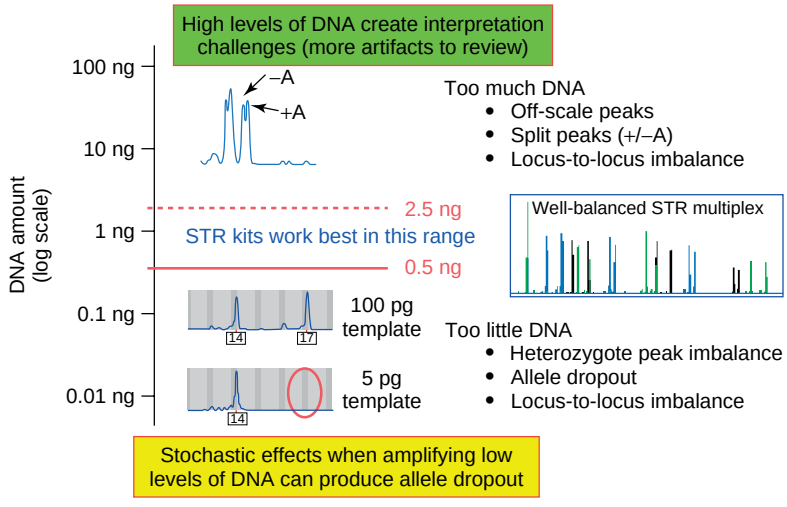


Figure 1 Impact of using nonoptimal amounts of template DNA for PCR amplification. [Derived from a presentation by Dr John Butler (US National Institute of Standards and Technology, Office of Law Enforcement Standards) at the Midwestern Academy of Forensic Sciences annual meeting in Indianapolis, Indiana on October 11, 2006]

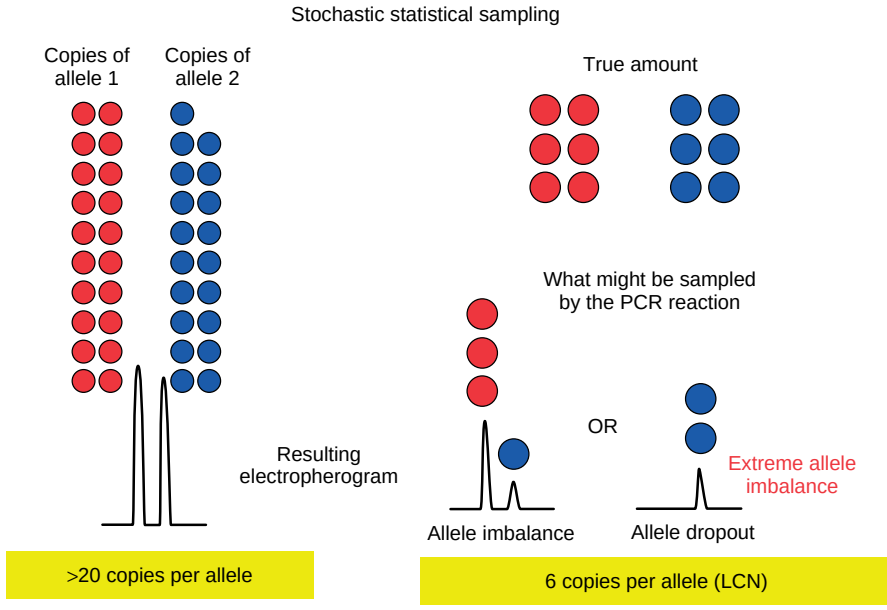


Figure 2 Stochastic statistical sampling during PCR amplification. When more than 20 copies of an allele are being amplified at the start of a PCR amplification it is unlikely that sampling errors will result in a large difference in the number of copies that are made of two different alleles. Sampling errors are much more likely to occur when fewer copies of the alleles are present at the start of the process. [Derived from a presentation by Dr John Butler (US National Institute of Standards and Technology, Office of Law Enforcement Standards) at the Midwestern Academy of Forensic Sciences annual meeting in Indianapolis, Indiana on October 11, 2006]

(Figure 2). With starting quantities of DNA arising from fewer than 150 human cells (and conceivably as few as five to ten cells), it is possible that one of two alleles at a locus will be amplified by the PCR process more than its counterpart (resulting in peak height imbalance or even allelic drop-out). It is also possible that stray alleles originating from just a few contaminating cells could be amplified preferentially, just by chance, relative to those that actually come from an evidence sample (resulting in allelic drop-in). Another commonly observed stochastic effect associated with small amounts of starting DNA template is an increase in the prevalence of a technical artifact known as "stutter" (which occurs when the enzyme responsible for making copies of the STR regions either slips forward or back during the amplification process and makes a copy that is either one repeat unit shorter or longer, respectively, than what was actually present in the template DNA). Quite simply, if there is not enough starting quantity of DNA prior to the time the PCR is started, the PCR process can produce results that are inaccurate and unreliable.

Avoiding Stochastic Effects

Because of the widely appreciated problems associated with stochastic effects in low copy number (LCN; sometimes also referred to as Low Template DNA or LTDNA) testing, the Applied Biosystems STR test kits normally fail when less than 125 pgs of DNA (that from roughly 20 diploid human cells) are used at the start of the process. This feature of the test kits specifically seeks to avoid these problems at the "stochastic limit of quantitation" (an amount of template that causes the chance of stochastic effects to be so likely as to make the test results incorrect) for the kits. The Technical Working Group on DNA Analysis and Methods (TWGDAM) validation of the AmpFISTR Blue test kit [2] explored the use of 27 and 30 PCR amplification cycles. This validation study for the test kit that was the precursor of the Profiler Plus and SGM+ test kit ultimately settled upon using 28 cycles so that quantities of DNA below 35 pg gave very low or no peaks – explicitly to avoid situations where peak imbalance results in

only one detectable allele from a heterozygous pair and concerns regarding trace contaminants could be minimized. It was assumed that 35 pg constituted an "analytical threshold".

Increasing the Sensitivity of Sensitive DNA Tests

One of the easiest ways that a PCR-based test can be made more sensitive is by simply increasing the amount of DNA replicating enzyme present and/or the number of PCR amplification cycles that are employed (Figure 3). One of the principal proponents of ultrasensitive DNA tests, the UK Forensic Science Service (FSS), in fact derives the increased sensitivity of their LCN tests primarily by increasing the number of rounds of PCR amplification to 34 cycles (six more than what is recommended by the manufacturer for the conventional use of the test kit that they use). Others have endeavored to generate information from lesser amounts of starting material by splitting samples and analyzing them multiple times and/or by purifying DNA from samples in different ways. While the details of these methodologies differ, they all suffer from stochastic effects due to the sampling errors (and ease of contamination and transfer) that small amounts of starting material entail. As a safeguard against the expectations of allelic drop-in (essentially, the presence of signal from unknown sources of contamination) and exaggerated stutter that increases in sensitivity brings, most laboratories

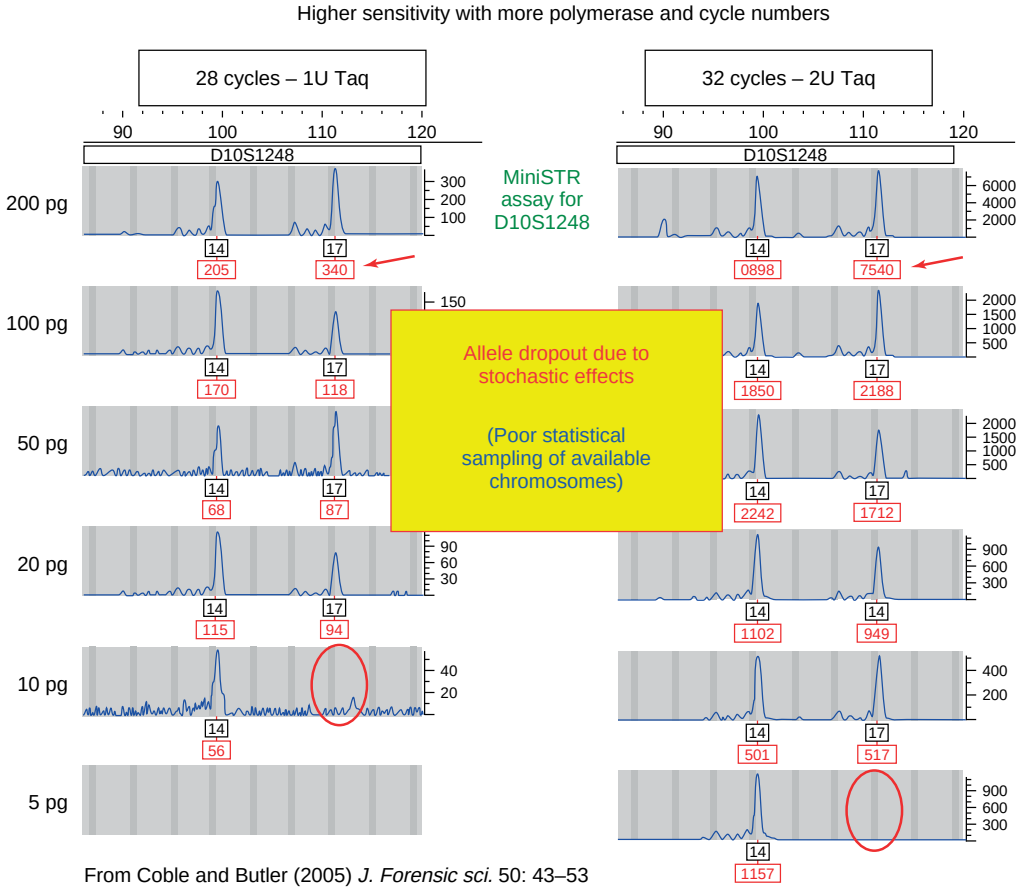


Figure 3 Higher scobnsitivity with more enzyme and rounds of PCR amplification. [Derived from a presentation by Dr John Butler (US National Institute of Standards and Technology, Office of Law Enforcement Standards) at the Midwestern Academy of Forensic Sciences annual meeting in Indianapolis, Indiana on October 11, 2006 (see also [3])]

that endeavor to perform analyses with less than recommended amounts of DNA template generate what are known as "consensus profiles". In such a consensus, a particular signal needs to be observed at least twice in multiple analyses of a sample before it is determined that the signal is associated with the sample rather than a contaminant. The reliance upon consensus profiles, especially when samples are analyzed only two times, actually results in an increase in the likelihood of problems associated with allelic drop-out which, in turn, results in reductions of the statistical weight attached to loci where only one allele is observed.

Increased Problems of Persistence and Transfer of DNA

Even under the very best of circumstances (virtually the opposite of those that require ultrasensitive DNA testing to be attempted), the presence of a DNA profile usually says nothing about the time frame or circumstances under which the DNA was transferred to an item. Further, the chance of "innocent" DNA transfer greatly increases as the amount of starting material for DNA profiling tests becomes smaller (e.g., [4]). Quite simply, even if one's DNA is found to be associated with an article of evidence, the great sensitivity of low template DNA testing causes very real problems to arise regarding both how and when that DNA was transferred.

DNA in quantities at LCN levels can be easily transferred (*see* **Transfer: DNA**; **Trace Evidence: Transfer, Persistence, and Value**) from one article to another (e.g., from evidence sample, onto the analyst's laboratory coat or gloves, then to another evidence sample; or by having been stored together in a single package) – without the contributor having any knowledge that the transfer(s) has occurred. No DNA tests are currently able to distinguish between secondary transfer (such as the transfer of DNA through contamination events) or DNA present due to direct contact with an object. Similarly, DNA tests are not currently capable of distinguishing in any way between the presence of DNA due to contamination (such as could very easily occur through storing or opening the objects in the same location as items obtained from an individual) or direct contact between an individual and the object. Given that LCN analyses can conceivably generate results from as little material as a few cells of an individual, the only way to be confident that results have not been obtained solely through contamination is to demonstrate conclusively with chain of custody (continuity) records that contamination is not even remotely possible. The greater the possibility of persistence and/or transfer of DNA from and between people and items, the greater the reduction in the probative value of DNA test results.

The Differing Paradigms of Conventional and LCN Testing

Conventional DNA profiling's greatest strength ultimately comes from its great ability to *either* include or to exclude an individual as a possible contributor to an evidence sample. In conventional PCR-based DNA profiling, the presence of alleles in an evidence sample that are not observed in a suspect and/or the presence of alleles in a suspect that are not observed in an evidence sample that is unlikely to have experienced allelic drop-out would normally exclude a suspect as a possible contributor to an evidence sample. But, because LCN testing operates at or beneath a stochastic limit of quantitation for a PCR-based amplification system, any number of such differences between evidence samples and a suspect's reference can be easily (and reasonably) attributed to allelic drop-out, allelic drop-in, or other stochastic effects (Figure 4). At the very least, each of these factors can lead to alternative interpretations of the evidence and, consequently, to a compounded reduction of the probative value of the test results (*see* **Mixture Interpretation: DNA**).

In conventional forensic DNA analysis, attempts are made to exclude a suspect, and only after failing to exclude, inferences are made regarding the rarity of the observed DNA profile. However, the statistical interpretation of LCN test results is significantly complicated by the expectation of stochastic effects. While inherent flexibility of interpretation of LCN testing results may be useful for the generation of investigative leads to which no statistical weights need to be attached, that flexibility of interpretation makes it unsuitable for establishing proof that an individual was directly associated with an evidentiary sample (let alone, what the time frame and circumstances of such an association may have been).

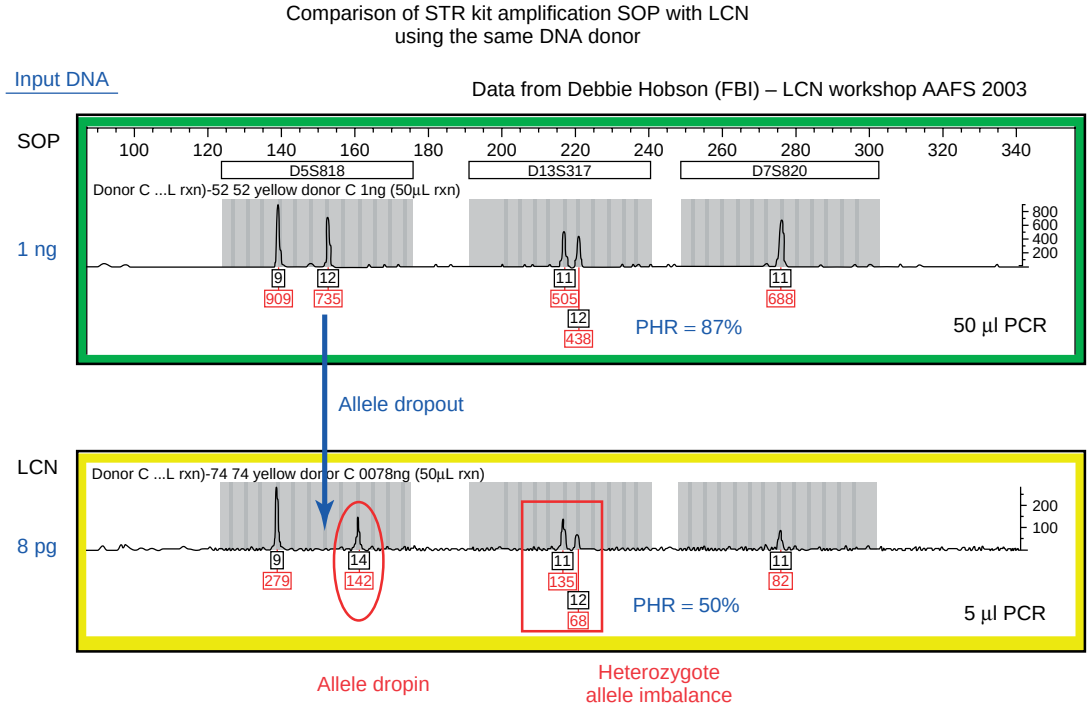


Figure 4 Comparison of conventional STR amplification and LCN amplification with the same test kit and source of human DNA. With conventional STR testing the true DNA profile of an individual can be determined directly from an electropherogram. LCN testing of the same individual with the same test kit gives rise to a significantly different electropherogram. [Derived from a presentation by Dr John Butler (US National Institute of Standards and Technology, Office of Law Enforcement Standards) at the Midwestern Academy of Forensic Sciences annual meeting in Indianapolis, Indiana on October 11, 2006]

Utility of Ultrasensitive DNA Tests

There is significant and ongoing debate regarding the utility and proper use of ultrasensitive DNA tests. A challenge to the reliability of LCN DNA profiling in the trial of Sean Hoey in Belfast Crown Court in Northern Ireland [5] prompted the United Kingdom's new Forensic Science Regulator to commission a review of low template DNA profiling techniques. That review [6] was released on 12 April, 2008 and concluded that the technologies collectively called low template DNA profiling are "robust" but accepts that what was offered as evidence in Sean Hoey's trial was not sufficient to establish the validity of the technique. The review also enumerates 21 recommendations (including the development of a consensus on the interpretation of test results and that more work be done to establish "best practices" for interpretation) that should be undertaken to improve the methodology. There are in fact things about LCN approaches upon which the reviewers and critics do agree. For instance, the International Society of Forensic Genetics similarly concludes that there is "a significant need for continuing education and research into" LCN reporting [7]. Also, Budowle *et al.* [8] caution that "Publicizing the potential of the application of LCN typing without describing its limitations may cause misunderstanding" which is consistent with the review's recommendations 1, 3, and 13. But given Budowle *et al.*'s [8] conclusion that "The method cannot be used for exculpatory purposes," the review's ultimate conclusion that LCN testing is "fit for purpose" leaves the important but unanswered question of "what *is* that purpose?"

Summary

Budowle *et al.* [8] observed that "Because of the successes encountered with STR typing, it was inevitable

that some individuals would endeavor to type samples containing very minute amounts of DNA." However, the stochastic effects associated with small amounts of template (e.g., allelic drop-out and drop-in, exaggerated peak height imbalance and stutter) coupled with the diminished ability to ascertain the tissue source of DNA samples or how long they have been associated with an article reduce the weight that can be attached to the finding of an ultrasensitive DNA profile match.

References

[1] Walsh, P.S., Erlich, H.A. & Higuchi, R. (1992). Preferential PCR amplification of alleles: mechanisms and solutions, *PCR Methods and Applications* **1**, 241–250.

[2] Wallin, J.M., Buoncristiani, M.R., Lazaruk, K.D., Fildes, N., Holt, C.L. & Walsh, P.S. (1998). TWGDAM validation of the AMpFISTR™ blue PCR amplification kit for forensic casework analysis, *Journal of Forensic Sciences* **43**(4), 854–870.

[3] Coble, M. & Butler, J. (2005). Characterization of new MiniSTR loci to aid analysis of degraded DNA, *Journal of Forensic Sciences* **50**(1), 43–53.

[4] Rutty, G.N. (2002). An investigation into the transference and survivability of human DNA following simulated manual strangulation with consideration of the problem of third party contamination, *International Journal of Legal Medicine* **116**, 170–173.

[5] *R v Hoey* (2007). NICC 49, 20 December, 2007.

[6] Caddy, B., Taylor, G.R. & Linacre, A.M.T. (2008). *Review of the Use of Low Copy Number DNA Analysis in Current Cases*, at http://police.homeoffice.gov.uk/publications/operational-policing/Review_of_Low_Template_DNA_1.pdf?view=Binary.

[7] Gill, P., Brenner, C.H., Buckleton, J.S., Carracedo, A., Krawczak, M., Mayr, W.R., Morling, N., Prinz, M., Schneider, P.M. & Weir, B.S. (2006). DNA Commission of the International Society of Forensic Genetics: recommendations on the interpretation of mixtures, *Forensic Science International* **160**, 90–101.

[8] Budowle, B., Hobson, D.L., Smerick, J.B. & Smith, J.A.L. (2001). *Low Copy Number – Consideration and Caution*, Proceedings of the 12th International Symposium on Human Identification, (www.promega.com/geneticidproc/ussymp12proc/contents/budowle.pdf).

DAN E. KRANE

LTDNA: Interpretation *see*

Interpretation: Low Template DNA

Luminol

Introduction

Ever since its forensic potential was first reported in 1928 [1], luminol (5-amino-2,3-dihydro-1,4-phthalazinedione or 3-aminophthalhydrazide, Figure 1) has been utilized for the detection of latent bloodstains during investigations involving violent crime. This use is based upon the emission of light when a reagent containing luminol and hydrogen peroxide is applied to the suspect stain, the resulting glow being due to the reaction of luminol with hemoglobin. The key events in the discovery, study, and use of luminol as a forensic reagent are presented in Figure 2. Despite the age of this presumptive test, there is still significant interest in the forensic community due to its high sensitivity and lack of damage on genetic material, allowing subsequent DNA analysis (*see* **DNA**). This interest has resulted in a number of recent publications, including modifications to the technique [2, 3], a new, commercially available formulation [4, 5], and studies on its performance as a forensic test [6, 7]. Recently, the authors carried out a comprehensive review of the forensic application of the luminol reaction as a presumptive test for latent blood detection, upon which this article is substantially based [8].

Luminol

Chemical and Physical Properties

The key chemical, physical, and toxicological characteristics of luminol are summarized in Table 1 [9, 10]. In relation to its use as a forensic reagent, solutions of luminol are very sensitive to light and the presence of metal cations; typically the solutions are

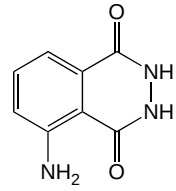


Figure 1 Chemical structure of luminol

stable only for periods of 8–12 h. Luminol, both in the pure form and in solution, is also thermally unstable, and therefore should be protected from high temperature [11].

The key property of luminol as a reagent for forensic science is its ability to undergo reactions which produce light. This phenomenon is termed *chemiluminescence* [12]; the fundamentals of which have been comprehensively reviewed in a number of textbooks and articles in recent years [13–16].

Table 1 Chemical, physical, and toxicological properties of luminol [9, 10]

Names	5-Amino-2,3-dihydro-1, -phthalazine-dione, *o*-aminophthalyl hydrazide, 3-aminophthalic hydrazide
Molecular formula	$C_8H_7N_3O_2$
Molecular mass	177.16 amu
Melting point	319–320 °C
pK_{a1}	6.74
pK_{a2}	15.10
Solubility in water	<0.1 g 100 ml^{-1} at room temperature
Physical properties	Yellow crystalline solid (grainy crystals)
General properties	Stable at room temperature, sensitive to light, combustible, incompatible with strong oxidizing agents, strong acids, strong bases, strong reducing agents, emits light on reaction with oxidizers (chemiluminescent)
Safety information and potential health effects	The toxicological properties have not been fully investigated in humans; mucosa irritation has been described: eyes, skin, respiratory tract, and gastrointestinal tract (with nausea, vomiting, and diarrhea). No data available about chronic effects. More information available at The National Toxicology Program (The National Institute of Environmental Health Sciences, NC, USA); Website http://ntp.niehs.nih.gov/index.cfm

Chemiluminescence

The luminol reaction is an example of direct chemiluminescence, which can be represented by:

$$A + B \longrightarrow [I]^* \longrightarrow \text{Products} + \text{Light} \qquad (1)$$

where A and B are reactants and $[I]^*$ is an excited-state intermediate. The excited-state intermediate returns to the ground state by a variety of routes. The light emitted has differing degrees of intensity, lifetime, and wavelength, with the latter parameter covering the spectrum from near ultraviolet, through the visible and into the near infrared.

The intensity of the chemiluminescence emission observed from a reaction depends upon both the reaction rate and the efficiency of the process generating the excited-state species. The use of very sensitive detectors and the almost complete absence of background emission have allowed the monitoring of even inefficient chemiluminescence reactions, such as the oxidative ultraweak chemiluminescent reactions in living cells [17].

The efficiency and wavelength of a chemiluminescence emission are greatly affected by the environment in which the reaction takes place. For solution-phase chemiluminescence, the factors that will affect the reaction are similar to those affecting normal fluorescence and phosphorescence.

For most analytical purposes, it is the chemiluminescence emission intensity (I_{CL}) that is measured, either as an integral over the lifetime of the emission or as a transient response. It is a function of both the efficiency and the rate of the reaction:

$$I_{CL} = \Phi_{CL}\left(\frac{dC}{dt}\right) \qquad (2)$$

where dC/dt is the rate of reaction (molecules reacting per second). Chemiluminescence reactions can occur very rapidly (<1 s) or extremely slowly (>1 day), according to the reaction and the conditions.

The Luminol Reaction

Luminol chemiluminescence has recently been reviewed by Barnett and Francis [13]. The excited-state intermediate has long been established as the 3-aminophalate dianion (Figure 3) [13]. In aqueous

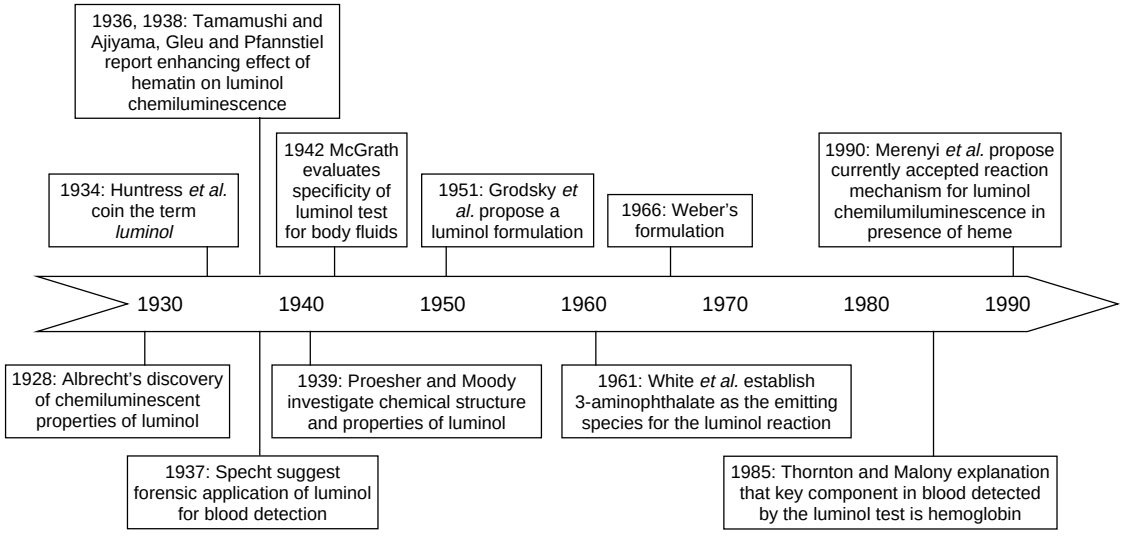


Figure 2 Key events in the discovery, study, and use of luminol as a forensic reagent

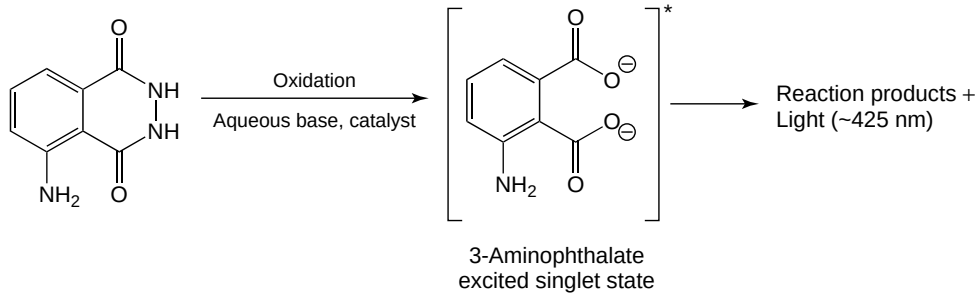


Figure 3 Luminol chemiluminescence reaction [Reproduced from Ref. 13. © Elsevier, 2005.]

solutions, the light observed ranges between blue-violet and blue-green, although the spectral range of emission is often rather broad and the observed maximum is dependent on several parameters of the reaction [18, 19]. This can include the presence of blood itself, which strongly absorbs at 420 nm, and may provide an inner-filter effect, thus shifting the observed maximum emission of luminol chemiluminescence to about 455 nm [3].

The light-producing pathway for the oxidation of luminol is a complex, multistep process and is dependent on several factors. These include pH, temperature and ionic strength of the reaction medium, and the reactive species that can be present in solution and interact with luminol, metal catalyst, or hydroxide ions [13]. For the luminol reaction, the exact role of the catalyst, which is required when the reaction is carried out in basic aqueous solution, and the reaction intermediates are not completely characterized. It is known that a wide range of other transition-metal catalysts and metal complexes catalyze the reaction. The optimum conditions of pH for the reaction depends on the identity of the catalyst used and varies between pH 8 and 11 [20], thus suggesting a multiplicity of potential catalysis mechanisms.

The luminol reaction mechanism has been the subject of significant discussion, and many suggested pathways have been postulated [21–29]. The current understanding of the probable reaction mechanism accounting for a number of findings from the aforementioned authors has been reviewed by Barnett and Francis [13]. A simplified scheme of the currently most accepted mechanism is presented in Figure 4 and is based on this work and on previous research by

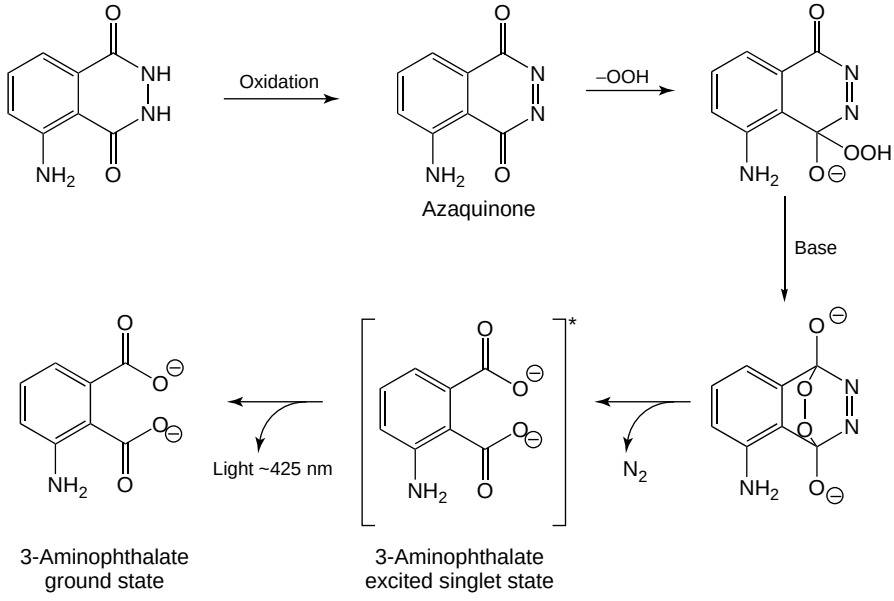


Figure 4 Simplified postulated reaction mechanism for luminol chemiluminescence

Merényi and collaborators during the 1980s [22–27]. For a more detailed discussion of the mechanism, the reader is directed to the authors' earlier review and references therein [8].

The key component within blood thought to be responsible for eliciting chemiluminescence with luminol is the iron-containing metalloprotein hemoglobin. This is the oxygen-carrying molecule found in the erythrocytes of all vertebrates and some invertebrates and is responsible for the red color of blood. Once outside an organism and deposited on a substrate, hemoglobin degrades [30], with the iron present within the molecule being oxidized from Fe^{2+} (ferrous) to Fe^{3+} (ferric). These ferric heme derivatives show the same catalytic properties and capability of participating in two-electron redox cycles as peroxidases, which are widely distributed especially in vegetables, their activity is termed a *pseudo-peroxidase* or *peroxidase-like*. This activity is commonly employed as the basis for many presumptive tests for blood, including luminol [29–33].

When a luminol formulation is applied on a bloodstain, ferric heme groups are able to catalyze both the decomposition reaction of peroxide and the oxidation of luminol and other substrates by peroxide [34–37]. These reactions are thought to happen as a result of the ability of the hydroxy–ferric–porphyrin (OH –Fe^{3+}–P) hematin group within hemoglobin to undergo a two-electron oxidation to a hydroxy–ferryl–porphyrin radical (OH –Fe^{4+}–P), which can return to the ferric porphyrin hematin state in two one-electron reduction steps *via* the hydroxy–ferryl–porphyrin (OH –Fe^{4+}–P) hematin group (Figure 5) [28, 37–40]. The catalytic process thus cycles between these three oxidation states of the hemoglobin, with the stable resting state being the ferric hematin. Alternative catalytic cycles and oxidizing species have been proposed (e.g., see Thornton and Maloney [28]), but the above cycle is now accepted by most researchers [37, 41].

The Luminol Reaction as a Presumptive Test for Blood

Operational Use

Luminol can be used to detect the presence of minor, unnoticed, or hidden bloodstains. It can disclose distribution, allowing bloodstain pattern evaluation and occasionally enabling the investigators to reconstruct some of the events of a crime by visualizing these patterns [42, 43]. Compared to other chemical-based

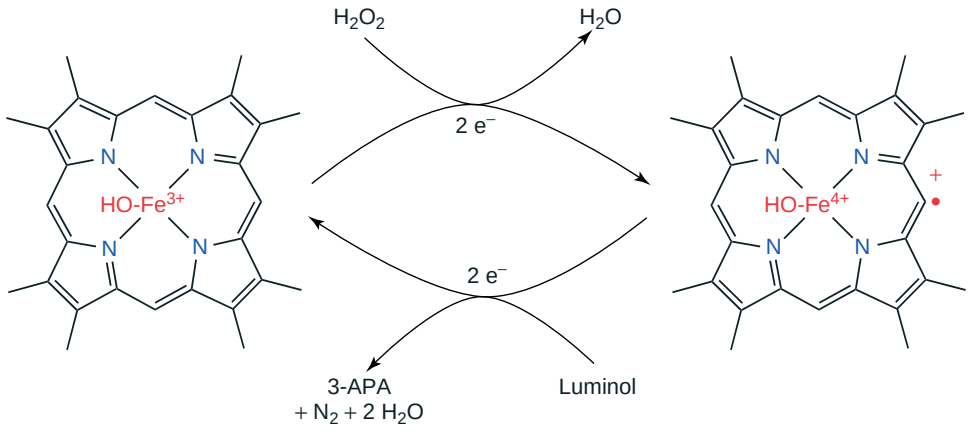


Figure 5 Iron redox cycling between different oxidation states when hematin is formed and sixth coordination position is occupied by –OH (P, porphyrin). In alkaline conditions and in the presence of a medium-strong oxidant such as hydrogen peroxide, the hydroxy–ferric–porphyrin (OH–Fe^{3+}–P) hematin group undergoes a two-electron oxidation to a hydroxy–ferryl–porphyrin radical (OH–Fe^{4+}–P)

tests and physical techniques for the detection of blood at crime scenes, luminol has the advantage of high sensitivity [7, 44, 45]. A recent evaluation reported that the luminol test could detect blood to a dilution of 1 : 100 000 [6]; earlier studies have reported that blood diluted down to a level of $1 : 10^6$ (1 μl of blood in 1 l of solution) was capable of being detected [33, 46, 47]. In addition, the luminol test does not significantly damage genetic material, especially when modern polymerase chain reaction (PCR) techniques are employed to analyze microsatellite DNA (*see* **DNA**). Only moderate adverse effects have been noted when other DNA testing procedures or serological markers were commonly used for identification purposes [6].

While several preparations of luminol have been described, including patented luminol molecule modifications or luminol blood-dependent chemiluminescence enhancers [28, 48], the formulations described by Grodsky *et al.* [47] and Weber [40], continue to be the most extensively used by forensic practitioners. This is due to their good performance, simplicity of preparation, low cost, and ready availability of the ingredients. These protocols are summarized in Table 2.

Regardless of the preparation, luminol solution is usually directly sprayed in completely dark environments. The light obtained can be photographed or filmed while the luminescent areas are marked in order to allow their detection once the light emission has faded [33, 49]. Examples of images of luminal-treated latent bloodstains are presented in Figure 6. Amplification of the luminolchemiluminescent emission by means of intensified cameras has been reported in a forensic context [28, 50] but is not in general use at crime scenes. Pretreatment of the surfaces possessing the stains with 2% hydrochloric acid (HCl) has been recommended by some authors [51, 52]. However, others have reported that this appears to decrease sensitivity by raising the background chemiluminescence level [42]. Acid treatment may also have a deleterious effect on subsequent DNA typing [33].

Owing to the potential irritant and harmful effects of luminol formulations (Table 1), and their application as aerosols, particular care has to be taken when conducting forensic testing for blood. Suitable personal protective equipment (goggles, respirators, gloves, and protective clothes) should be used by the operators when luminol is sprayed, and the area investigated should be ventilated after luminol application. The number of people assisting the operations should be limited to those strictly necessary [33].

Collection of the stains for further laboratory testing follows bloodstain location and photographic documentation, the method of collection depending on the nature of the substrate on which the stain is located (*see* **Biological Swabs**).

Table 2 Protocols for forensic luminol formulations

	Grodsky *et al.* [47]	Weber [40]
1.	Dissolve sodium perborate (3.5 g) in deionized water (500 ml) with stirring	Dissolve sodium hydroxide (8 g) in deionized water (500 ml) with stirring (stock solution A)
2.	Add luminol (0.5 g) and sodium carbonate (25 g) to the sodium perborate solution and stir until dissolved	Dissolve luminol (0.354 g) in the previously prepared sodium hydroxide solution (62.5 ml) with stirring (stock solution B)
3.	Decant solution into vaporizer or sprayer and use immediately	Dilute 30% hydrogen peroxide solution (10 ml) in deionized water (490 ml) (stock solution C)
4.		To prepare test solution mix 10 ml of each of the three stock solutions together and decant resulting mixture into a vaporizer or sprayer and use immediately
Note	Weighed out reagents can be kept for extended periods in plastic or glass containers prior to dissolution in water, luminol requires protection from light	Stock solutions can be stored in glass or plastic solutions at or below 4 °C away from direct light for extended periods



(a) (b)

Figure 6 Photographic images of luminol emission utilizing the Grodsky *et al.* formulation [47]: (a) tile floor where a man had been murdered and attempts made to clean blood from the crime scene; (b) carpet and tiles in toilet where small, barely noticeable bloodstains had been left in a murder case (note the false positive emission from the metal pipe connected to the WC (marked X)). Photographs courtesy of Francesco Cominetti, specialist photographer [Reproduced with permission from Ref. 47. © Northwestern University, 1951.]

Factors Influencing the Use of Luminol

While the luminol test itself is relatively simple, interpretation of results is more challenging. Factors that should be taken into consideration include the physical structure of the substrate upon which the bloodstains are found, the chemical composition of the substrate possessing the stains, and any other substances present on the substrate.

Physical Nature of Substrate. The first issue confronting the forensic practitioner when using the luminol test at a crime scene is a consideration of the physical nature of the substrates possessing the stains [33, 42, 53]. Substrates can be divided roughly into two groups: absorbent materials and nonabsorbent materials.

Absorbent materials encompass substrates such as carpeting, leather clothes, fabric clothes, roof linings, blankets, and also irregular, porous surfaces such as

wood-finish paneling, walls, and interstitial spaces between tiles or wood objects. These materials often can retain significant amounts of blood, maintaining relatively undegraded for many years. This is due to the rapid drying of blood, especially in domestic or covered environments, thus preventing its degradation by environmental biological agents such as bacterial hydrolytic enzymes. In addition, these substrates can protect blood from physical or chemical environmental agents such as solar radiation, moisture, and water, or cleaning attempts after the crime has been committed [33, 42]. It is possible with these substrates to apply multiple applications of the luminol reagent, without the risk of excessively diluting the stains, in order to best visualize and to successfully photograph the bloodstain pattern [33, 42].

Nonabsorbent substrates such as nontextured linoleum, vinyl, tile, glass, metal, and many others present more difficulties both in the reagent application and in the quality of chemiluminescence. These substrate surfaces are unable to effectively retain and store blood and, moreover, cannot prevent its degradation especially by physical and chemical agents. As clearly demonstrated by Lytle and Hedgecock [42], these surfaces are fairly easy to completely clean, and a mild washing attempt by water and soap leads to the removal of the bloodstains yielding almost nonexistent reaction with luminol. A further complication is that the application of luminol solutions to nonabsorbent surfaces can lead to the bloodstain pattern running, owing to the limited retention of the resulting solution by the smooth surface. [33]. Particular care should therefore be used when dealing with these substrates, particularly when they are nonhorizontal, in order to avoid the loss of the stains. Investigators should first use a minimum amount of luminol solution by rapidly spraying, preferably with a nebulizer, the suspected area, and avoiding further applications quickly photograph the emission [33, 42, 53].

Influence of Interfering Substances. There is a wide range of environmental, pharmaceutical as well as domestic and industrial substances that are able to affect luminol blood-induced chemiluminescence (Figure 6). False-negative results have not been described in the forensic literature; however, there are some substances that may suppress luminol chemiluminescence which could possibly come into contact with blood. These include polyphenolic derivatives [26, 54, 55] such as tannins, which are widely present in wood, and some amino acids [56]. More problematic are those compounds that provoke intensification or generation of a chemiluminescence emission even if blood is not present, leading to false-positive results. Owing to the possible presence of these substances at the crime scene, the luminol test must not be considered sufficiently specific to permit an unequivocal identification of blood [47, 53, 57–59]

Those compounds that generate luminol chemiluminescence, or enhance the luminol emission in the presence of bloodstains, can be divided into three major categories (Figure 7):

1. compounds showing a catalytic true peroxidase or peroxidase-like activity;
2. compounds with a high oxidizing capacity towards luminol; and
3. compounds with a complex chemical composition with an undefined action mechanism towards luminol mixture.

The first group encompasses inorganic or bioinorganic species and represents the major source of interferences with the forensic application of luminol. These compounds often show excellent catalyzing properties in redox reactions and are widely distributed in the environment and in plants. These substances include free metal ions, in most cases included in inorganic compounds such as rust or soils; biological complexes between metal ions and organic components (such as metal-porphyrins, and including bacterial or plant pigments) often within protein structures; and enzymes belonging to the oxidoreductases class such as horseradish peroxidases.

Commonly encountered examples of the second category of interferents are sodium hypochlorite, potassium permanganate, and iodine. These species are present in many household and industrial chemical solutions, including insecticides, cleaning agents, disinfectants, or antiseptics [53, 60–62].

The final category of interfering substances covers a range of compounds that can be found in materials such as domestic and commercial oils, various glues, carpets, sinks, automobile seats, paints, and varnishes, and many kinds of soils [11, 33, 63, 64]. These substances are often able to catalyze the luminol reaction almost as effectively as blood, but owing to their complex chemical composition, the exact

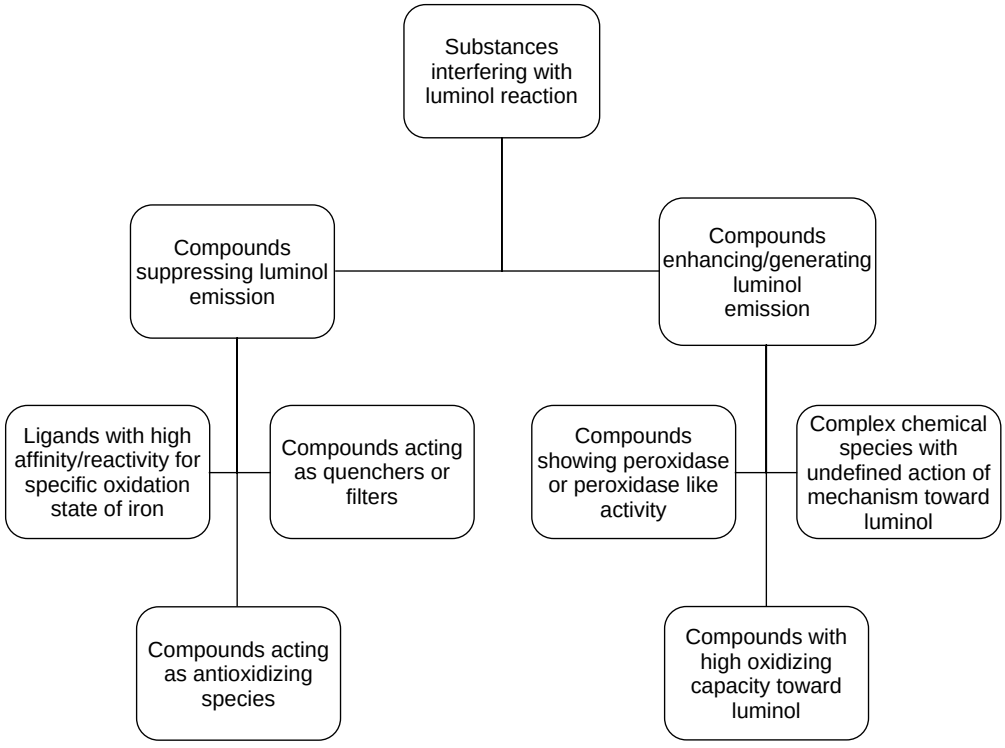


Figure 7 Classification of substances interfering with luminol chemiluminescence

mechanism underlying these interferences is not yet completely understood.

Quickenden and coworkers used instrumental methods to carry out extensive surveys of the effects of interferents with the luminol reaction [64, 65]. This included comprehensive studies of the luminol chemiluminescence emission elicited by a wide range of common, potentially interfering substances such as vegetable and fruit smears, pulps and juices, and household/industrial chemicals such as cleaning agents, insecticides, glues, paints, and varnishes. Of the 250 substances examined, they identified only a small number that produced chemiluminescence comparable to that of hemoglobin, including turnips, parsnips, horseradish, commercial bleach (sodium hypochlorite), copper metal, some furniture polishes, some enamel paints, and some interior fabrics from automobiles [66].

The serious issue of the hypochlorite interference effect with the luminol test [67] was also examined by this group [63]. They observed that when a person attempts to remove bloodstains by washing the area with water or sodium hypochlorite solution, depending on the thoroughness of the clean the effect on the luminescence spectrum could range from the complete absence of emission to various combinations of blood-initiated emission and hypochlorite-initiated emission (each peaking at its separate, respective wavelength), which might be expected if the cleaning process is not complete.

Finally, the same group examined a specific kind of crime scene, namely, the interior of an automobile. This study took into consideration both the effect of potential interferences from the internal fittings of the vehicle and also the effect of high temperature within the vehicle on the efficacy of the test [68]. Attempts to wash hemoglobin from the interior of the vehicles were tested using a variety of cleaning methods. It was found that there was little interference from materials within the three automobiles tested, although some surfaces did elicit weak luminol chemiluminescence. Attempts to remove hemoglobin with water alone were not successful; however, soapy water or a proprietary car cleaner removed

a significant proportion of the hemoglobin from the tested surface (ca. 90%). The use of cleaners and soapy water produced better results than water alone because they have the ability to solvate the globin proteins more effectively. The effect of increased car interior temperature led to improved sensitivity and chemiluminescence emission intensity of the test and was hypothesized to be due to thermal conversion of hemoglobin to methemoglobin in the presence of molecular oxygen.

Interpretation of Luminol Test Results

Luminol emission pattern interpretation can involve a qualitative statement of the luminescence pattern characteristics and an evaluation of emission spectra characteristics (maximum emission wavelength and emission intensity) [64–66]. In addition, attempts can be made to inhibit or at least reduce the interferences of chemiluminescence deriving from the reaction of luminol with substances other than blood by using chemical species followed by an emission intensity measurement [2, 3]. The latter approach, however, has only been employed successfully for hypochlorite bleach-induced chemiluminescence.

Generally visual examination is used when the luminol test is employed in a forensic situation. An experienced practitioner may distinguish the true blood-catalyzed chemiluminescence from that produced by other substances by the evaluation of parameters observable to the naked eye, such as emission intensity, duration, and spatial distribution. However, as this approach is subjective, it may lead to misinterpretation, and therefore caution should be exercised when using the test. Any confusion that may arise over a stain can usually be resolved by intelligent observation and, if necessary, by further testing [42], for example, by using a different presumptive test for blood, such as the immunochromatographic test for the confirmation of human blood presence [69].

In practice, false positives with metals are rarely a problem, as these can usually be anticipated or resolved by careful observation of the crime scene (see Figure 6). Interfering solid substances such as a metal object or surfaces that are coated homogeneously with these substances (e.g., some varnishes and paints) generally show different and distinguishable emission patterns with respect to both the spatial distribution and, often, the emission intensity of chemiluminescence. Upon reaction with metals, both emission kinetics and the intensity of chemiluminescence are rather characteristic: the emission will be intense but short. The reaction with blood will produce a less intense, longer-lasting even glow. Moreover, interfering chemiluminescence, especially from a solid object, such as a water valve, a knife, a copper pipe, floor, a carpet, or soil, will reproduce the shape and the dimension of the object. Emission patterns with blood will appear as spatters, wipes, smears, drag marks, or even footwear impressions.

The presence of hypochlorite-based bleaches on nonporous surfaces being sprayed is sometimes recognizable and can be identified by an experienced forensic practitioner, as it leads to bright flashes of chemiluminescence as opposed to the more gradual development of chemiluminescence by blood.

One operational advantage of the luminol test is the ability to highlight the presence of scattered, very small droplets of blood by the individual “sparkles” of blue chemiluminescence produced by each droplet. This makes this test easier to interpret than the other three common presumptive tests for blood (benzidine, phenolphthalein, and leuco-malachite chromogen tests) [7, 53].

Improvements to Luminol Formulations

A number of approaches to minimize luminol interferences and/or to increase the yield of the chemiluminescence emission have been described in the analytical chemistry literature. This has included the use of derivatives and analogs of luminol, a variation in the order of mixing of the reagents, the pretreatment of the substrate to be tested with chemical substances [58], and the addition to the luminol mixture preparation of chemical additives that selectively react with the putative interfering species reducing their availability for the reaction with luminol [54]. The successful use of a chemical species preventing luminol emission by interfering compounds has been reported by Kent and coworkers in their studies into reducing the effect of hypochlorite-containing bleaches [2]. Kent and coworkers found that certain strongly basic amines, such as 1,2-diaminothane, could inhibit the chemiluminescence due to hypochlorite under the alkaline conditions typical of forensic luminol tests (Grodsky’s or Weber’s formulations) [2]. A major disadvantage to this approach to reduce hypochlorite effect

on luminol emission was the toxicity of the amines involved. In a follow-up study, King and Miskelly confirmed the above results but they also found that the far less toxic amino acid glycine was nearly as effective [3].

In the above studies it was also noted that if the bloodstain could be left in air for a period of 1 to 2 days, the hypochlorite would decompose and thus no longer interfere with a standard luminol treatment. This effect has also been reported by Creamer *et al.* in 2005 [63], who found that the interference effect by bleach decreased if the area to be sprayed were left for several days allowing the bloodstains to thoroughly dry, as the hypochlorite decomposes, thus dissipating its effect on luminol emission.

Recently, a commercially available patented luminol-based formulation called Bluestar Forensic (ROC Import Group, Monte Carlo, Monaco) has become available and has been evaluated against traditional luminol preparations [4, 5]. From examination of the material safety data sheets available at the supplier's website (http://www.bluestar-forensic.com), the key differences in the formulation are the use of urea peroxide as an oxidant and sodium hydroxide to provide the basic conditions required for the reaction. These papers concluded that Bluestar Forensic provided the convenience of preparation (easy to mix in the field), that the chemiluminescence was sufficiently intense and long-lasting so that it could be visualized in the presence of some ambient light, and that the emission intensity was still reasonable when a bloodstain was resprayed. A recent comparison of luminol and Bluestar Forensic however found that while Bluestar Forensic gave good sensitivity, it showed poor specificity [6].

Acknowledgment

This article is substantially based on the authors' earlier review [8]. The authors would like to express their gratitude to Steve Gutowski from the Victoria Police Forensic Services Centre in Macleod, Victoria, Australia, to Donatella Pietraforte from the Cellular Biology and Neuroscience Department of the Superior Institute of Health in Rome, Italy, to Ron Nichols from the Department of Justice-Forensic Science Laboratory of Walnut Creek in California, USA, and to Dimitri Svistunenko Gordon Miskelly, University of Auckland from the Department of Biological Sciences of the University of Essex, UK. All these persons indirectly collaborated at various levels to the earlier review providing both valuable bibliography material and suggestions, comments, and important contributions. The authors would also like to thank Lisa Swann and Renee Jelly, both at Curtin University of Technology, for reviewing the draft article.

References

[1] Albrecht, H.O. (1928). Über die Chemiluminescenz des Aminophthalsäure Anhydrids, *Zeitschrift fuer Physikalische Chemie* **136**, 321–330.

[2] Kent, E.J.M., Elliot, D.A. & Miskelly, G.M. (2003). Inhibition of bleach-induced luminol chemiluminescence, *Journal of Forensic Sciences* **48**, 64–67.

[3] King, R. & Miskelly, G.M. (2005). The inhibition by amines and amino acids of bleach-induced luminol chemiluminescence during forensic screening for blood, *Talanta* **67**, 345–353.

[4] Blum, L.J., Esperança, P. & Rocquefelte, S. (2006). A new high-performance reagent and procedure for latent bloodstain detection based on luminol chemiluminescence, *Canadian Society of Forensic Science Journal* **39**, 81–100.

[5] Dilbeck, L. (2006). Use of Bluestar Forensic in Lieu of luminol at crime scenes, *Journal of Forensic Identification* **56**, 706–720.

[6] Tobe, S.S., Watson, N. & NicDaéid, N. (2007). Evaluation of Six Presumptive Tests for Blood, Their Specificity, Sensitivity, and Effect on High Molecular-Weight DNA, *Journal of Forensic Sciences* **52**, 102–109.

[7] Webb, J.L., Creamer, J.I. & Quickenden, T.I. (2006). A comparison of the presumptive luminol test for blood with four non-chemiluminescent forensic techniques, *Luminescence* **21**, 214–220.

[8] Barni, F., Lewis, S.W., Berti, A., Miskelly, G.M. & Lago, G. (2007). Forensic application of the luminol reaction as a presumptive test for latent blood detection, *Talanta* **72**, 896.

[9] Budavari, S. (1996). *The Merck Index*, 12th Edition, Merck & Co., Whithouse Station.

[10] Kricka, L.J. (1995). Chemiluminescence and bioluminescence, *Analytical Chemistry* **67**, 499R–502R.

[11] Stott, R.A.W. & Kricka, L.J. (1987). Purification of luminol for use in enhanced chemiluminescence immunoassay, in *Journal of Proceedings International Bioluminescence Chemiluminescence Symposium 4th Meeting*, J. Schoelmerich, ed, John Wiley & Sons, Chichester, UK, pp. 237–240.

[12] Wiedemann, E. (1888). Über Fluorescenz und Phosphorescenz, *Annalen der Physik Chemie* **34**, 446.

[13] Barnett, N.W. & Francis, P.S. (2005). Chemiluminescence: liquid-phase, *Encyclopedia of Analytical Science*, 2nd Edition, Elsevier Academic Press.

[14] García-Campaña, A.M., Baeyens, W.R.G., Cuadros–Rodríguez, L., Barrero, F.A., Bosque–Sendra, J.M. &

Gamíz–Gracia, L. (2002). Potential of chemiluminescence and bioluminescence in organic analysis, *Current Organic Chemistry* **6**, 1–20.

[15] Hastings, J.W. & Johnson, C.H. (2003). Bioluminescence and chemiluminescence, *Methods in Enzymology* **360**, 75–104.

[16] Roda, A., Guarigli, M., Michelini, E., Mirasoli, M. & Pasini, P. (2003). Analytical bioluminescence and chemiluminescence, *Analytical Chemistry* **75**, 463A–470A.

[17] Campbell, A.K. (1988). *Chemiluminescence: Principles and Applications in Biology and Medicine*, Ellis Horwood, Chichester.

[18] Roswell, D.F. & White, E.H. (1978). The chemiluminescence of luminol and related hydrazides, *Methods in Enzymology* **57**, 409–483.

[19] White, E.H. & Bursey, M.M. (1964). Chemiluminescence of luminol and related hydrazides: the light emission step [13], *Journal of the American Chemical Society* **86**, 941–942.

[20] Larena, A., Valero, M. & Bernabeu, E. (1983). Some chemiluminescence reactions in presence of different metal ions, *Optica Pura Y Aplicada* **16**, 91–96.

[21] Lind, J. & Merényi, G. (1981). Determination of the chemiluminescence quantum yield of luminol in rapid chemical reactions, *Chemical Physics Letters* **82**, 331–334.

[22] Eriksen, T.E., Lind, J. & Merényi, G. (1981). Oxidation of luminol by chlorine dioxide: formation of 5-aminophthalazine-1,4-dione, *Journal of the Chemical Society Faraday Transactions* **77**, 2125–2135.

[23] Lind, J., Merényi, G. & Eriksen, T.E. (1983). Chemiluminescence mechanism of cyclic hydrazides such as luminol in aqueous solutions, *Journal of the American Chemical Society* **105**, 7655–7661.

[24] Merényi, G. & Lind, J. (1980). Role of a peroxide intermediate in the chemiluminescence of luminol. A mechanistic study, *Journal of the American Chemical Society* **102**, 5830–5835.

[25] Merényi, G., Lind, J. & Eriksen, T.E. (1984). The equilibrium reaction of the luminol radical with oxygen and the one-electron-reduction potential of 5-aminophthalazine-1,4-dione, *The Journal of Physical Chemistry* **88**, 2320–2323.

[26] Merényi, G., Lind, J. & Eriksen, T.E. (1986). Nucleophilic addition to diazaquinones. Formation and breakdown of tetrahedral intermediates in relation to luminol chemiluminescence, *Journal of the American Chemical Society* **108**, 7716–7726.

[27] Merényi, G., Lind, J. & Eriksen, T.E. (1990). Luminol chemiluminescence: chemistry, excitation, emitter, *Journal of Bioluminescence and Chemiluminescence* **5**, 53–56.

[28] Thornton, J.I. & Maloney, R.S. (1985). The chemistry of the luminol reaction – where to from here?, *California Association of Criminalists Newsletter*, 9–16.

[29] White, E.H., Zafiriou, O., Kagi, H.H. & Hill, J.H.M. (1964). Chemiluminescence of luminol: the chemical reaction [12], *Journal of the American Chemical Society* **86**, 940–941.

[30] Gaensslen, R.E. (1983). *Sourcebook in Forensic Serology, Immunology, and Biochemistry*, U.S. Department of Justice, Washington, DC.

[31] Cox, M. (1991). A study of the sensitivity and specificity of four presumptive tests for blood, *Journal of Forensic Sciences* **36**, 1503–1511.

[32] Higaki, R.S. & Philp, W.M.S. (1976). A study of the sensitivity, stability and specificity of phenolphthalein as an indicator test for blood, *Canadian Society of Forensic Science Journal* **9**, 97–102.

[33] Laux, D.L. (2005). The detection of blood using luminol, in *Principles of Bloodstain Pattern Analysis: Theory and Practice*, S. James, P.E. Kish & T.P. Sutton, eds, CRC Press, Boca Raton, pp. 369–389.

[34] Chung, H.K., Dasgupta, P.K. & Marx, J.N. (1993). Spectrophotometric determination of H(2)O(2) with 1-anilinonaphthalene-8-sulfonic acid and 4-aminoantipyrine with hematin as catalyst, *Talanta* **40**, 981–988.

[35] Genfa, Z. & Dasgupta, P.K. (1992). Hematin as a peroxidase substitute in hydrogen peroxide determinations, *Analytical Chemistry* **64**, 517–522.

[36] Motsenbocker, M.A., Ichimori, Y. & Kondo, K. (1993). Metal porphyrin chemiluminescence reaction and application to immunoassay, *Analytical Chemistry* **65**, 397–402.

[37] Svistunenko, D.A. (2005). Reaction of heme containing proteins and enzymes with hydroperoxides: the radical view, *Biochimica Et Biophysica Acta* **1707**, 127–155.

[38] Dolphin, D. & Felton, R.H. (1974). Biochemical significance of porphyrin π cation radicals, *Accounts of Chemical Research* **7**, 26–32.

[39] Erdey, L., Pickering, W.F. & Wilson, C.L. (1962). Chemiluminescence of luminol and hemin, *Talanta* **9**, 371–375.

[40] Weber, K. (1966). Die Andwendung der Chemiluminescenz des Luminols in der Gerichtlichen Medizin und Toxicologie. I. Der Nachweis von Blutspuren, *Deutsche Zeitschrift fur die Gesamte Gerichtliche Medizin* **57**, 410–423.

[41] Baj, S. & Krawczyk, T. (2006). An investigation into the reaction of hemin-catalyzed luminol oxidation by peroxy compounds, *Journal of Photochemistry and Photobiology A* **183**, 111–120.

[42] Lytle, L.T. & Hedgecock, D.G. (1978). Chemiluminescence in the visualization of forensic bloodstains, *Journal of Forensic Sciences* **23**, 550–562.

[43] Pex, O.J. (2005). The use and limitations of luminol in bloodstain pattern analysis, *International Association of Bloodstain Pattern Analysts Newsletter*, December, 11–15.

[44] Budowle, B., Leggitt, J.L., Defenbaugh, D.A., Keys, K.M. & Malkiewicz, S.F. (2000). The presumative reagent fluorescein for detection of dilute bloodstains and subsequent STR typing of recovered DNA, *Journal of Forensic Sciences* **45**, 1090–1092.

[45] Vandenberg, N. & Oorschot, R.Av. (2006). The use of polilight® in the Detection of seminal fluid, saliva, and bloodstains and comparison with conventional chemical-based screening tests, *Journal of Forensic Sciences* **51**, 361–370.

[46] Fregeau, C.J., Germain, O. & Fourney, R.M. (2000). Fingerprint enhancement revisited and the effects of blood enhancement chemicals on subsequent Profiler Plus fluorescent short tandem repeat DNA analysis of fresh and aged bloody fingerprints, *Journal of Forensic Sciences* **45**, 354–380.

[47] Grodsky, M., Wright, K. & Kirk, P.L. (1951). Simplified preliminary blood testing: an improved technique and a comparative study of methods, *Journal of Criminal Law, Criminology, and Police Science* **42**, 95–104.

[48] Yamaguchi, M., Yoshida, H. & Nohta, H. (2002). Luminol-type chemiluminescence derivatization reagents for liquid chromatography and capillary electrophoresis, *Journal of Chromatography. A* **950**, 1–19.

[49] Lefebvre, G. (2005). creating luminol photographs with digital imaging – a method for photographing luminol reactions at crime scenes, *International Association of Bloodstain Pattern Analysts Newsletter*, June, 4–7.

[50] Thornton, J.I., Guarino, K., Rios, F.G. & Cashman, P.J. (1986). Enhancement of the luminol test by means of light amplification, *Journal of Forensic Sciences* **31**, 254–257.

[51] Ballou, S.M. (1995). Moves of murder, *Journal of Forensic Sciences* **40**, 675–680.

[52] Kirk, P. (1953). *Crime investigations*, Interscience, New York.

[53] Yeshion, T.E. (1991). The forensic application of luminol as a presumptive blood test, in *Proceedings of the 6th International Symposium on Bioluminescence and Chemiluminescence*, P.E. Stanley & L.J. Kricka, eds, John Wiley & Sons, Chichester, pp. 379–384.

[54] Arnhold, J., Mueller, S., Arnold, K. & Sonntag, K. (1993). Mechanisms of inhibition of chemiluminescence in the oxidation of luminol by sodium hypochlorite, *Journal of Bioluminescence and Chemiluminescence* **6**, 307–313.

[55] Sakagami ea, H. (1995). Hypochlorite scavenging activity of polyphenols, *Anticancer Research* **15**, 917–922.

[56] Triantis, T., Yannakopoulou, E., Nikokavoura, A., Dimotikali, D. & Papadopoulos, K. (2007). Chemiluminescent studies on the antioxidant activity of amino acids, *Analytica Chimica Acta* **591**, 106–111.

[57] Burke, B.A., Golestaneh, K. & Samson, H. (1999). My Dear Watson, Luminosity!-Or, Are Those Bloodstains?, *Journal of Chemical Education* **76**, 65–66.

[58] Ewetz, L. & Thore, A. (1975). Factors affecting the specificity of the luminol reaction with hematin compounds, *Analytical Biochemistry* **71**, 564–570.

[59] Specht, W. (1937). The chemiluminescence of hemin as a means of finding and recognising blood traces of forensic importance, *Angewandte Chemie* **50**, 155–157.

[60] Seitz, W.R. & Hercules, D.M. (1974). A quantitative study of chemiluminescence from the iodine-luminol reaction, *Journal of the American Chemical Society* **96**, 4094–4098.

[61] Seitz, W.R. & Hercules, D.M. (1973). Chemiluminescence analysis for trace elements, in *Chemiluminescence and Bioluminescence*, M.J. Cormier, D.M. Hercules & J. Lee, eds, Plenum Press, New York, pp. 427–449.

[62] Shen, J., Chen, L., Gao, W. & Zhou, Q. (1998). Application of iodine and iodide to chemiluminescence analysis of luminol, *Lihua Jianyan Huaxue Fence* **34**, 132–134.

[63] Creamer, J.I., Quickenden, T.I., Crichton, L.B., Robertson, P. & Ruhayel, R.A. (2005). Attempted cleaning of bloodstains and its effect on the forensic luminol test, *Luminescence* **20**, 411–413.

[64] Quickenden, T.I. & Creamer, J.I. (2001). A study of common interferences with the forensic luminol test for blood, *Luminescence* **16**, 295–298.

[65] Quickenden, T.I. & Cooper, P.D. (2001). Increasing the specificity of the forensic luminol test for blood, *Luminescence* **16**, 251–253.

[66] Creamer, J.I., Quickenden, T.I., Apanah, M.V., Kerr, K.A. & Robertson, P. (2003). A comprehensive experimental study of industrial, domestic and environmental interferences with the forensic luminol test for blood, *Luminescence* **18**, 193–198.

[67] Seitz, W.R. (1975). Chemiluminescence from the reaction between hypochlorite and luminol, *The Journal of Physical Chemistry* **79**, 101–106.

[68] Quickenden, T.I., Ennis, C.P. & Creamer, J.I. (2004). The forensic use of luminol chemiluminescence to detect traces of blood inside motor vehicles, *Luminescence* **19**, 271–277.

[69] Hochmeister, M.N., Budowle, B., Sparkes, R., Rudin, O., Gehrig, C., Thali, M., Schmidt, L., Cordier, A. & Dirnhofer, R.J. (1999). Validation studies of an immunochromatographic 1-step test for the forensic identification of human blood, *Journal of Forensic Sciences* **44**, 597–602.

Related Articles

Biological Swabs

DNA

Short Tandem Repeats

SIMON W. LEWIS AND FILIPPO BARNI

Wiley Encyclopedia of
FORENSIC SCIENCE

Volume 4
F–L

Editors-in-Chief

Allan Jamieson
The Forensic Institute, Glasgow, UK

Andre Moenssens
Forensics and Law Center, Columbia City, IN, USA

WILEY

(c) 2011 John Wiley & Sons, Inc. All Rights Reserved.



Malingering ***see*** Delusions, Hallucinations, Posttraumatic Stress Disorder

Malingering: Assessment of ***see*** Neuropsychological Assessment

Malingering: Forensic Evaluations

Malingering in Forensic Evaluations

The essential feature of malingering, as defined by the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*), is the "intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives such as avoiding military duty, avoiding work, obtaining financial compensation, evading criminal prosecution, or obtaining drugs . . ." [1] (p. 739). Malingering for the purposes of evading criminal prosecution or obtaining drugs are commonly seen in forensic settings and forensic mental health clinicians often face difficult decisions when evaluating and treating these individuals. In civil settings, malingering is often related to obtaining potentially large financial settlements.

In order to appropriately assess and manage malingering, clinicians must have in-depth knowledge of the following: (i) prevalence rates of malingering and how the *Diagnostic and Statistical Manual of Mental Disorders* Fourth Edition, Text Revision (*DSM-IV-R*) inadvertently encourages misdiagnosis and mismanagement of malingering-related issues, (ii) common reasons for countertransference toward malingering patients and strategies to avoid it, (iii) empirically validated measures of malingering, and (iv) clinical issues related with the evaluation of malingering.

Malingering and the DSM

Recent evidence [2–4] indicates that malingering is common in forensic settings with prevalence rates ranging from 10 to 29%. Even with the most conservative estimate of malingering, one thing is clear that forensic mental health practitioners cannot rely on the veracity of their patient's self-reports to establish diagnoses or effectively manage their patient's symptoms. Self-report distortion is especially salient in pretrial forensic evaluations where the stakes can be quite high.

For instance, individuals undergoing an evaluation of their mental status (i.e., insanity) at the time of offense, will, at times, embellish pre-existing psychopathology or fabricate mental health symptoms in

order to avoid prison sentence[5]. Likewise, competency to stand trial evaluees may feign psychopathology in an attempt to avoid trial altogether [6]. Others malinger for some reasons apart from trying to avoid prosecution. Such reasons include moving to a mental health unit where the environment is often viewed as calmer and safer [7] or to acquire psychotropic medications [8].

How clinicians view the issue of malingering can have significant influence on their treatment of bona fide patients and individuals who are feigning. Unfortunately, countertransference (see next section) often accompanies malingering and staff members frequently make assumptions about other aspects of the patient's presentation based on a malingering diagnosis. Current diagnostic nomenclature does little to disabuse professionals of this potential conflict.

The DSM [**1**] criteria indicate that malingering should be suspected if any of the following are present: (i) medicolegal context of presentation, (ii) marked discrepancy between the person's claimed stress or disability and the objective findings, (iii) lack of cooperation during the diagnostic evaluation and in complying with the diagnostic evaluation, and (iv) the presence of Antisocial Personality Disorder. In forensic settings, either one or several of these criteria are routinely present for reasons wholly unrelated to malingering. Reliance on these criteria can lead to the misclassification of patients with bona fide mental illnesses as feigning cases. In fact, exclusive reliance on DSM-IV TR criteria can result in up to 80% misclassification rate [9]. If a test for a medical disorder, such as human immunodeficiency virus (HIV), resulted in high false positive rate, the test would be discontinued forthwith.

Avoiding false positives with malingering cases must be the primary responsibility of the forensic examiner. Evaluators must be cognizant that certain evaluees are more at risk to be misdiagnosed as **malingerers**. Research has shown that bona fide patients with lower educational attainment and absence of previous mental health treatment are much more likely to be misclassified as **malingerers** [10]. Clinicians can avoid false positives by conducting comprehensive evaluations that include multiple data sources, continuing to evaluate for psychopathology or cognitive deficits in the presence of malingering, and not making facile assumptions based on incomplete data.

Malingering and Countertransference

An issue that frequently goes unchecked in malingering evaluations is the misuse and conveyance of diagnostic information. One prime example is the use of emotionally laden words by clinicians to describe a malingerer; a clear sign of countertransference [11]. I have overhead staff using terms like "liar", "fraud", or "fake" when describing an individual found to be malingering. These descriptors are clinically useless and result in untoward feelings toward the patient.

Even more distressing is that longitudinally, a malingering diagnosis can follow the patient and subsequent complaints may be dismissed as bogus and misattributed to malingering. As such, clinicians conducting forensic evaluations have an ethical responsibility to minimize harm by accurately reporting their findings [12]. When malingering is identified, discussing situational-specific aspects of the diagnosis would be the first step toward minimizing future consequences.

One way to avoid countertransference with malingering cases is through the use of explanatory models. Rogers [9, 13] introduced explanatory models as a counterperspective to the unitary perspective provided by the DSM. Rogers proposed three explanatory models: (i) criminological, (ii) adaptational, and (iii) pathogenic.

Criminological

The DSM is focused on the criminological model, which closely aligns malingering with antisocial behavior. The criminological model accounts for only a minority of malingering cases, which explains the high rate of false positives associated with the DSM criteria.

Adaptational

A prototypical analysis of the responses of forensic experts [14] reinforced that the majority of individuals who malinger in psycholegal evaluations do so as an adaptive response to adversarial circumstances. The adaptational model allows for the possibility that people facing difficult circumstances engage in deception designed to decrease the likelihood of a bad outcome (e.g., lengthy prison term).

Pathogenic

The pathogenic model posits that feigning can occur as a result of underlying psychopathology. Although

this model is infrequently used and has mainly fallen into disfavor, there is some support for this model derived from both a historical perspective and surveys of forensic practitioners [15].

Explanatory models are especially useful if clinicians consider alternative explanations for malingering. Using models to understand malingering can provide objectivity to the assessment process and eliminate emotionally laden statements regarding patients who are engaging in deception.

Measures that Detect Malingering of Psychopathology

The use of extant measures for evaluating feigning is necessary for accurate assessment. In reviewing measures used to systematically assess for the malingering of psychopathology, it is clear that clinicians have several available options. Given space constraints, I primarily focus on measures specifically designed to assess malingering, with the exception of the *Minnesota Multiphasic Personality Inventory Second Edition* (*MMPI-2*) [16]. Other multiscale inventories with malingering or response style scales such as the Personality Assessment Inventory (PAI) [17] and the Millon Clinical Multiaxial Inventory (MCMI) [18] are not discussed. It should be noted that both the PAI and MCMI-III possess significant psychometric limitations for detecting feigning, although recent research suggests that the PAI does possess utility for screening feigned presentations [3, 19].

- The Structured Interview of Reported Symptoms (SIRS), [16] is an interview designed to comprehensively assess malingering and related response styles. The SIRS is often employed as the criterion measure in feigning research due to its very strong psychometric properties and exceedingly low false positive rate [16]. Its eight primary scales focus on measuring symptom patterns consistent with malingering include rare, combinations, improbable and absurd, blatant, subtle, severity, selectivity, and reported versus observed. Rogers *et al.* [20] performed a confirmatory factor analysis (CFA) of the SIRS with a large sample of forensic referrals. The results obtained strongly support for a two-factor dimensional model: Spurious Presentation and Plausible Presentation. "Spurious Presentation" was consistent with "classic" malingering made up of atypical and rare symptoms. "Plausible Presentation" was defined by over-reporting actual symptoms of mental illness and issues related to psychological maladjustment. Considering these two factors, Spurious Presentation was focused on the overall content of the report while Plausible Presentation was focused on the magnitude of report.
- The Miller Forensic Assessment of Symptoms Test (M-FAST, [21]) is a 25-item structured interview that consists of seven scales designed to be used to screen for malingering. The M-FAST is closely modeled after the SIRS and consists of seven scales and a total score; the total score is especially effective for identifying patients who are potentially malingering and in need of a more thorough assessment [3]. The M-FAST was subjected to a CFA in order to assess if the two-factor structure found in the SIRS was applicable. Vitacco *et al.* [22] tested models of the M-FAST with 244 forensic patients and cross-validated those findings with 210 forensic patients. Results demonstrated that a single, parsimonious factor emerged that was closely aligned with the SIRS' Spurious Presentation ($r = 0.75$) and moderately related to Plausible Presentation ($r = 0.64$). The clear implication is that the M-FAST is most effective for identifying patients with atypical and rare symptoms.
- The Structured Inventory of Malingered Symptomatology [SIMS, [23]] is a 75-item true/false test composed of five scales: low intelligence, affective disorders, neurological impairment, psychosis, and amnesia. The SIMS evaluates a wide range of feigning aspects, including aspects of cognitive feigning. With the SIRS as a criterion for known-groups design, the SIMS effectively differentiated malingerers from nonmalingerers (*M* Cohen's $d = 2.25$) with excellent reliability estimates [3]. However, the use of self-report measure has inherent limitations in that aspects of interpersonal presentation are not evaluated. Given that interpersonal interactions are important in evaluating malingering, this limitation is noteworthy.
- The MMPI-2 [16] is a 567-item multiscale inventory designed to assess psychopathology. In addition to its clinical scales, the MMPI-2 contains specialized scales designed to evaluate issues related to response styles. These scales use validated strategies for assessing response styles

including (i) rare symptoms, (ii) erroneous stereotypes, (iii) symptom severity, or (iv) obvious and subtle symptoms. Rogers *et al.* completed a large meta-analysis of malingering on the MMPI-2 [24] and generally found robust effect sizes for many malingering scales. However, clinicians should be cautious in using the MMPI-2 to assess malingering for a wide range of cut scores that dramatically vary by disorder and study [24]. Research [25, 26] evaluating the Infrequency (F) and Infrequency Psychopathology [F(p)] scales with large groups of mental health and criminal-forensic patients found the presence of an underlying taxon for both scales.

Clinical Issues Related to Evaluating Malingering

A key consideration in the accurate assessment of malingering is the realization that mentally ill individuals can engage in deception. This is referred to as *partial malingering* [18], which is defined as the exaggeration of current symptomatology or reporting symptoms from a previous episode of mental illness. Partial malingering is common in forensic settings and can contribute to error when clinicians fail to recognize the presence of underlying mental disorder(s). Obtaining a thorough psychosocial history from the patient, interviewing collateral sources, and reviewing available records are useful tools in disentangling bona fide illness from feigned symptoms.

One potentially useful strategy to disentangle pathology and response styles is to employ a structured interview such as the schedule of affective disorders and schizophrenia-change version [(SADS-C), [27]]. The SADS-C provides an extensive evaluation of psychopathology [28], it has two scales (symptom combinations and symptom selectivity) that act as an effective screen for malingering [28, 29]. Ultimately, success may rely on the clinician's overall knowledge of psychopathology and response styles and the ability to integrate conflicting information regarding mental disorders and feigned presentation.

Another issue adding to the complexity of diagnosing malingering is the differentiation of malingering from factitious disorders. Similar to malingering, factitious disorders involve the intentional production of physical or psychological symptoms [1]. The key difference is the goal of factitious disorders is to assume the sick role in the absence of other incentives. Clinicians often struggle with assessing motivation and trying to determine the underlying rationale for the behavior. Differential diagnoses between malingering and factitious disorders should occur in two related steps: First, determine the occurrence of feigning using a validated instrument. If feigning is present, clinicians are then supposed to determine the underlying motivation [15, 30].

Neurocognitive Malingering

Malingering can extend into civil settings as well. Unlike forensic settings where individuals malinger to avoid trial or prison sentence, individuals often malinger in civil settings to receive monetary compensation. A prime example is personal injury litigation whereby an individual alleges that he or she received an injury of the mind or body. Given that compensation for personal injuries can be highly lucrative it is not surprising that rates of malingering in such cases have ranged from 20 to 35% [31].

Clearly the information presented earlier in this article regarding countertransference and assessment strategies remain salient; however, evaluating malingering in civil situations requires additional skill sets including assessing malingering on neuropsychological examinations (e.g., alleging brain damage). Malingering neurocognitive deficits is commonly seen when evaluating individuals in civil settings.

Cognitive Malingering

To understand cognitive malingering and know how to best diagnosis it, clinicians are advised to familiarize themselves with the criteria for malingering neurocognitive dysfunction (MND) functioning known as the *Slick criteria* [32]. The Slick criteria set forth in a seminal article [32] provided a definition of MND as "the volitional exaggeration or fabrication of cognitive dysfunction for the purpose of obtaining substantial material gain, or avoiding or escaping formal duty or responsibility. . ." (p. 552). The article allowed for gradients in MND ranging from definite to possible. These criteria are well reasonable and provide a fundamental foundation for diagnosing MND. To aid in the diagnostic process, measures specifically developed for MND are marketed and should be considered when cognitive malingering is expected.

Measures that Detect Cognitive Malingering

For the testing of MND, clinicians have many options; we focus only on few well-validated and frequently used measures due to space limitations. These measures often use chance responding for a forced-choice test and performing significantly lower than chance is indicative of poor effort and potential malingering. However, more sophisticated measures like the Validity Indicator Profile (VIP) [33] assesses for several issues relevant to MND and represent a major advancement in assessing for MND.

- Rey 15-Item Test consists of a 3×5 alphanumeric sequence that an individual is asked to reproduce. This is a simple measure that has continued to be very popular [34]; however, such popularity seems overstated and it may be best used to screen for cognitive malingering [35] as there have been reports of poor sensitivity, but good specificity. As such, this test does not appear appropriate for evaluating MND. An important consideration is not to use the test for individuals with mental retardation [35].
- Test of Memory Malingering [TOMM, [36]] is a forced-choice test, where the individual is presented with a series of pictures and later "forced" to pick which ones they previously viewed. The TOMM has shown some promise for use in forensic and civil evaluations [37, 38]; however, its cut score has been deemed to be high for some individuals with actual cognitive dysfunction and poor effort could be misinterpreted as malingering. Clinicians must differentiate reasons for poor performance, as low scores might be indicative of suboptimal effort.
- VIP [33] is a self-report test that has both verbal and nonverbal sections. The VIP has several advantages over its counterparts including three providing scores on three possibilities for poor performance: inconsistent, irrelevant, and suppressed. The latter is consistent with what would be considered MND. Over 1000 individuals were subjected to initial validation and 312 cases were cross validated [33]. The VIP has three scales, Inconsistent, Irrelevant, and Suppressed, to detect atypical performance. Notably, the Suppressed scale indicates formal malingering. The ability to differentiate among random responding, irrelevant responding, and malingering is the real strength of the VIP. The VIP has demonstrated excellent properties in evaluating cognitive malingering [39, 40].

Conclusions

Mental health practitioners are encouraged to receive specialized training to understand the complexities of malingering as fully as possible. In the context of assessing malingering, clinicians should strive for diagnostic accuracy with the following:

1. Use only reliable and valid instruments with a proven track record of successfully differentiating bona fide patients from individuals engaging in deception.
2. Untangle psychopathology and true cognitive deficits from malingering through comprehensive assessments that include an evaluation of both response styles and actual deficits. Improve diagnostic accuracy through the use of collateral sources and extensive record reviews to buttress psychological assessment findings.
3. Use explanatory models, especially the adaptational and pathogenic models, to generate an understanding of situational aspects of feigning.
4. Understand that malingering is not "all or none" and should be viewed on a continuum. A prime example is partial malingering where an individual with a bona fide illness exaggerates his or her symptoms.

References

[1] American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition, Text Revision (**DSM IV-R**), American Psychiatric Association Washington, DC.

[2] Cornell, D.G. & Hawk, G.L. (1989). Clinical presentation of malingerers diagnosed by experienced forensic psychologists, *Law and Human Behavior* **13**, 375–383.

[3] Vitacco, M.J., Rogers R., Gabel, J. & Munizza J. (2007). An evaluation of malingering screens with competency to stand trial patients: A known-groups comparison, *Law and Human Behavior*, **31**(3), 249–260.

[4] Boccaccini, M.T., Murrie, D.C. & Duncan, S.A. (2006). Screening for malingering in a criminal-forensic simple with the Personality Assessment Inventory, *Psychological Assessment* **18**, 415–423.

[5] Rogers, R. & Shuman, D.W. (2005). *Fundamentals of Forensic Practice: Mental Health and Criminal Law*, Springer, New York, NY.

[6] Rogers, R., Sewell, K.W., Grandjean, N.R. & Vitacco, M.J. (2002). The detection of feigned mental disorders on specific competency measures, *Psychological Assessment* **14**, 177–183.

[7] Vitacco, M.J. & Rogers, R. (2005). Assessment of malingering in correctional settings, in *Handbook of Correctional Mental Health*, C.L. Scott & J.B. Gerbasi, eds, American Psychiatric Press, Washington, DC, 133–154.

[8] Yates, B.D., Nordquist, C.R. & Schultz-Ross, A.R. (1996). Feigned psychiatric symptoms in the emergency room, *Psychiatric Services* **47**, 998–1000.

[9] Rogers, R. (1990). Models of feigned mental illness, *Professional Psychology* **21**, 182–188.

[10] Kucharski, T.L. & Duncan, S.A. (2006). Clinical and demographic characteristics of criminal defendants potentially misidentified by objective measures of malingering, *American Journal of Forensic Psychology* **24**, 5–20.

[11] Goldyne, A.J. (2007). Minimizing the influence of unconscious bias in evaluations: A practical guide, *Journal of the American Academy of Psychiatry and Law* **35**, 60–66.

[12] Ackerman, M. (2006). Forensic report writing, *Journal of Clinical Psychology* **62**, 59–72.

[13] Rogers, R., Sewell, K.W. & Goldstein, A. (1996). Explanatory models of malingering: A prototypical analysis, *Law and Human Behavior* **18**, 543–552.

[14] Rogers, R., Salekin, R.T., Sewell, K.W., Goldstein, A. & Leonard, K. (1998). A comparison of forensic and nonforensic malingerers: A prototypical analysis of explanatory models, *Law and Human Behavior* **22**, 353–367.

[15] Rogers, R., Bagby, R.M. & Dickens, S.E. (1993). *Structured Interview of Reported Symptoms, Professional Manual*, Psychological Assessment Resources, Odessa, FL.

[16] Butcher, J.N., Dahlstrom, W.G., Graham J.R., Tellegen A. & Kaemmer B. (1989). *Minnesota Multiphasic Personality Inventory-Second Edition Manual for administration and scoring*, University of Minnesota Press, Minneapolis, MN.

[17] Morey, L.C. (1991). *Personality Assessment Inventory*, Psychological Assessment Resources, Tampa, FL.

[18] Millon, T. (1994). *Millon Clinical Multiaxial Inventory-Third Edition (MCMI-III) Manual*, National Computer Systems, Minneapolis, MN.

[19] Hopwood, C.S., Morey, L.C., Rogers, R. & Sewell, K.W. (2007). Malingering on the personality assessment inventory: identification of specific feigned disorders, *Journal of Personality Assessment* **88**, 43–48.

[20] Rogers, R., Jackson, R.L., Sewell, K.W. & Salekin, K.L. (2005). Detection strategies for malingering: A confirmatory factor analysis of the SIRS, *Criminal Justice and Behavior* **32**, 511–525.

[21] Miller, H.A. (2001). *M-FAST: Miller-Forensic Assessment of Symptoms Test professional manual*. Psychological Assessment Resources, Odessa, FL.

[22] Vitacco, M.J., Jackson, R., Rogers, R., Neumann, C.S., Miller, H., Gabel, J. (2008). Detection strategies for malingering with the Miller Forensic Assessment of Symptoms Test: A confirmatory factor analysis of underlying dimensions, *Assessment*, **15**(1), 97–103.

[23] Widows, M.R. & Smith, G.P. (2004). *SIMS: Structured Inventory of Malingered Symptomatology Professional Manual*, Psychological Assessment Resources, Odessa, FL.

[24] Rogers, R., Sewell, K.W., Martin, M.A. & Vitacco, M.J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2, *Assessment* **10**, 160–177.

[25] Strong, D.R., Greene, R.L. & Schinka, J.A. (2000). A taxometric analysis of MMPI-2 infrequency scales [F and F(p)] in clinical settings, *Psychological Assessment* **12**, 166–173.

[26] Strong, D.R., Glassmire, D.M., Frederick, R.I. & Greene, R.L. (2006). Evaluating the latent structure of the MMPI-2 F(p) scale in a forensic sample: A taxometric analysis, *Psychological Assessment* **18**, 250–261.

[27] Spitzer, R.L. & Endicott, J.. *Schedule of Affective Disorders and Schizophrenia-Change Version*, Biometrics Research, New York.

[28] Rogers, R., Jackson, R.L., Salekin, K.L. & Neumann, C.S. (2003). Assessing Axis I symptomatology on the SADS-C in two correctional samples: The validation of subscales and a screen for malingered presentations, *Journal of Personality Assessment* **81**, 281–290.

[29] Rogers, R. (2001). *Handbook of Diagnostic and Structured Interviewing*, Guilford Press, New York, NY.

[30] Vitacco, M.J. (2008). Syndromes associated with deception, in *Clinical Assessment of Malingering and Deception*, 3rd Edition, R. Rogers, The Guilford Press, New York, NY, 39–50.

[31] Mittenberg, W., Patton, C., Canyock, D. & Condit, D. (2002). Base rates of malingering and symptom exaggeration, *Journal of Clinical and Experimental Neuropsychology* **24**, 1094–1102.

[32] Slick, D.J., Sherman, E. & Iverson, G.I. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research, *The Clinical Neuropsychologist* **13**, 545–561.

[33] Frederick, R.I. (1997). *Validity Indicator Profile, Test Manual*, Pearson Assessments, Minneapolis, MN.

[34] Slick, D.J., Jing, T.E., Strauss, E.H. & Hultsch, D.F. (2004). Detecting malingering: A survey of experts' practices, *Archives of Clinical Neuropsychology* **19**, 465–473.

[35] Reznek, L. (2005). The Rey 15-item memory test for malingering: A meta-analysis, *Brain Injury* **19**, 539–543.

[36] Tombaugh, T.N. (1996). *Test of Memory Malingering Professional Manual*, Multi-health Systems, New York.

[37] Greve, K.W., Bianchini, K.J. & Doane, B.M. (2006). Classification accuracy of the Test of Memory Malingering in traumatic brain injury: Results of a known-groups analysis, *Journal of Clinical and Experimental Neuropsychology* **28**, 1176–1190.

[38] Delain, S.L. & Ben-Porath, Y.S. (2003). Use of the TOMM in a criminal court assessment setting, *Assessment* **10**, 370–381.
[39] Frederick, R.I. (2002). Review of the Validity Indicator Profile, *Journal of Forensic Neuropsychology* **2**, 125–145.
[40] Frederick, R.I., Crosby, R.D. & Wynkoop, T.F. (2000). Performance of curve classification of invalid responding on the Validity Indicator Profile, *Archives of Clinical Neuropsychology* **15**, 281–300.

MICHAEL J. VITACCO

Malingering: Mental Retardation

see Mental Retardation: Death Penalty

Malpractice: Medical *see* Medical Malpractice

Malpractice Actions against Experts

Introduction

The term "malpractice," when describing expert behavior in professional dealings with third parties, has several connotations. In its popular understanding, it may be seen as a characterization of either negligent or intentional conduct that is considered harmful, inequitable, or fraudulent.

In professional practice, the term describes conduct that may expose the person guilty of such conduct to professional criticism by his peers in terms of sanctions within professional organizations to which the expert belongs. These sanctions may result in public or private reprimands, suspension or loss of membership privileges, expulsion from professional societies, and temporary or permanent loss of credentials or certifications. This form of malpractice is discussed in the separate article of **Ethics: Codes of Conduct for Expert Witnesses**.

The connotation in which the word "malpractice" is used in this article is as a legal term used in common law jurisdictions, particularly in the United States of America, to describe unlawful conduct that exposes an expert to possibly three different consequences: (i) a civil court action for damages caused to third parties while conducting professional duties; (ii) a criminal action brought by the government to punish defined criminal conduct such as fraud, theft, obstruction of justice, contempt of court, or similar crimes; or (iii) a quasi-criminal action brought by a governmental entity to restrict or remove a license to conduct a professional practice granted by public officials.

Since legal systems differ depending upon the country or jurisdiction, an in-depth exploration of malpractice becomes impossible within the confines of this work. For that reason, the possible legal consequences of professional misconduct are described only briefly here with reference to approaches taken by American courts and in the context of civil actions brought against experts.

Expert Malpractice – a Problem of National Scope

The development of a new cause of action designed to hold expert witnesses, like forensic scientists, doctors, and lawyers, responsible for their negligent professional behavior [1] has been driven by the growing recognition that expert negligence is not uncommon [2]. High-profile incidents have revealed failures in the application of some of the most well-established scientific techniques such as fingerprint identification [3, 4], DNA testing [5], and serologic analysis [6]. Other investigations have demonstrated that pathologists faked hundreds of autopsies [7] or committed grievous errors [8] in determining cause of death.

At present, the law does little to regulate the quality of expert testimony [9]. Nontort solutions offered by the scientific and legal communities to curb expert abuses include capping expert witness fees [10]; prescreening experts; using only court-appointed experts [11]; adhering to strict ethics codes [12]; instituting peer review [13]; and establishing a

science court [14]. Additionally, it has been suggested that fraudulent experts be prosecuted [15].

The conventional wisdom long held that the principal safeguard against errant expert testimony is the opportunity for adversarial cross-examination [16]. In reality, however, most lawyers do a woefully inadequate job in cross-examining experts [17] due to a general reluctance to challenge experts in their own fields and improper trial preparation [18]. Finally, the vast majority of civil and criminal cases are settled or plea bargained prior to trial so that the expert may never be subjected to rigorous questioning during the adversary process.

To date, none of the nontort solutions offered to curb expert abuses have succeeded in accomplishing their goal. While attempts at self-regulation and court supervision may provide some deterrence, they do not compensate individuals harmed by negligent experts. This lack of effective solutions to expert negligence or intentional professional misconduct has led to the developing area of expert malpractice tort law.

The Tort of Expert Malpractice

The expert witness malpractice cause of action offers the most effective means of achieving the twin goals of compensating injured individuals and deterring future misconduct. As a result, the last two decades have seen a substantial increase in the number of tort actions against experts and their employers.

The four elements of an expert malpractice claim are (i) the existence of a duty owed to the plaintiff arising out of the relationship between the expert and the plaintiff; (ii) a negligent act or omission by the expert in breach of that duty; (iii) causation; and (iv) damages [19].

The premise of the cause of action is that, first, expert witnesses owe a duty to their clients. However, the duty does not end there. Expert witnesses also owe a duty to any foreseeable plaintiff who may be affected by the expert's conduct and who are likely to suffer damages due to a negligently rendered opinion. These duties are based upon their professional knowledge and skills and are similar to the duties owed by a doctor to a patient and a lawyer to a client. In this way, the specter of malpractice encourages experts to be careful, accurate, and compliant with "quality control" measures.

The standard of care for a forensic scientist is that of the reasonably prudent practitioner in the relevant scientific field [20]. Standards of professional practice and ethical codes help define the standard of care, and most disciplines within the forensic sciences have adopted such standards. In order to prevail, a plaintiff must prove that the expert did not adhere to the standard of a reasonably prudent expert in rendering an opinion, conducting an examination, or giving testimony. Ordinarily, an independent evaluation by a disinterested expert in the same field will be required to determine whether an expert deviated from the required standard of care.

A crucial element of the tort of malpractice is causation. Causation tests whether the defendant's actions were in fact connected to the plaintiff's injury, and whether the connection was close enough to allow compensation to the injured party. In some cases, it will be readily apparent that an expert's testimony alone "caused" the wrong. This is especially true when the expert evidence is the only determinative evidence presented in the litigation [21]. Studies have demonstrated that, despite jury instructions to the contrary, jurors give expert testimony greater weight than other evidence [22]. Thus, it is clear that financial injury to a potential plaintiff or conviction and incarceration of a potentially innocent individual who is prosecuted on the basis of an expert's opinion evidence,[a] are reasonably foreseeable consequences of expert negligence or intentional misconduct.

When a plaintiff proves that an expert has committed malpractice, the measure of damages that may be awarded include, though are not limited to (1) the difference between a full verdict of proven loss and the reduced verdict resulting from the expert's testimony; (2) the difference between a full settlement and the reduced settlement that resulted from the expert's misconduct; (3) the cost of the expert's investigation; and (4) the attorneys' fees for responding to the expert's testimony and in proving the misconduct [23].

Although traditionally experts were afforded absolute immunity in trial testimony and trial preparation, expert witness malpractice causes of action are gaining momentum [24]. Courts in New Jersey [25], Connecticut [26], Texas [27], California [28], Pennsylvania [29], Massachusetts [30], Louisiana [31], Vermont [32], and Missouri [33] are among the growing number of jurisdictions that have allowed plaintiffs to sue experts for malpractice. This trend has induced defendants to settle cases even when

the jurisdiction has yet to recognize the cause of action [34].

Despite this trend, some jurisdictions continue to adhere to a policy of absolute immunity for expert witnesses [35]. Although witness immunity is an exception to the general rules of liability, and is traditionally extremely narrow in scope, a few courts have nevertheless shielded experts from civil liability for ordinary negligence by reasoning that (i) negligent mistakes or inaccuracies do not constitute perjury, or (ii) testimony and reports provided to courts are privileged [36]. Other courts have held that the expert witness who gives opinion evidence is the court's witness, and therefore enjoys immunity against all post-trial damage claims whether sued by a party or non-party to the action [37]. Such limitations are increasingly rare, however, and no court shields erring expert witnesses from perjury charges for willful deceptions, or from damage actions where the expert's conduct involved intentional or grossly negligent conduct.

There is a compelling argument to be made against expert immunity. First, the doctrine of immunity was not created to bar a suit against a professional who negligently performs services [38]. Moreover, when an expert is accused of malpractice, the real complaint is not with the testimony provided in court, but rather with the negligently produced out-of-court work product. By testifying, the expert merely publishes his negligence to the court. Absolute immunity should not be afforded to experts, who are neither judges nor their adjuncts, but merely third party participants in litigation. And the courts, the legal profession, and the forensic disciplines recognize that the trend is firmly toward permitting claims for damages resulting from negligent expert testimony.

Concerns that the proliferation of expert malpractice suits will have a chilling effect on the supply of willing forensic experts are misplaced. While the emergence of such a cause of action may chase the habitually negligent or incompetent expert from the field, this is, of course, a salutary by-product of the legal trend. Any additional impact on the supply of experts, or in the fees charged for their services, is not so compelling as to justify a public policy against recognizing the cause of action.

Conclusion

The interests of our system of justice in expert accountability and the full and accurate development of evidence in civil and criminal litigation are not served by protecting the incompetent or dishonest expert. The justice system as a whole benefits when expert malpractice actions are permitted, and the forensic sciences should enjoy greater respect and admiration when it is known that their members are accountable for their misdeeds.

While not universally recognized in American courts, malpractice actions against experts have been permitted in more and more jurisdictions and are occurring in greater numbers. This trend is likely to continue in the foreseeable future.

End Notes

[a.] Courts have awarded plaintiffs damages for illegal confinement due to legal malpractice, rejecting the argument that estimating the value of a person's loss of liberty is speculative. Geddie v. St. Paul Fire and Marine Ins. Co., 354 So. 2d 718 (La. App. 1978); Holliday v. Jones, 264 Cal. Rptr. 448 (Cal. App. 4 Dist. 1989) (awarding damages for emotional distress as a result of wrongful incarceration due to professional malpractice); In re Investig. of W. Va. St. Police Crime Lab, 438 S.E.2d, 509 (underlying civil suit settled for the state's $1 million insurance policy limit); see also, Restatement (Third) of the Law Governing Lawyers: Liability for Professional Negligence and Breach of Fiduciary Duty § 53 cmt. (g) (2000) ("emotional distress damages are . . . ordinarily recoverable when misconduct causes a client's imprisonment").

References

[1] Hanson, R.K. (1996). Witness immunity under attack: disarming "Hired Guns", *Wake Forest Law Review* **31**, 497, 508–509; (Reviewing arguments in support of and in opposition to witness immunity for experts). See also Henderson Garcia, C. (1991). Expert witness malpractice: a solution to the problem of the negligent expert witness, *Mississippi College Law Review* **12**, 39.

[2] Cooley, C.M. (2004). Reforming the forensic science community to avert the ultimate injustice, *Stanford Law and Policy Review* **15**, 381, 395–396 (Listing 21 recent exonerations where the underlying conviction was based on inaccurate or completely false testimony by forensic examiners). See also Giannelli, P.C. (2002). Fabricated reports, *Criminal Justice* **16**, 49 (discussing some of the recent "experts" who consistently fabricated reports and testimonies to achieve a particular bias).

[3] Stacey, R.B. (2004). A report on the erroneous fingerprint individualization in the Madrid train bombing case, *Journal of Forensic Identification* **54**, 706.

[4] Wax, S.T. & Schatz, C.J. (2004). A multitude of errors: The Brandon Mayfield case, *The Champion Magazine*, Sept./Oct.,6.

[5] Murray v. State, 692 So.2d 157, 159 (Fla. 1997); (Finding that State's expert "affirmatively misled" the trial court as to the general acceptance or reliability of PCR DNA methodology). See also Madigan, N. (2003). Houston's troubled DNA crime lab faces growing scrutiny, *The New York Times*, I20 (Detailing the results of an audit of the police DNA lab that found a number of problems with its methods; in response to the audit, the district attorney ordered a review of all convictions based on the lab's DNA analysis).

[6] State v. Woodall, 385 S.E.2d 253 (1989). Overturning rape conviction where serologist's evidence shown to be erroneous; In re Investig. of W. Va. St. Police Crime Lab., 438 S.E.2d 501, 509 (W. Va. 1993) (Report recommending that "[d]ue to the undisputed nature of the overwhelming evidence of misconduct on the part of [state serologist] Zain, . . ." 134 prisoners and parolees in whose cases the serologist had testified should be permitted to file petitions for post-conviction habeas corpus).

[7] Fricker, R.L. (1993). Pathologist's plea adds to turmoil, *ABA Journal* **79**, 24, 24. Fricker, R.L. (1993). Reasonable doubts, *ABA Journal* **79**, 39, 44 (the pathologist was reported to have a "reputation for providing the type of forensic evidence prosecutors needed," though his conclusions were later deemed "impossible" by qualified reviewing medical examiners).

[8] Nordheimer, J. (1993). In New Jersey slip-ups show autopsy system deficiencies, *The New York Times*, A1. This article describes a number of flawed autopsies by county medical examiners in two different New Jersey counties, including one case where a pathologist described bullet entrance and exit wounds, and its track through the brain, where it was later established that death was due to "blunt force injury" and that no evidence of a bullet wound existed, and a second case where a medical examiner concluded that a woman had died from alcohol poisoning and exposure, when a later autopsy established the woman had been strangled and raped.

[9] Peterson, J.L. & Murdock, J.E. (1989). Forensic science ethics: developing an integrated system of support and enforcement, *Journal of Forensic Sciences* **34**, 749. See also Austin v. American Assn. of Neurological Surgeons, 253 F.3d 967, 973 (7th Cir., 2001) ("[I]t is well known that expert witnesses are often paid very handsome fees, and common sense suggests that a financial stake can influence an expert's testimony, especially when the testimony is technical and esoteric and hence difficult to refute in terms intelligible to judges and jurors. More policing of expert witnesses is required, not less").

[10] Iowa Code Ann. § 622.72 (West 2006).

[11] Schroeder, O.C. & LeBlang, T.R. (2000). Court appointed experts, *Forensic Sciences* **1**, 18–1, 18–1–18–20, Cyril Wecht, Ed. Matthew Bender, N. Y. Explaining the role of the court-appointed expert; the statutory authorization for the court-appointed expert; the procedure for the appointment of the expert; the methods to discover the court-appointed expert's opinion; the weight accorded to the court-appointed expert's testimony; the use of court-appointed experts in civil and criminal trials; and finally, the constitutional issues related to appointing an expert; but see Struve, C.T. (2004). Doctors, the adversary system, and procedural reform in medical liability litigation, *Fordham Law Review* **72**, 943 (discussing a 1998 survey that indicated that in litigation involving complicated technical or scientific testimony only 16% of federal judges used court-appointed experts).

[12] http://www.cacnews.org/membership/handbook2006.pdf (Last accessed on Oct. 8. 2008).

[13] Carter, T. (2004). M.D. with a mission: a physician battles against colleagues he considers rogue expert witnesses, *ABA Journal* **90**, 40, 42. Advocating peer review in the courtroom for medical expert witnesses.

[14] Field, T.G., Jr., Kantrowitz, A., Cranor, C.F., Jacoby, I., Jasanoff, S., Mazur, A. & Cavicchi, J.R. (1993). Twenty-five year retrospective on the science court: a symposium, *Risk* **4**, 95–188. Containing a series of articles by advocates and detractors of the science court, including one by its "inventor"; see also Timmerbeil, S. (Spring 2003). The role of expert witnesses in German and U.S. civil litigation, *Annual Survey of International and Comparative Law* **9**, 163, 170 (discussing several proposals, including the science court, that addressed the concern that hired experts hindered the truth-seeking process of courts).

[15] In re Investig. of W. Va. St. Police Crime Lab, 438 S.E.2d, 501–509 (WV 1993). Harper, J. (1994). West Virginia court wants forensics expert prosecuted, *Houston Post* A22.

[16] Trower v. Jones, 520 N.E.2d 297 (Ill. 1988).

[17] Dowd, K.M. (1988). Expert Witnesses: Criminologist in the Courtroom, *The New England Journal on Criminal and Civil Confinement* **14**, 169, 171. Reviewing Anderson, P.A. & Winfree, L.T. (1987). *Expert Witnesses: Criminologist in the Courtroom*; Convicted by *Juries, Exonerated by Science: Cases Studies in the Use of DNA Evidence to Establish Innocence After Trial*, NIJ Research Report (June 1996) (Arguing that even improved science will not remedy the problem of inadequate legal counsel; of twenty-eight cases addressed in the study, had defense counsel sought the opinion of a competent expert or simply reviewed the case notes of the state's expert witnesses prior to trial, then the inconsistencies and inadequacies of these flawed testimonies could have been brought to light during the trial) (Available at http://www.ncjrs.org/txtfiles/dnaevid.txt (last visited on February 13, 2008)).

[18] Haddad, F.E. (1996). Admissibility of expert testimony, *Forensic Sciences* **1**(36), 1–21, 1–23, Cyril Wecht, Ed., Matthew Bender, N.Y. Stressing that the keys to an effective cross examination are (1) preparation and (2) becoming knowledgeable in the particular field.

[19] Keeton, W.P., Dobbs, D.B., Keeton, R.E. (1984). *Prosser and Keeton on Torts*, § 30, 5th Edition, Owen, D.G. (eds), West. 164–165. Dobbs, D.B. The Law of Torts § 114, 269 (West 2000); see also Daerr-Bannon, K.L. (2003). Cause of action for negligence or malpractice of expert witness, *Causes of Action 2d* **17,** 263 (adding that in a malpractice case against a friendly expert, the plaintiff must also prove that witness immunity is not applicable to the facts to avoid dismissal).

[20] LLMD of Michigan, Inc. v. Jackson-Cross Co., 740 A.2d 186, 191 (Pa. 1999). [t]he judicial process will be enhanced only by requiring that an expert witness render services to the degree of care, skill and proficiency commonly exercised by the ordinarily skillful, careful and prudent members of their profession; Masterson, L.R. (1998). Witness immunity or malpractice liability for professionals hired as experts? *The Review of Litigation* **17**, 393, 393.

[21] Frank, R.S. (1987). The essential commitment for a forensic scientist, *Journal of Forensic Sciences* **32**, 5. [t]he impact of the forensic scientist's conclusions affords no room for error, because such an error may be the direct cause of an injustice.

[22] Ludwig, K. & Fontaine, G. (1978). Effect of witnesses' expertness and manner of delivery of testimony on verdicts of simulated jurors, *Psychological Reports* **42**, 955. There are many cases that express concern that the special aura of reliability and credibility that surrounds an expert witness will cause the jury to neglect their fact-finding role. State v. Johnson, 681 N.W.2d 901, 906 (Wis. 2001); State v. Ward, 138 S.W.3d 245, 270 (Tenn. Crim. App. 2003); Franco v. State, 25 S.W.3d 26, 29 (Tex. App. 2000).

[23] Hansen, M. (2000). Experts are liable, too, *ABA Journal* **86**, 17, 17.

[24] Davis v. Wallace, 565 S.E.2d 386, 389–90 (W. Va. 2002).

[25] Levine v. Wiss & Co., 478 A.2d 397, 399 (N.J. 1984). Denying a court-appointed expert witness immunity; Weiss, L.S. (2004). Expert witness malpractice actions: emerging trend or aberration? *Practical Litigator* **15**(2), 27, 37 (discussing cases from other jurisdictions that recognize witness immunity if the expert was appointed by the court).

[26] Pollock v. Pahjabi, 781 A.2d 518 (Conn. Super. Ct. 2000).

[27] James v. Brown, 637 S.W.2d 914 (Tex. 1982).

[28] Mattco Forge, Inc. v. Arthur Young & Co., 60 Cal. Rptr. 2d 780 (Cal. App. 2 Dist. 1997).

[29] LLMD of Michigan, Inc., 740 A.2d 191 (Pa. 1999).

[30] Boyes-Bogie v. Horvitz, 14 Mass. L. Rptr. 208 (Mass. Super. 2001).

[31] Marrogi v. A.A. Mathews, 805 So. 2d 1118 (La. 2002).

[32] Politi v. Tyler, 751 A.2d 788 (Vt. 2000).

[33] Murphy v. A.A. Mathews, 841 S.W.2d 671 (Mo. 1992).

[34] DeBenedictis, D. (1994). Off-target opinions, *ABA Journal* **80**, 76, 76. Hospital and national drug laboratory settled on the eve of trial for undisclosed sums for misdiagnosing toxins in a baby's blood, which had resulted in the mother's murder conviction and imprisonment; In re Investigation of West Virginia State Police Crime Lab, *supra*.

[35] Bruce v. Byrne-Stevens & Associates Engineers, Inc., 776 P.2d 666 (Wash. 1989) Weiss, L.S. *Practical Litigator* N. **15**(2), 217, 30–31 (listing Washington as the only state where witness immunity still controls "friendly" experts and providing a detailed description of the *Bruce* decision); Diehl v. Danuloff, 618 N.W.2d 83 (Mich. App. 2000)(court-appointed expert enjoys quasi-judicial immunity as "an arm of the trial court"); Otero v. Warnick, 614 N.W.2d 177 (Mich. App. 2000)(holding that forensic odontologist for county medical examiner was not liable to a former criminal defendant because (1) as an employee of the medical examiner, the expert owed no duty to criminal defendants in performing his official duties, and (2) expert's testimony at trial was absolutely privileged provided it was relevant, material, and pertinent to the issue being tried).

[36] Saks, M.J. (1989). Prevalence and impact of ethical problems in forensic science, *Journal of Forensic Sciences* **34**, 772. Containing a summary of some cases involving litigation against expert witnesses.

[37] Bailey v. Rogers, 631 S.W.2d 784 (Tex. App. 3 Dist. 1982).. Compare Mattco Forge, Inc. v. Arthur Young & Co., 6 Cal. Rptr. 2d 781 (Cal. App. 2 Dist. 1992), on appeal after remand, 45 Cal. Rptr. 2d 581 (Cal. App. 2 Dist. 1995), on subsequent appeal, 60 Cal. Rptr. 2d 780 (Cal. App. 2 Dist. 1997) (holding that the California Civil Code's litigation privilege does not protect a negligent expert witness from liability to the party who hired the witness, though it would still shield experts that are court appointed, and would also shield expert witnesses from suit by opposing parties; Murphy, 841 S.W.2d at 679 (also holding that under Missouri law, privilege does not protect a negligent expert witness from liability to the party who hired the witness, though it would still shield experts that are court appointed, and would also shield expert witnesses from suit by opposing parties); Marrogi, 805 So. 2d, 1132 (broadening the scope of expert witness malpractice to include not only pretrial litigation services but also the expert's actual testimony during trial).

[38] Murphy, 841 S.W.2d 679 (Mo. 1992).. Holding that the policy behind witness immunity is not advanced by offering immunity for incompetent experts retained by a party to perform professional services including trial testimony; Marrogi, 805 So. 2d, 1132 (holding that the policy behind witness immunity is not advanced by offering immunity for incompetent experts retained by a party to perform professional services including trial testimony).

Related Articles

Alcohol: Behavioral and Medical Effects
Cross-Examination of Experts
Daubert v. Merrell Dow Pharmaceuticals
Discovery of Expert Findings
Discovery: Depositions
Discovery: Discovery Motions
Discovery in the United States: Civil Cases
Discovery in the United States: Criminal Cases
Direct Examination of Experts
Ethics: Codes of Conduct for Expert Witnesses
Error Rates in Forensic Methods
Expert Opinion in Court: a Comparison of Approaches
Expert Witnesses: Selection and Investigation of Credentials
Expert Opinion: United States
Expert Opinion: Appeal v. Trial
Frye v. United States
Kumho Tire v. Carmichael
Peer Review as Affecting Opinion Evidence

CAROL HENDERSON AND KURT W. LENZ

Management: Crime Scene *see* Crime Scene Management

Mandated Treatment: Mental Health *see* Treatment, Mandated: Mental Health

Mandatory Treatment *see* Therapeutic Jurisprudence

Marijuana as a Controlled Drug *see* Cannabis

Marks or Impressions of Manufactured Items

Introduction

Marks and impressions are regularly encountered at scenes of crime and may be produced by a vast number of objects such as tools, tires, footwear, and fabric. Marks produced by tools (*see* **Toolmarks**), tires (*see* **Tire Impressions**), and footwear are covered in separate articles. Marks left by fabrics can be encountered in a variety of situations. However, of particular interest is the fact that investigators frequently encounter glove marks when looking for fingermarks (*see* **Mass Grave Investigation**). When a gloved individual presses his or her hand on a surface, a residual mark may be left showing the pattern of the glove. This can result from dirty or greasy gloves, or even gloves that have been previously "contaminated" with body secretions or cosmetics. Negative marks can also occur from the removal of material from the surface, dust, for example. Glove marks can be two-dimensional or tri-dimensional, depending on the malleable nature, or not, of the recipient substrate.

Similar to shoe marks and tire marks, glove marks can be made of a combination of manufacturing and acquired features. The identifying power of glove marks will therefore depend on the quality of the mark (e.g., clarity) as well as the rarity of the fabric pattern and construction combined with the presence, or not, of random acquired features such as holes or wear and tear.

Glove marks may also help the investigator establish the "path" taken by an individual around the scene of crime. Furthermore, when clothing is pressed against a smooth surface, a latent mark is produced resulting in another type of fabric mark [1]. This is prevalent in motor vehicle hit and run accidents, where the examination of fabric impressions occurs

when a segment of the vehicle comes into contact with the clothing of a pedestrian [2]. These marks may help determine the path taken by an individual, as well as the sequence of events that had taken place during the course of the crime.

Detection, Collection, and Examination of Fabric Impressions

Glove Marks

In the age of forensic science awareness, criminals are using gloves as a protective measure to prevent deposition of their fingerprints at a crime scene. Both leather gloves and fabric gloves are commonly utilized. Although leather gloves contain a natural fat, both types of gloves collect grease, sweat, and dirt through everyday wear. Further, at scenes of crime where blood is involved (such as homicide), the glove may be contaminated with blood. It is the presence of these materials that allow the pattern of the glove to be marked onto a surface [3]. Glove marks may be either latent or visible. Latent marks are "invisible" and can be found by performing a grazing angle search with a light source. It is important to remember that glove marks are much more fragile than fingermarks [1].

As glove marks are formed best on smooth surfaces, the powdering method can be used to increase the contrast between the latent deposit and the background surface. Either black or white powder may be used for the development process, which should be brushed on sparingly. If the powder is brushed on vaguely and in excess, the print detail will be destroyed [1]. Any glove mark found at the crime scene should be documented and evidence photography carried out. Photographing should be done with a camera set up on a tripod, taken perpendicular to the glove print to remove distortion and eliminate perspective (*see* **Crime Scene Photography: US Perspective**). The glove print must be photographed with and without a scale [3].

If possible, the entire object on which the glove print was found should be taken back to the laboratory for direct examination and comparison. However, if this is not possible, the glove print may be lifted with a gelatin lifter similar to that used for lifting fingermarks (*see* **Toolmarks**). If the glove mark is "negative" (such that the glove removes material off the surface, dust is an example [4]), it can be simply lifted using a black gelatin lifter.

A technique involving electrostatic attraction of the marks onto a plastic lifting film may also be used. This process is especially effective for faint glove marks on a colorful background [3]. Glove marks in blood can be enhanced and developed using reagents used for the detection and enhancement of fingermark in blood, such as amido black [5] for example, which is a stain sensitive to the presence of protein.

Comparison (or test) prints can be made from the suspect's glove, usually on glass or in some cases on the same material as the glove print deposit found at the crime scene. A number of comparison prints may be produced using varying degrees of pressure. Although producing a good-quality comparison print may be difficult, powders and polish should be avoided when making the print as minute details may be lost. Instead, breathing slightly on the finger of the glove may help improve results [3].

Fabric Marks

Fabric marks can be made with items such as clothing, socks, towels, and handkerchiefs. Although not as common as glove marks, fabric marks do arise during casework. Such marks are generally left on a surface if the fabric is thin and contains residual material such as moisture or is dirty. These marks are searched for and developed in the same way as for glove marks. Evidence photography is also carried out the same way as for glove marks; however, if the fabric print is adequately large, several photos must be taken.

The scale position and the angle of lighting should be altered for each photograph so that maximum detail of the fabric mark is documented [1]. Such marks should be compared with the suspect fabric exhibit, and the similarities and differences between the seams, structure of the fabric, and other oddities examined. For cases involving clothing impressions made on areas of a vehicle in a motor accident, one way of preserving the evidence is by photographing the fabric impression that was made on the segment of the vehicle. An exemplar is produced of the clothing article of interest, which is done using the previously described method. This exemplar is then used for direct 1 : 1 comparison with the photograph of the impression found on the motor vehicle [6].

Interpretation and Evidential Significance of Fabric Impressions

Glove Marks

When used as evidence, glove marks may demonstrate manufacturing (class) characteristics that show that the mark *could* have been made from a particular glove (i.e., nonexclusion). In some instances, glove marks show enough random acquired characteristics, such as distinctive holes, for example, that allow the examiner to conclude that a particular glove *did* leave the mark found at the scene (i.e., identification). Between these two conclusions, a qualified opinion may be given, where the value of the mark will depend on its quality as well as the rarity of the fabric pattern and construction combined with the presence, or not, of random acquired features. In any case, the presence of a meaningful (or unexplainable) difference between the mark and the test print allows the examiner to conclude that the mark *could not* have been made from a particular glove (i.e. exclusion).

In general, it is more common for an identification to be made with leather gloves than cloth gloves. Cloth gloves may exhibit tears and snags, or irregularities in the weave pattern due to the manufacturing process or material used. Leather gloves may show wrinkles and crease formations, which is more likely to leave a mark with a large number of comparative features. The presence of tears and cracks due to the random process of wear and tear may allow the mark to be identified, as exactly the same pattern of imperfections cannot be found on any two different gloves [3]. The difference between a cloth glove mark and a leather glove mark is the surface pattern. Leather gloves are made of the skin of an animal, which leaves unique marks on a surface not dissimilar to those left by fingerprints (Figure 1). Naturally, the patterns found on animal skin are not homogenous and vary from animal to animal. On the contrary, cloth marks are consistent for each type, and so the individual identification is more difficult to achieve. Furthermore, both leather and cloth gloves often mold to the finger tips of the individual, allowing another point of identification for the examiner. In rare cases, partial fingerprints can also be found within or superimposed with a glove print (occurring when there is a hole in the glove), obviously increasing the evidential value of the mark [1].

It should be noted that when a glove is seized in a case it should be collected and protected appropriately, as it may be possible to detect and examine finger marks or DNA recovered from inside the glove.

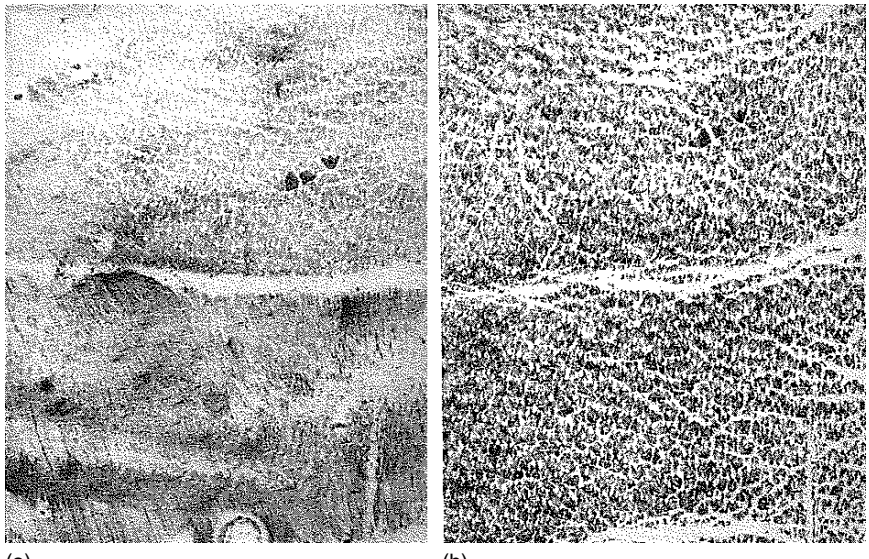


Figure 1 Comparison between a latent leather glove print (a) and the test print (b), showing the individual, minute characteristics of animal hide [1] [Reproduced with permission from Ref. 2. © IAI, 2000.]

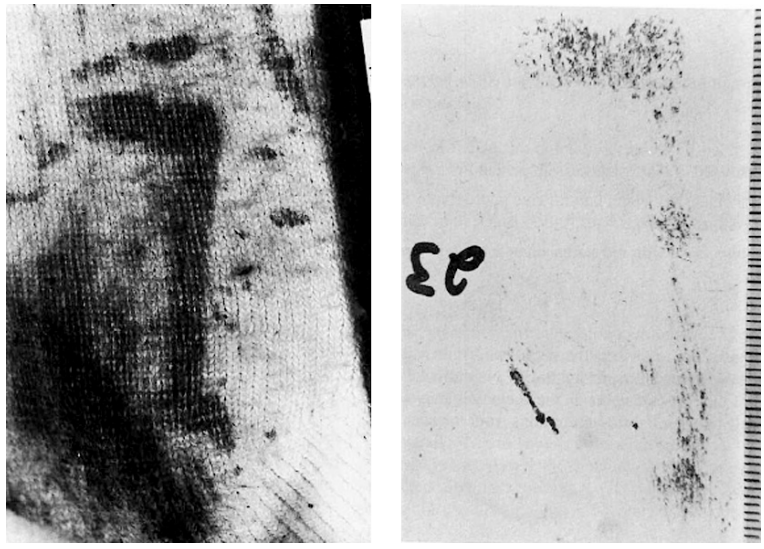


Figure 2 Comparison of bloodstain pattern on sock (reversed, 2 : 1 enlargement) and fabric impression enhanced with amido black stain (2 : 1 enlargement) [2] [Reproduced with permission from Ref. 6. © ASTM.]

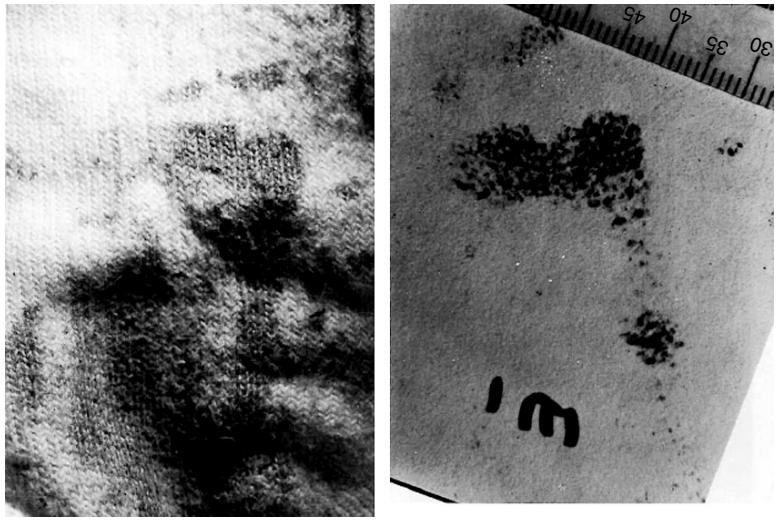


Figure 3 Comparison of the bloodstained sock and another fabric impression found at the crime scene, showing that the fabric tuft pattern on the sock was transferred onto the impression [Reproduced from Ref. 1. © Taylor and Francis Group, 2000.]

Fabric Prints

In most cases, fabric marks made by items such as towels and socks are hard to identify and individualize, reducing its evidential value. Unlike a glove mark, there is no definite boundary for the edges of the material. Comparing seam patterns or other surface patterns is difficult if the area on the item of interest that may have left the mark cannot be identified [1]. However, there has been a documented case involving the identification of fabric impressions made by a sock [2] (Figures 2, 3).

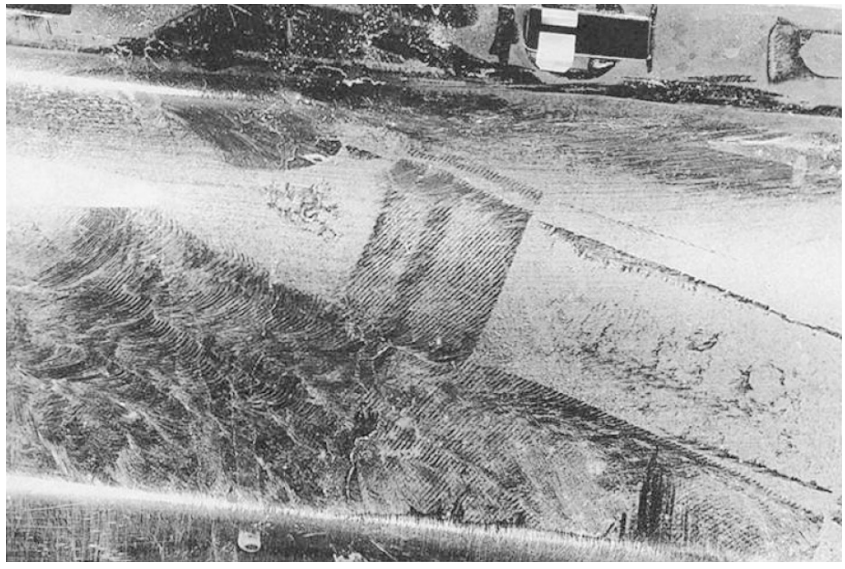


Figure 4 Impression found on the chrome bumper of a car after a motor vehicle accident involving a pedestrian [6] [Reproduced from Ref. 1. © Taylor and Francis Group, 2000.]

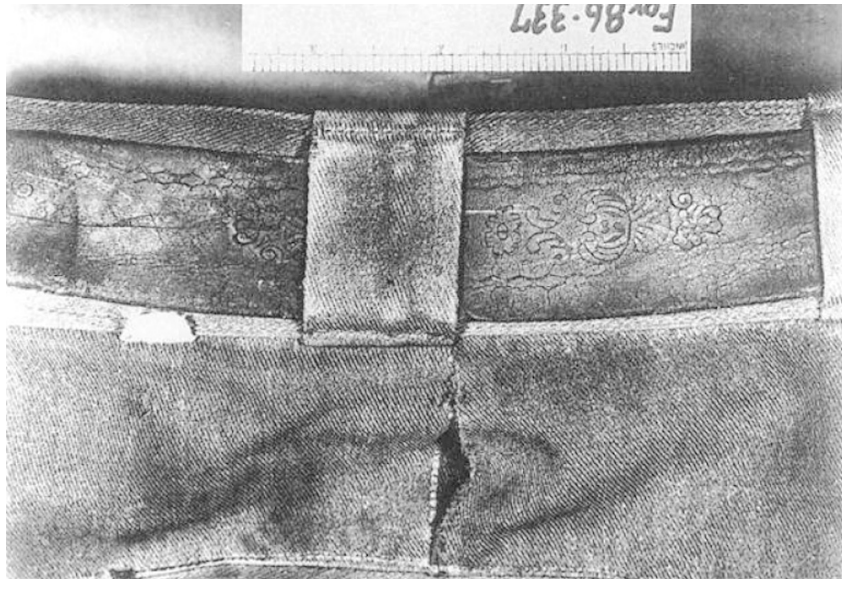


Figure 5 Pants and belt of the pedestrian, showing the pattern on the belt and the loop [6]

Fabric impressions were found at a homicide scene through the use of amido black stain (for marks in blood). After sequestering the suspect's bloodstained sock, it was discovered that the main bloodstain along with secondary bloodstains were in agreement with the impression found at the scene. Furthermore, the presence of four randomly placed raised tufts found on the sock corresponded with tuft markings found on the impression. The unique and random pattern as a result of the bloodstains and tufts was significant enough to trace the fabric impression to the suspect's socks [2].

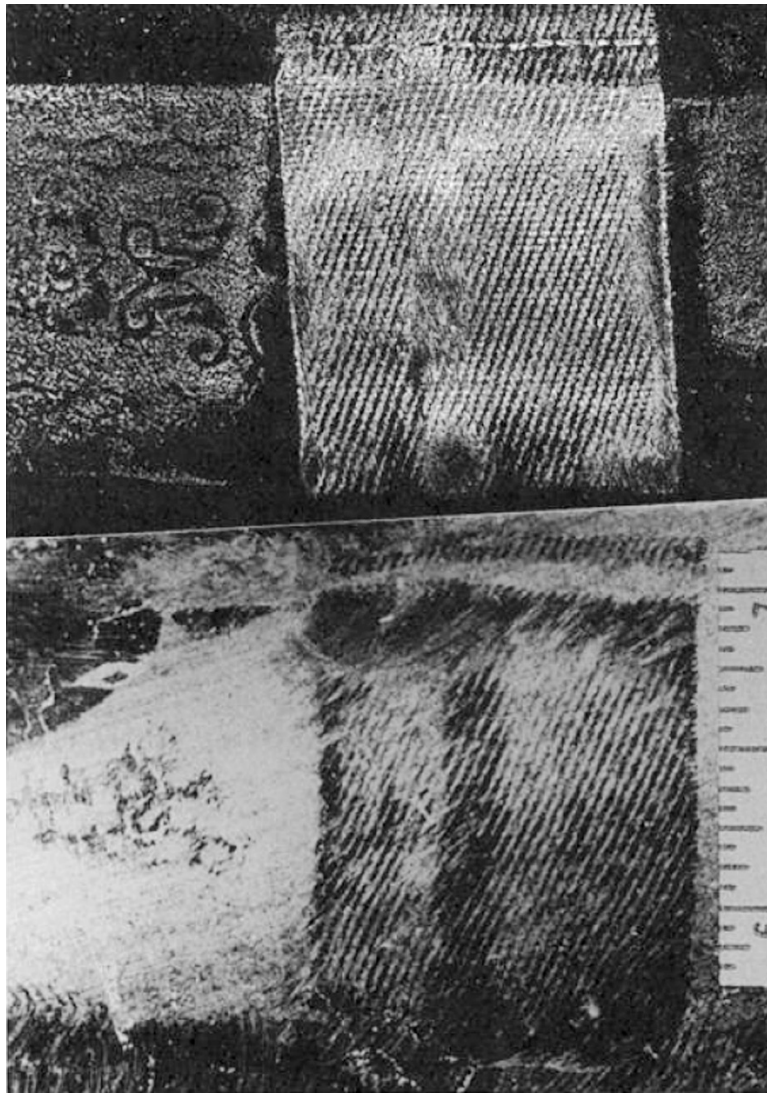


Figure 6 Juxtaposition of the belt and loop (top figure) with the corresponding region of fabric impression left on the chrome bumper (bottom figure) [6]

This example shows that the position and presence of unusual characteristics is enough to warrant an individual identification. The types of fabric impressions in such incidences as hit and run are also difficult to identify, as finding and preserving the fabric marks pressed into the vehicle is a tedious process. Although such impressions are not uncommon, it is rare to obtain evidence that permits identification of the clothing at the origin of the mark. However, an unusual case involving the individual identification of a clothing impression on a motor vehicle has been reported [6]. A pedestrian was hit by a motor vehicle, thereby causing his death. Upon examination of the chrome bumper an impression was found. The sites on the impression corresponded with the weave pattern on the pedestrian's pants and the stitching and loop hole of his belt. The combination of the size and position of the clothing items and the inscription found on the belt, which was also transferred onto the impression, was very unusual and was the basis

for the individual identification between the clothing and impression to be made [6] (Figures 4–6).

In most cases, fabric impressions provide class evidence, as it may be difficult to state that the marks had come from a given item excluding any other similar items. However, as for glove marks, the evidential value of fabric impressions will depend on the quality of the mark as well as the rarity of the fabric pattern and construction combined with the presence, or not, of random acquired features. As for glove marks, the presence of any meaningful (or unexplainable) difference between the fabric impression and the test print will allow the examiner to reach a conclusion of exclusion (i.e., the fabric impression *could not* have been made from the known fabric).

References

[1] Fisher, B.A.J. (2000). *Techniques of Crime Scene Investigation*, 6th Edition, CRC Press.
[2] Doller, W.D. (2000). An unusual case involving the individualization of fabric impressions made by a sock-clad foot, *Journal of Forensic Identification* **50**(5), 447–454.
[3] Nickell, J. & Fischer, J.F. (1999). *Crime Science: Methods of Forensic Detection*, The University Press of Kentucky.
[4] Cowger, J.F. (1992). *Friction Ridge Skin: Comparison and Identification of Fingerprints*, CRC Press.
[5] Hussain, J. & Pounds, C.A. (1989). *The Enhancement of Marks in Blood: Part ii, A Modified Amido Black Staining Technique*, Report no. 685, Central Research Establishment, June, 1989.
[6] Drummond, F.C. & Pizzola, P.A. (1990). An unusual case lnvolving the individualization of a clothing impression on a motor vehicle, *Journal of Forensic Sciences* **35**(3), 746–752.

Susan Luong and Claude Roux

Mass Grave Investigation

Mass Grave Definition

Mass graves are defined in various ways, most revolving around the number of people in a grave and what constitutes a "mass". This has ranged from two or more bodies that physically touch each other [1] to at least six individual bodies [2, 3]. Mant's [1] definition brings in the physical component that the bodies must be physically touching each other. Haglund [4] also feels that having remains in contact with each other is what makes mass graves different from other grave types, and uses that criterion in his definition.

Other definitions include the genesis of the grave. The United Nations Rapporteur interprets a mass grave as a location where three or more victims of extrajudicial, summary, or arbitrary executions were buried, not having died in combat or armed confrontations [5]. Schmitt [6, p. 279] defines a mass grave as "one that contains the remains of more than one victim who share some common trait connected with the cause and manner of death". He further goes on to state that "criminal mass graves contain the remains of a group of individuals who share some common trait that justified their assassinations in the eyes of the perpetrators" [6, p. 279]. A basic flaw in a definition of a mass grave that includes the genesis of the inhumation is that the investigator may never know whether they are dealing with a mass grave or not, if the genesis is not known. It is likely that this would lead to the need to find another term for such a grave where the reasons for the deaths are unknown.

A simple definition of a mass grave is that it contains four or more individuals buried at or about the same time. This is consistent with the definition of a mass murder in the *Crime Classification Manual* [7, p. 12] used by the Federal Bureau of Investigation. The Manual defines a mass murder as "a homicide involving four or more victims in one location and within one event" [7, p. 12]. A mass grave definition consistent with the standard definition of mass murder is simple and logical. The differentiation between a criminal mass grave and another type should rest on the result of the investigation, depending on whether the deaths were the result of a criminal act.

Mass Grave Taphonomy

Single and mass death sites have taphonomic differences. Graves with poor drainage and multiple layers of bodies tend to trap moisture, even if only the moisture from the decomposing remains themselves. This scenario accounts for the adipocere seen in mass graves created years or decades earlier. Undisturbed soils at the bottom of the grave trap water in the softer, unconsolidated soils of the grave fill and create

conditions in which adipocere develops. This keeps the bodies in a moist environment, and they tend to saponify, as do remains under water. Thus, bodies in the center of a large body mass may be well preserved after years of burial. Where bodies are laid side by side in a single layer, the decomposition fluids leach into the soil and the bodies decompose in a manner more similar to a body in a single grave.

Thus, mass graves tend to create their own microenvironments toward preservation [4, 8]. Mant [8, p. 33] notes that in a mass grave the bodies in the center of the grave are much better preserved than in the periphery of the grave. Mant did his research after World War II, exhuming graves for the Nuremberg Tribunal. Also related to World War II were the 1943 German excavations in Poland investigating mass graves in the Katyn Forest. A professor of forensic medicine and criminology at Breslay University summarized the condition of the bodies at Katyn as follows:

> The stages of decay were found to vary in accordance with the position of the bodies in the pits. Whilst mummification had taken place on the top and at the sides of the mass of bodies, a humid process could be observed caused by the damp nearer the center Buhtz 1943 in [9, p. 142].

The German investigators attributed the preservation of the remains to the microenvironment created by the mass grave:

> In the mass graves at Katyn the murdered captives were packed so tightly together (either dead or dying) and sealed over with copious quantities of solid soil, that the putrification process was slowed considerably. For example, in the "L" shaped grave ... the initial interment was estimated to be almost 3000 bodies. These bodies were packed together so tightly that decaying and decomposing fluids of each body penetrated, imbibed, and infiltrated other dead bodies within the grave

(Official Statement Concerning the Mass Murder at Katyn 1943 as quoted in [10, p. 53]).

While it might seem contradictory to have mummified tissue, skeletonized remains, and adipocere tissue in the same grave, it happens in mass graves and is dependent on the moisture available to the remains. Moisture is essential for both decomposition and adipocere formation. In talking about mass graves excavated after World War II:

> Adipocere formation depends on the hydrolysis of fat to fatty acids, and cannot take place without an adequate supply of water. This water is derived not only from the exterior of the body but also from the interior, dehydrating the underlying tissues and organs including the muscles themselves. Thus adipocere formation retards the action of putrefactive organisms.
>
> It was found at exhumation, as would be anticipated from the above, that there was an outer layer of adipocere surrounding a layer of mummified muscle tissue. Some mummified muscle tissue was usually found incorporated in the deepest layers of adipocere [8, p. 45].

In the best of circumstances, relating the amount of adipocere formation to postmortem interval is difficult. Adipocere may appear as early as a few days following death [10, p. 101] and has been observed in burials as old as 122 years [11, p. 470]. Sledzik and Micozzi [12, p. 485] state that "presence of adipocere is an artifact of decomposition suggesting that a minimum length of time has passed since death, but it may not always be useful in determining the postmortem interval." In mass graves in particular, adipocere formation tells little about the length of the postmortem interval.

Mass Grave Investigation Strategy

The investigation of larger graves has a number of logistical challenges not seen in investigating smaller graves [2, 4, 13]. Maintaining access to the central remains for body removal, while cleaning the surrounding remains, can be a challenge. The large amounts of fill in the grave mean that the disposal of the soil removed from the grave must be carefully planned. As discussed above, a large mass of bodies creates differences in decomposition, which become more noticeable the larger the mass of the remains. Removing, as well as analyzing, remains in all stages of decomposition adds to the equipment and personnel needed, both in the field and the morgue. Other large graves may not have large body masses that create the differences in decomposition, but share the logistical challenges of a large excavation, adding to the logistical challenge. The larger the grave, the greater the range of preservation likely to be present.

Locating mass graves, like smaller graves, is often dependant on witness testimony [6]. Also, the same remote sensing technologies may apply

(*see* **Length Measurement**), given the appropriate soil and topography conditions. In addition, the size of the graves makes them more likely to show up in aerial photography [2, pp. 120–121]. Given that large numbers of individuals were buried in the area and possibly executed there, and that it might have taken a number of people, as well as vehicles and equipment, to accomplish this, there is often surface evidence of a mass grave [6].

As with any archaeological feature, the outline and size should be defined before excavation. As the overburden is removed from a large grave, the excavator should look for the edges of the grave. This may be most easily done by placing two trenches, perpendicular to each other, over the grave (called *cross-trenching*). This should give the excavator an idea of the location of the sides of the grave. If the grave fill can be differentiated from the undisturbed soil, then the outline of the grave should be traced all around the edge. It sounds like common sense, but this will help keep the excavators from placing soil removed from one part of the grave on top of another part of the grave.

If the grave was dug using heavy machinery, it is probably going to take heavy machinery to remove the overburden to the depth of the remains. This will make it harder to look for tool marks or other indications of how the grave was dug. The excavators should look for teeth marks on the side of the grave showing the teeth of the original front-end loader or backhoe bucket. There may also be traces of a ramp that the machine used to access the bottom of the grave. On the ramp, or at the bottom, may be imprints of the tires, or treads, of the machine.

In some mass graves, the bodies are laid neatly side by side in rows in a long trench. In a case such as this, where the bodies were buried in an organized manner, the bodies can be numbered at one time, documented in sequence, and lifted in sequence.

Haglund and others [13] describe a method that worked satisfactorily on graves where the bodies were dumped in a disorganized manner. The excavators worked in two teams: a documentation team, and a body removal team. Working with groups of about 10 bodies at a time, the documentation team completed the mapping, notes, and photographs of the next 10 bodies to be removed on one side of the grave. If the bodies are not laid neatly in the grave, deciding on the next bodies to be removed takes some investigation to ensure which remains are least encumbered by other remains. Bodies were not numbered until they were ready to be removed, in order to create fewer errors in numbering. It is possible to label two portions of the same body with different case numbers when the intervening portions of the remains are covered with soil or other bodies.

The body removal team then removed the remains from the grave. On fleshed remains, this sometimes entailed lifting limbs of some bodies in order to free the target remains. Skeletal remains presented more of a challenge, and portions of a single body sometimes needed to be documented and partially removed to get to the remains below. When this is done, then both the portion removed and the articulating portion left in the ground need to be carefully labeled. No more than two skeletons on each area should be partially removed, or confusion will inevitably result. At no point should personnel in the team change, if skeletons have been partially removed. The same people that partially removed one skeleton should remove the remainder of that same skeleton.

In graves where heavy machinery was used to bury the remains, it is not uncommon to find disassociated body parts. This also happens at sites where the remains were buried after they were partially decomposed. Should isolated body parts be found, they should be catalogued separately from the complete sets of remains. A "partial remains" log can be started that numbers the assemblage, gives the location, photograph numbers, and gives a brief description of the remains. Including the partial remains in the same numbering system as the complete remains is sure to lead to confusion during the examination when the partial remains are reunited with the remaining body parts.

Determining a Charge

Mass graves are (thankfully) relatively rare in domestic criminal cases. They tend to result from mass fatality incidents, either natural or cultural. Only the cultural incidents become forensic cases. The most common of these are war crimes, crimes against humanity, and genocide, and so a brief definition of each are discussed.

Genocide was defined as a crime after World War II in the 1948 Convention on the Prevention and Punishment of the Crime of Genocide. The definition

of genocide set forth in the Convention is still definitive and used in the statutes of the Rwandan and Yugoslavian tribunals as well as the Rome Statute that established the International Criminal Court (ICC). The Convention defines genocide as a crime in either peace or war:

> any of the following acts committed with intent to destroy, in whole or in part, a national, ethnical, racial, or religious group, as such: killing members of the group; causing serious bodily or mental harm to members of the group; deliberately inflicting on the group conditions of life calculated to bring about its physical destruction in whole or in part; imposing measures intended to prevent births within the group; forcibly transferring children of the group to another group

(Convention on the Prevention and Punishment of the Crime of Genocide, United Nations. December 9, 1948).

The crime of genocide requires an intent to physically destroy a certain group of people. The court also requires that the group destroyed be a national, ethnical, racial, or religious group. Political dissidents and soldiers are not groups protected by the genocide convention. The forensic investigator considering genocide as a charge must attempt to determine whether there was intent to kill the deceased and that they belonged to a specific target group.

War crimes are violations of the international laws of war, a body of law known as the *International Humanitarian Law* (*IHL*). While limitations on armed conflict date at least as far back as the Chinese warrior Sun Tzu in the sixth century BC [14], p. 374], the Hague conventions of 1899 and 1907 codified much of the humanitarian law for the Western world. The Charter of the International Military Tribunal at Nuremberg defined war crimes as "violations of the laws or customs of war, including murder, ill-treatment, or deportation of civilians in occupied territory; murder or ill-treatment of prisoners of war (POWs); killing of hostages; plunder of public or private property; wanton destruction of municipalities; and devastation not militarily necessary" [14, p. 374].

Only intentional, grave breaches of the treaties are considered war crimes. If individuals violate the treaties, but do not commit a grave breach, then they have committed an illegal act, but not a war crime. Grave breaches of the conventions are defined as willful killing, torture or inhumane treatment (including medical experiments), willfully causing great suffering or serious injury to body or health, extensive destruction and appropriation of property not justified by military necessity and carried out unlawfully and wantonly, compelling a prisoner of war or civilian to serve in the forces of the hostile power, willfully depriving a prisoner of war or protected civilian of the rights of a fair and regular trial, unlawful deportation or transfer of a protected civilian, unlawful confinement of a protected civilian, and taking of hostages [14, p. 374]. In 1977, Protocol I of the Conventions expanded the definition of grave breaches to include certain medical experimentation, making civilians and undefended localities the object or inevitable victims of attack, the perfidious use of the Red Cross or Red Crescent emblem, transfer by an occupying power of parts of its population to occupied territory, unjustifiable delays in repatriation of POWs, apartheid, attack on historic monuments, and depriving protected persons of a fair trial [14, p. 374].

A crime against humanity has come to mean anything atrocious committed on a large scale [15, p. 107]. The term originated in the preamble of the 1907 Hague Convention, which was based on existing State practices that derived from the moral values that constituted the "laws of humanity". In 1945, the Nuremberg Charter, formally the Agreement for the Prosecution and Punishment of the Major War Criminals of the European Axis and Charter of the International Military Tribunal, first defined crimes against humanity in positive international law:

> Crimes against humanity: murder, extermination, enslavement, deportation, and other inhumane acts committed against civilian populations, before or during the war; or persecutions on political, racial or religious grounds in execution of or in connection with any crime within the jurisdiction of the Tribunal, whether or not in violation of the domestic law of the country where perpetrated [15, p. 107].

The statutes for the International Criminal Tribunal for the former Yugoslavia (ICTY) and the International Criminal Tribunal for Rwanda (ICTR) added rape and torture to the list of specific crimes included in crimes against humanity. The ICC added apartheid and enforced disappearances to the list [15, p. 108]. Crimes against humanity overlap with both genocide and war crimes. However, crimes against humanity can occur in either war or peace, unlike war crimes. Unlike genocide, they do not require

the intent to "destroy in whole or in part" a specific group, but rather are part of a "widespread and systematic attack directed against any civilian population" (Rome Statute of the ICC, UN document A/Conf. 183/9, 17 July 1998, Art. 7).

The investigator, then, must show that the subjects of the attack are civilians and the incident is part of a systematic or widespread pattern, rather than an isolated incident. Widespread, according the to the ICTR court, is not defined geographically, but rather as an attack carried out against a "multiplicity of victims" [16, pp. 34, 35].

In framing a charge, the existence of the grave and the location of remains are the first indicator that a criminal act may have occurred. This proves that these people are not refugees, that they are not alive somewhere but unaccounted for, but that they are, indeed, dead. The demography of the grave also becomes of paramount importance. A grave full of adult males, dressed in military uniforms, may result from a postbattle cleanup. However, a grave full of woman and children in civilian clothing, suggests that something else occurred.

The artifacts in the grave and personnel effects are the best indicators of whether the individuals belonged to a specific target group. To decide on a charge, it may, in fact, be unnecessary to identify individuals, although international human rights groups consider the exhumation of the remains without an attempt to identify them and return the bodies to the families extremely unethical [17]. To determine a charge, what is needed is the knowledge that the bodies did or did not belong to a specific national, religious, ethnic, or military group. Clothing, identification papers, jewelry, and artifacts with the bodies are crucial in this regard.

Corroboration of Witness Testimony

In the forensic examination of a mass grave, as in any forensic investigation, witness information is of paramount importance and the forensic evidence often plays only a corroborative role. As Schmitt [6, p. 280] says, mass graves are seldom a secret. Often there are survivors to the massacre who come forward and tell their story. In the scene reconstruction of a mass grave scenario, everything, from the topography of the area to the location of materials in the grave, plays a part.

The ecological artifacts, or "ecofacts", in and near the site are important in reconstructing the events at the site. In the investigation of World War II graves in the Katyn Forest, the trees on the graves were used to help date the graves themselves [9]. So, it is helpful to be aware of the natural features of a site area and what they mean in relationship to the activities at the site.

The location of cartridge cases can be used to trace where individual shooters stood and how they moved throughout the area, through firearms identification analysis [18]. In an investigation in the village of Koreme, in the Kurdish portion of Iraq, investigators plotted each cartridge case found in an execution area, where 27 young Kurdish men had allegedly been killed by an Iraqi military execution squad. Of a total of 124 cases collected, the firing pin analysis identified a minimum of 7 individual weapons used in the execution. The weapon used was an AK-47, or similar weapon, which normally has a detachable magazine containing 30 rounds. The firearms analysis showed that one weapon fired a minimum of 37 rounds, requiring reloading during the executions [19]. Scott [19] was able to use the map of the cartridge cases to show that several weapons moved closer to the victims as they fired. At least two rounds were fired within 10 m of the victims by the same weapon that reloaded. This type of detail in scene processing and analysis is crucial in scene reconstruction and can be used as strong corroboration in witness testimony.

Summary

Definitions of mass graves include factors such as the number of bodies, whether they touch each other, and how the bodies came to be in the grave. This article suggests a definition of mass graves parallel to the definition of mass murder: mass graves are graves of four or more individuals buried at or about the same time.

Mass graves are different from single graves in the changes in how bodies decompose when in a mass and the logistical challenges in excavation. When bodies are piled up on top of each other, they decompose much differently than in a single grave, owing to accumulated water and body fluids. This variation in decomposition stage, as well as the sheer size of the graves, can make the exhumations logistically challenging.

Criminal mass graves rarely result from simple homicide, but often are a result of genocide, war crimes, or crimes against humanity. Determining a charge relies on attention to the cause and manner of death of decedents, grave demographics, artifacts, and personal effects. Attention to how the grave is dug and bodies are arranged can help to corroborate witness testimony on the sequence of events and mechanics of execution and body disposal.

References

[1] Mant, A.K. (1987). Knowledge acquired from post-war exhumations, in *Death, Decay, and Reconstruction: Approaches to Archaeology and Forensic Sciences*, A. Boddington, A.N. Garland & R.C. Janaway, eds, Manchester University Press, London.

[2] Connor, M.A. (2007). *Forensic Methods: Excavation for the Archaeologist and Investigator*, Altimira Press, New York.

[3] Skinner, M. (1987). Planning the archaeological recovery of evidence from recent mass graves, *Forensic Sciences International* **34**, 267–287.

[4] Haglund, W.D. (2002). Recent mass graves: an introduction, in W.D. Haglund & M.H. Sorg, eds, *Advances in Forensic Taphonomy: Method, Theory, and Archaeological Perspectives*, CRC Press, Baton Rouge, pp. 243–261.

[5] United Nations (1991). *Manual on the Effective Prevention and Investigation of Extra-Legal, Arbitrary and Summary Executions*, United Nations, New York.

[6] Schmitt, S. (2002). Mass graves and the collection of forensic evidence: genocide, war crimes, and crimes against humanity, in *Advances in Forensic Taphonomy: Method, Theory, and Archaeological Perspectives*, W.D. Haglund & M.H. Sorg, eds, CRC Press, Baton Rouge, pp. 277–292.

[7] Douglas, J.E., Burgess, A.W., Burgess, A.G. & Ressler, R.K. (1992). *Crime Classification Manual: A Standard System for Investigating and Classifying Violent Crimes*, Jossey-Bass Publishers, San Francisco.

[8] Mant, A.K. (1950). *A Study of Exhumation Data*, MD Thesis, University of London, London.

[9] FitzGibbon, L. (1977). *Katyn Massacre*, Corgi Books, London.

[10] Gill-King, H. (1997). Chemical and ultrastructural aspects of decomposition, in *Forensic Taphonomy: The Postmortem Fate of Human Remains*, W.D. Haglund & M.H. Sorg, eds, CRC Press, Baton Rouge, pp. 93–108.

[11] Manhien, M.H. (1997). Decomposition rates of deliberate burials: a case study of preservation, in *Forensic Taphonomy: The Postmortem Fate of Human Remains*, W.D. Haglund & M.H. Sorg, eds, CRC Press, Baton Rouge, pp. 469–481.

[12] Sledzik, P.S. & Micozzi, M.S. (1997). Autopsied, embalmed, and preserved human remains: distinguishing features in forensic and historic contexts, in *Forensic Taphonomy: The Postmortem Fate of Human Remains*, W.D. Haglund & M.H. Sorg, eds, CRC Press, Baton Rouge, pp. 483–495.

[13] Haglund, W.D., Connor, M.A. & Scott, D.D. (2001). The archaeology of contemporary mass graves, *Historical Archaeology* **35**(1), 57–69.

[14] Ratner, S.R. (1999). Categories of war crimes, in *Crimes of War: What the Public Should Know*, R. Gutman & D. Reiff, eds, W.W. Norton & Company, New York, pp. 374–376.

[15] Bassiouni, M.C. (1999). Crimes against humanity, in *Crimes of War: What the Public Should Know*, R. Gutman & D. Reiff, eds, W.W. Norton & Company, New York, pp. 107–108.

[16] Human Rights Watch (2004). Genocide, war crimes, and crimes against humanity, *Topical Digests of the Case Law of the International Criminal Tribunal for Rwanda and the International Criminal Tribunal for the Former Yugoslavia*, Human Rights Watch, New York.

[17] International Commission of the Red Cross (2002). *Special Issue: Missing Persons. International Review of the Red Cross*, p. 848.

[18] Scott, D.D. & Connor, M.A. (1997). Context delicti: archaeological work in forensic context, in *Forensic Taphonomy: the Postmortem Fate of Human Remains*, W.D. Haglund & M.H. Sorg, eds, CRC Press, Baton Rouge, pp. 27–38.

[19] Scott, D.D. (1993). Firearms identification of the Koreme execution site, in *The Anfal Campaign in Iraqi Kurdistan: The Destruction of Koreme*, Human Rights Watch, Washington, DC, pp. 103–107.

Further Reading

Lauck, J.H. (1988). *Katyn Killings: In The Record*, The Kingston Press, Clifton.

MELISSA CONNOR

Mass Murder *see* Homicide: Multiple (Behavior)

Massacre *see* Homicide: Multiple (Behavior)

Materials Science

Materials Engineering and Failure Analysis

The discipline of materials engineering and science spans the scope of the physical and chemical makeup of materials, the strength and other mechanical properties that are critical to appropriate decisions in product design, the relationships between and among materials, and finally, the analysis of actual or potential failures.

The primary forensic question in materials engineering and science is usually why a materials failure occurred. The exact reason will depend on the application, but four fundamental questions given in the following are frequently considered:

- When did the material fail in relation to a sequence of events? Did it fail and cause an event, or did it fail as a result of the event?
- Was design or manufacturing a factor?
- Was the choice of material a factor?
- Was inadequate maintenance a factor?

The materials engineer answers these questions using specific training in the relationships among material structure, properties, and processing, and their affect on material performance. The materials engineer also uses a variety of characterization techniques to determine these relationships.

The materials engineer's knowledge overlaps with other fields. The materials engineer must be familiar, for example, with civil engineering principles to establish stress states in materials under loading conditions, chemical engineering to determine the effect of environment on material properties, and mechanical engineering to understand how a particular part fits into an overall design.

In the forensic aspect, the materials engineer applies his understanding of materials to identify the root cause of material failures. The engineer is frequently called on to provide opinions in cases of product liability due to a materials failure. Typical cases include the failure of a variety of products, automobile or aircraft accidents, and industrial or residential construction defects.

When evaluating a materials failure, the materials engineer considers the relevant material properties. Depending on the material in question and the type of failure, these properties can include mechanical properties such as strength, hardness, ductility, and toughness. Component design, stress state, and the effect of manufacturing defects, if present, must also be considered.

A variety of specialized references exist to assist the materials engineer in failure analysis. These include handbooks, testing protocols, and material specifications and standards.

Failure of Materials

The Stress–Strain Curve

Failure of materials in the extreme case is the fracture or permanent deformation of the material. Failure can also be any changes from the original or specified state of the material that prevent it from performing according to specification. Failure in this case includes mechanical changes such as temporary, elastic, or permanent, plastic, deformation, or the presence of defects that affect the expected mechanical properties.

To understand the processes of deformation and fracture, one must understand the mechanical properties of materials. The specific behavior of materials differs according to the type of material, but the defining properties are essentially the same. The important mechanical properties are given as follows:

- yield strength;
- tensile strength;
- percent elongation;
- Young's modulus;
- toughness;
- ductile or brittle behavior; and
- hardness.

These mechanical properties are tabulated for most engineering materials, including metals, plastics, and ceramics, and can be found in many print and on-line references, some of which are listed in the section "Information References" that deals with various information sources. For composites, bulk properties can be derived from the properties, volume fraction, and orientation of the constituents.

All of these properties, with the exception of hardness, can be directly determined from the stress–strain diagram of the material. A typical

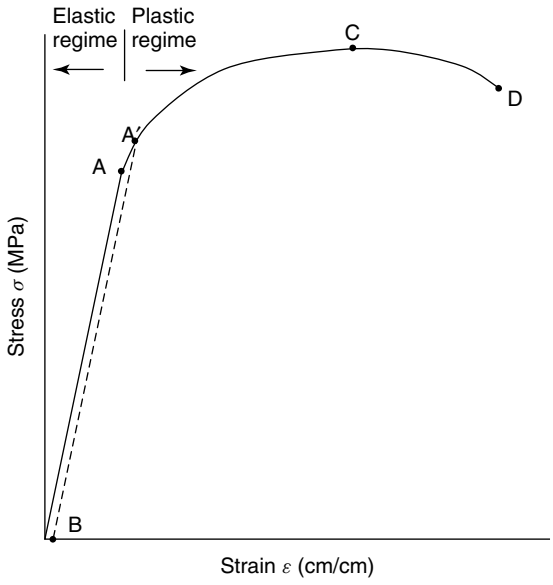


Figure 1 A typical stress–strain diagram for a metal

example of a stress–strain diagram for a metal is shown in Figure 1, and is discussed in the sections below.

The stress–strain diagram is obtained from a special type of materials characterization test called a *tensile test*. In a tensile test, an increasing load is applied to a specimen of the material being tested. As the load increases, the sample lengthens and deforms until fracture occurs. The changes in load and sample length with time are recorded, and converted into engineering stress and strain. From this data, the stress–strain diagram is created from which the mechanical properties of the material can be determined. Further details on the tensile test are provided in Section "Tensile Testing" (see Section "Materials Characterization").

Engineering Stress and Strain. Engineering stress, σ, is the amount of load, or force, per unit area applied to a material. Stress is commonly used instead of the force to remove the affect of material dimensions. For the same applied force, more deformation can be expected if the material being acted on has a smaller cross section, and less if the material has a larger cross section. The stress is defined as follows:

$$\sigma = \frac{F}{A_0} \tag{1}$$

where F is the load at any given time and A_0 is the initial cross-sectional area of the specimen. Stress is expressed in units of megapascals (MPa), or as pounds per square inch (psi), in the English system of units, where 1 MPa = 145 psi.

Engineering strain, ε, is the change in length of a material under stress, in the direction of that applied stress. It is the change in length at any time compared to the initial length, and is calculated as follows:

$$\varepsilon = \frac{l_i - l_0}{l_0} = \frac{\Delta l}{l_0} \tag{2}$$

where l_i is the length at any given time and l_0 is the initial specimen length. Strain is given in units of inches per inch (in./in.) or is stated as a percentage of the initial length ($\varepsilon \times 100\%$).

Deformation. When a load is applied to a material specimen, its dimensions will change in response. Depending on the amount of load, the deformation will be either temporary or permanent. For a small load, this deformation is temporary or elastic. Within the region of the material that is deforming elastically, inside the material itself, the bonds between atoms are being stretched. Atoms are displaced from their original positions and pulled away from each other, but no interatomic bonds are broken. In this case, when the load is removed, the bonds are no longer being stretched, and the atoms return to their equilibrium positions.

When a larger load is applied, the material will deform permanently or plastically. In this case, atoms are displaced far enough from their original positions that interatomic bonds are broken, and atoms move through the material. The movement of the atoms in response to the applied load results in permanent deformation.

These two regimes are shown on the stress–strain diagram of Figure 1. The linear region to the left of point A is the elastic region, and the nonlinear region to the right is the plastic region. As Figure 1 shows, plastic deformation only exists along with temporary elastic deformation.

The existence of the elastic and plastic deformation regimes has important implications for failure analysis. When an investigator examines a deformed component, the elastic portion of the deformation has recovered, and the actual deformation at the time of the event was greater than what is observed later.

The amount and significance of the elastic deformation can be determined from mechanical tests or from mechanical property data.

During a standard tensile test, changes in load or displacement are usually made slowly. When the load is applied rapidly, atoms have less time to move through the material. The effect is to make the material behave as if it were more brittle, in which case plastic deformation may be nonexistent.

Yield Strength. The yield strength of a material is the stress at which the material begins to deform plastically. Point A of Figure 1 is the demarcation point between elastic and plastic behavior. Below the yield stress, the material has only been elastically deformed. This elastic deformation can be compared to stretching a rubber band; if the stress is removed, the material returns to its original shape. If the yield stress at point A is exceeded, permanent deformation will occur. However, if a stress greater than the yield strength is applied and removed, recovery of the elastically deformed amount will take place; however, the plastic deformation will be permanent.

The exact point where the material behavior changes from elastic to plastic is not clear. Therefore, by convention, the yield strength is defined as the stress at a strain offset of 0.2% of the maximum measured strain before failure. As shown in Figure 1, the point on the x-axis where the strain is 0.2% of the maximum measured strain is located at point B. A line is drawn through that point, and parallel to the elastic portion of the stress–strain curve. The point, A′, where that line crosses the stress–strain curve is defined as the yield strength.

Tensile Strength. The tensile strength is the maximum tensile stress that the specimen can bear. On the stress–strain curve, this occurs at point C, the point of greatest stress on the curve. As the load is constantly increased during the tensile test, after point C, nonuniform deformation will begin to occur, along with necking of the sample, leading rapidly to fracture, at point D. In other words, if the stress at point C is reached, the material will fail completely unless the force is immediately reduced.

Percent Elongation. Percent elongation is a measure of the amount of deformation in the specimen. Specifically, it is the ratio of the change in specimen length to the initial length, and is calculated using the equation for strain (Equation 2).

Young's Modulus. Young's modulus, also called the *modulus of elasticity*, indicates how much stress can be applied to the material before permanent deformation occurs. A mechanical analogy is the deformation of a spring under tension. When a load is placed on a spring, the spring deforms in direct proportion to the load. If a load is removed, the spring returns to its original length. If the load exceeding the stiffness of the spring is applied, the spring is stretched beyond its capability to rebound, and is permanently deformed. The Young's modulus is the slope of the linear region of the curve, from zero up to the yield stress, or

$$E = \frac{\sigma_y}{\varepsilon} \tag{3}$$

where the strain is 0.02, by definition. The units of Young's modulus are megapascal.

Ductile or Brittle Behavior. The stress–strain curve can also reveal whether a material is ductile or brittle. As seen above, ductile materials, such as metals, undergo plastic deformation before failure; hence, failure occurs above the yield stress, and will result in permanent, plastic deformation of both the gross dimensions of the sample as well as the fracture surface. By contrast, with brittle materials little or no plastic deformation occurs. Typical stress–strain curves for ductile and brittle materials are shown in Figure 2.

Toughness. Toughness is a measure of the energy absorbed in the deformation until fracture of a material, and is measured as the area under the stress–strain curve. The units of toughness are joule per cubic meter.

Hardness. Hardness cannot be found directly from the stress–strain curve. It is determined by measuring the deformation resistance at the surface of the material. Several different methods are available for measuring hardness; however, all involve measuring the force required to press an indenter into a surface. Common methods are the Rockwell, Brinell, Knoop and Vickers hardness tests, and are specified by standards such as ASTM E18-05, E10-01, C730-98, and E92-82 [1].

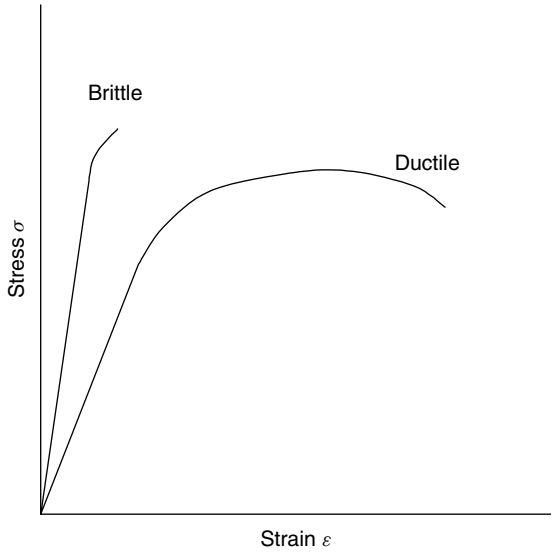


Figure 2 Typical stress–strain diagrams for a ductile and a brittle material

Hardness does correlate well with yield strength, since both yield strength and hardness measurements are taken when plastic deformation is just beginning. Tables and charts can be found in references such as the *ASM Metals Handbook* [2], which relate the yield strength of a metal to hardness.

Failure Modes

Fracture. The discussion until now has focused on ideal materials in a nominally defect-free state. Under those conditions, the material will fail when the tensile stress is exceeded. Materials in the real world can frequently contain defects or flaws, which may lower the strength of the material and lead to unexpected failure. The role of defects or flaws must be considered in failure analysis.

Defects can be material or geometric heterogeneities. Examples are inclusions, second phase particles, voids, machining marks, geometric variations, and miscellaneous surface flaws. These defects can act as stress concentrators, and by their geometry prevent the material from sustaining normal loading.

To demonstrate the effect of flaws on the strength of a material, a simple model of a flaw, a symmetrical crack, can be used. The crack is inside the material, and perpendicular to the direction of tensile loading. If the crack is small, it will be relatively stable. The size of the crack may remain constant, or it may grow slowly, depending on the applied stress, any cyclic loading, and the morphology of the material.

This crack growth is occurring by ductile fracture mechanisms, where the term ductile indicates that the material at the crack tip is deforming. If the crack grows, once it reaches a critical length with respect to the direction of the applied stress, it becomes unstable. At the point of instability, the crack will rapidly propagate through the material, resulting in a fracture. The unstable crack growth occurs by brittle fracture, with little or no plastic deformation occurring at the fracture surface.

Since materials without any bulk defects can sustain stress up to their yield strength without deforming, the yield strength would seem to be an appropriate design criterion. When defects are present, however, the material can fail below the yield strength. To account for this, during design the yield strength is usually multiplied by a safety factor. The designed safety factor varies according to the application, but may be 1.5–2 times the yield strength. A material whose yield strength is above the resulting design strength can then be selected.

Choosing a material with a yield strength exceeding the expected stress is intended to protect against failure in three different ways:

- If the maximum stress is slightly or occasionally greater than expected, the stress will still remain within the design envelope as long as a safety factor is used.
- If the yield stress of the material is slightly below the anticipated strength, perhaps due to a processing problem, impurities or internal flaws, the applied stress will remain below the actual yield stress.
- In some materials, not prone to brittle fracture, if the tensile stress is exceeded, the material will fail, catastrophically and without warning. Since yield stress is used as the design criterion, rather than the tensile stress, if the yield stress is exceeded, the part will deform, and thus may provide warning of an incipient failure.

Fatigue. Variable loads, cycling between large and small, or from tension to compression, can lead to failure at stresses below the normal design strength of the component. This is called *fatigue failure*, and is caused by the simultaneous action of cyclic stress,

tensile stress, and plastic strain. The cyclic stress starts the crack (nucleation); the tensile stress produces crack growth (propagation) until the remaining cross section of the part becomes too weak to sustain the load and quickly fails in overload.

The general features of a typical fracture surface can be seen in the photograph below (Figure 3).

The fluctuating tensile stress produces the striations shown in the figure, which are sometimes referred to as *beach marks* or *clamshell marks*. The presence of beach marks is a reliable evidence of fatigue crack propagation. However, depending on material and loading, the absence of fatigue striations does not mean that no fatigue occurred. The definitive call will depend upon the specifics of the material application and the events leading up to and following the failure, upon further examination under a high power optical microscope or scanning electron microscope (SEM) and upon fracture surface microstructure.

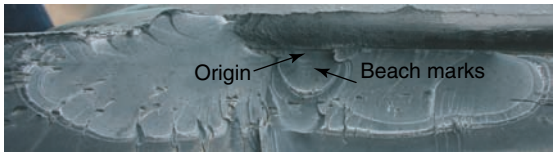


Figure 3 General features of a fatigue fracture, shown here in high-density polyethylene (HDPE). The fracture origin site can be located by tracing the center of curvature of the beach marks back to the root

In cases where cyclic loading is present, the component must be designed to accommodate cyclic loads by setting the maximum load below an endurance load defined below. The $S-N$ curve is used for this purpose, where S stands for stress and N stands for the number of cycles.

The $S-N$ curve for steel, shown in Figure 4, illustrates that, as the stress decreases, the number of cycles before failure increases. In this figure, if the stress is kept below 50% of the fracture strength, then the metal is said to have an "infinite life", or is at least able to survive more than 10 million cycles. This stress level is called the *endurance level*. The

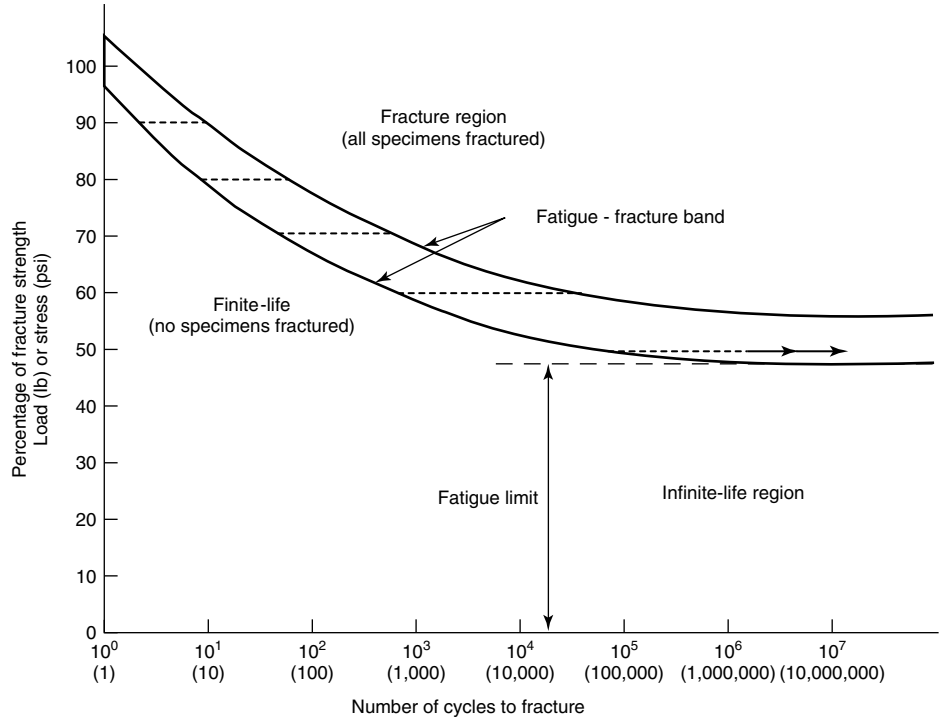


Figure 4 $S-N$ curves that typify fatigue test results for testing of medium-strength steels [Reproduced with permission from Ref. 3. © ASM International, 1986.]

American Society of Metals (ASM) Atlas of Fatigue Curves [3] may be referred to for the $S-N$ curves of particular metals.

Corrosion. Corrosion is an important mechanism of materials failure. Corrosion occurs when a metal interacts with the surrounding environment, and degrades as a result of electrochemical reaction. There are a number of specific types of corrosion; however, all follow the same basic process. Material is removed from the metal in the form of ions. This is called an *oxidizing*, or *anodic*, reaction. Ions are atoms that, in this case, have lost electrons. These lost electrons are taken up in a second reducing, or cathodic, reaction.

For corrosion to occur, three conditions described below must be present:

1. ionic pathway;
2. electrical pathway; and
3. a driving force.

The ionic pathway is generally moisture or an atmosphere with greater than 70% relative humidity. This allows the migration of the ions of corrosion.

The electrical pathway is any conducting pathway between the anode (oxidized area) and the cathode (reduced area).

The driving force may be two different metals in contact (galvanic driving force); a concentration difference such as an area of low oxygen potential on a pipe; or the difference stress in a metal part (so that the higher stress area tries to dissolve to lower its stress).

Generally, failure from corrosion is obvious in the initial inspection. To confirm corrosion and locate the root cause, it is useful to identify the required conditions listed above.

Materials Characterization

The forensic engineer or scientist specializing in materials has a variety of materials characterization techniques and protocols to use for explicitly testing or analyzing a failure. Standard testing protocols are defined by engineering and scientific organizations, such as the International Organization for Standardization (ISO) or the American Society of Testing and Materials (ASTM).

Mechanical Properties Characterization

Tensile Testing. Tensile testing methods, for example, ASTM E8, were originally developed to determine mechanical properties of metal specimens, although standard tests now exist for plastics (ASTM D638), ceramics (ISO 15490), and composites (ISO 527).

Two specimen configurations are preferred for the tensile test, one rectangular, and one cylindrical. For either configuration, the test specimen is tapered in the middle so that the cross-sectional dimensions are small compared to the length. The tapered section is called the *gauge length*. Owing to the smaller cross section in the gauge length, the stress is much higher there than anywhere else on the sample. Deformation is constrained to the gauge length, and failure will occur there, rather than where the testing fixture grips the specimen.

The sample is mounted in the tensile test instrument, held vertically by an upper and a lower grip. The lower grip is fixed, the upper on a movable crosshead. Load is applied by moving the upper crosshead upward. The specimen is elongated at a constant rate. Thus, the load and the cross-sectional area, and therefore the stress, change with time.

Internal load cells measure the applied force, from which the stress is calculated. Position sensors measure the crosshead location, from which the strain is determined. After the test is completed, the sample is removed from the fixture. In older test machines, load versus displacement is plotted continuously during the test with a pen plotter, and is then manually converted to a stress–strain curve. Modern instruments using computer control automatically generate the stress–strain curve.

Charpy Impact Test. The area under the stress–strain curve obtained in the tensile test corresponds to the toughness of the material. However, the area under the curve is dependent on the loading rate of the tensile test, which is typically very slow. The Charpy impact test, performed by ISO 179 or ASTM E23 methods, allows measurement of fracture toughness under dynamic conditions of rapid fracture.

A small rectangular sample with a cross section of 10 mm × 10 mm is machined with a center notch on one of the rectangular faces. Just as the smaller cross-sectional area of the gauge length in the tensile test specimen ensures that the fracture will occur within

that gauge length, the notch in the Charpy specimen ensures that impact fracture will occur at that notch.

The sample is placed in a test fixture consisting of a sample holder, a weighted pendulum, and a scale to measure the absorbed energy. The pendulum is raised and locked into place. On its release, the pendulum swings down and strikes the back, or unnotched, side of the specimen, breaking it.

The pendulum is almost frictionless, so without a specimen, the pendulum will swing back up to approximately the initial height. However, with a specimen in place, the pendulum swings to a lower height than it began at, as some of the kinetic energy of the pendulum is lost in breaking the specimen.

Therefore, the impact toughness measures the difference between the initial, potential, energy of the pendulum and its kinetic energy after breaking the sample. The Charpy impact energy is measured in foot-pounds.

Microstructural Characterization

Metallography. When alloying elements or impurities are present in a material, they may segregate into different regions, forming different structures or phases within the material. These structures can have a significant effect on the properties of the material. Many of these structures can be observed and identified with a microscope. If testing reveals that the mechanical properties of a material are different from the expected properties, it is frequently of interest to examine the structure of the material on a microscopic level. This examination often requires special sample preparation to reveal the microscopic features.

Originally developed to reveal the native microstructure of metals, cutting, mounting, polishing, and etching procedures exist to prepare specimens of any type of material for examination. The specimen preparation techniques are collectively grouped under the term *metallography*. The presence of flaws, such as inclusions or voids, and/or regions of differing elemental composition can frequently be identified by these methods. Preparation of a typical metal specimen is discussed below; preparation of other materials is substantially the same, although with different idiosyncratic challenges.

First, a specimen is cut to reveal a surface to examine. The typical specimen size is less than 1 inch by 1 inch. Cutting is best if done with a water- or oil-cooled cut-off wheel, to reduce any heat-induced changes. The mechanical cutting will necessarily create a zone of damage below the surface of the specimen, which will be removed later by polishing.

Once the specimen has been cut, it is encased in phenolic or an epoxy resin mixture. This hard mount will protect the sample and make polishing easier.

The mounted sample is then polished with various abrasives to remove the damaged surface zone. The polishing media starts out rough, with abrasive sandpaper of 120 or 240 grit size, and gets progressively finer, finishing with fine (0.3 μm) alumina or diamond polishing compound. In between polishing steps, the specimen surface is washed with soap and/or isopropyl alcohol and rinsed in an ultrasonic bath to remove any loose particles of the polishing media or sample fragments. It is also examined with a metallographic microscope to ensure the surface is smooth with no scratches.

Once a surface has reached the desired level of polish, chemical etching may be performed. Etching is the process of exposing a material to chemicals, such as acids, that preferentially remove materials of one phase or elemental composition with respect to other phases. This creates differences in surface height across the sample, which in turn, creates contrast during viewing with an optical microscope. Note that, owing to its different imaging technique, etching does not always produce contrast when the sample is viewed with an SEM.

Optical Microscopy. Information can frequently be gained about materials, especially fracture surfaces, from visual examination; however, it is frequently necessary to go to higher magnification to observe important features of a fracture surface or the distribution and morphology of different phases.

The typical magnification range of optical microscopy runs from 1 to 400 times the original size; 1000 times magnification is the upper limit of the magnification range for optical microscopes. When viewing fracture surfaces, which are typically rough, the depth of field, which is the ability of the microscope to keep the entire region being viewed in focus, can be a limiting factor. This is a function of the optics of the microscope, as well as of the behavior of light in general.

Electron Microscopy. When greater magnification than that provided by optical microscopy is required, an electron microscope can be used. The SEM

is a common instrument, and one that provides images that are intuitive and easy to understand. Magnification ranges from 20 to above 20 000 times the original size.

In an optical microscope, an image is created by light reflected from, or transmitted through, a specimen, the focused through a lens, and into the eye. In the SEM, the image is created by bombarding the specimen surface with electrons. A beam of these primary electrons is scanned across the surface of the specimen. The primary electrons either reflect back from the specimen surface (as backscattered electrons), or eject electrons from the atoms in the specimen (secondary electrons). A specialized detector then detects the backscattered and secondary electrons. The signal collected by the detector is converted into an electrical signal that is proportional to the detected intensity; this electrical signal is then converted to a gray scale image that is displayed on a computer monitor. Magnification occurs when the primary electron beam is scanned across a small area of the samples, and the image of the scanned surface region is displayed on a larger computer screen. This image can be displayed, printed, or saved in digital format.

Common Failure Analysis Procedures

Many features of forensic materials failure investigation are the same, despite the type(s) of materials involved. Identification of the root cause of failure is key, as careful documentation of the details of the investigation and preservation of the materials are involved. However, owing to their different mechanical behavior and failure modes, different procedures are commonly used in failure analysis of metals, ceramics, plastics, and composites. Some examples of procedures required in a forensic investigation are presented below.

- **Metals**

The fracture surface must be preserved because the fracture surface contains the information on where the failure originated and how it occurred. The fracture origin is the point from which cracking begins which may occur because of damage or a defect in that area. The fracture surface may reveal information about the time to failure, as in fast failure by overload or gradual fatigue leading to overload of the remaining metal. The fracture surface can be examined by processes such as optical microscopy or SEM but normally should not undergo any testing which is destructive in nature. For items too large to allow study of the surface without cutting, the surface can be replicated using a rubber mold material that allows the replica to be examined without destruction of the original. Collections of fracture surface photographs can be found in the ASM references discussed in Section "Information References".

- **Ceramics and Glass**

In glass, the direction of the fracture movement from the origin can be determined by observing the rib marks, also called *Wallner lines*, which are found on the fracture surface. The rib marks, which appear as concentric ridges, are almost always concave toward the origin, in the direction from which the fracture initiated. Owing to the brittle nature of glassy materials, the failure mechanisms are complex; a reference such as Failure Analysis of Brittle Materials [4] may be consulted for more details.

- **Plastics**

The fracture surface of hard plastics has failure information like a metal. Plastics, unlike a metal, may be more strongly affected by the chemical environment and temperature it is exposed to. A plastic may fracture by simple overload, fatigue, creep or stress cracking. Creep fracture occurs some time after the application of a load to the plastic. Stress cracking occurs under a comparatively small load when the plastic is exposed to a chemically active environment. In all cases, a properly preserved surface can provide information as to the cause of the failure.

- **Composites**

Depending on the type, shape, and amount of fiber reinforcement and the type and properties of the matrix containing the fibers, the properties and characteristics of composites widely varies. Composite materials where fibers, such as glass or carbon, are bonded together are found in high strength, lightweight structures such as aircraft and sport car bodies. Composites also are found in fiberglass ladders and wood siding products used in construction where wood fibers are held in place by resins. Each composite is generally unique in its creation and individual in its characteristics. The most common composite failures are due to delamination (separation) of the fiber mats because of failure of the

adhesive, or misorientation of the fiber mats due to design or manufacture.

Information References

A number of reference sources, including handbooks and textbook references, provide background and details on the materials science and engineering concepts introduced in this article. Callister's text [5] is popular for undergraduate courses in materials engineering, and provides an accessible overview of the relevant principles. Deiter [6] or Courtney [7] provide in depth information on metallurgy and failure modes of materials. For corrosion, Uhlig's text [8] is covers both corrosion mechanisms and prevention methods.

The expert will frequently consult the publications of the ASM [9]. ASM publishes a 21-volume series of handbooks, on various aspects of materials engineering and science, including but not limited to metals. Handbooks of particular interest are Volume 10 (Materials Characterization) [10], Volume 11 (Failure Analysis and Prevention) [11], Volume 13B (Corrosion: Materials) [12] and Volume 19 (Fatigue And Fracture) [13]. A selection of articles from these handbooks, specific to metals, is contained in the previously discussed Metals Handbook [2].

As previously discussed, accepted materials property testing methods are standardized by both the ASTM [1] and the ISO [14]. Descriptions of specific methods or standards are available for purchase through both organizations. For validation of methods and consistency of results, it is preferable to follow a standard method when an appropriate one is available.

References

[1] American Society of Testing and Materials (ASTM). Handbooks and standards are updated on a regular basis. The issuing organisations can provide information on the most current version. See www.astm.org for latest version information.

[2] Davis, J. (ed) (1999). *Metals Handbook*, 2nd Edition, ASM International.

[3] Boyer, H. (ed) (1986). *ASM Atlas of Fatigue Curves*, ASM International.

[4] Frechette, V. (1990). *Failure Analysis of Brittle Materials: Advances in Ceramics*, Wiley, Vol. 28.

[5] Callister, W. (2006). *Materials Science and Engineering: An Introduction*, 7th Edition, Wiley.

[6] Deiter, G. & Bacon, D. (1989). *Mechanical Metallurgy*, 3rd Edition, McGraw-Hill.

[7] Courtney, T. (2005). *Mechanical Behavior of Materials*, 2nd Edition, Waveland Pr.

[8] Uhlig, H. (1985). *Corrosion and Corrosion Control*, 3rd Edition, Wiley-Interscience.

[9] American Society of Metals (ASM). See www.asminternational.org for latest version information.

[10] Whan, R. (ed) (1986). *ASM Handbook Volume 10: Materials Characterization*, 9th Edition, ASM International.

[11] Powell, G. & Mahmoud, S. (1986). *ASM Handbook Volume 11: Failure Analysis and Prevention*, 9th Edition, ASM International.

[12] Cramer, S. (ed) (2005). *ASM Handbook Volume 13B: Corrosion: Materials*, ASM International.

[13] Dimatteo, N. (1996). *ASM Handbook Volume 19: Fatigue And Fracture*, ASM International.

[14] International Organization for Standardization (ISO). See www.iso.org for latest version information.

Related Articles

Microscopy: Light Microscopes

Reconstruction: Accident

ANASTASIA D. MICHEALS AND ROBERT N. ANDERSON

Matrix: DNA

Contemporary forensic DNA identification systems utilize a panel of fluorescent dyes to identify specific fragments, alleles, or nucleotides. Commonly, four or five dyes are used simultaneously. As the excitation of a particular fluorescent dye indicates a specific forensic outcome, it is essential that the detections systems have high specificity in their ability to isolate a particular dye or color. As there is some overlap in the emission spectra of the dyes that are simultaneously utilized, there is a need to filter out the "background" color that is drawn up with each individual dye. The technique applied is referred to as *multicomponent analysis* or *color deconvolution* [1].

A set of standards is run, each of which is labeled with one of the individual dyes that are to be utilized in the test system. Computer software is then able

to analyze the fluorescence data from each dye and assess the amount of overlap contributed by the other dyes of the system. As each dye analyzed generates data for each of the other dyes, collectively, the dataset is referred to as a *matrix* or *matrix file*. Matrix files differ for different sets of dyes and different electrophoresis systems and instruments. They are also subject to slight variation if there is a change in environmental conditions (such as temperature) and they should be regenerated for different batches of polymer or polyacrylamide (if using slab-gel systems).

The application of the matrix file occurs after the raw fluorescence data has been collected; it is usually applied automatically as one of the first steps of analysis. Applying the matrix has the effect of optimizing the signal from each dye, which sharpens the image and simplifies interpretation. A common artifact associated with this aspect of profile interpretation is known as *pull-up* and it occurs when there is incomplete separation of the spectra of two or more dyes. This creates spurious peaks of different colors that are effectively drawn up under the true peak. Pull-up is most likely to occur when the sample is overloaded, or off scale.

Reference

[1] Butler, J.M. (2005). *Forensic DNA Typing: Biology, Technology and Genetics of STR Markers*, Elsevier Academic Press, Burlington, MA.

SIMON J. WALSH

Medical Effects of Alcohol *see* Alcohol: Behavioral and Medical Effects

Medical Malpractice

Introduction

Medical malpractice charges are nearly as old as medicine. Already in the Codex Hammurapi (1700 B.C.), punishment in cases of medical malpractice was described [1, 2].

The *Constitutio Criminalis Carolina* (1532 A.D.) contains a separate chapter on medical malpractice (Chapter XXIX), and in this chapter, notions such as negligence, causality, and expert evidence can be found.

From the seventeenth to the nineteenth century it was mainly experts in forensic medicine who dealt with medical malpractice in their textbooks [1–3], e.g., Paolo Zacchia (1584–1659). In his book *Quaestionum–Medico legalium cura* (1621–1635) the sentence, "Medicus errat ommitendo et commitendo" can be found, which means that damage to the patient can not only be caused by doing something wrong but also by omitting to do the right. For a long time medical malpractice was a subject that was not frankly discussed within the medical community and public. A change was caused by the U.S. report "to err is human" (1999) [4], although

Table 1 Adverse events in hospitalized patients – an international comparison[(a)]

Study and place of study	Number of investigated cases	Adverse events (%)	Potentially preventable adverse events (%)
New York	31 000	3.7	–
Colorado and Utah	14 321	2.9	58
Australia	14 179	16.6	51
London	1014	10.8	46
New Zealand	6579	12.9	63

[(a)] According to [7]

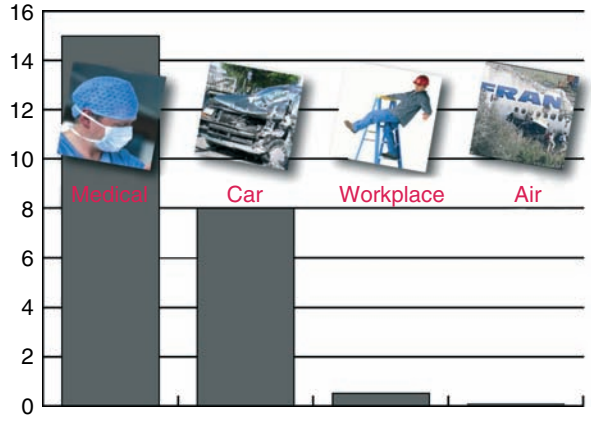


Figure 1 Incidence of annual accidental deaths in the United Kingdom (in thousands)

the message "Errare humanum est" is not new [5, 6]. But it was mainly the estimated cost caused by adverse events and medical malpractice that was surprising (Table 1). On the basis of Anglo-American studies, an estimation of adverse events in hospitalized patients in Germany was carried out. Taking into account the 16.5 million hospitalizations a year (in 2001), 31 600–83 000 deaths were estimated to be due to adverse events [7]. This would mean that more people would die as a consequence of wrong medical diagnosis or therapy than as a result of colon cancer (20 200), breast cancer (18 000), pneumonia (17 800), and traffic accidents (7700). For the United Kingdom, similar figures have been arrived at (Figure 1).

However, speaking about mishaps during medical care, adverse events, and medical malpractice requires a strict consideration of the terminology since epidemiologic research and law have different understandings of similar terms [8–12]. In the following, some aspects of medical malpractice will be addressed, which are based on international studies and also on experiences in Germany since some nationwide statistical data are available.

Definitions

Meanwhile, different institutions have proposed different definitions, e.g., the European Council, WHO, etc. [8, 9, 11–13].

Examples of definitions are given below:

An adverse event (AE) is a noxious and unintended response.

A preventable adverse event (PAE) is a noxious and unintended response that might have been prevented.

A negligent adverse event (NAE) is a noxious and unintended response due to a break of duty of care. An NAE is an equivalent of medical malpractice.

In penal law, medical malpractice is mainly defined as an AE (injury, harm) due to medical negligence. Medical negligence is defined as a preventable mistake caused by a breach of duty of care. Furthermore, there must be a causal connection between the mistake and injury, and in most jurisdictions this causal connection has to be proved without reasonable doubt.

In civil law, medical malpractice is defined as follows: "The defendant (doctor) owed duty of care to the plaintiff (patient). The doctor breaches the duty of care by failing to adhere to the standard of care expected. The standard is the quality that would be expected of a reasonable practitioner in similar circumstances. This breach of duty caused an injury to the patient." Furthermore, for epidemiologic research a definition of medical error is of importance.

Error may be error in planning or execution. An error of execution is the failure of a planned action to be completed as intended, whereas an error of planning is the use of a wrong plan to achieve an aim.

Epidemiology

Clear data on the epidemiology of medical malpractice are lacking. However, in the United States,

United Kingdom, and Australia, several studies were conducted concerning AE, PAE, and NAE but mostly only in hospitalized patients [14–36]. For ambulant medical care, hardly any data are available.

It is well known that only a small proportion of misadventures are on record, while the majority of misadventures or injuries do not become known [7, 11, 12, 37]. This fact is often illustrated by an iceberg model of accidents and errors (Figures 2 and 3).

The German Alliance of Patient Safety has carried out a systematic review of papers on the incidence of AEs, errors, etc. [33, 35]. Studies fulfilling the following criteria were included:

- original papers with data from January 1995 to December 2005;
- data collected on a well-defined reference group of patients;
- papers in which at least one of the following relevant criteria had been checked:

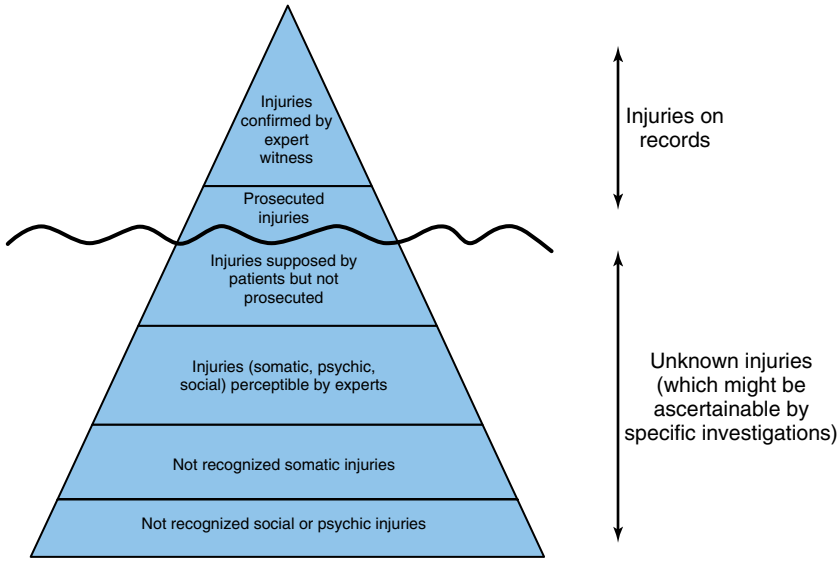


Figure 2 Types of injuries and probability of detection

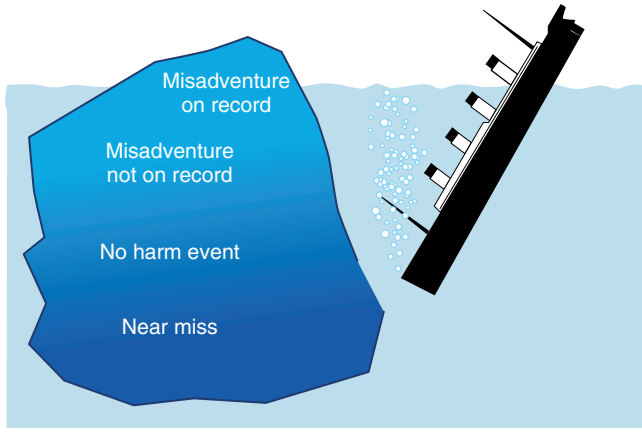


Figure 3 Iceberg model if accidents and errors

- AE
- PAE
- NAE
- errors
- near misses.

Each paper must contain a clear description of how the data were evaluated. Furthermore, clear data on factors like proportion, ratio, incidence rates must be evident. From more than 25 000 studies in PubMed and Embase, 151 studies from 25 countries with 7 686 166 patients fulfilled these criteria. The review revealed a dependency of the frequency of reported AEs, PAVs, and NAEs on the sample size: the higher the sample size, the lower the frequency. Furthermore, there was no influence of the geographic origin of the study; so the results may be more or less representative for countries with a "western" standard in health care. Among hospitalized patients, AEs can be expected in 5–10%, PAEs in 2–4%, NAEs in about 1%, and lethal outcome in about 0.1%. This would mean that on the basis of epidemiologic studies for hospitalized patients in Germany, 880 000–1 750 000 AEs, 350 000–700 000 PAEs, 175 000 NAEs, and 17 500 lethal cases would be expected. These data on the mortality due to AE and PAE were confirmed in a recent review of the German Alliance of Patient Safety [34]. According to this review, mortality due to PAE is 0.1% of all hospitalized patients. This would mean that in Germany with 17 million hospitalized patients per year, 17 000 lethal cases have to be expected. However, only a small proportion of these cases raise legal discussions. For most countries, data on the frequency of medical malpractice claims are not available. For Germany, it is estimated that only 1500–2000 cases a year are investigated by the public prosecutor [38]; these are mainly cases where death is thought to have occurred because of medical malpractice, and by a legal autopsy the cause and manner of death have to be cleared. In penal law, it is estimated that one investigation by the prosecutor is performed per 60 000 inhabitants, one piece of information of a prosecutor on 90 000 inhabitants [38]. On average, only eight cases per year are brought to a penal court, with four convictions and four stays of proceedings. For civil law, data are not available on the frequency of medical malpractice claims, but estimations speak of about 15 000 claims per year [39–41].

Every doctor is obliged to have a liability insurance. Data from the liability insurance companies are, however, not available [41, 42]. One insurance company with 108 000 insured doctors reported about 4500 incidents a year, with a settlement of cases in 30%, going to a civil court 10%, and medical malpractice confirmed at court in 4% [7].

In Germany, most claims of medical malpractice are dealt with at the arbitration committees of the medical councils. More than 30 years ago the medical councils formed these arbitration committees to make medical malpractice claims possible without applying to the courts [39, 43–50]. More than 10 000 cases per year are dealt with at the arbitration committees, and in 30% cases patient claims are confirmed.

Data of the Arbitration Committees

Since the data by the arbitration committees of the medical councils are well documented and evaluated either locally by the responsible medical chamber or nationwide by the German medical chamber, some details shall be addressed.

At the present moment, nine arbitration committees exist in Germany (Baden-Wuerttemberg, Bavaria, Hesse, Northrhine, Northern Germany, Saarland, Saxony, Westphalia, Rhineland-Palatinate). These arbitration committees have annual meetings and have recently evaluated their material nationwide (Medical Error Reporting System, MERS) [39, 43, 50]. The data of the nine arbitration committees are shown in Table 2. The largest arbitration board is that of the Northern German chambers in Hanover, and the smallest is that of Saarland. Table 2 gives details and numbers of applications per year, mode of settlement, and decisions [43].

Doctors in hospital are more often concerned with medical malpractice claims than those in private practice (Table 3). The most frequent complaints were on surgical therapy, followed by postoperative care, diagnostic imaging, informed consent, etc. (Table 4). The most frequent diagnoses resulting in malpractice claims are coxarthrosis, gonarthrosis, and fractures of lower leg and ankle (Table 5).

The bodily damages caused by medical malpractice can be classified as follows (Figure 4): no damage, minor damage, passing damage, permanent damage, and death [49]. Results concerning the severity of damages shown in Figure 4 are

Table 2 Official statistics on the work of the German committees of experts and arbitration boards (year 2004)[a]

I	Baden-Wuerttemberg	Bavaria	Hesse	Northrhine	Northern Germany	Saarland	Saxony	Westphalia –Lippe	Rhineland-Palatinate	Total
1. Total no. of applications of the past year	976	828	860	1793	4040	128	380	1777	362	11 144
2. No. of old applications not yet settled	677	729	747	2042	3887	79	143	1106	319	9729
3. No. of settled applications of last year	1057	770	793	1885	4211	119	395	1617	364	11 211
4. No. of applications still open at the end of the reporting year	596	787	813	1950	3716	88	128	1267	317	9662

(*continued overleaf*)

Table 2 (*continued*)

II	Baden-Wuerttemberg	Bavaria	Hesse	Northrhine	Northern Germany	Saarland	Saxony	Westphalia –Lippe	Rhineland-Palatinate	Total
Mode of settlement of claims of last year										
1. Withdrawn by applicant or not followed up because of lack of interest	72	87	91	158	345	13	3	103	39	911
2. Refused because of incompetence	137	30	14	110	68	6	2	94	17	478
3. Rejected because of unenforce-ability of error in treatment or doctor's duty to inform	24	9	8	0	1	0	0	8	2	52
4. Rejected because of the lapse of the application term	0	20	9	50	0	2	0	36	17	134
5. Rejected because hospital operator not governed by public law	1	0	0	0	7	0	0	3	5	16

or other case of public liability										
6. Rejected because of expert opinion	0	2	1	0	0	0	0	5	0	
7. Rejected because of preliminary proceedings, lawsuit, or legally binding court decision	17	12	13	30	42	3	8	25	2	152
8. Not decided because of appeal of one party	46	108	64	48	812	5	10	90	11	1194
9. Settled because of advising information	60	5	6	0	116	0	100	0	5	292
10. Rejected or not accepted for decision because of other reasons	0	4	8	111	0	5	8	43	20	199

(*continued overleaf*)

Table 2 (*continued*)

III	Baden-Wuerttemberg	Bavaria	Hesse	Northrhine	Northern Germany	Saarland	Saxony	Westphalia–Lippe	Rhineland-Palatinate	Total
Remaining cases accepted for decision on the merits										
1. Total	700	493	548	1378	2820	85	264	1210	246	7744
2. Error in doctor's duty to inform approved	6	12	14	25	25	0	2	4	1	89
3. Error in doctor's duty to inform questionable because of controversial facts	0	23	0	15	0	0	0	22	3	63
4. Medical malpractice and causality for injury approved	117	105	131	353	725	16	62	167	60	1736

5. Medical malpractice approved, causality negated	171	15	25	76	266	2	28	35	4	622
6. Medical malpractice approved, causality not clarified	0	7	6	28	0	0	0	16	6	63
7. Medical malpractice and error in doctor's duty to inform negated	406	359	370	830	1804	64	172	961	170	5136
8. Alternative verdict (as far as not included in 2 or 5)	0	6	2	51	0	0	0	5	2	66
9. Arbitration proposal (as far as not included in 1 and 7)	0	0	0	0	0	3	0	0	0	3

[a] Taken from [43]

Table 3 Medical malpractice and site of treatment[a]

Medical malpractice and site of treatment	Doctor's practice	Hospital
Site of treatment	2432	5303
Medical malpractice/inadequate informed consent confirmed	657	1336
Inadequate informed consent confirmed	27	45
Medical specialty 2006		
Doctor's practice		
General practitioner	389	
Orthopedic surgery	336	
General surgery	249	
Gynecology	239	
Internal medicine	218	
Traumatic surgery orthopedics	190	
Ophthalmology	149	
Radiology	98	
Dermatology and venereal diseases	79	
Urology	71	
Hospital		
Traumatic surgery orthopedics		1063
General surgery		943
Orthopedic surgery		629
Gynecology		418
Internal medicine		395
Anesthesiology and intensive care		179
Urology		166
Neurosurgery		164
ENT		134
Obstetrics		134

[a] Taken from [43]

Table 4 Claims by patients[a]

Claims by patients	2006	2005
Total no. of decisions on the merits	6751	7320
Total no. of claims (based on total no. of decision on the merits, maximum 4 claims/decision)	11 949	10 496
Most frequent claims		
Surgical therapy, performance	2998	
Postoperative therapy	861	
Diagnostic imaging	858	
Informed consent, risk	654	
Diagnostics, anamnesis, examination	636	
Therapy, conservative	599	
Diagnostics, general	539	
Therapy, pharmaceutics	477	
Indication	474	
Type of surgical therapy	397	

[a] Taken from [43]

Table 5 Most frequent diagnoses resulting in malpractice claims[a]

Most frequent diagnoses resulting in malpractice claims	2006
Decisions all together (total no.)	6751
Most frequent diagnoses	
Coxarthrosis (degenerative arthritis of hip joint)	225
Gonarthrosis	170
Fracture of lower leg and ankle	155
Fracture of forearm	130
Breast cancer	130
Intervertebral disk degeneration, lumbar	115
Traumatic injury of the knee	103
Deformity of toes/fingers	100
Degenerative injury of the knee	96
Femoral fracture	94

[a] Taken from [43] The table specifies the 10 most frequent diagnoses according to frequency. For all proceedings, a (correct) diagnosis (*ex post*) is referred to; for cases with more than one diagnosis, only the most important one is taken into consideration

based on the material of the arbitration committee in Northrhine [49].

The frequency of confirmed medical malpractice differs between doctors in private practice and hospital doctors as well as from discipline to discipline (Table 6). Medical malpractice is more often confirmed in doctors in private practice than in hospital doctors. Disciplines that are at special risks are general surgery, gynecology, and trauma surgery [43].

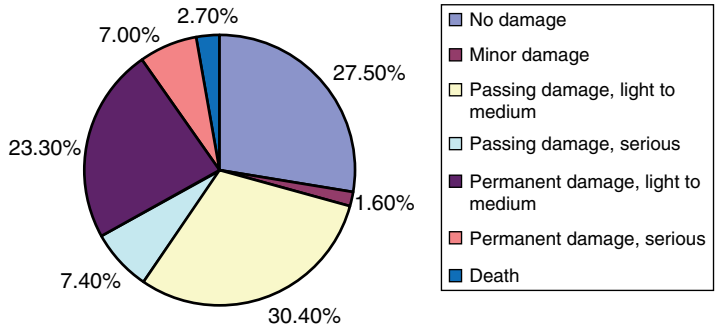


Figure 4 Bodily damage caused by medical malpractice (taken from [49]) [Reproduced with permission from Ref. 35. © Ärzteblatt, 2007.]

Medical malpractice claims have increased over the last few years (Figure 5). The error rate differs widely between the different arbitration boards (Figure 6). The arbitration board in North Rhine has the highest error rate (39%), and the board in Saarland the lowest (22%) [47].

While in Bavaria just 51 claims per 1 million residents are registered per year, other areas of Germany have many more claims per 1 million residents, e.g., North Rhine or Northern Germany [47] (Figure 7). While in Bavaria just 2 errors per 1000 physicians per year are confirmed, this rate is much higher in other German areas (Figure 8).

The probability to be charged for medical malpractice differs by the factor 3 for different regions of Germany. The probability that a medical malpractice is confirmed differs even by a factor of 6.

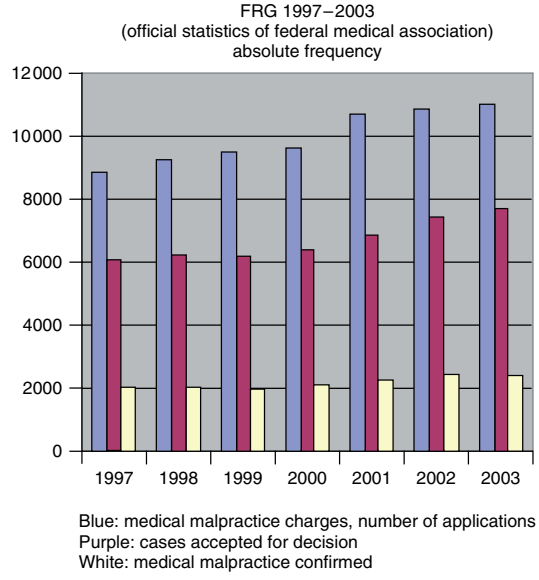


Figure 5 Medical malpractice claims, cases accepted for decision, and the number of confirmed medical malpractice over the years (from [47]) [Reproduced from Ref. 21. © Springer, 2005.]

Data from the Files of Institutes of Forensic Medicine

The arbitration committees deal predominantly with living patients [48, 49]. Lethal cases are a special subgroup and the best available data source are the files of the Institutes of Forensic Medicine [51–58]. Therefore, this special subgroup shall be addressed separately, especially since death is the severest outcome of medical malpractice.

From a separate retrospective analysis of medical malpractice claims in lethal cases, it is known that the number of cases has increased in the past years [57, 58]. Concerning medical malpractice claims in lethal cases, a doubling of cases could be observed (from 300 to 600 cases a year) in the cooperating German Institutes of Forensic Medicine (Figure 9). The autopsy rate due to medical malpractice claims varied between the institutes from 2.4 to 20%. That means that in some institutes every fifth autopsy was due to medical malpractice claims. Obviously, the prosecutor has a wide range of discretionary powers.

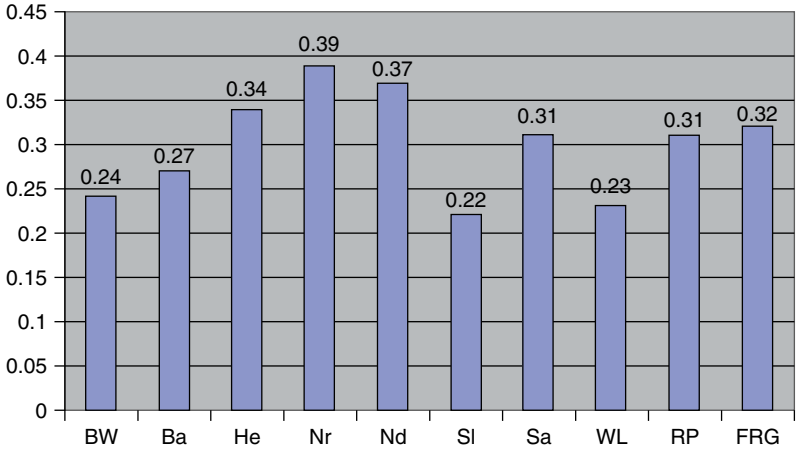
Hospital doctors are more often confronted with medical malpractice claims than doctors in private practice.

The medical disciplines concerned are surgery followed by internal medicine, general practice

Table 6 Medical malpractice charges: Arbitration committee North Rhine[a]

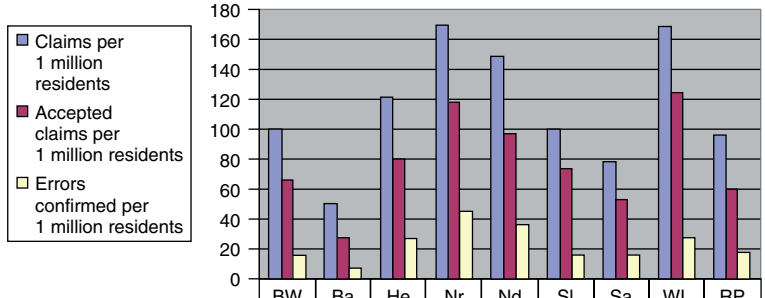
Medical disciplines and locations of medical care									
1.1.2005–31.12.2005	*n*	In %	Medical malpractice confirmed (*n*)	In %		*n*	In %	Medical malpractice confirmed (*n*)	In %
Doctors in private practice[b]	593	100.00	189	31.8	Doctors in hospitals[b]	1236	100.00	335	27.1
Orthopedic surgery[c]	86	14.50	17	19.7	General surgery[d]	264	21.36	85	32.2
General surgery[d]	76	12.82	29	38.1	Trauma surgery	175	14.16	57	32.5
Gynecology and obstetrics	70	11.80	27	38.6	Orthopedic surgery[c]	146	11.81	26	17.8
General practitioner	66	11.13	20	30.3	Gynecology and obstetrics	141	11.41	36	25.5
Internal medicine[d]	52	8.77	13	25	Internal medicine[d]	95	7.69	24	25.2
Urology	43	7.25	16	37.2	Urology	45	3.64	11	24.4
Radiology	30	5.06	22	73	Anesthesiology	37	2.99	7	18.9
ENT	30	5.06	6	20	Vascular surgery	37	2.99	8	21.6
Ophthalmology	26	4.38	2	7.7	Cardiology	32	2.59	7	21.8
Dermatology	18	3.04	5	27.7	ENT	30	2.43	6	20

[a] Taken from [49]
[b] One doctor per private practice
[c] Formerly orthopedics and reconstructive surgery
[d] Without subspecialization



BW, Baden-Wuerttemberg; Ba Bavaria; He, Hesse; Nr, North Rhine-Westphalia; WL, Westphalia Lippe; Nd, arbitration board of Northern Germany; Sl, Saarland; Sa, Saxony; RP, Rhineland-Palatinate; FRG, Federal Republic of Germany

Figure 6 Error rates according to different arbitration boards and committies of experts in Germany (according 1998–2003) [Reproduced from Ref. 21. © Springer, 2005.]



	BW	Ba	He	Nr	Nd	Sl	Sa	WL	RP
Claims per 1 million residents	99	51	122	170	149	99	79	169	96
Accepted claims per 1 million residents	66	28	80	118	97	74	53	125	59
Errors confirmed per 1 million residents	16	8	27	46	36	16	16	28	18

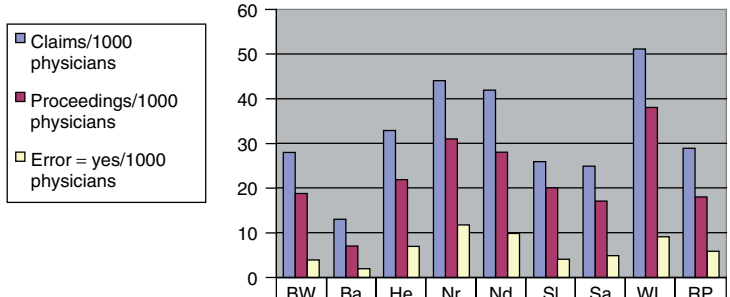
Figure 7 Claims, accepted claims, and confirmed errors per federal state per 1 million residents and year (averaged 1998–2003) (according to [47]) [Reproduced from Ref. 21. © Springer, 2005.]

and anesthesiology (Table 7). The cause of accusation was mostly conservative therapy, followed by surgical therapy, endoscopy, and intensive care (Table 8). There are different classifications of the types of mistakes, e.g. (see also Figure 10),

- medical malpractice, mistakes in information, documentation, and medication errors; and
- machine/medical product and organization.

In a separate evaluation of medical malpractice claims in lethal cases, medical malpractice charges were subdivided as follows:

1. negligence (omitting the necessary treatment)
   – insufficient diagnostics



	BW	Ba	He	Nr	Nd	Sl	Sa	WL	RP
Claims/1000 physicians	28	13	33	44	42	26	25	51	29
Proceedings/1000 physicians	19	7	22	31	28	20	17	38	18
Error = yes/1000 physicians	4	2	7	12	10	4	5	9	6

Figure 8 Claims, accepted claims, and errors per 1000 working physicians per year (averaged 1998–2003) (according to [47]) [Reproduced from Ref. 21. © Springer, 2005.]

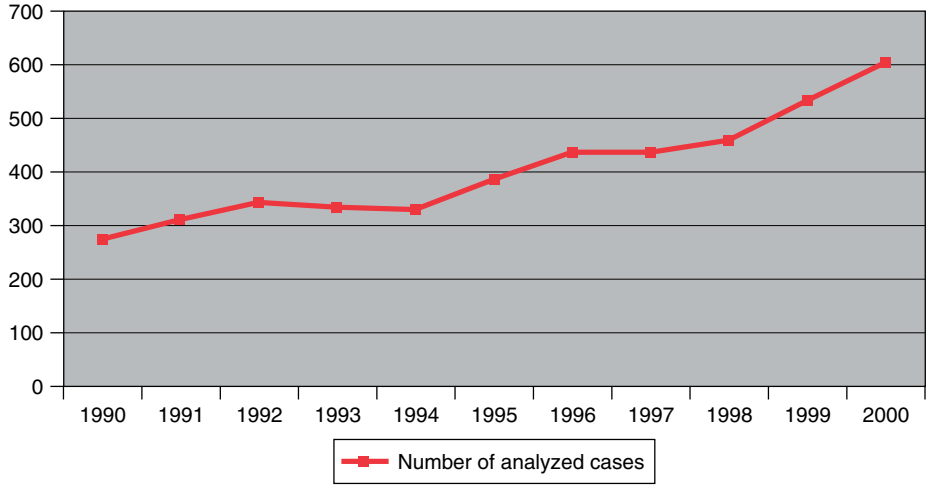


Figure 9 Number of autopsies in cases of suspected medical malpractice over the years

- delayed reaction to postoperative complications
- delayed admission to hospital and to intensive care unit

2. Complications in/or after surgery
   - intraoperative complications
   - exitus in tabula
   - complications concerning endoscopic operations
   - postoperative complications (bleeding, suture insufficiency, peritonitis)
3. Wrong treatment
   - transfusion reaction (transfusion of ABO-incompatible blood)
   - telephone diagnostics (therapeutical recommendations by telephone without visiting the patient)
   - further wrong treatment (retained instruments)
4. Mistake in care
   - insufficient prophylaxis of decubital ulcers
   - insufficient thrombosis prophylaxis

- wrong positioning during operation

5. Adverse drug event, medication errors
   - wrong drug
   - wrong dose
   - wrong application/administration
   - wrong frequency
   - disregarding drug allergy
   - misinterpretation of order given
   - illegible order.

Table 7 Medical disciplines concerned

Medical disciplines	Cases
Surgical disciplines	1272
Internal medicine	699
Practitioners	434
Anesthesiologists	156
Gynecologists and obstetricians	151
Orthopedic surgeons	126
Psychiatrists/neurologists	117
Pediatricians	86
ENT (ear, nose, and throat) specialists	72
Urologists	66
General practitioners	20

From patients and relatives, mostly the accusation is for negligence of a doctor, followed by complications within surgical therapy, wrong therapy, and medication errors (Table 9).

In nearly 40% of cases, the cause of the preliminary proceedings was the classification of the manner of death as "unclear" or "unnatural" in the death certificate (Table 10). Only in 20%, a complaint by relatives is the cause of the preliminary proceedings. Of special interest is the fact that medical malpractice is more often confirmed in doctors in private practice than in clinicians (Table 11). This is of importance since most epidemiological investigations on AE, PAE, and NAE are confined to hospitalized patients. For doctors in private practice

Table 8 Cause of accusation

Cause of accusation	
Conservative therapy	2604
Surgical therapy	1737
Endoscopy	232
Intensive care	88
Naturopathic treatment/alternative medicine	18

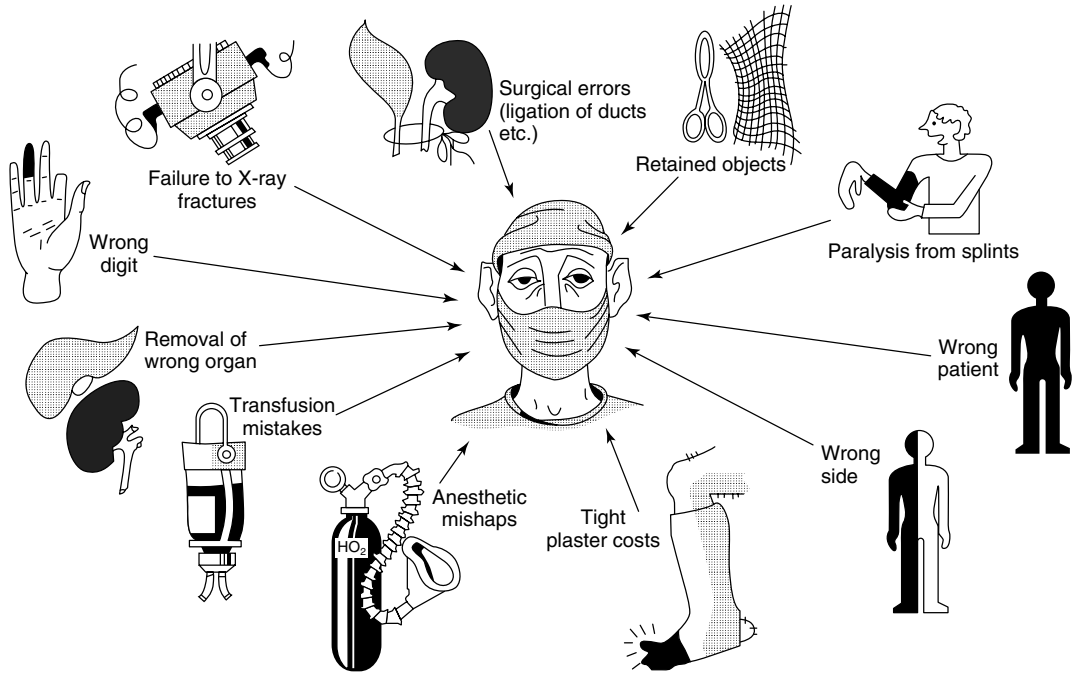


Figure 10 Types of medical malpractice (from Knight 1992)

Table 9 Classification of accusation

Type of accusation	Cases	Percentage (%)
Negligence	2158	48.5
Medication error, adverse event due to drug therapy	557	12.5
Complications within surgical therapy	1472	33.1
Wrong therapy	766	17.2
Mistake in care	320	7.2
Accusation not specified	153	3.4

Table 10 Causes of preliminary proceedings according to the analyzed documents

Cause of preliminary proceedings	Number of cases (%)
Manner of death "unclear" or "unnatural" in death certificate	1715 (38.5)
Cause of proceedings is unknown	1303 (29.3)
Complaint of offense by relatives (including friends and carers)	831 (18.7)
Complaint of a cotreating or post-treating physician	271 (6.1)
Complaint of offense by relatives as well as "manner of death unclear" or "unnatural" in the death certificate	190 (4.3)
No formal preliminary proceeding by the prosecutor, but proceeding to clear cause and manner of death	73 (1.6)
Self-complaint of the physician(s)	21 (0.5)
Complaint by the patient himself before death	18 (0.4)
Complaint by staff (especially nursing staff)	10 (0.2)
Anonymous complaint of offense	9 (0.2)
Other	9 (0.2)

and nursing staff, the rate of approved medical malpractice is 14.7 and 20.3%, respectively, much higher than for clinicians.

Table 11 Occupational group and number of approved medical malpractice and approved medical malpractice with approved causality for death

Occupational group	Number of approved medical malpractice	Approved medical malpractice with confirmed causality
Clinicians	7.8	3.5
Doctor in private practice	14.7	5.2
Nursing staff	20.3	9.3
Emergency service doctor	11.9	4.7
A group of doctors	12.6	4.0

Table 12 Distribution of malpractice charges by duration and site of treatment

Site of treatment	Treatment day 1	Treatment days 2–5	Treatment >days 5
Hospital	10	45	138
Home visit	58	8	26

Although surgery is the discipline with most accusations of medical malpractice, the rate of confirmed medical malpractice with approved causality is rather low, with 3.1%. Another difference between doctors in private practice and hospital doctors is that the interval between beginning of treatment and claim of medical malpractice is comparatively short in the case of doctors in private practice but longer in hospital doctors (Table 12). Apparently, death after short outpatient consultations place physicians at higher risk for malpractice charges.

In prospective clinical studies, medication errors make up a great part of AEs; medication errors are, however, underrepresented in other files of medical malpractice.

Adverse Drug Events

In epidemiological studies on adverse events, the most frequently seen iatrogenic injuries are nosocomial infections and adverse drug events [32, 59, 60]. However, these complications are not found in a comparable frequency in the files of the Institutes of Forensic Medicine, arbitration committees, liability

Table 13 Results of medical malpractice proceedings

	Ulsenheimer	Althoff/ Solbach 1984 (Aachen)	Mallach *et al.* 1993 (Tübingen)	Peters 2000 (Düsseldorf)[a]	Orben 2004	Bonn 2005[a]
Total	245	90	410	194 (297 accused)	601 (751 accused)	210 accused
Closing or acquittal	162 (66.1%)	80 (88.9%)	358 (87.3%)	89%	709 (94.4%)	183 (87.1%)
Sentence or settlement based on Section 153a StPO	66 (26.9%)	10 (11.1%)	52 (12.7%)	6%	42% (5.6%)	16 (7.6%)

[a] Partially the proceedings have been settled otherwise or have not yet been closed at the moment of data collection
Most preliminary proceedings are closed according to Section 170 Abs. 2 Strafprozessordnung (StPO; code of criminal procedure); that means that either no medical malpractice is evident, or the causal connection could not be proved with certainty. Only in 5–12% the doctor is either sentenced or the preliminary criminal proceeding is closed according to Section 153 a StPO (payment of an administrative fine). The higher frequency of sentences or settlement after paying an administrative fine in the material of Ulsenheimer (Table 13) is due to the fact that he, as a specialized solicitor, sees only cases that are quite severe

insurance companies, and health insurance companies [61–67].

Adverse drug reaction is a noxious and unintended response to a drug occurring at a conventional dose and used for disease in prophylaxis, diagnosis, therapy, or modification of physiological functions.

Adverse drug event is an adverse drug reaction and/or event caused by medication errors.

For instance, on the basis of a Scandinavian study [59], 50 000 lethal cases of adverse drug reactions would be expected in Germany just for hospitalizations in internal medicine; 28 000 fatal outcomes would be classified as preventable [60]. Obviously, there is a large gap between the frequency of fatalities based on epidemiological studies and those coming to the public attention. The types of medical malpractice in drug therapy are [65, 66] as follows:

- disregarding drug allergy
- mix-up of electrolyte solutions (KCl instead of NaCl)
- inadequate substitution of drug addicts
- wrong dose, especially in renal insufficiency, when antineoplastic drugs are given
- wrong administration of drugs (e.g., intrathecal administration of Vincristin when simultaneously different chemotherapeutics are given).

The typical patients in fatal adverse drug events are old, multidiseased patients using multiple drugs already on arrival at the hospital, very often receiving an additional four or more drugs during hospitalization. These patients have very frequently combined severe cardiac and pulmonary diseases.

Outcome of Medical Malpractice Claims

Not only are clear epidemiological data on the international frequency of medical malpractice claims missing from the literature but data on the outcome of such claims are also absent. For Germany, cases dealt with at the arbitration committees are well documented, and the success rate for medical malpractice claims is about 30% (see also Table 2 and Figures 5–8). In penal law, most preliminary criminal proceedings are closed according to Section 170 code of criminal procedure (i.e., that fault was either ruled out or could not be proved). Thus, there is either no medical malpractice, or the causal connection between medical malpractice and fatal outcome could not be proved with the necessary certainty [38, 68–71] (Table 12).

Conclusions

AEs in medical malpractice have gained importance because of their frequency as well as the socioeconomic costs and public and political attention. Organizations such as the Council of Europe or World

Health Organization have published recommendations on management of patient safety and prevention of AEs in health care [8, 13].

According to recent surveys of the German Alliance of Patient Safety among hospitalized patients, AEs can be expected in 5–10% of patients, PAEs in 2–4%, NAEs in 1%, and lethal outcome in about 0.1% [33–35].

Medical malpractice claims dealt with at the German arbitration committees are successful for the patient in one-third of the cases. The files of the Institutes of Forensic Medicine are indispensable for the epidemiology of AEs and improvement of patient safety for several reasons [58]:

First, the most dramatic cases are perhaps seen in forensic medicine – death due to medical malpractice[5, 6];
Second, since most studies on adverse events focus on hospitalized patients, the files of the forensic institutes cover the ambulant sector of health care as well.

Forensic pathologists can also contribute to an increase in patient safety since malpractice claims can provide a rich source of data concerning a small number of serious events [72]. A thorough evaluation of serious incidents, although less sophisticated than a root cause analysis, produces far more information than a usual hospital reporting system. Identification and evaluation of errors as well as reporting of errors may also contribute to the prevention of errors. This, among other medical disciplines, is also a task of forensic medicine [13].

References

[1] Dettmeyer, R., Preuß, J. & Madea, B. (2004). Malpractice – role of the forensic pathologist in Germany, *Forensic Science International* **144**, 265–267.

[2] Wagner, H.J. (1981). Zur historischen Entwicklung des Berichts "Ärztlicher Kunstfehler", *Zeitschrift Für Rechtsmedizin* **56**, 303–306.

[3] Virchow, R. (1870). *Kunstfehler der Ärzte*. Aktenstücke des Reichstags des Norddeutschen Bundes 5, 12–15, Berlin.

[4] Kohn, L.T., Corrigan, J.M. & Donaldson, M.S. (eds) (2001). *To Err is Human. Building a Safer Health System*, National Academy Press, Washington, DC.

[5] Madea, B. & Dettmeyer, R. (2006). Behandlungsfehler und Medizinschadensfälle – nicht nur ein Thema der Rechtsmedizin, *Editorial Rechtsmedizin* **16**, 353–354.

[6] Madea, B. & Dettmeyer, R. (2007). *Medizinschadensfälle und Patientensicherheit. Häufigkeit, Begutachtung, Prophylaxe*, Deutscher Ärzteverlag, Köln.

[7] Gutachten (2003). des Sachverständigenrates für die Konzertierte Aktion im Gesundheitswesen. Bundesrat-Drucksacke 143/03 vom 26.02.2003. Band I, Kapitel 4, S 131 ff.

[8] Council of Europe, Committee of Ministers (2006). *Recommendations REC7 of the Committee of Ministers to Member States on Medical Management of Patient Safety and Prevention of Adverse Events in Health Care*. http://www.coe.int/T/E/Social_Cohesion/Health. Cited 23 Oct 2006.

[9] Holzer, E., Thomeczek, C., Hauke, E., Conen, D. & Hochreithener, M.A. (2005). *Patientensicherheit. Leitfaden für den Umgang mit Risiken im Gesundheitswesen*. Facultas, Wien.

[10] Merry, A. & McCall Smith, A. (2001). *Errors, Medicine and the Law*, Cambridge University Press.

[11] Thomeczek, C., Bock, W., Conen, D., Ekkernkamp, A., Everz, D., Fischer, G., Gerlach, F., Gibis, B., Gramsch, E., Jonitz, G., Klakow-Franck, R., Oesingmann, U., Schirmer, H.-D., Smentkowski, U., Ziegler, M. & Ollenschläger, G. (2004). Das Glossar Patientensicherheit – Ein Beitrag zur Definitionsbestimmung und zum Verständnis der Thematik "Patientensicherheit" und "Fehler in der Medizin", *Gesundheitswesen* **66**, 833–840.

[12] Thomeczek, C., Rohe, J. & Ollenschläger, G. (2007). Das unerwünschte Ereignis in der Medizin, in *Medizinschadensfälle und Patientensicherheit. Häufigkeit – Begutachtung – Prophylaxe*, B. Madea & R. Dettmeyer (Hrsg) Deutscher Ärzte-Verlag, Köln, 13–20.

[13] World Alliance of Patient Safety (WHO) (2005). *Draft Guidelines for Adverse Event Reporting and Learning Systems*. http://www.who.int/patientsafety/events/05/Reporting_Guidelines.pdf.

[14] Bennett, R.G., O'Sullivan, J., DeVito, E.M. & Remsburg, R. (2000). The increasing medical malpractice risk related to pressure ulcers in the United States, *Journal of the American Geriatrics Society* **48**, 73–81.

[15] Brennan, T.A., Leape, L.L., Laird, N.M., Localio, R.A. & Hiatt, H.H. (1990). Incidence of adverse events and negligent care in hospitalized patients, *Transactions of the Association of American Physicians* 137–144.

[16] Brennan, T.A., Leape, L.L., Laird, N.M., Hebert, L., Localio, A.R., Lawthers, A.G., Newhouse, J.P., Weiler, P.C. & Hiatt, H.H. (1991). Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I, *New England Journal of Medicine* **324**(6), 370–376.

[17] Brennan, T.A., Sox, C.M. & Burstin, H.R. (1996). Relation between negligent adverse events and the outcomes of medical-malpractice litigation, *New England Journal of Medicine* **335**(26), 1963–1967.

[18] Brennan, T.A. (2000). The Institute of Medicine report on medical errors – could it do harm? *New England Journal of Medicine* **342**(15), 1123–1125.

[19] Brenner, J.R., Lucey, L.L., Smith, J.J. & Saunders, R. (1998). Radiology and medical malpractice claims: a report on the practice standards claims survey of the physician insurers association of America and the American college of radiology, *American Journal of Radiology* **171**, 19–22.

[20] Chopra, V., Bovill, J.G. & Spierdijk, J. (1990). Accidents, near accidents and complications during anaesthesia, *Anaesthesia* **45**, 3–6.

[21] Dean, B., Schachter, M., Vincent, C. & Barber, N. (2002). Causes of prescribing errors in hospital inpatients: a prospective study, *The Lancet* **359**, 1373–1378.

[22] Dean, B., Schachter, M., Vincent, C. & Barber, N. (2002). Prescribing errors in hospital inpatients: their incidence and clinical significance, *Quality and Safety in Health Care* **11**, 340–344.

[23] Ebbesen, J., Buajordet, I., Erikssen, J., Brors, O., Hilberg, T., Svaar, H. & Sandvik, L. (2001). Drug-related deaths in a department of internal medicine, *Archives of Internal Medicine* **161**, 2317–2323.

[24] Karcz, A., Holbrook, J., Auerbach, B., Blau, M.I., Bulat, P.I., Davidson, A., Docimo, A.B., Doyle, M.J., Erdos, M.S., Friedmann, M., Green, E.D., Hobbs, E.T., Iseke, R.J., Josephson, G.W., Kline, J., Moyer, P., Shea, D.J., Soslow, A.R., Testarmata, A.M. & Woodward, A.C. (1990). Preventability of malpractice claims in emergency medicine: a closed claims study, *Annals of Emergency Medicine* **19**(8), 865–873.

[25] Karcz, A., Korn, R., Burke, M.C., Caggiano, R., Doyle, M.J., Erdos, M.J., Green, E.D. & Williams, K. (1996). Malpractice claims against emergency physicians in Massachusetts: 1975–1993, *American Journal of Emergency Medicine* **14**(4), 341–345.

[26] Leape, L.L. (2000). Institute of medicine medical errors figures are not exaggerated, *Journal of the American Medical Association* **284**, 95–98.

[27] Leape, L.L. (2002). Reporting of adverse events, *New England Journal of Medicine* **347**(20), 1633–1638.

[28] Leape, L.L., Brennan, T.A., Laird, N.M., Lawthers, A.G., Localio, A.R., Barnes, B.A., Hebert, L., Newhouse, J.P., Weiler, P.C. & Hiatt, H. (1991). The nature of adverse events in hospitalized patients: results of the Harvard Medical Practice Study II, *New England Journal of Medicine* **324**(6), 377–384.

[29] Leape, L.L., Lawthers, A.G., Brennan, T.A. & Johnson, W.G. (1993). Preventing medical injury, *Quality Review Bulletin* 144–149.

[30] Leape, L.L., Bates, D.W., Cullen, D.J., Cooper, J., Demonaco, H.J., Gallivan, P.T., Hallisey, R., Ives, J., Laird, N., Laffel, G., Nemeskal, R., Petersen, L., Porter, K., Servi, D., Shea, B.F., Small, S.D., Sweitzer, B.J., Thomson, B.T. & Vander Vliet, M. (1995). System analysis of adverse drug events, *Journal of the American Medical Association* **274**, 35–43.

[31] Leape, L.L., Epstein, A.M. & Hamel, M.B. (2002). A series on patient safety, *New England Journal of Medicine* **347**(16), 1272–1274.

[32] Pirmohamed, M., James, S., Meakin, S., Grenn, C., Scott, A.K., Walley, T.J., Farrar, K., Park, B.K. & Breckenridge, A.M. (2004). Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients, *British Medical Journal* **329**, 15–19.

[33] Schrappe, M. (2006). *Aktionsbündnis Patientensicherheit*. Agenda Patientensicherheit 2006.

[34] Schrappe, M. (2007). *Aktionsbündnis Patientensicherheit*. Agenda Patientensicherheit 2007.

[35] Schrappe, M. & Lessing, C. (2007). Zur Häufigkeit von Medizinschadensfällen, in *Medizinschadensfälle und Patientensicherheit. Häufigkeit – Begutachtung – Prophylaxe*, B. Madea & R. Dettmeyer (Hrsg), eds, Deutscher Ärzte-Verlag, Köln, pp. 21–32.

[36] Studdert, D.M., Thomas, E.J., Burstin, H.R., Zbar, B.I.W., Orav, E.J. & Brennan, T.A. (2000). Negligent care and malpractice claiming behaviour in Utah and Colorado, *Medical Care* **38**(3), 250–260.

[37] Thomeczek, C. & Ollenschläger, G. (2006). Fehlermeldesysteme – aus jedem Fehler auch ein Nutzen? Bedeutung von Fehler-und "Incident-Reporting-Systems "in Industrie und Medizin, *Rechtsmedizin* **16**, 355–360.

[38] Orben, S.T. (2004). *Rechtliche Verantwortung für Behandlungsfehler. Hallesche Schriften zum Recht*, Carl Heymanns Verlag KG, Köln, Band 19.

[39] Berner, B. (2007). Tätigkeit der Gutachterkommissionen und Schlichtungsstellen in Deutschland, in *Medizinschadensfälle und Patientensicherheit. Häufigkeit – Begutachtung – Prophylaxe*, B. Madea & R. Dettmeyer (Hrsg), eds, Deutscher Ärzte-Verlag, Köln, pp. 33–38.

[40] Hansis, M.L. & Hansis, D.E. (2001). *Der ärztliche Behandlungsfehler*. 2 Aufl. ecomed Landsberg.

[41] Hansis, M.L. & Hart, D. (2001). Medizinische Behandlungsfehler in Deutschland, *Gesundheitsberichterstattung des Bundes* **4**, 1–15.

[42] Weidinger, P. (2007). Behandlungsfehlervorwürfe und Regulierungspraxis der Haftpflichtversicherer, in *Medizinschadensfälle und Patientensicherheit. Häufigkeit – Begutachtung – Prophylaxe*, B. Madea & R. Dettmeyer (Hrsg), eds, Deutscher Ärzte-Verlag, Köln, pp. 39–52.

[43] Bundesärztekammer (2005). Tätigkeitsbericht 2005. Kapitel 7, Ärztliche Berufsausübung, 368–370. For further information see also: Statistische Daten der BÄK. http://www.baek.de/downloads/gutachterkommissionenstatistik2006.pdf.

[44] Carstensen, G. (1990). Erfahrungen einer ärztlichen Gutachterkommission bei Behandlungsfehlern, *Zeitschrift für die Gesamte Versicherungswirtschaft* **79**, 42–53.

[45] Dettmeyer, R. & Madea, B. (2001). Iatrogene Schäden, Behandlungsfehler und Behandlungsfehlerbegutachtung, in *Handbuch Gerichtliche Medizin*, B. Madea & B. Brinkmann, eds, Springer-Verlag, Berlin-Heidelberg, New York, Tokio, Bd. 2, pp. 1457–1492.

[46] Eissler, M. (2004). Auswertung der Ergebnisse der Gutachterkommission für Fragen ärztlicher Haftpflicht bei der Landesärztekammer Baden-Württemberg für das Jahr 2002, *MedR Medizinrecht* 429–433.

[47] Eissler, M. (2005). Die Ergebnisse der Gutachterkommissionen und Schlichtungsstellen in Deutschland – ein bundesweiter Vergleich, *MedR Medizinrecht* **23**(5), 280–282.

[48] Laum, H.D. & Beck, L. (2003). Großes Interesse im Ausland an außergerichtlicher Schlichtung – Gutachterkommission für ärztliche Behandlungsfehler bei der Ärztekammer Nordrhein zieht positive Bilanz des Berichtszeitraums 2001/2002 – Kürzere Dauer der Verfahren, *Rheinisches Ärzteblatt* **57**, 19–20.

[49] Laum, H.D. & Beck, L. (2007). Vertrauensbeweis von Patienten und Ärzten. Die Ergebnisstatistik der Gutachterkommission spricht für die weiterhin hohe Akzeptanz ihrer Arbeit bei den Verfahrensbeteiligten, *Rheinisches Ärzteblatt* **1**, 15–18.

[50] Merten, M. (2007). Risikomanagement. Den Ursachen auf der Spur, *Deutsches Ärzteblatt* **104**, A1140–A1142.

[51] Bove, K.E. & Iery, C. (2004). The role of the autopsy in medical malpractice cases. I. A review of 99 appeals court decisions, *Archives of Pathology and Laboratory Medicine* **126**, 1023–1031.

[52] Dettmeyer, R., Egl, M. & Madea, B. (2005). Medical malpractice charges in Germany – role of the forensic pathologist in the preliminary criminal proceeding, *Journal of Forensic Sciences* **50**, 423–427.

[53] Kirch, W. & Schafii, C. (1996). Misdiagnosis at a university hospital in four medical eras, *Medicine (Baltimore)* **75**(1), 29–40.

[54] Kirch, W., Shapiro, F. & Fölsch, U.R. (2004). Health care quality: misdiagnosis at a university hospital in five medical eras. Autopsy-confirmed evaluation of 500 cases between 1959 and 1999/2000: a follow-up study, *Journal of Public Health* **12**, 154–161.

[55] Kricher, T., Nelson, J. & Burdo, H. (1985). The autopsy as a measure of accuracy of the death certificate, *New England Journal of Medicine* **313**, 1263 ff.

[56] Madea, B. & Dettmeyer, R. (2003). Ärztliche Leichenschau und Todesbescheinigung, *Deutsches Ärztebl* **100**, A3161–A3179.

[57] Preuß, J., Dettmeyer, R. & Madea, B. (2006). Begutachtung behaupteter letaler Behandlungsfehler im Fach Rechtsmedizin. Bundesweite Multicenterstudie, *Rechtsmedizin* **16**, 367–382.

[58] Preuß, J., Dettmeyer, R. & Madea, B. (2005). *Begutachtung Behaupteter Letaler und Nicht Letaler Behandlungsfehler im Fach Rechtsmedizin. Bundesweite Multicenterstudie im Auftrag des Bundesministeriums für Gesundheit und Soziales (BMGS)*. http://www.bmg.bund.de/cln_041/nn_599776/sid_22070317BA8CF91EC8A7A1BA5BEAD9D7/SharedDocs/Publikationen/Forschungsberichte/f-338,param=.html__nnn=true.

[59] Erster Deutscher Kongress für Patientensicherheit bei medikamentöser Therapie. Focus: Arzneimitteltherapie im Krankenhaus. 19.-20. April 2005,. Saarbrücken, unter anderem mit folgenden Vorträgen: Erikksen J: Tod durch Arzneimitteltherapie im Krankenhaus – Ergebnis und Konsequenzen der norwegischen Studie. Bates D: Medikamentionsfehler nicht negieren sondern vermeiden: Praxiserprobte Konzepte am Brigham and Womans Hospital in Boston – einem der 10 besten Krankenhäuser der USA.

[60] Schurrer, J.U. & Frölich, J.C. (2003). Zur Häufigkeit und Vermeidbarkeit von tödlichen unerwünschten Arzneimittelwirkungen, *Internist* **44**, 889–895.

[61] Lauterberg, J. & Mertens, A. (2007). Behandlungsfehler-Management in der Gesetzlichen Krankenversicherung am Beispiel der AOK, in *Medizinschadensfälle und Patientensicherheit. Häufigkeit – Begutachtung – Prophylaxe*, B. Madea & R. Dettmeyer (Hrsg), eds, Deutscher Ärzte-Verlag, Köln, pp. 57–64.

[62] Lignitz, E. & Mattig, W. (1989). *Der Iatrogene Schaden*, Akademie-Verlag, Berlin.

[63] Madea, B. (1994). Adverse drug reaction and medical malpractice, *Proceedings of the 16th Congress International Academy of Legal Medicine and Social Medicine*, Springer, Berlin.

[64] Madea, B. (1996). Rechtliche Aspekte der Arzneimitteltherapie – Aufklärung über Arzneimittel – Neben- und Wechselwirkungen, in *Innere Medizin und Recht*, B. Madea, U.J. Winter, M. Schwonzen & D. Radermacher (Hrsg), eds, Blackwell Wissenschaftsverlag, Berlin-Wien, pp. 28–49.

[65] Madea, B., Hennsge, C. & Lignitz, E. (1994). Fahrlässige Tötung durch medikamentöse therapie, *Rechtsmedizin* **4**, 123–131.

[66] Madea, B., Preuß, J., Musshoff, F. & Dettmeyer, R. (2006). Behandlungsfehlervorwürfe bei Arzneimitteltherapie – Gutachterliche Aspekte, in *Kausalität. Forensische Medizin, Toxikologie, Biologie, Biomechanik und Recht*, G. Kauert, D. Mebs & P. Schmidt (Hrsg), eds, Berliner Wissenschaftsverlag, pp. 77–99.

[67] Thomsen, H. (2006). Behandlungsfehler und Risikomanagement im AOK Institut Medizinschaden, *Rechtsmedizin* **16**, 361–366.

[68] Madea, B., Vennedey, Ch., Dettmeyer, R. & Preuß, J. (2006). Ausgang strafrechtlicher Ermittlungsverfahren gegen Ärzte wegen Verdachts eines Behandlungsfehlers, *The Deutsche Medizinische Wochenschrift* **131**(38), 2073–2078.

[69] Madea, B., Preuß, J., Vennedey, Ch. & Dettmeyer, R. (2007). Begutachtung von Behandlungsfehlervorwürfen im Strafverfahren, in *Medizinschadensfälle und Patientensicherheit. Häufigkeit – Begutachtung – Prophylaxe*, B. Madea & R. Dettmeyer (Hrsg), eds, Deutscher Ärzte-Verlag, Köln, pp. 105–120.

[70] Ulsenheimer, K. (1987). Ein gefährlicher Beruf: Strafverfahren gegen Ärzte, *Medizinrecht* **5**, 207–216.

[71] Ulsenheimer, K. (2007). Risikomanagement als Schadensprophylaxe aus der Sicht des Juristen, in *Medizinschadensfälle und Patientensicherheit. Häufigkeit – Begutachtung – Prophylaxe*, B. Madea & R. Dettmeyer (Hrsg), eds, Deutscher Ärzte-Verlag, Köln, pp. 185–196.

[72] Dettmeyer, R., Driever, F., Becker, A., Wiestler, O.D. & Madea, B. (2001). Fatal myeloencephalopathy due to

accidental intrathecal vincristin administration – a report of two cases, *Forensic Science International* **122**, 60–64.
[73] Knight, B. (1992). Legal Aspects of Medical Practice. *Churchhill Livingstone, Edinburgh*.

Further Reading

Faure, M. & Koziol, H. (Hrsg.) (eds) (2003). *Cases on Medical Malpractice in a Comparative Perspective*, Springer-Verlag, Heidelberg-Berlin.
Madea, B. & Waider, H. (1995). Ärztliche Offenbarungspflicht iatrogener Schädigungen, *Fortschritte der Medizin* **113**, 247–149.

Related Articles

Expert Opinion in Court: Civil Law Jurisdictions (France, Germany, Italy, and Spain)
Malpractice Actions against Experts

BURKHARD MADEA

Medical Records *see* Mental Status: Examination

Medical Treatment: Capacity to Consent to *see* Capacity to Consent to Medical Treatment

Memory *see* Deception: Truth Serum, Eyewitness Testimony

Memory: Adult *see* Eyewitness: Suggestibility of

Memory: Reconstructive

Memories Are Not Fixed

When we try to remember the past, we do not replay an event like a video camera might do. Memories are also not like books stored on a shelf, only to be opened and read the same each time. Rather we "reconstruct" what happened, sometimes by drawing inferences about what happened, or piecing together information that seems plausible. This is why memory is referred to as *reconstructive* in nature. Reconstruction is the process by which people build a memory of a past event at the time they are trying to remember it.

Sir Frederic C. Bartlett was one of the first researchers to champion the reconstructive nature of memory, as is described in his classic 1932 book *Remembering* [1]. To demonstrate how memory was a reconstructive process he had subjects read an unusual story and then asked them to recall it from memory at varying intervals that ranged from minutes to hours or even years. Sometimes he asked people to tell the story to another person, who told it to another and so on. Bartlett found that the story was often not well remembered, and that people made systematic errors when they tried to recall it or tell it to someone else. For example, they tried to make the story more coherent than it really was, and they often added details that were not part of the original story. He also found that as subjects reconstructed the story from memory, accurate recall was the exception rather than the rule.

Bartlett's experiments challenged the belief that memories were stored as static unchanging traces in the brain and that recall meant merely reexciting, or stimulating, these traces. The act of remembering is the end result of a complex process that begins when information is first presented and is stored in the mind. That complex process involves a number of stages that are fundamental to how memory works.

How Memory Works

There are three essential stages in the memory process: acquisition, retention, and retrieval [2]. Acquisition is the first stage: when an individual first experiences an event, devotes attention to it, and

encodes it into memory. Retention is the second stage and it refers to the passage of time that occurs between the original event and recollection of the event. The final stage of the process is retrieval, when the individual tries to remember the event (or reconstruct it). Errors can enter memory at any or all of these three stages.

During the acquisition stage, errors can be introduced when details of the event are not properly encoded and stored in long-term memory. In order to recall an event, an individual must first devote attention to it. When a person distorts information at the acquisition stage, the errors are sometimes called *constructive* errors. With poor lighting or long distances or a minimal chance to view the material, errors are more likely to occur.

Errors can also occur during the retention stage of memory. The length of time that passes between the initial encoding of an event and the recall of that event affects how much detail will be recalled. In 1885, Ebbinghaus conducted classic experiments documenting how recall deteriorates over time [3]. Using only himself as a subject, Ebbinghaus memorized lists of non-sense syllables and tested his recall over time. He found that forgetting occurs rapidly at first and then becomes more gradual as time goes on; these results were then plotted onto his now famous "forgetting curve".

The passage of time weakens our memory for an event and makes us more susceptible to errors. One way that errors can be introduced is through the acquisition of misleading information, a phenomenon known as the *misinformation effect* [4]. In studies of the misinformation effect, subjects witness a complex event such as a car accident and half are subsequently provided with misleading information about the event. In one study that used a car accident as the target event, participants in the misinformation group were given the suggestion that the traffic sign seen in the event was a yield sign instead of the stop sign that was actually seen. When later asked to provide a description of the initial event, participants who received the misleading information were less accurate than those who had not been exposed to any misinformation. These participants incorrectly claimed that they had seen a yield sign. The longer the time interval between witnessing the event and the introduction of the misinformation, the more memory for the initial event was impaired.

At the retrieval stage, memory for the target event can be altered by the way that questions are worded [2]. Even the change of a single word can affect memory for an event. In one study, participants were shown a short film of a car crash and then were given a questionnaire about the events that they had just seen. When participants were asked, "Did you see the broken headlight"?, significantly more of the participants reported that they had seen the broken headlight, a detail that was not in the film, than when participants were asked, "Did you see a broken headlight"? The change from "a" to "the" makes a significant difference because the use of the word "the" suggests that the item exists.

In another study on the influence of question wording, participants watched an automobile accident and then were asked about the speed of the cars involved in the accident. A question like "How fast were the cars going when they smashed into each other"? led to higher estimates of speed than the question "How fast were the cars going when they hit each other"? Moreover, the "smashed" question led people to be more likely to claim that they saw broken glass in the accident, a detail that did not occur. It is also important to note that once an individual recalls an event in a particular way, this is the new construction of the event.

These errors reveal something about the reconstructive nature of memory. When people hear the word "smashed", it is commensurate with a severe accident, one that occurred at higher speeds, and thus they infer that the speed was higher. Moreover, since broken glass is often associated with severe accidents they may infer that it occurred and think, in this case wrongly, that they saw it.

Another line of work that shows the reconstructive nature of memory is work on "pragmatic inference" [5]. A pragmatic inference is an assertion that leads people to believe that they experienced something that was not actually experienced. When people were presented with a statement like "The karate champion hit the cement block", they will often remember hearing "The karate champion broke the cement block". This is not a logical inference; the champ could have hit the block but not broken it. However, many people will draw the inference and misremember what they heard.

People also can come to remember autobiographical events that never happened. These wholly fabricated incidents are known as *rich false memories* [4].

In one study, the relatives of subjects created scenarios and implanted false memories of being lost for an extended time in a shopping mall and being rescued by an elderly person, at the age of six. This procedure for planting false memories has come to be known as the *lost-in-the-mall* technique. Using strong suggestion, people have been led to remember falsely that they were lost as a child, that they were attacked by an animal, or that they nearly drowned and had to be rescued by a lifeguard.

Rich false memories can also be planted through the use of other techniques such as guided imagination, suggestive dream interpretation, hypnosis, and the use of doctored photographs. These techniques have produced false memories in approximately 30% of the people who receive the suggestion. One criticism of the memory implantation studies is that the participants may be recalling actual memories, not false ones. In order to counter this criticism, researchers have successfully demonstrated that people can also be given impossible memories like meeting Bugs Bunny at Disneyland.

Source monitoring errors also highlight the reconstructive nature of memory. Source monitoring errors refer to the phenomenon of recalling someone or something but forgetting the source of the original object of recall [6]. For example, you may recognize a face but be unable to remember where you first encountered the person. In most cases, this failure to recall the source of the original memory does not have major implications, but for those accused of criminal acts, the consequences can be quite serious. In one known case, a psychologist was accused of rape based on the victim's detailed recollection of her attacker [3]. He was later cleared when it was discovered that he was on live television at the time, which the victim was watching during her attack. She recalled his face clearly but had misattributed the source.

Errors of source attribution can also occur when we are attempting to create something wholly our own. *Cryptomnesia* refers to the act of unknowingly attributing the work of someone else to oneself [3]. Without meaning to, people occasionally plagiarize other works or adopt the memories of other people as their own. Carl Jung reported on this phenomenon in the early 1900s noting that Friedrich Nietzsche had taken part of a story written by the German physician and poet Kerner. In a more recent case, Kaavya Viswanathan, a Harvard sophomore and author, claimed that her plagiarism from two novels by Megan McCafferty was entirely unconscious and unintentional.

On occasion, we sometimes unintentionally adopt the memories of siblings or close others. Common memories that have been adopted from someone else are for achievements or suffered misfortunes. We recall these events as happening to us, and are genuinely surprised when we discover that these memories are not ours. The memories that we are least likely to appropriate are those pertaining to wrongdoing, perhaps because of our self-serving biases. The adoption of another person's memory is not an intentional act but rather, it reflects the reconstructive nature of memory and how errors can insert themselves in our recollections. Knowing that these errors can occur unintentionally should make us less irritated and more forgiving when we observe them.

To reiterate, memories may sometimes become less clear over time, but they also change for other reasons. They are continually constructed and reconstructed, and for these reasons will often differ from genuine reality. Unfortunately, without independent objective corroboration, there is no way to assess whether a memory is accurate or not.

Applications

The courtroom is one arena where the reconstructive nature of memory is apparent. Eyewitnesses are frequently the cornerstone of the prosecution's case against a defendant, and provide compelling evidence to jurors (*see* **Eyewitness Testimony**; **Interrogative Suggestibility**; **Memory: Repressed**; **Eyewitness: Suggestibility of**). Yet eyewitness testimony is often mistaken, and such mistakes can be disastrous. Analyses of wrongful convictions, where individuals have been exonerated on the basis of DNA, have shown that the major cause is faulty eyewitness memory [7]. When a witness takes the stand and professes certainty about the defendant's guilt, it is powerful evidence. Unfortunately, confidence is not equal to accuracy and confidence can be inflated by many factors. One such factor is confirmatory feedback. When an individual is given positive feedback about the identification that they have made, they become more confident about their choice. The feedback need not even be verbal, as in a remark "Good job, you picked

the suspect". Even a nod can serve as positive feedback, and artificially inflate a witness's confidence, making them more compelling when they testify in court.

Rich false memories can also appear in the courtroom. In one case, a sheriff's deputy named Paul Ingram confessed to sexually abusing his daughters and being involved in a satanic cult that involved animal sacrifice and infanticide [8]. Ingram was highly suggestible and was imprisoned despite a lack of evidence indicating that the cult even existed. Sadly, Ingram is not the only individual who has been imprisoned or had a life destroyed by false memories. In the early 1980s, dozens of individuals were swept up in accusations of child abuse involving satanic cults. Most, but not all, of these people have been released from prison but the emotional scars of the accusations remain.

Memory is inherently a reconstructive process, one that works well in most cases but can occasionally result in errors with dire consequences. Since memory errors can have such serious consequences, it is important that we keep in mind that memory is malleable. When we seek to recall, we seek to reconstruct and errors can occur at any stage.

References

[1] Bartlett, F.C. (1932). *Remembering*, Cambridge University Press, London.

[2] Loftus, E.F. (1996). *Eyewitness Testimony*, Harvard University Press, Cambridge.

[3] Schacter, D.L. (2001). *The Seven Sins of Memory: How the Mind Forgets and Remembers*, Houghton Mifflin, Boston.

[4] Loftus, E.F. (2005). Planting misinformation in the human mind: a 30-year investigation into the malleability of memory, *Learning & Memory* **12**, 361–366.

[5] Chan, J.C.K. & McDermott, K.B. (2006). Remembering pragmatic inferences, *Applied Cognitive Psychology* **20**(5), 633–639.

[6] McNally, R.J. (2005). *Remembering Trauma*, The Belknap Press of Harvard University Press, Cambridge.

[7] Doyle, J.M. (2005). *True Witness: Cops, Courts, Science, and the Battle Against Misidentification*, Palgrave Macmillan, New York.

[8] Ofshe, R. & Watters, E. (1994). *Making monsters: False memories, Psychotherapy, and Sexual Hysteria*, University of California Press, Berkeley.

TIAMOYO PETERSON AND ELIZABETH F. LOFTUS

Memory: Repressed

Memory, Repressed

The term "repression" was popularized by Sigmund Freud in the late nineteenth century. By Freud's account, repression consisted of individuals making certain events that elicited painful or disruptive effects on the psyche inaccessible as a defense mechanism for the self. Although the term and various definitions of repression have existed for more than a century, there has been essentially no substantial proof that massive repression exists. There are, however, alternative explanations for what looks like "repression". Some individuals do not think about unpleasant experiences and are later reminded of them. This is not repression, but ordinary forgetting and remembering. Moreover, some individuals can have false memories planted in their minds as a result of suggestion, giving the appearance that an event was not being remembered (*see also* **Eyewitness: Suggestibility of**; **Hypnosis and Memory**; **Recollective Accuracy of Traumatic Memories**).

Despite the lack of verifiable proof of massive repression [1, 2], allegations of repression can have dramatic consequences. The existence of repressed and recovered memories became a fierce topic of debate due to alleged cases of childhood sexual abuse (CSA), many of which involved memories recovered by suggestive therapy techniques. A number of these cases resulted in legal battles and families being torn apart. This entry offers a brief overview and history of the repressed memory debate, legal claims and cases that resulted from the debate, and research that has attempted to examine the existence of repression.

Surge in Recovered Memories and their Effects

Media Contribution

Claims of recovered memories of CSA blossomed in the late 1980s and early 1990s. Accusations of CSA, including claims of satanic ritual abuse, peaked in the two-year period of 1991–1992. Since the early 1990s, the number of CSA accusations as a result of recovered memories has steadily declined. The media undoubtedly contributed to the surge in

repressed memories that were recovered in therapists' offices throughout North America. A notable case that was brought into the public eye as a result of the media coverage was a murder case involving George Franklin.

In 1990, George Franklin stood trial for a murder that had occurred more than 20 years earlier. Franklin was accused of murdering Susan Nason, a friend of his daughter, Eileen. Although Eileen was 8 years old at the time of her friend's murder, her testimony was key. Eileen testified that she had witnessed her father murder her friend and repressed the memory for 20 years. When her "memory" returned in the late 1980s, she contacted the police. George Franklin was found guilty of murder based solely on Eileen's testimony of her recovered memory. It was the first time that an American citizen had been tried and convicted of a murder on the basis of a recovered memory.[a]

In addition to the media coverage of trials involving recovered memories, a number of people began to write self-help books to assist victims of CSA. One of these books, *The Courage to Heal* by Ellen Bass and Laura Davis, encouraged victims to recover memories of abuse and to confront alleged molesters. There were a number of supporters of *The Courage to Heal*; however, there were also those who criticized its message. The critics contended that it caused widespread harm to many innocent people – both the alleged perpetrators, and the accusers whose lives may have been negatively affected by false accusations. Bass and Davis have no formal training in psychiatry or psychology, and some critics have argued that their book encourages the recovery of memories that may not be true. Because *The Courage to Heal* has sold over 800,000 copies, many people feel that it has contributed greatly to the surge in the number of recovered memory cases by engendering people to embrace their recovered memories, regardless of the existence of corroborative evidence of abuse.

Statute of Limitations

Another development that fueled the controversy over repressed memory concerned changes in how claims of repression were handled by the judicial system. Washington State was the first state to toll the statute of limitations for repressed memory cases. This meant that people with newly recovered memories of CSA had three years from the time of remembering to sue their parents, other relatives, or any alleged molesters. A number of other states followed Washington's example, and as a result, thousands of lawsuits were filed.

Most provisions applicable to victims of CSA fall into the categories of "minority tolling" and "delayed discovery doctrine". A tolling doctrine postpones the date from which a statutory period begins. In the case of minority tolling, a statute might run three years after the child turns 18, the legal age of majority. This delay is thought to be beneficial because some children may not feel comfortable coming forward to report abuse while still under the care of their abuser.

The delayed discovery doctrine has typically been used in cases involving medical malpractice. For example, the statute of discovery would be extended for a patient who had surgery and later realized that her abdominal pain resulted from her doctor failing to remove a sponge during a prior surgery. Thus, at the time it was determined that the sponge was causing the abdominal pain, the statute of limitations for a malpractice suit would begin to accrue. Those in support of applying the delayed discovery doctrine in recovered memory cases purport that such a medical malpractice claim is analogous to an individual who alleges recovery of CSA memories in that the source of their pain would be realized when the memories are recovered. Such an individual would have some number of years from the date they assert the memory was recovered in which to pursue legal action.

Because the statutes of limitations were tolled, a number of victims were able to sue their alleged abusers more easily. As a result, thousands of cases were brought to the courtroom. Soon there became another way in which repressed memories made their appearance in the court. People who were accused of CSA sued their accusers' therapists for planting false memories of childhood abuse that led to the original accusations. In addition, some patients who had "recovered" memories of abuse began to realize that those memories were not true. These patients retracted their claims and many of them sued their therapists. They are categorized into the group known as "Retractors".

Retractors

There have been a number of CSA accusers who have reestablished family relationships and acknowledged that their prior accusations were false. These

retractors typically blame their therapists for suggesting to them that they were victims of CSA and for encouraging memory recovery through questionable therapeutic techniques such as guided imagery, "truth serums" (sometimes termed *Amytal interviews*), and hypnosis.

In 1997, a landmark case ensued in which a retractor sued her therapist who helped her "recover" memories of CSA. Ms. Burgus, a patient of Dr. Braun's, originally sought treatment for postpartum depression, but was diagnosed as having multiple personalities. Dr. Braun believed that her symptoms resulted from sexual and ritual abuse including cannibalism and torture. He purported that this abuse was the cause of her forming multiple personalities. Even though Ms. Burgus had no recollection of the sexual and ritual abuse, Dr. Braun encouraged her to try and remember these instances of abuse through hypnosis. Ms. Burgus eventually realized that these allegations were not true. She retracted her previous abuse accusations and sued her former therapists, including Dr. Braun and the hospital. The lawsuit was settled for 10.6 million dollars.

False Memory Syndrome

Although therapy helps a number of people, it can have detrimental effects in cases involving recovered memories. This is particularly true in situations where there is no history of sexual abuse. For example, when a daughter accuses her father of child abuse, the family can be devastated. Often it forces other family members to choose sides; this has caused some families to become torn apart as a result of "recovered" memories (for the landmark legal case in this area see *Ramona v. Ramona*, No. 61898, California Superior Court, Napa County, 1994; reported in Johnston, Moira (1997) *Spectral Evidence*. Boston: Houghton Mifflin).

In addition to causing strain on family life, patients may develop new symptoms unrelated to the primary concern for which the patient sought therapy in the first place. These symptoms include false beliefs and memories of having been abused, a syndrome now referred to as *False Memory Syndrome*. This syndrome is characterized by flashbacks, which include detailed memories, and even hallucinations and delusions of the abuse. Thus, the patient believes that she has specific recollections of abuse that may not have occurred.

Families and professionals who saw the need to prevent the spread of False Memory Syndrome formed an organization in 1992 called the False Memory Syndrome Foundation (FMSF). The FMSF also supports the reconciliation of families who were torn apart by claims of repressed and recovered CSA.

Research in the Field

Like the topic itself, research in the field of repressed and recovered memories has been controversial. There are several different methodologies that have been used to study such alleged memories. The most common types of studies are retrospective, prospective, and case histories. Retrospective studies rely on the self-report of alleged abuse years after the actual event was said to have happened. Prospective studies use later self-reports of alleged abuse, but attempt to corroborate the abuse through various channels such as police reports or social worker reports. Lastly, case histories rely on data collected from an individual's personal experience. In general, most of the data has been collected from women who were alleged victims of CSA.

In one example of a retrospective study, researchers Briere and Conte [3] asked participants, who were recruited by their therapists, if there was ever a time in which they failed to remember their first instance of abuse. Those who reported forgetting the abuse at some point reported that the abuse occurred at a very young age compared to the reported age of the participants who claimed continuous memories of abuse. Overall, more than half of the participants reported that there was some point in time that they did not remember the abuse. However, from this study, it is problematic to conclude that repression exists. Not specifically remembering an event does not necessarily constitute massive repression. In fact, this lack of remembering can easily be accounted for by ordinary forgetting. Because of this, the results of this particular study cannot be satisfactorily interpreted, nor can they be considered adequate proof that memories can be repressed.

Another researcher, Melchert [4], conducted a retrospective study to clarify the confusing findings of earlier studies. In a nonclinical sample of individuals, of those reporting physical, emotional and/or sexual abuse, around half reported that they did not have

continuous memories of the abuse. Many participants in this study who claimed to not remember being abused later reported that they either consciously avoided thinking about the abuse, could have remembered the abuse if they had been reminded, or did not realize that the abusive acts against them were in fact abuse until a later time.

Another strategy for examining "repression" involves the use of a prospective design. Here, there is documentation made by an agency or individual (e.g., social worker) at the time of or shortly after the alleged CSA occurred. Some time after the abuse has been reported, individuals are contacted to determine what they remember about their prior claims of abuse.

In a prospective study conducted by Williams [5], women were contacted 17 years after CSA had been reported to authorities. Of these participants, 38% failed to recall the previously reported abuse to the interviewer. Thus, some consider this study to be a prime example of evidence that memories of CSA may be repressed. However, critics assert that participants may have remembered the abuse, but may have not wanted to discuss such an uncomfortable event with the interviewer. Additionally, in this study the younger the participant was at the time of the reported abuse, the less likely they were to remember the incident. This difference in remembering due to age could be explained by childhood amnesia, which is discussed later in this entry. More recently, in another prospective study conducted by Goodman and colleagues [6], only 8% of participants who had previously reported abuse failed to report remembering the abuse at the later interview. Even the rather low percentage could be the result of unwillingness to tell rather than inability to remember.

Another method for examining repression is with individual case histories. Some mental health professionals report on a patient who they claim had a repressed and recovered memory. The case of "Jane Doe" is a famous case history within the field of repressed and recovered memories [7, 8]. In the mid-1980s Jane Doe's parents filed for divorce and engaged in a vicious battle over custodial rights of Jane. As a result of this custody dispute, Jane, who was 6 years old at the time, was psychologically evaluated. A psychiatrist supported the abuse accusation, and Jane's father obtained custody while Jane's mother lost the rights of visitation.

In 1995, when Jane Doe was 17 years old, the same psychiatrist videotaped Jane again. At first Jane failed to remember being abused by her mother and then, during the same videotaped session, she appeared to spontaneously recover memories of the abuse. The psychiatrist asserted that this particular case was solid evidence of traumatic amnesia and recovery of memory. Following this claim, some critics who questioned the validity of this case history investigated other facts surrounding the reported abuse of Jane to determine whether this case offered conclusive evidence for repression. These investigators discovered that Jane had discussed the alleged abuse on many occasions between the first and second video tapings, thus undermining any claim of repression. Additionally, documented facts were brought to light that suggested that Jane may have not been abused by her mother at all.

Alternative Explanations for Repression

There are a number of possible sources from where richly detailed memory reports may come. The reports of victims who have claimed to recover memories could reflect true memories that have simply been forgotten by normal memory processes. The "recovery" of these memories may have been triggered by a retrieval cue. Another explanation is that the memories could reflect lies. Finally, it is possible that these memories are the result of therapists' suggestions and other activities in group therapy sessions that planted false beliefs.

True Memories

Instances where individuals claim to have recovered memories may actually entail ordinary forgetting and remembering. For example, an individual may remember an event that occurred at a much later date that they had not thought of in a long time. This true memory may be brought to mind by a trigger in the environment, such as a particular setting. However, this would be a case of ordinary forgetting and not repression and recovery of a memory. Additionally, failing to think about an event such as sexual abuse for any given amount of time does not necessarily mean that the memory for the event was repressed. In any given case in which a person reports recovering a memory, the statement could be an accurate reflection

of the individual's experience; however, it may be an instance of ordinary forgetting.

False Memories

It can be difficult to discriminate true memories from false ones; however, there are a few situations where we can say with near certainty that the reported recovered memory is false. For example, some people have claimed that they remembered abuse inflicted upon them before they were two years old. There have even been people who claim to have recovered memories of abuse that occurred when they were six months old. As adults we do not have concrete and reliable episodic memories for events that occurred in the first couple years of our lives. This phenomenon is known as "childhood amnesia". Therefore, it can be concluded that these very early "memories" are almost certainly false.

For those individuals whose recovered memories cannot be explained by childhood amnesia, research on suggestibility and the malleability of memory may explain their situations. Researchers have shown that humans are rather susceptible to forming false memories. In addition, a number of individuals believe those memories to be true and even embellish upon the details of those false memories. In research where rich false memories have been planted, a significant minority of subjects have been led through suggestion to believe that they had experiences like being lost in a shopping mall [9] or being attacked by a vicious animal [10]. These studies have shown that people are even susceptible to embracing false memories as their own. Once planted, the individual can report the false event with a great deal of detail, confidence, and emotion.

Conclusion

After decades of research in the field, there is little, if any, substantial evidence that massive repression exists. Moreover, there are other possible explanations for what might look like the recovery of a repressed memory. These include ordinary forgetting or being subjected to suggestible circumstances. Given the massive evidence that has been gathered on the malleability of memory, it should not be surprising that at least some individuals who purport they have recovered memories of past abuse may have actually formed false memories.

End Notes

[a.] George Franklin served six and half years in prison before his conviction was eventually overturned.

References

[1] Holmes, D.S. (1994). Is there evidence for repression? Doubtful, *The Harvard Mental Health Letter* **10**, 4–6.

[2] Pope Jr, H.G. & Hudson, J.I. (1995). Can individuals "repress" memories of childhood sexual abuse? An examination of the evidence, *Psychiatric Annals* **25**, 715–719.

[3] Briere, J. & Conte, J.R. (1993). Self-reported amnesia for abuse in adults molested as children, *Journal of Traumatic Stress* **6**, 21–31.

[4] Melchert, T.P. (1996). Childhood memory and a history of different forms of abuse, *Professional Psychology: Research and Practice* **27**, 438–446.

[5] Williams, L.M. (1994). Recall of childhood trauma: a prospective study of women's memories of child sexual abuse, *Journal of Consulting and Clinical Psychology* **62**, 1167–1176.

[6] Goodman, G.S., Ghetti, S., Quas, J.A., Edelstein, R.S., Alexander, K.W., Redlich, A.D., Cordon, I.M. & Jones, D.P.H. (2003). A prospective study of memory for child sexual abuse: new findings relevant to the repressed-memory controversy, *Psychological Science* **14**, 113–118.

[7] Loftus, E.F. & Guyer, M.J. (2002). Who abused Jane Doe? The hazards of the single case history, *Skeptical Inquirer* **26**(3), 24–32.

[8] Loftus, E.F. & Guyer, M.J. (2002). Who abused Jane Doe? The hazards of the single case history, *Skeptical Inquirer* **26**(4), 37–40, 44.

[9] Loftus, E.F. & Pickrell, J.E. (1995). The formation of false memories, *Psychiatric Annals* **25**, 720–725.

[10] Porter, S., Yuille, J.C. & Lehman, D.R. (1999). The nature of real, implanted, and fabricated memories for emotional childhood events: Implications for the recovered memory debate, *Law and Human Behavior* **23**(5), 517–537.

NICCI B. FOWLER, KALLY J. NELSON AND ELIZABETH F. LOFTUS

Memory Accuracy *see* Recollective Accuracy of Traumatic Memories

Memory of Adults: Suggestibility *see* Eyewitness: Suggestibility of

Memory of Children: Suggestibility *see* Children: Suggestibility of

Memory of Traumatic Events *see* Recollective Accuracy of Traumatic Memories

Mens Rea *see* Automatism as a Defense to Crime, Behavioral Science Evidence

Mental Defect *see* Behavioral Science Evidence

Mental Disease *see* Behavioral Science Evidence

Mental Health *see* Psychopathology: Terms and Trends

Mental Health: Special Services *see* Disaster Mental Health

Mental Health Courts

Introduction

In a perfect world we should not need mental health courts, let alone a literature about them. But we do. Realistically, in a less than perfect world we should perhaps aspire to see mental health courts as part of our future but not as the full "answer to the problem" as they are currently seen. "The problem" is that over the past 15–20 years the criminal courts have had to contend with ever-increasing numbers of mentally disordered individuals coming through the courthouse doors. This article describes the basics of mental health courts, how they operate, and examines how and why these courts have become an integral part of both the criminal justice and mental healthcare systems.

Background

The provision of mental healthcare services in most Western European and North American communities has witnessed a steady decline over the past number of decades. Beginning with the deinstitutionalization movement occurring in the later half of the twentieth century, adequate mental healthcare services became increasingly scarce. The reality today is no different; mental healthcare systems are generally underfunded and overexpended. There has been a movement toward community-based treatment of major mental illness, which in many instances appears to be insufficient. Despite what was promised, the money saved with the closure of hospitals has typically not been reinvested in community treatment. At the same time, while it is acknowledged that community-based mental healthcare is an important component of the mental healthcare system, it cannot address the needs of many of the more seriously afflicted individuals.

It goes without saying that decreasing mental healthcare services does not lessen the needs of those

members of our society who rely on such services. For these individuals closed doors and long-wait lists have offered little. Regrettably, those unable to receive adequate services often find themselves attracting the attention of the criminal justice system, leaving that system to "sort out the mess".

The effects of our emancipated mental healthcare system are varied; however, a site where it has had a particularly acute impact has been the criminal justice system. Being the social safety net of last resort, the criminal justice system has swelled with mentally disordered accused and has struggled to meet the rising demands placed upon it. In some jurisdictions, mentally disordered accused entering the criminal justice system has increased at a rate in excess of 10% per year over the past dozen years [1].

Unfortunately, the criminal justice system has become the surrogate mental healthcare provider in most societies. A *criminalization* of mental illness has occurred: a shifting of responsibility onto the criminal justice system for the provision of basic mental healthcare services.

Understandably, the criminal justice system has not delivered. The criminal justice system was neither designed nor intended to address society's responsibility to the mentally disordered individual. Even a basic familiarity with the traditional criminal justice system will reveal that it is no substitute. Undoubtedly, it is worse. Accused with mental disorders languish in detention centers and correctional facilities. They often fail to receive much needed treatment, typically feel alienated and marginalized, and generally have a difficult time regaining normal functioning once entwined in the system. The regression in treatment and rehabilitation, along with an overreliance on warehousing the mentally ill in penal settings, is a deplorable anachronism that harkens back to before the development of asylum care. Ironically, we presently find ourselves where we were situated 200 years ago trying to get the mentally ill out of the prisons and criminal justice system. We have come full circle.

Mental Health Courts: A Response

Mental health courts are a response to this reality. Recognizing the criminalization process that has occurred, these courts have sought to reverse the misplaced responsibility for the provision of mental healthcare services. These courts have various objectives, which include the targeting of accused with mental disorders, and have as their mandate one or more of the following objectives:

1. "diversion" of accused who have been charged with minor to moderately serious criminal offenses and offering them an alternative;
2. expediting the pretrial processes of assessing fitness to stand trial;
3. treatment of operative mental disorders; and
4. a slowing of the so-called revolving door.

Through successful participation in a treatment program overseen by a mental health court team, some accused can avoid conviction and sentence, as they are "diverted" back into the civil mental healthcare system. This model is based upon the earlier established drug courts. The court "intervenes" at a discrete point in time along the criminal justice continuum. Some jurisdictions have included diversion as a component of their mental health court but have attempted to also intervene at multiple postarrest junctures and to a much larger population of mentally disordered accused.

Mental Health Courts – A Broadly Defined Category

Despite sharing similar objectives, there are many models that claim the label of a "mental health court". Accordingly, when considering these courts it is important to gain an appreciation for the scope of what is being referred to. Different courts have different (or multiple) entry points in the criminal prosecution process, different entrance requirements, objectives, and outcomes. Nevertheless, very generally speaking, mental health courts are all attempting a rehabilitative response to what would otherwise have been criminally sanctioned behavior. The general philosophy driving this approach is quite simple – the traditional response to aberrant behavior where it is substantially the product of mental disorder is both ineffective and inappropriate.

Therapeutic Jurisprudence

In beginning to appreciate mental health courts, it is necessary to understand the principles of therapeutic jurisprudence (*see* **Therapeutic Jurisprudence**). It

is a theory which holds that the law should be administered and applied in a way that incorporates therapeutic goals. It advocates use the justice system in a manner which addresses the underlying factors that may lead an individual to come into contact with the law. It is a vehicle to obtaining a better societal response to proscribed behaviors.

This novel approach to criminal justice has particular relevance to mentally disordered accused in today's society.

We know that many mentally disordered accused often "end up" in the criminal justice system. Often they are charged with minor and nonviolent offenses. Many times the factors that place these accused in the justice system could more effectively be addressed in ways other than imposing traditional criminal justice sanctions. It is this normative foundation that explains why mental health courts often focus on treatment of mental health systems, housing, substance addictions, job training, and other matters in preference to such traditional options as jail, fines, and probation.

Restorative Justice

The creation of mental health courts also resonates with the principles and ideological foundation of the restorative justice movement. Restorative justice is a view of justice that centers on repairing the harm and relational disruption caused by criminal behavior. One of restorative justice's key tenets is to foster the involvement of all stakeholders, including the community, to restore victims and offenders to intact, contributing members of society. Mental health courts play a key role in this restorative process by rerouting accused to both treatment and a place in the community, and by enlisting the community to provide care and supervision for an afflicted individual deserving of both compassion and intervention. Given the frequency with which family and others close to a mentally disordered accused become his or her victims, a therapeutic outcome is more likely to reflect victims' wishes that offenders be dealt with in a compassionate, humane, and ultimately more socially protective way, notwithstanding their having been victimized by the accused.

The First Mental Health Courts

The first mental health courts began as grass root initiatives in the mid-1990s. Early versions, as mentioned above, found inspiration from the success of drug courts – an emerging brand of court dedicated to accused with substance addictions. On a very basic level, drug courts operate by offering accused an option: avoid serving a sentence for a drug-related offense by completing a drug-treatment program.

As judicial initiatives, drug courts were quickly lauded as helping to break addiction cycles. By adopting principles of therapeutic jurisprudence and focusing upon addiction as opposed to crime, these courts were assisting the participants avoid subsequent drug offenses. The success of the treatment model espoused by drug courts spawned a parallel idea for accused with mental disorders.

Like drug courts, the first mental health courts in the United States found promise in offering treatment instead of punishment. Their ability to effectively remove individuals from the criminal justice system through the provision of services demonstrated an important reality: for many accused, minor offenses may more accurately represent an inability to control or manage their mental health symptoms as opposed to deliberate criminality.

The increasing appeal of specialty courts and a growing awareness that the traditional criminal justice system was failing individuals with mental disorders have combined to legitimize the emergence of mental health courts. And come they have. The onset has been remarkable; despite only appearing within the past decade or so, they now number in excess of 200 within North America alone. Beyond the courts that currently exist, numerous jurisdictions are actively pursuing implementation.

The first known program in Canada that addressed the issue of mentally disordered accused in the criminal justice system appears to have been in Toronto, Ontario, Canada. The "Diversion of Mentally Disordered Accused" became a program that was officially part of the Crown Policy Manual in 1994. This program is most like the programs referred to as "mental health courts" in the United States. In addition to this program, a mental health court has been created to deal with a broader range of issues. Diversion is just one option in the court's range of options and therapeutic armamentarium. The mental health court operating in Toronto houses a diversion program but intervenes at various other postarrest junctures to assist all mentally disordered accused, regardless of the offense charged, whether or not "diversion" will ultimately become an option.

The Operation of a Mental Health Court

Understanding the operation of a mental health court requires an appreciation of both the mental health court team and the eligibility of accused for program participation.

While the "nuts and bolts" of mental health courts will vary, integral to the functioning of a mental health court is a multidisciplinary team approach. Judges and lawyers are supplemented by any number of psychiatrists, psychologists, case workers, and social workers who collaborate on how the particular needs of the accused can effectively be met. The various disciplines represented in these teams form a concentrated resource which facilitates the court's operation.

Each court maintains a formal or informal policy outlining the type of accused eligible for the court. Typically, participation in the "diversion" component of a mental health court is reserved for individuals with mental disorders charged with minor to moderately serious offenses. Nevertheless, certain courts also provide services that do not involve eligibility requirements. As an example, the Toronto mental health court addresses the pretrial issues of fitness to stand trial and treatment of underlying mental disorder for all accused, regardless of the seriousness of the offense. As well, the court may order assessments to determine the issue of criminal responsibility. On a voluntary basis it will also entertain resolutions and bail hearings.

In most mental health courts in the United States, eligible and consenting accused are given a choice: participate in a treatment program and have your criminal charges stayed, dropped, or reduced, or proceed in the regular stream. The treatment program is strictly voluntary, and accused are most often able to opt out at any time. The approach in Canada has been somewhat different. The primary focus is with respect to assessing fitness to stand trial and providing treatment. The accused's participation in this aspect of the court's operation is not voluntary. Thereafter, once fit to stand trial, whether the accused elects to remain with the court for a bail hearing, participate in "diversion", or resolve the matter with a guilty plea is the accused's option.

Accused who elect to participate in the mental health court will typically be required to comply with an individually tailored treatment program designed by the mental health court team.[a] The benefit of a multidisciplinary team is that treatment can take a variety of forms and is not limited to medication, but can include psychological therapies, educational training, occupational training, housing and access to social services, budgetary counseling, etc.

As previously noted, a primary goal of mental health courts is to reconnect and reintegrate individuals in need of treatment to the appropriate services. In this way, mental health courts form bridges to various services within the community. Assisting individuals manage their mental disorders through the provision of mental health and social services reduces the likelihood of subsequent offenses, and in this way mental health courts also seek to curb the disproportionately high rate of recidivism in this segment of the population. With a reduction in recidivism, the courts also make communities safer places to live. There are now studies that support the previously intuitive projection that mental health courts do, indeed, reduce recidivism rates [2].

That the behaviors that resulted in the accused's appearance before the court were not contrived by him or her to be criminal in nature in the usual sense, and more likely reflect one or more mental disorders, is a foundational premise of a mental health court. All of the court's personnel are inevitably guided by the premise that, but for the accused's mental disorder or condition (and attendant socioeconomic decline), he or she would not likely have become involved in the conduct before the court. The corollary to this premise is that no professional or worker from either the legal or mental health arenas should become involved in a mental health court unless he or she is philosophically oriented to a therapeutic outcome, and this is so notwithstanding public protection considerations.

Mental health courts consequently do not share the hard-hitting adversarial atmosphere associated with most criminal courts; actually, mental health courts hardly feel like courts at all. An informal atmosphere aimed at putting the participants at ease and fostering an environment conducive to discussion is a feature common to all mental health courts.

General Observations and Caveats

The recent emergence of mental health courts has transformed the criminal justice landscape for mentally disordered accused. While traditionally one general criminal court served any type of offender, today

mental health courts represent one of many specialized "problem-solving" courts in existence to serve specific types of offenders with special needs, including drug courts, domestic violence courts, aboriginal courts, and prostitution courts. The speed with which mental health courts have emerged cannot be understated. A mere 10 years ago none was reported to be in existence, while today well over 200 mental health courts exist in North America alone.

It can be stated that mental health courts have emerged out of frustration and an abhorrence of the realities faced by mentally disordered accused. Lengthy delays for fitness assessments, a cast of courtroom players untrained and/or unfamiliar with the realities of mental illness, and a prison system more prone to exacerbate than to manage, only begin to enumerate the list. These courts are often held to be direct commentaries on the failure of the traditional justice system to meet the needs of mentally disordered accused.

The mental health court movement in North America appears to be proceeding on the assumption that these courts represent the answer. In the United States, a congressional commitment to the funding of an additional 100 mental health courts in 2001 evidences this belief. Currently in Canada there exists a strong impetus in many provinces to create these courts.

Yet, it is ironic that if one were to design the optimal mental healthcare delivery system it is unlikely that too many professionals would conclude that judges, courts, and the criminal justice system at large are the best vehicle. Judges and lawyers as mental health service brokers? Unfortunately, our systems, both the mental health and criminal justice system, are the product of evolution rather than design. They have evolved like any other complex organism. Unable to turn back the clock, the question remains: what of mental health courts? The unfortunate reality is that mentally disordered accused are in ever-increasing numbers dropped at the courthouse door.

It is accepted that there will continue to be a need for mental health courts in the criminal justice system – but hopefully the need will be reduced in the context of our ultimate solution. Even if overhauls of the mental healthcare system reduce the numbers of mentally disordered accused coming in to the criminal justice system as they are presently recognized, it is clear that the courts should endeavor to respond more appropriately to this population. Mental health courts may be the mechanism for that better response within the criminal justice system.

The question to ask then is how to proceed; how, and what, should we learn from the short existence of mental health courts? A distillation of the accumulating literature [3] permits a listing of the features of mental health courts which should be avoided, those which we should strive to include, and general observations regarding mental health court's role in the justice system.

Mental health courts did not emerge in a vacuum. They represent an evolutionary response to an escalating need to deal more effectively with the growing numbers of mentally ill entering the criminal justice system. Their existence is a reflection of decades of erosion of the civil mental healthcare system's ability to completely fulfill its role and a testament to the expression "necessity is the mother of invention."

Mental health courts are an example of the justice system's ability to respond to a societal problem in a nontraditional, yet more effective, manner. The traditional punishment-based response of the criminal justice system to individuals who are in need of correction has failed both society and the mentally disordered accused and is, in fact, counterproductive. As an example of therapeutic jurisprudence, mental health courts attempt to get at the underlying root causes of undesirable behavior. They are an example of therapeutic jurisprudence in action.

Mental health courts should have no "entrance ticket" other than a willingness on the part of the accused to attempt change. It is inappropriate to require an accused entering a diversion program to enter a guilty plea or to accept criminal responsibility. It is antithetical to both the goal of "decriminalization" and the notion that the accused we are diverting are being diverted because their otherwise criminal activity is more reasonably seen as the product of their mental disorder. To dangle the prospect of leniency in front of the mentally disordered accused and promise that lenient outcome only if he pleads guilty completely undermines the voluntariness of the plea.

Participation in a diversion program should be voluntary. Setting aside the philosophical debate as to what is truly voluntary, an accused should be permitted to enter or leave a diversion program at will. It is well established that motivation is the lynchpin for change. It is equally well established that it is rather pointless to waste rehabilitative time,

resources, and effort on an individual who is not a willing participant.

Noncompliance with a diversion program should not attract criminal sanctions. For the reasons noted above, there is little point in *forcing* an accused to comply with a treatment regime. Noncompliance should, first of all, be visited with further reassurance and support. Perhaps alternative therapeutic courses should be considered. Relapse is a normal, expected feature of rehabilitation. It should not attract penalties. If after all supportive options have been exhausted the accused is showing no interest in participating, they should be permitted to withdraw and return to the regular prosecutory stream. A mentally disordered accused should not be punished for attempting a therapeutic avenue, but failing.

A second reason for not including criminal sanctions as a response to noncompliance is that it is inconsistent with the principal objective of decriminalizing the mentally disordered accused.

Upon completion of the diversion program the accused should avoid a criminal conviction. The objective is to "divert" the accused out of the criminal justice system. Therefore, once society's intervention has been successfully complied with, the accused should be rewarded with the avoidance of a criminal conviction. Again, this is consistent with the primary objective of decriminalizing the mentally disordered individual.

Diversion programs should, as much as possible, divert the accused back in to the civil mental healthcare system. In order to avoid a two-tier mental healthcare system with one perceived to be superior, it is seen as preferable to rely upon one universal system. Diversion programs constitute another entry portal, but the resources should not be dependent upon outstanding criminal charges. It should not be one system for the "bad" and one system for the "good". Such a system may actually serve to increase the arrest rates of mentally disordered individuals if the police believe that the only or best way to secure treatment for the individual is to introduce him to the criminal justice system.

The duration of a diversion program should be a function of clinical improvement rather than participation for a fixed period of time. If the objective of the diversion program is to reintegrate the accused back in to society once sufficiently stable, it does not seem logical to connect completion or "success" to a time line. The duration of time spent in the diversion program should naturally vary from individual to individual.

Counsel should be mandatory for all accused entering a diversion program. Given the problematic considerations of "capacity" to consent to assessment and treatment and the general issues of voluntariness, to some extent the court can take comfort when the accused is represented by counsel. For vulnerable accused, such as mentally disordered accused, the assignment of counsel is a layer of protection and comfort that should be included as a matter of routine.

The expectations of the diversion program should be explicitly stated. Of course, if an accused is to be a voluntary participant, it is imperative that the accused understand precisely what it is that he is agreeing to participate in. A very good practice adopted by some courts is the creation of a "contract" wherein the particulars of the diversion plan are clearly articulated and agreed to. This may serve a useful purpose in refreshing all party's memories as to what was agreed. Again, the agreement can be modified, as necessary, in that most rehabilitative programs are to a large extent a matter of trial and error. In addition to assisting the mentally disordered accused, clearly defined goals and objectives can provide focus for the service providers within the court.

Any jurisdiction contemplating a diversion program or mental health court should not see that as a singular or unidimensional response to the problem of mentally disordered individuals coming in to the criminal justice system. The need for a mental health court should often be seen as symptomatic of a "sick" civil mental healthcare system. A full review and overhaul of the civil mental healthcare system will also reduce the extent to which the mental health court will be relied upon as a principal delivery vehicle. It makes little sense to divert an accused back to a civil system that was not adequate in the first place. As well, it is apparent that precharge diversion and the identification of high-risk individuals may be preferable to postarrest diversion. Accordingly, mental health courts comprise only one response within a range of solutions. Optimally, these courts should simply form one part of a larger strategy; it is inappropriate to view mental health courts as a panacea to the range of problems that have given rise to the existence of these courts.

Every diversion program or mental health court should strive to include a plan for evaluation. A confluence of factors and barriers has served to

frustrate research into the functioning and efficacy of mental health courts. To counter this reality, it is important for mental health courts to be diligent in data collection and to collect data ranging from the volume of accused coming through the system to recidivism rates. Monitoring a program in a systematic way will provide useful information as to which components are efficacious and which ones are not.

A court-based diversion plan should incorporate as many community "partners" as possible. The success of a diversion program will inevitably be, in part, determined by how robustly the community is supporting the plan. In many ways the effectiveness of a mental health court can be predicted by the strength of the essential services found in the community. By having both forensic and nonforensic programs and community agencies in the diversion scheme, mental health courts will provide for a better transfer of the accused out of the criminal justice system and back in to the community.

All professionals participating in the diversion program or mental health court should receive specialized training. Training is an important aspect of a mental health court to ensure that participants are approaching the task of diversion from the same perspective, and having the same objectives ensures that the team will function more consistently. As an example, it is not uncommon for different factions within a mental health court to view "success" differently, such that the accused will be confronted with a confusing and frustrating set of expectations. Again, consistency and effectiveness require all mental health court staff to be fully apprised and "on the same page".

Trained judges, lawyers and prosecutors, clinicians, and other court personnel, are on the forefront of a new and challenging area of service provision and study. They are often in the best position to understand the complex relationship between mental disorder and criminality, and the plight of the mentally ill in the criminal justice system. Their knowledge and experience provides them with an opportunity to play a unique educative role within their own community, or other communities looking to put a mental health court in place.

Families and friends of the accused should be included in the diversion process to the greatest extent possible. Where possible, and where appropriate, mental health courts should strive to enlist the support of friends and family so that the accused will experience a consistently supportive environment within which he may progress. Individuals close to the accused can also be recruited as "therapists" and sources of information with respect to progress as well as relapse.

Upon completion of a diversion program a continuation of treatment and support in the community must be guaranteed. It is pointless to invest in a court-based diversion program only to have all the supports put in place withdrawn upon completion. Involvement with a mental health court should not lead to a dead end but hopefully should represent a bridge to the recipient of essential services on a going-forward basis. Again, this points to the desirability of having one universal mental healthcare system that supports all mentally disordered individuals regardless of whether they are presently before the courts.

Mental health courts should intervene at multiple junctures. It is submitted that a comprehensive mental health court should assist the accused at multiple junctures, not just "diversion". This view strikes as both obvious and inevitable in light of the various mental health concerns that often arise at any point from the inception to conclusion of a criminal prosecution. Commencing at the point of arrest, a specialty court can assist the mentally disordered accused with bail hearings, the assessment of fitness, treatment orders, keep-fit orders, uncomplicated criminal responsibility matters, disposition hearings, guilty pleas, and sentencing as well as the business of diversion. While diversions are typically focused on "outcomes", the mentally disordered accused has unique concerns at many junctures in the prosecution process which can be addressed by a mental health court.

End Notes

[a.] At the same time, some jurisdictions have a fixed 'program' of a fixed duration in which all candidates enroll. It is the author's view that individually fashioned regimens adjusted to the individual's particular needs are more likely to be successful.

References

[1] Schneider, R.D. (2000). 7*A Statistical Survey of Provincial and Territorial Review Boards*, Federal Department of Justice, Ottawa.

[2] Kaplan, A. (2007). Mental health courts reduce incarceration, save money, *Psychiatric Times* **24**(8), 1–3.
[3] Schneider, R.D., Bloom, H. & Heerema, M. (2006). *Mental Health Courts: Decriminalizing the Mentally Ill*, Irwin Law, Toronto.

Related Articles

Therapeutic Jurisprudence

RICHARD D. SCHNEIDER

Mental Health Evaluations: Malingering in *see* Malingering: Forensic Evaluations

Mental Health Practitioners: Threat to *see* Violence Risk Assessment for Mental Health Professionals

Mental Health Treatment: Mandated *see* Treatment, Mandated: Mental Health

Mental Health Treatment: Right to *see* Treatment, Right to: Mental Health

Mental Health Treatment: Right to Refuse *see* Treatment, Right to Refuse: Mental Health

Mental Illness *see* Psychopathology: Terms and Trends

Mental Illness: Criminalization of *see* Criminalization of the Mentally Ill

Mental Retardation

Over the years, there have existed many definitions of mental retardation, each of which has been revised based on the new knowledge and/or the changing perceptions of scholars in the field (see Reference [1] for an overview). At present, there exist three primary definitions that have been put forth by the following organizations: (i) American Association on Intellectual and Developmental Disabilities (AAIDD, formerly the American Association on Mental Retardation (AAMR)), (ii) the American Psychiatric Association (APA), and (iii) the World Health Organization (WHO). There are similarities among the three definitions, but there are also differences that can result in different ways of portraying the disorder. In the following paragraphs, the reader is provided a glimpse into the construct of mental retardation as put forth by each organization; for a more comprehensive description, the reader is encouraged to consult the primary sources (see below for details) and current authoritative texts, such as Jacobson *et al.* [2] and Switzky and Greenspan [3].

American Association on Intellectual and Developmental Disabilities (AAIDD)

According to the AAIDD [4], mental retardation is defined as "disability characterized by significant limitations both in the intellectual functioning and in adaptive behavior as expressed in conceptual, social, and practical adaptive skills. This disability originates before the age of 18" (p. 8). The AAIDD does not identify any one measure of intelligence as necessary for diagnostic purposes, nor does it require the use of a particular measure of adaptive behavior. Instead, the organization states that the measures must be standardized using the general population and be appropriate for their intended use [1].

As noted in the manual, the following five assumptions are essential to the application of this definition.

1. Limitations in present functioning must be considered within the context of community environments typical of the individual's age peers and culture.
2. Valid assessment considers cultural and linguistic diversity as well as differences in communication, sensory, motor, and behavioral factors.
3. Within an individual, limitations often coexist with strengths.
4. An important purpose of describing limitations is to develop a profile of needed supports.
5. With appropriate personalized supports over a sustained period, the life functioning of the person with mental retardation generally will improve (p. 8).

American Psychiatric Association (APA)

Similar to the definition put forth by the AAIDD, in the Diagnostic and Statistical Manual, fourth edition, text revision (DSM-IV-TR) the APA conceptualizes mental retardation as a disability that is characterized by deficits in both the intellectual functioning and adaptive behavior that are manifest before the age of 18. Specifically, the diagnosis of mental retardation is made if all of the following criteria are met [5].

1. Significantly subaverage intellectual functioning; an intelligence quotient (IQ) of approximately 70 or below on an individually administered IQ test (for infants, a clinical judgment of significantly subaverage intellectual functioning).
2. Concurrent deficits or impairments in present adaptive functioning (i.e., the person's effectiveness in meeting the standards expected for his or her age by his or her cultural group) in at least two of the following areas: communication, self-care, home living, social/interpersonal skills, use of community resources, self-direction, functional academic skills, work, leisure, and health and safety.
3. The onset is before the age 18 ([5]; p. 49).

A feature of the diagnosis that was dropped by the AAIDD but maintained by the APA is severity of impairment. Specifically, according to the APA, mental retardation falls along a continuum of severity. The four levels of severity identified by the APA are based on the measured IQ: (i) mild, (ii) moderate, (iii) severe, and (iv) profound. The diagnosis of mental retardation, severity unspecified may be applied when there is strong evidence to support the diagnosis, but the individual cannot be tested using a standardized measure of intelligence.

International Classification of Diseases (ICD; 10th Edition)

Unlike the definitions put forth by the AAIDD and the APA, the WHO provides only a description of the disorder and not the means by which to diagnose it [6]. As noted in the International Classification of Diseases (ICD)-10, mental retardation is

> . . . a condition of arrested or incomplete development of the mind, which is especially characterized by impairment of skills manifested during the developmental period, skills which contribute to the overall level of intelligence, i.e., cognitive, language, motor, and social abilities. Retardation can occur with or without any other mental or physical condition.
>
> Degrees of mental retardation are conventionally estimated by standardized intelligence tests. These can be supplemented by scales assessing social adaptation in a given environment. These measures provide an approximate indication of the degree of mental retardation. The diagnosis will also depend on the overall assessment of intellectual functioning by a skilled diagnostician.
>
> Intellectual abilities and social adaptation may change over time, and, however, poor, may improve as a result of training and rehabilitation. Diagnosis

should be based on the current levels of functioning (p. 369–370).

Similar to APA, ICD-10 describes mental retardation as a condition that falls along a continuum from mild to profound mental retardation based on the IQ score. Again, the ICD-10 does not provide specific information regarding how to diagnose the condition but instead provides a method by which mental retardation can be classified and traced via medical databases.

It is important to note that some experts in the field of mental retardation disagree with the way it is currently defined. For example, according to Greenspan and Switzky [7], "Mental retardation (MR) is an invented bureaucratic category, currently undergoing radical rethinking and likely renaming, that includes many biological-based brain disorders, but is itself determined through functional criteria (e.g., IQ below a certain level) that are purely arbitrary" (p. 19). Given the ever-changing models and definitions of intellectual disabilities, it is critical that individuals interested in mental retardation stay abreast of advancements in the field.

Etiology

Like many disorders, mental retardation does not have a single identifiable etiological factor. As described in the 2002 AAMR manual, the etiology of mental retardation is conceptualized as multifactorial and is composed of four broad categories of risk factors: (i) biomedical, (ii) social, (iii) behavioral, and (iv) educational. These four factors do not function in isolation, but instead are thought to interact across the life of the individual and across generations [4]. In fact, in each category there exist many distinct factors that can cause deficits in intellectual functioning. For instance, chromosomal disorders (e.g., Down syndrome), maternal illness, birth injury, and traumatic brain injury are all factors found in the biomedical domain. In the social domain are risk factors such as poverty, domestic violence, lack of adequate stimulation, and institutionalization. Maternal drug and alcohol use, parental smoking, parental rejection of the caretaking role, and social deprivation are examples of behavioral risk factors. Finally, parental intellectual ability, lack of preparation for parenthood, inadequate education services, and inadequate family support have been identified as educational risk factors.

In their review of epidemiological studies on the etiology of mental retardation, McLaren and Bryson [8] found that as many as 50% of the population of individuals diagnosed with mental retardation had more than one causal risk factor. It is important to note that the presence of any one risk factor, or perhaps even a combination of many risk factors, does not mean that an individual has mental retardation. The diagnosis of mental retardation requires observable impairments in functional abilities, regardless of etiology, and in about half of the cases of mental retardation, the etiology is not known [9]. It is important to be aware that the absence of an identifiable risk factor or etiology does not mean that an individual does not have mental retardation.

Assessment of Mental Retardation

The Intelligence Quotient. As is clear from the definitions of mental retardation, an individual's level of intellectual functioning must be established before a diagnosis can be made. To do this, an individual must be administered a standardized measure of intellectual functioning that provides an IQ. Although there are many tests of intelligence, the most commonly used, individually administered, and well-standardized tests are the *Wechsler Adult Intelligence Scale – Fourth Edition* [10] the *Wechsler Intelligence Scale for Children – Fourth Edition* [11], and the *Stanford-Binet Intelligence Scales – Fifth Edition* (SB5) [1]. For diagnostic purposes, the most important index of functional ability is the overall, or Full Scale, IQ.

The AAIDD and APA both require an IQ of approximately 70 or below for a diagnosis of mental retardation. A score of 70 places an individual approximately two standard deviations below the mean of the general population at the time the test was developed.[a] When accounting for standard error of measurement, this cutoff score may be as high as 75.

Adaptive Functioning. As previously noted, in addition to the IQ criterion, a diagnosis of mental retardation requires concomitant deficits in adaptive behavior. In brief, adaptive behaviors are everyday skills, such as walking, talking, grooming, cooking, cleaning, and participating in school or work. These abilities are learned over time in the context of one's home and community, and they represent

skills that are necessary to function within that context. It is important to emphasize that adaptive behaviors develop over the course of time and with experience, and thus individuals are evaluated against their same age peers. To measure adaptive skills, adaptive behavior scales have been developed and normed on individuals with and without intellectual disabilities. These scales require that an informant, typically a parent, teacher, or other individual who is very familiar with the individual's daily level of functioning, rate the person of interest on a variety of skills. For instance, the informant may rate the extent to which the individual follows directions or balances a checkbook along a continuum ranging from "Never Does or Can't Do" to "Always Does or Can Do Without Assistance".

For the diagnosis of mental retardation, the AAIDD definition requires that "significant limitations in adaptive behavior should be established through the use of standardized test measures normed on the general population, including people with disabilities and people without disabilities. On these standardized measures, significant limitations in adaptive behavior are operationally defined as performance that is at least two standard deviations below the mean on either (a) one of the following three types of adaptive behavior: conceptual, social, or practical, or (b) an overall score on a standardized measure of conceptual, social, and practical skills" ([4], p. 14). Like the AAIDD, the APA [5] recommends the use of standardized measures in the assessment of adaptive behavior. That said, it is well recognized that in some instances quantitative analysis of functional abilities is either inappropriate or does not provide enough information to make a determination. In addition to standardized, quantitative measures, both the AAIDD and the APA state that the final determination of mental retardation requires the use of multiple sources of information and clinical judgment.

There exist numerous scales of adaptive behavior that can be used for the purposes of diagnosis, classification, and planning for supports; no single measure is best for all the three. Examples of adaptive functioning scales commonly used in the assessment of mental retardation include the Scales of Independent Behavior – Revised Full Scale (SIB-R) [12], the *Adaptive Behavior Assessment System – Second Edition* (*ABAS-II*) [13], and the *Vineland Adaptive Behavior Scales – Second Edition* (*VABS-II*) [14]. It is important to note that, regardless of the measure chosen, the utility of the information obtained is limited by the quality of the information provided by the rater [4, 15].

Mental Retardation and Everyday Life

As previously mentioned, mental retardation is characterized by problems in everyday life or adaptive functioning and the extent of these difficulties varies with the severity of retardation. Difficulties in adaptive functioning may be evident in social, conceptual, and/or practical areas [4]. There exist many misconceptions about how to conceptualize adaptive behavior for diagnostic purposes. First, some believe that adaptive behavior is measured by estimates of abilities or potential, but it is the individual's actual performance that is important. Second, adaptive behavior is typical behavior that reflects an individual's ability to function on a day-to-day basis. It is not measured by isolated successes or failures. Third, adaptive behavior is performance in one's community, not in restricted settings, such as prison or therapeutic treatment programs. Fourth, the definition of mental retardation does not require that a cause of impairment be identified; diagnosis is determined by the presence of significant deficits in intellectual functioning and concomitant deficits in adaptive behavior that are evident before the age of 18.

As previously noted, there are hundreds of identified causes of mental retardation, but the cause for any individual is often not known. This is particularly true for individuals with IQs in the 60s who usually do not have a medical syndrome that is known and do not have an obviously different appearance or manner. For instance, people with mild mental retardation have significant limitations in achieving independence, but they have fewer difficulties and require less support or assistance than people with more severe levels of retardation.

With regard to support, it is typical for individuals with mild mental retardation to receive informal help from family members or neighbors, rather than from provider agencies, and it is quite likely that they have, or have had, gainful employment in jobs that require limited decision making or judgment. In most instances, individuals with mild mental retardation require support from others to manage money and make major life decisions. Few people

in this category require a legal guardian, and many marry and have meaningful social relationships. In fact, it is common for these individuals to be active participants in their communities, and as a result, they are vulnerable to exploitation by others. Further, it is more common for people with mild or perhaps moderate mental retardation to be engaged in criminal activity, than it is for those with more severe levels of impairment.

As the degree of mental retardation becomes more severe, the need for formal supports from an agency, such as a local mental health center or other provider, becomes greater. People with moderate or severe mental retardation are more likely to require supervision for most, if not all, of their day. As a result of this greater supervision and less personal freedom, they are less likely to be exploited in the community or to become involved in criminal activity. Nevertheless, these individuals are subject to exploitation, often by care providers. For instance, men may be exploited for their labor, or women may be exploited for sexual favors.

In contrast to individuals with mild mental retardation, individuals with more severe degrees of impairment are likely to have an identifiable medical or environmental cause for their retardation (e.g., Down syndrome and/or fetal alcohol syndrome) and to need specialized medical care. These individuals are also more likely to have serious behavior problems that require professional services. Problems of aggression, self-injury, self-stimulation, property destruction, and other disruptive behavior can limit individuals' opportunities for integrated community living. These problem behaviors are sometimes associated with mental illnesses that coexist with mental retardation [16].

Services and Supports

The services and supports available today are more varied and individualized than in the past; the following are just a few examples. The federal Individuals with Disabilities Education Improvement Act (IDEIA) of 2004, Public Law 108–446, mandates individualized services from birth to the age of 22. The Early Intervention Program for Infants and Toddlers with Disabilities (Part C) under IDEIA serves children from birth to the age of two. The IDEIA also mandates services and describes procedures for services during the school years through age 22. Students are provided an Individualized Education Program (IEP) and a transition plan to smooth the path from school to community living and work. Children and adults in the United States identified with mental retardation are eligible for Medicaid [17], which provides medical and a wide variety of other lifetime services.

In adulthood, many people with mild mental retardation are not identified by any service agency and, thus, receive only informal supports from family, neighbors, and employers. This is not surprising as these individuals show fewer deficits and demonstrate a higher level of skill in some of the more obvious areas of functioning (e.g., work, leisure, and social relationships). It is important to note that the majority of individuals with mental retardation, at all levels of functioning, reside with their parents or other family members throughout their lives. Residential care is typically reserved for more impaired individuals.

The most significant trend in services in recent years has been away from uniform programs that are designed to serve large numbers of people with disabilities and toward individualized services that reflect the desires and abilities of each person. People who receive formal services from a public or private provider agency typically participate in a person-centered planning process. The result is a written, person-centered plan (PCP) that describes the services and supports best suited to that individual. The plan typically includes residential, work, and personal living considerations for the individual, and it is reviewed and revised regularly. PCPs are characterized by emphasis on living in the most integrated setting that is appropriate to the person and honoring the choices of the person with regard to residence, work, social supports, health care, community participation, recreation, and other personal preferences.

Personal planning commonly occurs in the area of residential care. Over the past few decades, the emphasis has moved from "housing" individuals with mental retardation in large institutions, toward providing them with opportunities for more independent living with individualized services in smaller settings [18]. In keeping with one's PCP, the most appropriate residential option is chosen from a range that includes small group homes, supported apartments, semi-independent living, and home ownership. All of these models of residential services require trained staff members who provide individualized supports.

A significant challenge to fulfilling the ideal of individualized residential living is the need for direct support personnel who can provide varied individualized services, often around the clock.

With regard to employment, state vocational rehabilitation agencies assist people in obtaining competitive or supported employment. These services begin in adolescence with planning for transition from school to work. People with mild mental retardation typically have basic functional academic skills and do not have physical disabilities that would limit their work opportunities. These individuals often thrive in jobs that do not require abstract thinking or decision making. They may, however, require more supervision than the average worker, and they may require support in managing their money.

Individuals with moderate or severe mental retardation are often good candidates for supported employment. In this model, the worker has assistance in learning and in performing the job for as long as needed. In some instances, the assistant or job coach can be faded out as the worker becomes more independent. In other instances, the job coach may be needed permanently. As the severity of mental retardation increases, the individual is more likely to have significant motor limitations (e.g., cerebral palsy), significant health problems (e.g., seizure disorder), or significant communication problems that interfere with work and require more assistance and/or shorter work hours. Regardless of severity of disability, the trend in employment services in recent years has been toward competitive or supported employment in real community jobs and away from sheltered workshops and other work in isolated and separate settings.

The use of community services provides a final example of how individual planning or person-centered planning can be implemented for the betterment of the individual. Services and supports in a PCP may come from "natural supports", such as family members or neighbors, or from formal provider agencies. Medicaid largely dictates the rules for services for such agencies in the United States, and the sources of funding are typically a mix of Medicaid, state, and local government, private payers, and charity. The PCP is intended to support the greatest degree of personal independence that the person wishes and can achieve. This goal may be difficult to achieve, because many adults with mental retardation are socially isolated. Their natural support network is limited to their families and paid service providers. To expand the individual's social network, the PCP describes strategies to increase the individual's participation in meaningful activities each day. Adults with mental retardation have typically led very sedentary lifestyles, and the result has been prevalent obesity and associated health problems. The PCP is designed to prevent such problems through a healthy and active lifestyle that includes access to health care and emphasis on personal choice.

In summary, the construct of mental retardation is ever evolving. At the time of writing, the 11th edition of the AAIDD manual is underway and is likely to recommend changes in methods of assessment, methods of prevention, and the identification and implementation of needed supports. Though there is a new DSM under construction, it is unlikely that the two agencies will come together to create one conceptualization of mental retardation. Historically, the DSM has followed in the footsteps of AAIDD, and based on the timeline,[b] it is likely that the fifth version of the DSM will closely mirror AAIDD's 2002 conceptualization.

End Notes

[a.] In recent years it has been shown that IQ scores for the general population, as measured by the Wechsler scales, have been increasing by 0.3 points per year. This phenomenon has been termed the *Flynn effect* and has resulted in much debate regarding the accuracy of IQ scores over time and how best to interpret scores obtained on tests with outdated norms. The interested reader may want to refer to the following publications for a more detailed description of the phenomenon and the controversy surrounding it [19].

[b.] DSM-V is expected to be commercially available in 2011.

References

[1] Roid, G.H. (2003). *Stanford-Binet Intelligence Scales*, 5th Edition, Riverside, Itaska, IL.

[2] Jacobson, J.W., Mulick, J.A. & Rojahn, J. (eds) (2007). *Handbook of Intellectual and Developmental Disabilities*, Springer, New York.

[3] Switzky, H. & Greenspan, S. (eds) (2006). *What is Mental Retardation? Ideas for an Evolving Disability in the 21st Century*, American Association on Mental Retardation, Washington, DC.

[4] American Association on Mental Retardation (2002). *Mental Retardation: Definition, Classification, and Systems of Supports*, 10th Edition, American Association on Mental Retardation, Washington, DC.
[5] American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition, text revision, American Psychiatric Association, Washington, DC.
[6] World Health Organization (1993). *International Statistical Classification of Diseases and Related Health Problems*, 10th Edition, World Health Organization, Geneva.
[7] Greenspan, S. & Switzky, H. (2003). Execution exemption should be based on actual vulnerability not disability label, *Ethics and Behavior* **13**, 19–26.
[8] McLaren, J. & Bryson, S.E. (1987). Review of recent epidemiological studies of mental retardation: prevalence, associated disorders, and etiology, *American Journal on Mental Retardation* **92**, 243–254.
[9] McDermott, S., Durkin, M.S., Schupf, N. & Stein, Z.A. (2007). Epidemiology and etiology of mental retardation, in *Handbook of Intellectual and Developmental Disabilities*, J.W. Jacobson, J.A. Mulick & J. Rojahn, eds, Springer, New York, pp. 3–40.
[10] Wechsler, D. (2008). *Wechsler Adult Intelligence Scale*, 4th Edition, Psychological Corporation, Harcourt Brace, San Antonio, TX.
[11] Wechsler, D. (2003). *Wechsler Intelligence Scale for Children*, 4th Edition, Psychological Corporation, Harcourt Brace, San Antonio, TX.
[12] Bruininks, R., Woodcock, R.W., Weatherman, R.F. & Hill, B.K. (1996). *SIB-R: Scales of Independent Behavior – Revised*, Riverside, Itaska, IL.
[13] Harrison, P.L. & Oakland, T. (2003). *ABAS-II: Adaptive Behavior Assessment System*, 2nd Edition, Psychological Corporation, San Antonio, TX.
[14] Sparrow, S.S., Balla, D.A. & Cicchetti, D.A. (2005). *Vineland Adaptive Behavior Scales*, 2nd Edition, Pearson Assessments, Minneapolis.
[15] Beail, N. (2003). Utility of the Vineland Adaptive Behavior Scales in diagnosis and research with adults who have mental retardation, *Mental Retardation* **41**, 286–289.
[16] Fletcher, R., Loschen, E., Stavrakaki, C. & First, M. (eds) (2007). *Diagnostic Manual – Intellectual Disability (DM-ID): A Textbook of Diagnosis of Mental Disorders in Persons with Intellectual Disability*, NADD Press, Kingston, NY.
[17] National Research Council (2002). Mental retardation: determining eligibility for social security benefits. Committee on Disability Determination for Mental Retardation, in *Division of Behavioral and Social Sciences and Education*, D.J. Reschly, T.G. Myers & C.R. Hartel, eds, National Academy Press, Washington, DC.
[18] Prouty, R., Lakin, C. & Coucouvanis, K. (2007). In 2006, fewer than 30% of persons receiving out-of-home residential supports lived in homes of more than six residents, *Intellectual and Developmental Disabilities* **45**, 289–292.
[19] Kanaya, T., Scullin, M.H. & Ceci, S.J. (2003). The Flynn Effect and U.S. policies: the impact of rising IQ scores on American society, *The American Psychologist* **58**, 778–790.

KAREN L. SALEKIN AND J. GREGORY OLLEY

Mental Retardation: Death Penalty

In 2002, the US Supreme Court held in a 6–3 decision in the landmark case of *Atkins v. Virginia* [1] that the execution of offenders with mental retardation violates the Eighth Amendment's prohibition against cruel and unusual punishments. The Court reversed its 5–4 decision 13 years earlier in *Penry v. Lynaugh* [2], which held that the Eighth Amendment's ban on cruel and unusual punishments did not categorically prohibit the execution of offenders with mental retardation.

The Court's holding in *Atkins* was based on several stems of reasoning, including public policy justifications in support of capital punishment and the extent to which they apply to individuals who have mental retardation [3]. The Court found that a "national consensus" had developed against the execution of individuals with mental retardation based on the fact that (i) many state legislatures had previously voted "overwhelmingly" in favor of the prohibition, (ii) there was a "complete absence" of states that had reinstated the death penalty for offenders with mental retardation, and (iii) the practice of executing offenders with mental retardation was uncommon, even in those states that allowed such executions.

In the 13 years between the *Penry* and *Atkins* decisions, only five states (i.e., Alabama, Louisiana, South Carolina, Texas, and Virginia) had carried out executions of offenders who were suspected to have mental retardation [1, 4], which was documented to be as many as 44 individuals [5]. The Court remarked that it was not the number of states that prohibited the practice that was significant, but it was the "consistency of the direction of change" that the Court found to be "powerful evidence that today our society views mentally retarded offenders as categorically less culpable than the average criminal". The Court also

considered opposition to such executions expressed by religious and professional organizations, disapproval by the world community, and polling data that showed a "widespread consensus" among Americans who were against such executions.

Citing the scientific literature extant at the time [6, 7], the Court noted that although those with mental retardation may be competent to stand trial and know the wrongfulness of their criminal acts, because of their impairments, "they have diminished capacities to understand and process information, to communicate, to abstract from mistakes and learn from experience, to engage in logical reasoning, to control impulses, and to understand reactions of others". Drawing on other research findings in the field [8, 9], the Court also noted that individuals with mental retardation "often act on impulse rather than pursuant to a premeditated plan, and that in group settings they are followers rather than leaders". The Court concluded that the deficiencies of those with mental retardation "do not warrant an exemption from criminal sanctions, but do diminish their personal culpability".

The Court also considered how the underlying justifications of the death penalty – retribution and deterrence – apply to offenders with mental retardation [3]. The Court reasoned that "the lesser culpability of the mentally retarded offender surely does not merit that form of retribution". Similarly, the Court recognized that because of their cognitive and behavioral impairments that render them less morally culpable, those with mental retardation are also less likely able to "process the information of the possibility of execution as a penalty and, as a result, control their conduct based upon that information." As such, the Court further noted, "Nor will exempting the mentally retarded from execution lessen the death penalty's deterrent effect with respect to offenders who are not mentally retarded."

Finally, the Court expressed concern that as a group, defendants with mental retardation "face a special risk of wrongful execution," not only because they may confess to crimes they did not commit [9] but also because of their lesser ability to provide defense counsel with meaningful assistance, their tendency to serve as poor witnesses, and their demeanor, which may create an erroneous impression that they lack remorse for their crimes.

Following the *Atkins* decision, courts in death penalty jurisdictions have begun to call upon mental health professionals to evaluate offenders in capital murder cases to assist the trier of fact in making a determination of mental retardation. Mental health professionals may be requested to render an expert opinion about mental retardation in so-called "*Atkins* hearings" at different stages of legal proceedings, which may occur pretrial, at the sentencing phase, or during postconviction appeals. However, the assessment of mental retardation in death penalty cases presents many challenges for forensic experts. In addition, the results of informal surveys of psychologists' professional practices in *Atkins* cases suggest great variability in the assessment methods used to diagnose mental retardation [10, 11].

A comprehensive review of the assessment of mental retardation in death penalty cases is beyond the scope of this chapter. The focus here is to (i) provide an overview of death penalty statutes on mental retardation in the United States, (ii) describe the contemporary definitions of mental retardation, (iii) discuss the diagnosis of mental retardation in death penalty cases, and (iv) review the available research on malingered mental retardation.

Death Penalty Statutes on Mental Retardation

Currently, there are 37 states, in addition to the federal government and the US military, which have the death penalty [12] (*see also* **Death Penalty and Age**). The 13 states that do not have the death penalty are Alaska, Hawaii, Iowa, Maine, Massachusetts, Michigan, Minnesota, New York, North Dakota, Rhode Island, Vermont, West Virginia, and Wisconsin. The District of Columbia also does not have the death penalty. At the time of *Penry*, only two states, Georgia and Maryland, had enacted laws that prohibited the execution of individuals with mental retardation. In the 13 years between *Penry* and *Atkins*, 16 other states enacted statutes barring such executions [1, 12]. The 18 states that had statutes that barred execution of those with mental retardation at the time of *Atkins* were Arizona, Arkansas, Colorado, Connecticut, Florida, Georgia, Indiana, Kansas, Kentucky, Maryland, Missouri, Nebraska, New Mexico, New York (except for murder by a prisoner), North Carolina, South Dakota, Tennessee, and Washington [12]. In 1988, when the Congress enacted legislation reinstating the federal death penalty, it excluded the execution of those with mental retardation [3].

After the *Atkins* ruling, eight more death penalty states revised their statutes to provide definitions of mental retardation. The states are California, Delaware, Idaho, Illinois, Louisiana, Nevada, Utah, and Virginia [12]. Because the *Atkins* Court provided little guidance to the states on how to develop rules to determine which offenders have mental retardation, there is significant variability across these pre- and post-*Atkins* state statutes. Twelve death penalty states apparently have yet to develop statutes for determining mental retardation in *Atkins* cases: Alabama, Mississippi, Montana, New Hampshire, New Jersey, Ohio, Oregon, Oklahoma, Pennsylvania, South Carolina, Texas, and Wyoming. Although some of these states have statutes that define mental retardation for the purpose of civil commitment and/or guardianship, it is unclear how these statutes may apply in *Atkins* cases. Other states, such as Mississippi, have adopted the *Atkins* decision in case law [13].

Definitions of Mental Retardation

Mental retardation has been defined by professional organizations in the field (*see also* **Mental Retardation**). The American Association on Intellectual and Developmental Disabilities (AAIDD) (formerly the American Association on Mental Retardation (AAMR)) has published four major versions of its definition since 1961. The 2002 version is the most recent. The American Psychiatric Association (APA) has also published four major versions of its *Diagnostic and Statistical Manual of Mental Disorders*. The 2000 version (*DSM-IV-TR*) [14] is the most recent.

Many state legislatures enacted statutes based on the 1983 American Association on Mental Deficiency (former name for AAMR) definition of mental retardation [3]: "Mental retardation refers to significantly subaverage general intellectual functioning existing concurrently with deficits in adaptive behavior and manifested during the developmental period" [15]. In 1992, the AAMR revised its definition with an emphasis on refining the adaptive functioning component of the previous version:

> Mental retardation refers to substantial limitations in present functioning. It is characterized by significantly subaverage intellectual functioning, existing concurrently with related limitations in two or more of the following applicable adaptive skill areas: communication, self-care, home living, social skills, community use, self-direction, health and safety, functional academics, leisure, and work. Mental retardation manifests before age 18 [16].

The 1992 AAMR definition was one of the definitions cited in *Atkins* and was adopted by several state legislatures in the 1990s [3]. However, it has been criticized for lacking empirical research support and theoretical grounding [17]. The APA's current definition in the *DSM-IV-TR* [14] contains language similar to the 1992 AAMR definition and also was one of the definitions cited in *Atkins*. The AAMR revised its definition in 2002 [18] and this version also focuses on describing adaptive functioning: "Mental retardation is a disability characterized by significant limitations both in intellectual functioning and in adaptive behavior as expressed in conceptual, social, and practical adaptive skills. This disability originates before age 18." Ellis [3] has suggested that the 2002 AAMR definition is most appropriate, because it contains the three essential components of all definitions cited in the *Atkins* decision. The 2002 AAMR definition also has been described as being more consistent with contemporary thinking and research on the assessment of adaptive behavior [19].

Diagnosis of Mental Retardation in Death Penalty Cases

Regardless of which definition is used to diagnose mental retardation, it is essential that the assessment methods used are consistent with the standards of professional practice [3, 20], particularly in capital murder cases, because there is no other assessment in which the stakes are higher. Because the population of individuals with mental retardation consists mostly of those who function in the mild range of impairment [14] and their impairments are often not immediately observable, accurate diagnosis for this subpopulation is difficult [19]. Some commentators [21, 22] have suggested that misdiagnosis may stem from a lack of understanding of the definition of mental retardation and failure to properly assess each diagnostic criterion.

Despite differences across the definitions of mental retardation, all have three common clinical components: (i) significant deficits in intellectual functioning; (ii) related deficits in adaptive functioning; and (iii) manifestation of deficits during the developmental period.

Intellectual Functioning

Current definitions of mental retardation require that an individual demonstrate significant limitations in intellectual functioning as measured by standardized psychometric instruments [14, 18]. This is operationally defined in the *DSM-IV-TR* [14] definition as an intelligence quotient (IQ) score of ~70 or below, or two standard deviations below the statistical mean. Because of the measurement error associated with intelligence test scores (±5 points), it is possible to diagnose mental retardation based on an IQ score of 75 or below [14].

There are three individually administered intelligence tests that are generally accepted measures of mental retardation for adults. They are the *Wechsler Adult Intelligence Scale – Fourth Edition* (*WAIS-IV*) [23]; the *Stanford-Binet Intelligence Scale – Fifth Edition* (*SB-5*) [24]; and the *Kaufman Adolescent and Adult Intelligence Test* (*KAAIT*) [25]. Group-administered intelligence tests, such as the *Revised Beta* [26], have been widely used as screening tests in correctional facilities [21, 27]. However, only global measures of intelligence are appropriate for diagnosing mental retardation [19] and are considered by many practitioners as the "gold standard" in assessments of mental retardation in death penalty cases [10].

The interpretation of IQ scores depends heavily upon the examiner's clinical judgment. Factors related to the examinee's behavior during the testing process, such as fatigue, effort, motivation, and attempts to malinger intellectual deficits can threaten the validity of scores. In addition, practice effects can be caused by repeated administrations of the same intelligence test in a short period. This problem may occur frequently in *Atkins* proceedings in which multiple experts administer the same intelligence test to offenders within a relatively brief timeframe. It is not uncommon to find differences in an individual's IQ scores over time. Such differences in IQ scores may occur on multiple administrations of the same test, scores on different editions of the same test, and scores on different tests [27].

Comparisons of IQ scores also can be misleading because of the "Flynn effect", which means there is a tendency for IQ scores to increase in the general population over time [28–30]. Thus, an individual's IQ may be artificially increased as a function of when the intelligence test was administered. The Flynn effect has become a topic of much debate, particularly with regard to the practice of adjusting individual IQ scores in capital cases. Some authors [31–34] have advocated modifying individual test scores to correct for the Flynn effect. Others [35] have argued that such modification of individual IQ scores is not an accepted professional practice. Harcourt Assessment, the publisher of the Wechsler tests, does not endorse the recommendation to modify *WAIS-III* scores to account for the Flynn effect [36].

Adaptive Functioning

The *adaptive behavior prong of the diagnosis of mental retardation* suggests that intellectual deficits are accompanied by real-world disabling effects on an individual's functioning [3]. Adaptive behavior has been described as the most problematic part of the definition of mental retardation [19] and may also be the least understood [37]. Some psychologists and legal professionals tend to view adaptive behavior only in terms of practical daily living skills. However, for those who function in the mild range of mental retardation, deficits are more likely to exist in the areas of social and conceptual skills [19].

Several standardized instruments have been developed to assess adaptive behavior, such as the *Vineland Adaptive Behavior Scales II* (*VABS II*) [38], *Adaptive Behavior Assessment System – Second Edition* (*ABAS-II*) [39], and the *Scales of Independent Behavior – Revised* (*SIB-R*) [40], each of which has been normed on the general population including those with intellectual disabilities. Although these instruments are commonly used in the field, they have been criticized for inadequately assessing the constructs of gullibility and naiveté, which are common characteristics of individuals with mental retardation [41]. Some experts occasionally use the *Street Survival Skills Questionnaire* (*SSSQ*) [42] in death penalty cases to assess defendants' adaptive functioning [10]. The *SSSQ* has been criticized [19] as an inappropriate measure of adaptive deficits because it is a test of knowledge, rather than performance, and it emphasizes practical skills and not conceptual or social skills.

Most adaptive behavior scales are intended to measure an individual's current functioning in the community. This creates methodological problems for assessments of adaptive functioning with incarcerated populations, particularly those who have been on

death row for many years. In such cases, the examiner must perform a retrospective assessment of adaptive functioning. However, concerns exist regarding the validity of retrospective assessments of adaptive behavior [43]. Some authors [44] have called for the development of a "penologically normed" instrument to assess adaptive functioning for incarcerated populations. To date, there is no such instrument available, and information about an inmate's prison behavior normed on other inmates would not contribute to a valid diagnosis of mental retardation. Because there is no perfect method to retrospectively assess adaptive behavior, experts must rely on a mixture of imperfect information and clinical judgment [19].

Age of Onset

The "third" component of all definitions of mental retardation requires that the disability be manifested during the developmental period [14, 18]. Many states have defined this as prior to age 18, although some states have extended the age to 22 [3, 12]. This prong of the definition distinguishes mental retardation from other disabilities that may occur later in life, such as traumatic brain injury and dementia, and it also helps to identify defendants in death penalty cases who may attempt to feign mental retardation [3, 45].

Malingered Mental Retardation

In his dissenting opinion in *Atkins* [1], Justice Scalia expressed concern about the possibility that individuals can "readily feign" mental retardation. This issue also has been raised in Mississippi case law, which has explicitly addressed the need for assessments of malingering in *Atkins* claims. In *Foster v. State* [46], the Mississippi Supreme Court required the use of the *Minnesota Multiphasic Personality Inventory-2* (*MMPI-2*) [47] to assess malingering in *Atkins* cases. However, the *MMPI-2* was designed to identify malingered mental illness, not mental retardation. It also has been criticized as an invalid measure of malingering for those with mental retardation because of its required eighth grade reading level [48]. Nevertheless, great pressure has been placed upon psychologists to assess malingering in *Atkins* cases (*see also* **Malingering: Forensic Evaluations**).

Malingered Cognitive Deficits

Few studies have investigated the validity of malingering measures for those with mental retardation. The studies that are available have produced mixed results. In a study by Hurley and Deal [49] using a sample of individuals with IQ scores between 50 and 78, subjects were administered four measures of malingering, which included one that assesses feigned psychiatric disorders, the *Structured Interview of Reported Symptoms* (*SIRS*) [50], and three measures of malingered memory: the *Test of Memory Malingering* (*TOMM*) [51], the *Rey 15-Item Memory Test* [52], and the *Rey Dot Counting Test* (*RDCT*) [53]. The findings did not support the use of three of the four measures with a population with mental retardation. The authors recommended that one (*RDCT*) undergo further evaluation for its effectiveness as a screening measure. On the other hand, a study by Simon [54] with a sample of 21 adjudicated forensic inpatients with comorbid Axis I disorders and mental retardation supported the use of the *TOMM* [51] for assessing malingering with individuals with mental retardation.

Goldberg and Miller [55] administered the *Rey-15 Item Memory Test* [52] to individuals with severe psychiatric disorders and those with mental retardation. Although the cut scores worked well for the psychiatric patients, they were ineffective for those with mental retardation. Similarly, Hayes *et al.* [56] administered three measures of malingering, including the *Rey 15 Item Memory Test* [52], the *M-Test* [57], and the *Rey Dot Counting Test* [53] to 38 subjects in a maximum-security forensic hospital who had been diagnosed with mental retardation. On the basis of the results, the researchers concluded that the battery of tests contributed nothing to the identification of malingering in defendants with mental retardation. Conversely, Schretlen and Arkowitz [58] used a combination of measures, including validity scales of the *MMPI-2* [47], two scores on the *Bender Gestalt* [59], and an experimental measure to identify individuals feigning insanity or mental retardation. The researchers found that the combination of measures accurately identified most of the subjects who feigned mental retardation.

Although several more recent studies investigating the use of effort measures with populations of persons with mental retardation have been published in the last year, these studied also have produced rather mixed results [60–62].

Malingered Adaptive Deficits

To date, only one study has investigated the susceptibility of measures of adaptive behavior to malingered responding. Doane and Salekin [63] investigated whether collateral informants could feign adaptive deficits on the *ABAS-II* [39] and the *SIB-R* [40] within the context of a death penalty case. The results indicated that both measures were susceptible to feigning by collateral informants. The *ABAS-II* was extremely vulnerable to exaggeration of deficits, and this response style was not easily detected. Subjects were not as adept at feigning mental retardation on the *SIB-R*, and the response style was easily detected.

In summary, few studies have investigated the validity of assessments of malingered mental retardation, especially malingered adaptive deficits. Similarly, most instruments designed to detect malingered cognitive deficits lack normative data for populations who have mental retardation and have not been validated for the purpose of detecting malingered mental retardation. Therefore, concerns regarding false positives exist with such instruments, because individuals with mental retardation may produce scores suggestive of malingering when they are not. Further research is needed to determine how to appropriately assess both malingered adaptive skills and feigned intellectual deficits in populations with mental retardation.

Conclusion

In light of the US Supreme Court's decision in *Atkins v. Virginia* [1], courts in death penalty jurisdictions are faced with the challenge of determining whether offenders in capital murder cases have mental retardation. Bonnie [64] observed that one of the "striking aspects" of the Court's decision in *Atkins* is that the constitutional prohibition against the execution of offenders with mental retardation is framed in the language of a "clinical diagnosis". Thus, expert opinions by mental health professionals play a critical role in the legal determination of mental retardation in *Atkins* cases, because the court's decision may depend heavily on an expert's diagnosis.

Standards that govern the admissibility of scientific evidence [65] not only require that forensic experts' methodology be generally accepted in the field, but expert opinions also must be based on knowledge, skill, experience, training, and education. However, few mental health professionals have extensive training in both forensic evaluation and mental retardation [66]. In addition, the mandates of some *Atkins* statutes defy principles of psychological assessment and can create ethical dilemmas for psychologists who serve in the role of experts in such cases [67].

In 2005, Division 33 of the American Psychological Association (Mental Retardation and Developmental Disabilities) formed an *Ad Hoc* Committee [68] to identify issues related to mental retardation and the death penalty and to clarify psychologists' role in *Atkins* proceedings. Several seminal articles have since been published that describe the controversies surrounding assessments in *Atkins* cases and provide recommendations for best practice [66, 69, 70]. Experts in the fields of forensic evaluation and mental retardation must continue working together to ensure that assessments of mental retardation in death penalty cases meet the highest standard of professional practice, as the outcome of these evaluations is literally a matter of life or death.

References

[1] Atkins v. Virginia, 536 U.S. 304 (2002).

[2] Penry v. Lynaugh, 492 U.S. 302 (1989).

[3] Ellis, J.W. (2003). Mental retardation and the death penalty: A guide to legislative issues, *Mental and Physical Disability Law Reporter* **27**, 11–24.

[4] Keyes, D.W., Edwards, W.J. & Perske, R. (1997). People with mental retardation are dying legally, *Mental Retardation* **35**, 59–63.

[5] Keyes, D.W., Edwards, W.J. & Perske, R. (2002). People with mental retardation are dying, legally: At least 44 have been executed, *Mental Retardation* **40**, 243–244.

[6] McGee, J. & Menolascino, F. (1992). The evaluation of defendants with mental retardation in the criminal justice system, in *The Criminal Justice System and Mental Retardation*, R.W. Conley, R. Luckasson & G.N. Bouthilet, eds, Paul Brookes Publications, Baltimore.

[7] Appelbaum, K.L. & Appelbaum, P.S. (1994). Criminal justice related competencies in defendants with mental retardation, *Journal of Psychiatry and Law* **14**, 483–503.

[8] Ellis, J. & Luckasson, R.A. (1985). Mentally retarded criminal defendants, *George Washington Law Review* **53**, 414–493.

[9] Everington, C. & Fulero, S.M. (1999). Competence to confess: Measuring understanding and suggestibility of defendants with mental retardation, *Mental Retardation* **37**, 212–220.

[10] Macvaugh, G.S. & Grisso, T. (2006). Assessment of mental retardation in death row inmates: A survey of

professional practices, *Paper Presented at the Conference of the American Psychology-Law Society, March 2006*, St. Petersburg.
[11] Everington, C. & Olley, J.G. (2004). An analysis of forensic psychological evaluations in capital cases involving defendants with mental retardation: Has Atkins made a difference? *Paper Presented at the Conference of the American Psychology-Law Society, March 2004*, Scottsdale.
[12] Death Penalty Information Center. Retrieved from website: www.deathpenaltyinfo.org (accessed on, 2007).
[13] Chase v. State, 873 So.2d 1029 (Miss. 2004).
[14] American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition, Text Revision. American Psychiatric Association, Washington, DC.
[15] American Association on Mental Deficiency (1983). *Classification in Mental Retardation*, 8th Edition, American Association on Mental Deficiency, Washington, DC.
[16] American Association on Mental Retardation (1992). *Mental Retardation: Definition, Classification, and Systems of Supports*, 9th Edition, American Association on Mental Retardation, Washington, DC.
[17] Greenspan, S. (1997). Dead manual walking? Why the 1992 definition needs redoing, *Education and Training in Mental Retardation and Developmental Disabilities* **32**, 179–190.
[18] American Association on Mental Retardation (2002). *Mental Retardation: Definition, Classification, and Systems of Supports*, 10th Edition, American Association on Mental Retardation, Washington, DC.
[19] Everington, C. & Olley, J.G. (2008). Implications of *Atkins v. Virginia*: Issues in defining and diagnosing mental retardation, *Journal of Forensic Psychology Practice* **8**(1), 1–23.
[20] Olvera, D.R., Dever, R.B. & Earnest, M.A. (2000). Mental retardation and sentences for murder: Comparison of two recent court cases, *Mental Retardation* **38**, 228–233.
[21] Baroff, G.S. (1991). Establishing mental retardation in capital cases: A potential matter of life and death, *Mental Retardation* **29**, 343–349.
[22] Keyes, D.W., Edwards, W.J. & Derning, T.J. (1998). Mitigating mental retardation in capital cases: Finding the "invisible" defendant, *Mental and Physical Disability Law Reporter* **22**, 529–539.
[23] Wechsler, D. (2008). *Wechsler Adult Intelligence Scale*, 4th Edition, Psychological Corporation, San Antonio.
[24] Roid, G.H. (2003). *Stanford-Binet Intelligence Scales*, 5th Edition, Riverside, Itaska.
[25] Kaufman, A.S. & Kaufman, N.L. (1993). *Kaufman Adolescent and Adult Intelligence Test*, American Guidance Service, Circle Pines.
[26] Kellogg, C.E. & Morton, N.W. (1978). *The Revised Beta Examination (Beta II)*, The Psychological Corporation, San Antonio.
[27] Baroff, G.S. (2003). Establishing mental retardation in capital cases: An update, *Mental Retardation* **41**, 198–202.
[28] Flynn, J.R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978, *Psychological Bulletin* **95**, 29–51.
[29] Flynn, J.R. (1998). WAIS-III and WISC-III IQ gains in the United States from 1972 to 1995: How to compensate for obsolete norms, *Perceptual and Motor Skills* **86**, 1231–1239.
[30] Kanaya, T., Scullin, M. & Ceci, S. (2003). The Flynn effect and U.S. policies, *American Psychologist* **58**, 778–790.
[31] Flynn, J.R. (2006). Tethering the elephant: capital cases, IQ, and the Flynn Effect, *Psychology Public Policy and Law* **12**, 170–189.
[32] Greenspan, S. (2006). Issues in the use of the "Flynn Effect" to adjust IQ scores when diagnosing MR, *Psychology in Mental Retardation and Developmental Disabilities* **31**, 3–7.
[33] Greenspan, S. (2007). Flynn-adjustment is a matter of basic fairness: Response to Roger B. Moore, Jr, *Psychology in Mental Retardation and Developmental Disabilities* **32**, 7–8.
[34] Flynn, J.R. (2007). Capital offenders and the death sentence: A scandal that must be addressed, *Psychology in Mental Retardation and Developmental Disabilities* **32**, 3–7.
[35] Moore, R.B. (2006). Modification of individual's IQ scores is not accepted professional practice, *Psychology in Mental Retardation and Developmental Disabilities* **32**, 11–12.
[36] Weiss, L. (2007). *Technical Report: Response to Flynn*, Harcourt Assessment, Retrieved from website: http://harcourtassessment.com/NR/rdonlyres/98BBF5D2-F0E8-4DF6-87E2-51D0CD6EE98C/0/WAISIII_TR_lr.pdf.
[37] Everington, C. & Keyes, D. (1999). Diagnosing mental retardation in criminal proceedings: The importance of documenting adaptive behavior, *Forensic Examiner* **8**, 31–34.
[38] Sparrow, S.S., Balla, D.A. & Cicchetti, D. (2005). *Vineland Adaptive Behavior Scales II*, Pearson Assessments, Minneapolis.
[39] Harrison, P.L. & Oakland, T. (2003). *ABAS II: Adaptive Behavior Assessment System*, 2nd Edition, The Psychological Corporation, San Antonio.
[40] Bruininks, R.H., Woodcock, R.W., Weatherman, R.F. & Hill, B.K. (1996). *SIB-R: Scales of Independent Behavior – Revised*, Riverside, Itaska.
[41] Greenspan, S. & Switzky, H.N. (2003). Execution exemption should be based on actual vulnerability, not disability label, *Ethics and Behavior* **13**, 19–26.
[42] Lindenhoker, D. & McCarron, L. (1983). *The Street Survival Skills Questionnaire*, McCarron-Dial Systems, Dallas.
[43] Brodsky, S.L. & Galloway, V.A. (2003). Ethical and professional demands for forensic mental health professionals in the post-*Atkins* era, *Ethics and Behavior* **13**, 3–9.
[44] Weiss, K.J., Haskins, B.H. & Hauser, M.J. (2004). Commentary: *Atkins* and clinical practice, *Journal of*

The American Academy of Psychiatry and the Law **32**, 309–313.
[45] Bonnie, R.J. & Gustafson, K. (2007). The challenge of implementing *Atkins v. Virginia*: How legislatures and courts can promote accurate assessments and adjudications of mental retardation in death penalty cases, *Richmond Law Review* **41**, 811–860.
[46] Foster v. State, 848 So.2d 175 (Miss. 2003).
[47] Hathaway, S. & McKinley, J.C. (1989). *The Minnesota Multiphasic Personality Inventory (MMPI-2)*, Merrill/Prentice Hall, Columbus, Ohio.
[48] Keyes, D.W. (2004). Use of the Minnesota multiphasic personality inventory (MMPI) to identify malingering mental retardation, *Mental Retardation* **42**, 151–153.
[49] Hurley, K.E. & Deal, W.P. (2006). Assessment instruments measuring malingering used with individuals who have mental retardation: Potential problems and issues, *Mental Retardation* **44**, 112–119.
[50] Rogers, R., Bagby, R.M. & Dickens, S.E. (1992). *Structured Interview of Reported Symptoms: Professional Manual*, Psychological Assessment Resources, Odessa, FL.
[51] Tombaugh, T.N. (1996). *Test of Memory Malingering (TOMM) Manual*, Multi-Health Systems, New York.
[52] Lezak, M. (1995). *Neuropsychological Assessment*, 3rd Edition, Oxford University Press, New York.
[53] Boone, K., Lu, P. & Herzberg, D. (2002). *The Dot Counting Test*, Western Psychological Services, Los Angeles.
[54] Simon, M.J. (2007). Performance of mentally retarded forensic patients on the Test of Memory Malingering, *Journal of Clinical Psychology* **63**, 339–344.
[55] Goldberg, J.O. & Miller, H.R. (1986). Performance of psychiatric inpatients and intellectually deficient individuals on a task that assesses the validity of memory complaints, *Journal of Clinical Psychology* **42**, 43–46.
[56] Hayes, J.S., Hale, D.B. & Gouvier, W.M. (1997). Do tests predict malingering in defendants with mental retardation? *Journal of Psychology* **131**, 575–576.
[57] Beaber, R.J., Marston, A., Michelli, J. & Mills, M.J. (1985). A brief test for measuring malingering in schizophrenic individuals, *American Journal of Psychiatry* **142**, 1478–1481.
[58] Schretlen, D. & Arkowitz, H. (1990). A psychological test battery to detect prison inmates who fake insanity or mental retardation, *Behavioral Sciences and The Law* **8**, 75–84.
[59] Bender, L. (1938). *A Visual Motor Gestalt Test and Its Clinical Use*, The Orthopsychiatric Association, New York.
[60] Graue, L.O., Berry, D.T.R., Clark, J.A., Sollman, M.J. Cardi, M., Hopkins, J. & Werline, D. (2007). Identification of feigned mental retardation using the new generation of malingering detection instruments: Preliminary findings, *The Clinical Neuropsychologist* **21**, 929–942.
[61] Marshall, P. & Happe, M. (2007). The performance of individuals with mental retardation on cognitive tests assessing effort and motivation, *The Clinical Neuropsychologist* **21**, 826–840.
[62] Dean, A.C., Victor, T.L., Boone, K.B. & Arnold, G. (2008). The relationship of IQ to effort test performance, *The Clinical Neuropsychologist* **22**, 705–722.
[63] Doane, B.M. & Salekin, K.L. (in press). Susceptibility of current adaptive behavior measures to feigned deficits, *Law and Human Behavior*.
[64] Bonnie, R.J. (2004). The American Psychiatric Association's resource document on mental retardation and capital sentencing: Implementing *Atkins v. Virginia*, *Journal of The American Academy of Psychiatry and the Law* **32**, 304–308.
[65] Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).
[66] Olley, J.G. (2006). The assessment of adaptive behavior in adult forensic cases: Part 2: The importance of adaptive behavior, *Psychology in Mental Retardation and Developmental Disabilities* **32**, 7–8.
[67] Duvall, J.C. & Morris, R.J. (2006). Assessing mental retardation in death penalty cases: Critical issues for psychology and psychological practice, *Professional Psychology, Research and Practice* **37**, 658–665.
[68] Olley, J.G., Greenspan, S. & Switzky, H. (2006). Division 33 *ad hoc* committee on mental retardation and the death penalty, *Psychology in Mental Retardation and Developmental Disabilities* **31**, 11–13.
[69] Olley, J.G. (2006). The assessment of adaptive behavior in adult forensic cases: Part 1, *Psychology in Mental Retardation and Developmental Disabilities* **32**, 2–4.
[70] Olley, J.G. (2007). The assessment of adaptive behavior in adult forensic cases: Part 3: Sources of adaptive behavior information, *Psychology in Mental Retardation and Developmental Disabilities* **33**, 3–6.

GILBERT S. MACVAUGH, III, KAREN L. SALEKIN AND J. GREGORY OLLEY

Mental Status: Examination

Introduction

No complex activity can be studied without good bookkeeping. Yet even the best system demands adequate judgment and individual initiative.
–Adolf Meyer "Motto" [1]

The Mental Status Examination (MSE) refers to the structured reporting of information systematically obtained by the clinician regarding an individual's mental state at the time of the interview. The MSE is usually presented as a separate section of the evaluation, distinct from the individual's biographical history or anamnesis (childhood, education, employment, relationships, and other history) and distinct from information regarding past psychiatric history, substance abuse, and history of medical illness. The MSE summarizes findings that are obtained through systematic assessment at the time of the interview. The examiner may rely upon direct observation as well as the individual's spontaneous statements, responses to inquiry regarding present symptoms and responses to specific test questions (of cognitive function, for example).

Historical Background of the Mental Status Examination

Although the format of the MSE was codified by the Swiss-American psychiatrist Adolf Meyer in 1902 [1, 2] in a privately printed pamphlet, some treatises from the nineteenth century present case reports or vignettes by describing the patient's symptomatology and findings on examination in a structured manner, using the categories that comprise the MSE as we know it today. As recently as the 1960 s, psychiatrists debated the value of the MSE, leading up to studies of the reliability of findings on different elements of the MSE. Today, in the English-speaking world, the usefulness of the MSE is accepted by consensus and it is taught as a standard part of the training of psychologists and psychiatrists.

Purposes of the Mental Status Examination

The primary purpose of performing a MSE is to aid in establishing a diagnosis of the individual's current state. This may include the determination of acute symptomatology in the context of longstanding illness as well as the assessment of the individual's risk of imminent harm to himself or others. Although the MSE is generally performed the most thoroughly by mental health professionals, the MSE is relevant in medical specialties other than psychiatry, because a patient's mental state may have bearing on the reliability of information given in the subjective account of his or her history, medical or otherwise ([3], pp. 5–6).

The aim of the MSE as recorded in a clinical treatment note or a forensic report is to communicate information to a reader. The MSE can be viewed as analogous to the physical examination, in that it provides a structured assessment of the person's present condition and a coherent, organized presentation of the data. A psychiatrist may make use of the MSE to remind herself on a future office visit of how the patient appeared the time before. A doctor on call may refer to the MSE to determine whether a patient was stable earlier in the day; he will then record his own examination along with the details of his management of the situation, so that the patient's psychiatrist will know the details of the patient's condition overnight.

In the forensic setting, a MSE conveys the evaluee's mental state at the time of the examination and thereby provides a context and support for the expert's findings in the evaluation. In an assessment of Competence to Stand Trial, for example, the recording of the defendant's present mental condition allows the expert to refer back to these elements in the writing of her opinion. In adversarial proceedings, the MSE allows the trier of fact to assess whether differences in expert opinion may be in part explained by differences in mental state at separate points in time.

Conducting the Mental Status Examination

Similarities Between the MSE in Clinical Treatment and Forensic Settings

Both types of interview require clinical skill in order to obtain the necessary information to complete an assessment. Information pertaining to the MSE is obtained in the context of a larger interview process that includes the history of past psychological symptoms, treatment, and personal history. Although clinicians view the MSE as a specific examination that takes place during a particular portion of the interview (usually after the individual's account of her history and pertinent background information) the clinician obtains data that contribute to the MSE throughout the entire face-to-face contact time with the person being evaluated. Development of rapport is important as is tact and a nonjudgmental attitude, as the interviewee may be reluctant to speak about symptoms such as hallucinations, delusions,

fears, or suicidal ideas (SI). The clinician should also be attentive to his own potential bias or his personal reactions to the individual, which may lead to avoidance of necessary exploration of the interviewee's responses to questions.

Differences Between the MSE in Clinical Treatment and Forensic Settings

Clinical treatment interviews and forensic evaluations serve different purposes, and the MSE is partly shaped by elements specific to each context. A succinct summary of the major differences between treatment interviews and forensic interviews may be found in Melton *et al.* [4] The differences most pertinent to the MSE *per se* concern the examiner's attitude and the degree of validity of the interviewee's self-report in forensic settings. These are addressed briefly below:

Confidentiality. Unlike the treatment interview, a forensic interview is likely to result in a written report to a legal entity (attorney, court, or administrative agency) responsible for deciding an issue that has required clarification by an expert. Despite the nonconfidentiality warning given at the beginning of the interview, the examiner's necessary use of clinical interviewing skills may put the evaluee at ease, leading the evaluee to have the impression that the examiner is in a helping role. This may encourage the individual to reveal information about current symptoms that he may later realize is not in his best interest. On the contrary, some evaluees are acutely aware of the nonconfidential nature of the content of the MSE and may purposefully or unintentionally misrepresent, distort or falsify their responses to the examiner's questions, or withhold information (*see* **Malingering: Forensic Evaluations**). One of the ethical challenges of forensic practice involves the task of maintaining an appropriate balance between scepticism and neutrality, on the one hand, and empathy and tact on the other.

Therapeutic Versus Medicolegal Aims. Related to the issue of malingering is the awareness on the part of both examiner and evaluee of the legal questions that form the backdrop and purpose of the interview. Unlike the treatment interview, with its emphasis on therapeutic aims and problem solving in the patient's current circumstances, the MSE in a forensic interview contributes to forming an opinion on a legal issue. Additionally, the forensic question posed to the evaluator sometimes entails an assessment of cause and effect, further changing the role of the MSE. An evaluee may believe that it is in his best interest to exaggerate symptoms reported in the MSE, because he is asserting emotional damages in a tort claim. Or an evaluee may seek to give the impression of having few if any pathological symptoms at present in order to support a claim in favor of fitness for return to work.

Categories of the Mental Status Examination

An essential feature of the MSE is its systematic nature. The clinician may observe a number of elements important in the MSE during the subjective, narrative portion of the interview, while the interviewee is giving an account of his past history and the issues leading up to the request for evaluation, either with the aim of treatment or for legal purposes. Nevertheless, the systematic organization of the MSE allows the examiner to review the major components of the individual's current mental state in a more thorough fashion, and helps to avoid the omission of questions regarding significant areas of mental functioning that would otherwise not be formally assessed.

Table 1 represents the major categories of the MSE, which are usually assessed by most clinicians in the English-speaking community. We define in broad outline the meanings of the technical terms that describe these categories, referring the reader to more comprehensive, detailed treatises which are available [3, 5].

The panoply of tests and possible questions that a clinician may use to assess elements of each of the categories described in Table 1 is wide, and each examiner has his preferences. There is no standardized set of tests and questions for performing the MSE, though, as stated above, consensus exists regarding the areas of mental functioning that should be included.

A Note on the Mini Mental State Examination

Although a number of the tests used to assess items on the cognitive portion of the MSE overlap with or

Table 1 Main categories of the mental status examination

Appearance, attitude, and behavior	This section of the report includes description of the individual's physical appearance (clothing, tattoos, or piercing), hygiene and physical movement (such as tics, visible side effects of medications, agitation, or slowness). The individual's degree of cooperation with the interview and his interaction style may also be noted.
Mood and affect	Mood refers to the person's description of her general emotional state (e.g., sad, happy, or nervous) and may be reported as a direct quote. Affect refers to the clinician's observation of the outward emotional expression, including the appearance of sadness or elation, range and variability of expression, and the appropriateness of emotions in relation to the content of the individual's speech.
Speech and language	A person's speech may be described on the basis of ease of flow, spontaneity, rate, and volume. The examiner may also comment on whether the individual is difficult to interrupt or speaks as though under pressure. Included here are comments on speech and language abnormalities, such as stuttering, slurring, or use of invented words.
Thought process and content	Thought process refers to the flow of ideas and whether the thinking is disjointed or coherent. The examiner may note whether the individual has a tendency to give answers in a circuitous manner or tends to express vaguely associated thoughts without coming back to the topic at hand. Thought content refers to the person's preoccupations, obsessions or delusions, as well as nondelusional paranoid or grandiose ideas. This section also includes the presence or absence of suicidal and homicidal ideas.
Perception	The presence or absence of perceptual abnormalities is noted here and may include hallucinations (when the individual reports a perception involving one of the five senses, in the absence of real stimuli) or illusions (when a true stimulus is misinterpreted or distorted).
Cognition	This section includes the person's level of consciousness (awake, alert, or sleepy), assessment of immediate and short-term memory, concentration and ability for abstract thinking. The individual's awareness of current events and general knowledge is also reported here.
Insight and judgment	Insight refers to the individual's understanding of his condition. The term is primarily used to describe his awareness and acceptance of having a psychiatric disorder, though it may also be used to indicate awareness of his legal situation. Judgement refers to the person's ability to use logical reasoning to respond to situations, and is often formally tested through hypothetical scenarios as well as assessed through observation of recent behavior and responses in the interview.

are identical to items on the rating scale known as the *Mini Mental State Examination* (*MMSE*), these two evaluations should not be confused. The MMSE was devised by Folstein *et al.* in 1975 [6] as a means of providing rapid and clinically feasible assessment of whether an individual may be suffering from an organic brain disorder. The MMSE is scored on a scale of a maximum of 30 points, and provides an impression as to whether further formal testing is required. It does not assess the noncognitive areas covered in the MSE. The full MMSE is not considered to be an inherent part of the MSE, though some clinicians may choose to perform it routinely in each evaluation by personal preference.

The Mental Status Examination in Medical Records

In the recording of a MSE in the medical record, the examiner usually assumes that any reader has the

same level of technical knowledge as the examiner himself. The writer often uses commonly accepted abbreviations (AH for auditory hallucinations, SI for suicidal ideas) and frequently does not write in complete sentences (Affect – constricted, inappropriate).

The level of detail included in the MSE varies according to the clinician's familiarity with the patient and according to the findings that may bear emphasis when a record is being created to communicate with other practitioners. Upon first contact in the outpatient setting, or at the time of admission, the MSE is recorded in detail. At subsequent outpatient follow-up visits or on subsequent evaluations in the hospital, an MSE may cover very briefly those portions of the examination that are unremarkable or unchanged. A change in clinical condition, or emergency coverage, on call for example, may stimulate the examiner to record greater detail. In general, the majority of clinicians comment on each of the main categories of the MSE in the record following any assessment of the patient. It is common practice to remark on the presence or absence of suicidal or homicidal ideas during the interview in making a record of each contact with the patient.

It is important to keep in mind that although mental health professionals tend to be aware that what they write in the medical record may someday appear in court, legal proceedings are not the main purpose of the medical record. An attorney should not attempt to interpret abbreviations, scrawls, and shorthand that are rife in the medical record without the assistance of a medical expert. Although this may seem obvious, the point also extends to the gathering of data, in that attorneys should not attempt to sort and select medical records, which are often obtained from multiple sources and can be voluminous, prior to forwarding the material to an expert for technical assessment. A set of documents that is as complete as possible – and as legible as possible (i.e., sufficiently dark photocopies) – is indispensable to the expert in formulating an opinion.

The Mental Status Examination in Forensic Reports

The recording of the MSE in the context of legal proceedings follows guidelines for all forensic reports: technical terms must be "translated" into language comprehensible to a lay reader remaining all the while coherent for assessment by another forensic examiner who may be called upon to comment on the findings. Because the MSE represents one of the most highly technical parts of the forensic report, the task of translation into lay language poses particular challenges. Descriptions should be sufficiently detailed so as to evoke for the reader a mental impression of the state of the individual at the time of evaluation. In the case of positive (abnormal) findings or findings that differ from prior descriptions of the person, examples are useful.

An MSE should be performed at each meeting with the evaluee, though repetition of the cognitive examination is usually unnecessary unless there are intervening factors between interviews (e.g., changes in medication, or an observed difference in performance in the interview). This MSE should be recorded for each interview date, in order to convey the consistency or lack thereof in the individual's mental state overtime. If the evaluee is seen on three occasions, the examiner should comment on the MSE on each of the three dates, briefly if there is no significant change, and at greater length if there are major differences.

The narrative description of the MSE in a forensic report should follow the same outline, including the same categories as that of the clinical MSE. This ensures that another clinical examiner, hired by the opposing side or brought into the case at a later point by the same attorney, is able to follow a predictable sequence familiar to all clinicians.

The writing of the MSE is also an occasion for the examiner to identify the evidence that will be included to support the determination of a diagnosis and opinion. The MSE is a chance for the examiner to systematically assess the evidence that is present at the time of the interview, in reference to the legal questions on which he has been asked to opine.

Is the Mental Status Examination "Objective"?

> *The river where you set your foot just now is gone – those waters giving way to this, now this.*
>
> *– Heraclitus, "Fragments"* [7]

We stated earlier that the MSE is analogous to the physical examination in other branches of medicine. However, psychiatric symptoms are in general more

apt to vary between evaluations than physical findings such as a tumor or a broken bone. More so than in general medicine, findings on the psychiatric examination may vary between examiners and at different points in time. This is relevant in legal proceedings, where discrepant expert opinions are subject to comparison and cross-examination.

In a 1961 article, Rosenzweig *et al.* presented findings from a study of the agreement between evaluators and consistency over time in findings on the MSE. They found that "while reliability (agreement) was not significantly influenced by individual bias in interpretation of concepts or by individual capacity to make observations it was significantly influenced by individual differences in interviewing technique" ([8], p. 1107). This suggests that if examiners are observing the same phenomenon, they are likely to record it in a similar manner, but that those elements of the MSE that require the examiner to elicit specific responses are susceptible to greater variation, depending upon the interviewer's focus and approach in asking questions.

Numerous factors may influence the result of the MSE, including the interviewee's degree of fatigue or willingness to be forthcoming, the examiner's interview style and technique, the setting where the interview occurs, etc. The evaluee's mental condition is subject to change over time, due to worsening or improvement of his mental illness, the effects of medication and the presence or absence of life stressors. The MSE is an assessment of present mental state and to some extent, differing expert opinions may reflect differing data at the time of evaluation. Nevertheless, as Ross and Leichner point out, "[t]he MSE is one of the tightest areas of clinical psychiatry, and its assessment is reliable both when the assesor's responses are restricted to highly discrete, operationalized items and in global, qualitative assessment" ([9], p. 111) even where the examiners' vocabulary in the description of details vary.

Conclusion

The MSE is a useful means of organizing clinical data concerning an individual's mental condition at the time of examination. The MSE refers to a particular time in the clinical interview when present state information is reviewed in a systematic manner, as well as to the written recording of this information, along with other observations obtained during face-to-face contact with the patient or forensic evaluee.

References

[1] Meyer, A.. (1951). Outlines of examinations; privately printed, 1918, in *The Collected Papers of Adolf Meyer*, *Volume III: Medical Teaching*, E.E. Winters, ed, The Johns Hopkins Press, Baltimore, pp. 224–258.

[2] Keller, M.B. & Manschreck, T.C. (1981). The bedside mental status examination – reliability and validity, *Comprehensive Psychiatry* **22**(5), 500–511.

[3] Trzepacz, P.T. & Baker, R.W. (1993). *The Psychiatric Mental Status Examination*, Oxford University Press, New York.

[4] Melton, G.B., Petrila, J., Poythress, N.G. & Slobogin, C. (1997). *Psychological Evaluations for the Courts: A Handbook for Mental Health Professionals and Lawyers*, 2nd Edition, The Guilford Press, New York, p. 42.

[5] Othmer, E., Othmer, S.C., Othmer, J.P. (2005). *Psychiatric Interview, History, and Mental Status Examination. Kaplan & Sadock's Comprehensive Textbook of Psychiatry*, 8th Edition, B.J. Sadock & V.A. Sadock, eds, Lippincott Williams & Wilkins, Philadelphia, pp. 794–826.

[6] Folstein, M.F., Folstein, S.E. & McHugh, P.R. (1975). 'Mini-mental state': a practical method for grading the cognitive state of patients for the clinician, *Journal of Psychiatric Research* **12**(3), 189–198.

[7] Haxton, B. (translator) (2001/2003). *Heraclitus. 41 Fragments*, Penguin Books, New York.

[8] Rosenzweig, N., Vandenberg, S.G., Moore, K. & Dukay, A. (1961). A study of the reliability of the mental status examination, *American Journal of Psychiatry* **117**, 1102–1108.

[9] Ross, C.A. & Leichner, P. (1988). Residents performance on the mental status examination, *Canadian Journal of Psychiatry* **33**, 108–111.

Further Reading

Taylor, M.A., Abrams, R., Faber, R. & Almy, G. (1980). Cognitive tasks in the mental status examination, *Journal of Nervous and Mental Disease* **168**(3), 167–170.

SUZANNE YANG AND DELANEY M. SMITH

Methanol ***see*** Alcohol: Use, Abuse, Tolerance, and Dependency

Method: Error *see* Error Rates in Forensic Methods

Microchemistry

Microchemistry, also known as *chemical microscopy*, was defined by Emile Chamot in elementary chemical microscopy as, "the application of the microscope to the solution of chemical problems". He also explained that microchemistry is chemistry on a small scale. The American microchemist, Benedetti-Pichler, defined microchemistry as the development, correlation, and systematization of the methods for handling small quantities of material, and for the observation of their properties. The scientific discipline of microchemistry has a rich and storied history going back to Pliny (23–79 AD), who describes an iron sulfate test. Microchemistry has been practiced for over 180 years, long before the development of instrumental methods. Microchemical tests provide a rapid and inexpensive way to obtain chemical data from microscopic size particles. Many rather simple tests have been developed to qualitatively analyze cations, anions, and functional groups. Francois-Vincent Raspail (1794–1878) is often recognized as the founder of chemical microscopy and the first true microchemist. In 1827, Raspail published results of his work on silica in Spongilla and calcium oxalate in the starch of monocotyledons. It is within this publication that he introduces the term *chemical microscopy*. Friedrich Emich (1860–1940) and Fritz Pregl (1860–1930) from the Technical University of Graz, Austria were the preeminent microchemists in Europe. The earliest complete book dealing with microchemistry was from an American, Theodore G. Wormley, The Microchemistry of Poisons, in 1867. There are many other exceptional scientists that have advanced chemical microscopy, an incomplete list of these individuals include E. Boricky, A. Streng, K. Haushofer, C. Klement, A. Renard, H. Behrens, A.C. Hutsse, C. Hinrichs, N. Schoorl, J. Donau, P.D.C. Kley, O. Tunmann, H. Molisch, C.W. Mason, L. Rosenthaler, L. Kofler, A. Kofler, F. Schneider, W. McCrone, H. Schaeffer, H. Keune, C. Fulton, E. Jungreis, Palenik, S., and J. Delly [1].

Practicing scientists working in chemistry, microanalysis, and serology can rely on many sources of literature in microchemistry. The classic texts by E.M. Chamot and C. W. Mason, Volumes 1 and 2, the Handbook of Chemical Microscopy [2] are essential. They were originally published in 1930–1931, and have gone through several editions and reprints. C. Fulton's Modern Microcrystal Test for Drugs [3] is a usefully text employed in drug identification. E. Jungreis' book, Spot Test Analysis, defines tests for clinical, environmental, forensic, and geochemical applications [4]. A large volume of periodical literature is available. There are three journals devoted to microchemistry: Mikrochemie (1923–1952), Mikrochimica Acta (1953-), and the Microchemical Journal (1957-). The Microscope (1937-) is an excellent source of literature on microchemistry, such as J. Hollifield's 2003 article, "Characterization of Squaric Acid Precipitates" [5] and T. Hopen and J. Kilbourn's article "Characterization and Identification of Water-Soluble Explosives" [6]. Many useful microchemical tests and spot tests have been developed over the years and incorporated into books, articles, and manuals in areas as diverse as paper and textile fiber analysis, explosives, drug chemistry, and food microscopy (*see* **Explosion Debris: Laboratory Analysis of**). Spot tests are color or precipitate reactions resulting from mixing an unknown and a test reagent on a ceramic or glass plate having a series of depressions (a spot plate). The fifth edition of the *Merck Index* describes over 4510 named reactions [7].

Long before the advent of instrumental chemical analysis "chemical microscopes" were employed to critically examine and identify materials in the physical world. These microscopes were based upon the petrographic microscope (used in the examination of mineralogical specimens) design first developed by Henry Fox Talbot in the 1840s. The microscope was equipped with Nicol prisms that polarized light. A circular rotating stage with a Vernier scale was employed to measure crystallographic angles. These microscopes are essentially the same type used today, referred to as *polarized light microscopes* (*PLM*), and are used in all modern forensic laboratories around the world (*see* **Microscopy: Low Power**; **Microscopy: High Power**). For most observations, a 10× objective and a 10× ocular, giving a total magnification of 100× is adequate. Occasionally, a 40× objective may be used giving a total magnification

of 400×. Scientists must have a good foundation in the basic principles of polarized light microscopy to understand what they are looking at and the terminology used to describe the crystals. Even no reactions or the formation of a precipitate but no crystals tell something about the material being tested.

Other equipment useful in microchemical testing include known chemically pure reagents, a small glass rod with a 3–4-mm smooth tip, glass or porcelain spot plates, disposable capillary pipettes, microscope slides, and spatulas [8].

Microchemical tests, also known as *microcrystal tests*, are used in both organic and inorganic characterization and identification of unknowns. E. G. C. Clarke states, in Isolation and Identification of Drugs [9], the value of crystal tests: "The microcrystal test is unsuitable as a primary method of identification of an unknown compound, as it does not lend itself to form the basis of an identification scheme. Its real value is as a means of final identification to confirm a provisional diagnosis made from chromatographic or spectrometric evidence, its extreme simplicity, the rapidity of which may be performed, and its high degree of specificity, rendering it ideal for this purpose". When properly performed, microcrystal testing is a useful tool for the analytical chemist. As with any other analytical tool it has its advantages and limitations. It is fast, requires very little sample, is often nondestructive (in that the tested material can be recovered), can distinguish among the isomers of some compounds, and, for many compounds, is highly specific. Some disadvantages are that when first starting, the analyst must test many compounds to learn what the products are and which reagents work best. Analysts must learn how to be consistent in reporting what they see, i.e., the written description of the crystals or drawings of the crystals must be reported in the same manner, there is no printout of the results (although with the advent of digital microphotography this problem can be eliminated), and some closely related compounds may give the same type of crystals.

The basic techniques of microcrystal testing and the examination of many substances encountered in forensic analysis are well documented in Volume II of Chamot and Mason. Microcrystal testing of materials requires getting the material in question to react with specific reagent(s) to obtain insight as to the composition of the material or to identify the substance. The nature of the material being examined, and, often, experience determine the method used. In inorganic microchemical analysis, ions of the material being tested need to be brought into solution. For most reactions water is used as the solvent. The solubility of the test material gives clues in identification. For example, sodium chloride rapidly dissolves in water whereas sulfur does not. Other times the solvent may need to be slightly basic or acidic. To perform a solubility test, a single particle (or a few particles) is placed on a microscope slide next to a very small drop of solvent. The solvent can be first placed on a glass rod and gently tapped to the microscope slide releasing the solvent. While using the stereobinocular microscope one of the test particles is pushed into the solvent and the solubility, if any, is recorded. If solubility is not observed, gentle heating over a small flame and quick microscopical observation reveal the status of solubility. Another direct method to observe solubility is "huffing" on a single crystal with warm breath. The heat and moisture from the mouth may be all that is needed to dissolve a particle. Chamot and Mason give three microscopical methods for the addition of reagents to test drops. Method I is the method most commonly used. The reagent and test particle are dissolved in separate drops of water on a microscope slide. They are joined together using a glass rod forming a narrow channel between the two drops. Diffusion occurs as the two drops contact each other forming precipitates and, or, characteristic crystals. Method II introduces a large crystal of the reagent to the test drop to form crystals. In method III, the reagent solution is drawn in a narrow channel across a dry film, obtained by evaporating to dryness a solution of the substance to be tested. A drop of the test solution is placed on a clean slide and a drop of the reagent is then placed near the dried material. Using a glass rod or a platinum needle, a small line of the reagent solution is pulled across the dried film being careful not to inundate the entire dried area and examine the crystals. The hanging drop method is another elegant test using sublimation to form crystals. A drop of the test solution, a portion of the dried material, or material containing the unknown substance is placed on a microscope slide, and a cut glass ring from a test tube is placed over it. A drop of another material that causes the sample to volatilize is added. A drop of the test reagent is placed on a coverslip, which is then inverted and placed over the opening (Figure 1). The volatilized material rises and reacts with the reagent. This is a good method when one has a mixture and

cannot develop identifiable crystals with any of the other methods and the compound of interest is known to volatilize.

Coupled with microcrystalline tests are spot tests (color tests), the basis of many tests by law enforcement agencies to test street drugs to determine what they may be. However, in the hands of the trained microscopist, many of these tests are very specific and provide detailed information. F. Feigl's books on inorganic and organic spot tests are some of the principle reference books useful when conducting these tests [10, 11]. The Merck Manual, fifth edition, is another good reference for finding tests for specific materials. These tests can be used to determine the presence of metals, chemical compounds (inorganic and organic), certain elements, anions, cations, materials in food products, building materials, poisons, plant material, etc. Feigl based his tests on reactions that produced a macroscopic color; however,

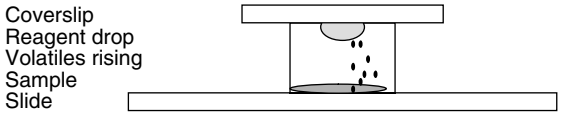


Figure 1 Schematic diagram of a sublimation cell used in the hanging drop method

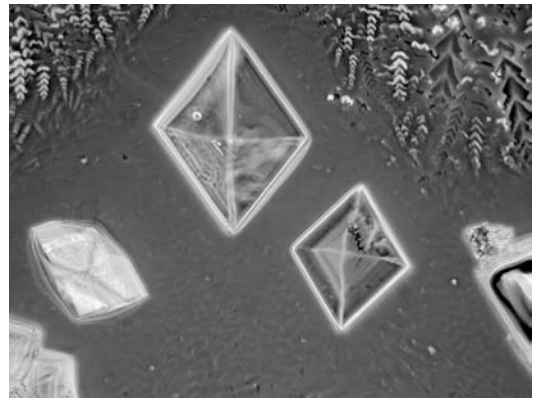


Figure 2 Recrystallization of sodium nitrate from a drop of water. Original magnification 200×

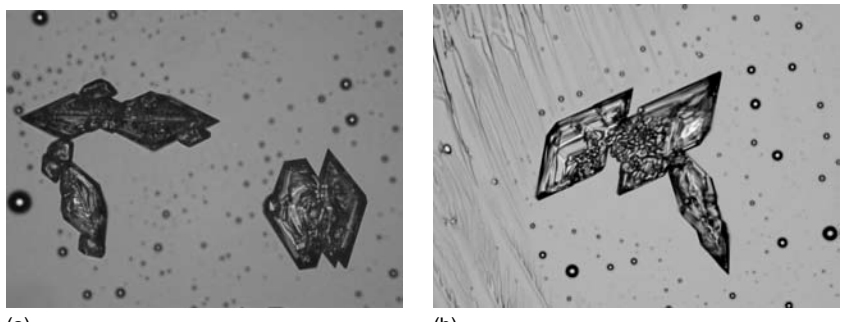


Figure 3 Sulfur recrystallized from chloroform. Original magnification 200×

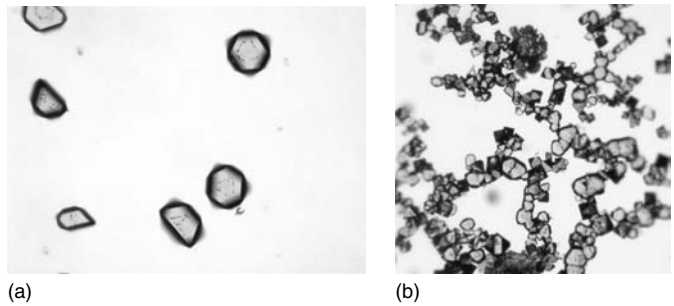


Figure 4 Isotropic octahedral crystals indicate the presence of ammonium. (a) magnification 200×; (b) 100×

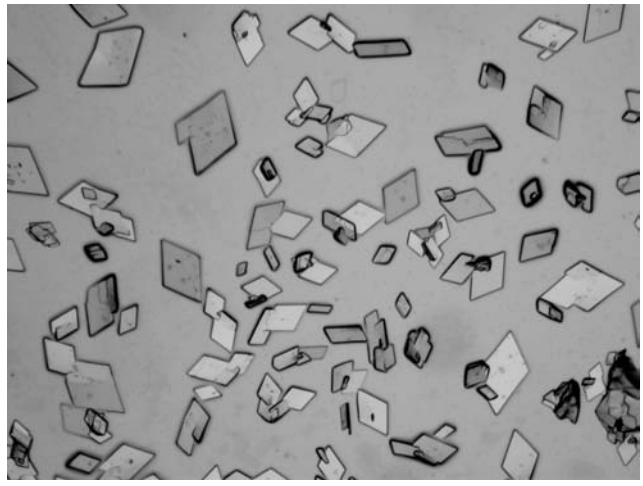


Figure 5 Yellow organic paint pigment using the sublimation cell. Original magnification 200×



Figure 6 Takayama test for the presence of blood. Original magnification 200×

when these tests are conducted on a microscale, the reactions need to be viewed using the PLM. Often the colored product forms a precipitate that may crystallize forming distinctive crystals that can be characterized.

Microchemical tests and spot tests are used by forensic scientists in a number of disciplines. Examples of some of these tests are included with photomicrographs to convey the science and art of microscopy.

In the analysis of explosive residue, microchemical tests are often used (*see* **Explosion Debris: Laboratory Analysis of**). Inorganic water-soluble explosive oxidizers can be identified by the direct addition of a particle into a small drop of water on a microscope slide. By allowing a known particle to dissolve and recrystallize, characteristic well-formed crystals of the oxidizer can be quickly identified (Figure 2). Sulfur, an ingredient in black powder, can be recrystallized from methylene chloride or chloroform and its characteristic bipyramidal crystals observed (Figure 3). Ammonium can be identified by the hanging drop method. A few grains of the test chemical such as ammonium perchlorate, are placed

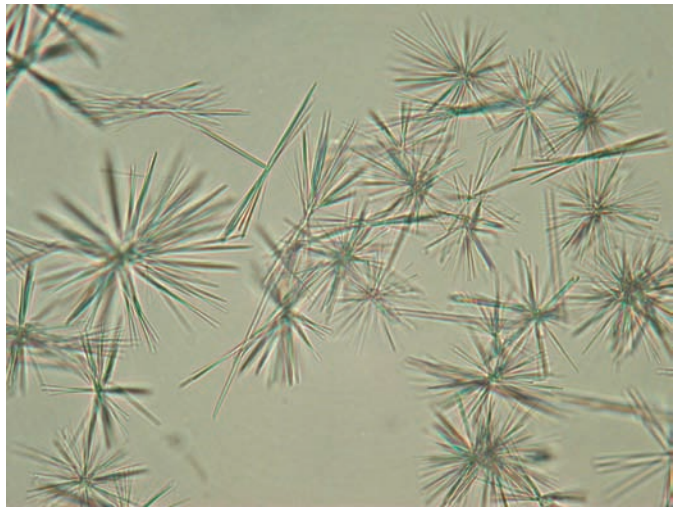


Figure 7 Colorless needles of gypsum, which formed in the presence of calcium. Original magnification 200×



Figure 8 A group of reagents used in drug chemistry

onto a microscope slide and covered with one drop of 2N sodium hydroxide. One drop of platinic chloride solution is placed on the underside of a coverslip and then placed over the glass ring. Soon isotropic octahedra indicative for the presence of ammonium, forms (Figure 4). Many organic paint pigments can be identified by placing a test particle on a microscope slide that is covered with a small cut test tube ring of glass in which a coverglass is placed on top (*see* **Paint**). The slide is placed on a hot plate and as the sample heats crystalline sublimation products form on the underside of the coverglass (Figure 5).

In the examination of microscopic flakes of material suspected to be blood, the Takayama test can be used. A portion of a suspected blood flake is placed on a microscope slide and covered with a coverslip. One or two drops of the Takayama reagent (prepared using sodium hydroxide, pyridine, glucose, and distilled water) are instilled under the coverslip. The slide is warmed for approximately 30 s on a hot

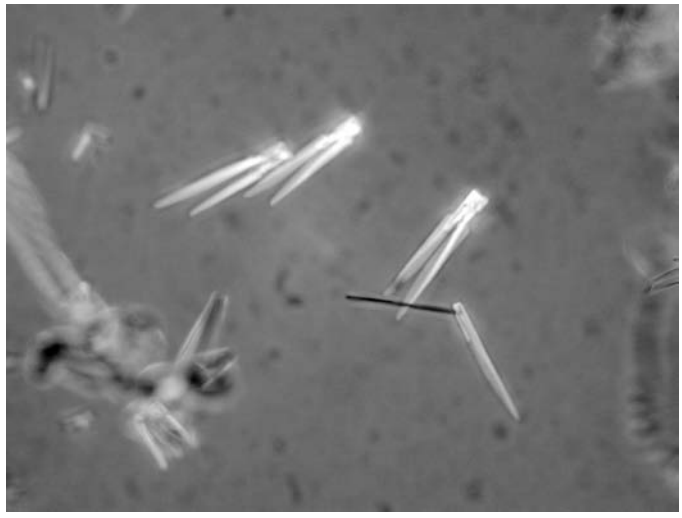


Figure 9 Gold chloride solution in phosphoric acid showing distinctive "clothe-pin" crystals indicative of methamphetamine. Original magnification 400×

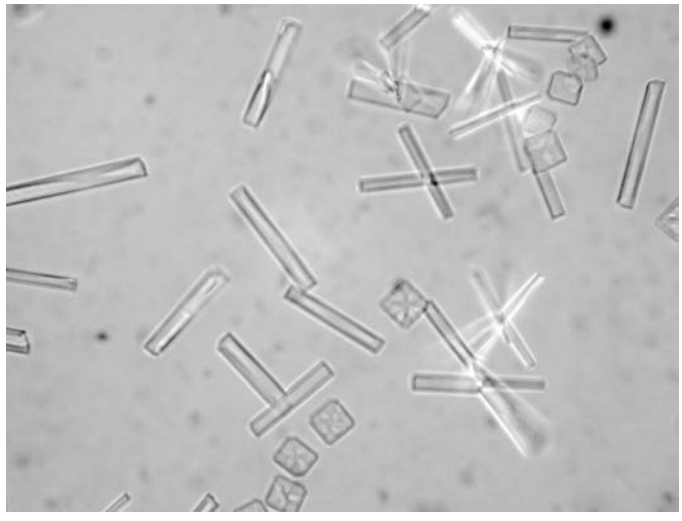


Figure 10 Gold chloride solution in phosphoric acid showing distinctive crystals indicative of dl-amphetamine. Original magnification 400×

plate. The presence of needle-shaped or rhomboid hemochromogen crystals (Figure 6) indicates a positive test for the presence of blood. This test is reported to react positive with as little as 0.001 ml of blood or 0.1 mg of hemoglobin. To test for the presence of calcium, a particle of suspected calcite ($CaCO_3$) is dissolved in a small drop of dilute hydrochloric acid. The evolution of carbon dioxide bubbles indicates CO_2 is being liberated. Using method I, a drop of dilute sulfuric acid is added. Observation along the edges yields birefringent colorless needles of gypsum ($CaSO_4 2H_2O$), which form in the presence of calcium (Figure 7). Many microchemical tests and reagents have been developed for the identification of drugs (Figure 8). Commonly used microchemical tests in drug chemistry include the use of gold

chloride in phosphoric acid to confirm the presence of methamphetamine (Figure 9) and dl-amphetamine (Figure 10).

In spite of the continuing value of microcrystalline tests, lack of training and modern tendency to rely on instrumentation results in them being used less in forensic laboratories than some years ago.

References

[1] Delly, J.G. (2006). *The Literature of Classical Microchemistry, Spot Tests, and Chemical Microscopy*, The Eyepoint, http://www.Modern Microscopy.com.

[2] Chamot, E.M. (1938/1940). *Handbook of Chemical Microscopy*, John Wiley & Sons, New York, Vol. 1 and 2.

[3] Fulton, C. (1969). Modern microcrystal test for drugs, in *The Identification of Organic Compounds by Microcrystalloscopic Chemistry*, Wiley Interscience, New York.

[4] Jungreis, E. (1997). *Spot Test Analysis: Clinical, Environmental, Forensic, and Geochemical Applications*, 2nd Edition, John Wiley & Sons.

[5] Hollifield, J. (2003). Characterization of squaric acid precipitates, *The Microscope* **51**(2), 81–103.

[6] Hopen, J.T. & Kilbourn, J.H. (1985). Characterization and identification of water soluble explosives, *The Microscope* **33**, 1–22.

[7] Merck (1940). Chemical, clinico-chemical reactions, tests and reagents, in *The Merck Index*, A 367 page table (623–990), 5th Edition, Merck Publishing.

[8] Delly, J.G. (2006). *Essentials of Polarized Light Microscopy*, 3rd Edition, McCrone Associates, College of Microscopy.

[9] Clarke, E.G.C. (1969). *Isolation and Identification of Drugs*, The Pharmaceutical Press, London, Reprinted 1971.

[10] Feigl, F. (1958). *Spot Tests in Inorganic Analysis*, 5th Edition, E.O. Ralph, ed, Elsevier, Amsterdam.

[11] Feigl, F. (1966). Spot tests in organic analysis, in *Collaboration with Vinzenz Anger*, E.O. Ralph, ed, 7th Edition, Elsevier, Amsterdam.

WILLIAM M. SCHNECK

Microsatellites

Prior to the 1980s, forensic genetic testing targeted polymorphic protein and blood group marker systems, for example ABO blood grouping and human leukocyte antigens (HLAs). These methods were limited by their low level of polymorphism and the relatively large amount of high-quality biological material required [1]. Also, there was limited analysis of biological material other than blood.

In 1980, Wyman and White used the restriction nuclease *Eco*RI to demonstrate variation in length between individuals at one particular noncoding locus on chromosome 14 [2]. The digested DNA was separated on a gel and detected by Southern blot analysis using a radioactive probe. This type of analysis was called *restriction fragment length polymorphism* (RFLP). This was the first example of a DNA marker that could be used to distinguish between individuals.

In 1985, Jeffreys *et al.* described a new type of polymorphic DNA marker termed *minisatellites* or *variable number tandem repeats* (VNTRs). Minisatellites are composed of sequences varying from 15 to 50 bp in length, repeated tandemly up to a total length of 20 kb [3]. It was this DNA technology that was first used in a criminal investigation (see Case study 1).

Case Study 1 This is a study of two murder investigations involving teenage girls in Leicester, England, in the mid-1980s. Owing to the case circumstances, investigators thought they were perpetrated by the same individual. Subsequently, a 17-year-old male confessed to the latter murder but denied involvement in the first. Using minisatellite technology, Professor Jeffreys was able to show that the biological materiel was left behind by the same person and that the man who had previously confessed could not be this person. A mass screen was undertaken with DNA sampled from 600 men from surrounding villages. A local baker, Colin Pitchfork, was eventually charged and convicted of the murders [4].

However, minisatellite analysis using RFLP was time consuming, difficult to interpret, and required over 50 ng of DNA.

In the 1990s, forensic DNA analysis advanced again to utilize PCR technology. PCR was originally applied forensically to HLA DQα and subsequently to minisatellites (known as *amplified fragment length polymorphisms*, *AMP FLPs*). PCR analysis of

VNTRs was quicker and required only approximately 5 ng of DNA; however, the resulting fragments of DNA were still relatively large [5].

By reducing the targeted repeat sequence, the amplified product is smaller and therefore less prone to be affected by sample degradation as many forensic samples may be. For this reason, microsatellites or short tandem repeats (STRs) amplified via the PCR became the preferred method for many forensic laboratories in the mid-1990s (*see* **Short Tandem Repeats**). STRs are lengths of DNA 2–5 nucleotides in length repeated tandemly. Their short length, variability between different individuals, and their ability to be amplified via PCR make them extremely useful as forensic DNA markers. Refer to **Short Tandem Repeats** for a more thorough discussion.

References

[1] Gill, P., Jeffreys, A.J. & Werrett, D.J. (1985). Forensic application of DNA 'Fingerprints', *Nature* **318**, 577–579.

[2] Wyman, A.R. & White, R. (1980). A highly polymorphic locus in human DNA, *Proceedings of the National Academy of Sciences* **77**(11), 6754–6758.

[3] Jeffreys, A.J., Wilson, V. & Thein, S.L. (1985). Hypervariable 'minisatellite' regions in human DNA, *Nature* **314**, 67–73.

[4] Napper, R. (2000). A national DNA database. The United Kingdom experience, *Australian Journal of Forensic Sciences* **32**, 65–70.

[5] Helmuth, R., Fildes, N., Blake, E., Luce, M.C., Chimera, J. & Madej, R. (1990). HLA-DQα allele and genotype frequencies in various human populations, determined by using enzymatic amplification and oligonucleotide probes, *American Journal of Human Genetics* **47**, 515–523.

JO-ANNE BRIGHT

Microscopy: FTIR

Introduction

Infrared spectroscopy is a technique widely practiced by forensic scientists and analytical chemists in general. The principle behind the technique is to shine infrared radiation onto a specimen and record the way in which it absorbs the radiation. The way in which this is done is by using a device called an *infrared spectrometer*, which, in it simplest implementation, irradiates the specimen at any one time with only a single wavelength of light in the infrared range. After absorption at that wavelength has been measured and recorded the spectrometer moves on to another wavelength, absorbance is recorded, and so on, until the entire infrared range has been covered. In classical instruments, this requires the spectrometer to make use of a device that works like a prism, and slits that reject radiation around the wavelength being measured. A graph of the magnitude of absorbance at each wavelength is referred to as that *specimen's infrared spectrum*. In general, other types of light (or electromagnetic radiation) can be absorbed by specimens as well, for example visible light is absorbed by a number of objects, resulting in a physical manifestation we know as color. The reason why infrared radiation, in particular, is useful in an analytical chemistry sense is because its absorbance arises from the interaction between the radiation and the many different connections (or bonds) between the atoms that make up the molecules that make up the specimen. What distinguishes one type of molecule from another is the types of atoms they contain and the arrangement in which the atoms are bonded to each other. As infrared spectroscopy is very sensitive to the bonds and the atoms present in molecules, it is a very powerful technique for analyzing materials to identify the molecules that make it up. By comparison, visible light absorption is much less sensitive to the atoms and bonds present in molecules, therefore it has much less analytical power. This can be appreciated by considering, for example, three objects with exactly the same shade of red, such as an apple, paint on a car, and a piece of dyed fabric. All three objects exhibit identical visible light spectra, but as the molecules of the chemicals responsible for the color are all different, their patterns of absorption of infrared light (i.e., their infrared spectra) are all different.

The infrared absorbance experiment described earlier is conducted in the laboratory by shining a beam of infrared radiation of about 10–15 mm in diameter onto the specimen. This is not a limitation if the objective of the exercise is to analyze a drug seizure or a large sample of paint or liquid. However, a substantial amount of evidentiary material is small or microscopic. A variant of infrared

spectroscopy, which arises because there is a lower limit to the size of specimen that can be handled by standard laboratory equipment, is called *infrared microscopy* (or more correctly *infrared microspectrometry*). In infrared microspectroscopy, the infrared beam of large diameter is reduced in size to about 0.2 mm, and then caused to interact with the microscopic specimen. The infrared radiation emerging from the specimen and its vicinity is magnified and passed on to a detector. Three general types of equipment can be used for microspectroscopy: a beam condenser, which is an accessory that is attached to a standard spectrometer and makes use of its infrared beam and detector; an infrared microscope accessory, which also attaches to a standard spectrometer, but carries its own detector; and an infrared microspectrometer, which is an instrument that only operates on the microscopic domain. Beam condensers, although inexpensive, are difficult to use for trace evidence and do not produce good results for very small specimens, for this reason they are not commonly used in modern forensic laboratories. The other two accessories produce a much smaller beam at the specimen. The beam is circular, but it can be trimmed into square or rectangular shapes by curtains that can be drawn across the beam (these are referred to as a *diaphragm* or an *aperture*) before it is focused onto the specimen. This is important because fibers and many microscopic particles are not circular. Perhaps the most important feature as far as the analyst is concerned is that infrared microscope accessories and microspectrometers allow the specimen to be viewed and moved around at the focal point of the condensed infrared beam, and allow the size and orientation of diaphragms to be adjusted so that a region of interest within the specimen is analyzed, rather than the entire specimen.

In forensic science infrared microspectroscopy is used for the examination of microscopic particles of paint, rubber, explosives, and plastic, individual textile fibers, individual crystals of drugs and diluents, and individual textile fibers. Although it can be used for the examination of minerals (such as extenders in paint), it is not used for the examination of glass or metallic particles. Although there were early attempts at infrared microspectroscopy, its widespread popularity has only emerged in the last two decades owing to the widespread release of Fourier transform spectrometers. In conventional spectrometers, as described earlier, only single wavelengths of radiation are presented to the specimen at any one time. The wavelength selection process is not very efficient, with the result that only a very dim beam interacts with the specimen. If this beam is passed through a microscope to interact with a very tiny specimen, it is very difficult to measure the small levels of absorbance involved. In Fourier transform instruments, the entire range of infrared radiation is presented at high brightness to the specimen. The signal received from the specimen is a sum of all absorbances over all wavelengths over the entire infrared region, and a special mathematical procedure, called *Fourier transformation*, is used to "unscramble" the summed data into the infrared spectrum (*see* **Paint**; **Examination of Fibers and Textiles**; **Drug Analysis**).

Forensic scientists use infrared spectroscopy in two, nonexclusive general ways. Firstly, for the purposes of identification, the infrared spectrum of an unknown specimen is searched against entries on a spectral database until a "match" is found. In the case of a mixture of compounds, infrared spectroscopy is relatively straightforward in that the spectrum of the mixture is simply a direct combination or sum of the spectra of the individual components. Secondly, infrared spectroscopy is also used in a comparative sense, that is, to examine whether two pieces of material might share a common origin (for example comparison of a foreign fiber retrieved from a victim and fibers from a garment worn by a suspect). Some limitations of infrared spectroscopy are that some materials do not absorb infrared light at all (for example, common salt, metals, etc.), it is not a good technique for the detection of traces within a material (for example, explosives residues within water or soil), and sometimes it can be difficult to distinguish between members of a particular class of materials (for example, within the nylon textile fiber class, it is difficult to distinguish between nylon 6 and nylon 6, 6).

Infrared Transmission Microspectroscopy

Figure 1 depicts the important features found in an infrared microspectrometer or a microscope accessory functioning in transmission mode. The "lenses" in these devices are manufactured from paired convex and concave mirrors, because infrared radiation is absorbed by glass, and it is impractical to make

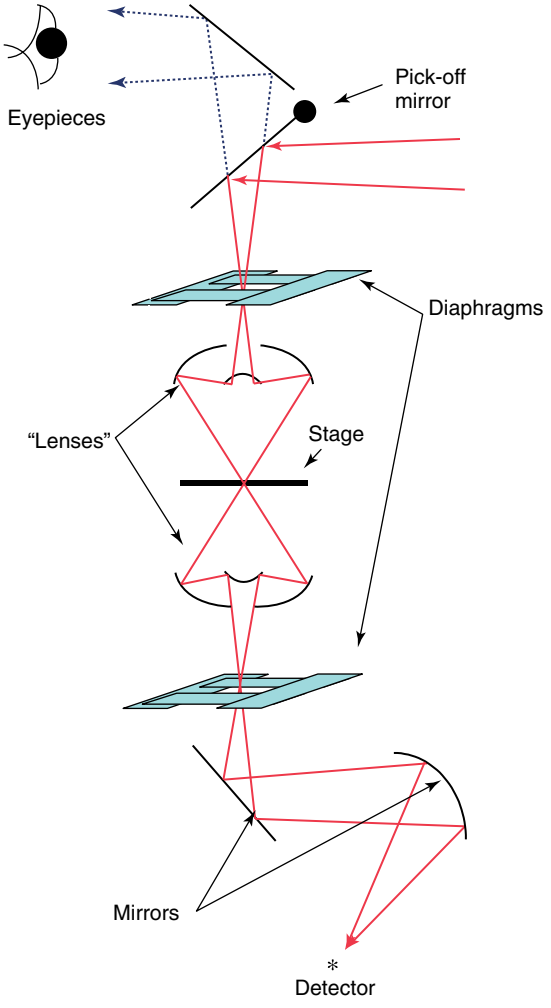


Figure 1 Schematic diagram of a typical infrared microspectrometer or infrared microscope accessory. The solid arrows and lines indicate the path for infrared radiation, while the dashed lines and arrows indicate the path for visible light to the observer once the pickoff mirror has been swung to the vertical position. Each "lens" comprises a concave mirror and a convex mirror depicted by the black lines. Not all microscopes have the diaphragm situated between the stage and the detector (termed a redundant diaphragm) [Reproduced from Ref. 3. © Taylor and Francis Group, 1999.]

lenses from material that is transparent to infrared radiation. Each diaphragm is made from four thin sheets of material opaque to infrared radiation that can be moved independently across the beam. At the center of the device is a stage upon which the specimen is mounted. X, Y, and Z controls allow the stage to be moved in three orthogonal directions in space so that the specimen can be placed into the center of the beam and brought into the focal point of the beam. The pickoff mirror can be swung into the position shown in Figure 1, which allows the infrared beam to pass through the specimen for analysis, or it can be swung vertically so that the specimen can be viewed through the eyepieces at high magnification, positioned appropriately, and a diaphragm set up. Further discussion of infrared microscopes and their application to forensic science can be found in [1–3].

Figures 2(a and b) illustrate the function of the diaphragm. Figure 2(a) depicts the situation where the diaphragm curtains are fully retracted, which results in a circular beam of maximum diameter illuminating the specimen. Obviously a substantial portion of the beam reaches the detector having not passed through the specimen. Radiation that has not passed through the specimen but reaches the detector is referred to as "stray' radiation and its intrusion into spectroscopy is not desirable. To reduce the amount of stray radiation the diaphragm can be configured as shown in Figure 2(b), where the specimen is illuminated with a small, appropriately shaped spot of radiation. Figure 3 depicts two spectra recorded from the same single polyethylene terephthalate (also know as PET or "polyests") textile fiber. The spectrum in (a) was recorded with a diaphragm established as depicted in Figure 2(a) (i.e., high stray light), while that in (b) was recorded using a diaphragm as shown in Figure 2(b) (lower stray light). The effect of stray light is that peaks do not attain their true relative heights, and this is called *photometric inaccuracy*.

Although the spectrum in Figure 3(b) is good, it is still not perfect photometrically; this arises because of the effects of a phenomenon known as *diffraction*. When any radiation is caused to pass through an opening or passes close by an opaque edge, bending of the radiation (diffraction) takes place. The amount of bending is inversely proportional to the frequency of the illuminating radiation; infrared being of a frequency lower than visible light bends more, and low frequency infrared (i.e., that recorded at the "right hand side" of spectra) bends more than high frequency infrared. As a consequence, even though the diaphragm in Figure 2(b) casts a sharp shadow on the specimen to the human eye (which operates

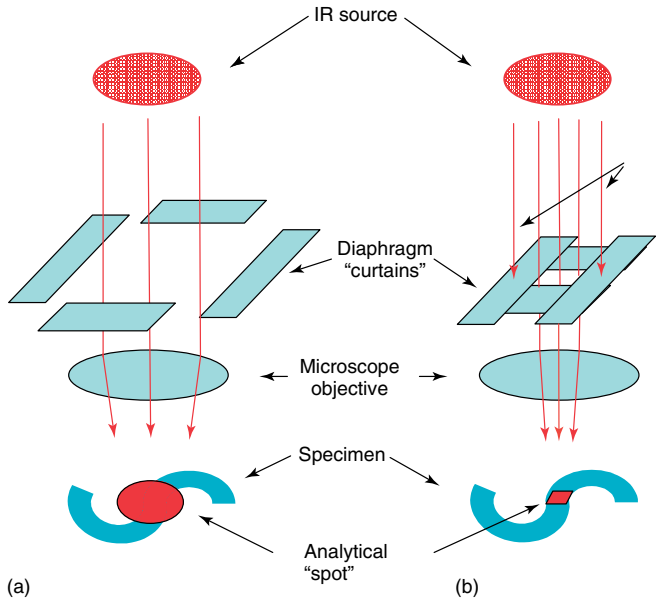


Figure 2 (a) Shows an infrared microscope with a textile fiber on the specimen stage. The curtains of the diaphragm are moved back away from the infrared beam path and as a consequence a spot of maximum size falls on to the fiber. Much of the beam does not interact with the specimen and goes on to reach the detector as stray radiation. (b) Depicts the situation where the curtains of the diaphragm are adjusted to trim off stray radiation [Reproduced from Ref.3. © Taylor and Francis Group, 1999.]

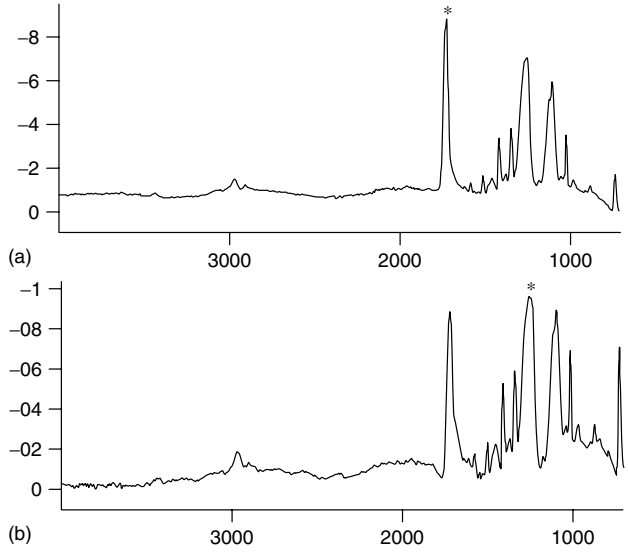


Figure 3 Spectrum (b) was collected from a single polyethyleneterephthalate textile fiber using an infrared microscope configured as shown in Figure 2(a) (i.e., with no diaphragm). Spectrum (a) was acquired from the same fiber with the microscope configured as shown in Figure 2(b). Note the differences between the two spectra with regard to the absolute absorbance values for the peaks (e.g., the two largest peaks, marked with asterisks, have values of about 0.1 and 0.8), of the two largest peaks in each spectrum (marked with an asterisk), the fact that the two maxima are at different frequencies, and the apparent "magnification" of small peaks in spectrum (b) [Reproduced with permission from reference [3], p. 189.]

using comparatively high frequency visible radiation), when infrared radiation is used to illuminate the specimen the shadow is not distinct. Figure 4 is a refinement of Figure 2(b) showing diffracted radiation and an approximation of the "shadow" of two opposite edges of the diaphragm imaged through the optics of an infrared microscope as "seen" by the specimen. Diffraction causes more stray light to reach the specimen than the size of the diaphragm would suggest. Any attempt to completely eliminate stray light is accompanied by compromises and diminishing returns. It is of course possible to reduce the size of the diaphragm to compensate for diffraction, but as indicated in Figure 4, diffraction does not stop at a certain point beyond the edge of the diaphragm, it continues indefinitely, albeit at a rapidly diminishing magnitude. For relatively large specimens (greater than about 30–40 μm in size), it is an effective tactic to configure a diaphragm to be about 5–10 μm smaller than the specimen and eliminate a large proportion of the stray light; reducing the size further reduces stray light further, but by ever diminishing returns. This is not recommended for specimens much smaller than about 30 μm as the reduction in the size of the diaphragm starts to severely attenuate the intensity of the radiation presented to the specimen, which is translated into a reduction in the signal to noise ratio of the spectral data, with the result that small peaks can disappear into the baseline. Depending upon the microscope used, a different diaphragm configuration (called *dual remote diaphragms*, as depicted in Figure 1) can be employed to improve the situation, but in any event, for very small specimens a compromise must be drawn between signal to noise ratio and rejection of stray light; it must be assumed that some residual stray light is present and its potential spectral contribution appreciated. In forensic comparisons it is a wise precaution to ensure that spectra of questioned and reference specimens are collected under conditions of equivalent stray light (i.e., specimens of approximately the same dimensions, and diaphragms of approximately the

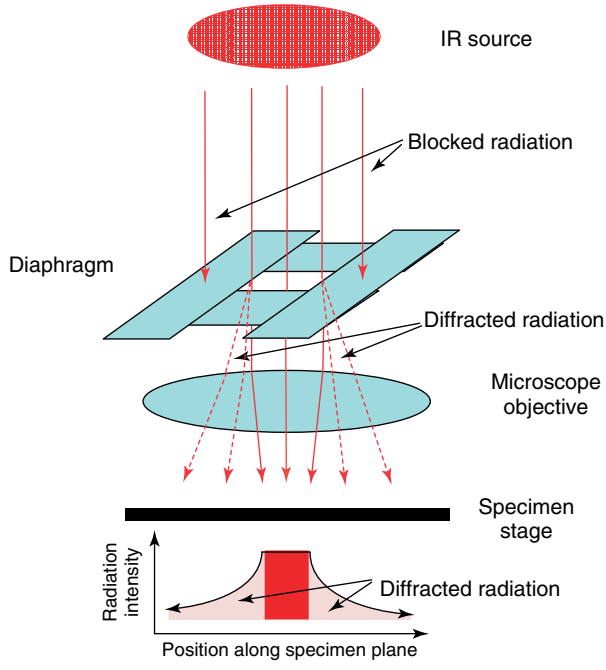


Figure 4 This diagram shows diffracted infrared radiation (dotted red arrows) that "bends" past the diaphragm established for a specimen as shown in Figure 2(b). An exaggerated approximation of the intensity of radiation across the specimen stage is depicted in the graph below. Instead of a sharp shadow (i.e., a rapid drop to zero intensity) beyond the edge of the fiber a more gradual drop to zero takes place. All the radiation that contributes to the pink zone in the graph reaches the detector as stray radiation

same dimensions). If these precautions are taken, the analyst can be certain that spectral differences between different specimens reflect a genuine compositional difference between the specimens and not spectral artifacts (i.e., these precautions guard against Type 1 errors, or false elimination of association, stray light effects are very unlikely to lead to Type 2 errors, or false association). Efforts should also be taken to minimize stray light during the recording of spectra for spectral libraries and chemometric datasets, otherwise poor matching/clustering might result.

In the case of a homogeneous specimen, such as a chip of single layer paint, the main consequence of diffraction in infrared microspectroscopy is photometric inaccuracy arising from stray light. In the case of a heterogeneous specimen, such as a cross-section of multilayer paint, the consequences of diffraction are a little more serious. Figure 5 refines the situation in Figure 4 with the diaphragm established to isolate the light gray central paint layer from adjacent dark gray and clear layers. As in Figure 4, diffracted infrared radiation strikes beyond where the diaphragm suggests it should. In this situation, however, instead of the diffracted radiation traveling through air and reaching the detector as stray radiation, it travels through adjacent paint layers and on to the detector, with the result that the acquired spectrum mostly relates to the blue layer but contains contributions from the adjacent layers. Again, with large specimens it is desirable to arrange the diaphragm well away from any interface with adjacent layers, but again it is also not an effective option with smaller specimens. For additional material relating to the effects of diffraction and stray radiation upon infrared microspectroscopy see [3–5].

For infrared transmission analysis of polymeric materials, and others that strongly absorb infrared radiation, it is desirable that the thickness of the specimen does not exceed 10–20 μm. If the specimen is much thicker than, then the possibility arises that the infrared beam at some frequencies will be completely absorbed, and this leads to photometric inaccuracies much greater than those arising from stray light and diffraction. It is usually the case that forensic specimens, even single textile fibers, exceed the desirable thickness, so it is usual in infrared transmission microspectroscopy to treat the sample in some way to make it thinner. The use of a microtome, crushing the specimen between diamond anvils, pressing or rolling it, or carefully slicing sections away using a scalpel or microknife are all effective. A collateral benefit of crushing or rolling the specimen is that it makes it wider; this allows for the usage of bigger diaphragms and therefore allows for better rejection of stray light.

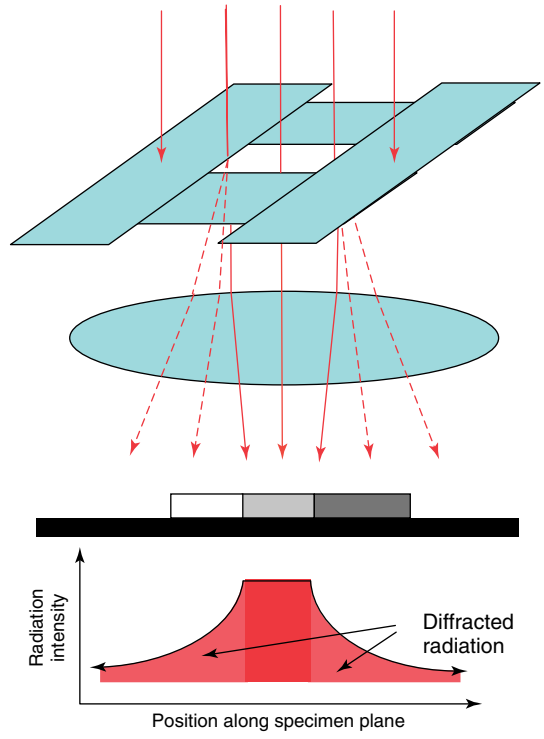


Figure 5 This diagram shows a cross section of three-layer paint placed under an infrared microscope with the diaphragm established on the interfaces between the light gray central layer and the adjacent clear and dark gray layers. As in Figure 4, the diffracted radiation and its intensity across the sample stage are depicted as dotted lines and light shading, respectively. Diffracted radiation passes through the clear and dark gray layers, carrying their spectral characteristics to the detector

Infrared Reflectance Microspectroscopy

Infrared microspectroscopy is not restricted to transmission techniques. By the use of mirrors it is possible to collect the infrared radiation reflected from the surface of the specimen and transfer it to the detector to acquire reflectance spectral data. For thin or transparent specimens, it is also possible to mount them on

a highly reflective substrate, such as aluminum foil or a gold coated glass microscope slide, and collect not only the radiation reflected from the surface of the specimen but also the radiation that has passed through the specimen, reflected off the substrate, and then traveled back through the specimen. This technique is sometimes referred to as *transflection*, as it is a mixture of transmission and reflectance spectroscopy. In this technique, the specimen thickness is effectively doubled as far as transmission is concerned, therefore it is a very useful method for very thin specimens. Reflection (and transflection) spectroscopy produces baseline artifacts (dips) that can be corrected using a mathematical function called a *Kramers–Kronig transformation*. Most spectrometer manufacturers offer software applications that perform this transformation.

Another technique of use to the forensic scientist is attenuated total reflectance (ATR) spectroscopy. In this technique, the beam from the microspectrometer is directed into a special crystal and caused to reflect internally off a face of the crystal. At the point of reflection a small proportion of the beam exists beyond the face of the crystal. This so-called evanescent wave can interact with material in contact with the surface of the crystal, and if the material absorbs in the infrared region then the reflected beam is slightly attenuated as a result. Figure 6 illustrates the essential features of ATR spectroscopy. The reflected beam is measured with and without the material in contact with the crystal and the difference between the two states yields the ATR spectrum. The distance that the evanescent wave emerges from the crystal depends upon the frequency of the incident radiation; low frequency radiation emerges further than high frequency radiation. As a result, low frequency infrared radiation interacts with a specimen placed in contact with the crystal more strongly compared to high frequency infrared radiation. Compared to a transmission spectrum of the same substance, an ATR spectrum shows an ever increasing intensification of peaks toward the right hand side of the spectrum (i.e., to the low frequency end). This means that if ATR spectra are to be compared against transmission spectra, there has to be some correction of the ATR data; this also can be done using software. The actual distance that the evanescent wave emerges from the surface of the crystal is very small (the order of microns), and therefore it only penetrates a very short distance into the specimen. Obviously ATR spectroscopy is very surface sensitive; this must be borne in mind if specimens are contaminated on their surface, weathered, or laminated. In practice, ATR microspectroscopy is very simple. Microscopes can be configured with an accessory crystal fitted to the objective, or special objectives can be purchased. The specimen is placed on a strong substrate onto the stage of the microscope, the region of the specimen to be analyzed is selected and then, using the Z control of the stage, the specimen is brought into contact with the ATR crystal. It is preferable that the microscope has some means by which the pressure brought to bear on the specimen is controlled, otherwise it is possible to damage components by using too much force, or collect poor data as a result of insufficient pressure.

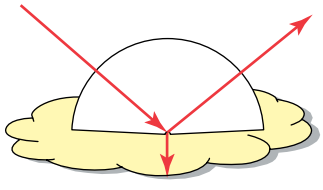


Figure 6 This diagram depicts a hemispherical attenuated total reflectance crystal of an infrared microscope pressed against a piece of paint. The infrared beam traveling from the source to the detector via the crystal is depicted by the solid arrows. At the point of reflection off the face of the crystal a small portion of the beam emerges from the face of the crystal (the evanescent wave, depicted as the small arrow) and interacts with the paint

Infrared Microspectral Imaging

Infrared microspectral imaging involves collection of many infrared spectra at many positions (called *pixels*) across the specimen The image is produced by selecting a frequency of interest (or a frequency range of interest), software then represents each pixel as a coloured spot the brightness of which is proportional to the infrared absorption at the frequency of interest at each pixel. As the image originates from the abundance and distribution of functional groups (measured by means of their characteristic infrared group frequencies) within the specimen, it illustrates the spatial distribution of chemically equivalent and chemically distinct regions in the specimen. Infrared microspectral imaging, therefore, can be very informative. As the image relates directly to

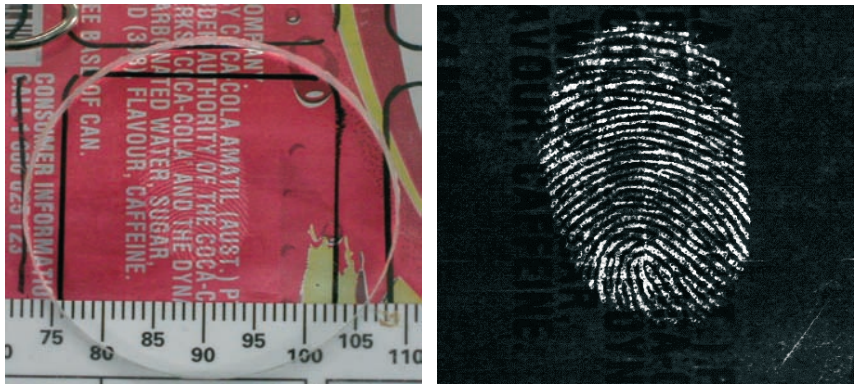


Figure 7 (a) The image was collected using visible light. It shows a fingerprint (circled) deposited on a Coke™ beverage can that has been treated with Superglue™ resin. (b) The image is a mosaic of over 1000 individual 700 μm^2 infrared images of the fingerprint displaying absorbance intensity at 1789 cm^{-1} (the C=O stretch frequency of Superglue™), Image supplied by B Reedy and M Tahtouh, Centre of Forensic Science, University of Technology Sydney, Sydney, Australia [Reproduced with permission from B Reedy and M Tahtouh, University of Technology Sydney.]

chemical information within the specimen, infrared microspectral imaging is most frequently referred to as *infrared chemical imaging*, or sometimes as *hyperspectral imaging*.

Prior to about 1995, the only way to construct chemical images was to place the specimen under an infrared microscope, establish a diaphragm delineating a certain portion of the specimen (say a 20 $\mu m \times 20\,\mu m$ area), collect a spectrum of that pixel, move the specimen 20 μm in some direction and record another pixel spectrum, and keep repeating the process until the desired area on the specimen has been analyzed. Even though this was usually accomplished with the aid of an automated microscope stage, it was a time-consuming process and low fidelity images were usually produced. In modern instruments, instead of a single detector an array of many detectors (such as 256, 1024, or 4096) arranged in a square pattern (called a *focal plane array*) is used to collect a set of spectra simultaneously from a square array of pixels (such as 256, 1024, or 4096) in the specimen. As all spectra are acquired simultaneously, focal plane array mapping is very rapid. Furthermore, images can be acquired at high spatial resolution without a requirement for apertures. If it is important to image an area larger than the field of view of the microscope then a motorized stage can be employed to move the specimen to an adjacent location and collect another data image set. Data image sets can then be displayed together to form a mosaic image of the specimen. The potential that microspectral infrared imaging offers to forensic science has been demonstrated through its successful application to fingerprint visualization [6, 7] and paint examination [8]. Figure 7 shows the greatly enhanced contrast possible with infrared microspectral imaging. The image on the left shows a fingerprint enhanced by Superglue™ fuming and photographed using visible light. The fingerprint (circled) is barely visible. The infrared image, on the right, displays the relative intensity of infrared absorbance at 1789 cm^{-1} at each pixel in a mosaic of over 1000 smaller infrared images. The infrared image has such high contrast between the background (a Coke™ beverage can) and the fingerprint because the background paint has very little absorbance at 1789 cm^{-1} while the superglue has a very strong absorbance.

References

[1] Humecki, H.J. (ed) (1995). Practical spectroscopy, *Practical Guide to Infrared Microspectroscopy*, Marcel Dekker, New York, Vol. 19.

[2] Roush, P.B. (ed) (1987). *The Design, Sample Handling, and Applications of Infrared Microscopes*, ASTM STP 949, American Society for Testing and Materials, Philadelphia.

[3] Kirkbride, K.P. & Tungol, M.W. (1999). Infrared microspectroscopy, in *Forensic Examination of Fibers, 2nd Edition*, J. Robertson & M. Grieve, eds, Taylor & Francis, London, pp. 179–222.

[4] Sommer, A.J. & Katon, J.E. (1991). Diffraction-induced stray light in infrared microspectroscopy and its effects on spatial resolution, *Applied Spectroscopy* **45**, 1633–1640.
[5] Messerschmidt, R.G. (1995). Minimizing optical nonlinearities in infrared microspectroscopy, In *Practical Guide to Infrared Microspectroscopy*, *Practical Spectroscopy*, H.J. Humecki, ed, Marcel Dekker, New York, Vol. 19, pp. 1–39.
[6] Tahtouh, M., Despland, P., Shimmon, R., Kalman, J.R. & Reedy, B.J. (2007). The application of infrared chemical imaging to the detection and enhancement of latent fingerprints: method optimization and further findings, *Journal of Forensic Science* **52**, 1089–1096.
[7] Crane, N.J., Bartick, E.G., Schwartz Perlman, R. & Huffman, S. (2007). Infrared spectroscopic imaging for noninvasive detection of latent fingerprints, *Journal of Forensic Science* **52**, 48–53.
[8] Flynn, K., O'Leary, R., Lennard, C., Roux, C. & Reedy, B.J. (2005). Forensic applications of infrared chemical imaging: multi-layered paint chips, *Journal of Forensic Science* **50**, 832–841.

K. PAUL KIRKBRIDE

Microscopy: Hair *see* Hair: Microscopic Analysis

Microscopy: High Power

A microscope is an instrument designed to extend man's visual capability, i.e., to make visible even the minute details that cannot be seen with the naked eye [1].

A *compound* microscope is one that provides magnification in two stages by means of an *objective* and an *eyepiece*. The term is not restricted to high magnification or high power microscopes, as low power microscopes are also compound microscopes. However, when the term *compound microscope* is used, many people think of the classical transmitted light microscope such as that shown in Figure 1. Today such an instrument can be used in a wide variety of forms, including inverted microscopes and with numerous forms of transmitted and epi-illumination allowing the observation of surface features and the ability to "see through" the specimen. In order to obtain successful outcomes, sample preparation is usually required. This can be complex and involve making thin sections (or cell preparations) and using a variety of chemical staining techniques.

It is beyond the scope of this article to explain the optical theory underlying image formation in a compound microscope. The reader is referred to [2] for the more detailed treatment of this topic. The end result is that the observer normally sees a magnified, virtual, and inverted image.

The term *high power* refers to the ability to visualize the specimen at overall magnifications of up to $1,000\times$. This is achieved with objective lens magnification of up to $100\times$ and an eyepiece magnification of generally $10\times$. The normal total magnification is not necessarily the *effective* magnification as the latter depends on the microscope being properly set up to achieve the full *resolving power* of the objective.

The RMS Dictionary defines resolving power as "the ability to make points or lines which are closely adjacent in an object distinguishable in an image. High resolving power implies that the resolved distance is small".

In order to achieve maximum resolution, the full numerical aperture (NA) of lens systems has to be achieved. Figure 2 shows the concept of NA. Briefly, if the maximum possible resolving power of an objective is to be achieved, then the maximum cone of light entering the objective must be achieved. This requires that the sub stage *condenser* is properly focused.

The highest possible NA can be achieved only with an oil immersion objective for the reasons shown in Figure 2. Maximum NA depends on the microscope being properly set up which includes correct illumination of the object. Two forms of illumination to be commonly used are called, *critical illumination* and *Köhler illumination*, where the latter is the preferred method of illumination. Assuming the microscope is correctly set up, then NA critically depends on the type and quality of objective lens. Typically, objective lenses are complex multilens systems with correction for spherical and chromatic aberrations. A typical nonresearch instrument has *achromat* lenses. These are corrected for red and blue color aberration. *Apochromatic* lenses also are corrected for blue–violet color aberration. It is also possible to correct for field curvature to provide a flat

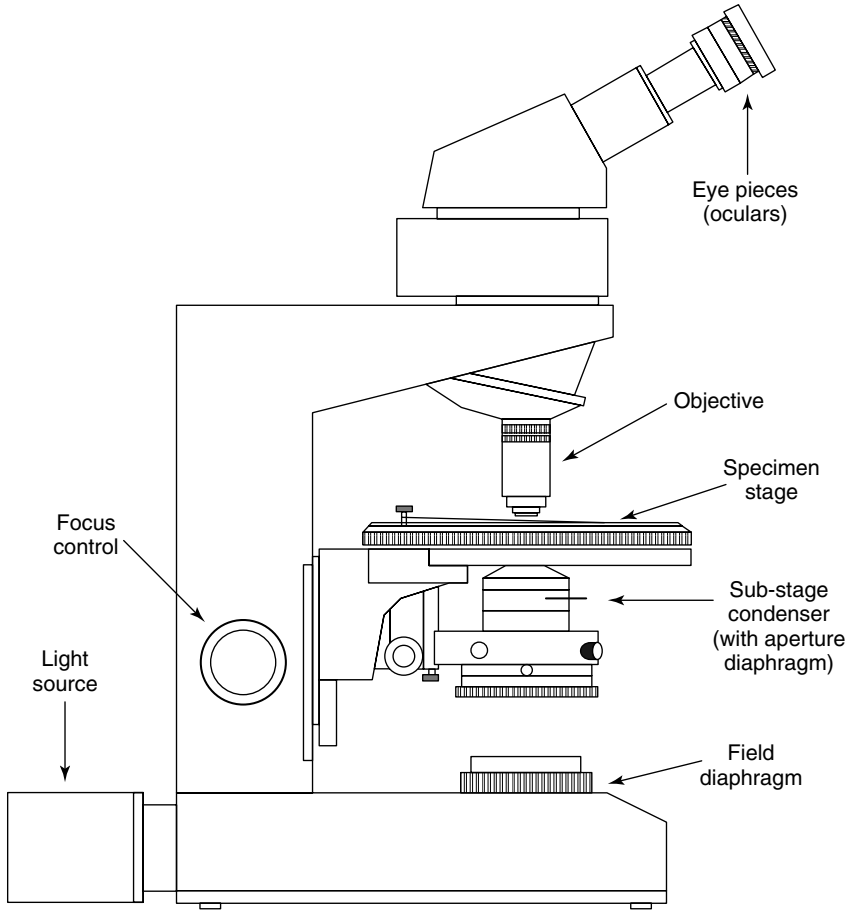


Figure 1 Main components of a modern compound microscope

field. Such lenses are called *plan*. Hence, in a high grade research compound microscope, the objectives are plan apochromat lenses with an NA in the order of 1.00 for 40× and 1.30 for 100× oil immersion. A typical achromat, dry objective at 40× would have an NA of 0.60–0.65.

The substage condenser also has a multilens system. Low cost microscopes have a two lens Abbe type condenser and higher cost microscopes have a more complex aplanatic condenser which is corrected for spherical aberration. To achieve objective NA of 1.4, a fully corrected oil immersion achromatic type condenser is required.

Finally, several types of *eyepiece* are also available. Typically today, a *compensating* eyepiece is found in which the eyepiece is designed to correct for lateral chromatic error.

Modern microscopes are designed and manufactured for ease of use and have features such as *parcentration* (ensuring that as objectives are changed the center of the visual field remains the same) and *parfocality* (ensuring focusing can be achieved as objectives are changed with a minimum of fine adjustment).

In most modern microscopes, the illuminating system is built into the base of the microscope. In low cost microscopes this is usually a precentred halogen bulb. In more expensive microscopes, a tungsten light source is common with the capability to center the light sources.

The standard set up for routine high power microscopic examination is *brightfield* microscopy in which direct light passes through the specimen, enters the objective, and illuminates the background

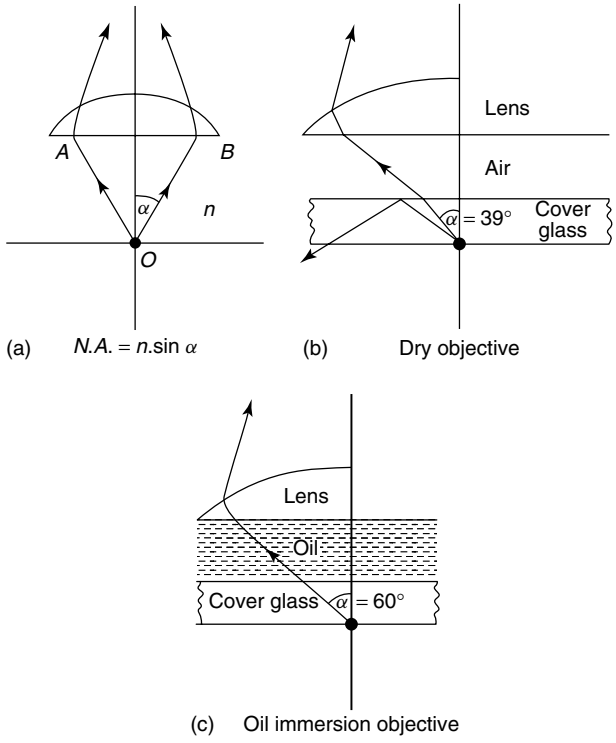


Figure 2 (a) A diagram to represent the concept of numerical aperture (NA), n is the refractive index of the medium (usually air) between the object O and the lens, α is half the angle of acceptance AOB. (b) Path of light rays in a dry objective of NA 0.95. The maximum value of α is 39° due to refraction at the air/cover glass interface. Rays with a greater obliquity are totally reflected at the surface of the cover glass. (c) Path of light rays in an oil immersion objective of NA 1.3. The value of α is now increased to 60° since there is now no air/glass interface refraction [Reproduced with permission from Edward Arnold Ltd. © 1976 [3].]

against which the image is seen. In order to visualize the image, there has to be a difference in the refractive index (RI) of the specimen and the mountant in which the specimen is placed; usually covered by a glass cover slip. For example, the measurement of the RI of glass is based on using a mountant whose RI changes as it is heated; for hair microscopy, the choice of mountant is critical if the internal structure of the hair is to be seen. For some samples, it can help to increase contrast by closing the iris diaphragm in the sub stage condenser. However, for very transparent specimens, there are a number of specialist techniques such as *dark field* microscopy and *phase contrast* microscopy, which provide images with a high degree of contrast. These techniques, especially phase contrast, are widely applied in biology but are not frequently used in routine forensic work.

Other specialist forms of microscopy include *fluorescence* microscopy and *polarizing* microscopy. Epi or incident fluorescence is commonly used in the forensic examination of fibers. Transmission fluorescence has been applied in aspects of forensic biology. It is beyond the scope of this article to describe the theory of polarizing microscopy. In brief, it makes use of the *optical* properties of some samples, (anisotropic) to "twist" polarized light. Casartelli [4] has described in simple terms how a polarized light microscope works. Although there are differing views within the forensic community as to the role of polarizing microscopy, there is no doubt that polarizing microscopy is still a valuable technique adding an analytical dimension to the microscope. The McCrone Particle Atlas [5] gives an insight into the extremely wide range of possible applications of this technique for forensic work.

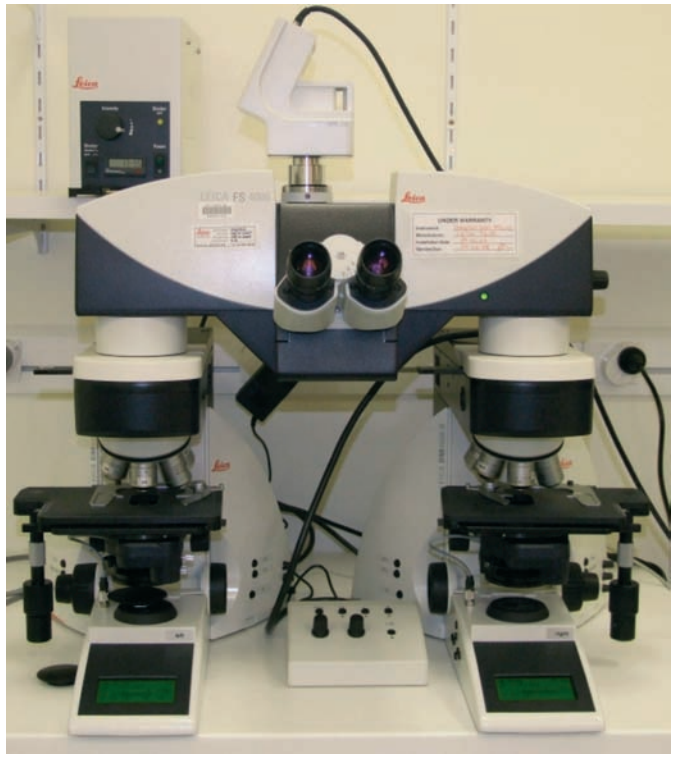


Figure 3 Comparison microscope

Finally, the forensic world makes use of comparison microscopy (see Figure 3) in which two microscope systems are linked by a system of prisms to present their images into a single binocular comparison eyepiece such that two images can be directly compared side by side. This technique is used in the forensic examination of fiber and hairs. Most of the current modern microscopes have the capability of being linked to image capture systems.

In conclusion, the compound light microscope is a versatile and indispensable piece of equipment. As a general recommendation, "simple is best". A less expensive microscope with limited options is often fit for purpose and is more preferred for routine use. This is especially the case where very little formal training is offered in basic microscopy. To achieve even close to the theoretical performance of any microscope, it must be kept clean and regularly serviced and maintained. Add to this an understanding of the basic theory of image formation, and how to use a microscope, and there is no reason why anyone cannot make effective use of microscopy.

References

[1] Anon (1989). *RMS Dictionary of Light Microscopy*, Oxford University Press – Royal Microscopical Society.

[2] De Forest, P.R. (2002). Foundation of forensic microscopy, in *Forensic Science Handbook*, 2nd Edition, R. Saferstein, ed, Prentice Hall, New Jersey, Vol. 1, Chapter 5, pp. 215–319.

[3] Bradbury, S.L. (1976). *The Optical Microscope in Biology*, Arnold, London.

[4] Casartelli, J.D. (1969). *Microscopy for Students*, 2nd Edition, McGraw Hill, London.

[5] McCrone, W.C., Draftz, R.G. & Delly, J.G. (1967). *The Particle Atlas*, Ann Arbor Science Publishers.

Further Reading

Houck, M.M., Bowen, R. An argument for light microscopy – a review of forensic microscopy for trace evidence analysis. *Forensic Science Review* **17**, 1–15. 2005.

JAMES ROBERTSON

Microscopy: Light Microscopes

Introduction

This article is intended to cover the fundamentals of light microscopy. A microscopical examination may be the only instrumental technique needed to identify a piece of evidence, provide direction for further analysis, or compliment other instrumental techniques.

The microscope is used in many forensic disciplines including firearms identification, serological examinations, drug chemistry, and trace evidence. It is in the discipline of trace evidence where the light microscope has found its greatest use in the characterization, identification, and comparison of particulate material. The trace evidence analyst may encounter limitless types of particulate matter including fibers, hairs, paint, explosives, pollen, soil, glass, and tape. The use of the light microscope provides valuable information in the examination of trace particulate and its exclusion in any examination would be inconceivable.

Stereomicroscopy

Stereomicroscopes, sometimes referred to as *dissecting microscopes*, are microscopes that produce a magnified image of a specimen that is upright, laterally correct and with a perception of depth – exactly the way the eye sees in normal vision. The stereomicroscope is of particular importance in forensic examinations because the 3-D image, the extended depth of field, and long working distance facilitate inspection, sorting, preparation, and dissection of evidence. In many cases, these instruments can also provide enough information to identify the sample.

Design

In normal vision or when using a stereomicroscope, the brain combines two off-axis images to generate one image that is not only two dimensional but also has the perception of depth. There are two types of stereomicroscopes produced today – the *Greenough* design and the *Common Main Objective* design.

Greenough Stereomicroscope. The Greenough stereomicroscope dates from the later part of the nineteenth century and uses two completely independent optical axes that are set approximately 14° apart (Figure 1). Because the left and right optical paths must maintain critical alignment throughout the entire optical path, it is impossible for the user to insert and remove accessory items such as coaxial illuminators, epi-fluorescence or add a port for documentation. However, it is possible to produce highly corrected optics, including apochromatic lenses, at relatively low cost. Many laboratory-grade stereomicroscopes are Greenough instruments and are quite adequate for examinations and sorting of evidence.

Common Main Objective Stereomicroscope. The common main objective (CMO) stereomicroscope features two parallel optical axes (one for each eye) from the eyepieces down through the main body. At the common main objective, each optical path turns inward to view the specimen at slightly different angles (Figure 2). This type of microscope also produces a stereoscopic 3-D image, but with an important advantage over the Greenough design.



Figure 1 A modern day Greenough stereomicroscope



Figure 2 A modern day common main objective stereomicroscope

Since the optical paths are parallel, the image in a CMO instrument that is above the main body and below the observation tube is projected at infinity. Thus, the CMO stereomicroscope can accommodate a wide range of accessories, which can be easily inserted and removed by the user. These accessories include multiple photo/video ports, coaxial illuminators, and epi-fluorescence illuminators, often in combinations. Additionally, the observation tube that contains the eyepieces can be easily interchanged to accept other special purpose observation tubes that have eyepieces set at unique angles, eyepiece tubes that tilt for ergonomic advantages, or eyepiece tubes with long extensions to increase the distance of the observer from the specimen.

Magnification and Resolution

The total magnification of a stereomicroscope is a function of the eyepiece magnification, the main objective (if a CMO instrument) or the supplementary objective (if a Greenough instrument), the magnification factor of any intermediate accessory, and the magnification setting of the main body of the microscope. It is possible with some CMO stereomicroscopes to extend the total visual magnification to nearly 1000X, although the resolving power of any stereomicroscope is limited primarily to the numerical aperture of the main objective, which is at most approximately 0.2 and yields a maximum useful visual magnification of about 200X. Refer to the section "Brightfield" for a more detailed discussion on magnification, numerical aperture, and resolution.

Accessories

The accessories for CMO stereomicroscopes are extensive and allow these instruments to examine the internal structures of evidence in transmitted light and their surface details in incident light. Illumination sources include fiber optic bundles and ringlights as well as fixed and adjustable LED sources. Components for polarization, darkfield, epi-fluorescence, and digital documentation may be added to expand the functionality. Stereomicroscopes may be used with compact incident light stands for the examination of small pieces of evidence or they may be attached to elaborate heavy-duty stands for the detailed inspection of clothing or other large objects such as an automobile.

Limitations of the Stereomicroscope

Because the image projected through either side of a stereomicroscope is not perpendicular to the specimen plane, the left and right edges of a flat sample are out of focus when the center is in sharp focus. This is generally not a noticeable problem at lower magnifications and when the instrument is used for observation of samples that have a significant amount of surface relief. However, when the stereomicroscope is used for video discussion or documentation, only one side of the stereo image is diverted to the camera. Since either side is off-axis, accommodations must be made for this nonperpendicular view. With CMO instruments, the major manufacturers offer carriers that shift the position of one side of the microscope over the center of the main objective. The view obtained with this uniaxial instrument is perpendicular to the specimen plane and uses the center portion of the main objective which offers more optical correction than the lens periphery. This uniaxial position also provides the most accuracy when an eyepiece micrometer or computer software is used for measuring.

Brightfield Microscopy

A brightfield (biological) microscope with ordinary (unpolarized) illumination can be utilized to characterize, identify, and compare many types of particles encountered by the trace evidence microscopist. Although we see later that a polarized light microscope will give additional properties one can use to characterize and identify particles, unpolarized light may be employed to characterize physical features and to determine average optical properties of particles. Hairs, diatoms, and pollen grains are a few examples of particles that can be identified solely based on their physical characteristics when utilizing brightfield microscopy. Also, unpolarized light can be used to determine the refractive index (RI) and the color of particles that have a single RI (isotropic) as well as determine the average RI and the average color of particles that have more than one RI (anisotropic).

Concepts in the Design of the Light Microscope

To effectively use a microscope, a microscopist must know how to configure and adjust a microscope plus have a basic understanding of light and how it interacts with matter. The following discussion of microscopy is intended to be a simple primer that provides an introduction to light and the components of a contemporary compound microscope.

Light. The portion of the electromagnetic spectrum that the human eye detects as light is the range of wavelengths of energy from approximately 380 to 750 nm. The eye recognizes the shorter wavelengths of light as the colors violet to blue, the intermediate wavelengths as the colors green to yellow, and the longer wavelengths as the colors orange to red. Furthermore, the human eye interprets an increase in the amplitude of this wave as an increase in intensity or brightness of the light (Figure 3).

The interaction of light with matter provides specific information that can be used to characterize and identify particles during a microscopical examination. Also, the interaction of light with the components of a microscope directly affects the quality of the information revealed in the resulting magnified image. The principal interactions of light with matter which are discussed here are reflection, refraction, and dispersion.

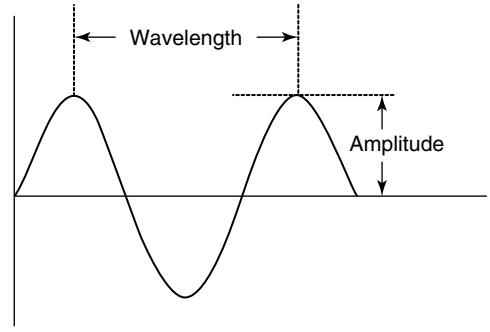


Figure 3 Diagram Figure showing wavelength and amplitude for a ray of light

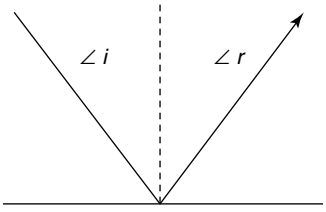


Figure 4 Specular reflection

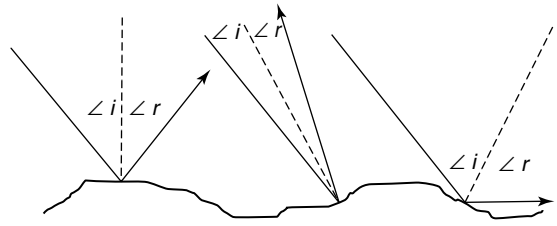


Figure 5 Diffuse reflection

Reflection. Specular reflection occurs when light strikes a smooth surface (Figure 4). The angle at which the light leaves the surface is equal to the angle at which it arrives and, therefore, the angle of reflection (r) is equal to the angle of incidence (i).

Diffuse reflection occurs when light strikes an irregular surface and the rays are reflected in different directions. Diffuse reflection still obeys the laws of reflection as the angle of reflection is still equal to the angle of incidence (Figure 5). How light is reflected off a material using incident light can provide valuable information when characterizing and comparing samples.

Refraction. Refractive index (n) is the ratio of the speed of light in a vacuum to the speed of light in a transparent medium. When light enters a medium having a different RI at any angle other than perpendicular to the surface, the light will not only change velocity but also direction. This change in direction, or bending of light, is known as *refraction*. Refraction is governed by Snell's law that states

$$n_r/n_i = \sin i/\sin r \qquad (1)$$

where n_r is the RI of the refractive medium, n_i is the RI of the incident medium, sin i is the sine of the angle of incidence, and sin r is the sine of the refraction.

Therefore, as light passes from a medium of lower to one of a higher RI, the incident ray of light will be refracted toward the normal to the interface then away from the normal to the interface as it passes from a higher to lower RI medium (Figure 6). Furthermore, as the difference between the two refractive indices increases, the amount the light ray is refracted also increases. The refraction of light as it passes through the optical components of a microscope is an essential function of the microscope lenses in the formation of a magnified image.

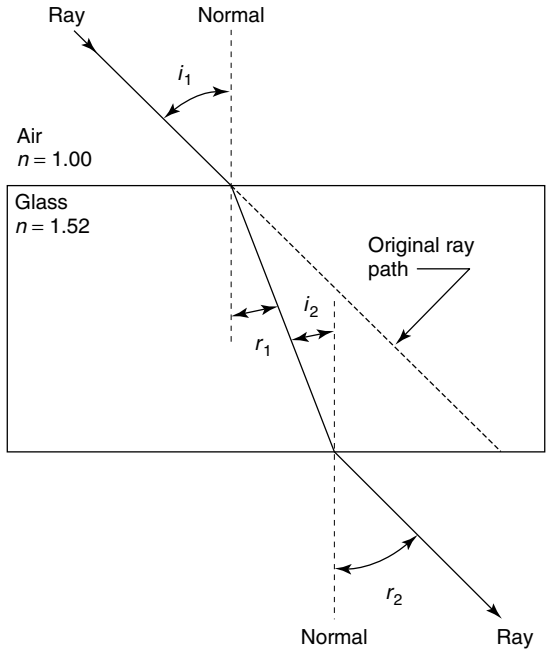


Figure 6 Refraction of light

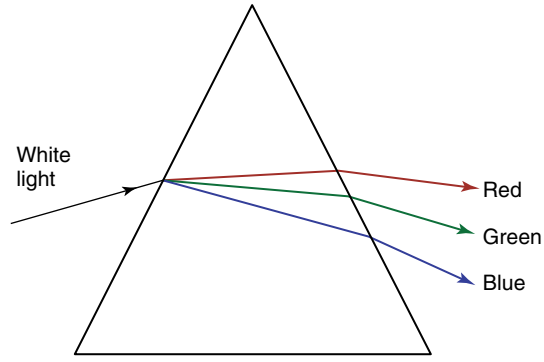


Figure 7 Dispersion of white light through a glass prism

Dispersion. Dispersion is the change in RI of a transparent substance as a function of the wavelength of light resulting in the separation of white light into its component wavelengths or colors (Figure 7). The RI is usually lower for longer wavelengths (red) and higher for shorter wavelengths (blue). The dispersion of white light as it passes through the optical elements of a microscope results in a degraded image quality and requires correction or compensation for this aberration.

Magnification. A simple magnifier helps the eye focus an object that is closer than 10 in (25 cm), which is the normal reading distance for the eye. When the object is moved closer to the eye, its visual angle is increased and the object can be focused by the eye when a positive lens is placed between the eye and the object (Figure 8). Magnification can be expressed in several ways but one simple relationship is:

$$\text{Magnification} = \text{image size}/\text{object size} \qquad (2)$$

A compound microscope uses an objective lens positioned close to the specimen to produce a magnified real image of the object. The eyepiece functions as a simple magnifier and further magnifies the real image produced by the objective. Therefore, the final magnification to the observer is the product of the objective magnification, the eyepiece magnification, and any intermediate tube factor that may be present in the microscope.

Objectives. The objective is the lens system nearest the specimen and can have a series of complex lens

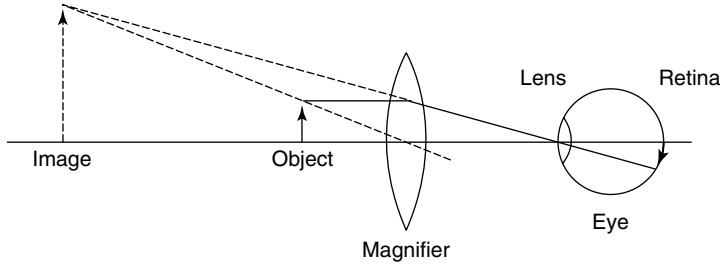


Figure 8 Diagram of simple magnification

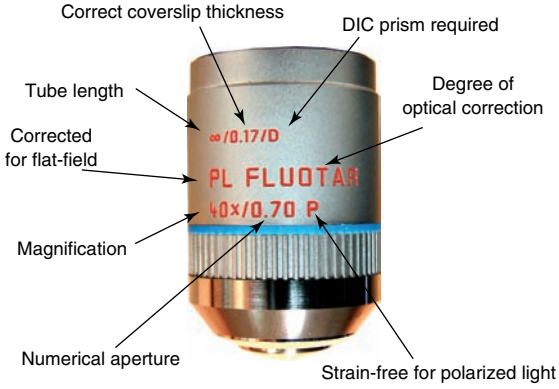


Figure 9 Objective nomenclature

elements. The objective provides the primary magnification and in part defines the resolution of the final image. Objectives are designed as having different magnifications and numerical apertures. The resolution of the microscope, which is the ability to discern fine structure or minute particles spaced closely together, is a function of the objective and the condenser. The nomenclature engraved on the body of the objective provides useful information regarding magnification, numerical aperture, and intended usage (Figure 9). For example, a 0.17 value means that the objective must be used with a coverglass of 0.17 mm thickness or the image will not be as sharp due to spherical aberration.

Eyepieces. The eyepiece provides a magnified secondary image of the primary image produced by the objective. Eyepieces having different magnifications and field-of-view indexes are available. The field-of-view index of the microscope eyepiece is the diameter in millimeters of an internal diaphragm which is used to match the level of correction of the objectives being used. Eyepieces with wide field-of-view indexes require objectives with more extensive correction for flatness of field or the periphery of the image will be out of focus. The 10X magnification eyepiece is the most widely used eyepiece and is available with a field-of-view index from 18 to 25.

Condenser. The primary function of the substage condenser in a transmitted light microscope is to concentrate and focus the light evenly across the field of view. Laboratory and research grade microscopes must be able to achieve Köhler illumination which requires an adjustable condenser. Condensers for brightfield observation techniques may span the entire magnification range from a 1.25X objective to a 100X objective and have a mechanism that permits a top condensing lens to be swung in the light path for objective magnifications of 10X and above. This dual function allows the condenser to evenly illuminate large fields at low magnification and smaller fields at higher magnification.

Special use condensers for phase contrast, polarization, darkfield, and differential interference contrast (DIC) provide the necessary components to achieve these special contrasting techniques. Since the highest numerical aperture possible with a dry condensing system is 0.90, condensers are available that allow immersion oil to be placed on the underside of the specimen slide to couple the condenser optically to the specimen allowing condenser numerical apertures up to 1.4.

Illuminators. Most microscopes today use built-in or directly coupled illumination sources, often with a power supply providing variable intensity control that is integrated into the microscope stand. The primary illumination source for transmitted light

is a low-voltage tungsten halogen bulb with high intensity wattages from 20 to 100. Light emitting diodes (LEDs) are being used in some microscopes because they produce a bright white light with an exceptionally long life.

Although a routine bench microscope may not have provisions for adjustment of the filament, the light produced is sufficient to properly illuminate the specimen and render high quality images. Research level instruments have the lamp located in an attached housing that provides for critically positioning the filament in the center of the optical axis, independently controlling a secondary image of the filament produced by a reflecting mirror behind the bulb, and precisely focusing the filament for Köhler illumination.

Numerical Aperture. As the focal length of the front element of an objective becomes shorter (closer to the object) as is required to increase the magnification of the objective, the angle of acceptance of the image forming rays (angular aperture) is increased (Figure 10). The numerical aperture is a measure of the light gathering ability of a lens system and determines the resolving power and depth of field. Generally, objectives with the shortest working distance have the highest magnification and the greatest angular apertures (AA). The numerical aperture can be calculated using the formula:

$$NA = n \sin \frac{AA}{2} \tag{3}$$

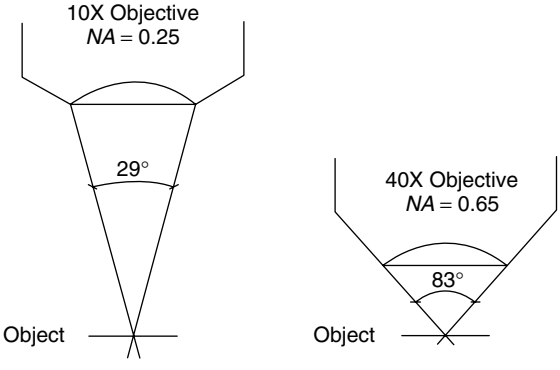


Figure 10 Diagram showing acceptance angle for two objectives having a different magnification and numerical aperture

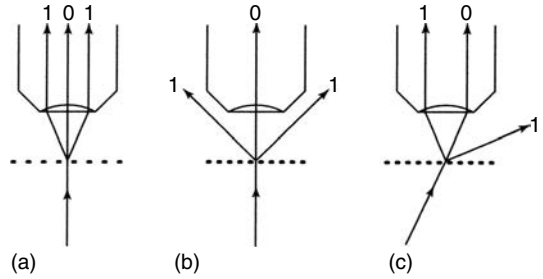


Figure 11 Diffraction of light through fine structure

where n is the RI of the space between the front element of the objective (or top element of the condenser) and the microscope slide.

Diffraction. In the late nineteenth century, Ernst Abbe demonstrated that there is a correlation among diffraction, numerical aperture, resolution, and the RI of the medium between the specimen and the front lens of the objective. A series of diffracted rays are produced when light interacts with a specimen and as the detail in the specimen becomes finer, the angle of the diffracted rays increase (Figure 11). To resolve detail in a specimen, one needs to capture the direct ray and at least one of the first-order diffracted rays. However, the image quality improves as more of the diffracted rays are captured. Also, the shorter wavelengths of light (blue) produce a smaller diffraction angle than the longer wavelengths of light (red) resulting in greater resolution for shorter wavelengths.

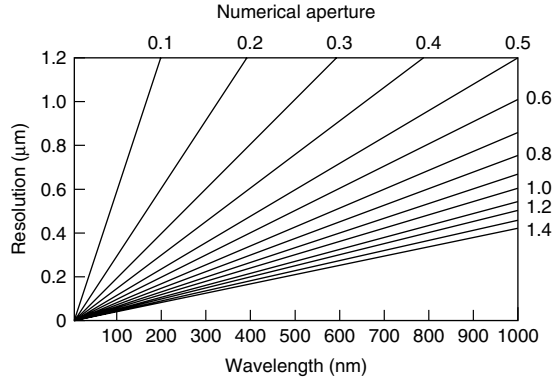


Figure 12 Diagram showing the relationship among wavelength, numerical aperture, and resolution

Resolving Power. The numerical aperture of an objective and the condenser along with the wavelength of light used to illuminate the sample determines the resolving power of the microscope. Resolving power is the ability of an optical system to make two closely spaced points (fine detail) visible as separate entities. Therefore, a resolving power of 1 μm means that two point-shaped particles at a 1 μm distance from each other will just be distinguishable as separate objects. A graph depicting the relationship among wavelength, numerical aperture, and resolving power is provided in Figure 12.

Contrast. If a magnified image is to be of value, the details of the structure of the specimen must be observable. While a microscope may be resolving the fine details of a specimen, these details are of no value unless there is sufficient contrast for the eye to detect the structure. Contrast may be enhanced by mounting the sample in a medium with a different RI, by the use of colored stains, or by simply closing down the aperture diaphragm. However, as the aperture diaphragm is closed down, the resolution significantly decreases. It is a good practice to observe specimens with the aperture diaphragm in different positions. Also, optical or illumination contrasting techniques such as darkfield, Rheinberg, phase contrast, differential interference contrast (DIC), or modulation contrast can be employed. It is the appropriate combination of resolution, contrast, and magnification that provides the microscopist with an image with sufficient detail to provide useful information.

Finite and Infinity Microscopes. The two designs for compound microscopes are finite tube length and infinity corrected tube length (Figure 13). While both systems can produce high-quality images, the infinity corrected system is now the most commonly available optical design for laboratory and research grade instruments. A microscope with infinity corrected optics allows for accessory components to be added such as incident light illuminators for opaque specimens or epi-fluorescence illumination, polarizers, and compensators without degrading the image. Many contemporary bench microscopes and all older instruments are of the finite tube length design.

Objective Aberrations. A perfect converging lens causes all of the light rays emitting from a point on one side of a lens to be converged (focused) on the

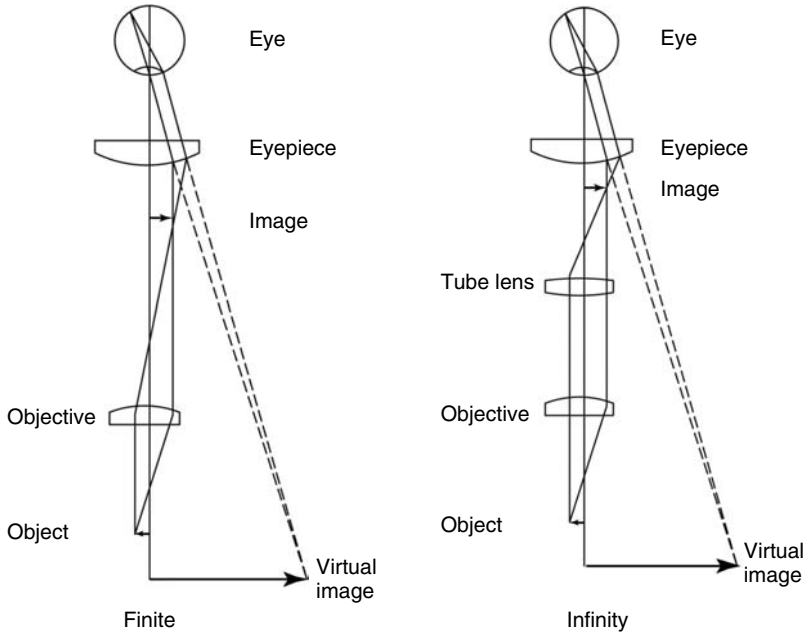


Figure 13 Ray path through a finite corrected microscope (left) and an infinity corrected microscope (right)

other side of the lens and produces an exact image of the original point. In practice, all lenses suffer from aberrations and, in fact, these imperfections impair the ability of the lens to produce that exact copy of an object. Of the six primary aberrations, spherical and chromatic affect the entire field while coma, astigmatism, curvature of field, and distortion affect the quality of the off-axis or peripheral image. To a degree, all of these aberrations can be reduced to a level at which they no longer adversely interfere with visual observation. The extent to which a microscope is corrected for these aberrations determines its performance and subsequent cost.

Particle Characterization – Unpolarized Light

Before any microscopical examination can begin, one must ensure that the microscope is properly aligned and adjusted for Köhler illumination. Depending on the microscope being utilized, the centering and focusing of the lamp filament may not be possible to obtain true Köhler illumination. This is because many manufacturers have eliminated the ability to center and focus the lamp filament and have introduced a diffuser into the light path. Once Köhler illumination is achieved, particles are characterized and identified based on the properties that are observed and measured. This is first accomplished using unpolarized light or with plane polarized light utilizing a single polar and finally with crossed (two) polars.

Sample Preparation. Samples are prepared for microscopical examination by placing them between a microscope slide and coverglass (coverslip) and using a liquid (mounting medium) of choice. The mounting medium may be a temporary medium that will evaporate away such as water or a suitable organic liquid. Xylene was commonly used in past years as a temporary medium, but its use today has been restricted in some laboratories. A stain may also be used as mounting medium. Some common stains that may be used include: Graff C stain or Selliger's stain to characterize and identify cellulose fibers; iodine solution or Betadine solution to detect starch grains and gelatinized starch particles; and malachite green in nitrobenzene to detect and identify clay particles.

A standard oil having a known RI is another mounting medium that can be used, such as Cargille refractive index liquids (Cargille Laboratories, Cedar Grove, NJ, USA). A standard oil having a RI of 1.66 is usually employed for general particle characterization and for the examination of the heavy mineral fraction (density $> 2.89\,\mathrm{g\,cc^{-1}}$) in a soil sample. A liquid having a RI = 1.550 is usually favored to characterize and identify minerals in the light fraction (density $< 2.89\,\mathrm{g\,cc^{-1}}$). Many fiber microscopists prefer a liquid having a RI = 1.525.

There are many permanent mounting media to choose from including Permount™, Meltmount™, Entellan™, and Norland Optical Adhesive™. Aroclor™ has been used in the past years as a mounting medium but is not readily available today. As can be seen, there are numerous mounting media to choose from and which mounting medium one will use greatly depends on the sample and the information one wishes to obtain from the examination.

Particle Morphology. Under examination, the first characteristic one will notice is the physical features of the particles. General terms should be used to describe particles unless one understands precise descriptions that are used in specialized fields (crystallography, wood taxonomy, mineralogy, paleontology, etc.) and can be applied correctly. Also, the microscopist will project a more visual image of the particle observed under the microscope using descriptive terms such as irregular, spherical, acicular, granular, platy, fibrous, or tabular. Also, interfacial angles between crystal faces can be measured if present.

Particle Size. The next characteristic one may determine is particle size. The size of a particle may help to distinguish between two particles that have similar morphologies such as corn starch and rice starch, the former having an average diameter twice that of latter. Also, particle size is an important consideration when comparing particles contained in two different samples.

Particle size can be either estimated or measured precisely. The particle size is estimated by comparing the particle to the diameter of the field of view. A 10X objective has a field of view of approximately 2.2 mm or 2200 μm, whereas a 40X objective has a field of view around 0.54 mm or 540 μm. Precise particle measurements can be made using a calibrated eyepiece scale, which is located in the front focal plane of the eyepiece and is therefore superimposed on the field of view. One can calibrate the eyepiece

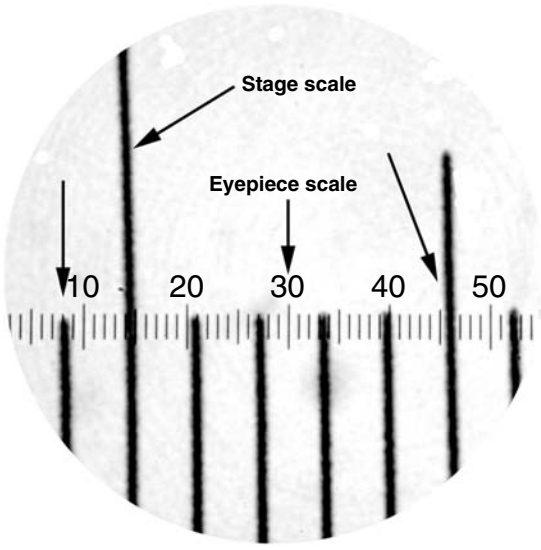


Figure 14 Image of the eyepiece scale superimposed over the stage micrometer scale

scale using a stage micrometer. For the stage micrometer shown in Figure 14, each division is equal to 10 μm. Note that 42 eyepiece divisions equals 6 stage micrometer divisions or 60 μm. Therefore, each eyepiece scale division for this objective equals 1.4 μm. Since the diameter of the field of view varies with each objective, the eyepiece scale must be calibrated for each objective so the particle size can be measured accurately using any objective.

Refractive Index. Refractive index as previously defined is the ratio of the velocity of light in a vacuum to the velocity of light passing through a transparent medium. The RI will always be greater than 1.00 and determination of the RI (or RIs) is an important aspect in the identification of transparent particles. Also, it is important to remember that the RI varies with the wavelength of light (dispersion) and it is understood that the RI value usually reported is at 589 nm (n_D) unless otherwise noted.

The degree of observed contrast (dark edges) of a particle is a measure of the RI of the particles relative to the RI of the medium. The particle contrast will increase as the difference in RI increases between the particle and the medium. If the particle has colored borders it means that the particle and medium have the same RI at some wavelength near or in the visible region of light. If the particle is "invisible" then the liquid and the particle have the same RI.

With a little practice, one can judge the difference between the RI of the particle and the RI of the liquid by the amount of contrast observed. If the contrast is low, one can estimate that the particle and the liquid have a RI difference that ranges from ±0 to 0.04. If the contrast is moderate, then the difference of the two RIs is estimated to be somewhere between ±0.04 and 0.12. If the contrast is high, the difference between the particle and liquid RIs is greater than 0.12. To "calibrate" your eye, one should estimate the amount of contrast in a consistent manner with the aperture diaphragm closed down (smallest opening).

The degree of contrast observed for a particle in a liquid does not demonstrate if the particle has a higher or lower RI than the mounting medium. Even though there are several ways to determine which has a higher RI, the Becke line technique is commonly conducted to quickly determine whether the particle or the liquid has the higher RI. The Becke line is a bright halo that is observed near the boundary of a particle which moves in and out of the particle as the particle is brought through good focus. This halo is the result of the concentration of light due to the refraction (light changing direction) at the interface of the particle and mounting medium having different refractive indices. The light will be bent toward the normal (an imaginary line perpendicular to the surface) as it enters a transparent medium of higher RI and away from the normal as it enters a medium of lower RI. As you focus above the plane of best focus (increase the distance between the objective and particle) the Becke line will move into the particle if it has a higher RI or will move into the liquid if it has a higher RI (Figure 15). The aperture diaphragm should be closed down when conducting the Becke line test and the movement of the fine focus must be slight.

Isotropic particles (noncrystalline particles and cubic crystals) have a single RI and this index will be observed whether unpolarized or plane polarized light is employed. However, anisotropic particles (crystals with more than one RI) have either two principal RIs (uniaxial) or three principal RIs (biaxial). The two principal RIs for uniaxial crystals are noted as ω and ε. For biaxial crystals, the three principal RIs are noted as α, β, and γ. Also, anisotropic fibers have two different refractive indices: RI parallel to the length ($n_{\parallel}$) and perpendicular to the length ($n_{\perp}$).

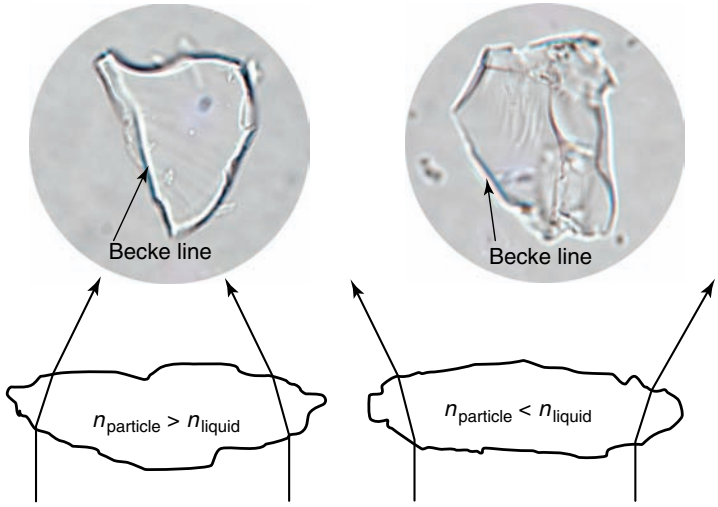


Figure 15 Ray diagram and images showing the Becke line slightly above best focus for a particle having a higher refractive index than the mounting medium (left) and for a particle having a refractive index lower than the mounting medium (right)

With unpolarized light we will observe an average RI (n_{iso}) for anisotropic particles. For most anisotropic crystals, the average RI can be calculated using the formula:

$$n_{iso} = \frac{1}{3}(2\omega + \varepsilon) \quad \text{or} \quad n_{iso} = \frac{1}{3}(\alpha + \beta + \gamma) \quad (4)$$

And for anisotropic fibers:

$$n_{iso} = \frac{1}{3}(n_{ll} + 2n_{\perp}) \quad (5)$$

One advantage of looking at the average RI (n_{iso}) of anisotropic particles is that the average RI will be observed regardless of the orientation of the particle. Determining and comparing the average RI of anisotropic particles is a useful technique when characterizing a sample containing different types of textile fibers as well as for the detection of trace components in a sample. A list of n_{iso} values for fibers is provided by Gaudette in "The Forensic Aspects of Textile Fiber Examination". If a particle/fiber of interest is found by using unpolarized light, polarized light (plane or crossed) can be quickly inserted into the light path so additional optical properties can be determined.

Dispersion Staining. Dispersion staining is a powerful technique that one can employ to find the RI relationship between transparent particles and the mounting medium. Dispersion staining colors are observed when a particle and a liquid have different dispersion curves (plot of RI with wavelength) but have a common RI. This technique is useful for the identification of particles where the RI data is known and is especially useful for the detection of a component in a "needle-in-a-haystack" type of examination.

The dispersion staining technique employs an opaque annular stop or central stop located in the back focal plane of the objective (Figure 16). This stop enhances the color that is observed at the border of the particle and liquid. The color of the particle at the border will depend on where the liquid and the particle have the same RI. Wavelengths of light at the intersection of the dispersion curves for the particle and liquid will not be deviated and will pass through the opening of the annular stop but will be blocked if employing the central stop. The other wavelengths of light will be deviated and be blocked by the annular stop but will pass through the opening of the central stop if employed. The degree of deviation of the wavelengths of light will depend on the differences in the dispersion curves of the particle and liquid. As one would expect, the dispersion staining colors become more vivid if the particle and liquid have very different dispersion curves (Figure 17).

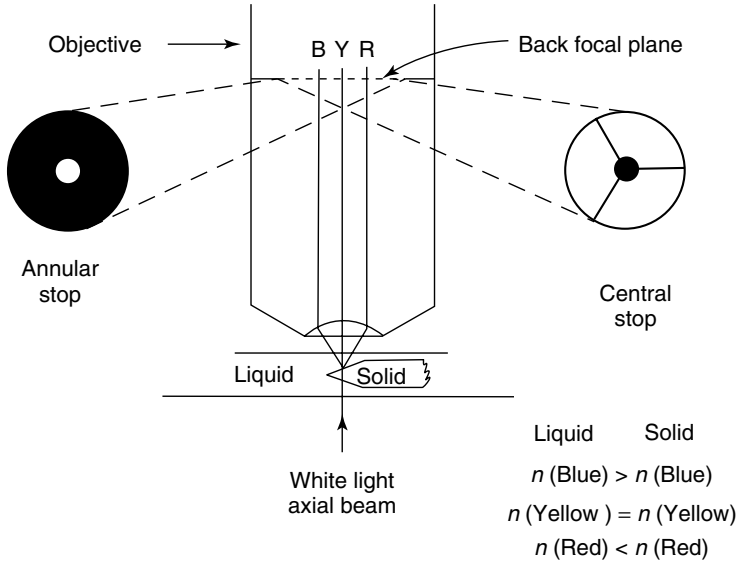


Figure 16 Diagram of the dispersion staining objective with the annular and central stop [Reproduced with permission from McCrone Research Institute.]

Table 1 Dispersion staining colors. λ_0 notes where the particle and the liquid have the same RI[(a)]

		Central stop colors	
Matching λ_0 (nm)	Annular stop colors	In focus	Becke line
<420	Blue–black	Light yellow	Faint gold + violet
430	Blue–violet	Yellow	Faint gold + violet
455	Blue	Golden yellow	Faint gold + violet
485	Blue–green	Golden–magenta	Yellow + violet
520	Green	Red–magenta	Violet + orange
560	Yellow–green	Magenta	Blue violet + red orange
595	Yellow	Blue–magenta	Blue + red
625	Orange	Blue	Blue
660	Orange–red	Blue–green	Green
>680	Brown–red	Pale blue	Pale green

[(a)]Reproduced with permission from McCrone Research Institute

The central stop method is generally preferred since the complementary dispersion staining colors are observed on a darkfield. Table 1 shows the colors observed at various matching wavelengths using the annular stop and central stop.

One can plot the dispersion curve for a particle by plotting the matching RI observed in several different liquids (Figure 18). To accomplish this, one must know the dispersion data for the liquids. This plot will provide the RI not only at n_D but also at n_F (486 nm) and n_C (656 nm), the other two refractive indices commonly listed in literature.

Alternately, one can plot a dispersion staining curve in a series of liquids and obtain the n_D value for a particle without knowing the dispersion data for the liquids. This is accomplished by plotting a horizontal line at the n_D value for the different liquids (Figure 19). However, only the RI at 589 nm (n_D) has any direct RI value to literature values. The RIs at the other wavelengths along the dispersion staining

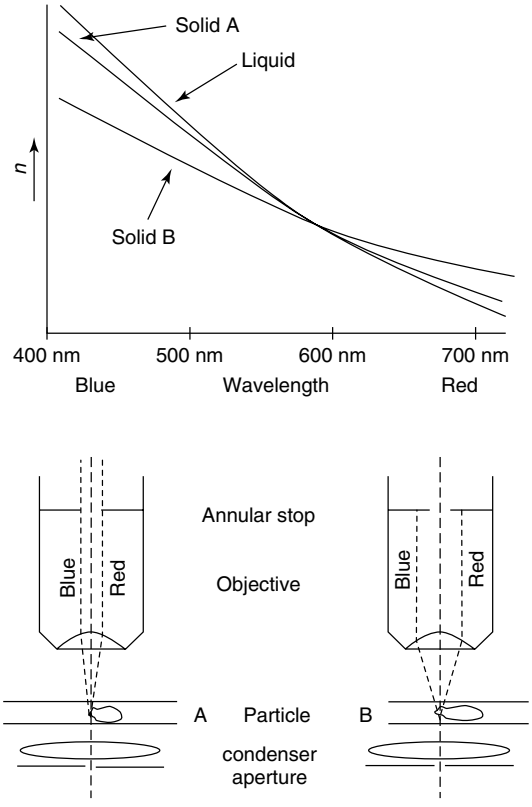


Figure 17 Ray paths through a dispersion staining objective using an annular stop corresponding to the dispersion curves for solid A and for solid B mounted in the same liquid

curve do not provide the actual RI of the particle. Therefore, to determine the actual RI for the particle at any wavelength other than n_D (589 nm), one needs to replot the dispersion staining curve using the actual known dispersion data for the liquids. In addition, dispersion curves are generally plotted with a linear wavelength scale whereas for dispersion staining curves the wavelength scale is plotted as $1/\lambda^2$. Plotting the wavelength as $1/\lambda^2$ generally provides a straighter line for the dispersion staining curve. For more information on dispersion staining, one should consult the *Polarized Light Microscopy* manual by McCrone, McCrone, and Delly.

Other Characteristics. If present, other characteristics may be noted including (but not limited to) color, transparency, and surface features. The color of the particle may be the first feature noted and can be observed with transmitted and/or reflected light. The color of a particle is the result of selective absorption of certain wavelengths of light. Only the average transmission color will be observed with unpolarized light for transparent particles with more than one RI. It can be seen later that the use of plane polarized light (single polar) will provide additional information for colored particles which have more than one RI.

Some particles may not transmit all the illuminating light and are translucent rather than transparent. This translucent characteristic may be due to optical discontinuities within the particle which scatter light through refraction and/or reflection as light travels through the particle. These optical discontinuities may be the result of different orientation of internal grain structure, impurities, or inclusions such as air bubbles. It may also be the result of surface features which are more apparent when the RI of the particle and liquid differ greatly.

The surface of the particle is another important feature used to characterize and compare particles. Transmitted light is used to observe the surface of a transparent or translucent particle by focusing on the top and bottom surfaces. However, with opaque particles, a reflected light illuminator is employed. General descriptive terms may include smooth, dimpled, fractured, rough, pitted, and nodular.

Polarized Light Microscopy

The polarized light microscope (PLM) is a powerful analytical instrument for the forensic trace analyst. The polarizing microscope can be applied to almost any type of trace evidence and will increase the number of identification characteristics one can determine for a particle under investigation. The PLM differs from a biological microscope in that it has a polarizer (polar) located in the substage condenser, an analyzer (second polar) located above the stage, a circular stage, strain free objectives, a crossline eyepiece, and a slot for inserting compensators in the body tube. The most common compensator is the full-wave (530 nm) compensator, but it is desirable to have a quarter-wave (137 nm) compensator and a variable compensator. Also, the stage and/or objectives should be centerable.

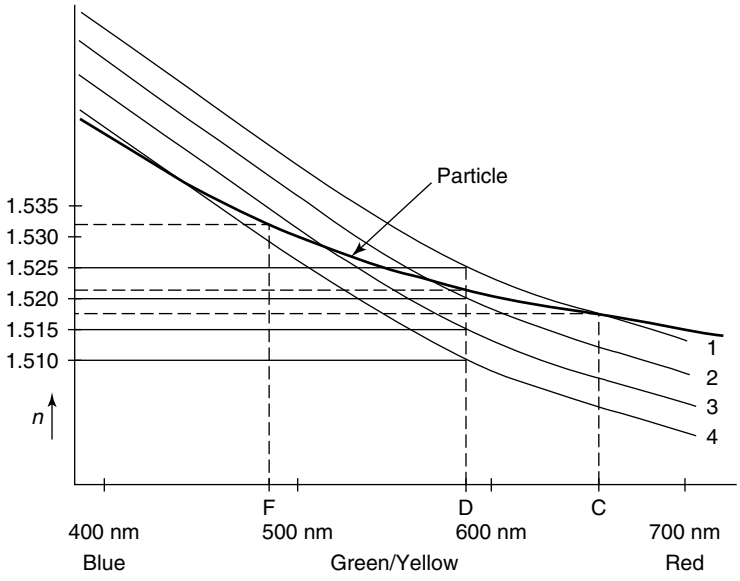


Figure 18 Dispersion curve plot for a particle mounted in four different liquids. The refractive index values are $n_F = 1.532$, $n_D = 1.522$, and $n_C = 1.517$

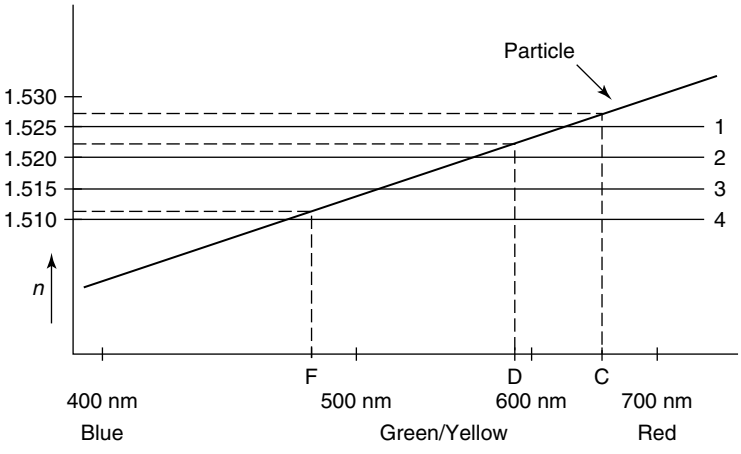


Figure 19 Dispersion staining curve for the particle in Figure 16. The dispersion staining values are $n_F = 1.511$, $n_D = 1.522$, and $n_C = 1.527$

Particle Characterization – Single Polar

As noted in the section "Brightfield", without any polarizer (polar) in the system one can observe only the average color and RI of anisotropic particles. This is because ordinary light is unpolarized and vibrates in a multitude of different directions perpendicular to the propagation direction of travel. Since optical properties will vary with the vibration of the light in an anisotropic particle, one must employ plane polarized light to observe and measure these variations.

Plane polarized light is obtained by using a polarizing filter. When unpolarized light passes through a polarizing filter, the light will exit vibrating in a single (privileged) direction perpendicular to the direction of travel (Figure 20). Now with the employment of

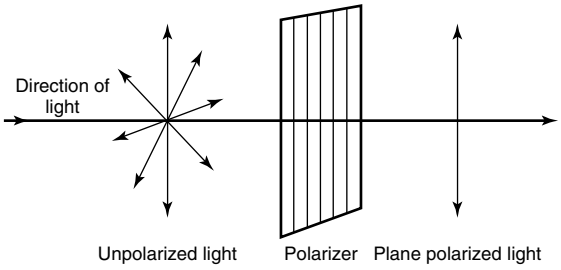


Figure 20 Unpolarized light passing through a polarizing filter resulting in plane polarized light

a single polar, one will be able to measure different optical properties of an anisotropic material as a function of the vibration direction of the light passing through the particle. Therefore, utilizing PLM will increase the number of optical properties that one can determine when characterizing and identifying particles. Before one measures the different properties of a particle, it is necessary to have a well centered stage and check the alignment of the polarizer and analyzer which are normally aligned perpendicular to each other.

Refractive Indices. With plane polarized light, more than one RI can be observed for anisotropic particles. The RIs one will observe will depend on the orientation of the particle and the alignment of the RIs with the polarizer. Fibers are discussed first since they lay flat on the microscope slide and show their two principle indices: the RI parallel ($n_{\parallel}$) to the fiber length and the RI perpendicular ($n_{\perp}$) to the fiber. When the length of the fiber is aligned parallel to the polarizer, the $n_{\parallel}$ RI will be observed (Figure 21a). If one rotates the stage and aligns the fiber perpendicular to the polarizer, than the $n_{\perp}$ will be observed (Figure 21b). If one rotates the stage so neither of the RIs coincides with the polarizer, than the light will be vectorially split between the two RIs and will be split equally when the fiber is aligned at a 45° angle to the polarizer (Figure 21c). When the fiber is aligned parallel or perpendicular to the polarizer, the respective RIs can be estimated or precisely determined using immersion methods or by employing the dispersion staining technique. Note that the polyester fiber shows higher contrast when the fiber is aligned east–west (parallel RI) and is a quick way to check that your polarizer is also aligned east–west.

Unlike fibers, which show their longitudinal view, most crystals show a randomly oriented view when mounted on a microscope slide. The optical indicatrix helps us to understand how the RI varies depending on the vibration direction of light through a crystal. The indicatrix is a 3-D model whose radii are proportional to the RI values perpendicular, or nearly so, to the direction of light and, therefore, assumes a different solid geometrical form depending on the RIs of a crystal. Thus, for an isotropic (cubic or noncrystalline) particle which has a single refractive index, all the radii would be of equal length and will result in the simplest indicatrix having the form of a sphere (Figure 22). If one would cut a section at any angle through the center of the indicatrix, a circular section would be obtained having a circumference equidistant at all points from the center of the sphere. Therefore, one would observe a single RI for light traveling perpendicular to any circular section no matter the vibration direction of the light.

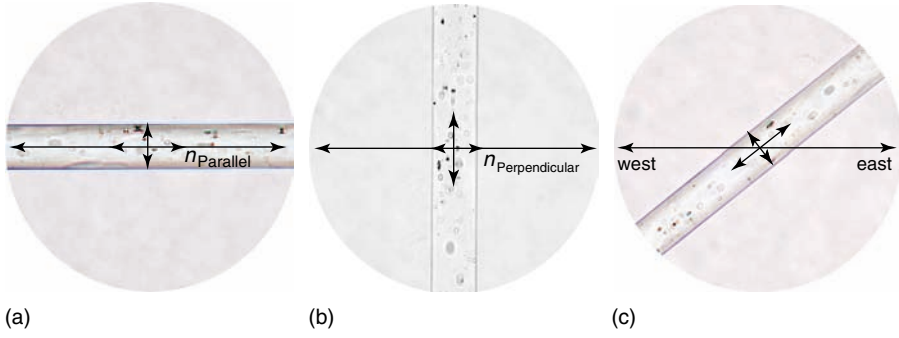


Figure 21 Three different orientations of a polyester fiber viewed with an east-west polarizer. Note the chance of contrast in the different views. The RI of mounting medium is 1.525

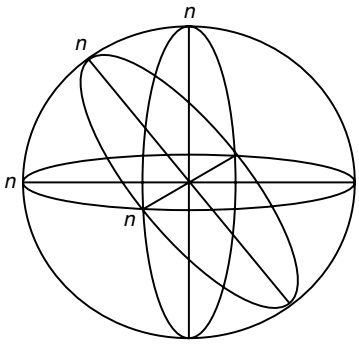


Figure 22 Isotropic indicatrix

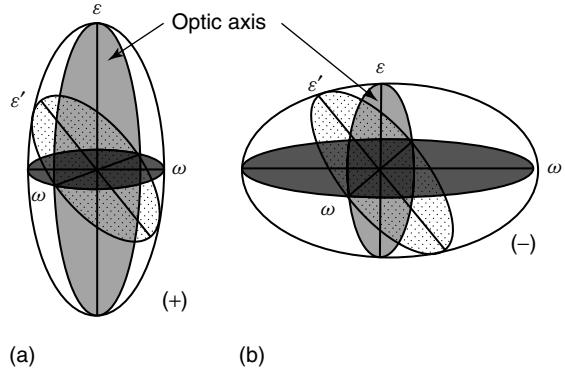


Figure 23 Uniaxial indicatrix: (a) prolate ellipse where $\varepsilon > \omega$ (positive optic sign) and (b) oblate ellipse where $\varepsilon < \omega$ (negative optic sign)

For uniaxial (tetragonal and hexagonal) crystals, the indicatrix takes on an elliptical form with the maximum and minimum magnitude of the radii equal to the two principal indices, ε and ω. If the $\varepsilon > \omega$ (+) optic sign the shape of the ellipsoid will be prolate (Figure 23a) and if $\varepsilon < \omega$ (−) optic sign the shape of the ellipsoid will be oblate (Figure 23b).

For all uniaxial crystals, the ε RI coincides with the optic axis. The optic axis is a unique axis and if one cuts a section though the center of the indicatrix perpendicular to the optic axis, a circular section (shaded dark gray) will be obtained as shown in Figure 23. Plane polarized light traveling along the optic axis will vibrate in this circular section containing only the ω RI. Therefore, a uniaxial crystal in this orientation would behave as an isotropic crystal since only the ω RI will be observed (Figure 24a) and the contrast for the crystal will remain constant with plane polarized light as you rotate the stage.

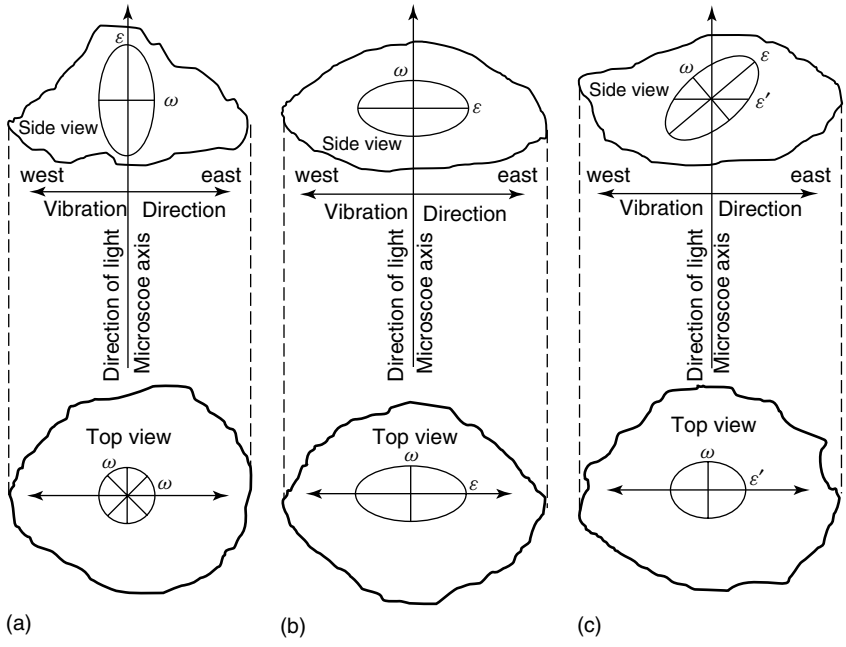


Figure 24 Three different orientations with the uniaxial indicatrix superimposed in the different views

For a section cut through the center of a uniaxial indicatrix parallel to the two principal indices, ε and ω, a principal ellipse (shaded light gray) will be obtained as seen in Figure 23. Therefore, if the propagation direction of the plane polarized light is perpendicular to this section, one will observe either the ε RI or the ω RI when one or the other is aligned parallel to the vibration direction of the polarized light (Figure 24b). If the stage is rotated so neither principal RI is parallel with the polarizer, as previously discussed, the plane polarized light will be split vectorially between the two principal RIs and equally when they are positioned at a 45° angle from the polarizer. Also, with a crystal in this orientation the contrast will change as the stage is rotated and the degree of contrast change will be dependent on the differences in the RIs between ε, ω and the mounting medium.

An ellipse will be obtained if a section (textured plane) is cut at a diagonal angle through the center of an indicatrix as shown in Figure 23. This elliptical cross section will contain the ω and ε' RIs. An ε' RI can be any multitude of RIs between ε and ω and which ε' RI will be observed is dependent on the orientation of the crystal. If the propagation direction of plane polarized light is perpendicular to an elliptical section, one will observe either the ω RI or the ε' RI when one or the other is aligned parallel to the vibration direction of the polarized light (Figure 24c). If the stage is rotated so neither RI is parallel with the polarizer, the plane polarized light will be split vectorially between the two RIs and equally when they are positioned at a 45° angle from the polarizer. Also, as mentioned above, when one rotates the stage with a crystal in this orientation the contrast will change and to what degree will depend on the differences in the RIs of ω, ε' and the mounting medium. How to easily locate the RIs for a randomly oriented crystal is discussed in the crossed polarized light section.

For biaxial (orthorhombic, monoclinic, and triclinic) crystals, the indicatrix is a triaxial ellipsoid defined by the three principal indices α, β, and γ. By definition, α is the lowest RI, γ is the highest RI, and β is the intermediate RI that lies somewhere between α and γ. As one can see in Figure 25a, light traveling parallel to γ will vibrate in the plane (shaded dark gray) containing α and β, light traveling parallel to α will vibrate in the plane (shaded gray) containing β and γ, and light traveling parallel to β will vibrate in the plane (shaded light gray) containing α and γ. In the plane (shaded light gray) containing α, the lowest RI, and γ, the highest RI, there are radii equal to β RI. Therefore, there are two sets of radii that would describe two circular planes containing only the β RI in the biaxial indicatrix (Figure 25b). Light traveling perpendicular to the two circular planes would travel along the optic axes which lie in the ellipsoid plane containing α and γ. If the β RI is closer to α RI, the crystal has a positive (+) optic sign (Figure 25c). However, if the β RI is closer to the γ RI, the crystal has a negative (−) optic sign (Figure 25d). The RIs between α and β are noted as α' and RIs between β and γ are noted as γ'.

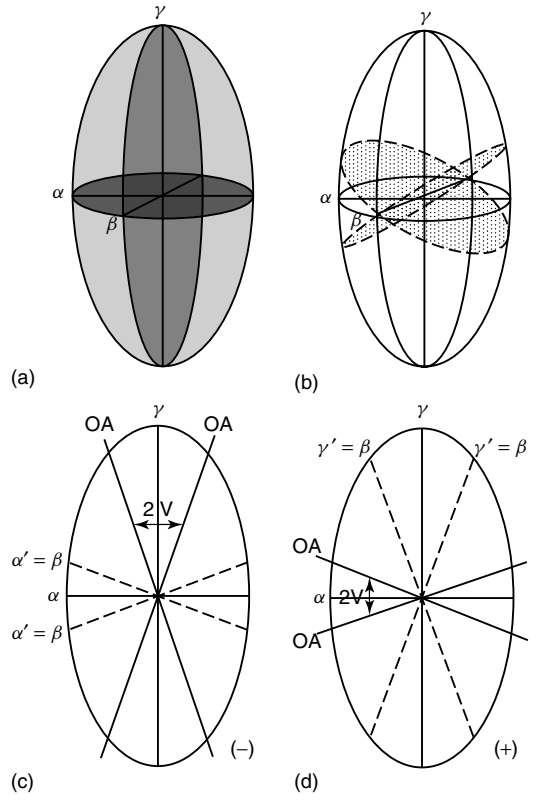


Figure 25 Biaxial Indicatrix

As with uniaxial crystals, one can superimpose the RIs within a crystal to understand and/or predict optical characteristics that one might observe during the examination of a particle (Figure 26). Again, how to easily locate the RIs for a randomly oriented crystal will be discussed in the crossed polarized section.

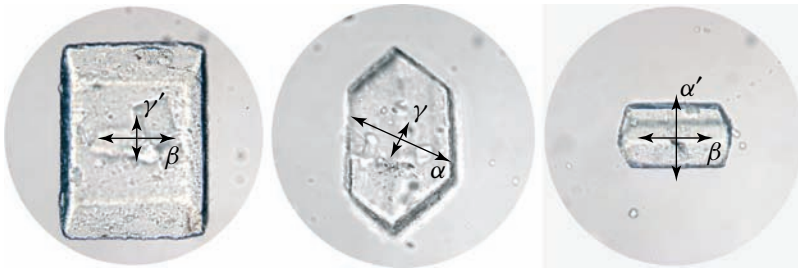


Figure 26 Three views of a sucrose crystal with the RIs superimposed in the different views

The RIs of anisotropic particles can be estimated or precisely determined using immersion methods or by employing the dispersion staining technique.

Color – Pleochroism and Dichroism. Pleochroism observed with plane polarized light provides an additional diagnostic characteristic. Pleochroism is the general term for the phenomenon of colored anisotropic particles that absorb different wavelengths of light depending on the vibration direction through the particle. Dichroism is the proper term that refers to uniaxial crystals and many dyed fibers that show two colors with plane polarized light. All colored anisotropic particles show pleochroism but at times the change in color may be too faint to detect. Colored isotropic particles and colorless particles do not show pleochroism.

Employing plane polarized light, all one needs to do to observe pleochroism/dichorism is to rotate the stage while viewing a colored anisotropic crystal or a dyed fiber which displays this effect (Figure 27). When observing pleochroism, one should note the change of color corresponding to any physical features such as a crystal face, cleavage plane, or fiber length. The strongest absorption color usually corresponds to the highest RI direction but the asbestos mineral crocidolite is an exception to this rule.

Particle Characterization – Crossed Polars

With crossed polars, one will greatly increase the number of identification characteristics that can be determined for a particle under investigation. Crossing the polars will allow one to determine the degree of birefringence, observe different types of extinction, determine the sign of elongation, and observe interference figures. As previously discussed, it is very important to insure that the polars are properly aligned with the eyepiece crosslines and the stage is well centered.

Wave Interference. Two waves having the same wavelength and traveling along the same propagation direction will interfere to produce a resultant wave having the same wavelength whose amplitude is the vector sum of the original two waves. However, it is important to understand the difference in how the resultant wave and interference colors are produced by two rays vibrating in the same plane or by two waves after vibrating in perpendicular planes.

Retardation of Waves Polarized in the Same Plane. If two waves, having the same wavelength and traveling in the same direction, are vibrating in the same plane and one wave is retarded from the other

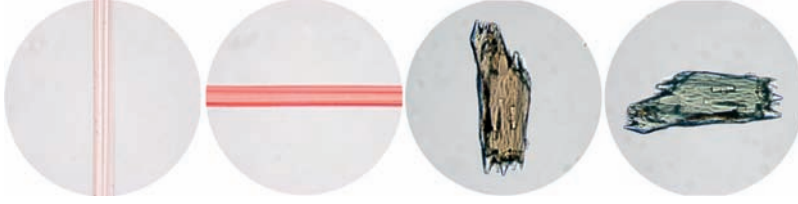


Figure 27 Dyed fiber showing dichorism (left two images) and a hyperstene crystal showing pleochroism (right two images)

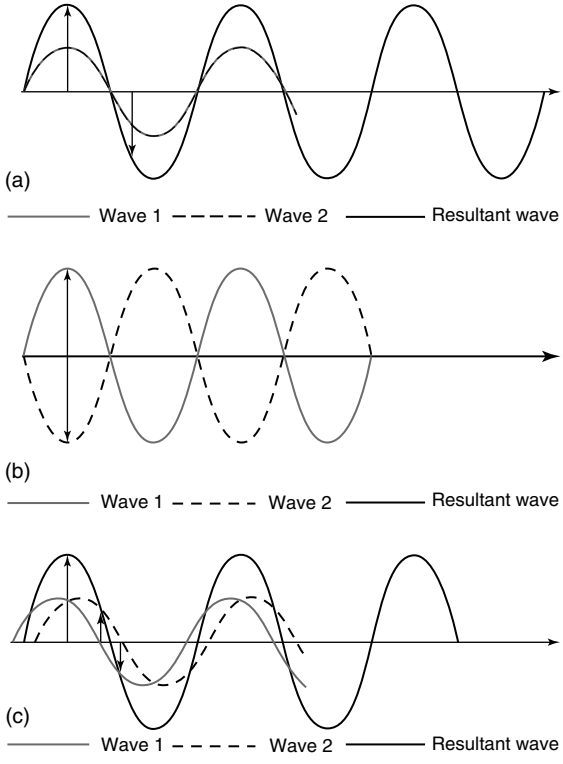


Figure 28 Interference between two waves traveling in the same plane and along the same path produce the resultant wave shown when one wave is retarded by (a) $n\lambda$, (b) $n\lambda + \lambda/2$, and (c) $n\lambda = \lambda/8$. [Illustration by David Diener]

wave resulting in a path difference 1λ, 2λ, or any $n\lambda$ (where n is an integer) the two waves will be in phase. This will result in constructive interference and the resultant wave will appear brighter since the amplitude of the resultant wave is twice the amplitude of the original rays (Figure 28a). However, if two waves, having the same wavelength and traveling in the same direction, are vibrating in the same plane and one wave is retarded from the other wave so the path difference is $\lambda/2$, $1\lambda + \lambda/2$, $2\lambda + \lambda/2$, or any $n\lambda + \lambda/2$, the two waves will destructively interfere. The resultant wave will have zero amplitude and this wavelength will be eliminated (Figure 28b). Furthermore, if two waves, having the same wavelength and traveling in the same direction, are vibrating in the same plane and one ray is retarded some distance other than $n\lambda$ or $n\lambda + \lambda/2$, the resultant wave produced will be a combination of both constructive and destructive interference (Figure 28c).

Interference Colors Produced by Two Waves Vibrating in the Same Plane. If two beams of white light interfere in the same plane as described above, some wavelengths will be eliminated while other wavelengths will be intensified. As a result, a series of colors known as *interference colors* will be produced and the resultant color will be dependent on the phase difference for the different wavelengths of light. This phenomenon produces what is commonly referred to as *thin film interference colors*, also known as *Newton's colors*. These colors are frequently seen when white light incident on a thin transparent film is reflected from the top and bottom surfaces, resulting in the reflected rays interfering with each other. Soap bubbles, a layer of oil on water, and thin deposits on a reflective substrate are a few examples that commonly show thin film interference colors.

Retardation of Waves Polarized in the Perpendicular Planes. When plane polarized light enters an anisotropic crystal which is oriented so the two indices (major or prime) are at a 45° angle from the polarizer, the wave will be resolved and vectorially split equally into two waves traveling in two mutually perpendicular planes. As a result, the slow wave (higher RI) will be retarded from the fast wave (lower RI). When both rays exit the crystal, they will continue to vibrate in perpendicular planes and will interfere with each other at the analyzer to produce a resultant wave vibrating in a single plane that is the vector sum of the two parent waves. If the two waves for a particular wavelength have a path difference of 1λ, 2λ, 3λ, or any $n\lambda$, the resultant wave will vibrate in the same plane as the original wave. The analyzer, which is a second polarizer located above the crystal whose privileged vibration direction is perpendicular to the polarizer, will block this wavelength of light (Figure 29a). However, if for a particular wavelength the two waves vibrating in two mutually perpendicular planes have a path difference of $\lambda/2$, $1\lambda + \lambda/2$, $2\lambda + \lambda/2$, or any $n\lambda + \lambda/2$, the resultant wave will vibrate in a plane particular to the polarizer vibration direction. Therefore, the analyzer will allow this wavelength of light to be transmitted through the analyzer (Figure 29b).

Furthermore, when a path difference for a particular wavelength has been separated $\lambda/4$, $1\lambda + \lambda/4$,

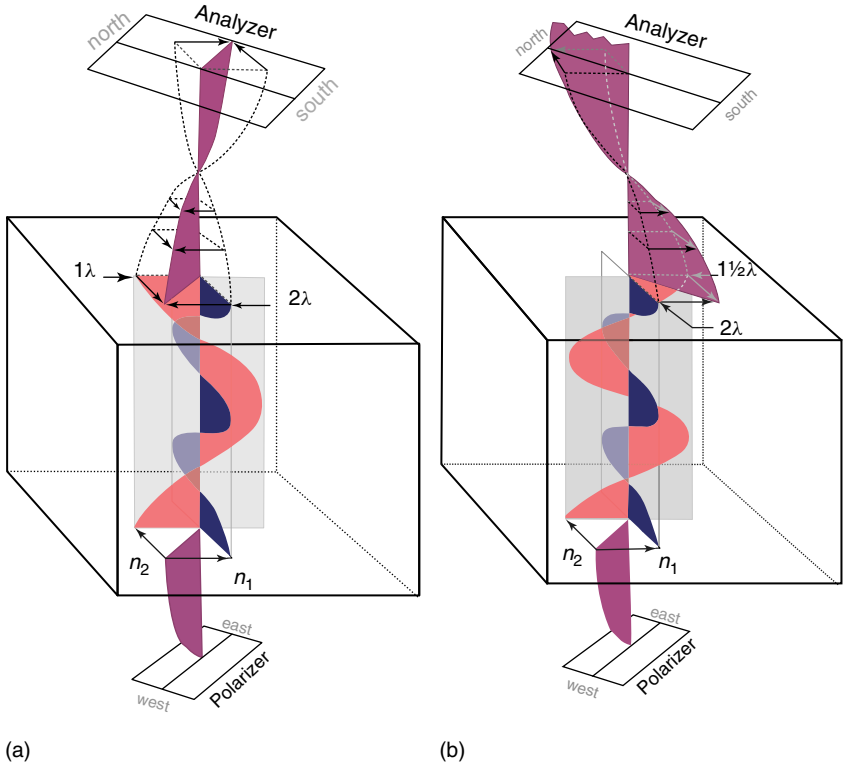


Figure 29 Interference (a) between two waves after emergence from a crystal when one wave is retarded by 1λ. Note the resultant wave is polarized in the same plane as the polarizer and would be completely be extinguished by the analyzer. Interference (b) between two waves after emergence from a crystal when one wave is retarded by 1/2λ. Note the resultant wave is polarized in the same plane as the analyzer and would be completely be transmitted by the analyzer. For illustration purposes, the waves are shown interfering before they reach the analyzer which allows for an easier understanding of the interaction of light when it reaches the analyzer. [Illustration by David Diener]

$2\lambda + \lambda/4$, or any $n\lambda + \lambda/4$, the resultant wave will vibrate in a circular, cork screw fashion and only 50% of this wavelength will be passed (transmitted) by the analyzer. And, if the path difference is any other value than $n\lambda$, $n\lambda + \lambda/2$, $n\lambda + \lambda/4$, which is normally the case, the resultant wave will vibrate in an elliptical fashion and some percentage of the resultant wave will be passed by the analyzer. The percent transmission of light by the analyzer for different phase differences is illustrated in Figure 30.

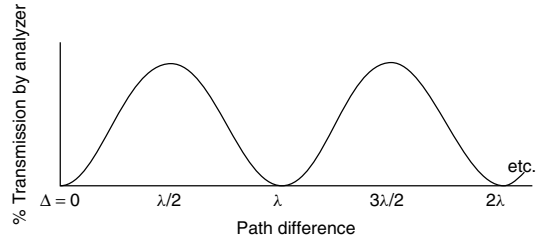


Figure 30 Percent transmission of light by the analyzer depending on the phase difference (retardation) after passage through a birefringent crystal with crossed polars

Interference Colors Produced by Two Waves Vibrating in Perpendicular Planes. If illuminated with white light, an anisotropic crystal viewed between crossed polars where the light vibrates along the two indices, a series of polarization (interference) colors will be observed as a result of unequal transmission by the analyzer of the different wavelengths of light as previously described. The interference colors of the particle will be brightest when the two indices are positioned at a 45° angle from the

vibration direction of the polarizer. These interference colors are commonly referred to as *polarization colors* or *retardation colors*.

Birefringence. Birefringence is the numerical difference of the two refractive indices ($n_2 - n_1$) for an anisotropic particle in a particular view. These indices for a crystal may be principal or prime indices and for a fiber ($n_{\parallel} - n_{\perp}$). When the birefringence value is stated in literature, it is the maximum difference between the indices of a fiber or crystal unless otherwise stated. The interference color one sees between crossed polars is due to retardation (r) of the different wavelengths of light and is the result of the relationship between birefringence (B) and thickness (T) of the material under investigation. It is very important to remember that the thickness of the material is not the linear distance measured in the plane of the preparation but the thickness along the microscope axis.

This relationship is expressed as $r = 1000BT$ and is plotted graphically in the Michel–Levy chart that was first published in 1889, and a microscopist can determine any one of the values if the other two are known (Figure 31). Referring to the Michel–Levy chart, one can see that thickness increases as you move up the left ordinate. The retardation (path difference) increases as you go to the right and the colors become paler. The birefringence for a particle can be determined by moving up the diagonal line where an observed retardation color and known thickness intersect. Glass fibers of any thickness would be black on a black field with crossed polars since glass has one RI and light travels through the particle at a single velocity. However, a 10 μm fiber showing a first-order red color would have a birefringence of 0.055

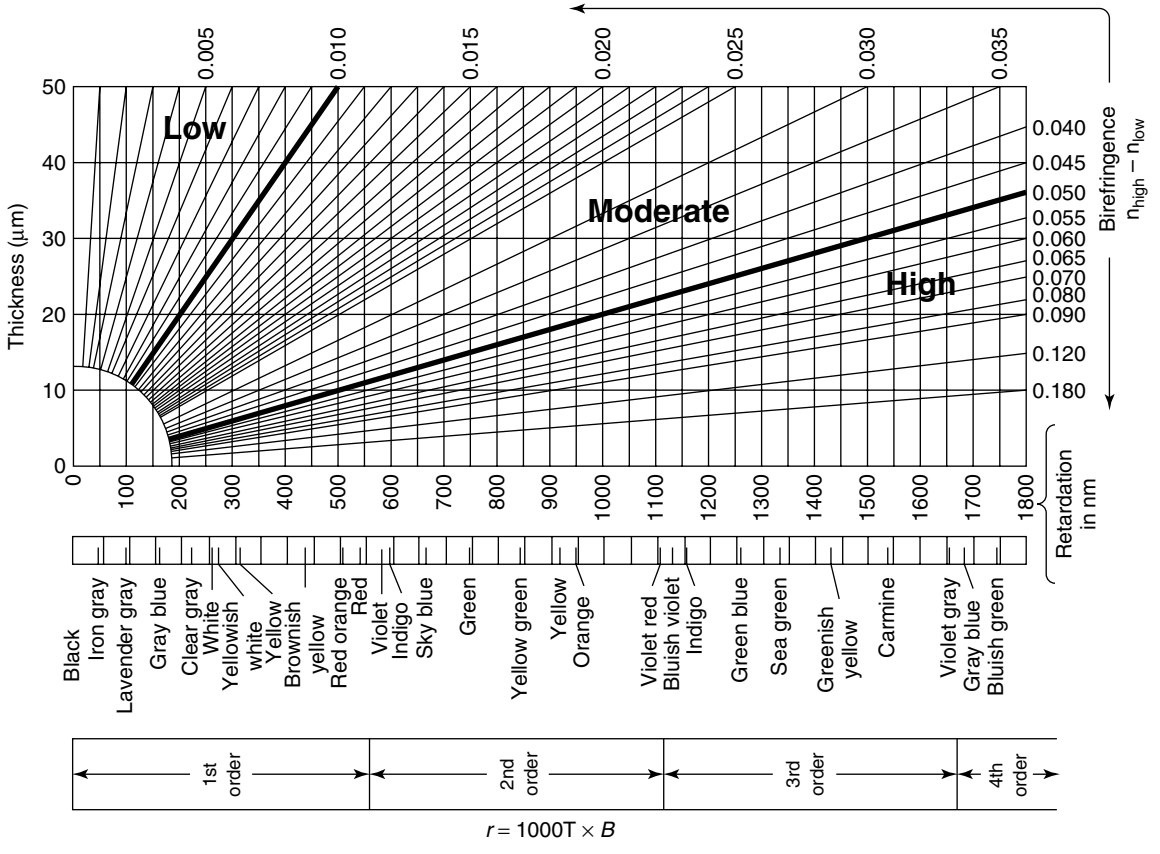


Figure 31 A black and white Michel–Lévy chart showing the relationship among retardation, thickness, and birefringence. One can obtain a color Michel–Lévy chart from most of the references listed in the bibliography

and a 10 μm fiber showing a third-order red color would have a birefringence of 0.170. Some particles will not follow the Michel–Levy series of retardation colors but will show anomalous polarization colors and when recognized it is a unique identifying characteristic.

Extinction. Particles will be brightest between crossed polars when the two indices are in a particular view. As one rotates the stage, the particle will usually go dark every 90°, a circumstance known as *extinction*. There are different types of extinction one can observe for a particle (Figure 32). Most particles will show complete or uniform extinction and may show parallel, symmetrical, or oblique extinction.

Some particles will not uniformly go to complete extinction but will show nonuniform extinction. Undulose extinction shows areas of extinction that move progressively across the particle in a fanlike motion as the stage is rotated. Polycrystalline crystals will show areas of extinction due to the random orientation of the crystals to each other.

Other particles may not show any extinction. Cotton fibers may not show good extinction due to the different orientations of the overlaying top and bottom layer of the collapsed fiber. Some particles in a particular view may show dispersion of extinction and the different "extinction" colors are dependent on the wavelength of light (Figure 33). When recognized, it is a unique identifying characteristic.

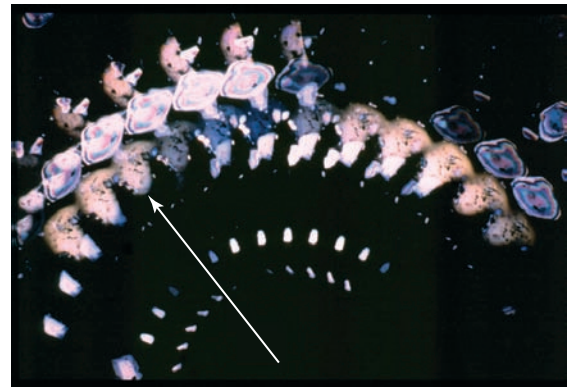


Figure 33 Photomicrograph of RDX by Dr. Walter C. McCrone with a crystal (arrow) showing dispersed extinction when viewed with crossed polars. The photomicrograph is a series of 12 exposures with the stage being rotated several degrees between each exposure. Note the crystal does not go to complete extinction but displays a series of colors from blue to purple to yellow. The RDX crystal above the crystal showing dispersed extinction has a different orientation and shows complete extinction [Reproduced with permission from McCrone Research Institute.]

Sign of Elongation. By convention, a fiber has a positive (+) sign of elongation if $n_{\parallel} > n_{\perp}$ and a negative (−) sign of elongation if $n_{\parallel} < n_{\perp}$. One can quickly determine if the fiber has a (+) or (−) sign of elongation by using a compensator. Compensators are usually marked with an arrow with a γ or Z symbol

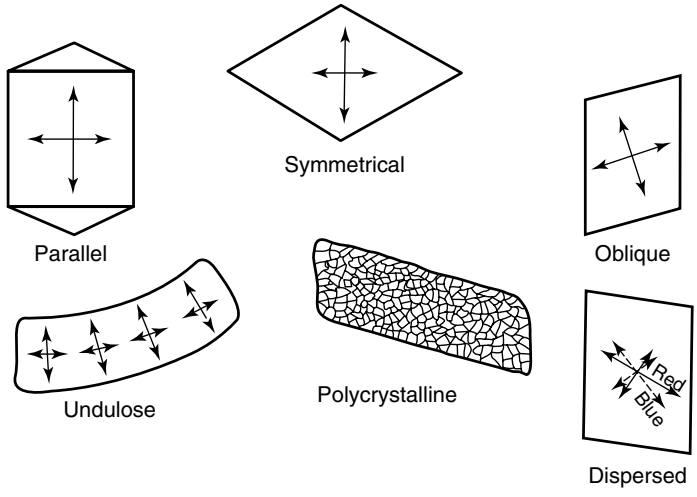


Figure 32 Different types of extinction [Reproduced with permission from McCrone Research Institute.]

to indicate the slow component (higher RI) direction. Commonly used compensators to determine the sign of elongation include the first-order red plate having a retardation of 530 nm and the quarter-wave plate having a retardation of 137 nm.

Utilizing crossed polars, the retardation colors are noted for a fiber that is aligned parallel to the slow component direction of the compensator. When a first-order red plate is inserted, the retardation will change by 530 nm and the two retardations (fiber and compensator) will be added or subtracted (lower RI from the higher RI). Addition means the two slow components are parallel and the fiber has a (+) sign of elongation (i.e., $n_{\parallel} > n_{\perp}$). Subtraction means the slow component of the fiber is perpendicular to the slow component of the compensator and the fiber has a (−) sign of elongation (i.e., $n_{\parallel} < n_{\perp}$).

For example, if a gray blue (150 nm) retardation color is observed with crossed polars for a fiber aligned parallel to the slow component of the compensator and turns blue (680 nm) when the 530 nm compensator is inserted, then the two retardations have added and the fiber has a (+) sign of elongation. However, the fiber has a negative sign of elongation if it turns yellow (380 nm). Higher-order retardation colors observed with crossed polars will appear more pale when the compensator is inserted for a fiber that has a (+) sign of elongation and brighter for a (−) sign of elongation. Wollastonite, a fibrous mineral, may show (+) or (−) sign of elongation because the β RI is parallel to the length of the fiber. When α (or α') RI is perpendicular to the length, the fiber will have a (+) sign of elongation and a (−) sign of elongation when γ (or γ') RI is perpendicular to the fiber length.

Interference Figures. An interference figure is an important diagnostic feature that is formed by rays that travel along different directions within the crystal. Interference figures are observed by examining the back focal plane of the objective using a Bertrand lens, a phase telescope, or by removing an eyepiece and inserting a pinhole eye cap. Unfortunately, most crystals have orientations that provide interference figures that are of little diagnostic value. However, if one can recognize a crystal in the orientation that will provide a useful figure, one can determine a number of optical properties such as if the crystal is uniaxial or biaxial, the location of the principal refractive indices, and the optic sign. Low retardation colors observed on highly birefringent crystals provide the best interference figures.

An objective having a numerical aperture of at least 0.65 is required to obtain an interference figure and the aperture diaphragm needs to be opened to provide a full cone of light to the objective. An example of a centered uniaxial interference figure is provided and a centered biaxial figure is provided in Figure 34. One should be aware of the value of interference figures even though a more in depth discussion is not possible here but one can investigate a number of references listed in the bibliography such as *Optical Crystallography* by F. Donald Bloss for more information.

Fusion Methods

Fusion methods many times can be included when characterizing a sample. Most of the time, a sample can be heated over an alcohol lamp and characteristic crystals that develop from the melt state can be observed. Also, a hot stage can be attached to the microscope and used to accurately determine the melting point of a particle under investigation. Many microscopists use the hot stage to distinguish between polymers like low density and high density polyethylene as well as nylon 6 and nylon 6,6.

Fusion methods can be used to characterize not only organic compounds but also some inorganic compounds. For example, ammonium nitrate has four crystal phases above room temperature: an orthorhombic crystal phase from room temperature to 32 °C, a second biaxial crystal phase from 32 to 84 °C, a tetragonal crystal phase from 84 to 125 °C, and finally a cubic crystal phase which melts at 170 °C. Using a polarized light microscope, it is truly a spectacular sight to view a sample going from room temperature to a melted state and from a melted

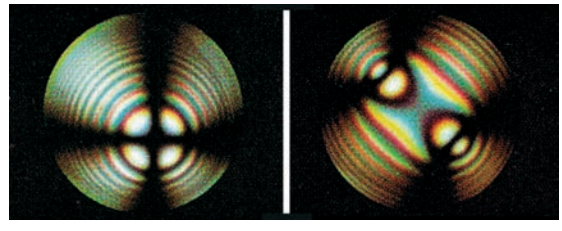


Figure 34 Interference figures: uniaxial (left) and biaxial (right) [Reproduced with permission from McCrone Research Institute.]

state back to room temperature. For more information on fusion methods one should consult the *Polarized Light Microscopy* manual by McCrone, McCrone and Delly as well as *Fusion Methods* by McCrone.

Microchemical Tests

Microchemical tests aid in the identification of particles under investigation and should not be overlooked by the forensic microscopist. These tests can be applied to both inorganic compounds (e.g., explosives) and organic compounds (e.g., drugs) and many times a microchemical test will provide more information about a particle under investigation (e.g., valence state) than other modern instrumental methods. Microchemical tests are particularly useful because they are quickly and easily conducted, their sensitivity requires only a minute amount of sample, and they are easy to perform.

Most microchemical tests are based on the formation of characteristics crystals that develop using a selected reagent and are normally carried out on a microscope slide without a coverslip using a 10X objective. An in-depth discussion on microchemical methods is not possible in this article and one should consult references such as the *Handbook of Chemical Microscopy, Volume 1 and 2*, by Chamot and Mason as well as "Microchemical Reactions in Particle Identification" by S. Palenik in the *Particle Atlas, Volume 5*, for a detailed presentation on microchemical methods.

Fluorescence Microscopy

Since the early 1980s, fluorescence microscopy has become a common analytical tool employed in the trace evidence section of many forensic laboratories. This technique is a sensitive and simple analytical tool that can be employed to characterize and compare many different types of trace evidence such as paints, cosmetics, fibers, adhesives, plastics, and mineral grains in soil samples. Fluorescence microscopy is a nondestructive method that can quickly provide information which cannot be obtained by other types of instrumentation. This article addresses only reflected or incident light (also referred to as *Epi*) fluorescence since it is the type of fluorescence microscopy used today. A simple diagram of a reflected light fluorescence microscope showing the ray paths for illuminating and imaging the specimen is provided in Figure 35.

Fluorescence

Luminescence is the term used to describe nonthermal radiation of longer wavelength (lower energy) that is emitted when energy of shorter wavelength (higher energy) is absorbed by some materials. The fluorescent material may be inorganic or organic and the emitted radiation (luminescence) is usually in the visible region of light. The energy stimulation can be ion, X-ray, electron, or photon radiation. Luminescence can also be produced by chemical reactions

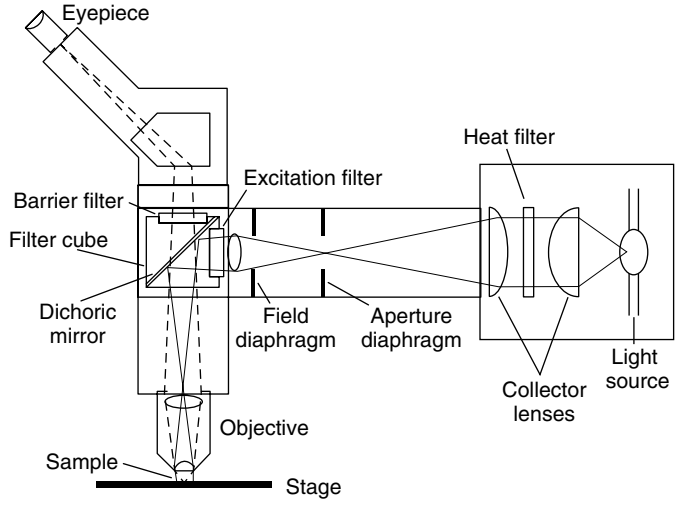


Figure 35 Schematic diagram showing the illuminating and imagining ray path through a fluorescent microscope

or by mechanical action. Depending on the type of primary excitation, different descriptive terms may be used to describe the luminescence: e.g., ionoluminescence (ion), Roentgenoluminescence (X-ray), cathodoluminescence (electron), and photoluminescence (photon).

Luminescence can be further subdivided into phosphorescence and fluorescence. Phosphorescence is generally characterized by continued luminescence after the excitation stops and fluorescence ceases instantaneously when the excitation stops. Fluorescence can also be subdivided into primary fluorescence (auto) and secondary fluorescence. Primary fluorescence is the ability to emit luminescence without any treatment or staining. Secondary fluorescence requires staining the specimen with a dye (e.g., fluorchrome). Normally, only primary fluorescence will be of interest in the examination of trace evidence.

Illuminators

There are two types of light sources used for fluorescence microscopy which are the pressure mercury lamp and the xenon lamp. The high-pressure mercury lamp produces a rather noncontinuous energy distribution with most of its energy output concentrated in several narrow bands (Figure 36). The xenon lamp produces a more continuous spectrum from the ultraviolet to the near infrared and closely resembles characteristics of visible light (Figure 37). In addition to being able to produce a more continuous energy spectrum, xenon lamps can be switched on and off without waiting for the lamp to warm-up or cool down. Various output powers are available for these

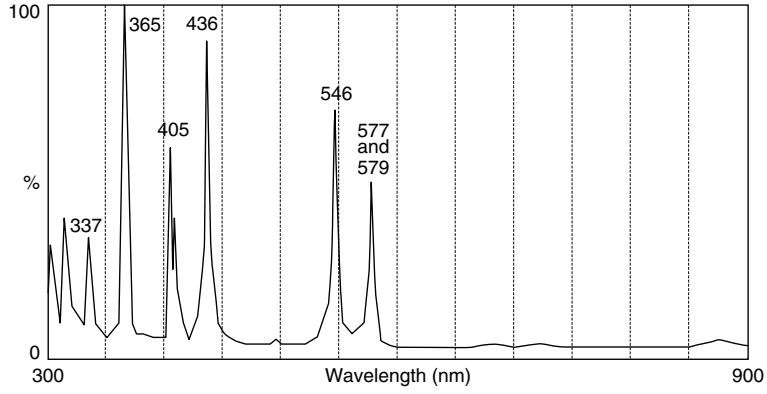


Figure 36 Spectral output for mercury lamp

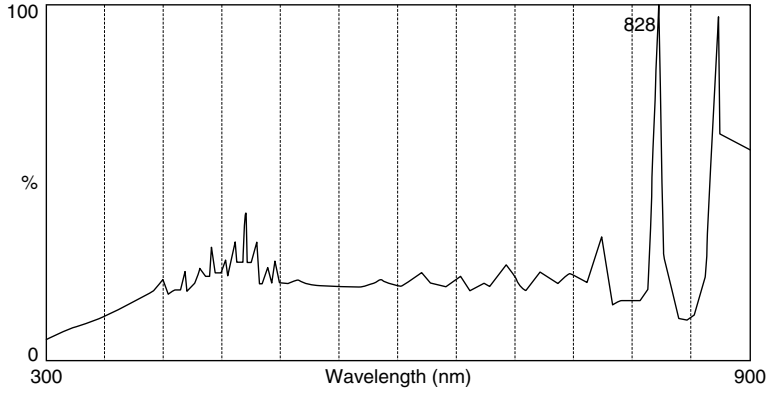


Figure 37 Spectral output for xenon lamp

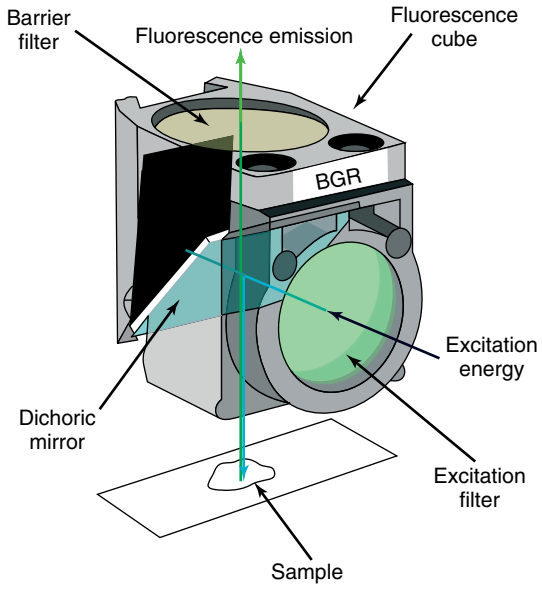


Figure 38 Cutaway view showing the components of a fluorescent filter cube

lamps, but generally 100 W mercury lamps and 75 W xenon lamps are used for the examination of trace evidence samples. Before starting any examination, one needs to ensure that the microscope has been properly aligned and adjusted so that the illumination field provides the best image quality.

Optical Cube

Besides the alignment of the microscope, the proper selection of the filters contained in the cube is an important factor in determining the intensity of the fluorescence image observed though the microscope. The optical cube or filter cube contains the excitation filter, the dichroic mirror (DM), and the barrier filter (Figure 38). Manufacturers today offer a variety of fluorescence cubes containing different filter and mirror combinations, and it is important to understand the purpose of each component of the cube. The filters and mirror in a cube allow certain wavelengths of light to be transmitted, reflected, or absorbed and the correct configuration will maximize the fluorescence intensity of the sample being examined.

Excitation Filter. The excitation or exciter filter (EX) allows shorter wavelengths of light from the

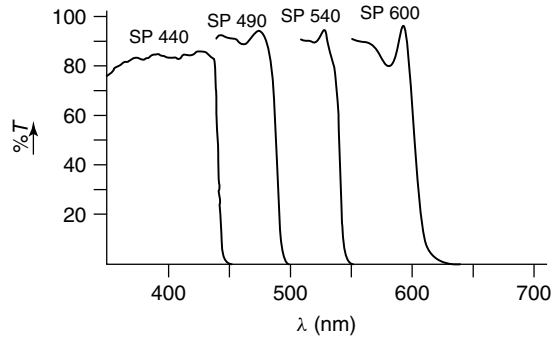


Figure 39 Typical transmission curves for several shortpass excitation filters. The nomenclature denotes the cut-off wavelength at 50% maximum transmission

illuminator to pass through while longer wavelengths are suppressed. As one would expect, the purpose of this filter is to transmit only the wavelengths of illumination that will effectively excite the specimen and eliminate longer wavelengths of light that would obscure any fluorescence from the specimen. The filter may have simple nomenclature designating the wavelength range that is passed (transmitted) such as UV or U for ultraviolet, B for blue, G for green or may have a more detailed nomenclature specifying the wavelength range that is passed by the filter. Shortpass (SP) filters that transmit shorter wavelengths yet block longer wavelengths were used in the past for the excitation filter. The nomenclature typically used for these filters denotes the cut-off wavelength that is located at 50% maximum transmission (Figure 39). Today, however, bandpass (BP) filters, or bandpass interference filters (BPIF) are utilized almost exclusively for the excitation filter. BP filters typically have a nomenclature which denotes the center wavelength of the transmission and may note the bandwidth at half of the maximum transmission (Figure 40). BP filters can further be designated as wideband or narrowband filters. As suggested by S. Palenik in "Microscopical Examination of Fibers", wideband filters would be a better choice for the forensic microscopist since they are considering a wide range of materials that may fluoresce over a broad range of wavelengths.

Dichroic Mirror. The shorter wavelengths of excitation illumination will be reflected toward the specimen by the DM, also referred to as the *chromatic beam splitter* (BS). The DM is positioned at a 45°

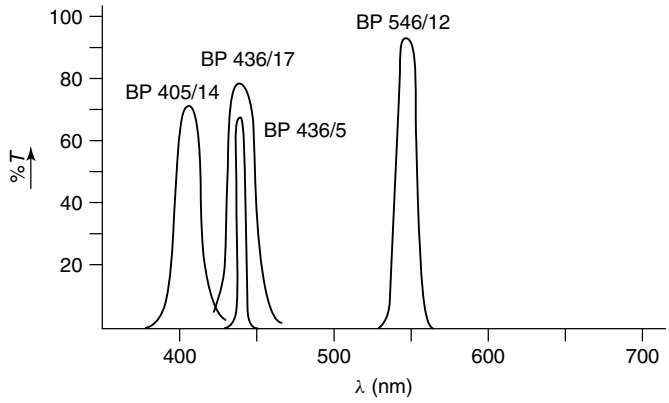


Figure 40 Typical transmission curves for several bandpass excitation filters. The nomenclature denotes the center wavelength of transmission and the bandwidth at 50% maximum transmission

in the optical light path. At the same time, the unwanted longer wavelengths of light will be transmitted through the DM and absorbed on the interior black coated surface of the cube. Once the excitation wavelength reaches the specimen, the emitted fluorescent wavelengths from the sample and some of the unabsorbed excitation wavelength(s) of light will be reflected back toward the DM. The longer fluorescent wavelengths will be transmitted through the DM toward the eyepiece and the unabsorbed and unwanted excitation wavelengths will be reflected back to the lamp. This unique ability of the DM to reflect the shorter wavelengths and pass longer wavelengths is accomplished with the use of a thin-film interference coated piece of glass. Current DMs can pass and reflect light with 90% efficiency. The typical nomenclature for a DM will denote the wavelength (located at 50% maximum transmission) at which shorter wavelengths will be reflected by the mirror and longer wavelengths will be passed (Figure 41).

Barrier Filter. The last filter in the cube is the barrier (BA) or emitter (EM) filter. Since the wavelength ranges transmitted for the excitation filter and the DM may slightly overlap, the barrier filter is needed to block any harmful ultraviolet excitation light and effectively transmit only the fluorescence emitted by the sample. A barrier filter can be either a longpass (LP) filter or BP filter. Longpass filters allow wavelengths longer than their cut-on point to be transmitted and the nomenclature for these filters denotes the cut-on wavelength that is located at 50% maximum transmission (Figure 42).

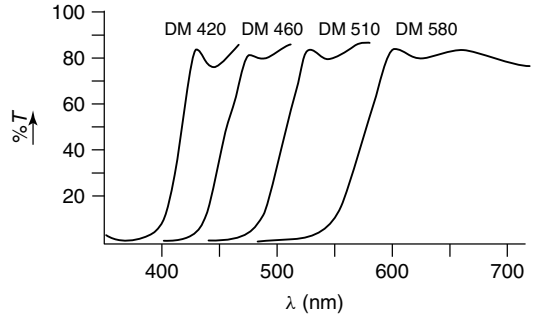


Figure 41 Typical transmission curves for several dichroic mirrors. The nomenclature denotes the cut-on wavelength at 50% maximum transmission

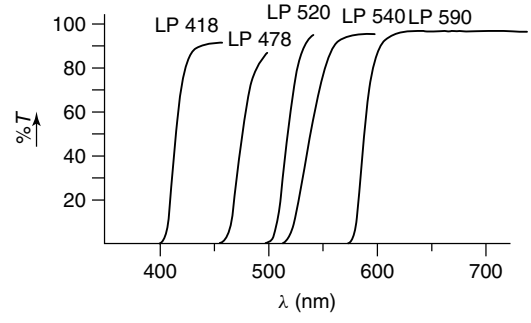


Figure 42 Typical transmission curves for several barrier filters. The nomenclature denotes the cut-on wavelength at 50% maximum transmission

A simple ray diagram showing the function of each component in a filter is provided in Figure 43

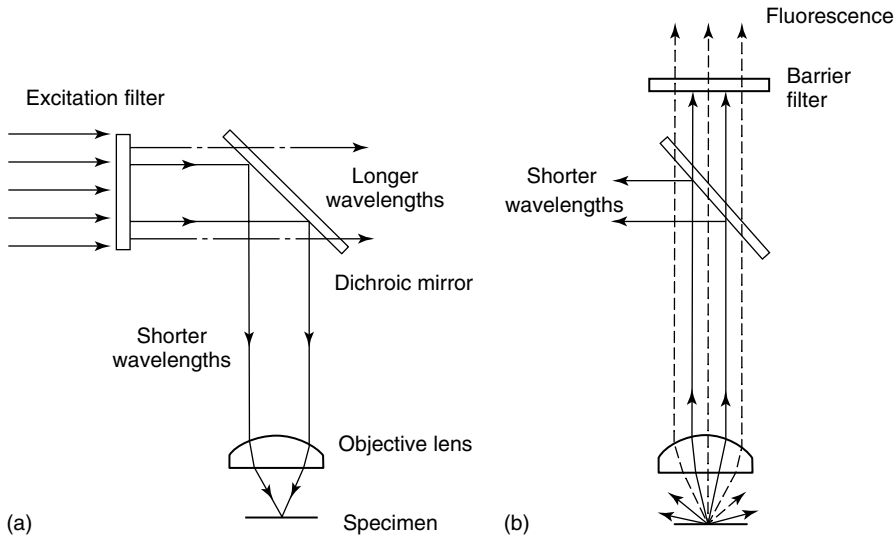


Figure 43 Schematic diagram showing the function for the different components of a fluorescence filter cube. Excitation rays (a) and fluorescence emission (b)

and an example of a typical transmission curve for a filter cube is provided in Figure 44.

Cathodoluminescence

Furthermore, it is worth noting that the use of cathodoluminescence (CL) microscopy is gaining popularity in forensic laboratories. As mentioned at the beginning of this section, cathodoluminescence is the fluorescence (luminescence) obtained by excitation of the sample using electrons instead of photons. Accessories for the scanning election microscope (SEM) can be obtained to observe cathodoluminescence or one can outfit a light microscope with a small vacuum specimen chamber containing a cold-cathode electron gun. An example of a paint sample observed utilizing reflected visible light, fluorescence, and cathodoluminescence is provided in Figure 45. For more information on cathodoluminescence, one may consult "Applications of Cathodoluminescence in Forensic Science" by S. Palenik and J. Buscaglia in *Forensic Analysis on the Cutting Edge: New Methods for Trace Evidence Analysis*.

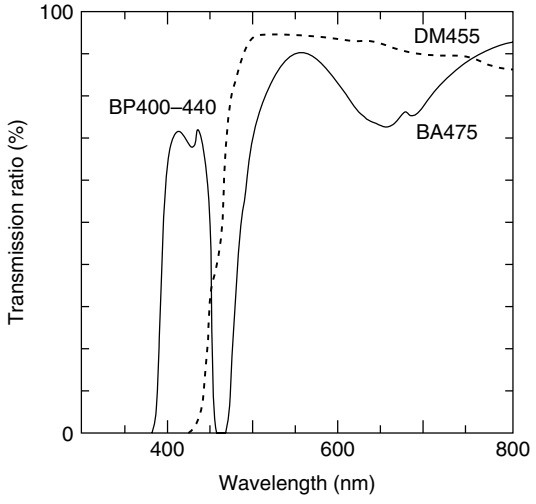


Figure 44 Transmission curves for a typical fluorescence filter cube

Comparison Microscopy

Comparison microscopes are used extensively in a forensic laboratory to compare pieces of evidence to a known specimen. Although comparison microscopes are employed in the examination of bullets, cartridge cases, or toolmarks, this type of examination is also used to compare trace evidence such as hairs, fibers, paint chips, and other types of particles that may

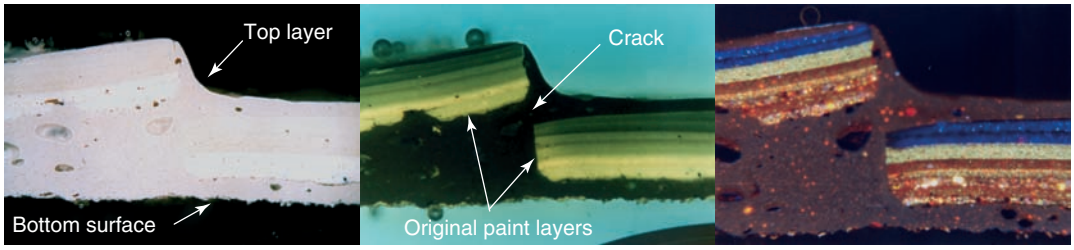


Figure 45 Embedded cross section of a multilayered white paint sample from an old building. The last application of white paint flowed through the crack in the paint and migrated under the older layers of paint. Images (from left to right) are reflected (ordinary) light, fluorescence, and cathodoluminescence

become separated from a source and accidentally left at the scene of a crime.

Design

The most critical part of a comparison microscope is the optical bridge that connects the optical paths of two side-by-side microscope systems. This bridge consists of multiple prisms arranged in such a manner to allow the magnified images from the left and right sides to be viewed simultaneously with a fine "hair-line" split between the two fields of view. Depending on the complexity of the instrument, the prisms can be adjusted either mechanically or by motors so that the hair-line is moved horizontally in the field of view to expose more or less of one side of the image or to superimpose the left and right images over each other.

The comparison microscope used for firearms, toolmarks and documents is configured for low visual magnifications from 2X to 120X and uses incident (reflected) illumination. A system used for trace evidence comparisons uses two compound microscopes usually equipped for intermediate visual magnifications from 40X to 600X and with brightfield and polarized transmitted illumination (Figure 46). Microscopes for fiber comparisons are now often supplied with epi-fluorescence components for comparison of known and questioned materials at selected excitation and emission wavelengths.



Figure 46 Modern trace evidence comparison microscope

Acknowledgment

The authors would like to thank Ms Natasha Neel with the Bureau of Alcohol, Tobacco, Firearms, and Explosives and Mr Christopher Taylor with the US Army Criminal Investigation Laboratory who reviewed many versions of this article and provided a number of valuable suggestions. A special thanks to Mr Stephen Garten with the Bureau of Alcohol, Tobacco, Firearms, and Explosives who helped prepare figures, compile references, and review various versions of this document. Also, the authors dedicate this article to the late Dr Walter C. McCrone, our mentor and good friend, who encouraged and helped many forensic microscopists worldwide.

Further Reading

Stereomicroscopy

DeForest, P.R. (2005). Foundations of forensic microscopy, in *Forensic Science Handbook*, R. Saferstein, ed, Prentice Hall, Englewood Cliffs, pp. 416–528.

McCrone, W.C. (1991). Light microscopy, in *Physical Methods of Chemistry*, B.W. Rossiter & J.F. Hamilton, eds, John Wiley & Sons, New York.

McCrone, W., McCrone, L. & Delly, J. (1978). *Polarized Light Microscopy*, McCrone Research Institute, Chicago.

Needham, G. (1958). *The Practical Use of the Microscope*, Springfield, Thomas.

Nikon (2007). Nikon microscopyu: stereomicroscopy [Online], http://www.microscopyu.com/articles/stereomicroscopy/index.html.
Petraco, N. & Kubic, T. (2004). *Microscopy for Criminalists, Chemists, and Conservators*, CRC Press, Boca Raton.

Brightfield Microscopy

Abramowitz, M. (2003). *Microscope Basics and Beyond*, Olympus America, New York, Vol. 1.
Bradbury, S. & Bracegirdle, B. (1998). *Introduction to Light Microscopy*, BIOS Scientific Publishers, Oxford.
Davidson, M. & Abramowitz, M. Optical microscopy [Online], http://www.olympusmicro.com/primer/microscopy.pdf.
DeForest, P.R. (2005). Foundations of forensic microscopy, in *Forensic Science Handbook*, R. Saferstein, ed, Prentice Hall, Englewood Cliffs, pp. 416–528.
Houck, M. & Siegel, J. (2006). Microscopy, *Fundamentals of Forensic Science*, Academic Press, Amsterdam, pp. 79–98.
McCrone, W.C. (1991). Light microscopy, in *Physical Methods of Chemistry*, B.W. Rossiter & J.F. Hamilton, eds, John Wiley & Sons, New York, pp. 343–443.
McCrone, W., McCrone, L. & Delly, J. (1978). *Polarized Light Microscopy*, McCrone Research Institute, Chicago.
Needham, G. (1958). *The Practical Use of the Microscope*, Springfield, Thomas.
New York Microscopical Society (1989). *Glossary of Microscopical Terms and Definitions*, 2nd Edition, New York Microscopical Society, New York (this is currently available from the McCrone Research Institute in Chicago).
Nikon (2007). Nikon microscopyu: introduction to polarized microscopy [Online], http://www.microscopyu.com/articles/polarized/polarizedintro.html.
Nikon (2007). Nikon microscopyu: concepts and formulas in microscopy [Online], http://www.microscopyu.com/articles/formulas/formulasindex.html.
Petraco, N. & Kubic, T. (2004). *Microscopy for Criminalists, Chemists, and Conservators*, CRC Press, Boca Raton.
Zeiler, H. (1973). *The Optical Performance of the Light Microscope*, 2 parts, McCrone Research Institute, Chicago.

Polarized Light Microscopy

Benedetti-Pichler, A. (1942). *Introduction of the Microtechnique of Inorganic Analysis*, John Wiley & Sons, New York.
Bloss, F. (1999). *An Introduction to the Methods of Optical Crystallography*, Mineralogical Society of America, Washington, DC.
Chamot, E. & Mason, C. (1931). *Handbook of Chemical Microscopy*, Wiley, New York (Vol. 2 is currently available from The McCrone Research Institute in Chicago) 2 vols.
DeForest, P.R. (2005). Foundations of forensic microscopy, in *Forensic Science Handbook*, R. Saferstein, ed, Prentice Hall, Englewood Cliffs, pp. 416–528.
Dyar, M.D., Gunter, M.E. & Tasa, D. (2006). Optical mineralogy, in *Mineral and Optical Mineralogy*, Mineralogical Society of America, pp. 81–102.
Hartshorne, N. & Stuart, A. (1970). *Crystals and the Polarizing Microscope*, American Elsevier Science, New York.
Houck, M. & Bowen, R. (2005). An argument for light microscopy – a review of forensic microscopy for trace evidence analysis, *Forensic Science Review* **17**(1), 1.
Houck, M. & Siegel, J. (2006). Microscopy, in *Fundamentals of Forensic Science*, Academic Press, Amsterdam, pp. 79–98.
Kile, D. (2003). *The Petrographic Microscope: Evolution of a Mineralogical Research Instrument*, The Mineralogical Record, Arizona.
McCrone, W. (1957). *Fusion Methods in Chemical Microscopy*, Interscience Publishers, New York (this is currently available from the McCrone Research Institute in Chicago).
McCrone, W.C. (1979). Particle analysis in the crime laboratory, in *The Particle Atlas*, W.C. Mc.Crone, J.G. Delly & S.J., Palenik, eds, Ann Arbor Science Publishers, Ann Arbor, p. 1379 (this is currently available on CD from the McCrone Research Institute in Chicago) Vol. 5.
McCrone, W.C. (1987). *Asbestos Identification*, Ann Arbor Science Publishers, Ann Arbor.
McCrone, W.C. (1991). Light microscopy, in *Physical Methods of Chemistry*, B.W. Rossiter & J.F. Hamilton, eds, John Wiley & Sons, New York, pp. 343–443.
McCrone, W., McCrone, L. & Delly, J. (1978). *Polarized Light Microscopy*, McCrone Research Institute, Chicago.
Nikon (2007). Nikon microscopyu: introduction to polarized microscopy [Online], http://www.microscopyu.com/articles/polarized/polarizedintro.html.
Olympus (2007). Olympus microscopy resource center: specialized microscopy techniques – polarized light microscopy [Online], http://www.olympusmicro.com/primer/techniques/polarized/polarizedhome.html.
Palenik, S.J. (1979). Microchemical reactions in particle identification, in *The Particle Atlas*, W.C. Mc.Crone, J.G. Delly & S.J. Palinik, eds, Ann Arbor Science Publishers, Ann Arbor, pp. 1175–1184 (this is currently available on CD from the McCrone Research Institute in Chicago) Vol. **5**.
Palenik, S. (1988). Microscopy and microchemistry of physical evidence, in *Forensic Science Handbook*, R. Saferstien, ed, Prentice Hall, Englewood Cliffs, pp. 161–208.
Petraco, N. & Kubic, T. (2004). *Microscopy for Criminalists, Chemists, and Conservators*, CRC Press, Boca Raton.
Stoiber, R. & Morse, S. (1994). *Crystal Identification with the Polarizing Microscope*, Chapman and Hall, New York.
Viney, C. (1990). *Transmitted Polarized Light Microscopy*, McCrone Research Institute, Chicago.
Wahlstrom, E. (1979). *Optical Crystallography*, John Wiley and Sons, New York.

Fluorescence Microscopy

Abramowitz, M. (1993). *Fluorescence Microscopy: The Essentials*, Olympus American, New York.
Birk, G. (1984). *Instrumentation and Techniques for Fluorescence Microscopy*, Wild Leitz (Australia) Pty, Sydney.
Herman, B. (1998). *Fluorescence Microscopy*, BIOS Scientific Publishers, Oxford, 2 Vols.

Holz, H. *Worthwhile Facts About Fluorescence Microscopy*, Zeiss Publication.
Marshall, D. (1988). *Cathodoluminescence of Geological Materials*, Unwin Hyman, London.
Nikon (2007). Nikon microscopyu: fluorescence microscopy [Online], http://www.microscopyu.com/articles/fluorescence/index.html.
Olympus (2007). Olympus microscopy resource center: specialized microscopy techniques – fluorescence microscopy [Online], http://www.olympusmicro.com/primer/techniques/fluorescence/fluorhome.html.
Palenik, S. (1999). Microscopical examination of fibers, in *Forensic Examination of Fibres*, J. Robertson & M. Grieve, eds, CRC Press, Boca Raton, p. 153.
Palenik, C. & Buscaglia, J. (2007). Applications of cathodoluminescence in forensic science, in *Forensic Analysis on the Cutting Edge*, R. Blackledge, ed, John Wiley & Sons, New Jersey, pp. 141–174.
Reichman, J. (2000). *Handbook of Optical Filters for Fluorescence Microscopy*, Chroma Technology Corp, Rockingham.
Semrock 2000. 2007 Catalog.
Spring, K. & Davidson, M. Introduction to fluorescence microscopy [Online], http://www.microscopyu.com/articles/fluorescence/fluorescenceintro.html.

Comparison Microscopy

Ernst Leitz Wetzlar (1986). *Bulletin for the Forensic Laboratory – Comparison Microscope*, R. Beck, ed, Ernst Leitz Wetzlar.
Needham, G. (1958). *The Practical Use of the Microscope*, Springfield, Thomas.
Petraco, N. & Kubic, T. (2004). *Microscopy for Criminalists, Chemists, and Conservators*, CRC Press, Boca Raton.

THOMAS J. HOPEN AND MALCOLM DAVIS

Microscopy: Low Power

Microscope is an instrument designed to extend man's visual capability, i.e., to make visible even the minute details that can not be seen with the naked eye [1].

On the basis of the early pioneering study on microscopy in the sixteenth century, the first real microscope was made in the seventeenth century by Antony van Leeuwenhoek. Nonetheless, Leeuwenhoek, using this simple microscope, which was little more than a magnifying glass, discovered bacteria and spermatozoa and he built 400 microscopes in his lifetime! From this "single lens" microscope evolved the *compound microscope*, which simply means more than one lens so that the image magnified by one lens can be magnified by another. Today the term *microscope* is generally used to refer to a compound microscope, where the lens closer to the object being viewed is the *objective* lens while that closer to the eye of the observer is the *eyepiece*. Despite the apparent complexity of research level microscopes, they are in essence simple. They are also probably the most poorly understood and badly used of all laboratory instruments. The stereo microscope is a compound microscope. A typical high-quality stereo microscope is shown in Figure 1. In fact it comprises two compound microscopes, which are aligned side-by-side at the correct visual angle to provide a true stereoscopic image. Most stereo microscopes in laboratories today have a common main objective, and as the "workings" of the microscope are housed in a single body, it is not obvious that there are two optical systems present. Older Greenough type systems had two completely separate optical systems. Not all low-magnification binocular (two eyepieces) microscopes are stereo microscopes. Low-cost, low-magnification microscopes for students often have a single objective and image forming system, with a beam splitter delivering an image to both eyepieces. These microscopes do not produce a stereo image.

Current stereo microscopes have many attractive properties:

- they have a long working distance (the distance between the specimen and the objective lens);
- an upright nonreversed image (i.e., the image is seen in the same direction as it sits on the stage, facilitating sample movement and magnification);
- the image has a genuine three dimensional feel;
- little sample preparation usually required;
- various types of illumination can be used including transmitted light and overhead or incident light of several types (ring light, angled reflected light);
- large field of view; and
- can be attached to an arm enabling the examination of larger objects.

Modern instruments can also be set up for polarizing and fluorescence applications.

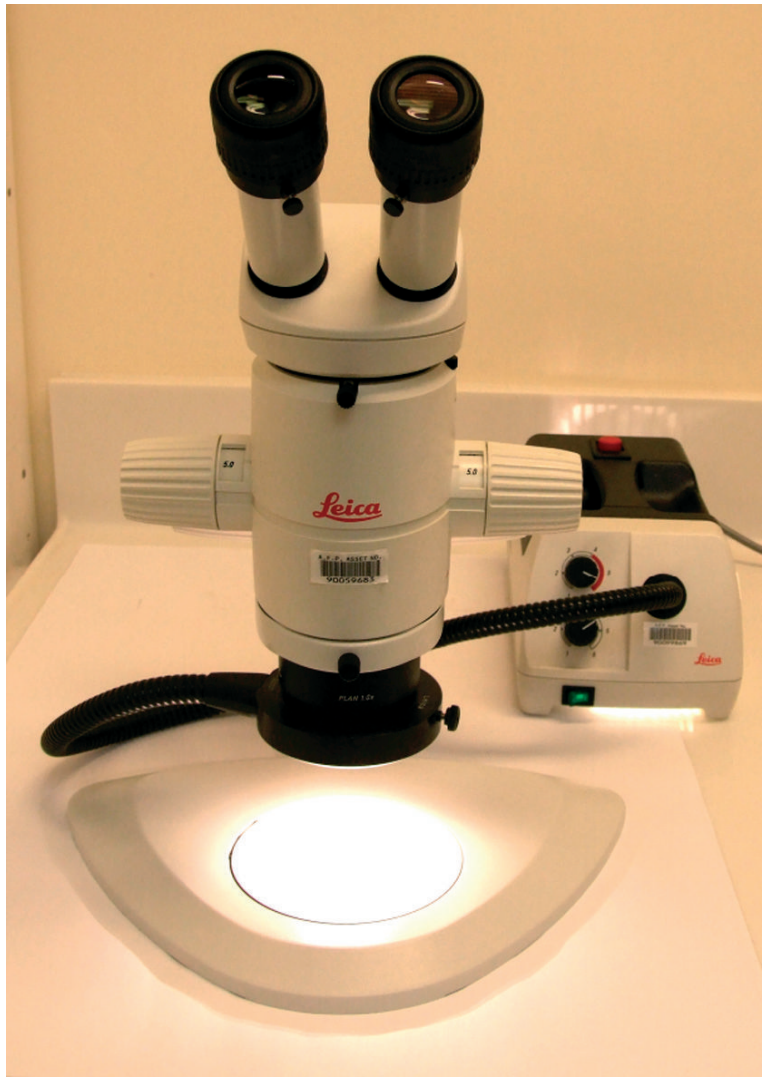


Figure 1 Stereo microscope

The useful magnification range is typically between 2.5× and a maximum of 100×. More expensive instruments often have a zoom capability.

The stereo microscope can be considered to be an extension of the human eye allowing the observer to see more details. In forensic work, stereo microscopes are mainly used to investigate the microscopic details of large objects, such as clothing, down to small fibers, soil, plant material, glass, paint, and other possible physical evidence contained in recovered debris.

There are probably no areas of forensic endeavor where low-power microscopy does not have a role to play.

With most new instruments, video or still digital cameras can be attached enabling the examiner to record images for inclusion in case notes or in case reports.

Training in the use of low-power microscopes is strongly recommended if they are to be used to their full potential. However, even a relatively inexperienced individual can make reasonably effective use

of simple low-power microscopes. This is not the case with high-power microscopes.

Reference

[1] Anon (1989). *RMS Dictionary of Light Microscopy*, Oxford University Press – Royal Microscopical Society.

Further Reading

De Forest, P.R. (2002). Foundation of forensic microscopy, in *Forensic Science Handbook*, 2nd Edition, R. Saferstein, ed, Prentice Hall, New Jersey, Vol. 1, Chapter 5, pp. 215–319.
Houck, M.M. & Bowen, R. (2005). An argument for light microscopy – a review of forensic microscopy for trace evidence analysis. *Forensic Science Review* **17**, 1–15.

JAMES ROBERTSON

Microscopy: Scanning Electron Microscopy

Introduction

The invention of the electron microscope gave rise to the discovery of infinite entities, so far invisible, by optical microscopes. The use of electrons made it possible to overcome the optical limits of light. Nowadays, electron microscopy is a well-established technique and is widely used for forensic science purposes. With resolution in the nanometer and Angstrom range, electron microscopy and scanning probe microscopy are two of the most powerful microscopy techniques available today.

According to the imaging mechanism, the family of electron microscopes consists of the transmission electron microscope (TEM), the scanning transmission microscope (STEM), the scanning electron microscope (SEM), and the environmental SEM (ESEM). The Nobel prize-winning TEM was first developed, followed by the SEMs [1, 2]. The TEMs, TEM as well as STEM, are less in use for forensic observations because of the limitation of specimen thickness ($<1\,\mu$m thick).

The first SEMs were constructed in the 1930s and 1940s and have since then been constantly developed into very high-resolution instruments and this development lead to the invention of the ESEM, which allowed the observation of specimens that were, until then, impossible to submit to a classical SEM because of its high-vacuum environment.

The SEM is a very versatile instrument for the visualization of very small objects, the study of surface morphology, thanks to its resolution and excellent depth of field as well as its spectroscopic possibilities (Figure 1a, b). Like optical microscopes, the SEM measures only the X and Y dimensions of a sample, contrary to scanning probe microscopes (SPMs) that measure the Z dimension as well.

The SEM operates with an electron beam that interacts with the sample and generates several signals that are detected and processed to form the image [3]. Given that the electrons interact with the molecules of air, the electron beam has to be produced and travel in a vacuum, avoid trajectory deviations.

SEM: Functioning and Image Mechanisms

SEMs consist of five main parts: the electron source, the electron column with a series of electromagnetic lenses, the sample chamber, the signal processing part constituted of several detectors, the signal processing devices, and the image visualization devices such as a monitor (Figure 2). An elaborated pumping system is attached to the microscope, keeping the microscope and its parts under vacuum (ranging from 10^{-4} to about 10^{-10} Torr).

Electron Source

The electron beam can be generated by either a thermoionic source or a field emission (FE) source. The first kind produces electrons when heated and the second type generates electrons when an intense electric field is applied to it. These sources are part of an assembly – the electron gun. The two different systems cannot be interchanged. However, the choice of the source influences the performance of the SEM, i.e., the quality of its electron beam as well as its probe size. The FE source produces a monochromatic beam (small electron energy spread), where thermoionic sources have a greater electron

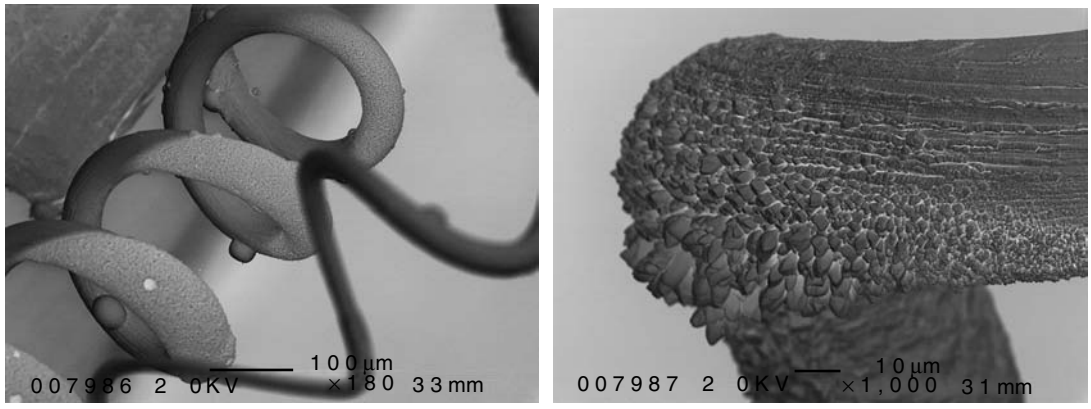


Figure 1 The high depth of field and high resolution allow having a sharp general view (a) or detail view (b) of a filament of a broken car light bulb

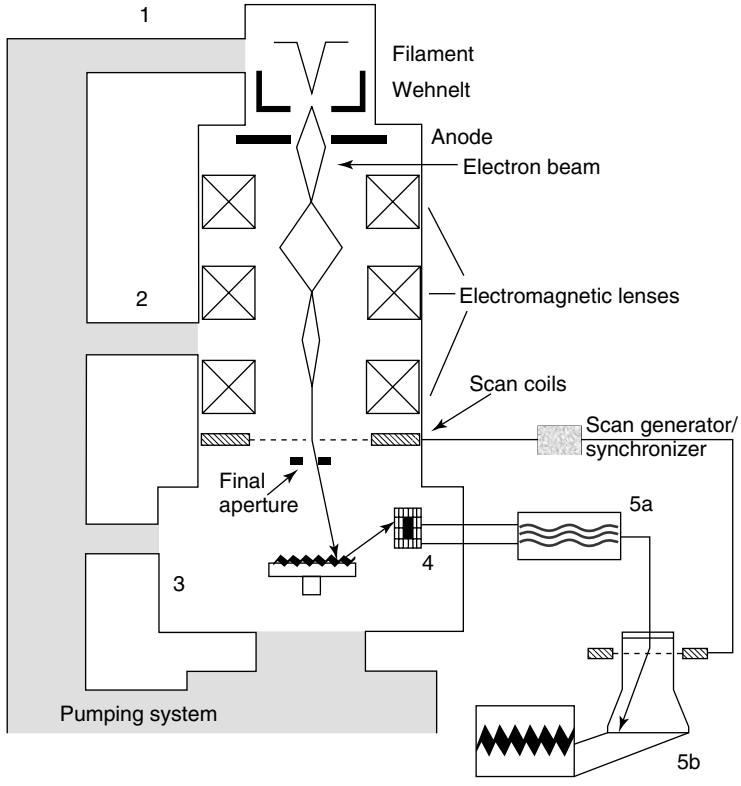


Figure 2 Schema of a conventional SEM: 1. Electron source (thermoionic gun); 2. electron column; 3. sample chamber; 4. detectors; 5a. Signal processing devices (e.g., photomultiplier, amplifier, etc.); and 5b. visualization system (monitor and computer)

energy range. The gun operates under vacuum is common to both the systems (10^{-4}–10^{-9} Torr).

The function of the electron gun, independent of the type, is to focus the electrons coming off the source and to direct this beam into the electron column and finally to the sample. To achieve this, the electron source is incorporated into a gun assembly, which acts as a lens to focus the emitted electrons. The assembly basically consists of a cathode and an anode. However, the design of the gun is different for thermoionic sources and field emission sources.

Thermoionic Gun. This gun consists of three elements: the electron emitting material (commonly called *filament*), the Wehnelt, and the anode (Figure 3).

These three parts form the so-called triode where the filament is the cathode. The Wehnelt has a potential that is more negative – the bias voltage – than the cathode itself. The bias voltage is variable and is used for controlling the emission of electrons from the filament. A high bias voltage restricts the emission to a small area, thereby reducing the total emitted current. A low bias voltage increases the size of the emitting area on the filament and thus the total emission current. The Wehnelt creates a negative field that makes the emitted electron converge to a point called a crossover situated between the Wehnelt and the anode (Figure 3). This crossover determines the size of the final probe diameter and thereby the resolution of the microscope. The Wehnelt acts as a simple electrostatic lens – the first lens in the microscope. The crossover is determined by the type of filament, the electric field between the cathode and the anode, and the exit angles of the electrons from the filament. The heating current and the Wehnelt bias voltage have to be set in order to operate at saturation condition (i.e., the smallest beam diameter with an even bright spot).

The filament in a thermoionic gun has to be either of a high-melting point material or with a low work function (i.e., necessary energy for the electrons to escape). In practice, tungsten with its high melting temperature (3660 K) and rare-earth boride crystal such as lanthanum hexaboride (LaB_6) or cerium hexaboride (CeB_6) are used. The tungsten filament has either a hairpin shape or is pointed, to minimize the emitting area. Tungsten filaments are heated directly, where rare-earth boride crystals are indirectly heated by a bonding material surrounding the crystal, such as carbon or rhenium. This type of filament can be used in standard tungsten SEM if the vacuum system is upgraded. Tungsten filaments need a vacuum of around 10^{-4} Torr where boride crystals require a higher vacuum (around 10^{-6}–10^{-7} Torr).

Field Emission Gun. FE guns are much simpler than thermoionic ones. They consist of a cathode and two anodes (Figure 4).

A very fine tungsten tip with a radius of <100 nm constitutes the cathode. This sharp tip is submitted to a strong electric field (>10^{-7} V/cm) that lowers the work function barrier sufficiently for electrons to tunnel out of the tungsten. This electric field is

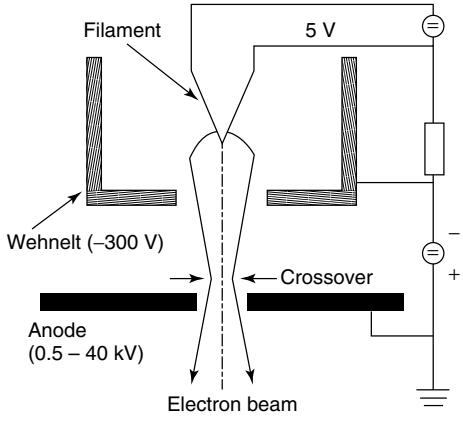


Figure 3 Schema of a thermoionic gun with a hair-pin-shaped filament

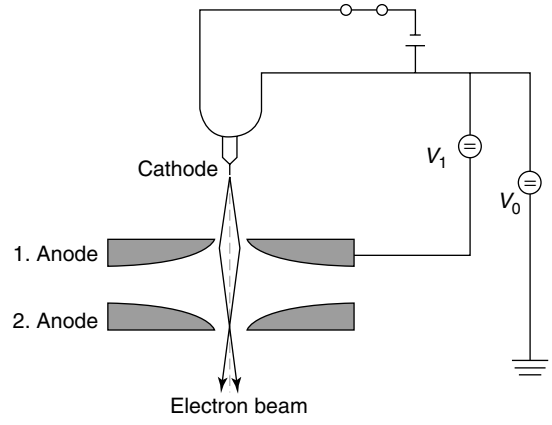


Figure 4 Schema of a cold field emission gun (V_0, acceleration bias; V_1, extraction bias)

Table 1 Comparison of the resolution in function of the electron gun system:

kV	Tungsten (nm)	LaB_6 (nm)	Field emission	
			CFE (nm)	TFE (nm)
0.5	80	50	15	
1.0	40	25	5	15
30.0	4	3	1.5	2

(CFE: cold field emission; TFE: thermal field emission). The spatial resolution depends on the type of electron gun, the electron-optical system and the extent to which the material interacts with the electron beam. The values are indicative and depend on the instrument

achieved by the first anode. The work of the second anode is to accelerate the electrons to 100 kV or more. The combined fields of the anodes act like a more refined electrostatic lens to produce a crossover. The second anode controls the effective probe size and position. In this system, the tungsten is operated at ambient temperature and that is why the process is called a *cold field emission*. This requires a high vacuum of about 10^{-9} Torr, preventing the tungsten tip from rapid surface contamination. The vacuum can be lowered when heating the tip. The thermal energy assists in the electron emission and the surface contaminations are not the same. For such *thermal field emission*, the tungsten tip is coated of zirconia (ZrO_2), which lowers the work function of the tungsten and therefore enhances the electron emission. This makes it possible to use a broader tip than for the cold FE. This type is called *Schottky* emitters.

Electron Column

Once the electron beam has left the electron gun, it is guided to the sample by a series of electromagnetic lenses. The lenses direct and condense the electron beam to its final size of about 2–80 nm (depending on the acceleration voltage). The electron beam is divergent after passing through the anode plate and must be collimated by condenser lenses and apertures into a relatively parallel stream. This is achieved by electromagnetic lenses. In fact, the lenses beyond the electron gun demagnify the image of the crossover in the electron gun. Electrons can be directed and focused by electrostatic fields (e.g., electron gun) or magnetic fields. Most electron microscopes are equipped with electromagnetic lenses because of their lower inherent aberrations. The resolution of a SEM is a function of the final electron beam diameter and the acceleration voltage and, as a consequence, of the type of electron gun. The resolution of a FE-SEM is better than that of a thermoionic gun SEM (Table 1). The limit of resolution is reached when the current within the electron probe is insufficient to produce a usable signal. Given the small size of the probe beam, the depth of field is high compared to a conventional light microscope (hundreds of times more).

Lenses. The lenses in electron microscopes are electromagnetic lenses (except the gun lens in the FE gun that is an electrostatic lens). A magnetic lens consists of a coil of copper wires inside the iron pole pieces. A current through the coils creates a magnetic field (force lines are symbolized by the circles in Figure 5) in the bore of the pole pieces. Where the circuit is interrupted (the gap), the magnetic field goes out into the vacuum and creates the lens field that is used for focusing the electron beam.

The rotationally symmetric magnetic field is inhomogeneous, i.e., it is weak in the center of the gap and becomes stronger close to the bore. Therefore, electrons close to the center are less strongly deflected than those passing the lens far from the axis. The overall effect is that a beam of parallel electrons is focused into a spot (so-called *crossover*). An electron passing through this field perpendicular to these field lines moves in a curved trajectory because there is a force on the electron created by the electron's movement in the magnetic field. It results in a helical trajectory and a rotation of the image. The focal length of the lens is controlled by the current in the lens coil. The focal length decreases as the current increases (the electrons are more deviated) or the acceleration voltage decreases.

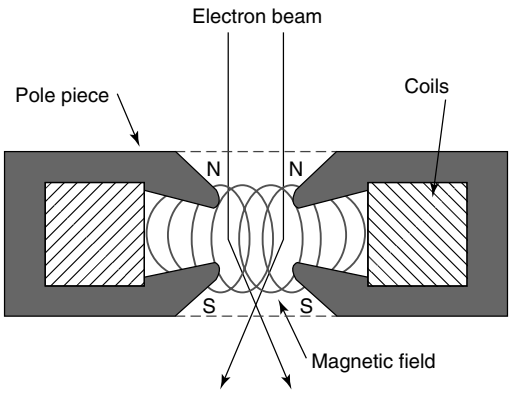


Figure 5 Schema of an electromagnetic lens (cross section)

The SEMs employ one or several condenser lenses to demagnify the diameter of the electron beam. The final lens in the column is the objective lens. Its role is to focus the image by controlling the movement of the probe crossover along the optical axis (Z-axis) of the column. The design of this lens generally incorporates space for the scanning coils, the stigmator (for correction of astigmatic lens defaults), and the beam-limiting aperture. This lens either includes the specimen (i.e., the specimen is placed inside the lens) or not (i.e., the specimen is placed outside the lens, in the case of large specimens).

Scanning Devices. In a SEM, the electron beam is not stationary on the specimen but moves in a raster on the specimen surface. The electron probe scans across the specimen in two perpendicular (X and Y) directions.

The deflection system is integrated in the last lens, the objective lens, or is situated just above it (Figure 2). The scanning coils, mounted in sets of two above another, create magnetic fields (positive and negative). The electrons in the beam will be attracted by the positive field and repelled by the other, leading to a deflection toward the positive coil. The beam moves over the specimen surface line by line in an X, Y raster covering a rectangular area. The raster is generated by the line and frame generators that feed the scan coils with the appropriate current leading to magnetic fields of different strength according to the scan size. The scan process is continuous and once one sequence is finished, a new frame is immediately started.

The output of the two scan generators is also applied to the deflection coils of a display device, typically a TV-type cathode-ray tube (CRT) on which the SEM image will appear. Since the electron beams in the SEM column and in the CRT are scanning in synchronism, for every point on the specimen (within the raster-scanned area) there is a corresponding point on the display screen. In a modern SEM, the scan signals are generated digitally by computer. The image is acquired pixel by pixel.

The signal which modulates the image brightness can be derived from any property of the specimen which is caused by (or changes in response to) electron bombardment (cf. below "image mechanism").

Magnification Process. It is convenient here to consider the magnification process. Unlike light microscopes or TEMs, the magnification is not achieved by means of lenses but is the ratio of the linear size of the viewing screen to the linear size of the raster on the specimen. The image on the screen has a fixed size given by the display screen and is bigger than the raster area on the specimen; the magnification (M) is

$$\mathrm{M} = \frac{\text{Scan distance* in the image}}{\text{Scan distance* on the specimen}}$$

(* length of the scanned line)

Magnification is therefore controlled by the current supplied to the X, Y scanning coils, and not by objective lens power. The magnification of a SEM ranges from $10\times$ or less to $1\,200\,000\times$ or more (depending on the SEM type and the operation mode, i.e., high, low, or environmental vacuum) (Figure 6).

Sample Chamber

Unlike the sample stage of a conventional optical microscope, the sample chamber in a SEM is a confined space and cannot be directly operated by hands. The chamber is constructed in order to be directed by the means of devices allowing moving, tilting (range of -5 to $\pm 70^{\circ}$), and rotating the specimen. This requires a very fine tuned mechanical system that provides small and continuous motions (mm to $<0.01\,\mu\mathrm{m}$) in the X, Y, and Z directions as well as

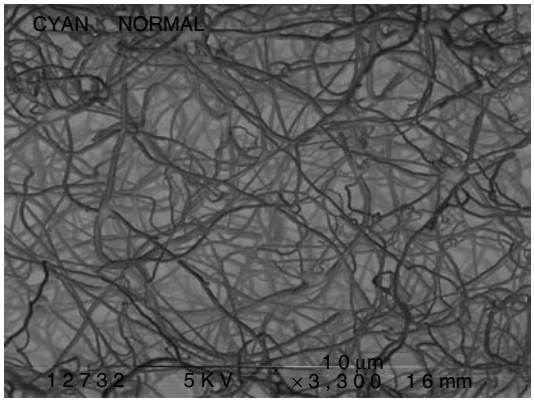


Figure 6 Cyanoacrylate deposit on a fingerprint ridge

an almost absolute stability of the object. The eucentric goniometer stage is nowadays fully motorized. The size of the chamber determines the size of the specimen. In general, forensic SEMs have a rather big chamber (up to around 400 mm left to right) in which a big object can fit and be moved (moving area up to about 150 mm × 150 mm for X/Y movements and up to 65 mm for Z movements). Standard SEMs have a rather small chamber, allowing movements of 10 mm × 10 mm. If the sample to submit has bigger dimensions, the area of interest has to be cut out (modification of the evidence). Small objects are normally mounted on specimen holders of different shapes and sizes (often with diameters of around 12 mm) according the type of stage. The chamber is constructed in order to withstand the necessary vacuum for the electron beam (high vacuum around 10^{-4} Torr or more). It also hosts all devices necessary for the caption and processing of the output signals.

Electron Interaction and Image Mechanisms

The incident electron beam will generate several reactions once it penetrates the specimen surface. The beam strikes the specimen point by point. Several interactions occur, providing the signals that form the image or give material information about the specimen. The incident electrons interact both elastically and inelastically with the specimen, forming the limiting interaction volume from which the various types of radiation emerge, including backscattered, secondary and absorbed electrons, characteristic and Bremsstrahlung X-rays,

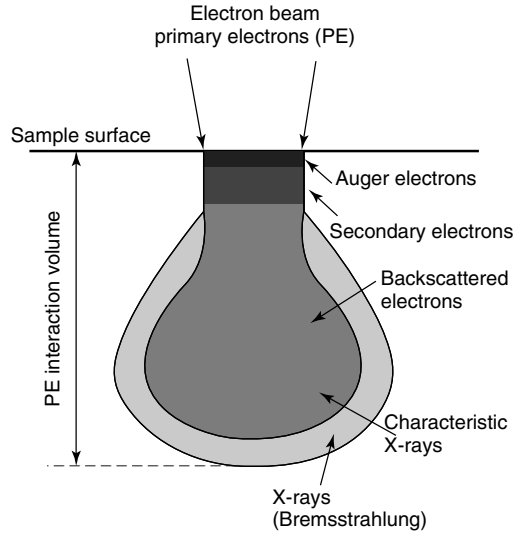


Figure 7 The interaction volume depends on several parameters (acceleration voltage, atomic number of the sample elements, and the sample density)

and in some materials, cathodoluminescence radiation (emission of photons), and Auger electrons (Figure 7).

By measuring the magnitudes of these signals with suitable detectors, a determination of certain properties of the specimen (topography, composition, crystallography, electrical conductivity, etc.) can be made at the single location where the electron beam strikes. All interactions occur together but can be selected individually by using the appropriate detector. Not all detectors are usually present on a single instrument. The size, depth, and shape of the interaction volume depend on the density of the material being investigated and the beam velocity (acceleration voltage) (Figure 8). The interaction volume takes a shape between a pear or teardrop and a hemisphere. The resulting signals are either electrons or photons. The first category is primarily used for the image and the second category of signals is mainly taken for spectroscopic information.

Imaging Signals. Secondary electrons (SEs) are the most common signal used for investigations of surface morphology. They are produced as a result of interactions, i.e., inelastic scatter between primary electrons (incident beam electrons) and the weakly bound electrons in the conduction band of the sample.

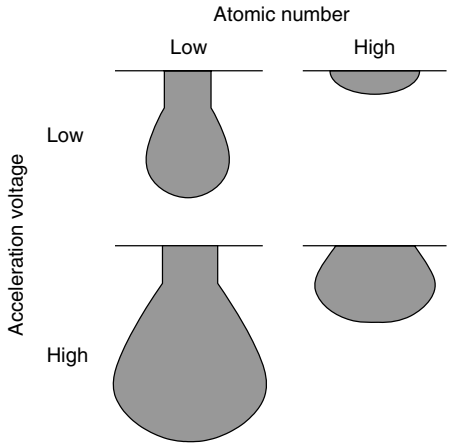


Figure 8 Influence of the acceleration voltage of the primary electrons in function of the atomic number of the sample elements

These electrons are knocked out of their orbitals and, if sufficient near the sample surface, escape as SEs. The SEs are low-energy electrons (<50 eV) and only those formed near the sample surface (within a few nanometers) have enough energy to escape and to be detected (Figure 7). The energy is obtained when they collide with a primary electron.

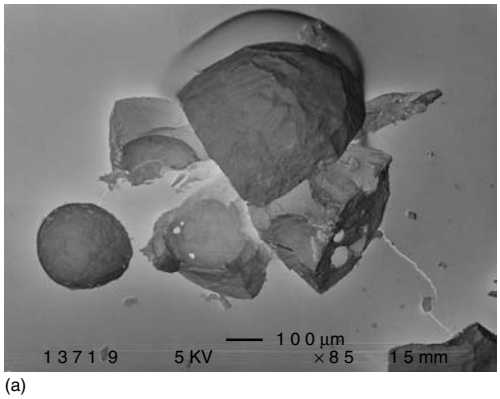
The other phenomenon that occurs is the elastic scatter between the primary electrons and the atoms of the sample. The electrons are diffracted by the atomic fields within the specimen and backscatter out of the sample surface. Because their trajectory changes they might travel out of the specimen. The backscattered electrons (BSEs) have energy of up to that of the primary electrons and are emitted from approximately the top 40% of the area of the teardrop or hemisphere (Figure 7). The BSE come from deeper regions in the specimen and because of the diffraction (spread of the electron beam) within the specimen material the image resolution is inferior.

An SE reproduces the morphology of the surface, whereas the BSE creates shadows (topography information) because of the straight trajectory and atomic number contrast (composition information) (Figure 9a, b).

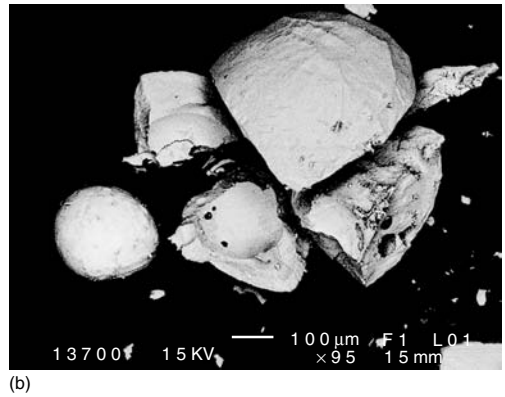
BSE are often used in analytical SEM along with the spectra made from the characteristic X-rays. Because the intensity of the BSE signal is strongly related to the atomic number (Z) of the specimen, BSE images can provide information about the distribution of different elements in the sample as well as the crystallographic structure of the specimen.

The caption of both the signals (SE and BSE) can be achieved by one or two separate detectors. The most commonly used detector is the *Everhart–Thornley electron detector*, which is sensitive to both the signals [4]. It consists of a scintillator, a light pipe, and a photomultiplier tube. This type of detector is not very efficient for BSE signals. That is why often a second dedicated detector is installed for the caption of only the high-energy BSE. Scintillator backscatter detectors, solid-state diode detectors (semiconductors), or channel plate detectors can be fitted.

The amplified electrical signal output is displayed as a two-dimensional intensity distribution on an analogue cathode ray tube (monitor), or subjected



(a)



(b)

Figure 9 Comparison between a SE and a BSE image: the surface details of soldering particles in SE mode (a) disappear in BSE mode (b)

to analogue-to-digital conversion and displayed and saved as a digital image. The brightness of the signal depends on the number of SE and/or BSE electrons reaching the detector.

Spectroscopic Signals: X-Rays, Auger Electrons, and Light. The other signals (characteristic X-rays, cathodoluminescence radiation, and Auger electrons) that occur when bombarding a material with an electron beam are used less for imaging purposes than they are for analytical needs.

The most frequently used analytical method in association with the SEM is an X-ray spectrometer. X-rays are emitted when the electron beam removes an inner shell electron from the sample, causing a higher energy electron to fill the shell and release energy in form of photons (X-ray) or Auger electrons. These X-rays are characteristic of the difference in energy between the two shells and of the atomic structure of the element from which they are emitted. Therefore, these X-rays are characteristic for each element and are used to identify the composition and measure the abundance of elements in the sample (Figure 7). The emitted X-rays are detected and processed by two methods by using two characteristics associated with light (wave and energy quantum). The same applies to X-rays. They can be detected either by their wavelength or by their energy. The latter technique, energy dispersive X-ray spectroscopy (EDX), is most commonly associated to the SEM. In the case of the detection of the X-ray energy, the photons are detected by a semiconductor (silicon doped with lithium) that converts the energy of the X-ray into a proportional electrical charge. This analogue impulsion is amplified, digitalized, and then treated in a multiple channel analyzer. Each impulsion is attributed according to its charge to a predefined channel. The information is accumulated in a histogram and displayed on a monitor. The simultaneous counting of all the photons in the range of 0–20 keV allows to acquire in one step a spectrum of all elements, within limits, of a sample (Figure 11b). This gives a quick overview of a sample composition. However, the accuracy of EDX spectrum is affected by several limitations. The qualitative analysis does not include all the elements from the periodic table. It is restricted from boron ($Z = 5$) to uranium ($Z = 92$). Another limitation is given by its lack of specificity, with respect to the oxidation state of complex anions. Quantitative analysis is compromised by secondary fluorescence (higher energy X-rays excite the emission of X-rays of lower energy in a sample) causing the emission of more X-rays of an element than really present and overlapping peaks. A semi-quantitative analysis is more appropriate for this type of X-ray analysis, contrary to the wavelength dispersive X-ray spectroscopy (WDX).

This method allows the measuring of the wavelength of the emitted X-ray from the sample. X-ray diffraction is a powerful tool to study both the X-ray spectra and the arrangement of atoms in crystals. The X-ray wavelength is determined by irradiating a known single crystal at a precise angle. The single crystal diffracts the photons in accordance to the Bragg's law and is collected by a detector (proportional counter or scintillation counter). The crystal is positioned in equal distance to the specimen and to the detector. The crystal with known spacing is chosen. These planes reflect different wavelengths at different angles. The detector can discriminate one angle from another and attribute the corresponding wavelength. Only one element at the time can be determined. WDX counts only the X-rays of a single wavelength, not producing a broad spectrum of wavelengths or energies. Several crystals are necessary to cover the whole range of X-rays. This generally means that the element must be known to find a crystal capable of diffracting it properly. Although this technique is more sensitive than the EDX, it is less applicable in forensic science, because the sample composition has to be known and its surface has to be plane and polished, parameters that are encountered only rarely among forensic samples.

Although SEM-EDX has few restrictions, it is a very versatile and efficient technique for elemental analysis. The combination of high magnification and the possibility to analyze the target is very useful, especially in forensic science.

The next two spectroscopic techniques are encountered less in forensic applications, but they can be associated with the SEM. Instead of X-ray emission, the excess energy is transferred to a third electron from a further outer shell, prompting its ejection. This ejected species is called an *Auger electron* (Figure 7), and the method for its analysis is known as *Auger electron spectroscopy* (*AES*). The Auger electron has a specific characteristic energy that depends on the type of atom and the chemical environment in which the atom was located. Auger spectrometry is a widely used surface analysis technique.

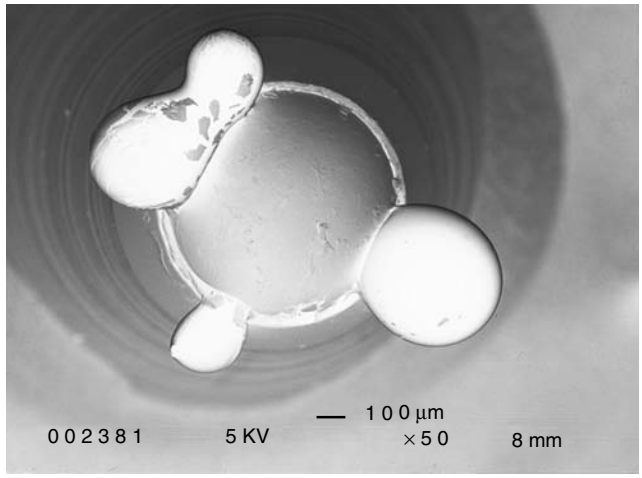


Figure 10 The influence of a high vacuum on a wet sample: the ink contained in the ink reservoir is pressed out of the ball housing and forms the drops around the tip of a ballpoint pen

Inelastic scatter can generate the emission of light (UV–visible–IR), the so-called *cathodoluminescence*. This phenomenon occurs when an electron is promoted from the valence band into the conduction band, leaving behind a hole. When an electron and a hole recombine, it is possible for a photon to be emitted. It happens for materials with a band gap structure and is useful for the characterization of semiconductors, insulators, certain minerals, glasses, and biological specimens. The radiation is detected with an appropriate photomultiplier for the photon-energy range of interest.

Sample Preparation

As mentioned above, classical SEMs operate in a vacuum environment and with an electron beam. This has a dramatic influence on a sample that cannot withstand the pressure of about 10^{-4} Torr or more (e.g., hydrated, oily, and out gassing samples) and/or is not conductive (Figure 10).

The chamber size might also have its influence and need a resizing of the sample, which might be a handicap for forensic investigations (disruption of the sample integrity). The problem of non-conductivity can be overcome by applying a conductive coat to the sample. Non-oxidation metals such as gold, platinum, palladium, and carbon are used. They are conductive and chemically stable (no oxidation) and inert. The metals are deposited on the sample by low-vacuum sputter coating or by high vacuum evaporation (carbon). The coat is some nanometers thick. Carbon coating is appropriate if the sample has not only to be visualized but also to be analyzed (carbon is not detected by EDX). The coating prevents the accumulation of static electric charge on the specimen during electron irradiation. Another reason for coating, even when there is more than enough conductivity, is to increase signal and improve contrast and resolution. For hydrated samples, they have to be dried or frozen before being submitted to the chamber vacuum, which may modify its morphology. If coating, drying, or freezing is required, then the SEM observation has to be done at the end of an observation sequence. Optical characteristics are altered by the coat. To overcome both the problems (vacuum incompatibility and insulating surface) without modifying the sample, one has to use a low-voltage beam of the FE-SEM and/or an ESEM. The latter even allows visualizing the samples in a nearly atmospheric pressure.

Environmental SEM

The ESEM has been invented with the aim to develop a SEM that accepts virtually any specimen regardless of composition and provides a high-resolution SE image (Figure 11a). The ESEM unites all this requirements. Commercial ESEMs have been introduced in the late 1980s and early 1990s. Ten years of research

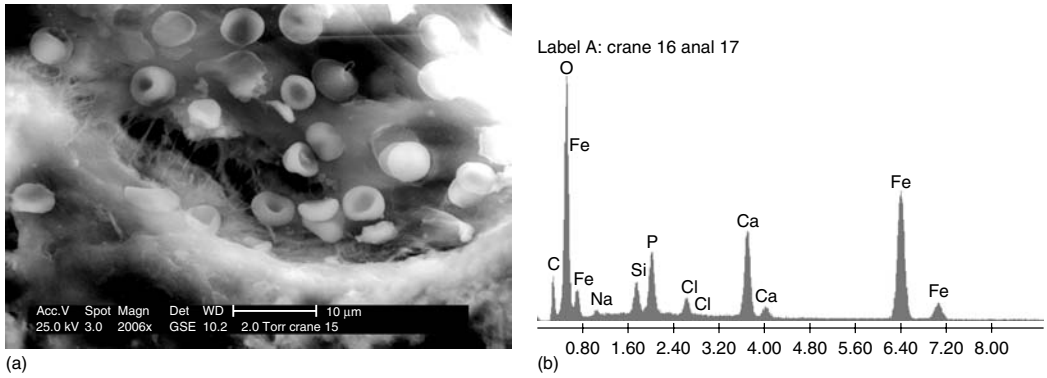


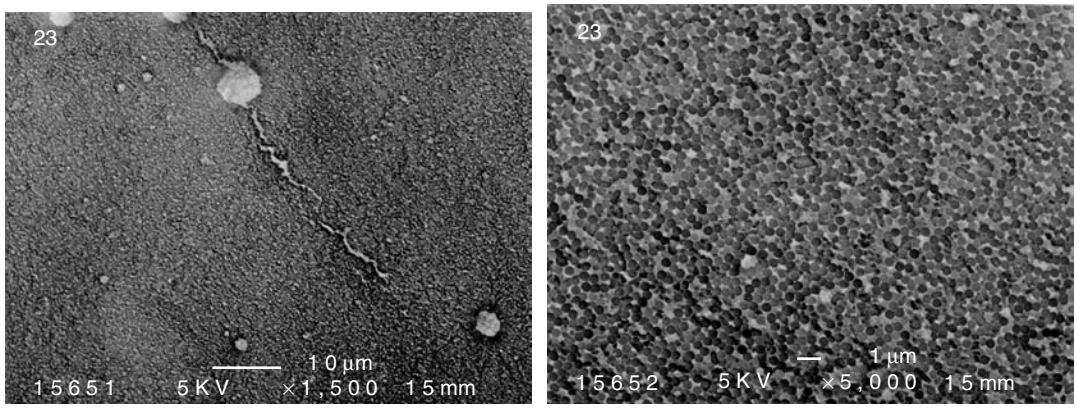
Figure 11 ESEM image (a) of hemoglobin cells around a crane fracture and their EDX analysis (b)

and development were necessary for the creation of such an instrument as well as for the discovery of a new signal detection method [5]. Although the ESEM has partially the same setup as the conventional SEM, the main problem was to find a solution allowing the use of an electron beam in a high-pressure ambience without contaminating the electron gun chamber and the electron column and ultimately to still detect the emitted signals (SE, BSE, X-rays, etc.). The problem of keeping the upper parts (gun and electron column) of the SEM in high vacuum is overcome through the use of a differential pumping technique. A series of chambers at increasing pressure along the beam path are linked by small apertures. Dedicated pumps at each stage maintain the required pressure gradient and stop contaminants from reaching the clean upper column. The sample chamber can then be maintained at a relatively high pressure (10–20 Torr) while the electron optical column is differentially pumped to keep the vacuum adequately high at the electron gun. Several stages of differential pumping system are built in the final aperture lens. The pressure is regulated by the flow of gas (inert gas and/or water vapor). The high-pressure region around the sample in the ESEM neutralizes charge and provides an amplification of the SE signal. Positively charged ions generated by the beam interactions with the gas help to neutralize the negative charge on the specimen surface. The pressure of gas in the chamber can be controlled, and the type of gas used can be varied according to the need. The detectors of the emitted signals are different from those described for the conventional SEM. The conventional SE detector (Everhart–Thornley detector) cannot be used in the presence of gas because of an electrical discharge (arcing) caused by the kilovolt bias associated with this detector. In lieu of this, an environmental secondary detector (ESD) is used [6]. It takes advantage of the ionization behavior of the low-pressure gases found in the specimen environment. Gas ionization is induced by a moderate electric field, forcing collision between highly mobile SEs and neutral gas molecules. It results in a multiplication of the secondary signal and at the same time the positive ions neutralize the electrical charges at the specimen surface.

For the backscattered electron, the conventional BSE detection means have been adapted to operate in the gaseous conditions of the ESEM. The characteristic elemental X-rays also produced in the ESEM can be detected by the same detectors used in the conventional SEM (Figure 11a, b).

Modern ESEMs have a "three-in-one" concept. Three different vacuum modes (high, low, and environmental vacuum) can be chosen with one instrument according to the specimen vacuum sensitivity. They are equipped with the conventional detectors (when working in high vacuum; 10^{-4} Torr and more an Everhart–Thornley for SE, solid state for BSE) as well as the dedicated environmental detectors as described for low vacuum. These instruments still have an excellent resolution range (3.5 nm at 30 kV in all vacuum modes, <15 nm in low-vacuum mode at 3 kV).

The possibility to work in a vacuum of about 20 Torr makes it possible to add some optional devices such as a Peltier cooled stage (−20 to +50 °C), a heating stage (for temperatures up to 1500 °C) and



(a) (b)

Figure 12 Aspect of a gel pen ink stroke (a) and (b) under different magnifications

a cryo transfer system (for *in situ* investigations) for the observation of sample behavior under temperature changes.

Forensic Utility of SEMs

With the commercialization of ESEM, there is no limit to its use in forensic science. The ESEM has demonstrated its near universal applicability, combining the high resolution and depth of field of the SEM with the flexibility and ease of the light microscope. Whenever something small has to be studied, the SEM is certainly the microscope to choose [7]. The only limit is that optical properties and characteristics cannot be studied, as well as the aspect of the sample in transparency. In case of a non-pigmented water-based ink stroke leaving no deposit on the paper (when the ink is completely absorbed by the fibers), it cannot be seen by SEM. This limits the application for line crossing problems [8]. See also the relevant article about intersecting lines in this encyclopedia.

The magnification range from macro to micro allows zooming in and out of details of the sample, with an extreme ease and with a resolution and depth of field never obtained by the conventional light microscopes (Figure 12a, b).

The additional spectroscopic tools make the SEM a powerful analytical instrument, allowing a target compositional analysis. The area of interest can be selected and analyzed. This makes the technique essential for gunshot residues (GSRs), because of its capability to analyze discrete particles in the sample. The analysis is nowadays fully automated and the position of each analyzed particle is recorded. This makes the localization and verification of each analyzed particle very convenient. An advantage is that the particles are directly taken from the support (e.g., suspect) by the SEM holder equipped with a carbon-coated adhesive. No further sample preparation is required prior to the SEM/EDX analysis. This minimizes the risk to lose particles while handling.

The elements constituting the GSR can originate from the primer, the bullet, a coating or jacket on the bullet, cartridge components, and previous residues in the barrel. The characteristic particles for Sinoxid primers contain the metal elements lead, barium, antimony, and calcium [9, 10]. The heavy metal particles in a lead-free primer (e.g., Sintox™ ammunition commercialized by Dynamit Nobel AG) are replaced by the elements titanium and zinc [10, 11]. Particles from the bullet are composed of copper, zinc, nickel, mild steel, and lead (in metallic form).

The samples usually contain many other particles that are of no particular interest (contaminations from the support); the particles of interest have to be "selected" before being analyzed. This is done by the backscattered detector, which is calibrated in relation to the Z (atomic number) and not to the video intensity. The particles are detected based on their average atomic number falling within a predefined Z range in relation to primer compositions, which tend to have particles with intrinsically high average Z. Therefore, searching for "bright particles only"

is eliminated. The older systems are based on this method, which of course lacks selectivity. The GSR particles are 0.5–10 μm in size (usually 0.5–2 μm). Therefore, the automated search and analysis of a sample, i.e., a defined area on the sample holder surface, take quite some time but is much faster and reliable than a manual search of primer residue particles.

In case of a positive result (i.e., particles containing the above-mentioned elements in combination), its interpretation is not evident. The combinations Pb/Sb/Ba and Sb/Ba are considered to be characteristic for Sinoxid primer residues [12]. However, other sources might produce particles with same combinations [13–15]. In view of these findings, the morphological aspect of the found particles becomes important and might allow to support stronger the hypothesis of GSR particles than particles generated from another source than a firearm. GSR particles have a spherical morphology. Not to forget also the memory effect when firing different ammunitions from the same firearm that might produce mixed compositions of GSR particles [10, 11, 16]. More details about GSR can be found as a separate specific contribution in this encyclopedia.

The combined abilities of the SEM to resolve fine structures and determine their elemental composition are an indispensable aid in the examination of small items of trace evidence. There is no doubt that the SEM is extremely useful and the list of its forensic applications is endless. However, the only inconvenience of the technique is still its price.

References

[1] Ruska, E. (1934). Ueber Fortschritte im Bau und in der Leistung des magnetischen Elektronenmikroskopes, *Zeitschrift für Physik* **87**, 580–602.

[2] Ruska, E. (1980). *The Early Development of Electron Lenses and Electron Microscopy*, S. Hirzel Verlag, Stuttgart.

[3] Goldstein, J., Newbury, D., Echlin, P., Joy, D., Lyman, C., Echlin, P., Lifshin, E., Sawyer, L. & Michael, J. (2003). *Scanning Electron Microscopy and X-ray Microanalysis*, Springer, New York.

[4] Everhart, T.E. & Thornley, R.F.M. (1960). Wide-band detector for micro-microampere low-energy electron currents, *Journal of Scientific Instruments* **37**, 246–248.

[5] Danilatos, G.D. (1988). Foundations of environmental scanning electron microscopy, *Advances in Electronics and Electron Physics* **71**, 109–250.

[6] Danilatos, G.D. (1990). Theory of the gaseous detector device in the ESEM, *Advances in Electronics and Electron Physics* **78**, 1–102.

[7] Mazzella, W.D. & Khanmy-Vital, A. (2003). A study to investigate the evidential value of bluer gel pen inks, *Journal of Forensic Sciences* **48**, 419–424.

[8] Khanmy-Vital, A., Kasas, S. & Dietler, G. (2001). The use of atomic force microscopy to determine the sequence of crossed lines, *Problems of Forensic Sciences* **66**, 401–412.

[9] Zeichner, A. & Levin, N. (1997). More on the uniqueness of gunshot residue (GSR) particles, *Journal of Forensic Sciences* **42**, 1027–1028.

[10] Khanmy, A. & Gallusser, A. (1995). Influence of weapon cleaning on the gunshot residues from heavy metal free ammunition, *Advances in Forensic Sciences* **3**, 60–65.

[11] Gunaratnam, L. & Himberg, K. (1994). The identification of gunshot residues particles from lead-free Sintox ammunition, *Journal of Forensic Sciences* **39**, 532–536.

[12] Wallace, J.S. & McQuillan, J. (1984). Discharge residues from cartridge-operated industrial tools, *Journal of Forensic Science Society* **24**, 495–508.

[13] Mosher, P.V., McVicar, M.J., Randall, E.D. & Sild, E.H. (1998). Gunshot similar particles produced by fireworks, *Canadian Society of Forensic Science Journal* **31**, 157–168.

[14] Garofano, L., Capra, M., Ferrari, F., Bizzaro, G.P., DiTullio, D., Dell'Olio, M. & Ghitti, A. (1999). Gunshot residue – further studies on particles of environmental and occupational origin, *Forensic Science International* **103**, 1–21.

[15] Torre, C., Mattutino, G., Vasino, V. & Robino, C. (2002). Brake linings: a source of non-GSR particles containing lead, barium and antimony, *Journal of Forensic Sciences* **47**, 494–504.

[16] Zeichner, A., Levin, N. & Springer, E. (1991). GSR particles formed by using different types of ammunition in the same firearm, *Journal of Forensic Sciences* **36**, 1020–1026.

AITA KHANMY-VITAL

Mini-STRs

Disclaimer: The opinions and assertions contained herein are solely those of the authors and are not to be construed as official or as views of the US Department of Defense, the US Department of the Army, or the Armed Forces Institute of Pathology.

For well over a decade, short tandem repeat (STR) markers have played an important role in advancing the field of forensic DNA typing (*see* **Short Tandem Repeats**). In the United States, there are two commercial companies that produce multiplex STR kits used by forensic laboratories: Applied Biosystems (Foster City, CA) and Promega Corporation (Madison, WI). Each company has produced a "megaplex" autosomal STR kit containing 16 markers: the 13 Combined DNA Index System (CODIS) loci [1, 2]. 2 markers specific to each company's kit, and the amelogenin marker for sex determination. Both 16-plex kits have amplicon sizes ranging from 100 to 450 base pairs (bp). The availability of commercial kits has helped to standardize the STR markers used by the forensic community.

The conventional STR kits perform well when an optimal quantity (approximately 1 ng) of high-quality DNA is used, exhibiting peak height balance both within loci (heterozygous alleles) and between each marker. However, biological evidence at crime scenes is often exposed to the elements and/or microbial agents that may cause DNA to degrade. Inhibitor molecules, such as heme in blood or humic acid in soil, can also copurify with DNA and prohibit the generation of a full STR profile. Degradation and inhibition may result in amplification failure especially of high-molecular weight loci, resulting in a "partial profile" in which only a subset of the core 13 CODIS loci are obtained (*see* **DNA: Degraded Samples**). The loss of multiple forensic markers as a result of degradation and inhibition reduces the overall statistical significance of any observed match. If a biological sample is too highly degraded for STR analysis, the forensic scientist may have to turn to mitochondrial DNA (mtDNA), which can be expensive and time consuming to test, and may result in limited information, especially for common mtDNA haplotypes (*see* **Mitochondrial DNA: Profiling**).

One strategy to recover the genetic information lost due to DNA degradation would be to reduce the size of the PCR (Polymerase Chain Reaction amplicons for the "larger" loci. By designing PCR primers to bind closer to the core tandem repeat (Figure 1), it is possible to create a smaller PCR product, a "mini-STR", while still retaining the core repeat information. The reduced-size PCR amplicon then has a greater chance (relative to conventional STR primers) of generating a profile when genetic material is degraded.

Emergence of Mini-STRs as a Forensic Tool

During the mid-1990s, Dr John Butler (presently at the National Institute of Standards and Technology, NIST) and his colleagues at GeneTrace Systems developed a system to rapidly genotype STR profiles using matrix-assisted laser desorption/ionization time-of-flight (MALDI/TOF) mass spectrometry. The rapid detection of multiple unlabeled PCR products using MALDI/TOF represented a significant reduction in sample processing and analysis time as compared to standard STR profiling. One limitation of the MALDI/TOF technology, however, was related to the size of PCR fragment analyzed; amplicons larger than 140 bp were difficult to resolve with the method. To make the commonly used forensic STR

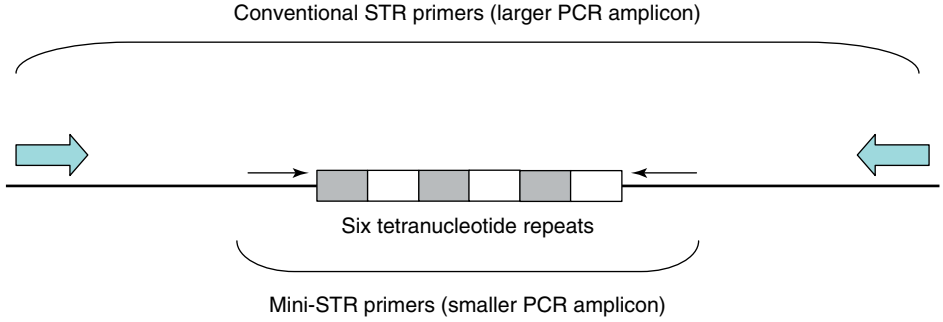


Figure 1 An example of a mini-STR compared to conventional STR primers. The mini-STR primers are adjacent to the six nucleotide repeats producing a smaller PCR product compared to the conventional STR primers

markers more amenable to MALDI/TOF detection, Butler and colleagues redesigned the primers to create smaller amplicons. Significantly, Butler and colleagues noted in their 1998 publication that these reduced-size amplicons could be utilized to increase the recovery of STR profiles from degraded samples [3, 4].

A few years later, laboratories began publishing information on reduced-size PCR amplicons for (mostly CODIS) loci [5–8]. These early studies demonstrated the forensic potential of mini-STRs, sometimes with very significant results. For example, Hellmann *et al.* [5] reduced the amplicon size of the TPOX locus by 160 bp to test telogen hairs. This group was able to recover the correct genotype over 77% of the time compared to only about 18% of the time with conventional primers – a fourfold improvement.

Additional advances in mini-STR research and application also came as a result of the terrorist attack on the World Trade Center (WTC) towers on September 11, 2001. Shortly after the attacks, the New York Office of the Chief Medical Examiner (NY-OCME) was charged with the task of identifying the nearly 3000 victims of the disaster. Knowing that many of the samples recovered from the crime scene would be highly degraded and of poor quality, the NY-OCME investigated new technologies, such as single nucleotide polymorphism (SNP) typing and reduced-size STR amplicons, to increase the success rate on the compromised samples. Through work at NIST and a private laboratory (the Bode Technology Group) all of the core CODIS markers were redesigned as mini-STRs [9]. Protocols were developed and validated at NIST [9], and samples from the disaster were tested at the Bode Technology Group [10]. Mini-STRs proved to be valuable for the identification efforts: nearly 20% of the 850 DNA-only identifications from the WTC disaster were made using mini-STRs [11].

Following the success of mini-STRs in recovering genetic information from CODIS loci, additional research into other markers of forensic interest has included autosomal loci unlinked to the CODIS markers [12]. Additional markers not linked to the CODIS loci can be useful for resolving complex paternity cases such as incest. A set of X-chromosome mini-STRs [13] have been developed for select forensic scenarios where linked X-STR markers can provide additional discrimination relative to autosomal loci. Finally, Y-chromosome mini-STRs [14] have been developed to amplify male-specific loci (*see* **Y-Chromosome Short Tandem Repeats**).

In 2007, the first commercial mini-STR kit, AmpFℓSTR® MiniFiler™ (Applied Biosystems), became available to the forensic community. A 9-plex, MiniFiler™ includes amelogenin and eight of the largest loci from the AmpFℓSTR® Identifiler® and SGM Plus® (Applied Biosystems) STR typing kits. The MiniFiler™ amplicon sizes generally range up to 250 bp (290 bp for the extended FGA locus alleles), and represent an approximate 30–200 bp size reduction as compared to the Identifiler® and SGM Plus® amplicons for the included loci. In addition to the reduced-size amplicons, the MiniFiler™ amplification protocol, which increases (relative to other AmpFℓSTR® kits) the number of thermal cycles to 30, and proprietary amplification buffer, which already contains the DNA polymerase, are kit optimizations designed to overcome inhibition and recover loci from compromised samples. MiniFiler™ is thus intended for use as a supplement to other megaplex STR kits when an insufficient number of loci are recovered, or as a stand-alone kit when extremely limited sample quantities would prevent multiple attempts to develop a genetic profile [15].

A second commercial mini-STR kit, PowerPlex® S5 (Promega Corporation), is also now available to forensic practitioners. Like MiniFiler™, the S5 kit utilizes small amplicons (under 260 bp) and a proprietary amplification buffer optimized to overcome inhibition. In contrast to MiniFiler™, however, PowerPlex® S5 includes only four loci in addition to amelogenin. Rather than as a means to recover additional loci when large-amplicon loci fail, the S5 kit is marketed as a reliable, lower-cost exclusionary tool for criminal casework and population screening.

Challenges to Mini-STR Development

With the availability of commercially available mini-STR kits, many forensic laboratories can now take advantage of using these markers on casework samples without the need to produce "in-house" multiplexes that must pass high quality control standards for use in forensic casework. However, challenges remain to the development and use of both in-house

and commercial mini-STR kits. These include repeat size limitations, limited multiplex real estate, and the potential for profile discordance as a result of primer design.

Some of the current CODIS loci are difficult to develop as true mini-STRs because of the size of the repeat region. For example, the FGA locus has a rather large allele spread – ranging from 17 to 33 repeats (and up to 51 repeats among the extended alleles). A large allele spread is advantageous for a high-diversity marker, but successful amplification of a 50 tetranucleotide repeat (200 bp) along with an additional 40–50 bp for the flanking primer sequence (240–250 bp total amplicon) in a highly degraded sample is unlikely. Ideally, mini-STR multiplexes would be composed of markers with amplicons sizes no larger than 150 bp.

However optimal from a profile recovery standpoint, the restriction of mini-STRs to amplicons sizes less than 150 bp greatly limits the number of markers that can be tested in a four or five fluorescent dye system. To overcome the limited multiplex real estate, a few multiplexes, with each dye channel containing 1–2 mini-STR markers, could be constructed for use in tandem. However, a disadvantage to typing with multiple assays is the consumption of additional, often limited, DNA template. This could be especially problematic when typing samples under "low copy number" conditions where multiple replicates are required for data analysis and interpretation [16, 17] (*see* **Low Copy Number DNA**). Although a greater number of fluorescent dyes could be used to increase the number of loci amplified in a single reaction, this is not at present a viable option because of capillary electrophoresis (CE) detection limitations.

Alternately, nonnucleotide linker molecules that shift the electrophoretic mobility of the amplified product could be used to construct a mini-STR multiplex with more than 1–2 markers per dye channel. For example, suppose that two mini-STR loci, A and B, have the amplicon size ranges 72–112 and 96–136 bp, respectively. In constructing a multiplex, it would be necessary to position these two loci in separate dye channels since there is potential for overlap between the higher molecular weight alleles of locus A and the lower molecular weight alleles of locus B. Using non-nucleotide linkers equivalent to approximately 2.5 bp [18], the addition of 10 such linkers between the primer and the fluorescent dye for locus B would shift the apparent size of the amplicon by +25 bp. As a result, locus B would migrate at 121–161 bp and provide enough separation to be included in the same dye channel as locus A. Although non-nucleotide linker molecules can increase the electrophoretic separation between overlapping loci as described, there is a limit to the number of linkers that can be added to a primer and still result in consistently high-quality data.

Primer design presents an additional challenge to the development of mini-STRs from standard STR loci. A necessity for the creation of reduced-size STR amplicons is a clean sequence region flanking the repeat motif. Not all of the industry-standard STR markers have clean flanking regions, which can make the design of mini-STRs a challenge, and may result in nonconcordant alleles when typing the same loci with different STR multiplexes. For example, the CODIS marker D7S820 has a polyT stretch found 13 nucleotide bases downstream of the core tetranucleotide repeat. Designing a primer adjacent to the repeat and avoiding the polyT repeat would potentially miss a number of microvariants created by the addition (X.1) or subtraction (X.3) of thymidine nucleotides in this region. In other words, if a sample tested with a conventional kit typed as a 15.1 allele, a mini-STR designed to avoid this polyT region (through placement of an amplification primer between the core repeat and the T-stretch) would result in a 15 allele.

There are several additional ways in which newly designed STR primers may result in allele discordance with standard STR typing kits. For example, suppose that at some STR locus an individual is genotyped as 4, 5 using a commercial kit. We shall focus only on the chromosome possessing the "5" allele (Figure 2, bottom allele). Interestingly, this particular allele has six core repeats. but between the core repeat and the forward primer of the commercial kit is a 4 bp deletion on the chromosome. The net effect of this deletion (equivalent to a single tetranucleotide repeat) is that the commercial kit scores this allele as having five repeats (Figure 2A; the sample types as a 4, 5). If the newly designed mini-STR primer is designed to hybridize between the commercial kit primer and the 4 bp deletion, then both genotypes will be concordant (Figure 2B; the sample types as a 4, 5). If, however, the mini-STR primer binding site includes the 4 bp deletion region, it is possible that little or no hybridization will occur. This would be especially true if the 3′ end of the

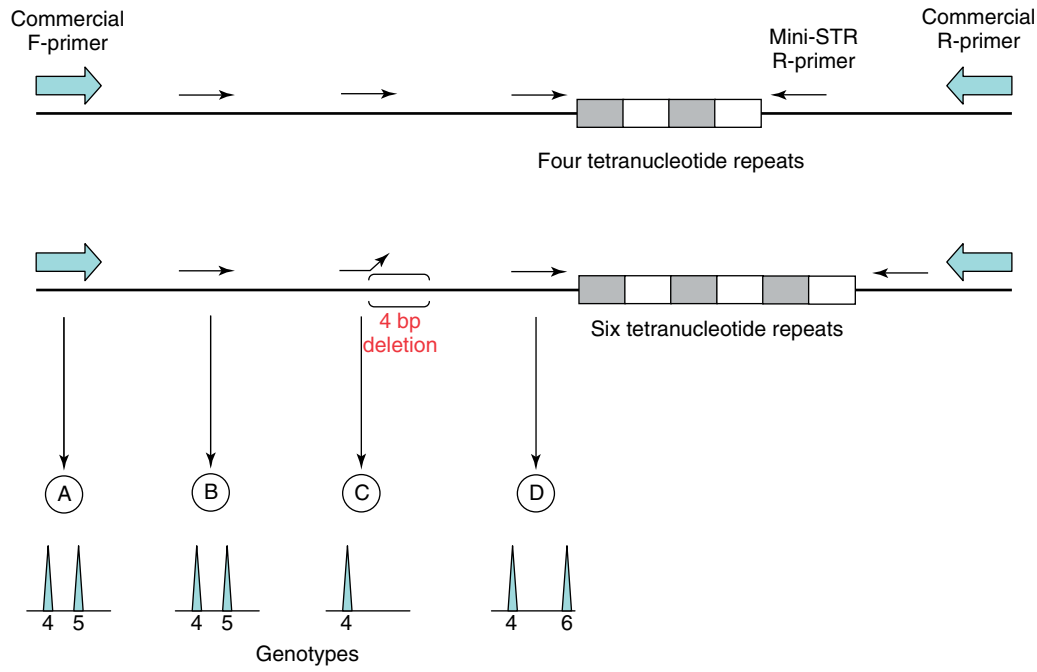


Figure 2 An example of discordance among mini-STR primers compared to a commercial kit. The top chromosome has four tetranucleotide repeats and the bottom chromosome has six tetranucleotide repeats. In the flanking region of the forward primer, there is a 4 bp deletion. Amplification using the commercial STR primers (A) results in a genotype of 4, 5 the deletion effectively "subtracts" one of the core repeats. (B) The mini-STR primer binding site is also outside of the 4 bp deletion region, and produce a 4, 5 genotype. (C) The 3′ end of the mini-STR primer produces no PCR product from the bottom chromosome resulting in a "null allele" genotype of 4, 4 and be discordant with the conventional kit. (D) The mini-STR primer is adjacent to the core repeat and is well within the 4 bp deletion, the genotype will be 4, 6 and be discordant with the conventional kit

mini-STR primer did not match the target sequence on the chromosome. If a "null allele" were to result due to the primer binding site mutation (Figure 2C; the sample types as 4, 4), the mini-STR genotype will be discordant with the kit genotype. Finally, if the mini-STR primer is adjacent to the core repeat and thus well within the 4 bp deletion, the genotype resulting from the mini-STR amplification will be 4, 6 (Figure 2D) and again discordant with the conventional kit.

Because of the mutations that result in insertions, deletions, and SNPs in the flanking regions of forensic STRs, laboratories should always test any new mini-STR primer set against the conventional kit on a set of population samples to assure concordance between the two systems. In most cases discordance is not a major issue since typically only one allele of the total 26 CODIS alleles would be affected. That is, a moderate stringency database search would still identify a profile typed with the conventional kit when compared to the mini-STR-generated profile with a single allele difference.

Conclusions

Mini-STRs have already proven effective for the recovery of genetic information from highly compromised samples, such as those encountered in mass disasters (*see* **Disaster Victim Identification**). Beyond the increased power of discrimination offered by the recovery of more genetic loci, mini-STRs have also proven valuable for the sorting and reassociation of highly commingled remains [17] from mass graves according to the International Commission on Missing Persons [19]. The European forensic community has recently recommended that current forensic loci should be reduced in size as much as possible and

that new mini-STRs be incorporated into the next generation STR multiplexes to increase the number of shared Interpol loci [20, 21]. In the future, mini-STRs may play an important role in the analysis of forensic evidence normally reserved for mtDNA analysis, such as skeletal remains [22] and shed telogen hairs [23]. The potential utility of mini-STRs is high and their frequency of use in forensic casework is likely to increase as both new and familiar autosomal, X-chromosome, and Y-chromosome loci are designed with reduced-size amplicons.

References

[1] FBI's Combined DNA Index System (CODIS) Homepage; http://www.fbi.gov/hq/lab/codis/index1.htm.

[2] Budowle, B, Moretti, T.T., Niezgoda, S.J. & Brown, B.L. (1998). CODIS and PCR-based short tandem repeat loci: law enforcement tools. *Proceedings of the Second European Symposium on Human Identification*, June 1998. Innsbruck, Austria, Madison, WI. Promega Corporation. pp. 73–88. http;//www.promega.com/geneticidproc/eusymp2proc/17.pdf.

[3] Butler, J.M., Li, J., Shaler, T.A., Monforte, J.A. & Becker, C.H. (1998). Reliable genotyping of short tandem repeat loci without an allelic ladder using time-of-flight mass spectrometry, *International Journal of Legal Medicine* **112**(1), 45–49.

[4] Butler, J.M. & Becker, C.H. (2001). *Improved Analysis of DNA Short Tandem Repeats with Time-of-Flight Mass Spectrometry*, Science and Technology Research Report, National Institute of Justice, available at: http://www.ncjrs.gov/pdffiles1/nij/188292.pdf.

[5] Hellmann, A., Rohleder, U., Schmitter, H. & Wittig, M. (2001). STR typing of human telogen hairs–a new approach, *International Journal of Legal Medicine* **114**(4–5), 269–273.

[6] Wiegand, P. & Kleiber, M. (2001). Less is more–length reduction of STR amplicons using redesigned primers, *International Journal of Legal Medicine* **114**(4–5), 285–287.

[7] Ohtaki, H., Yamamoto, T., Yoshimoto, T., Uchihi, R., Ooshima, C., Katsumata, Y. & Tokunaga, K. (2002). A powerful, novel, multiplex typing system for six short tandem repeat loci and the allele frequency distributions in two Japanese regional populations, *Electrophoresis* **23**(19), 3332–3340.

[8] Tsukada, K., Takayanagi, K., Asamura, H., Ota, M. & Fukushima, H. (2002). Multiplex short tandem repeat typing in degraded samples using newly designed primers for the TH01,CSF1PO,and vWA loci, *Legal Medicine (Tokyo)* **4**, 239–245.

[9] Butler, J.M., Shen, Y. & McCord, B.R. (2003). The development of reduced size STR amplicons as tools for analysis of degraded DNA, *Journal of Forensic Science* **48**(5), 1054–1064.

[10] Holland, M.M., Cave, C.A., Holland, C.A. & Bille, T.W. (2003). Development of a quality, high throughput DNA analysis procedure for skeletal samples to assist with the identification of victims from the World Trade Center attacks, *Croatian Medical Journal* **44**, 264–272.

[11] Biesecker, L.G., Bailey-Wilson, J.E., Ballantyne, J., Baum, H., Bieber, F.R., Brenner, C., Budowle, B., Butler, J.M., Carmody, G., Conneally, P.M., Duceman, B., Eisenberg, A., Forman, L., Kidd, K.K., Leclair, B., Niezgoda, S., Parsons, T.J., Pugh, E., Shaler, R., Sherry, S.T., Sozer, A. & Walsh, A. (2005). DNA Identifications after the 9/11 World Trade Center attack, *Science* **310**(5751), 1122–1123.

[12] Coble, M.D. & Butler, J.M. (2005). Characterization of new miniSTR loci to aid analysis of degraded DNA, *Journal of Forensic Sciences* **50**(1), 43–53.

[13] Asamura, H., Sakai, H., Kobayashi, K., Ota, M. & Fukushima, H. (2006). MiniX-STR multiplex system population study in Japan and application to degraded DNA analysis, *International Journal of Legal Medicine* **120**(3), 174–181.

[14] Park, M.J., Lee, H.Y., Chung, U., Kang, S.C. & Shin, K.J. (2007). Y-STR analysis of degraded DNA using reduced-size amplicons, *International Journal of Legal Medicine* **121**(2), 152–157.

[15] Mulero, J.J., Chang, C.W., Lagacé, R.E., Wang, D.Y., Bas, J.L., McMahon, T.P. & Hennessy, L.K. (2008). Development and validation of the AmpFlSTR MiniFiler PCR Amplification Kit: a MiniSTR multiplex for the analysis of degraded and/or PCR inhibited DNA, *Journal of Forensic Science* **53**(4), 838–852.

[16] Gill, P., Whitaker, J., Flaxman, C., Brown, N. & Buckleton, J. (2000). An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA, *Forensic Science International* **112**(1), 17–40.

[17] Irwin, J.A., Leney, M.D., Loreille, O., Barritt, S.M., Christensen, A.F., Holland, T.D., Smith, B.C. & Parsons, T.J. (2007). Application of low copy number STR typing to the identification of aged, degraded skeletal remains, *Journal of Forensic Science* **52**(6), 1322–1327.

[18] Butler, J.M. (2005). *Forensic DNA Typing: Biology and Technology behind STR Markers*, 2nd Edition, Academic Press, London.

[19] Parsons, T.J., Huel, R., Davoren, J., Katzmarzyk, C., Milos, A., Selmanovic, A., Smajlovic, L., Coble, M.D. & Rizvic, A. (2007). Application of novel "mini-amplicon" STR multiplexes to high volume casework on degraded skeletal remains, *Forensic Science International: Genetics* **1**, 175–179.

[20] Gill, P., Fereday, L., Morling, N. & Schneider, P.M. (2006a). The evolution of DNA databases – recommendations for new European loci, *Forensic Science International* **156**, 242–244.

[21] Gill, P., Fereday, L., Morling, N. & Schneider, P.M. (2006b). Letter to editor – new multiplexes for Europe: amendments and clarification of strategic development, *Forensic Science International* **163**, 155–157.
[22] Opel, K.L., Chung, D.T., Drabek, J., Tatarek, N.E., Jantz, L.M. & McCord, B.R. (2006). The application of miniplex primer sets in the analysis of degraded DNA from human skeletal remains, *Journal of Forensic Sciences* **51**(2), 351–356.
[23] Müller, K., Klein, R., Miltner, E. & Wiegand, P. (2007). Improved STR typing of telogen hair root and hair shaft DNA, *Electrophoresis* **28**(16), 2835–2842.

MICHAEL D. COBLE AND REBECCA S. JUST

Miranda Rights: Capacity to Waive *see* Capacity to Waive Miranda Rights

Miranda Warnings *see* Deception: Truth Serum

Missing Persons and Paternity: DNA

Introduction

Matching DNA profiles have proved to be of great value in forensic science when they link a person to evidence associated with a crime. The chance of a coincidental match between the profiles of two different people is so small that matches are rightly regarded as strong evidence. This has led to the increasing use of searches for a particular profile in databases of profiles previously obtained by law enforcement agencies. These forensic applications make little appeal to an essential nature of DNA profiles – genetic markers are heritable, meaning that a parent passes on one copy of his or her marker genes to a child. This leads to the similarity of profiles between related individuals and the application of DNA profiling to parentage disputes and missing person searches. The strength of the evidence from matching or partially matching profiles depends in large degree on the degree of relatedness of the people whose DNA is examined and calculations can be complex. There is simplification, however, in the situation where the people are not inbred and where their population can be regarded as being in a state of evolutionary equilibrium. These simplifications are described in this article.

The calculations described here are all within the framework of likelihood ratios of the probabilities of an observed set of DNA profiles under alternative hypotheses about the relatedness of the people whose DNA has been examined.

Measures of Inbreeding and Relatedness

The use of DNA profiles in parentage and missing persons situations rests on the comparisons of profiles from people who may be related. The greater the degree of relatedness, the more likely it is that profiles will be similar but it is necessary to be able to quantify such statements. Are father and son more related than are two brothers? How can profiles from two men be used to favor the explanation that they are brothers over the claim they are unrelated? Such questions require a means for attaching numbers to the degree of relatedness, and these numbers refer to the chance that alleles, the components of DNA profiles, are identical by descent.

People are related when they have ancestors in common and this means that relatives may share genes – the genes that descended from those ancestors. Two copies, or alleles, of the same genetic marker that have descended from the same ancestral allele are said to be identical by descent (ibd). Two individuals that have ibd alleles are said to be related, and individuals that receive ibd alleles from their parents are said to be inbred. In the classical theory of population genetics, there is an implied reference population: ibd alleles are copies of the same allele in the reference population and their histories further back in time are not considered. The probability that an allele taken at random from one individual is ibd

to an allele taken at random from another individual is the coancestry coefficient θ of those two individuals. The inbreeding coefficient F of a child has the same value as the coancestry θ of its parents.

There is a simple "path-counting" method for determining coancestry coefficients. For individuals X and Y, the number of individuals in the path linking them through their common ancestor(s) A is written as α. This number includes the individuals themselves. If F_A is the inbreeding coefficient of the ancestor, generally zero, the coancestry coefficient is given by $\theta_X = \sum_A (0.5)^\alpha (1 + F_A)$. For people related on only the maternal or paternal side of their family, there is likely to be only one ancestor A and the simplest example is that of parent Y and child X. Their common ancestor is Y and their coancestry is $(0.5)^2 = 0.25$. If Z is one of the parents of Y, then the grandparent (Z)–grandchild (X) coancestry is $(0.5)^3 = 0.125$, and so on. The two paths linking full-sibs X and Y through their parents G and H are XGY and XHY so $\theta_{XY} = (0.5)^3 + (0.5)^3 = 0.25$.

The equality of the coancestries for parent–child and full-sib pairs suggest that a more detailed quantification of relatedness is needed, and the first step is to use the number of pairs of alleles two relatives have that are ibd. There can be zero, one, or two such pairs depending on whether the individuals are unrelated, unilineal relatives (e.g., parent–child) or bilinear relatives (e.g., full-sibs). This description requires that an individual's two alleles are not themselves ibd. Calculating the probabilities k_0, k_1, k_2 that two individuals have 0, 1, 2 ibd pairs is fairly straightforward. If X has alleles a, b and Y has alleles c, d then it is necessary to ask how many of the four pairs ac, ad, bc, bd are ibd. If c is the allele that is copied by parent Y to transmit to child X, then it must be that either a or b is ibd to c. There are no other ibd relationships and $k_0 = 0, k_1 = 1, k_2 = 0$. If a, c are the alleles that full-sibs have received from parent G and b, d are the alleles they receive from parent H, then each of these pairs has equal chances of being copies of the same or different parental alleles and so being ibd or not. This leads to $k_0 = 0.25, k_1 = 0.50, k_2 = 0.25$. The coancestry coefficient θ is a summary of the three k coefficients: $\theta = k_1/4 + k_2/2$. Values of these measures are shown in Table 1 for some common relatives. Note that they all refer to the unobservable property of identity by descent. It is the expected values for those degrees of relatedness that are shown in the table.

Table 1 Identity by descent measures for noninbred relatives

Relationship	k_2	k_1	k_0	θ
Identical twins	1	0	0	$\frac{1}{2}$
Full sibs	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$
Parent/child	0	1	0	$\frac{1}{4}$
Half-sibs	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{8}$
First cousins	0	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{1}{16}$
Unrelated	0	0	1	0

The k-coefficients are sufficient for describing the relatedness for noninbred individuals. There is a more elaborate set of 15 coefficients for the situation with inbreeding when any two or three of four of the alleles carried by two people may be ibd. These coefficients were discussed in a review by Weir *et al.* [1]. For a random-mating population and genetic markers that have reached a state of equilibrium between the opposing evolutionary force of genetic drift that reduces genetic variation and the force of mutation that increases variation, all 15 of these more complicated ibd measures can be expressed in terms of the quantity θ. This θ is the probability that any two alleles in a population are ibd and it refers to the effect of past evolutionary events rather than immediate family membership. The probability that any three alleles are ibd is $2\theta^2/(1 + \theta)$ and the chance that any four alleles are ibd is $6\theta^2/[(1 + \theta)(1 + 2\theta)]$. The probability that two pairs of alleles are ibd (whether or not all four are ibd) is $\theta^2(1 + 5\theta)/[(1 + \theta)(1 + 2\theta)]$. These results were all derived by Evett and Weir [2].

Frequencies of sets of alleles

Setting up a set of identity measures was the first step in allowing observed DNA profiles to be used to make inferences about relatedness of the people represented by those profiles. Identity measures describe the relationship and they can also be used to express how likely it is that two people with a specified degree of relatedness have particular profiles. In other words, if the relationship is known, then the profile

types can be predicted. Some results from statistical theory can then be used to reverse the argument: if the profiles are known, what can be inferred about the relatedness?

Identity by descent cannot be observed, but the ibd probabilities allow the probabilities of sets of alleles to be expressed as functions of allele frequencies. Without any inbreeding or relatedness, the probability P_{ii} that an individual is homozygous for allele A_i is given by the Hardy–Weinberg result $P_{ii} = p_i^2$ where p_i is the frequency of the allele. The genotype frequency P_{ij} for heterozygote A_iA_j is $P_{ij} = 2p_ip_j$, $j \neq i$.

For individuals with inbreeding level F, these Hardy–Weinberg frequencies are modified to $P_{ii} = p_i^2 + Fp_i(1-p_i)$ and $P_{ij} = 2(1-F)p_ip_j$. Of more importance in the present context are the probabilities that two related individuals have specified genotype frequencies. The probability that two noninbred members of the same family are both homozygous A_iA_i has to take into account that they may share zero, one, or two pairs of alleles ibd. This means that their four copies of allele A_i may represent four, three, or two independent (non-ibd) alleles. The required probability is

$$\Pr(A_iA_i, A_iA_i) = k_0p_i^4 + k_1p_i^3 + k_2p_i^2 \qquad (1)$$

and the complete set of probabilities is shown in Table 2.

To take into account the evolutionary relatedness quantified by the parameter θ, there is a very convenient result first described in the forensic context by Balding and Nichols [3]. The probability that an allele is of type A when n_A of the previous n alleles examined were of that type is $[n_A\theta + (1-\theta)p_A]/[1+(n-1)\theta]$ and this can be called the *Dirichlet sampling formula*. For the first allele examined, the probability is just p_A. The probability an allele is A, given that an A has already been seen, is $[\theta + (1-\theta)p_A]$ so the chance they are both A is $p_A[\theta + (1-\theta)p_A] = p_A^2 + \theta p_A(1-p_A)$. The "background" or "evolutionary" relatedness produces the same effect as an inbreeding coefficient of $F = \theta$.

Table 2 Genotype-pair probabilities for noninbred relatives (different subscripts denote different alleles)

Genotype pair	Probability
A_iA_i, A_iA_i	$k_2p_i^2 + k_1p_i^3 + k_0p_i^4$
A_iA_i, A_jA_j	$k_0p_i^2p_j^2$
A_iA_i, A_iA_j	$k_1p_i^2p_j + 2k_0p_i^3p_j$
A_iA_i, A_jA_k	$2k_0p_i^2p_jp_k$
A_iA_j, A_iA_j	$2k_2p_ip_j + k_1p_ip_j(p_i+p_j) + 4k_0p_i^2p_j^2$
A_iA_j, A_iA_k	$k_1p_ip_jp_k + 4k_0p_i^2p_jp_k$
A_iA_j, A_kA_l	$4k_0p_ip_jp_kp_l$

The probability that an allele is A, given that the previous two alleles were A, is $[2\theta + (1-\theta)p_A]/(1+\theta)$ and this is of immediate use in paternity testing: it is the probability that a random man provides paternal allele A, given that the alleged father is homozygous for the allele (but see Table 4).

As one further example, the probability that an allele is A, given that the previous three alleles examined were also A, is $[3\theta + (1-\theta)p_A]/(1+2\theta)$. In the forensic context where a crime stain, left by the perpetrator, is of type AA this leads to the probability that a suspect is also of type AA:

$$\Pr(AA|AA) = \Pr(A|AAA)\Pr(A|AA) = \frac{[3\theta + (1-\theta)p_A][2\theta + (1-\theta)p_A]}{(1+\theta)(1+2\theta)} \qquad (2)$$

Note the use of the conditional probability symbol |.

The results just described for relatives, meaning familial relationships among people in the same family or evolutionary relatedness among people in the same population, are employed in the likelihood ratios for comparing alternative hypotheses about sets of observed DNA profiles. Whether the context is in the forensic situation involving a suspect and a crime stain, or the parentage situation of a child and an alleged father, or the missing person situation involving a stain or some remains and the relatives of the missing person, there is the evidence E of two or more DNA profiles and there are alternative hypotheses H that specify the relationship(s) among the sources of those profiles. One hypothesis, H_p, may be identity in the forensic case, father–child in the parentage case, and sibships in the missing person case. An alternative, H_d, may be unrelatedness in each case. The strength of the evidence is expressed as a likelihood ratio $\mathrm{LR} = \Pr(E|H_p)/\Pr(E|H_d)$.

The identity measures and expressions for joint genotypic probabilities also allow another type of calculation that has recently arisen in studies of databases [4]. When all profiles in a database are

compared to all other profiles, there are often quite striking degrees of similarity observed. It is helpful to be able to predict the degree of such similarity, taking into account familial and evolutionary relatedness, in order to determine whether or not the observed similarities indicate any unusual features of the database. To date, the observations have been consistent with expectations [4], and these expectations are now described.

The probability P_2 that two profiles match is [4]

$$\begin{aligned} P_2 &= \sum_i \Pr(A_iA_i,\, A_iA_i) + \sum_i \sum_{j\neq i} \Pr(A_iA_j,\, A_iA_j) \\ &= \sum_i \Pr(A_iA_iA_iA_i) + 2\sum_i \sum_{j\neq i} \Pr(A_iA_iA_jA_j) \\ &= \frac{1}{D}[6\theta^3 + \theta^2(1-\theta)(2+9S_2) + \\ &\quad 2\theta(1-\theta)^2(2S_2+S_3) + (1-\theta)^3(2S_2^2 - S_4)] \end{aligned} \tag{3}$$

The first line specifies the genotypes, the second shows the corresponding sets of alleles, and the third shows the value from the Dirichlet assumption. Random mating is assumed for the second line. The third line employs the notation $S_k = \sum_i p_i^k, k = 2, 3, 4$ and $D = (1+\theta)(1+2\theta)$. Partial matches occur when two individuals share one allele at a locus, rather than the two required for a match. The probability that two individuals partially match is

$$\begin{aligned} P_1 &= 2\sum_i \sum_{j\neq i} \Pr(A_iA_i,\, A_iA_j) \\ &\quad + \sum_i \sum_{j\neq i} \sum_{k\neq i,j} \Pr(A_iA_j,\, A_iA_k) \\ &= 4\sum_i \sum_{j\neq i} \Pr(A_iA_iA_iA_j) \\ &\quad + 4\sum_i \sum_{j\neq i} \sum_{k\neq i,j} \Pr(A_iA_iA_jA_k) \\ &= \frac{1}{D}[8\theta^2(1-\theta)(1-S_2) + 4\theta(1-\theta)^2(1-S_3) \\ &\quad + 4(1-\theta)^3(S_2 - S_3 - S_2^2 + S_4)] \end{aligned} \tag{4}$$

with the same meaning for the three rows as for P_2. Finally, for two individuals to mismatch, i.e., have no alleles in common,

$$\begin{aligned} P_0 &= \sum_i \sum_{j\neq i} \Pr(A_iA_i,\, A_jA_j) \\ &\quad + 2\sum_i \sum_{j\neq i} \sum_{k\neq i,j} \Pr(A_iA_i,\, A_jA_k) \\ &\quad + \sum_i \sum_{j\neq i} \sum_{k\neq i,j} \sum_{l\neq i,j,k} \Pr(A_iA_j,\, A_kA_l) \\ &= \sum_i \sum_{j\neq i} \Pr(A_iA_iA_jA_j) \\ &\quad + 2\sum_i \sum_{j\neq i} \sum_{k\neq i,j} \Pr(A_iA_iA_jA_k) \\ &\quad + \sum_i \sum_{j\neq i} \sum_{k\neq i,j} \sum_{l\neq i,j,k} \Pr(A_iA_jA_kA_l) \\ &= \frac{1}{D}[\theta^2(1-\theta)(1-S_2) \\ &\quad + 2\theta(1-\theta)^2(1-2S_2+S_3) \\ &\quad + (1-\theta)^3(1-4S_2+4S_3+2S_2^2-3S_4)] \end{aligned} \tag{5}$$

If, in addition to membership in the same population, two individuals have family relatedness described by k_0, k_1, k_2:

$$\begin{aligned} &\Pr(\text{Match}) = k_2 + k_1[\theta + (1-\theta)S_2] + k_0P_2 \\ &\Pr(\text{Partial Match}) = k_1(1-\theta)(1-S_2) + k_0P_1 \\ &\Pr(\text{Mismatch}) = k_0P_0 \end{aligned} \tag{6}$$

It is important to stress that these matching probabilities P_0, P_1, P_2 do not refer to specific profiles. They represent similarities over all profile types and so they apply to studies of databases rather than to specific parentage or missing person situations.

Parentage Testing

The usual situation in paternity disputes is that mother, child, and alleged father are genotyped. The alleged father is declared "not excluded" if he carries an allele that is inferred to be the child's paternal allele and the strength of the evidence against him is quantified as the paternity index *PI*. The two simplest explanations for the genetic evidence E are

H_p: the alleged father is the father.
H_d: the alleged father is not the father.

and the paternity index is

$$PI = \frac{\Pr(E|H_p)}{\Pr(E|H_d)} \tag{7}$$

If there is a prior probability π_0 of paternity, the posterior probability π should be changed to Bayes' theorem:

$$\frac{\pi}{1-\pi} = PI \times \frac{\pi_0}{1-\pi_0} \tag{8}$$

The PI can be expressed in terms of probability of genotype G_C of child, conditional on genotypes G_M, G_{AF} of mother, and alleged father:

$$\begin{aligned} PI &= \frac{\Pr(G_C|G_M, G_{AF}, H_p)\Pr(G_M, G_{AF}|H_p)}{\Pr(G_C|G_M, G_{AF}, H_d)\Pr(G_M, G_{AF}|H_d)} \\ &= \frac{\Pr(G_C|G_M, G_{AF}, H_p)}{\Pr(G_C|G_M, G_{AF}, H_d)} \end{aligned} \tag{9}$$

since the adult probabilities do not depend on the hypotheses. Provided the mother and alleged father are not related, so that the child's maternal and paternal alleles A_M and A_P are independent, it is more convenient to work with the alleles than the child's genotype. Noting that A_M depends only on the mother's genotype:

$$\begin{aligned} PI &= \frac{\Pr(A_M A_P|G_M, G_{AF}, H_p)}{\Pr(A_M A_P|G_M, G_{AF}, H_d)} \\ &= \frac{\Pr(A_M|G_M, H_p)\Pr(A_P|A_M, G_M, G_{AF}, H_p)}{\Pr(A_M|G_M, H_d)\Pr(A_P|G_M, G_{AF}, H_d)} \\ &= \frac{\Pr(A_P|A_M, G_M, G_{AF}, H_p)}{\Pr(A_P|G_M, G_{AF}, H_d)} \end{aligned} \tag{10}$$

because $\Pr(A_M|G_M)$ does not depend on H.

Suppose first that familial or evolutionary relatedness is not considered. Under explanation H_p, the alleged father has provided the paternal allele, of type A_i say, and the probability of the allele is 1.0 or 0.5, depending on whether he is homozygous or heterozygous for that allele. Under explanation H_d some other man, the true father TF, has provided the paternal allele and the probability of this unknown man providing the paternal allele is just the population allele frequency p_i. The PI is $1/p_i$ or $1/(2p_i)$ for homozygous A_iA_i or heterozygous A_iA_j alleged fathers.

Alleged father related to true father

There are situations, including those of incest, where the alternative hypotheses involve relatives. There may be reason to suspect that either the alleged father or his brother is the true father of a child. The two hypotheses become

H_p: the alleged father is the father.
H_d: the alleged father is related to the father.

and then it is necessary to determine the probability of the paternal allele, given that it came from a relative of the alleged father. Provided there is no inbreeding, the calculations follow from the joint probabilities for relatives given in Table 2 and are shown in Table 3. These lead to the probabilities of paternal allele A_i under H_d that it came from a relative, and the PI values are $1/[2\theta_{AT} + (1-2\theta_{AT})p_i]$ for homozygous alleged fathers and $1/\{2[\theta_{AT} + (1-2\theta_{AT})p_i]\}$ for heterozygous alleged fathers. The quantity θ_{AT} is the coancestry of alleged and (under H_d) true fathers.

It is also possible to construct a likelihood ratio for a different pair of alternative explanations for situations when the alleged father is either deceased or

Table 3 Paternity index calculations when H_d is that alleged father is related to the father. (The paternal allele is A_i and A_j, A_k are any other distinct alleles.)

Alleged father	$\Pr(A_i\|H_p)$	Relative	Pr(Relative|Alleged father)	$\Pr(A_i\|H_d)$
A_iA_i	1.0	A_iA_i	$(k_0p_i^4 + k_1p_i^3 + k_2p_i^2)/p_i^2$	1.0
		A_iA_j	$(2k_0p_i^3p_j + k_1p_i^2p_j)/p_i^2$	0.5
A_iA_j	0.5	A_iA_i	$(2k_0p_i^3p_j + k_1p_i^2p_j)/(2p_ip_j)$	1.0
		A_iA_j	$[4k_0p_i^2p_j^2 + k_1p_ip_j(p_i+p_j) + 2k_2p_ip_j]/(2p_ip_j)$	0.5
		A_iA_k	$(4k_0p_i^2p_jp_k + k_1p_ip_jp_k)/(2p_ip_j)$	0.5

otherwise not available for testing but his relative can be tested. If X is the tested man, the hypotheses are

H_p: X is a relative of the father Y.
H_d: X is unrelated to the father Y.

If the degree of relatedness in H_p is specified by θ_{XY}, then the *PI*, often called the *Avuncular Index*, when the paternal allele is A_i is $[1 - 2\theta_{XY} + 2\theta_{XY}/p_i]$ when the tested man is homozygous A_iA_i, $[1 - 2\theta_{XY} + \theta_{XY}/p_i]$ when the tested man is heterozygous A_iA_j, and it is $(1 - 2\theta_{XY})$ when the tested man does not carry A_i.

Evolutionary relatedness

For populations in which there is a (low) level of relatedness of individuals because of the evolutionary history of the population, there is a need to consider the relatedness of mother, alleged father, and father. This does not affect $\Pr(A_P|G_M, G_{AF}, H_p)$ because that is determined by the genotype of the alleged father. Under H_d, however, the Dirichlet sampling formula can be used. The probability of the paternal allele depends on the four alleles already seen: those of the mother and the alleged father. The resulting *PI* values are shown in Table 4.

Table 4 Paternity Index values for a population with evolutionary relatedness. (Different subscripts denote different alleles)

G_M	G_C	A_M	A_P	G_{AF}	PI
A_iA_i	A_iA_i	A_i	A_i	A_iA_i	$\frac{1+3\theta}{4\theta+(1-\theta)p_i}$
				A_iA_j	$\frac{1+3\theta}{2[3\theta+(1-\theta)p_i]}$
	A_iA_j	A_i	A_j	A_jA_j	$\frac{1+3\theta}{2\theta+(1-\theta)p_j}$
				A_iA_j	$\frac{1+3\theta}{2[\theta+(1-\theta)p_j]}$
				A_jA_k	$\frac{1+3\theta}{2[\theta+(1-\theta)p_j]}$
A_iA_k	A_iA_i	A_i	A_i	A_iA_i	$\frac{1+3\theta}{3\theta+(1-\theta)p_i}$
				A_iA_k	$\frac{1+3\theta}{2[2\theta+(1-\theta)p_i]}$
	A_iA_j	A_i	A_j	A_jA_j	$\frac{1+3\theta}{2\theta+(1-\theta)p_j}$
				A_iA_j	$\frac{1+3\theta}{2[\theta+(1-\theta)p_j]}$
				A_jA_l	$\frac{1+3\theta}{2[\theta+(1-\theta)p_j]}$

Missing Person Calculations

Many of the issues involved in missing person calculations are the same as those for paternity disputes. Instead of a paternal allele being known, a biological sample from the missing person is available. Suppose a person is missing; the genetic evidence E consists of the genotype from a sample that has come from some person X who may be the missing person Y, together with the genotypes from the spouse M and child C of the missing person. Two explanations of the evidence are

H_p: the sample is from the missing person.
H_d: the sample is not from the missing person.

A general approach for calculating the likelihood ratio is to work with probabilities of genotypes conditional on those in the previous generation(s):

$$\begin{aligned} LR &= \frac{\Pr(E|H_p)}{\Pr(E|H_d)} \\ &= \frac{\Pr(G_C, G_M, G_X|H_p)}{\Pr(G_C, G_M, X|H_d)} \\ &= \frac{\Pr(G_C|G_M, G_X, H_p)\Pr(G_M, G_X|H_p)}{\Pr(G_C|G_M, G_X, H_d)\Pr(G_M, G_X|H_d)} \\ &= \frac{\Pr(G_C|G_M, G_X, H_p)}{\Pr(G_C|G_M, H_d)} \end{aligned} \tag{11}$$

since the genotype of the child does not depend on that of X when H_d is true (ignoring evolutionary relatedness within the population). This likelihood ratio is the same as in the paternity case where X is alleged to be the father of child C who has mother M. Similar extensions can be made to allow for X to be a relative of the missing person, or to allow for evolutionary relatedness among all members of a population.

It may be the case that people apart from the spouse and child of the missing person are typed. The general procedure is the same: the probabilities of the set of observed genotypes under two explanations are compared. Suppose the parents P, Q as well as the child C and spouse M of the missing person Y are

typed, and that a sample is available that has come from some person X thought under H_p to be Y. Under explanation H_d, the sample from X did not come from Y, and therefore the genotype of X does not depend on the genotypes of P and Q and the genotype of C does not depend on the genotype of X.

$$\begin{aligned} LR &= \frac{\Pr(E|H_p)}{\Pr(E|H_d)} \\ &= \frac{\Pr(C, M, X, P, Q|H_p)}{\Pr(C, M, X, P, Q|H_d)} \\ &= \frac{\Pr(C|M, X, P, Q, H_p)\Pr(M, X, P, Q|H_p)}{\Pr(C|M, X, P, Q, H_d)\Pr(M, X, P, Q|H_d)} \\ &= \frac{\Pr(C|M, X, H_p)\dfrac{\Pr(M, X|P, Q, H_p)}{\Pr(P, Q|H_p)}}{\Pr(C|M, P, Q, H_d)\dfrac{\Pr(M, X|P, Q, H_d)}{\Pr(P, Q|H_d)}} \\ &= \frac{\Pr(C|M, X, H_p)\Pr(M|H_p)\Pr(X|P, Q, H_p)}{\Pr(C|M, P, Q, H_d)\Pr(M|H_d)\Pr(X|H_d)} \\ &= \frac{\Pr(C|M, X, H_p)\Pr(X|P, Q, H_p)}{\Pr(C|M, P, Q, H_d)\Pr(X|H_d)} \end{aligned} \tag{12}$$

An example is shown in Table 5.

Evolutionary relatedness can be accounted for by modifying the terms involving allele frequencies. In this case, the term $\Pr(X|H_d)$ needs to take into account the alleles already seen in P, Q, M, and C. For the example in Table 5, where $X = A_1A_3$, this probability is that of obtaining A_1 after having seen two copies of A_1 and A_2 and one copy of A_3, A_4, A_5 and A_6, and then of obtaining A_3 after having seen three copies of A_1, two copies of A_2, and one copy of A_3, A_4, A_5 and A_6. From the Dirichlet sampling formula, this probability is

$$\begin{aligned} &\Pr(A_1A_3|A_1A_1A_2A_2A_3A_4A_5A_6) \\ &= \frac{2\theta + (1-\theta)p_1}{1+8\theta}\frac{\theta + (1-\theta)p_3}{1+7\theta} \end{aligned} \tag{13}$$

Table 5 An example of a missing person calculation.

Child	G_C	A_1A_2
Sample	G_X	A_1A_3
Spouse	G_M	A_2A_4
Mother	G_P	A_1A_5
Father	G_Q	A_3A_6
$\Pr(C\|M, X, H_p)$	=	$\frac{1}{4}$
$\Pr(X\|P, Q, H_p)$	=	$\frac{1}{4}$
$\Pr(C\|M, P, Q, H_d)$	=	$\frac{1}{8}$
$\Pr(X\|H_d)$	=	$2p_1p_3$
LR	=	$\frac{1}{4p_1p_3}$

As a final example, consider the case where profiles are available from one parent P, four siblings S, the spouse M, and a child C of a missing person, as well as from a sample X that may be from that missing person. Sample profiles are shown in Table 6. Write the evidence as $E = (C, M, X, S, P)$ and the hypotheses H_p, H_d that X is or is not from the missing person. It is necessary to introduce the untyped parent Q and add over all possible genotypes for this parent that are consistent with P and S under H_d, and consistent with P, S, X under H_p. The probability of any specific genotype of Q does not depend on the hypotheses. The general procedure is still to write the probabilities for people conditional on those in the previous generations:

$$\begin{aligned} LR &= \frac{\Pr(C, M, X, S, P|H_p)}{\Pr(C, M, X, S, P|H_d)} \\ &= \frac{\sum_Q \Pr(C, M, X, S, P|Q, H_p)\Pr(Q|H_p)}{\sum_Q \Pr(C, M, X, S, P|Q, H_d)\Pr(Q|H_d)} \\ &= \frac{\sum_Q \Pr(C|M, X, S, P, Q, H_p)\dfrac{\Pr(M, X, S, P|Q, H_p)\Pr(Q)}{\Pr(M, X, S, P|Q, H_d)\Pr(Q)}}{\sum_Q \Pr(C|M, X, S, P, Q, H_d)} \\ &= \frac{\sum_Q \Pr(C|MXH_p)\dfrac{\Pr(M|H_p)}{\Pr(XSP|QH_p)}\Pr(Q)}{\sum_Q \Pr(C|MSPQH_d)\dfrac{\Pr(M|H_d)}{\Pr(X|H_d)\Pr(SP|QH_d)}\Pr(Q)} \\ &+ \frac{\sum_Q \Pr(C|MXH_p)\dfrac{\Pr(XS|PQH_p)}{\Pr(P|QH_p)}\Pr(Q)}{\sum_Q \Pr(C|MPQH_d)\dfrac{\Pr(X|H_d)}{\Pr(S|PQH_d)\Pr(P|QH_d)}\Pr(Q)} \\ &= \frac{\Pr(C|M, X, H_p)\sum_Q \Pr(X, S|P, Q, H_p)\Pr(Q)}{\Pr(X|H_d)\sum_Q \Pr(C|M, P, Q, H_d)\Pr(S|P, Q, H_d)\Pr(Q)} \end{aligned} \tag{14}$$

Table 6 An example of a missing person calculation

P	mother, with genotype A_3A_4
S	sibs, with genotypes A_2A_4, A_2A_4, A_2A_4, A_3A_4
Q	untyped father who must have genotype A_2A_3 or A_2A_4
M	spouse, with genotype A_5A_6
C	child, with genotype A_3A_5
X	sample, with genotype A_3A_3

$$\Pr(C|M, X, H_p) = 1/4$$

$$\Pr(X, S|P, Q, H_p) = \begin{cases} \dfrac{1}{1024}, & Q = A_2A_3 \\ 0, & Q = A_2A_4 \end{cases}$$

$$\Pr(X|H_d) = p_3^2$$

$$\Pr(C|M, P, Q, H_d) = \begin{cases} \dfrac{1}{2}, & Q = A_2A_3 \\ \dfrac{1}{4}, & Q = A_2A_4 \end{cases}$$

$$\Pr(S|P, Q) = \begin{cases} \dfrac{1}{256}, & Q = A_2A_3 \\ \dfrac{1}{256}, & Q = A_2A_4 \end{cases}$$

$$\Pr(Q) = \begin{cases} 2p_2p_3, & Q = A_2A_3, \\ 2p_2p_4, & Q = A_2A_4 \end{cases}$$

$$LR = \frac{(1/4) \times (1/1024) \times 2p_2p_3}{(p_3^2)[(1/2) \times (1/256)2p_2p_3 + (1/4) \times (1/256)2p_2p_4]} = \frac{1}{4p_3(2p_3 + p_4)}$$

Details for the specific profiles are shown in Table 6.

Discussion

Interpreting DNA evidence for situations involving parentage or missing person identification rests on the profile probabilities for sets of related individuals. When relatedness is a consequence of membership in the same family, there is a set of three parameters that are the probabilities two people share zero, one, or two pairs of ibd alleles. For the relatedness resulting from the shared evolutionary history of all members of a population, there is a convenient formulation in terms of a general coancestry coefficient θ. This formulation has become widely used in single-contributor forensic calculations and it should also be used in paternity calculations. The identification of remains in missing person situations can be quite complicated when many family members are typed, but the use of likelihood ratios to compare evidence probabilities under alternative hypotheses provides a general approach. The probabilities need to be written for individuals in one generation conditional on individuals in previous generations.

Relatedness coefficients allow the probabilities of DNA profiles for sets of individuals to be written out explicitly. Although increasing relatedness is expected to result in increasing profile similarity, the probability expressions in Tables 1 and 2 make it clear that even unrelated people may have very similar profiles, whereas related people may have quite dissimilar profiles. For example, unrelated people may both be homozygous A_1A_1 at a locus and full-sibs may have completely different genotypes A_1A_1, A_2A_2. It is more appropriate to compare evidence probabilities under alternative hypotheses than it is to have arbitrary rules that deny the possibility of relatedness, once some threshold level of allelic dissimilarity is reached.

Acknowledgment

This work was supported in part by NIH grant GM 75091.

References

[1] Weir, B.S., Anderson, A.D. & Hepler, A.B. (2006). Genetic relatedness analysis: modern data and new challenges, *Nature Reviews Genetics* **7**, 771–780.

[2] Evett, I.W. & Weir, B.S. (1998). *Interpreting DNA evidence–statistical genetics for forensic scientists*, Sinuaer Associates, Sunderland.

[3] Balding, D.J. & Nichols, R.A. (1994). DNA profile match probability calculations: how to allow for population stratification, relatedness, database selection and single bands, *Forensic Science International* **64**, 125–140.

[4] Weir, B.S. (2007). The rarity of DNA profiles, *Annals of Applied Statistics* **1**, 358–370.

Related Articles

Mitochondrial DNA: Interpretation

Short Tandem Repeats: Interpretation

BRUCE S. WEIR

Mistaken Identification *see* Eyewitness Lineups: Identification from

Mitigation Testimony

Despite a growing consensus against the death penalty in the United States, there still exist a large number of people that believe in the use of capital punishment. Problematic to some is the holding in *Witherspoon v. Illinois* [1], which guaranteed that only those individuals that are willing to impose the death penalty be allowed to serve as jurors in capital litigation. As such, imposition of the death penalty may not be readily influenced by issues that have led to the decrease in support for the death penalty (e.g., wrongful executions, lack of deterrent value), and these individuals may be more inclined to give the death sentence than would many people in the general population.

In 1979, the United States Supreme Court (USSC) upheld the use of guided discretion in the application of the death sentence for specific crimes in a bifurcated trial; the first stage requires the jury to determine guilt or innocence, and the second to determine sentence after consideration of aggravating and mitigating circumstances (see, for example, [2–4]). In brief, aggravating circumstances are factors that define and narrow the class of defendants eligible for the death penalty (*Zant v. Stephens* [5]), and mitigating circumstances are factors that decrease a capital defendant's culpability to the level at which the death penalty is considered undeserved. In order to proceed with the option of death, the state must prove that one or more statutory aggravating factors exists.

In most states, aggravating circumstances are delineated by statute and include factors that can be separated into four categories: (i) defendant characteristics (e.g., whether the defendant was previously convicted of another capital offense or a felony involving the use or threat of violence to the person not limited to the following factors of the crime and/or the defendant); (ii) elements of the crime (e.g., the defendant knowingly created a great risk of death to many persons; the capital offense was especially heinous, atrocious, or cruel compared to other capital offenses); (iii) motive for the crime (e.g., the capital offense was committed for pecuniary gain; the capital offense was committed to disrupt or hinder the lawful exercise of any governmental function or the enforcement of laws; and (iv) victim characteristics (e.g., the murdered individual was an on-duty peace officer who was killed in the course of performing his official duties and the defendant knew, or should have known, that the victim was a peace officer; the defendant was an adult and the murdered person was an unborn child in the womb at any stage of its development) [6]. While the number of statutory aggravating circumstances varies by state, all states, however, place the burden of proof at the level of beyond a reasonable doubt.

In contrast to aggravating factors, mitigating factors are not limited to those defined by statute, but instead include "any aspect of character or record, and any circumstance of the offense that might serve as a basis for a sentence less than death" [7]. In its decision in *Wiggins v. Smith* [8] the USSC unequivocally stated that failure to investigate and present mitigating evidence during the sentencing phase of a capital trial violates the defendant's Sixth Amendment right to effective assistance of counsel. In *Wiggins*, the defense had found, but failed to present, mitigating circumstances including, but not limited to, severe neglect and physical abuse, sexual abuse, and borderline mental retardation. In the majority opinion, Justice Sandra Day O'Connor opined that "Had the jury been able to place [Wiggins's] excruciating life history on the mitigating side of the

scale, there is a reasonable probability that at least one juror would have struck a different balance" (p. 536).

Jury Decision Making in Capital Litigation

At the most basic level, the concepts of mitigating and aggravating circumstances are easily understood: a mitigating circumstance lessens a defendant's moral culpability and goes against a sentence of death, and an aggravating circumstance increases a defendant's moral culpability and may be used to support the imposition of the death sentence. Despite the apparent simplicity of the constructs, research has consistently shown that jurors' ability to apply these concepts during capital litigation is impaired (e.g., [9–13]). In their evaluation of comprehension of capital sentencing instructions in California, Haney and Lynch [14] identified significant problems in this arena. In their sample of college educated individuals, only 15% were able to provide a legally correct definition of aggravation, 12% could provide a correct definition of mitigation, and a mere 8% provided correct definitions of both terms. Perhaps the most concerning findings were that 30% of the sample provided completely incorrect definitions of mitigation and, after being read jury instructions three times, 11% of the sample was unable to provide a definition of mitigation. In 1997, Haney and Lynch [15] again evaluated the comprehension of jury instructions with respect to the value of providing jurors with explicit definitions of the constructs. The results of this study demonstrated that the inclusion of explicit definitions did nothing to improve overall comprehension.

Garvey [16] addressed an important question in capital litigation, specifically, what impact, if any, do certain factors have on jury decision making in capital cases. Using data from the Capital Jury Project (CJP)[a], Garvey found that the following factors made it more likely that a juror would impose the death penalty: (i) the murder was particularly heinous; (ii) the victim of murder was a child; (iii) the defendant lacked remorse; and (iv) the defendant was identified to be a risk for future dangerousness. Conversely, doubt regarding guilt, a defendant's youthfulness, the presence of mental retardation, and other factors that are outside of the defendant's control (e.g., mental illness) were found to be strong mitigators. Of import was that jurors assigned almost no weight to developmental factors such as child abuse or a background of extreme poverty.

Though studies on lay persons' knowledge regarding developmental risk factors have not yet been conducted, it is possible that the limited significance that jurors placed on these factors is related to lack of understanding regarding the long-term impact of negative life events. In a study that evaluated juror's perceptions of expert testimony, Sundby [17] found that most jurors held negative views about experts, and in particular defense experts. The data demonstrated the presence of three major concerns regarding defense experts, who provide mitigation testimony: (i) jurors believed that such experts lack objectivity; (ii) jurors do not believe that experts have the ability to explain human behavior; and (iii) jurors believe that experts fail to make the critical connection between the psychological principles and the defendant's behavior. Jurors clearly need more than just a recitation of traumatic life experiences and the delivery of a sad story. What jurors need is an explanation of how adverse life events impacted the development of the defendant and how such adversity can be linked to their status as a capital defendant.

Risk, Protective Factors, and Resilience: A Developmental Approach to Mitigation

The sheer magnitude of the outcome of the sentencing decision in capital litigation warrants exploration into the best method to communicate jurors. It is the opinion of this writer that presentation of a defendant's life history within the context of theories of developmental psychopathology is critical, as it provides jurors with a context to understand the defendant's life history. The rationale behind this belief lies in the fact that the capital defendant is similar to any other person in that they have a developmental history. Their developmental trajectory was influenced and shaped by numerous intervening events that either served to hinder or enhance positive growth and development.

In the field of developmental psychology, these life events have been identified under the terms *risk factors* and *protective factors*, respectively. From the developmental perspective, "maladaption is viewed as evolving through the successive adaptations of persons in their environments. It is not something a person 'has' or an ineluctable expression of an

endogenous pathogen. It is the complex result of a myriad of risk and protective factors operating overtime" [18, p. 251].

Risk factors

At the most basic level, a risk factor is a predictor that has been scientifically demonstrated to have a strong link to adverse outcomes such as delinquency, adult antisocial behavior, substance abuse, unemployment, and violence. A recent study [19] investigated the precursors to serious and chronic delinquency, as well as youth violence, and from these data were able to identify a number of risk and protective factors. Included among the many risk factors were such things as perinatal difficulties, family history of criminal behavior and substance abuse, early exposure to violence, economic deprivation, media portrayals of violence, academic failure and lack of commitment to school, and low intelligence.

It is well accepted that there is no single risk factor, or a set of risk factors, that is necessary or sufficient to produce an adverse outcome. In fact, years of research has unequivocally shown that outcomes worsen as the number of risk factors increases (*see*, for example, [20–24]). For example, longitudinal studies have shown that children and adolescents diagnosed with conduct disorder demonstrate broad-based dysfunction over multiple domains. In his review of the literature, Kazdin [25] linked seven categories of dysfunction in adulthood to a prior diagnosis of conduct disorder. Specifically, data from multiple studies have shown that these individuals are more likely to (i) experience psychiatric disturbances; (ii) engage in criminal behavior; (iii) have limited occupational success; (iv) function poorly in school; (v) have impaired marital relationships; (vi) isolate from others; and (vii) experience poor physical health. Not surprisingly, these areas of dysfunction are common among death-row inmates [26].

Protective Factors

While the lives of capital murder defendants tend to be laden with risk factors such as poverty, early exposure to illicit drugs and violence, physical abuse, and ineffective parenting [26], in even the most disadvantaged of systems, an individual is typically exposed to one or more protective factors (e.g., a teacher or coach who takes an interest in the child; an opportunity to engage in community recreation; a family member who notices danger in the home and takes the child in). For the past three decades, the importance of prevention has been recognized, and researchers and policy makers have been dedicated to finding ways to increase access to protective factors; the goal of such intervention is to foster resilience in at-risk children and adolescents (see below for a presentation of the concept of resilience).

The push toward prevention is exemplified in a study funded by the National Institute on Drug Abuse that identified protective factors that can help prevent high-risk youths from engaging in delinquent behavior and illicit substance use [27]. The authors of this study found that resistance to both illicit substance use and delinquency was directly related to the accumulation of protective factors across multiple domains of an adolescent's life. Some of the protective factors identified in the study include appropriate parental supervision, mutual connectedness between parent and child, a commitment to education by both child and parent, association with a peer group that has conventional values, parental approval of one's peer group, positive self-esteem, and child involvement in pro-social activities.

Resiliency

Perhaps the main obstacle to presenting mitigation testimony is the fact that jurors, though similar in that they have matured in response to internal and external forces, are different from capital litigants in that they have somehow succeeded (even if only marginally so) in the presence of some adversity. In the developmental psychology literature, this construct has been referred to as *resiliency* and has been used to explain how individuals that were raised in the same environment can demonstrate highly disparate life trajectories.

Resilience has been defined as a "dynamic developmental process reflecting evidence of adaptation despite significant life adversity" [28]. As previously mentioned, studies have demonstrated that the presence of any single risk factor does not cause adverse outcomes. Instead, it is the convergence of risk factors that leads to widespread dysfunction, and it is the presence of protective factors that fosters resilience (see, for example, [29, 30]).

It has become apparent that resilience in adulthood is related to a number of factors such as having

had more resources and fewer adversities early in life. Exposure to effective parenting, having had a greater number of positive relationships with adult role models, and there being a longer time between the birth of siblings have been found to buffer the negative impact of adverse life events (e.g., [31, 32]). These findings, among others, underscore the critical role of protective factors in the developmental process.

Expert Testimony on Mitigation

Searching for and understanding the role of protective factors has dual purposes in mitigation. First, like most people, the capital defendant will have had exposure to one or more protective factors over the course of their life. This may be the influence of a teacher, involvement in a coping skills program, or perhaps the presence of supportive parents. The usefulness of the identification of protective factors lies not in their presence, but in their impact on the life course of the defendant. It may be that there were opportunities for change via counseling or involvement in extracurricular activities, but that these protective factors were not powerful enough to overcome the impact of the risk factors. Second, there are cases in which protective factors are absent from the life of the defendant and this fact must be made apparent to the jury.

Furthermore, the court must be made aware that the defendant's life trajectory may have been different had someone come forward and intervened on behalf of the individual, or had their life experience included positive influences such as access to prosocial peer groups, financial stability, or a safe living environment. During the presentation of mitigation, it is important to clearly delineate what could have been done to change the defendant's life course, what was done to assist in the process of change, and what tools were readily available yet not implemented.

With respect to expert testimony on mitigation, it is imperative that the mental health practitioner is knowledgeable of the risk factors experienced by the defendant and be able to discuss the impact of these risk factors in relation to interventions that were or were not implemented. Of course, individual risk factors would need to be evaluated over the life course, and be discussed with respect to cumulative stress and the interaction among factors. While perhaps not obvious, the issue of resiliency is of great import in capital mitigation. Clearly, millions of people experience adverse life events, but it is only the rare few who engage in capital crime. In order for mitigation to be successful, the trier of fact must be convinced that the defendant's experience of similar life events was unique, and that it was the uniqueness of the defendant's response that provides an explanation for the violent behavior.

In addition to having a comprehensive understanding of developmental psychopathology, it is critical that the mental health professional is aware that juror decision making does not occur in a vacuum; instead, information is acquired, manipulated, and maintained in the context of each juror's life experience. Similar to decision making in other realms, individuals that are asked to make decisions in the legal arena are doing so, at least in part, on the basis of schemas. In other words, the knowledge that they have accumulated from previous interactions over many years has a direct influence on how jurors think and respond (e.g., [33]). As such, the strength of mitigation evidence is at least in part related to the relationship between the defendant, the crime, and the pre-existing schemata of each juror.

In situations in which jurors' schemata are based on biased or inaccurate information that are detrimental to the defendant, the power of mitigation testimony is likely to be greatly reduced. As such, mental health professionals must enter the courtroom ready to discuss stereotypes and to present information that challenges the myths that may surround the character and life history of the defendant.

Putting it all in Context: *Moore v. Parker*

The lack of understanding of the complexity of presenting mitigating circumstances was highlighted in a recent decision by the US Court of Appeals for the Sixth Circuit. In the case of *Moore v. Parker* [34], the Court affirmed the district court's denial of Moore's writ of *habeas corpus* that was based on ineffective assistance of counsel. With regard to the penalty phase, Moore brought forth the following claims: "his attorneys erred by (1) allegedly spending only about three percent of their preparation time on the penalty phase; (2) remaining unaware of ninety-five letters sent to the first court supporting

him, which could have led them to more mitigating evidence and (3) not having another psychologist examine him after the first on they selected proved to be a fraud."

In response, Judges Boggs and Cook found that the testimony of four witnesses that testified to Moore's troubled upbringing (i.e., the defendant, Moore's aunt, Moore's prison "boss", and a Reverend that had only a brief conversation with Moore) and one psychologist (not hired to conduct a mitigation evaluation) was sufficient for the purposes of mitigation.[b] Judge Boggs affirmed the lower court's ruling, at least in part, on the basis of the finding that strong mitigators were presented to the jury. Included among them were "severe abuse and neglect", "his mother stabbing his father and to his having grown up in numerous foster homes and institutions", "his mother's alcoholism and his father's abusiveness", "his abuse and neglect in foster homes and institutions," and "watching his father hitting his mother so hard that his father broke his hand in three places." Judges Boggs and Cook agreed with the district court in their view that the testimony of the psychologist "cast Moore as an easily angered, impulsive, out of control emotional leech with poor judgment" and such testimony would have been detrimental to his case.

What is pointed out in the dissent is that Moore's dangerousness was not placed in the context of his chaotic and abusive upbringing. Furthermore, Judge Martin unequivocally stated that had the jury been presented with a "graphic description of Moore's atrocious childhood" there exists a reasonable probability that the jury may have come to a different decision at sentencing.

As aptly discussed in Douglas Rau's review of *Moore v. Parker* [35], it is clear that Judges Boggs and Cook were focused on whether or not mitigation testimony would engender sympathy from the jury. In their narrow conceptualization of the goal of mitigation, the judges lost sight of the fact that sympathy does not arise solely from presentation of positive defendant attributes in the face of adversity. To the contrary, in a capital case sympathy is more likely to be engendered by the presentation of the horrible and unfathomable aspects of an individual's upbringing that help to explain how the defendant came to be the potentially dangerous, frightening, and unpredictable person that committed murder. As stated by Judge Martin, expressed mitigation specialists are "qualitatively different from lay witnesses and can translate complex information into testimony that will assist the jury in reaching its determination."

End Notes

[a.] The Capital Jury Project is an ongoing program of research on the decision making of capital jurors. The project began in 1991 by a consortium of university-based researchers and is supported with funds from the National Science Foundation.

[b.] In his dissent, Judge Martin noted that only Moore and his aunt spoke of the issue of childhood, not four witnesses as noted by Judges Boggs and Cook.

References

[1] Witherspoon v. Illinois, 391 U.S. 510 (1968).
[2] Gregg v. Georgia, 428 U.S. 153 (1976).
[3] Jurek v. Texas, 428 U.S. 262 (1976).
[4] Proffitt v. Florida, 428 U.S. 242 (1976).
[5] Zant v. Stephens, 462 U.S. 862, 878 (1983).
[6] Acker, J.R. & Lanier, C.S. (1994). "Parsing this lexicon of death": aggravating factors in capital sentencing statutes, *Criminal Law Bulletin* **30**, 107–153.
[7] Lockett v. Ohio, 438 U.S. 586 (1978).
[8] Wiggins v. Smith, 539 U.S. 510 (2003).
[9] Bentele, U. & Bowers, W.J. (2002). How Jurors decide on death: guilt is overwhelming; aggravation requires death; and mitigation is no excuse, *Brooklyn Law Review* **66**, 1013–1080.
[10] Costanzo, M. & Costanzo, S. (1992). Jury decision making in the capital penalty phase: legal assumptions, empirical findings, and a research agenda, *Law and Human Behavior* **16**, 185–201.
[11] Diamond, S.S. (1993). Instructing on death: psychologists, juries, and judges, *American Psychologist* **48**, 423–434.
[12] Haney, C., Sontag, L. & Costanzo, S. (1994). Deciding to take a life: capital juries, sentencing instructions, and the jurisprudence of death, *Journal of Social Issues* **50**, 149–176.
[13] Lynch, M. & Haney, C. (2000). Discrimination and instructional comprehension: guided discretion, racial bias, and the death penalty, *Law and Human Behavior* **24**, 337–358.
[14] Haney, C. & Lynch, M. (1994). Comprehending life and death matters: a preliminary study of California's capital penalty instructions, *Law and Human Behavior* **18**, 411–436.
[15] Haney, C. & Lynch, M. (1997). Clarifying life and death matters: an analysis of instructional comprehension and penalty phase arguments, *Law and Human Behavior* **21**, 575–595.

[16] Garvey, Stephen P. (1998). Aggravation and mitigation in capital cases: what do jurors think? *Columbia Law Review* **98**, 1538.
[17] Sundby, S.E. (1994). The jury as critic: an empirical look at how capital juries perceive expert and lay testimony, *Virginia Law Review* **83**, 1109–1188.
[18] Sroufe, L.A. (1997). Psychopathology as an outcome of development, *Development and Psychopathology* **9**, 251–268.
[19] Herrenkohl, T.I., Maguin, E. & Hill, K.G. (2000). Developmental risk factors for youth violence, *Journal of Adolescent Health* **26**, 176–186.
[20] Coie, J.D., Watt, N.F., West, S.G., Hawkins, J.D., Asarnow, J.R. & Markman, H.J., Ramey, S.L., Shure, M.B. Long, B. (1993). The science of prevention, *American Psychologist* **48**, 1013–1022.
[21] Deater-Deckard, K., Dodge, K. & Bates, J.E. (1998). Multiple risk factors in the development of externalizing behavior problems: group and individual differences, *Development and Psychopathology* **10**, 469–493.
[22] Egeland, B., Carlson, E. & Sroufe, L.A. (1993). Resilience as process, *Development and Psychopathology* **5**, 517–528.
[23] Garmezy, N. & Masten, A.S. (1991). The protective role of competence indicators in children at risk, in *Life-Span Developmental Psychology: Perspectives on Stress and Coping*, E.M. Cummings, A.L. Greene & K.H. Karraker, eds, Lawrence Erlbaum Associates, Hillsdale, England, pp. 151–174.
[24] Masten, A.S. & Wright, M.O. (1997). Cumulative risk and protection models of child maltreatment, in *Multiple Victimization of Children: Conceptual, Developmental, Research and Treatment Issues*, B.B.R. Rossman & M.S. Rosenberg, eds, Haworth Press, Binghampton, pp. 7–30.
[25] Kazdin, A.E. (1997). Conduct disorder across the lifespan, in *Developmental Psychopathology: Perspectives on Adjustment, Risk, and Disorder*, S.S. Luthar, J.A. Burack, D. Cicchetti & J.R. Weisz, eds, Cambridge University Press, pp. 248–272.
[26] Cunningham, M.D. & Vigen, M.P. (2002). Death row inmate characteristics, adjustment, and confinement: a critical review of the literature, *Behavioral Sciences and the Law* **20**, 191–210.
[27] Smith, C., Lizotte, A.J., Thornberry, T.P. & Krohn, M.D. (1995). Resilient youth: identifying factors that prevent high-risk youth from engaging in delinquency and drug use, in *Delinquency and Disrepute in the Life Course*, J. Hagan, ed, Greenwich, CT, pp. 217–247.
[28] Rau, D. (2007). The scope of mitigation in the death penalty. *Journal of the American Academy of Psychiatry and the Law* **35**, 135–136.
[29] Cicchetti, D. (2003). Forward, in *Resilience and Vulnerability: Adaptation in the Context of Childhood Adversities*, S.S. Luthar, ed, Cambridge University Press, New York, p. XX.
[30] Garmezy, N. (1985). Stress-resistant children: the search for protective factors, in *Recent Research in Developmental Psychopathology*, J.E. Stevenson, ed, (*Journal of Child Psychology and Psychiatry* Book Supplement No. 4), Pergamon Press, Oxford, pp. 213–233.
[31] Rutter, M. (1985). Resilience in the face of adversity: protective factors and resistance to psychiatric disorder, *British Journal of Psychiatry* **147**, 598–611.
[32] Masten, A.S. & Powell, J.L. (2003). A resilience framework for research, policy, and practice, in *Resilience and Vulnerability: Adaptation in the Context of Childhood Adversities*, S.S. Luthar, ed, Cambridge University Press, New York, pp. 1–25.
[33] Werner, E. & Smith, R. (2001). *Journey From Childhood to Midlife: Risk, Resiliency, & Recovery*, Cornell University Press, New York.
[34] Fiske, S.T. & Taylor, S.E. (1984). *Social Cognition*, Addison-Wesley, Reading.
[35] Moore v. Parker, 425 F.3d 250 (6th Cir., 2005).

Further Reading

People v. Taylor, No. 123 (N.Y. Ct. App. filed Oct. 23, 2007).
People v. LaValle (3 N.Y.3d 88).
Baze v. Rees, 217 S.W. 3d 207, 209 (Ky. 2006), cert. granted, 76 U.S.L.W. 3154 (U.S. Sept. 25, 2007). (No. 07–5439).
http://www.gallup.com
National Omnibus Poll (2007). as cited in http://www.deathpenaltyinfo.org/article.php?did = 2163#DPIC07).

KAREN L. SALEKIN

Mitochondrial DNA: Interpretation

Introduction

The general principles of interpretation of mtDNA evidence have been detailed previously [1]; we follow the same principles and nomenclature here. Any mtDNA investigation seeks to address whether or not the donor of a reference sample (K) could be the donor, or a maternal relative of the donor, of a questioned sample (Q). The sequences obtained from the samples K and Q are referred to as SK and SQ, respectively.

Despite intense interest and many attempts to detect it, paternal inheritance of human mtDNA has only been demonstrated reliably in a single instance ([2]; for a critique of studies claiming to demonstrate

paternal inheritance of mtDNA, see [3]). Therefore, for the purposes of forensic interpretation of mtDNA evidence, we assume strict maternal inheritance.

What Constitutes a Match between *K* and *Q*?

If the result of the mtDNA analysis is that *K* and *Q* have the same sequence, this observation supports the contention that these samples have a common maternal origin. If *K* and *Q* have different sequences, then the contention of a common maternal origin is not supported by the observation. However, if the difference between *K* and *Q* is slight, the interpretation is not straightforward as the mutation rate of mtDNA is an order of magnitude greater than that of nuclear DNA. Differences in mtDNA sequence can be observed both between maternal relatives (e.g., [4]) and within a single individual (e.g., [5–9]). Two main approaches to evaluating whether an "inclusion", an "exclusion", or an "inconclusive" result is reported have been adopted. The first is a simple ruleset based on the number of differences between the samples: if no differences between the samples are observed, an inclusion is reported, if one difference is observed, the result is declared inconclusive and if two or more differences are observed, an exclusion is reported [10].

In the second approach, the general framework of the likelihood ratio (LR) is applied [1, 11, 12]:

$$\frac{\text{Probability of the evidence} \mid K \text{ and } Q \text{ have the same maternal origin}}{\text{Probability of the evidence} \mid K \text{ and } Q \text{ have different maternal origins}} = \frac{\Pr(E|H_p)}{\Pr(E|H_d)} \quad (1)$$

In this approach, any uncertainty about whether or not the mtDNA samples originated from the same maternal lineage is reflected in the numerator of the LR. Thus, if the sequences match exactly, the numerator would approach 1; if the sequences differ at many positions, the numerator tends toward 0, resulting in exclusion. For sequences, which differ at a single base or a small number of bases, the numerator is intermediate, depending on the mutation rate at the particular base(s) in question, the body tissue from which *K* and *Q* originate, and the number of generations separating the donors of these samples. Data regarding the mutability of individual mtDNA bases is accumulating, with some of the most useful information coming from phylogenetic analyses, which enable distinction to be made between bases that are highly polymorphic due to recurring mutations at highly mutable sites and those that are polymorphic because an ancient mutation has reached appreciable frequency in one or more populations (e.g., [13–16]). Several large studies of heteroplasmic segregation within (e.g., [5–9]) and between (e.g., [4]) individuals provide data to inform the assessment of impact of the body tissue and that of the generational divide, respectively.

Commonly Used Methods for Assessing Evidential Strength When an Inclusion is Reported

All methods for assessing evidential strength rely on population databases to gauge how common or rare is the sequence in question. Databases are considered in detail in the section "Databases".

The Counting Method

Many practitioners rely on the "counting method" to illustrate to the court the relative strength of a mtDNA match. In this method, the number of sequences matching that of the crime scene and suspect observed in the available databases is reported, together with the size of the databases. This method has the advantage of simplicity, and relies on no population genetics assumptions. For common haplotypes, it gives a relatively good estimate of the match probability, providing the database is relevant. However, for haplotypes that have not previously been observed in the database, the practitioner must ensure that the court is not left with the impression that failure to observe a haplotype in a database of, for example, several thousand sequences, indicates that the strength of the observed match is on a par with an autosomal DNA match such as a short tandem repeat (STR) match.

Upper Bound Frequency Estimation

Some have therefore extended the approach to calculate an upper 95% confidence interval for the

frequency estimate. Following Holland and Parsons [12], this is calculated as

$$p = 1 - \alpha^{1/n} \tag{2}$$

where $\alpha = 0.05$ for a 95% confidence interval.

Confidence intervals can also be calculated for frequency estimates for haplotypes that have previously been observed. Again following [12], this is calculated as

$$p \pm \sqrt{\frac{p(1-p)}{n}} \tag{3}$$

where n is the database size.

However, this method is unlikely to give a robust estimate of the confidence interval in most instances, as the normal approximation is poor when p is small [12].

An alternative approach is to add the observations of the sequence in the case to the database, in an approach similar to that advocated by Balding and Nichols [17] for addressing sampling error. Under the assumption of innocence, the profile from the crime sample and that from the reference sample do not originate from the same individual; the calculation is therefore

$$p = \frac{(x+2)}{(n+2)} \tag{4}$$

where x is the number of observations of the sequence and n is the database size [1].

Buckleton *et al.* [11] have suggested that the correct estimator for haploid genomes should be:

$$p = \frac{(x+1)}{(n+2)} \tag{5}$$

However, we contend that the estimator is intended to provide a common sense method to ensure that the strength of evidence is not overstated, given that the sequence has been observed (twice) in the case, even if not in the reference database, so prefer to adopt the more conservative and intuitive formula as originally stated [1]. Nevertheless, when either is fed into an LR, the result converges (Tables 1 and 2).

Table 1 $SQ = SK =$ [146C, 16129A, 16189C] (quoted as differences to the revised Cambridge Reference Sequence [18]); database searches carried out using www.empop.org

Method	Sampling correction	Count in Austrian Caucasian database	Result from Austrian Caucasian data	Count in Western European Caucasian database	Result from Western European Caucasian database
Counting method	None	0/273	No matching sequences found in 273 Austrians	0/3830	No matching sequences found in 3830 Western Europeans
Upper bound frequency estimation	$p = 1 - \alpha^{1/n}$ [12]	0/273	0.011	0/3830	0.0008
	$p = (x+2)/(n+2)$ [1]	0/273	0.007	0/3830	0.0005
	$p = (x+1)/(n+2)$ [11]	0/273	0.004	0/3830	0.0003
Likelihood ratio	$p = 1 - \alpha^{1/n}$ [12] $LR = 1/P$	0/273	91.6	0/3830	1279
	$p = (x+2)/(n+2)$ [1] $LR = 1/P$	0/273	137	0/3830	1916
	$LR = (n+3)/(x+2)$ $p = (x+1)/(n+2)$ [11]	0/273	138	0/3830	1916
Database structure method [17]	None	0/101	392	0/3830	Not calculated as κ unknown

Table 2 $SQ = SK =$ [263G, 340T, 523del, 524del] (quoted as differences to the revised Cambridge Reference Sequence [18])

Method	Sampling correction	Count in Austrian Caucasian database	Result from Austrian Caucasian data	Count in Western European Caucasian database	Result from Western European Caucasian database
Counting method	None	1/273	One matching sequence found in 273 Austrians	1/3830	One matching sequence found in 3830 Western Europeans
Upper bound frequency estimation	$p + \sqrt{(p(1-p)/n)}$ [12]	1/273	0.007	1/3830	0.0005
	$p = (x+2)/(n+2)$ [1]	1/273	0.011	1/3830	0.0008
	$p = (x+1)/(n+2)$ [11]	1/273	0.007	1/3830	0.0005
Likelihood Ratio	$p + \sqrt{(p(1-p)/n)}$ [12] $LR = 1/P$	1/273	137	1/3830	1915
	$p = (x+2)/(n+2)$ [1] $LR = 1/P$	1/273	92	1/3830	1277
	$LR = (n+3)/(x+2)$ $p = (x+1)/(n+2)$ [11]	1/273	92	1/3830	1278
Database structure method [19]	*n/a*			Not calculated as this method is for previously unobserved haplotypes only	

The Likelihood Ratio Approach

The general formulation of the LR is given in the section "What Constitutes a Match between K and Q". In adopting the LR approach, an estimate of the likelihood of the evidence conditional on K and Q having different maternal origins must come from some estimate of the frequency of the observed sequence in a relevant population. This estimate may be achieved *via* the counting method, or an extension of this method to correct for sampling error as described in the previous section.

The LR approach has the significant advantage of enabling the evidence to be considered as a whole: evaluation of whether or not K and Q match (section "What Constitutes a Match between K and Q") and evaluation of how frequently the sequences occur are weighed against each other, not considered in isolation. The usefulness of such an approach is best illustrated by examples as follows (updated from [1]):

1. *SK* and *SQ* match exactly, and there are no sequences the same or very similar to these in the database. The findings therefore support the proposition that K and Q have the same maternal origin. The counting method, a frequency estimate or a likelihood ratio (corrected for sampling error and distribution of mtDNA types in the population if necessary) could be used to assist the court in assessing the strength of support for this hypothesis.
2. *SQ* and *SK* differ by a single, frequently mutating base; neither sequence has previously been observed in the database, but there are other sequences in the database differing by a single base. The findings support the proposition that *SQ* and *SK* have a common maternal origin, but

the strength of this support is lower than that in example 1; this reduction is by a factor approximating the substitution rate, plus an additional small reduction to allow for the possibility that one of the similar sequences in the database may have the same maternal origin, albeit invoking an additional mutation.

3. *SQ* and *SK* are identical, both showing a heteroplasmic base at a site not known to be a mutation "hotspot";[a] the sequence (with either base at the heteroplasmic position) has not previously been observed in the database. The findings support the proposition that *Q* and *K* have the same maternal origin; this support is enhanced relative to that in example 1, by a factor approximating the substitution rate at the heteroplasmic base.
4. *SQ* and *SK* differ by a single base at which substitution has previously been observed, but which does not appear to be a mutation "hotspot"; *SQ* has been observed once in a database of 100 sequences, *SK* has not been observed before and there is an additional sequence in the database that differs from *SK* at a mutation "hotspot". There is significant uncertainty surrounding estimates for both the numerator and the denominator, but on balance, the *LR* is not likely to be significantly different to 1 and the result would be reported as inconclusive.
5. *SQ* and *SK* differ at a position that appears from phylogenetic and familial mutation rate estimates to be stable; *SQ* and *SK* have both been observed several times in the database. The findings support the proposition that *Q* and *K* have different maternal origins, and an exclusion is effectively reported.

Methods Not Yet Commonly Adopted

Strength of Evidence Calculation Based on Database Structure (Proportion of Singleton Haplotypes)

Brenner [19] proposed a method for calculating the evidential strength of rare Y chromosome haplotypes; this method is equally applicable to mtDNA evidential strength calculations where the crime sample has a previously unobserved haplotype

$$\Pr(T = SQ) \leq \frac{(1-\kappa)}{n} \tag{6}$$

where T is the haplotype of an innocent suspect for whom the probability of a chance match is being calculated; SQ is the crime scene profile; and K is the proportion of samples in the database that are singletons.

Hence, for example in a database where $\kappa = 80\%$,

$$\Pr(T = S) \leq \frac{1}{5n} \tag{7}$$

which in the simplest case of an exact match between crime scene profile and suspect profile results in a likelihood ratio

$$LR = \frac{1}{\Pr(T = S)} \geq 5n \tag{8}$$

Thus, as long as the crime sample is added to the database, the *LR* can be significantly greater than the size of the database. Incorporating sampling corrections and development of the method to allow for heteroplasmy and mutation have not yet been reported.

Phylogenetic Method for Calculating Match Probability

A method for evaluating match probabilities, taking into account the genealogical tree, which reflects the ancestral relationships between the sequences in a database has been reported by Wilson *et al.* [20]. This type of model can potentially take account of demographic issues such as population structure, thus avoiding the criticisms of Salas *et al.* [13], who also advocate phylogenetic approaches.

In essence, the coalescent model [21] and models for population growth and subdivision, together with mutation rate information specify the Bayesian prior distributions for the genealogical tree underlying a sample of DNA sequences. Inference about aspects of this tree is on the basis of its posterior distribution, given the observed data.

The authors used Monte Carlo Markov Chain (MCMC) algorithms for the analysis, in which the tree consisted of n (database size) $+1$ (crime scene profile) observed sequences. A new branch connecting an unobserved sequence, x, of an alternative possible culprit with an unknown mtDNA sequence with the tree *via* a new node, z, was introduced (see Figure 1; from [20]). At each iteration of the algorithm, the probability that the sequence x matches s was calculated, conditional on the location and state

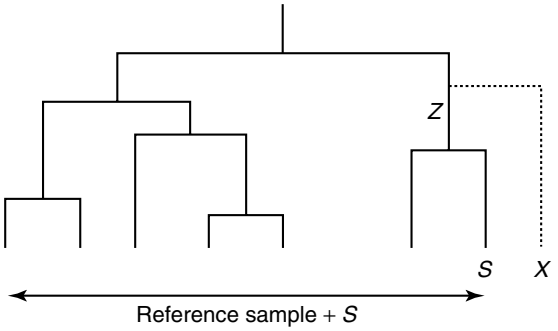


Figure 1 From [20]: Representation of the maternal genealogy of $n = 6$ individuals in an anonymous reference database, together with the suspect, s, and a further individual, x, regarded as an alternative suspect of unknown mtDNA type: the node labeled z corresponds to the most recent woman ancestral to both x and at least one of the other $n = 1$ individuals [Reproduced from Ref.16. © Elsevier, 2005.]

of the node z. The average of these conditional probabilities over 10^5 MCMC outputs approximates the match probability (P_m).

Using this algorithm under the standard coalescent (no population growth or subdivision), this match probability was calculated for the following:

1. the most common haplotype;
2. a haplotype not previously observed, but very similar to the most common haplotype; and
3. a haplotype not previously observed and very different to any observed haplotype.

from a dataset of mtDNA minisequence haplotypes in British Caucasians, Afro-Caribbeans, and Asians [22]. These match probabilities were compared with frequency estimates generated using the counting method amongst $n + 1$ observed sequences, and very similar results were obtained. Counter intuitively the match probability of the dissimilar haplotype in the Afro-Caribbean and Asian datasets was higher than that for the haplotype most similar to the most common haplotype.

The authors concluded that the counting method was usually, although not always, a conservative method. The coalescent model with growth gave similar results, but the model with population splitting was not applied. It would be a profitable area of research for mtDNA interpretation to use similar phylogenetic methods to study the effects of population substructure and establish the circumstances under which the counting method is not conservative.

Databases

Database Relevance and Availability

The availability of relevant databases is critical to the forensic interpretation of mitochondrial data. Whereas the shuffling of preexisting variants by recombination is the principal source of the discriminating power of autosomal STR multiplexes, mtDNA variation only arises through *de novo* mutation events, which are destined to remain associated with the genetic background of the progenitor molecule. Consequently the frequency of a mitochondrial haplotype cannot be predicted by simple multiplication of the relevant allele frequencies at each polymorphic site as for autosomal STRs, but must be determined by direct observation of the entire haplotype within a database.

The strength of mitochondrial evidence is heavily dependant upon the size of the database, particularly when a novel haplotype is involved; unfortunately the haploid nature of the mitochondrial genome also increases its susceptibility to genetic drift leading to higher F_{st} values and greater variation between geographically and ethnically distinct populations. This in turn necessitates the construction of databases for each subpopulation that may give rise to an evidential sample in order that the appropriate strength of evidence can be derived. Furthermore, the greater cost and complexity of sequencing combined with the difficulty of acquiring samples from the appropriate populations generally means that the available databases are often small or derive from populations that do not faithfully reflect the composition of the potential source gene pool. Many laboratories have little option but to make use of databases from areas outside of their jurisdiction. In cosmopolitan populations, where significant migration over a number of generations is the rule, it is likely that the typically general population databases available are sufficiently representative of the population from which the offender originated. Studies of mtDNA variability within and between European populations have confirmed this trend for most European cosmopolitan populations [23], although the resolution of this analysis would not detect localized differences in, for example, small village populations.

Mitochondrial sequence databases covering differing portions of the genome have been compiled for both forensic and academic purposes typically covering both HVI and HVII regions (e.g., Federal Bureau of Investigation (FBI)/Scientific Working Group on DNA Analysis Methods (SWGDAM): http://www.fbi.gov/hq/lab/fsc/backissu/april2002/miller1.htm [24] & European DNA Profiling Group (EDNAP) mtDNA Population Database (EMPOP): http://www.empop.org [25]); Mitomap: http://www.mitomap.org/ [26]) with a rapidly increasing but still limited number of whole genome sequences becoming available (http://www.genpat.uu.se/mtDB/ [27]). Although it is becoming clear that many highly informative Single Nucleotide Polymorphisms (SNPs) are located within the coding regions and much effort is being applied to detecting these in an efficient way (e.g., [28]), database size and the limited quantities of crime stain material generally confines casework sequencing to the hypervariable regions, which are most likely to discriminate between potential sources. Variant sites are typically recorded as changes from the revised Cambridge Reference Sequence [18] allowing comparisons between databases, but one should note that the span of the regions examined may not correspond precisely between data sets. Phylogenetic studies on the basis of this accumulated data have highlighted great variability in mutation rate, subsequent persistence, and distribution of variation at different bases (e.g., [13–16]). The two hypervariable regions being noncoding have fewer functional constraints and therefore can tolerate the greatest polymorphism; some sites are particularly unstable and are frequently heteroplasmic, including two poly C regions, which are inclined to slippage. Such regions are variable upon many branches of the phylogenetic tree due to recurrent mutation, while many slower mutating sites are polymorphic only in a single lineage.

Knowledge of the phylogenetic relationships between populations on the basis of their patterns of haplogroup sharing has greatly increased our understanding of the complex genetic origins of ethnic subpopulations. Many mitochondrial clades are strongly associated with geographic regions and have been used to trace the migration of humans from Africa to colonize the world. Demographic studies have highlighted several instances in which different patterns of genetic contribution are apparent for females (mtDNA) and males (Y chromosome), this being apparent in many Hispanic populations where the indigenous female contribution is very strong, but much reduced with respect to Y chromosomes, where European origins predominate but the extent to which this is apparent depends on the population being examined [29, 30].

Recent periods of mass migration, either forced or by choice, have greatly altered the mtDNA gene pool of many regions but care must be taken to ensure that databases, which appear to share similar ethnic origins, either from the continent of origin or resulting from dispersals to other parts of the globe, are indeed reliable substitutes. Similarly tribal or caste differences may exist even among indigenous people, which should be accounted for by ensuring that sampling accurately reflects the current composition of the area in question.

Database formats range from simple tables in published papers through to sophisticated online databases. The level of checking prior to publication may also vary from a single automated sequence read per sample through to duplication of sequencing using primers on both strands upon templates derived from separate duplicate extractions or polymerase chain reactions (PCRs) with witnessing of all manual stages. Online databases such as FBI/SWGDAM (http://www.fbi.gov/hq/lab/fsc/backissu/april2002/miller1.htm) and EMPOP (http://www.empop.org) are openly accessible and have been updated with additional samples or corrections as appropriate, indeed the EMPOP database includes the sequence electropherograms to aid confirmation of profile quality and accuracy [31]. Full details of the sample's geographic and ethnic origin are submitted so that its relevance to an enquiry can be fully determined. Furthermore, all potential contributors have to correctly genotype a number of quality control samples before their samples are uploaded and every haplotype is checked using the Network phylogenetic software package (http://www.fluxus-technology.com/) to screen for potential sequencing or transcription errors.

Concerns Regarding Database Quality

A series of publications have used the phylogenetic approach to highlight the likely occurrence of errors in the majority of databases (e.g., [32–34]). While genuine recurrence of mutations at hypervariable positions results in small reticulations within

phylogenetic networks in which the temporal order of mutations cannot be determined, when groups of closely spaced variants reoccur on widely separated branches of the network, this has usually provided evidence of artificial recombination resulting from errors during the transcription of data or mix-ups with tubes, which have "recombined" HVI and HVII regions from different individuals [35]. When individual slow-mutating sites show variants on widely separated branches this highlights sequences, which should be checked for accuracy but may in fact be authentic. The majority of errors are thought to occur during transcription of data to the published tables and their frequency is therefore reduced by software packages that reduce the amount of human intervention. Other potential causes include degraded samples in which contaminant DNA overwhelms an unduplicated PCR reaction, or a primer binding site polymorphism, which knocks out amplification of the authentic sequence.

Further sources of error include base shifts in tabulated data, where the positions relative to the revised cambridge reference sequence (rCRS) are incorrect; these can involve shifts of a single base, by several positions (often 10 or 100), or into an adjacent column. Phantom mutations, where unlikely transversions appear in sequences may be owing to transcription errors or problems with the sequencing chemistry, which result in hard to interpret positions [33]. Reference bias, where a change from the rCRS is overlooked and therefore goes unreported, and base mis-scoring when transversions are mis-recorded as the more common transitions also occur frequently (all discussed in detail in [32]). Virtually all of these errors could be avoided by duplication of the sequence using primers on both strands, the independent comparison of these sequences with the reference and their recording in two different formats e.g., motif and dot table, which can subsequently be compared electronically to highlight any discrepancies [31, 34].

As errors have been highlighted by these publications in widely used databases such as the SWGDAM Mitochondrial DNA Database, much effort has been made to purge these and test what effect they may have had on reports produced with the imperfect data. Budowle *et al.* [36] argue that sequencing errors have largely been eliminated owing to improved chemistries, equipment and software and the widespread adoption of duplication by forensic laboratories as outlined in International Society for Forensic Genetics (ISFG) guidelines [37]. Using up-to-date casework techniques discrepancies are exceedingly rare demonstrating that these safeguards successfully eliminate the vast majority of potential errors. They also conclude that the impact on reporting is marginal. The likely effect of any error is to create a new haplotype that does not already exist in nature, therefore assuming the error rate is low it merely reduces the true size of the database by the proportion of profiles, which contain errors. Although there is no justification for complacency and every effort should be made to rigorously control and improve the accuracy of databases used in forensics, it is unlikely that the error levels that had existed would have counteracted the conservative allowances that are made for database size.

Database Size

How large should a database be? Clearly this depends on many factors, primarily the diversity within the population. This has been modeled by Pereira *et al.* [38] using samples from a database of 549 individuals from across Portugal in comparison with 1 200 Germans from a single village [39]. While haplotype diversity peaked with samples of 300 in Portugal and 400 in the German sample, the number of haplotypes continued to rise almost linearly. The proportion of Portuguese haplotypes that were unique for HVI declined from around two-thirds when the sample size was 50 to about a third in the full database of 549, while for HVII the decline was much faster from 39.0% to 14.9% as the smaller number of polymorphic sites in this region are subject to a higher rate of recurrent mutation. When both HV regions are considered together the proportion of unique sequences remained above 50% even in the full database and modeling suggests that a sample size in excess of 1 300 would be required to reduce the proportion of novel haplotype observations below 5% among each subsequent addition of 100 samples. Given that the Portuguese population is less diverse than Central Europeans, it is clear that in most instances databases ideally need to number in the thousands before they cease to reveal a substantial number of previously unobserved haplotypes.

Worked Examples

To illustrate the practical impact of using the different interpretation methods and databases described earlier, the methods detailed have been applied to two example sequences, using two databases, both drawn from the EMPOP database: 273 Austrian Caucasians and 3830 Western Europeans. The phylogenetic method was not included, as the algorithms were developed for minisequence data, not full sequence data; computationally, full sequence data are more challenging. Additionally, for the database structure method, only 101 Austrian Caucasians were considered, as the structure of this dataset had previously been published [40].

End Notes

[a.] Hotspots are regions of the genome (usually single bases or short homopolymeric runs) that have greatly increased mutation rates to the extent that the same mutation is found in several different haplotypic backgrounds due to recurrent mutation.

References

[1] Tully, G., Bär, W., Brinkmann, B., Carracedo, A., Gill, P., Morling, N., Parson, W. & Schneider, P. (2001). Considerations by the European DNA profiling (EDNAP) group on the working practices nomenclature and interpretation of mitochondrial DNA profiles, *Forensic Science International* **124**, 83–91.

[2] Schwartz, M. & Vissing, J. (2002). Paternal inheritance of mitochondrial DNA, *New England Journal of Medicine* **347**, 576–580.

[3] Bandelt, H.-J., Kong, Q.-P., Parson, W. & Salas, A. (2005). More evidence for non-maternal inheritance of mitochondrial DNA? *Journal of Medical Genetics* **42**, 957–960.

[4] Parsons, T.J., Muniec, D.S., Sullivan, K., Woodyatt, N., Alliston-Greiner, R. Wilson, M.R., Berry, D.L., Holland, K.A., Weedn, V.W., Gill, P. & Holland, M.M. (1997). A high observed substitution rate in the human mitochondrial DNA control region, *Nature Genetics* **15**, 363–368.

[5] Sullivan, K.M., Alliston-Greiner, R., Archampong, F.I.A., Piercy, R., Tully, G., Gill, P. & Lloyd-Davies, C. (1996). A single difference In mtDNA control region sequence observed between hair shaft and reference samples from a single donor, *Proceedings of the Seventh International Symposium on Human Identification, Scottsdale (AZ), Promega Corporation*, pp. 126–129.

[6] Wilson, M.R., Polansky, D., Replogle, J., DiZinno, J.A. & Budowle, B. (1997). A family exhibiting heteroplasmy in the human mitochondrial DNA control region reveals both somatic mosaicism and pronounced segregation of mitotypes, *Human Genetics* **100**, 167–171.

[7] Calloway, C.D., Reynolds, R.L., Herrin, G.L. & Anderson Jr, W.W. (2000). The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age, *American Journal of Human Genetics* **66**, 1384–1397.

[8] Tully, L.A., Parsons, T.J., Steighner, R.J., Holland, M.M., Marino, M.A. & Prenger, V.L. (2000). A sensitive DGGE assay reveals a high frequency of heteroplasmy in hypervariable region one of the human mitochondrial DNA control region, *American Journal of Human Genetics* **67**, 432–443.

[9] Paneto, G.G., Martins, J.A., Longo, L.V., Pereira, G.A., Freschi, A. & Alvarenga, V.L. (2007). Heteroplasmy in hair: differences among hair and blood from the same individuals are still a matter of debate, *Forensic Science International* **173**, 117–121.

[10] Scientific Working Group on DNA Analysis Methods (2003). Guidelines for mitochondrial DNA (mtDNA) nucleotide sequence interpretation, *Forensic Science Communications* **5**, at http://www.fbi.gov/hq/lab/fsc/backissu/april2003/swgdammitodna.htm.

[11] Buckleton, J., Walsh, S. & Harbison, S. (2005). Nonautosomal forensic markers, in *Forensic DNA Evidence Interpretation*, J. Buckleton, C.M. Trigg & S. Walsh, eds, CRC Press, Florida, pp. 299–339.

[12] Holland, M. & Parsons, T. (1999). Mitochondrial DNA sequence analysis – validation and use for forensic casework, *Forensic Science Review* **11**, 21–50.

[13] Salas, A., Bandelt, H.-J., Macaulay, V. & Richards, M.B. (2007). Phylogenetic investigations: the role of trees in forensic genetics, *Forensic Science International* **168**, 1–13.

[14] Galtier, N., Enard, D., Radondy, Y., Bazin, E. & Belkhir, K. (2006). Mutation hotspots in mammalian mitochondrial DNA, *Genome Research* **16**, 215–222.

[15] Gurven, M. (2000). How can we distinguish between mutational "hotspots" and "old sites" in human mtDNA samples? *Human Biology* **72**, 455–471.

[16] Allard, M.W., Polanskey, D., Miller, K., Wilson, M.R., Monson, K.L. & Budowle, B. (2005). Characterization of human control region sequences of the African American SWGDAM forensic mtDNA data set, *Forensic Science International* **148**, 169–179.

[17] Balding, D.J. & Nichols, R.A. (1994). DNA profile match probability calculation: how to allow for population stratification relatedness database selection and single bands, *Forensic Science International* **64**, 125–120.

[18] Andrews, R., Kubacka, I., Chinnery, P., Lightowlers, R., Turnbull, D. & Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, *Nature Genetics* **23**(2), 147.

[19] Brenner, C. (2006). Evidential strength of a rare haplotype, *Oral presentation at: DNA in Forensics*, 28-30 September, Innsbruck.

[20] Wilson, I.J., Weale, M.E. & Balding, D.J. (2003). Inferences from DNA data: population histories evolutionary processes and forensic match probabilities, *Journal of the Royal Statistical Society A* **166**, 155–201.
[21] Kingman, J.F.C. (1982). The coalescent, *Stochastic Processes and their Applications* **13**, 235–248.
[22] Tully, G., Sullivan, K.M., Nixon, P., Stones, R.E. & Gill, P. (1996). Rapid detection of mitochondrial DNA sequence polymorphisms using multiplex solid-phase fluorescent minisequencing, *Genomics* **16**, 97–159.
[23] Melton, T., Wilson, M., Batzer, M. & Stoneking, M. (1997). Extent of heterogeneity in mitochondrial DNA of European populations, *Journal of Forensic Science* **42**, 437–446.
[24] Monson, K.L., Miller, K.W.P., Wilson, M.R., DiZinno, J.A. & Budowle, B. (2002). The mtDNA Population Database: an integrated software and database resource for forensic comparison, *Forensic Science Communications* http://www.fbi.gov/hq/lab/fsc/backissu/april2002/miller1.htm **4**.
[25] Parson, W., Brandstätter, A., Alonso, A., Brandt, N., Brinkmann, B., Carracedo, A., Corach, D., Froment, O., Furac, I., Grzybowski, T., Hedberg, K., Keyser-Tracqui, C., Kupiec, T., Lutz-Bonengel, S., Mevag, B., Ploski, R., Schmitter, H., Schneider, P., Syndercombe-Court, D., Sørensen, E., Thew, H., Tully, G. & Scheithauer, R. (2004). The EDNAP mitochondrial DNA population database (EMPOP) collaborative exercises: organization results and perspectives, *Forensic Science International* **139**(2–3), 215–226.
[26] Ruiz-Pesini, E., Lott, M.T., Procaccio, V., Poole, J.C., Brandon, M.C. & Mishmar, D. (2007). An enhanced MITOMAP with a global mtDNA mutational phylogeny, *Nucleic Acids Research* **35**(Database issue), D823–D828.
[27] Ingman, M. & Gyllensten, U. (2006). mtDB: Human Mitochondrial Genome Database a resource for population genetics and medical sciences, *Nucleic Acids Research* **34**(Database issue), D749–D751.
[28] Coble, M.D., Vallone, P.M., Just, R.S., Diegoli, T.M., Smith, B.C. & Parsons, T.J. (2006). Effective strategies for forensic analysis in the mitochondrial DNA coding region, *International Journal of Legal Medicine* **120**, 27–32.
[29] Allard, M.W., Polanskey, D., Wilson, M.R., Monson, K.L. & Budowle, B. (2006). Evaluation of variation in control region sequences for Hispanic Individuals in the SWGDAM mtDNA data set, *Journal of Forensic Science* **51**(3), 566–573.
[30] Hammer, M.F., Chamberlain, V.F., Kearney, V.F., Stover, D., Zhang, G. & Karafet, T. (2006). Population structure of Y chromosome SNP haplogroups in the United States and forensic implications for constructing Y chromosome STR databases, *Forensic Science International* **164**(1), 45–55.
[31] Brandstatter, A., Niederstatter, H., Pavlic, M., Grubwieser, P. & Parson, W. (2007). Generating population data for the EMPOP Database – An overview of the mtDNA sequencing and data evaluation processes considering 273 Austrian control region sequences as example, *Forensic Science International* **166**, 164–175.
[32] Bandelt, H.J., Lahermo, P., Richards, M. & Macaulay, V. (2001). Detecting errors in mtDNA data by phylogenetic analysis, *International Journal of Legal Medicine* **115**, 64–69.
[33] Brandstatter, A., Sanger, T., Lutz-Bonengel, S., Parson, W., Beraud-Colomb, E. & Wen, B.J. (2005). Phantom mutation hotspots in human mitochondrial DNA, *Electrophoresis* **26**(18), 3414–3429.
[34] Salas, A., Carracedo, C., Macaulay, V., Richards, M. & Bandelt, H.-J. (2005). A practical guide to mitochondrial DNA error prevention in clinical forensic and population genetics, *Biochemical and Biophysical Research Communications* **335**, 891–899.
[35] Bandelt, H.J., Salas, A. & Lutz-Bonengel, S. (2004). Artificial recombination in forensic mtDNA population databases, *International Journal of Legal Medicine* **118**(5), 267–273.
[36] Budowle, B., Polanskey, D., Allard, M.W. & Chakraborty, C. (2004). Addressing the use of phylogenetics for the identification of sequences in error in the SWGDAM Mitochondrial DNA Database, *Journal of Forensic Science* **49**(6), 1256–1261.
[37] Bär, W., Brinkmann, B., Budowle, B., Carracedo, A., Gill, P. & Holland, M. (2000). DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing, *International Journal of Legal Medicine* **113**(4), 193–196.
[38] Pereira, L., Cunha, C. & Amorim, A. (2004). Predicting sampling saturation of mtDNA haplotypes: an application to an enlarged Portuguese database, *International Journal of Legal Medicine.* **118**, 132–136.
[39] Pfeiffer, H., Forster, P., Ortmann, C. & Brinkmann, B. (2001). The results of an mtDNA study of 1200 inhabitants of a German village in comparison to other Caucasian databases and its relevance for forensic casework, *International Journal of Legal Medicine* **114**, 169–172.
[40] Parson, W., Parsons, T.J., Scheithauer, R. & Holland, M.M. (1998). Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case, *International Journal of Legal Medicine* **111**, 124–132.

Related Articles

Databases

Mitochondrial DNA: Profiling

Short Tandem Repeats

GILLIAN TULLY AND JON WETTON

Mitochondrial DNA: Profiling

Introduction

Mitochondrial deoxyribonucleic acid (mtDNA) analysis is a routine adjunct to crime scene investigation. Introduced in the early 1990s to aid with the identification of military remains, and implemented since then by the Armed Forces DNA Identification Laboratory in Rockville, Maryland, in all military conflicts [1], it was first used in criminal justice proceedings by the FBI in 1996 in *Tennessee v. Paul William Ware* [2]. In that case, a single hair located in the throat of a victim linked Ware to a homicide. MtDNA provides a valuable locus for forensic DNA typing in certain circumstances, especially for skeletal remains, shed hairs, or hair fragments, and degraded samples of all types. In general, it is used when short tandem repeat (STR) testing is not possible owing to limited or degraded nuclear DNA because mtDNA is naturally abundant and resistant to degradation [3]. MtDNA cannot be a unique identifier, as can nuclear DNA, due to its pattern of maternal inheritance, nor does it have the statistical power of a nuclear DNA match; as such, it is used as supplementary circumstantial evidence in criminal cases. MtDNA analysis has played a large role in the identification of missing persons when applied to skeletal remains.

Mitochondrial DNA Biology

Deoxyribonucleic acid (DNA) is found in two locations in all human cells except red blood cells (*see* **DNA**). Nuclear deoxyribonucleic acid (nuDNA), inherited from both parents, makes up 26 pairs of chromosomes in the nucleus. Mitochondrial deoxyribonucleic acid (mtDNA), inherited only from the mother, is located in the mitochondria, small, peanut-shaped cytoplasmic organelles that generate cellular energy. The full complement of nuDNA has about 3 billion of the four chemical bases of DNA (adenine, guanine, thymine, and cytosine, abbreviated as A, G, T, and C; also known as *nucleotides*) in a linear array within the chromosomes. Human mtDNA contains approximately 16 569 nucleotides in a small circular molecule. Whereas each cell contains two copies of nuDNA, there are hundreds to thousands of mtDNA molecules within dozens to hundreds of mitochondria per cell, depending on the particular tissue. With some exceptions, all tissues in an individual are homogeneous for the single mtDNA type, or sequence, of DNA nucleotides in that individual's mtDNA molecules.

The mtDNA types, sequences, or "profiles" present among humans have been generated by mutational changes occurring in this DNA over many generations. All living humans can be linked by their profiles into a single large tree and share a common maternal ancestor; those individuals with similar types are most closely clustered in the tree. MtDNA reflects the biogeographical ancestry of a particular maternal lineage. For example, certain types may be readily recognizable as having originated in Asia, Africa, or Europe. MtDNA is passed intact from a mother to all her children; males inherit their mother's mtDNA but do not pass it on to their children. For this reason, all maternally related individuals share the same mtDNA profile.

The natural abundance of mtDNA is the key to its forensic utility. MtDNA recovery from small or degraded biological samples is greater than nuDNA recovery owing to the high copy number (there are many mitochondria in a cell, but one nucleus) and because the molecule's small circular structure may protect it from damage by heat, humidity, acidity, and ultraviolet (UV) light. In addition, nuDNA in naturally shed hairs and hair shafts is extremely limited even when these samples are freshly collected [4].

The mtDNA molecule codes for 13 proteins, two ribosomal RNAs, and 22 transfer RNAs, and also contains a 1122 bp "non-coding" region, sometimes called *the control region* or *D-loop*, which is forensically informative [5]. The DNA sequence differs so much among individuals in two "hypervariable" sections of the control region that the likelihood of choosing two people at random with the same mtDNA sequence is very low [6]. About 8–12 nucleotide differences would be observed between two maternally unrelated individuals for the regions that are analyzed, comprised of DNA sequence about 700 bp long [7].

Candidates for forensic mtDNA typing analyses are as follows: (i) shed hairs with no follicle, tissue, or root bulb attached, (ii) hair shaft fragments, (iii) bones or teeth that have been subjected to long periods of high acidity, high temperature, or

high humidity, (iv) stain or swab material that has been unsuccessfully typed for nuDNA markers, and (v) tissue (skin, muscle, organ) that has been unsuccessfully typed for nuDNA markers. While mtDNA typing of blood, semen, and saliva crime scene stains from clothing and floors is possible, it is likely that mixtures will be obtained due to the extreme sensitivity of this form of typing in samples that, unlike hairs and bones, are difficult to clean before DNA extraction. On the other hand, degraded samples collected from near-sterile or UV radiation–exposed surfaces, such as the exterior of a vehicle, may easily provide single-source mtDNA profiles.

The Analytical Process

An mtDNA analysis begins when total genomic DNA is extracted from biological material such as a tooth, blood sample, or hair. Extraction methods, such as grinding of hair or bone or complete dissolution of proteinaceous material are designed and validated in order to optimize the yield of DNA for specific sample types. In addition, the external surfaces of samples are thoroughly cleaned via sanding with a rotary drill bit (for bones) or ultrasonic water baths (for hairs). Approximately 2 cm of a single hair is used in the average case, though success has been achieved with much less; for skeletal remains, approximately 0.1–0.5 g of bone is used [8, 9].

Following extraction, the polymerase chain reaction (PCR) is used to amplify the two hypervariable portions of the noncoding region using flanking primers. Primers are small bits of DNA that identify and hybridize to or adhere to the ends of the region one wishes to PCR amplify, therefore targeting a region for amplification. Primer pairs have been designed and manufactured to encompass virtually any region of mtDNA in humans, and may include "mini-primer pairs" that can recover the smallest fragments of DNA from a degraded sample, usually under 150 bp in length [10] (*see also* **DNA: Degraded Samples**).

Because the natural abundance of mtDNA as well as the creation of PCR product introduces many copies of mtDNA into the laboratory, care is taken to eliminate the introduction of exogenous (contaminating) DNA during both the extraction and amplification steps by methods such as the use of prepackaged sterile equipment and reagents, aerosol-resistant barrier pipette tips, gloves, masks, and lab coats, separation of pre- and post-amplification areas in the lab using dedicated reagents for each, ultraviolet irradiation of equipment, and autoclaving of tubes and reagent stocks. In forensic casework, questioned samples are processed at different times than known samples and in different laboratory rooms. Most importantly, several negative controls that would indicate the presence of contamination introduced during testing are run in parallel with all samples. Overall, contamination is more of a concern for the mtDNA laboratory than the nuclear DNA laboratory, but each laboratory determines through internal validation studies how contamination and its control, detection, and interpretation can impact casework and still result in a defensible outcome [11].

When adequate amounts of PCR product are amplified from the two hypervariable regions, as determined from either a yield gel or other detection method, sequencing reactions are performed. These chemical reactions use each PCR product as a template to create a new complementary strand of DNA in which some nucleotides are labeled with dye. The strands created at this stage are then separated according to size by an automated sequencer that uses a laser to "read" the sequence. Where possible, the sequences of both hypervariable regions (called *HV1* and *HV2*) are determined on both strands of the double-stranded DNA molecule and in overlapping PCR products, with sufficient redundancy to confirm the nucleotide sequence that characterizes that particular sample.

Two forensic analysts independently assemble the mtDNA sequence and then compare it to a standard published reference sequence called *the revised Cambridge Reference Sequence* (rCRS, [12, 13]), denoting all the nucleotide differences. The entire process is then repeated with a known sample, usually blood, saliva, or a buccal swab, collected from a known individual. The sequences from both samples, about 780 nucleotides each, are compared to determine if they match. Depending on data quality or ambiguities, portions of the analysis may be repeated.

Heteroplasmy

Heteroplasmy is defined as the presence of two or more types of mtDNA within an individual

[14]. There are two forms of heteroplasmy: length heteroplasmy and site (or sequence) heteroplasmy. The baseline state of mtDNA composition in humans, with the exception of individuals with tissue-specific mitochondrial diseases, is homoplasmy. That is, the overwhelming majority of mtDNA-containing cells within an individual contain the same 16 569 bp mtDNA molecule throughout. The exception to this dominant state of uniformity is the frequent occurrence of length heteroplasmy in certain control region strings of cytosine residues ("C-stretches"), which creates populations of mtDNA molecules in each cell that differ slightly in length. Length heteroplasmy occurs in about 50% of all individuals. Site or sequence heteroplasmy, where at a single nucleotide address there are two different DNA bases such as T and C, has been observed in approximately 1% of blood samples and 10–15% of hair samples, with other tissues such as bone believed to be intermediate with respect to frequency. Some degree of heteroplasmy exists in all individuals; precisely how detectable and abundant it is becomes the focus of the forensic DNA practitioner. Guidelines are derived from each laboratory's validation studies to allow for conservative interpretations of heteroplasmy such that false failures to exclude cannot occur when it is present.

Interpretation

The FBI's Scientific Working Group on DNA Analysis Methods (SWGDAM) Guidelines for mtDNA Nucleotide Sequence Interpretation ((http://www.fbi.gov/hq/lab/fsc/backissu/april2003/swgdammitodna.htm) state:

> The following guidelines may be used in most cases:
>
> - Exclusion
>
> If there are two or more nucleotide differences between the questioned and known samples, the samples can be excluded as originating from the same person or maternal lineage.
>
> - Inconclusive
>
> If there is one nucleotide difference between the questioned and known samples, the result will be inconclusive.
>
> - Cannot Exclude
>
> If the sequences from questioned and known samples under comparison have a common base at each position or a common length variant in the HV2 C-stretch, the samples cannot be excluded as originating from the same person or maternal lineage.

In the event of a "cannot exclude" result, the SWGDAM mtDNA database (http://www.fbi.gov/hq/lab/fsc/backissu/april2002/miller1.htm) is searched for the mitochondrial sequence that has been observed for the samples. At present, the SWGDAM database of human mtDNA sequences has around 5000 sequences available for a search of a casework sequence, but an increase to over 10 000 samples is expected within one to two years. The current convention in the event of a failure to exclude is for the analyst to report the number of times the observed sequence is present in the database in order to estimate its relative frequency in the population. A frequency statistic is calculated, and a 95 or 99% confidence interval is placed around the estimated frequency to account for the inherent uncertainty in the frequency calculation since one would never be able to type all living humans [1]. This convention of using the upper-bound frequency provides a very conservative approach to estimating how many individuals at most would be expected to have a particular type. For example, a novel type (one that has not been observed previously in the database) would be estimated to occur at most in 6 in 10 000 individuals based on the size of the database currently being used.

While around 60% of all mtDNA types appear to be rare (occurring a single time) in each of the FBI's ethnic sub-databases (individuals claiming to be African or of African origin, Asian or of Asian origin, Caucasian or of European origin, or Hispanic), there is one mtDNA sequence that is seen in around 7% of Caucasians. Interestingly, the frequency of this type has remained fairly stable as the database has grown over the last 10 years. However, almost two-thirds of the newly typed samples have novel sequences; therefore, all the mtDNA variation present in the general human population has not yet been identified. One exception to the high diversity that is present within ethnic databases has been observed: due to founder effects, mtDNA diversity in native Americans is limited, and certain types in this group are observed at high frequencies [15].

In addition to the SWGDAM database, over 20 000 human mtDNA control region sequences

have been characterized by anthropologists and forensic scientists (see the many references to human mtDNA populations that have been studied at http://www.mitomap.org/, as well as [16]). In general, the pattern observed in most populations around the world, with the exception of some populations of anthropological interest, is that most of the sequences are uncommon, and few types are present at frequencies greater than 1%. Because of this fact, it is possible to exclude greater than 99% of a population as potential contributors of a sample in most cases, except where one is dealing with a more common type. In contrast, a multilocus nuclear DNA typing profile provides vastly superior discriminatory power and statistics that permit source attribution. For this reason, mtDNA can never provide the resolution of individuality that nuDNA typing can.

As mentioned above, mtDNA is maternally inherited, so that any maternally related individuals would be expected to share the same mtDNA sequence. This fact is useful in cases where a long deceased or missing individual is not available to provide a reference sample but any living maternal relative might do so. Because of meiotic recombination and the diploid (biparental) inheritance of nuDNA, the reconstruction of a nuDNA profile from even first-degree relatives of a missing individual is rarely this straightforward. This feature has allowed some important historical mysteries to be solved, such as the identification of the remains of the assassinated Romanov family [17], the Vietnam Unknown Soldier [18], and American outlaw gunfighter Wild Bill Longley [19]. However, the maternal inheritance pattern of mtDNA might also be considered problematic. Because all individuals in a maternal lineage share the same mtDNA sequence, mtDNA cannot be considered a unique identifier. In fact, apparently unrelated individuals might share an unknown maternal relative at some distant point in the past. For this reason, it is important that judges, attorneys, and juries in criminal proceedings are educated about the maternal inheritance pattern of mtDNA.

Nonforensic Uses

While mtDNA is useful for forensic examinations, it has also been used extensively in two other major scientific realms. First, there are a number of serious inherited diseases caused by deleterious mutations in gene-coding regions of the mtDNA molecule [20]. In addition, molecular anthropologists have been using mtDNA for three decades to examine both the extent of genetic variation in humans and the relatedness of populations all over the world [21]. An mtDNA maternal inheritance pattern can reveal ancient population histories, which might include migration patterns, expansion dates, and geographic homelands. MtDNA has been recovered from several Neanderthal skeletons, and the resulting population genetics studies have allowed anthropologists to conclude that modern humans do not share a close relationship with Neanderthals in the human evolutionary tree [22]. The general methods for performing all mtDNA analyses, including forensic methods, are identical to those used in molecular biology laboratories all over the world for studying DNA from any living organism. There are several thousand published articles on mtDNA available at http://www.mitomap.org/.

Laboratory Practices

MtDNA analysis is offered by the FBI in their Quantico facility as well as in four regional laboratories. Several private commercial labs also offer testing, as do the Armed Forces DNA Identification Lab, University of North Texas Health Science Center, Office of the Chief Medical Examiner in New York, and California Department of Justice. Most, if not all, of these laboratories are guided in application of mtDNA analysis by federal Quality Assurance Standards for DNA testing and various accrediting bodies.

MtDNA matching has become an extremely valuable resource for the FBI's National Missing Person DNA Database program. Under this program, maternal relatives submit samples for typing and inclusion to the database for eventual comparison to profiles obtained from recovered skeletal remains.

Although mtDNA is still undergoing admissibility hearings in some jurisdictions, several hundred cases have been litigated in over one-half of the United States since 1996. All convictions in which mtDNA has played a role have been upheld at the appellate level (*see also* ***Frye v. United States***). For more

information on court cases and appellate decisions, see http://www.denverda.org/DNA/Mitochondrial_DNA_Legal_Decisions.htm.

A complete tutorial on mtDNA is available online at http://dna.gov/training/otc/ through the President's DNA Initiative. This course was specifically developed for officers of the court.

References

[1] Holland, M.M. & Parsons, T.J. (1999). Mitochondrial DNA sequence analysis – validation and use for forensic casework, *Forensic Science Reviews* **11**, 21–50.

[2] Davis, C.L. (1998). Mitochondrial DNA: State of Tennessee v. Paul Ware, *Profiles in DNA* **1**, 6–7.

[3] Budowle, B., Adams, D.E., Comey, C.C. & Merrill, C.R. (1990). Mitochondrial DNA: a possible genetic material suitable for forensic analysis, in *Advances in Forensic Sciences*, H.C. Lee & R.E. Gaensslen, eds, Year Book Medical Publishers, Chicago, IL, pp. 76–97.

[4] Wilson, M.R., Polanskey, D., Butler, J., DiZinno, J.A., Replogle, J. & Budowle, B. (1995). Extraction, PCR amplification and sequencing of mitochondrial DNA from human hair shafts, *Biotechniques* **18**, 662–669.

[5] Taanman, J.-W. (1999). The mitochondrial genome: structure, transcription, translation, and replication, *Biochimica et Biophysica Acta* **1410**, 103–123.

[6] Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. & Wilson, A.C. (1991). African populations and the evolution of human mitochondrial DNA, *Science*, **253**, 1503–1507.

[7] Budowle, B., Wilson, M.R., DiZinno, J.A., Stauffer, C., Fasano, M.A., Holland, M.M. & Monson, K.L. (1999). Mitochondrial DNA regions HVI and HVII population data, *Forensic Science International* **103**, 23–35.

[8] Melton, T. & Nelson, K. (2005). Forensic mitochondrial DNA analysis of 691 casework hairs, *Journal of Forensic Sciences* **50**, 73–80.

[9] Nelson, K. & Melton, T. (2007). Forensic mitochondrial DNA analysis of 116 casework skeletal samples, *Journal of Forensic Sciences* **52**, 557–561.

[10] Gabriel, M.N., Huffine, E.F., Ryan, J.H., Holland, M.M. & Parsons, T.J. (2001). Improved mtDNA sequence analysis of forensic remains using a "mini-primer set" amplification strategy, *Journal of Forensic Sciences* **46**, 247–253.

[11] Carracedo, A., Bär, W., Lincoln, P., Mayr, W., Morling, N., Olaisen, B., Schneider, P., Budowle, B., Brinkmann, B., Gill, P., Holland, M., Tully, G. & Wilson, M. (2000). DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing, *Forensic Science International* **110**, 79–85.

[12] Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M. & Howell, N. (1999). Reanalysis and revision of the Cambridge Reference Sequence for human mitochondrial DNA. *Nature Genetics* **23**, 147.

[13] Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C. Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. & Young, I.G. (1981). Sequence and organization of the human mitochondrial genome, *Nature* **290**, 457–465.

[14] Melton, T. (2004). Mitochondrial DNA heteroplasmy, *Forensic Science Reviews* **16**, 1–20.

[15] Budowle, B., Allard, M.W., Fisher, C.L., Isenberg, A.R., Monson, K.L., Stewart, J.E.B., Wilson, M.R. & Miller, K.W.P. (2000). HVI and HVII mitochondrial DNA data in Apaches and Navajos, *International Journal of Legal Medicine* **116**, 212–215.

[16] Parson, W., Brandstätter, A., Alonso, A., Brandt, N., Brinkmann, B., Carracedo, A., Corach, D., Froment, O., Furac, I., Grzybowski, T., Hedberg, K., Keyser-Tracqui, C., Kupiec, T., Lutz-Bonengel, S., Mevag, B., Ploski, R., Schmitter, H., Schneider, P., Syndercombe-Comb, D., Sorensen, E., Thew, H., Tully, G. & Scheithauer, R. (2004). The EDNAP mitochondrial DNA population database (EMPOP) collaborative exercises: organisation, results, and perspectives, *Forensic Science International* **139**, 215–226.

[17] Gill, P., Ivanov, P.L., Kimpton, C., Piercy, R., Benson, N., Tully, G. Evett, I., Hagelberg, E. & Sullivan, K. (1994). Identification of the remains of the Romanov family by DNA analysis, *Nature Genetics* **6**, 130–135.

[18] Daoudi, Y., Morgan, M., Diefenbach, C., Ryan, J., Johnson, T., Conklin, G., Duncan, K., Smigielski, K., Huffine, E., Rankin, D., Mann, R., Holland, C., McElfresh, K., Canik, J., Armbrustmacher, V. & Holland, M. (1998). Identification of the Vietnam Tomb of the Unknown Soldier: the many roles of mitochondrial DNA, in *Proceedings of the Ninth International Symposium on Human Identification*. Promega Corporation, Scottsdale, AZ.

[19] Owsley, D.W., Ellwood, B.B. & Melton, T. (2006). Search for the grave of William Preston Longley, hanged Texas gunfighter, *Historical Archaeology* **40**, 50–63.

[20] Wallace, D.C., Brown, M.D. & Lott, M.T. (1999). Mitochondrial DNA variation in human evolution and disease, *Gene* **238**, 211–230.

[21] Stoneking, M. (1990). Mitochondrial DNA variation and human evolution, in *Human Genome Evolution*, M. Jackson, T. Strachan & G. Dover, eds, BIOS Scientific Publishers, Oxford, pp. 263–281.

[22] Krings, M., Stone, A., Schmitz, R.W., Krainitzki, H., Stoneking, M. & Pääbo, S. (1997). Neandertal DNA sequences and the origin of modern humans, *Cell* **90**, 1–12.

TERRY MELTON

Mixture Interpretation: DNA

Introduction

A mixed DNA profile is obtained when two or more contributors provided biological material to an evidentiary sample that has been analyzed. Here, only routine STR analysis techniques will be considered. With a STR profile, the presence of a mixture becomes obvious when more than two alleles are detected at a given locus.

DNA mixture interpretation is a complex and contentious field. There are a number of unsettled issues in the interpretation of mixtures and current methods, while adequate, leave clear room for improvement. In this section, we will outline current methods for interpretation and also highlight those areas where further research would benefit our science.

The primary division in modern interpretation methods lies between those who seek to

1. extract from the mixed DNA profile, when feasible, a single (unmixed) profile of interest and report it using the statistical techniques used for single stain calculation (*see* **Short Tandem Repeats: Interpretation**);
2. report an exclusion probability (often termed *random man not excluded* (RMNE) or *combined power of exclusion* (CPE)); and
3. report a likelihood ratio.

Much discussion has occurred debating these points, with views often very strongly held. This is unusual as the pros and cons of each method are well known and little new information has been added to the debate in recent years.

The peak heights present in mixtures are assumed to be approximately proportional to the amount of DNA from each contributor, and there is considerable empirical evidence for this. However, the correlation is only approximate and may be affected by degradation, stutter, and stochastic effects. The larger of two contributors is termed the *major contributor* and is assumed to be the source of the largest peaks. The smaller contributor is termed the *minor*. In a three-person mixture, the contributors may be present in quantities described by the terms *major, minor*, and *trace*.

It is also possible that two contributors are present in approximately equal proportion. In such cases, the mixture may be termed a 1 : 1 mixture.

Treating the Mixture as a Single Stain

If the profile in question is a clear major minor and the major is the profile of evidential interest, then it is acceptable to interpret the mixture as if it was an unmixed stain as long as care is taken regarding a potential factor of 2. Consider the situation where the stain may be explained as mixture of the suspect as the major and an unknown minor then under the defense hypothesis two unknowns are required to explain the mixture. In such cases, the unmixed stain calculation will err in favor of the prosecution by a factor of 2. This is known as the *Whitaker effect* [1–4].

Random Man Not Excluded

A frequentist method in common use is often termed *Random Man not Excluded*. In its simplest implementation it gives the probability that a random man has both his alleles contained within the mixture. Hence, for a mixture that shows the alleles abc and d the genotypes aa, ab, ac, ad, bb, bc, bd, cc, cd, and dd are not excluded.

Formally, if the mixture has alleles $A_1 \ldots A_n$ then the exclusion probability at locus l, (PE_l) is $PE_l = 1 - \left(\sum_{i=1}^{n} p(A_i)\right)^2$ if Hardy–Weinberg equilibrium is assumed. By writing $\sum_{i=1}^{n} p(A_i) = p$, we can obtain $PE_l = 1 - p^2$. If Hardy–Weinberg equilibrium is not assumed, Budowle gives $PE_l = 1 - \left(\sum_{i=1}^{n} p(A_i)\right)^2 - \theta \sum_{i=1}^{n} p(A_i)\left(1 - \sum_{i=1}^{n} p(A_i)\right)$. We can write this as $PE_l = 1 - p^2 - \theta p(1-p)$. The use of the equivalent of NRC II recommendation 4.1 [5] leads to $PE_l = 1 - p^2 - \theta \sum_{i=1}^{n} p_i(1-p_i)$, which differs slightly. The PE across multiple loci (PE) is calculated as $PE = 1 - \prod_l (1 - PE_l)$.

The advantages of such an approach are that

1. it does not assume a number of contributors;
2. it can report a statistic for mixtures that are lower quality than reportable under an LR approach; and
3. it is easier to explain.

Extensions to this approach essentially use judgment or a set of rules to generate a list of possible

unknown genotypes. The assumptions used to generate this list may include assuming a number of contributors and a conditioning profile. Once the list of possible genotypes is generated, the combined probability of inclusion (CPI, the complement of CPE) is generated by summing the estimated probabilities of the included genotypes. RMNE is assigned as 1-CPI. Typically, the genotype probabilities are estimated using the product rule but this is not a requirement.

There is a debate at the adequacy of using RMNE as a meaningful statistics (see EFS entry by Balding). If we agree that a balanced and fair assessment of the evidence is only achieved through the use of a likelihood ratio, hence by addressing the evidence both under H_p and H_d, then RMNE statistics does not represent an appropriate expression of the weight of evidence. This is because (i) the numerator is not assessed and (ii) the proposition for the denominator implies a loose concept of a random man and the evidence is assessed without taking into account of the profile of the suspect at hand.

Likelihood Ratio-Based Approaches

The first stage in the development of a likelihood ratio for mixtures is the establishment of the two hypotheses. Even using sublevel 1 hypotheses (see EFS entry by this can be problematic and requires the use of judgment and an awareness of the case circumstances [6–11]). Issues include, but are not restricted to

- assigning a number of contributors and
- determining whether any contributors can be safely assumed to be present under both the prosecution and defense hypotheses. Contributors that are assumed to be present are termed *conditioning* contributors.

One of the more common two person mixtures is one that can be explained as containing the DNA of the suspect and the complainant. In such a case, the hypotheses may be

H_p: the DNA in the crime stain is from the suspect and the complainant and

H_d: the DNA in the crime stain is from an unknown person and the complainant.

In this case, the conditioning contributor is the complainant. The unknown person is termed *the unknown*.

If E is used to represent the DNA results obtained in the case, then the LR is $p(\mathrm{E}|H_p)$ divided by $p(\mathrm{E}|H_d)$.

Assigning the Number of Contributors

Assignment of the number of contributors is undertaken from the number of alleles observed with some attention paid to peak balances. The number of alleles per locus is an important parameter but it should not be the only criterion used [12, 13]. It is important to have an understanding of stuttering and of genetic phenomena such as trisomy, gene duplication, and somatic mutation [4]. Mixtures where an evidential profile is sufficiently low in peak height that dropout may have occurred should be treated as low-template profiles. Hence most of the discussion here will assume that dropout is not an issue (for more information on such artifacts, see EFS entry on low template DNA).

Typically, the number of contributors is assigned as the minimum number required to explain the mixed profile. Hence a profile containing at most two alleles per locus is assigned as containing one contributor unless peak balances suggest that a second contributor is present. A profile containing at most four alleles per locus is assigned as a two-person mixture. Profiles containing five or more alleles are termed *higher order mixtures*.

Interpretation of higher order mixtures can be problematic but is simplified if conditioning contributors may be assumed and if one profile is clearly a major contributor. If it cannot be simplified, it is advised to refrain from undertaking any calculation.

Once the number of contributors is assigned, the hypotheses generated and the presence of any conditioning contributors determined that the next step is to list the possible genotypes for any unknown contributors. This may be achieved either considering peak height or not.

Considering Only Qualitative Data

If no consideration is to be made of peak height, then the methods of Weir *et al.*, Fung and Hu, and Curran *et al.* return the likelihood ratio either utilizing the product rule (Weir *et al.*), NRC II recommendation

4.1 (Fung and Hu) or recommendation 4.2 (Curran *et al.*) [14–20].

Although LR computation varies in function of the propositions involved, this discussion can be pursued on the basis of a case with a two-person mixture that can be explained as containing a complainant and a suspect (H_p) or as the complainant and an unknown person (H_d). The arguments can be extended without loss of generality to other pairs of propositions. In this case, the LR is always in the form of one divided by the sum of the random match probabilities for the collection of all the possible genotypes for the unknown person. The summation under the denominator assessment implies an equally weighted contribution for each potential genotype. This is justified by the fact that the model explicitly does not take into account peak areas.

The Curran *et al.* method is embodied in the software DNAMIX II and is suitable for casework use. It does not take account of peak heights and cannot handle dropout. Since it does not utilize all the available information it has the expected performance of understating the evidence if H_p is true and overstating it if H_d is true. Recall that this is the expected performance of any method that does not utilize all the available evidence. Certain rules based on peak height may be applied to identify those situations where the approach may be nonconservative and appropriate action may then be taken. This is especially true when the peak areas would suggest that a combination of the designated contributors is not supported if the H_p is true. By default, the above methods assign a probability of 1 to the numerator of the likelihood ratio and concentrate on the estimation of the denominator. That default value may not always be supported by the data.

Beecham and Weir (in draft) have developed a method to add a sampling uncertainty consideration to the method of Curran *et al.* [14] This approach is embodied the software DNAMIX III (available at http://statgen.ncsu.edu/~gwbeecha/).

The work by Fung and Hu [19] led also to the development of dedicated software, also in case involving relatives [21] (available at http://www.hku.hk/statistics/EasyDNA/).

Considering Peak Heights

If peak heights are to be considered then the list of possible unknown genotypes is modified by eliminating those combinations that are expected to give a poor fit to the peak heights. Consider the situation where we have a conditioning profile, say the complainant, who has genotype ab. The peak heights $a = 1000, b = 2000, c = 1000$ are very unlikely if the unknown contributor is genotype ac. Hence the genotype ac is eliminated as a possibility for the unknown. This elimination process may proceed by the judgment of an experience caseworker or according to a set of rules. One set of such rules has been published [22] and is embodied in the software suite Fss-I^3. Three types of rules are typically used on the basis of (i) expected peak imbalance, (ii) expected mixing proportion between the two contributors, and (iii) expected allelic dropout. Currently, appropriate treatments have been given to two-person mixtures only. In this case, the LR is on the form of one divided by the sum of the random match probabilities for the collection of all the genotypes for the unknown person that have passed the rules.

The situation where a minor evidential profile is present in an assumed two-person mixture and dropout is possible can be handled by careful application of judgment or the rules. Typically, possibilities for the unknown are expanded to include one or two unobserved alleles. The possibility of two unobserved alleles returns a likelihood ratio of 1 for this locus and is not always conservative (see the recent case against Garside and Bates [23]). The possibility of one unobserved allele is typically handled using the $2p$ rule [24] or its subpopulation corrected variants. However, any profile where dropout is possible is really low-template DNA (LTDNA) and the LTDNA style of interpretation should be applied [25].

Once the set of possible unknown genotypes has been determined the likelihood ratio may be developed using either the product rule or one of the subpopulation correction options (see EFS entry by Balding). The approaches outlined above are commonly referred to as *binary*, because the decision on accepting or rejecting a given combination amount to assign a weight of 1 or 0 in the likelihood ratio calculation. It is not a full probabilistic approach as early advocated by Evett, Gill and Lambert [26] and continued by Perlin and Szabady [27], but it constitutes a promising step toward it. However, the use of a binary model should be undertaken having in mind a clear understanding of its limits. As mentioned for the techniques not involving peak areas, the analyst needs to make sure that the genotype of

the suspect is the best supported option (taking into account peak areas) under the numerator proposition and that all reasonable genotypes are included in the denominator [28, 29].

Standard

There has been a substantial move amongst the European Network of Forensic Science Institutes toward setting standards for interpretation of mixtures, see [28, 30].

Issues in Mixture Interpretation

Currently, there are a number of areas in DNA mixture interpretation, which are under active research. This includes methods to handle the uncertainty in the number of contributors, dealing with dropout, automation of the process, and a more effective use of the peak height information. Of particular interest is the research involving Bayesian networks that may bring intuitive ways to handle the complexity of mixture interpretation [31].

References

[1] Evett, I.W. (1987). On meaningful questions: a two-trace transfer problem, *Journal of the Forensic Science Society* **27**, 375–381.

[2] Triggs, C.M. & Buckleton, J. (2003). The two trace transfer problem revisited, *Science and Justice* **43**(3), 127–134.

[3] Meester, R. & Sjerps, M. (2003). The evidential value in the DNA database search controversy and the two-stain problem, *Biometrics* **59**(3), 727–732.

[4] Buckleton, J.S., Triggs, C.M. & Walsh, S.J. (2004). *DNA Evidence*, CRC Press, Boca Raton.

[5] NRC_II (1996). *National Research Council Committee on DNA Forensic Science, The Evaluation of Forensic DNA Evidence*, National Academy Press, Washington, DC.

[6] Cook, R., Evett, I.W., Jackson, G., Jones, P.J. & Lambert, J.A. (1998). A hierarchy of propositions: deciding which level to address in casework, *Science and Justice* **38**(4), 231–240.

[7] Cook, R., Evett, I.W., Jackson, G., Jones, P.J. & Lambert, J.A. (1999). Case pre-assessment and review in a two-way transfer case, *Science and Justice* **39**(2), 103–111.

[8] Cook, R., Evett, I.W., Jackson, G., Jones, P.P. & Lambert, J.A. (1998). A model for case assessment and interpretation, *Science and Justice* **38**(3), 151–156.

[9] Cook, R., Evett, I.W., Jackson, G. & Rogers, M. (1993). A workshop approach to improving the understanding of the significance of fibres evidence, *Journal of the Forensic Science Society* **33**(3), 149–152.

[10] Evett, I.W., Gill, P.D., Jackson, G., Whitaker, J. & Champod, C. (2002). Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks, *Journal of Forensic Sciences* **47**(3), 520–530.

[11] Evett, I.W., Jackson, G. & Lambert, J.A. (2000). More on the hierarchy of propositions: exploring the distinction between explanations and propositions, *Science & Justice* **40**(1), 3–10.

[12] Buckleton, J.S., Curran, J.M. & Gill, P. (2007). Towards understanding the effect of uncertainty in the number of contributors to DNA stains, *Forensic Science International Genetics* **1**(1), 20–28.

[13] Paoletti, D.R., Doom, T.E., Krane, C.M., Raymer, M.L. & Krane, D.E. (2005). Empirical analysis of the STR profiles resulting from conceptual mixtures, *Journal of Forensic Sciences* **50**, 1361–1366.

[14] Curran, J.M., Triggs, C.M., Buckleton, J.S. & Weir, B.S. (1999). Interpreting DNA mixtures in structured populations, *Journal of Forensic Sciences* **44**(5), 987–995.

[15] Weir, B.S., Triggs, C.M., Starling, L., Stowell, L.I., Walsh, K.A.J. & Buckleton, J.S. (1997). Interpreting DNA mixtures, *Journal of Forensic Sciences* **42**(2), 213–222.

[16] Evett, I.W. & Weir, B.S. (1998). *Interpreting DNA Evidence – Statistical Genetics for Forensic Scientists*, Sinauer Associates, Sunderland.

[17] Evett, I.W., Buffery, C., Willott, G. & Stoney, D.A. (1991). A guide to interpreting single locus profiles of DNA mixtures in forensic cases, *Journal of the Forensic Science Society* **31**(1), 41–47.

[18] Fung, W.K. & Hu, Y.Q. (2000). Interpreting DNA mixtures based on the NRC-II recommendation 4.1, *Forensic Science Communications* **2**(4), http://www.fbi.gov/programs/lab/fsc/backissu/oct2000/fung.html.

[19] Fung, W.K. & Hu, Y.Q. (2001). The evaluation of mixed stains from different ethnic origins: general result and common cases, *International Journal of Legal Medicine* **115**, 48–53.

[20] Fung, W.K. & Hu, Y.Q. (2002). The statistical evaluation of DNA mixtures with contributors from different ethnic groups, *International Journal of Legal Medicine* **116**, 79–86.

[21] Hu, Y.Q. & Fung, W.K. (2003). Evaluating forensic DNA mixtures with contributors of different structured ethnic origin: a computer software, *International Journal of Legal Medicine* **117**(4), 248–249.

[22] Bill, M., Gill, P., Curran, J., Clayton, T., Pinchin, R., Healy, M. & Buckleton, J. (2005). PENDULUM – a guideline based approach to the interpretation of STR mixtures, *Forensic Science International* **148**, 181–189.

[23] R. v Garside and Bates, (2006). *EWCA Crim 1395*, Royal Courts of Justice, London.

[24] Buckleton, J. & Triggs, C.M. (2006). Is the 2p rule always conservative? *Forensic Science International* **159**, 206–209.
[25] Gill, P., Whitaker, J.P., Flaxman, C., Brown, N. & Buckleton, J.S. (2000). An investigation of the rigor of interpretation rules for STR's derived from less that 100 pg of DNA, *Forensic Science International* **112**(1), 17–40.
[26] Evett, I.W., Gill, P.D. & Lambert, J.A. (1998). Taking account of peak areas when interpreting mixed DNA profiles, *Journal of Forensic Sciences* **43**(1), 62–69.
[27] Perlin, M.W. & Szabady, B. (2001). Linear mixture analysis: a mathematical approach to resolving mixed DNA samples, *Journal of Forensic Sciences* **46**(6), 1372–1377.
[28] Gill, P., Brown, R.M., Fairley, M., Lee, L., Smyth, M., Simpson, N., Irwin, B., Dunlop, J., Greenhalgh, M., Way, K., Westacott, E.J., Ferguson, S.J., Ford, L.V., Clayton, T. & Guiness, J. (2008). National recommendations of the technical UK DNA working group on mixture interpretation for the NDNAD and for court going purposes, *Forensic Science International: Genetics* **2**(1), 76–82.
[29] Clayton, T.M. & Buckleton, J.S. (2004). Mixtures, in *Forensic DNA Evidence Interpretation*, CRC Press, Boca Raton, pp. 217–274.
[30] Gill, P., Brenner, C.H., Buckleton, J.S., Carracedo, A., Krawczak, M., Mayr, W.R., Morling, N., Prinz, M., Schneider, P.M. & Weir, B.S. (2006). DNA commission of the international society of forensic genetics: recommendations on the interpretation of mixtures, *Forensic Science International* **160**, 90–101.
[31] Cowell, R.G., Lauritzen, S.L. & Mortera, J. (2008). Probabilistic modelling for DNA mixture analysis, *Forensic Science International: Genetics Supplement Series* **1**(1), 640–642.

JOHN S. BUCKLETON

M'Naghten Standard of Insanity ***see*** Insanity: Defense

M'Naghten's Rule ***see*** Behavioral Science Evidence

Modeling: Fire ***see*** Fire Modeling and Its Application in Fire Investigation

Modeling: Trauma Causation in Vehicle Accidents ***see*** Trauma Causation: Analysis of Automotive

Molestation ***see*** Child Sexual Abuse Accommodation

Morphogenesis of Fingerprints ***see*** Friction Ridge Skin: Morphogenesis and Overview

MtDNA ***see*** Mitochondrial DNA: Profiling

Multiple Personalities ***see*** Dissociative Disorders



Narcoanalysis ***see*** Deception: Truth Serum

Narcotics: Amphetamine ***see*** Amphetamine

Narcotics: Benzodiazepines ***see*** Benzodiazepines

Narcotics: Cannabis ***see*** Cannabis

Narcotics: Cocaine ***see*** Cocaine

Narcotics: Opiods ***see*** Opioids

Natural Causes of Sudden Death: Noncardiac

Respiratory

Pneumonia

Prior to the discovery of antibiotics, one-third of all people who developed pneumonia subsequently died from the infection. Until the introduction of antibiotics, pneumonia was the most common cause of death in the United States and, perhaps, worldwide. Since the advent of antibiotics, the situation has changed radically.

More often than not, pneumonia-related deaths do not occur quickly. Respiratory and/or systemic symptoms and signs are often present before death occurs. Occasionally, the forensic pathologist may be confronted with cases of sudden unexpected death (SUD), which are the sequel of a fulminant course of a previously undiagnosed pneumonia.

Children represent a special population; in a series of 265 autopsies of children with SUD, the main cause was disease of the respiratory system (total 113 cases, 42.6%, included lobular pneumonia (28.3%), aspiration pneumonia (24.8%), and viral pneumonia (16.8%). The high predominance of respiratory deaths is probably explained by the relative immaturity of children's lungs, and the fact that the body's resistance is reduced [1].

Many cases of SUD in infancy can be adequately explained after performing a careful postmortem

examination (*see* **Autopsy**). Vennemann *et al.* [2] reported that, in Germany, 2910 infants died in 2004, and of these 394 babies died suddenly and unexpectedly. The authors of that report investigated a three-year population-based, case–control study, in Germany (1998–2001). A total of 455 deaths were reviewed and 51 (11.2%) were found to be unexplained. Most of these deaths were due to respiratory or generalized sepsis, and, of course, a number of different viruses and bacteria have been suggested as causative agents of sudden infant death syndrome (SIDS), though this connection has never been conclusively established.

Bacterial Pneumonia

Acute bacterial infection of the lungs is still one of the commonest causes of death, especially in the very young and very old, but often its presence is due to the existence of some other secondary to some other debilitating process. For example, bacterial pneumonia is a major cause of morbidity and mortality among transplant recipients.

The onset of bacterial pneumonia varies from sudden to gradual. The patient experiences shaking chills, high fever, sweating, shortness of breath, chest pain, and a cough that produces thick, greenish or yellow phlegm. Acute respiratory insufficiency and/or severe sepsis may cause death in these subjects [3]. And, while there are case reports describing rapid symptom progression, SUD from pneumonia is rare [4, 5].

Some evidence for respiratory tract infection is common among infants dying suddenly. For example, Lin reported that in his autopsy series of sudden infant death, 28.3% of the decedents had lobular pneumonia [1]. Epidemiologic evidence indicates that SIDS is associated with *Bordetella pertussis* infection. Nicoll and Gardner [6] examined postperinatal infant deaths resulting from respiratory causes and SIDS. All these cases occurred in England and Wales between 1968 and 1984; they estimated an excess mortality rate for undiagnosed pertussis of 460–700 deaths. Similarly, Cherry calculated 362 excess infant deaths caused by pertussis [7]. Data from Sweden and Norway indicate a direct correlation between the incidence of pertussis and the occurrence of SIDS [8]. In another study, Heininger *et al.* [9] enrolled 254 infants with SUDs; an autopsy diagnosis of SIDS was made in 76% of the infants. In the remaining subjects, causes of death were respiratory or other infections (14%), congenital anomalies or organ failures (4%), aspiration (2%), or accidents or traumatic events (4%). During a standardized autopsy, nasopharyngeal specimens and tracheal specimens were obtained for polymerase chain reaction (PCR) assays to detect B pertussis. PCR results were positive for B pertussis for 12 case subjects (5.1%) (all with SIDS or respiratory infections).

Generally, bronchopneumonia and pneumonia are characterized by widespread patchy areas of inflammation that begin as a widely dispersed bronchitis and bronchiolitis; focal areas of pneumonia then develop in the centers of the acini. The consolidated areas are generally larger and more numerous in the lower lobes where they may be several millimeters across. Small beads of yellow mucous can often be expressed from the bronchioles on the cut surface of the lung. In severe cases, patches of consolidation may become confluent. Once the bacteria reach the alveoli they elicit an acute inflammation, with copious exudation of fluid and migration of neutrophils into the alveoli [10]. In lobar pneumonia (as in the case of pneumococcal infection), the changes are uniform throughout the affected lobe.

Viral Pneumonia

Half of all cases of pneumonia are believed to be caused by viruses. More viruses are being identified as the cause of respiratory infection, and though most attack the upper respiratory tract, some produce pneumonia, especially in children. Most cases of viral pneumonia are mild and get better without treatment, but some cases are more serious and require hospitalization. People at risk for more serious viral pneumonia typically have impaired immune systems such as people with HIV, transplant patients, young children (especially those with heart defects), the elderly, and those that immunocompromised for whatever reason.

Viral pneumonia is caused by several different viruses, including influenza, parainfluenza, adenovirus, rhinovirus, herpes simplex virus, respiratory syncytial virus, hantavirus, and cytomegalovirus. The initial symptoms of viral pneumonia are the same as influenza symptoms: fever, dry cough, headache, muscle pain, and weakness. Increasing breathlessness may also occur; fever may be present. Viral pneumonia may be complicated by an invasion of bacteria,

with all the typical symptoms of bacterial pneumonia. Sometimes the clinical course is quite uneventful and viral infection only recognized with the occurrence of SUD.

Cases of fatal varicella pneumonia have been described in adult patients [11–13]. Fatal influenza infection can be caused by influenza pneumonia alone, by respiratory complications caused by bacterial superinfection [14], or by extrapulmonary influenza-associated manifestations such as myocarditis or encephalopathy. Complication rates of influenza infection, as well as for all the other virus infections, are highest in immunosuppressed patients, diabetics or individuals with other severe metabolic illnesses, and people older than 65 years. Case reports of fatal, sudden death due to influenza pneumonia have been described in the literature [14, 15].

Many viral infections may result in life-threatening complications in healthy children and many cases of a SUD of children due to viral pneumonia have been reported [16–20].

It can become difficult to unequivocally diagnose the underlying disease, especially if incipient stages and/or general infections are involved. Interstitial pneumonia, as caused by viruses, is defined histologically, but in its early stages the changes can be nonspecific; detection of the virus would therefore be required to verify the diagnosis [21, 22]. In viral pneumonia, the lungs appear bulky, and may be hyperemic (overfilled with blood). Bloodstained, frothy fluid oozes freely from the cut surface. Areas of hemorrhage are present and may be extensive. The mucosa of the bronchial tree is very hyperemic. The histological pattern is characterized by interstitial pneumonitis, with extensive alveolar collapse, and filling of remaining alveoli with fluid and desquamated epithelial cells. Alveolar epithelial damage is common in viral pneumonia; it causes the formation of hyaline membranes [10].

Mycoplasma Pneumonia

The vast spectrum of diseases caused by *Mycoplasma pneumoniae* includes pharyngitis, sinusitis, tracheobronchitis, pneumonia, myocarditis, pancreatitis, hepatitis, arthritis, and meningoencephalitis. Such infections are generally benign and often run a subclinical course; it is estimated that less than 5% of the cases of mycoplasma pneumonia are severe enough to require admission to a hospital. A very small number of fatal sudden mycoplasma pneumonia cases have been reported [23, 24].

Mycoplasma pneumonia generally produces widespread bronchiolitis and interstitial pneumonia similar to that caused by many respiratory viruses. The bronchiolitis sometimes progresses to epithelial ulceration. Lymphocytic infiltration of the walls of alveolar ducts and alveoli characterizes the interstitial pneumonia while edema fluid, red blood cells, and macrophages are found in many groups of alveoli and some alveoli may contain hyaline membranes [10].

Acute Interstitial Pneumonitis

A particular form of interstitial pneumonia is acute interstitial pneumonitis (AIP), a fulminant disease culminating in acute respiratory failure and often death. First described by Hamman and Rich [25], AIP is a life-threatening respiratory disease of unknown cause that occurs in patients of both genders equally; generally, it occurs in previously healthy people with no significant medical history. Patients report a prodromal (before main symptoms) illness lasting several days with fever, nonproductive cough, malaise, followed by the acute onset of progressive shortness of breath, which rapidly evolves to respiratory failure. Other signs and symptoms that may be present include cyanosis, crackles, wheezing, hemoptysis, diaphoresis, and hypotension [26, 27]. The radiological and histological patterns are those of acute respiratory distress syndrome (ARDS); however, AIP differs from ARDS because, in the former, there is an absence of known inciting events and multiorgan failure does not occur. While mortality in all patients with ARDS has decreased over the years, there have been no analogous reports of an improvement in survival rates for patients with AIP; in every reported series, the hospital mortality has been >50% [28].

We have previously reported a fatal case of AIP in a 15-year-old boy with sudden death [29]. The postmortem diagnosis was made on the basis of the convergence of clinical onset and course, anatomical, pathological, and radiological findings, as well as the exclusion of each known cause of ARDS, in particular infections, confirmed by negative results of microbiological analysis, and of immunohistochemical and laboratory tests for the detection of more common respiratory virus.

The exact pathogenesis of AIP remains a mystery; thus, the diagnosis of AIP remains a diagnosis of exclusion. The differential diagnosis includes infectious pneumonia, ARDS, drug-induced lung disease, acute pancreatitis, congestive heart failure (CHF), connective-tissue disease, and other acute forms of interstitial lung disease.

The histological features of AIP, as seen in open lung biopsies or autopsies, are those of diffuse alveolar damage (DAD), a nonspecific pattern of acute lung injury deriving from numerous causes, which shows both epithelial and endothelial injury. Histological lung examination generally shows the typical findings of DAD: alveolar septa mildly thickened by edema and capillary congestion, alveolar edema, hyaline membranes lining the denuded alveolar walls, hyperplastic Type 2 pneumocytes, alveolar infiltrates of polymorphonuclear neutrophilic leukocytes, pigmented macrophages, monocytes and plasma cells, and fibrin thrombi in small arteries.

Thromboembolism

Thromboembolism encompasses two interrelated conditions that are part of the same spectrum, deep venous/vein thrombosis (DVT) and pulmonary embolism (PE). Obstruction of blood flow to one or more arteries of the lung occurs when a thrombus lodged in a pulmonary vessel breaks free and travels to the lung. Almost all clinically important PEs are the result of a DVT occurring in the deep veins of the lower extremities, proximal to and including the popliteal veins. However, emboli also can originate from the pelvic veins, the inferior vena cava, and even the upper extremities. PE and thromboembolic disease represent major health problems worldwide. The diagnosis is often difficult to establish and is frequently missed. De Bakey [30] published a review after critical evaluation of more than 375 000 postmortem cases and more than 3 000 000 clinical cases of PE that had been reported around the world over the preceding half a century. He concluded that there was great confusion over true incidence of PE. Fifty years later, the true incidence of thromboembolic disease is still not known [31].

Risk factors for venous thromboembolism include patients age, surgery, trauma, hospital or nursing home confinement, active malignant neoplasm with, or without, concurrent chemotherapy, central vein catheterization or transvenous pacemaker, prior superficial vein thrombosis, varicose veins, and neurological disease with extremity paresis. Compared to residents in the community, hospitalized residents have over a 150-fold increased incidence of acute venous thromboembolism. Those hospitalized for medical illness and those hospitalized for surgery account for cases of thromboembolism in almost equal proportions. The incidence in nursing home residences independently accounts for over one-tenth of all venous thromboembolism diseases in the community. The risk among surgery patients can be further stratified on the basis of patients age, type of surgery, and the presence of active cancer. The incidence of postoperative venous thromboembolism is increased for surgery patients who are 65 years of age or older. High-risk surgical procedures include neurosurgery, major orthopedic surgery of the leg, thoracic, abdominal or pelvic surgery for malignancy, renal transplantation, and cardiovascular surgery. Active cancer accounts for almost 20% of all cases of venous thromboembolism occurring in the community. The risk appears to be higher for patients with pancreatic cancer, lymphoma, malignant brain tumors, cancer of the liver, leukemia, and colorectal and other digestive cancers. Prior superficial vein thrombosis is an independent risk factor for subsequent DVT or PE, remote from the episode of superficial thrombophlebitis. Long-haul (>6 h) air travel is associated with a slightly increased risk for venous thromboembolism. Among women, additional risk factors for venous thromboembolism include oral contraceptive use and hormone therapy, pregnancy, and the postpartum period. Other conditions associated with venous thromboembolism include heparin-induced thrombocytopenia, myeloproliferative disorders (especially polycythemia rubra vera and essential thrombocythemia), intravascular coagulation and fibrinolysis/disseminated intravascular coagulation (ICF/DIC), nephrotic syndrome, paroxysmal nocturnal hemoglobinuria, thromboangiitis obliterans (Buerger's disease), thrombotic thrombocytopenic purpura, Bechet's syndrome, systemic lupus erythematosis, inflammatory bowel disease, homocystinuria, and possibly hyperhomocysteinemia [32]. Heart failure is another well-known risk factor for thromboembolism.

The spectrum of disease produced by thromboembolism ranges from clinically unsuspected or clinically unimportant to massive embolism causing death. Untreated acute proximal DVT accounts

for clinical PE in 33–50% of patients. Mortality in untreated PE is approximately 30%: about one-third of PE cases are fatal. Data indicate that more than one half of all cases of thromboembolism remain undiagnosed. The clinical diagnosis of pulmonary thromboembolism is notoriously inaccurate, with many cases either wrongly diagnosed (overdiagnosed) or missed (underdiagnosed). Autopsy is still regarded as the diagnostic gold standard (*see* **Autopsy**). The accuracy of antemortem diagnosis of PE is within the range of just 10–30% [31]. DVT and PE are very often undiagnosed in life, especially in children dying SUDs.

The prevalence of PE at autopsy varies widely in different published autopsy series, from not less than 10% in unselected material [33]; in 20% of all hospitalized patients [34]; in 20–60% of adult autopsies [35]; in more than half of all autopsies [36]; in about 10% of adults who die suddenly in hospital [37]; approximately 3.5% of all hospital deaths [38]; 1% in the general population of hospital patients; and 30% in patients dying after severe burns, trauma, or fractures [39]. The incidence of PE at autopsy is strongly influenced by the nature of the population surveyed (age and nature of patients, i.e., surgical, oncological, gerontological), by the mode of selection for postmortem examination and, in particular, by the care with which the autopsy is performed [31].

At autopsy, gross examination will reveal embolus in the main pulmonary vascular tree; segmental or subsegmental pulmonary arteries are affected in most of the cases and are more prevalent in nonfatal pulmonary thromboembolism (PTE). On the other hand, emboli in the main arteries, as well as in arteries of the trunk and above, are more prevalent in fatal PTE. Pulmonary infarctions that appear as dark areas where the air space has become filled with blood are frequently seen. They are commonly multiple.

Thrombosis of the deep leg veins is often detected at gross examination as the source of embolism. However, the absence of thrombus from these veins does not exclude them as the site of thrombosis since dislodgement and embolization may leave no residual disease.

Amniotic Fluid Embolism

Amniotic fluid embolism (AFE) is a rare but potentially fatal complication of pregnancy with an incidence approximately between 1 in 8000 and 1 in 80 000 pregnancies. The first reported case of AFE was documented in 1926 [40]. It was first recognized as a syndrome in 1941, when two investigators described the presence of fetal mucin and squamous cells during postmortem examination of the pulmonary vasculature in women who had unexplained obstetric deaths [41]. Since then, many studies, case reports, and series have been published in an attempt to elucidate the etiology, risk factors, and pathogenesis of this obstetric complication [42].

The true incidence of AFE is not known. There are many reasons for this. Firstly, there are a number of inaccuracies in reporting the cause of maternal death. Secondly, the presentation can be variable and the number of nonfatal or subclinical episodes is unclear. Finally, in those patients who do survive or in whom no autopsy was performed, it is often quite difficult to confirm the diagnosis [43]. AFE is one of the leading causes of death during labor and the first few postpartum hours, and it remains a deadly, unpreventable, and unpredictable obstetric emergency. Recent data suggest mortality rates approach 61%. Most patients do not survive the initial course and die within five days. Of those patients who survive, neurologic impairment is common [44]. Other authors refer an associated mortality rate up to 90% [45]. The fetal mortality rate, although better than the maternal rate, is a dismal 21%. Fifty percent of the surviving neonates experience permanent neurological injury. Predisposing factors traditionally associated with AFE include placental abruption, uterine overdistention, fetal death, trauma, tumultuous or oxytocin-stimulated labor, multiparity, advanced maternal age, and rupture of membranes. However, in numerous demonstrated cases of AFE, none of these conditions or demographic characteristics were present [46].

The syndrome appears to be initiated after maternal intravascular exposure to fetal material that is contained in the amniotic fluid. Many hypotheses have been proposed to explain the pathophysiology of this condition. Clark *et al.* [47] contend that AFE more closely resembles an anaphylactic reaction to fetal debris than an embolic event, and they propose the term *anaphylactoid syndrome of pregnancy*. Recently, bradykinin, released in association with DIC, has been considered as an important contributor to the severe hypotension that is manifested with AFE [48, 49]. The literature also contains evidence that AFE may not be a true embolic event resulting from

the physical obstruction of the pulmonary vasculature. The high degree of variability in symptoms, the lack of characteristic findings on radiological exam, the absence of a dose–response effect on symptoms, and the occasional occurrence of coagulopathies are not entirely consistent with a physical block to the circulation, at least not as the main mechanism of disease. An alternative hypothesis is that fetal antigen leaking into the maternal circulation might activate the complement cascade. This rare immune response may be initiated by an obscure pathological antigen, or by common antigens presented uncommonly – in amount, timing, or frequency of entry into the maternal circulation [50, 51].

Presenting symptoms may vary; common clinical features include shortness of breath, altered mental status followed by sudden cardiovascular collapse, DIC, and maternal death.

The entry criteria in the National Registry for AFE [47] include

1. acute hypotension or cardiac arrest;
2. acute hypoxia, defined as dyspnea, cyanosis, or respiratory arrest;
3. coagulopathy, defined as laboratory evidence of intravascular consumption or fibrinolysis or severe clinical hemorrhage in the absence of other explanations;
4. onset during dilatation and evacuation, labor, cesarean delivery, or within 30 min postpartum;
5. absence of any other significant confounding condition or potential explanation for the signs and symptoms observed.

Although labor and delivery appear to be the greatest risk period [52], AFE has been reported in the second trimester and as having a delayed presentation [53], up to 36 h postpartum [54].

The diagnosis of AFE is very difficult even at autopsy (*see* **Autopsy**; **Histology**). It is a commonly held misconception that the presence of fetal debris in the pulmonary circulation is diagnostic of an amniotic fluid embolus. In fact, fetal debris can be found in the pulmonary circulation in a predominance of patients who underwent a normal labor, and AFE is only identified in 78% of those patients who meet the criteria for the diagnosis of AFE [43, 47, 55].

Some authors have suggested that AFE may be the result of anaphylactic reactions to fetal antigens, and that the major part of this clinical syndrome is the result of mast cell degranulation followed by the release of histamine, tryptase, and other mediators. It has been suggested that the identification and distribution of mast cell tryptase within the pulmonary tree is a useful criteria for the diagnosis of fatal AFE. In a previous report, Fineschi *et al.* [56] demonstrated a numerical increase of pulmonary mast cells in the subjects who died of AFE compared with that of the control groups. Many other reports add further supporting evidence to the concept of AFE as an anaphylactoid syndrome of pregnancy [57–59].

Macroscopic autopsy findings in AFE are usually insignificant; neither are macroscopic findings generally found either. Postmortem diagnosis is based on histological demonstration of amniotic fluid components (fetal squames, lanugo, and mucin) in pulmonary blood vessels that can be demonstrated by various staining methods (Alcian blue, Attwood). Evaluation of pulmonary mast cell tryptase using immunohistochemistry as well as measurement of serum tryptase level as markers of mast cell degranulation are recommended in all cases of suspected AFE-related death.

Morbid Obesity

Obesity and overweight are defined as an accumulation of excess body fat, to an extent that may impair health. A crude population measure of excess fat is the body mass index (BMI), a person's weight (in kilograms) divided by the square of his or her height (in meters). WHO defines overweight as a BMI of 25 or more, and obesity as a BMI of 30 or greater. Obesity is a complex pathological condition that affects virtually all age and socioeconomic groups and threatens to overwhelm both developed and developing countries. WHO's latest projections indicate that globally, in 2005, approximately 1.6 billion adults (age 15+) were overweight, and that at least 400 million adults were obese. WHO further projects that by 2015, approximately 2.3 billion adults will be overweight and more than 700 million will be obese. At least 20 million children under the age of 5 years were overweight globally in 2005. Contrary to conventional wisdom, the obesity epidemic is not restricted to industrialized societies; in developing countries, it is estimated that over 115 million people suffer from obesity-related problems.

Overweight and obesity lead to serious health consequences. Risk increases progressively as BMI increases. In fact, there is evidence that, on a population level, the risk of chronic disease increases progressively as average BMI increases above 21. Obesity is associated with numerous comorbidities such as cardiovascular disease (CVD), Type 2 diabetes, hypertension, certain cancers, and sleep apnea [60]. A relationship exists between BMI and all-cause mortality also in adolescence [61]. In conclusion, the mortality associated with untreated morbid obesity is significant, manyfold that of the normal population and exceeds the risk of surgical intervention [62].

Even in the absence of comorbidity, when adipose tissue accumulates in excess, a variety of adaptations/alterations in cardiac structure and function occur, as do changes in metabolism [63], Obesity may affect the heart through its influence on known risk factors such as dyslipidemia, hypertension, glucose intolerance, inflammatory markers, obstructive sleep apnea/hypoventilation, and the prothrombotic state, as well as through yet-unrecognized mechanisms. As a whole, overweight/obesity predisposes, or is at least associated with, numerous cardiac complications such as coronary heart disease (CHD), heart failure, and sudden death through its impact on the cardiovascular system (*see* **Cardiac and Natural Causes of Sudden Death**).

Obese subjects have an increased risk of arrhythmias and sudden death, even in the absence of cardiac dysfunction, and the risk of sudden cardiac death with increasing weight is seen in both genders. In the Framingham study, the annual sudden cardiac mortality rate in obese men and women was estimated to be ≈40 times higher than the rate of unexplained cardiac arrest in a matched nonobese population.

Moreover, a prolonged QT_c interval was observed in ≈30% of subjects with impaired glucose tolerance, and there is a positive association between BMI and QT_c. When visceral obesity or insulin levels increase, sympathovagal balance may be the best explanation for changes in QT_c. The clinical significance of obesity-associated QT prolongation and the mechanisms involved remain speculative. Moreover, because extremely obese patients often have a dilated cardiomyopathy, fatal arrhythmias may be the most frequent cause of death [61]. In conclusion, subjects with morbid obesity have high rates of sudden, unexpected cardiac deaths and the literature is rich in reports of cases of SUD in morbidly obese subjects [64–66].

Hematological Disorders

Hemoglobinopathies

Hemoglobinopathy is a type of genetic defect that results in abnormal structure of one of the globin chains of the hemoglobin molecule. The most common hemoglobinopathy is sickle-cell disease (SCD). In the United States today, 1 of every 650 African-Americans (0.15%), greater than 50 000 individuals, suffers from sickle-cell disease. In addition, almost 2 million patients, representing 8% of the African-American population, are affected by sickle-cell trait being heterozygous for the sickle-cell gene [67].

The most common causes of death in the sickle-cell population include infection/sepsis, acute chest syndrome (ACS) (clinical term defined by new pulmonary infiltrate on chest radiography, accompanied by fever, chest pain, and variety of respiratory symptoms, including wheezing, coughing, and tachypnea), sudden cardiac death, cerebrovascular accident, and renal failure. Among all reported causes of death in sickle-cell disease, age-specific patterns are well documented. In sickle-cell infants and children, sepsis, predominantly pneumococcal, represents a leading cause of death, followed by acute splenic sequestration. In the adult sickle-cell population, mortality is much more commonly associated with ACS, sickle-cell pain crises, and/or cardiac failure [68–70].

More than 20% of sickle-cell patients develop fatal acute or chronic pulmonary complications. The acute pulmonary sequelae, including ACS, thromboembolism, lung edema, fat/bone marrow embolism, and vaso-occlusive crisis, account for a large proportion of sudden deaths among sickle-cell patients [71, 72]. In a recent report on 306 autopsies of patients with sickle-cell disease, the most common cause of death for all sickle variants and for all age groups was infection (33–48%). The terminal infection was heralded by upper respiratory tract syndromes in 72.6% and by gastroenteritis in 13.7%. The most frequent portal of entry in children was the respiratory tract but, in adults, it was a site of severe or chronic organ injury. Other causes of death included stroke 9.8%, therapy complications 7.0%, splenic sequestration 6.6%, pulmonary emboli/thrombi 4.9%, renal failure 4.1%, pulmonary hypertension (PHT) 2.9%,

hepatic failure 0.8%, massive hemolysis/red cell aplasia 0.4%, and left ventricular failure 0.4% [73].

Sudden and unexpected death in this population is not rare [74]; Manci *et al.* reported that death was frequently sudden and unexpected (40.8%) or occurred within 24 h after presentation (28.4%), and was usually associated with acute events (63.3%) [73]. Darbari *et al.* [75] reported that leading circumstances of death in 141 autopsies of adult SCD included PHT (26.2%), renal failure (22.6%), infection (18.4%), thromboembolism (14.9%), cardiac diagnoses (12.0%), cirrhosis (11.3%), pneumonia or ACS (9.9%), bleeding (7.8%), and iron overload (7.0%); in 23.4% of cases, death was sudden.

In sickle-cell disease, sudden death frequently occurs during exertion and is characterized by rhabdomyolysis, heat stroke, and cardiac arrhythmia; this entity has been widely reported in literature where many cases of sudden death associated with sickle-cell disease or sickle-cell trait have been described in affected soldiers during military basic training, in athletes during training, and in previous healthy subjects during exertion [76–83].

Military recruits, pilots, and subjects exposed to hypoxic stress such as high altitude, and experiencing sudden cardiorespiratory collapse as a result of sickle-cell trait have been described [84].

Many hypotheses have been offered to explain the high rate of sudden death in SCD. Involvement of autonomic nervous dysfunction in sudden death has been reported in various diseases and it has been suggested that this may be the case in sickle-cell anemia as well [85]. Heat stress, dehydration, viral illness, and poor physical conditioning have all been identified as factors, which may contribute to exertional rhabdomyolysis and sudden death in SCD, suggesting a multifactorial etiology [80]. Dehydration, hypoxia, acidosis, and physical exertion are known aggravating factors for sudden death in sickle-cell trait, because that can initiate a vaso-occlusive sickle-cell crisis. Others have proposed a role for some coexisting disease such as diabetes [86].

Of particular interest to the forensic pathologist, illicit drug use, especially cocaine and morphine, both of which are recognized triggers of global hypoxic events, may lead to the sickling of red blood cells, with consequent vaso-occlusion and lung edema has been and, in fact, such a sequence has been reported as explaining a case of sudden death [87].

Only few autopsy studies of sickle-cell patients have addressed patterns of pathologic pulmonary involvement that are very often recognizable in SCD [70, 88, 89]. When pulmonary involvement is present, it is usually in the form of PHT, first detected at autopsy with various grade of changes (from reversible to irreversible), consisting in mild-to-severe medial hypertrophy and muscularization of pulmonary arterioles. The lungs from many of these individuals manifest intimal fibroelastosis of small- and medium-sized pulmonary arteries. Small muscular arteries can show dilatation and may form plexiform lesions. Cardiomegaly may be also present. In Graham's study [67], almost two-third of the cases had significant pulmonary findings at autopsy (71.4%). The most frequent postmortem pulmonary findings were pulmonary edema (47.6%), followed by thromboembolism (38.1%), fat emboli (33.3%), PHT, grades I–IV (33.3%), and microvascular vaso-occlusive thrombi (28.5%). Oppenheimer and Esterly found autopsy evidence of pulmonary thromboemboli in 66%.

In conclusion, data from literature indicate a high rate of occurrence of pulmonary findings (fat embolism, PHT, and cardiac right ventricular hypertrophy) in a large percentage of SDC patients presenting with sudden death. Histological examination may reveal widespread vaso-occlusive sickling of red blood cells in the brain, heart, liver, kidneys, adrenal glands, thyroid gland, intramyocardial coronary arteries, skeletal muscles, pancreas, testis, and spleen.

Leukemia

Leukemia is a neoplastic disease that arises in blood-forming tissue such as the bone marrow and causes large numbers of blood cells to be produced and enter the bloodstream. The leukemias form a very heterogeneous group of neoplasms, which differ both in clinical course (acute or chronic), and blood cell line that affected. Depending on these factors, patients with leukemia may have a number of symptoms, such as fevers or night sweats, frequent infections, weakness or tiredness, headache, bleeding and bruising easily, pain in the bones or joints, swelling or discomfort in the abdomen (from an enlarged spleen), swollen lymph nodes, especially in the neck or armpit, and weight loss.

SUD secondary to undiagnosed neoplasia is extremely uncommon in adults, with a reported

incidence between 0.17% [90] and 0.58% [91]. Acute leukemia, as well as bronchogenic carcinoma, gastric adenocarcinoma, and adenocarcinoma of the urinary bladder, are the most common tumors causing SUD in adults [92]. Death has been attributed to a variety of mechanisms, including hemorrhage, thromboembolism, and widespread tumor dissemination.

Cases of SUDs due to splenic rupture in subjects with undiagnosed chronic leukemia have been described [93]: acute cardiovascular failure [94], intracardiac thrombosis [95], and pulmonary complications [96].

In acute leukemia, symptoms appear rapidly and progress quickly. As in chronic leukemia, sudden death is reported both in adults and children [97], sometimes occurring after a period of complete remission of the disease [98, 99]. Sudden unexpected death due to neoplastic disease in infancy and childhood (SUDNIC) is extremely uncommon. The most common causes of SUDNIC are tumors affecting vital structures, such as the heart and brainstem, and include primary cardiac tumors, primary CNS neoplasms, and fatal CNS hemorrhage secondary to hematopoietic neoplasms [100, 101]. Other less common causes of sudden death in children with leukemia may involve intracerebral infiltration or hyperviscosity and leukostasis [101]. Whybourne *et al.* [102] reported a case in which sudden death was the first manifestation of lymphoblastic leukemia in a 16-week-old boy; at autopsy, no evidence of intracranial hemorrhage was found. Infiltration of multiple organs (myocardium, lungs, etc.) was detected. The mechanism of mortem was retraced to the diffuse infiltration of myocardium probably predisposing to arrythmias, favored also by the diffuse infiltration of the lungs leading to a condition of hypoxia. Recently, *Somers et al.* [103] reviewed a 20-year autopsy cases ($n = 4926$); 8 cases of SUDNIC were identified in which postmortem diagnoses included 2 cases of acute leukemia (1 myelogenous, 1 lymphoblastic). Finally, it has recently been recognized that some form of treatment for leukemia may, themselves, cause SUD.

In conclusion, data from literature demonstrate that infants, children, and adults may have minimal or no symptoms in the presence of leukemic disease and highlight the need for a thorough autopsy examination in all cases of SUD. Arsenic tetroxide is now often used in the treatment of acute promyelocytic leukemia, and it appears to be an effective agent. However, its use is sometimes accompanied by sudden death because arsenic disrupts the normal function of potassium channels, leading to QT prolongation and torsades des pointes [104].

Metabolic Disorders

Diabetes

Diabetes mellitus (DM) is an etiologically and clinically heterogeneous group of metabolic disorders that share the common feature of hyperglycemia. Long-term hyperglycemia produces tissue damage, which ultimately manifests itself as microvascular disease, macrovascular disease, and neuropathy. Since 1991, when the British Diabetic Association initated a study of unexplained deaths in young patients with Type 1 DM [105], evidence has correlated the so-called dead-in-bed syndrome with diabetes. What characterized this kind of death is that it occurs at night and the patients are found dead in an undisturbed bed the next morning. Patients had gone to bed in apparent good health, and subsequent autopsy consistently revealed no cause of death (of course, the study was undertaken before it was possible to screen ion channel disease postmortem, and no doubt that some of the children included in the original study did suffer from some form of inherited long QT syndrome (LQTS) syndrome. See below). Nocturnal hypoglycemia is common among Type 1 diabetic patients; however, the precise mechanism by which hypoglycemia may cause sudden death remains uncertain. The suddenness of death in these patients implicated a possible cardiac cause. Many hypotheses have been proposed to explain the "dead-in-bed syndrome": undetected autonomic neuropathy, hypoglycemia-induced disturbances in cardiac electrophysiology, and other factors are probably involved in the sudden death of young diabetic patients.

However, the definite cause of the "dead-in-bed" syndrome continues to be a subject of further speculation and investigation (Figure 1) [106].

Cardiovascular autonomic neuropathy (CAN) encompasses damage to the autonomic nerve fibers that innervate the heart and blood vessels, resulting in abnormalities in heart rate control and vascular dynamics; it has been widely studied as a risk factor for sudden death in patients with DM [107, 108].

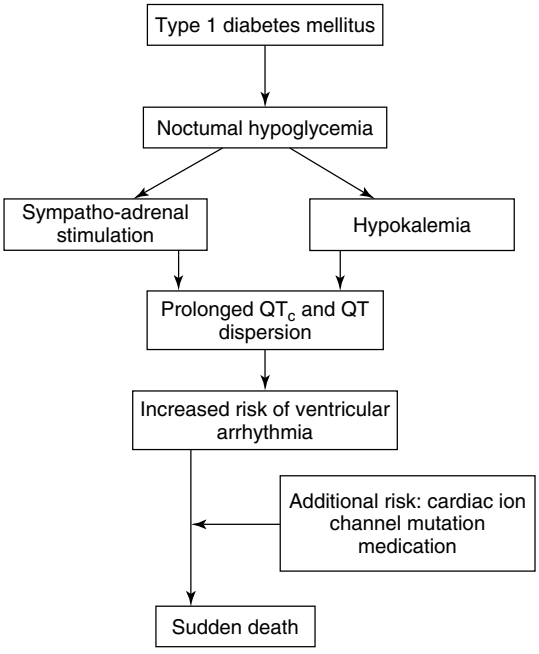


Figure 1 Hypothesis for "dead-in-bed syndrome" in Type 1 diabetes mellitus

One potential cause may be severe but asymptomatic ischemia, which can induce lethal arrhythmias. QT prolongation may also predispose individuals to life-threatening cardiac arrhythmias and sudden death. It has been shown that predisposition to arrhythmias and an association with mortality may also be related to intracardiac sympathetic imbalance [109, 110]. However, the significance of CAN as an independent cause of sudden death has been questioned recently. In the Rochester Diabetic Neuropathy Study, the investigators found that all cases of sudden death in individuals with and without diabetes had severe coronary artery disease (CAD) or left ventricular dysfunction. They suggested that although CAN could be a contributing factor, it was not a significant independent cause of sudden death. Heart failure is, however, common in individuals with diabetes; it is identified in these patients by the presence of neuropathy, even in those without evidence of CAD or left ventricular dysfunction [108]. The association of CAN in the absence of coronary disease and cardiomyopathy requires further study [111].

It is well known that DM is an independent risk for CVD: the association between diabetes and incidence of CAD is widely reported in the literature. In the United States, diabetes is referred to as the most prevalent factor placing patients at risk for coronary events [112]. Furthermore, patients with DM have increased risk for ventricular arrhythmia that is thought to be secondary to CAD or CHF. Susceptibility to dysrhythmias in diabetic hearts is a very debatable issue. Diabetes is an independent risk for cardiac arrythmias in addition to other CVDs [113, 114]. QT prolongation is associated with SUD in the diabetic populations [115]. In a large study, Mohaved *et al.* [116] founded that patients with DM have significantly higher prevalence of ventricular fibrillation independent of CAD or CHF, which, in part, may explain the higher risk of sudden death in these patients.

Furthermore, patients with DM have a hypercoagulable state that is of multifactorial origin. Many coagulation factors, such as fibrinogen, d-dimer, and von Willebrand factor may become elevated in diabetics. At the same time, there is decreased fibrinolysis along with platelet hyperaggrebility and endothelial dysfunction. Tsai *et al.* [117] founded DM as an independent risk for venous thromboembolism in a large population including both genders. Mohaved *et al.* [118] founded that DM is strongly associated with PE and PHT independent of CAD, CHF, or smoking.

In cases where diabetes is related to the occurrence of sudden death, gross necropsy examination may reveal signs of coronary plaques and coronary stenoses. The kidneys may show smaller size and granular surface. Histologically, the earliest detectable change in glomerular structure is the thickening of the basement membrane. Postmortem biochemistry of vitreous humor is a worthwhile adjunct to routine postmortem screening. Apart from the determination of glucose level in vitreous fluid, the determination of hemoglobin A1c considered as a definitive indicator of prolonged hyperglycemia has been proposed as a useful tool in postmortem diagnosis of diabetes [119].

Alcoholic Ketoacidosis

The entity of alcoholic ketoacidosis (AKA) was first described by Dillon *et al.* [120]. AKA affects chronic alcoholics and is characterized by metabolic acidosis with increased anionic gap. A typical patient with AKA has a history of chronic alcohol abuse

(*see* **Alcohol: Use, Abuse, Tolerance, and Dependency**; **Alcohol: Interaction with Other Drugs**; **Alcohol: Behavioral and Medical Effects**) and recent binge drinking, followed by the abrupt cessation of alcohol consumption. Clinical findings of AKA are very similar to those of diabetic ketoacidosis, but hyperglycemia and glycosuria are generally absent. AKA is due to the accumulation of D-β-hydroxybutyrate and acetoacetic acid. The accumulation is probably the result of various factors such as volume depletion (vomiting, decreased fluid intake) and starvation having a lipolytic effect. As the previous alcoholic intake has an inhibitory effect on fatty acid oxidation, a higher level of fatty acids will be available. When fatty acids are presented to the liver faster than they can be oxidized, it results in a surplus of acetyl–CoA. The excess of acetyl–CoA is converted into β-hydroxy-β-methyl-glutaryl-CoA, which is cleaved to form acetyl–CoA and free acetoacetic acid. Some of the acetoacetic acid is reduced to D-β-hydroxybutyrate and a small amount is decarboxylated to acetone [121].

In uncomplicated AKA, the prognosis may be good, but several complications could be present (lactic acidosis, acute pancreatitis, Wernike's encephalopathy, etc.). AKA may be a life-threatening condition and it is a significant factor in at most a small minority of alcoholic deaths [122]. Fatalities of chronic alcoholics where the cause of death could not be determined by thorough autopsy, histology and toxicology including determination of alcohol concentration are described [123]; in those cases, ketoacidosis and lactic acidosis were assumed to be the cause of death. In such cases of unexpected and unexplained deaths in chronic alcoholics, measurement of ketone bodies in vitreous humor or pericardial fluid using clinical laboratory methodologies is recommended. In particular, β-hydroxybutyric acid (β-HBA) should be considered as the diagnostic marker of choice for the postmortem determination of AKA and as the cause of death [124].

Hypoglycemia

Since the introduction of insulin into diabetes care in 1921, hypoglycemia (HoG) emerged as a very complex concern. Some medications (all insulins, all insulin secretagogs) can cause HoG. There is particular concern about the drug metformin: most experts suggest that metformin monotherapy does not cause clinically relevant HoG under usual circumstances, but it might under unusual circumstances. Such circumstances include combined therapy with metformin and insulin or insulin secretagogs (I/IS).

Many conditions make patients more sensitive to I/IS and predispose to HoG: leanness and youth, physical activity, first trimester of pregnancy and the immediate postpartum period, impaired liver function and alcoholism (depleted glycogen stores), impaired renal function (prolonged half-life insulin), counterregulatory hormone failure (adrenals, glucagon, and growth hormone), autonomous neuropathy, previous hypoglycemia, being unaware of hypoglycemia, total parental nutrition, and poor metabolic control [125].

The scientific literature quotes very different values for the incidence of mortality due to HoG, which ranges in insulin-treated diabetic patients (high-risk populations) from 2 to 4% [125].

The exact mechanism by which hypoglycemia causes sudden death remains uncertain. One explanation of HoG-related sudden death is that hypoglycemia directly causes disturbances in cardiac electrophysiology, which may provoke malignant tachydysrhythmias. There have been reports of premature ventricular contractions, atrial arrhythmias, and ischemic ECG changes during hypoglycemia [126–128]. Moreover, there is evidence that insulin-induced hypoglycemia causes an acquired form of long QT syndrome. In addition to QT_c prolongation during hypoglycemia, QT dispersion (the difference between the longest and shortest QT interval on a 12-lead ECG) increases [129, 130].

In conclusion, the sudden death of a diabetic patient treated with I/IS creates a unique diagnostic problem. A high level of suspicion based on circumstantial evidence must alert the forensic pathologists on hypoglycemia as a probable cause of death (Figure 2) [125].

Postmortem diagnosis of HoG as a cause of sudden death can be proved by testing glucose levels in blood or spinal or vitreous fluid. The measurement of the ketone body β-hydroxybutyrate (β-HBA) may also help in the investigation of cases where hypoglycemia is suspected [131].

Thyroiditis

In forensic medicine, the study of thyroid function and disease is important in cases of sudden death [132–135].

Consideration	Lower risk, less probability	Higher risk, greater probability
Age	Older	Young
Duration of diabetes	Short	Long
Control as judged by the levels of CBG, hemoglobin AIc, and fructosamine	Suboptimal	Optimal
Adjusts to late meals and activity	Yes	No
History of frequent episodes of hypoglycemia	No	Yes
History of being unaware of hypoglycemia	No	Yes
Keeps a CBG diary	Yes	No
Keeps follow-up appointments	Yes	No
Frequency of CBG testing	Frequent	Rare or never

Figure 2 Considerations for assessing risk of hypoglycemia or, if applicable, estimating the probability that hypoglycemia caused idiopathic or accidental sudden death

Cases of sudden death associated with undiagnosed chronic thyroiditis are described [136–138]. An unusual case of fatal heatstroke in a young woman discovered unconscious in a sauna has been reported in which a preexisting Hashimotos thyroiditis was revealed at autopsy [139]. Moreover, an association between lymphocytic myocarditis and lymphocytic thyroiditis in a case of sudden death involving a 40-year-old man with no known medical history has also been reported [140].

Reports in the literature also raise the possibility of thyroid disease, in particular silent (painless) thyroiditis, may be an underlying cause of SUDs. Routine microscopy of the thyroid gland is therefore advocated in cases of SUD. It has also been suggested that postmortem measures of thyroid hormones (thyroxine [T4] and triiodothyronine [T3]) and the pituitary hormone known as *thyroid-stimulating hormone* (*TSH*) may be useful diagnostic tools in cases of SUDs [141].

Anaphylaxis

Anaphylaxis was first described in the scientific literature about 100 years ago by Portier and Richet, who reported that their attempts to immunize dogs against the sting of jellyfish with actinia extract instead brought about an acute anaphylactic episode. In the extreme or classic form, anaphylaxis typically involves the cutaneous, respiratory, cardiovascular, and gastrointestinal systems, target organs all heavily populated with mast cells. Its presentation is often

more enigmatic, with variable target organ involvement and expression of symptoms.

In 1998, a Joint Task Force on Practice Parameters defined anaphylaxis as an "immediate systemic reaction caused by rapid, IgE-mediated immune release of potent mediators from tissue mast cells and peripheral basophils". Aggregation of Fc*e*RI by allergen-driven cross-linking of receptor-bound IgE activates mast cells and basophils to release mediators that induce the pathophysiologic features of the anaphylactic response. Initial sensitization occurs through a highly coordinated series of steps involving a variety of cell types and mediators, which is affected by environmental exposure and complex genetic factors. Anaphylactic reactions are distinguished from anaphylactoid reactions, which "mimic signs and symptoms of anaphylaxis, but are caused by non-IgE-mediated release of potent mediators from mast cells and basophils" [142].

The most common etiologies of anaphylactic reactions include allergic responses to food, drug (medications, biologics, and vaccines), insect sting, and latex. Also exercise (jogging, walking, tennis, dancing, etc.) can lead to typical anaphylaxis. The pathogenesis and true incidence of exercise-induced anaphylaxis remain unknown. Finally, another form of anaphylaxis consists of the so-called idiopathic anaphylaxis, whose diagnosis is one of exclusion. The exact incidence of idiopathic anaphylaxis is unknown, but several studies estimate that nearly 20% of cases of anaphylaxis are idiopathic. There are no clinically distinguishing features (although 33% of cases are nocturnal), and it may be fatal.

Anaphylaxis is an uncommon cause of sudden death. In many cases, no specific macroscopic or microscopic findings are detected at autopsy. The most common finding is nonspecific pulmonary congestion and edema, but features suggesting an allergic reaction (cutaneous erythema, urticaria, laryngeal and/or pharyngeal edema) may be present. The presence of petechial hemorrhages is another nonspecific postmortem finding as is brain swelling. Generally, histological examinations will disclose mucosal edema, inflammation with eosinophilia, epithelial sloughing in the bronchial tree, cerebral edema, and polyvisceral stasis [143, 144].

In the presence of a clinical history suggestive for anaphylaxis-related death, postmortem measurement of serum tryptase levels can be a useful diagnostic aid. Tryptase, a neutral protease, is the major protein component of mast cell secretory granules. An elevated serum tryptase level is considered to be a specific marker for systemic mast cell activation, which is a central feature of anaphylaxis. In addition to this, because this enzyme has a longer serum half-life than other chemical mediators like histamine, the measurement of serum tryptase level is reported to be useful for postmortem as well as antemortem diagnosis of anaphylaxis [145–147]. Finally, in SIDS, one possible explanation hypothesized some 40 years ago is an anaphylactic reaction to protein components such as cow's milk [148, 149]. It was suggested that sensitized infants might suffer a fatal anaphylactic reaction if recently ingested cow's milk was regurgitated and inhaled during sleep. The question is still under debate [150]. New evidence is provided for an increased degree of mast cell activation in infants whose death was classified as SIDS. The stimuli for mast cell degranulation in these infants remains unclear, but the potential for anaphylaxis to have occurred in a proportion of cases of SIDS must be considered [151–154].

Infection

Massive adrenal hemorrhage (Waterhouse–Friderichsen syndrome) is an uncommon, but usually fatal, consequence of overwhelming sepsis [155]. Despite the predominant association with meningococcal infection, there are numerous other well-recognized etiologies, including sepsis resulting from other organisms, and noninfectious causes, such as anticoagulant treatment, antiphospholipid syndrome, trauma, and spontaneously occurring or postoperative adrenal hemorrhage. Gram-negative organisms, including klebsiella, pasturella, and *Hemophilus influenzae*, have all been reported to precipitate the syndrome. There are fewer reports of Waterhouse–Friderichsen syndrome following infection with Gram-positive bacteria. The clinical picture is generally characterized by sudden onset shock, pyrexia, cyanosis, dyspnea, and purpura, with adrenal hemorrhage. Unexpected death can occur as sequel of a rapidly progressive course of a previously undiagnosed infection [156–159].

At necropsy skin rashes will be evident, as will signs of upper airway infection, bronchopneumonia, and evidence of meningitis. Hemorrhage, elsewhere, may coexist with a massive adrenal hemorrhage,

which is a more characteristic finding. Microscopic examinations generally reveal hemorrhages involving all layers of the adrenal glands with secondary cell necrosis. Measurement of serum procalcitonin concentration (PCT) may be a useful aid [160]. In fact, it is well known that in bacteremia and sepsis, PCT levels are highly elevated ($>10\,\mathrm{ng\ ml^{-1}}$ in sepsis), and therefore PCT is a well-established clinical parameter in the diagnosis of systemic infection of bacterial origin [161].

References

[1] Lin, S. (2006). Analysis of 265 autopsies of sudden death in children, *Journal of Clinical Forensic Medicine* **13**(6–8), 293–295.

[2] Vennemann, M., Bajanowski, T., Butterfass-Bahloul, T., Sauerland C., Jorch, G., Brinkmann, B. & Mitchell, E.A. (2007). Do risk factors differ between explained sudden unexpected death in infancy and sudden infant death syndrome?, *Archives of Disease in Childhood* **92**(2), 133–136.

[3] Kellum, J.A., Kong, L., Fink, M.P., Weissfeld, L.A., Yealy, D.M., Pinsky, M.R., Fine, J., Krichevsky, A., Delude, R.L. & Angus, D.C. (2007). Understanding the inflammatory cytokine response in pneumonia and sepsis, results of the Genetic and Inflammatory Markers of Sepsis (GenIMS) Study, *Archives of internal medicine* **167**(15), 1655–1663.

[4] Escoffery, C.T. & Shirley, S.E. (2002). Causes of sudden natural death in Jamaica, a medicolegal (coroner's) autopsy study from the University Hospital of the West Indies, *Forensic Science International* **129**(2), 116–121.

[5] Batalis, N.I., Caplan, M.J. & Schandl, C.A. (2007). Acute deaths in nonpregnant adults due to invasive streptococcal infections, *The American Journal of Forensic Medicine and Pathology* **28**(1), 63–68.

[6] Nicoll, A. & Gardner, A. (1988). Whooping cough and unrecognised postperinatal mortality, *Archives of Disease in Childhood* **63**, 41–47.

[7] Cherry, J.D. (1984). The epidemiology of pertussis and pertussis immunization in the United Kingdom and the United States: a comparative study, *Current Problems in Pediatrics* **14**, 1–78.

[8] Lindgren, C., Milerad, J. & Lagercrantz, H. (1997). Sudden infant death and prevalence of whooping cough in the Swedish and Norwegian communities, *European Journal of Pediatrics* **156**, 405–409.

[9] Heininger, U., Kleemann, W.J. & Cherry, J.D. (2004). A controlled study of the relationship between *Bordetella pertussis* infections and sudden unexpected eaths among German infants, *Pediatrics* **114**(1), e9–e15.

[10] Corrin, B. (2000). *Pathology of the Lungs*, Churchill Livingstone Edition.

[11] Burton, G.G., Sayer, W.J. & Lillington, G.A. (1966). Varicella pneumonitis in adults: frequency of sudden death, *Diseases of the Chest* **50**(2), 179–185.

[12] Gregorakos, L., Myrianthefs, P., Markou, N., Chroni, D. & Sakagianni, E. (2002). Severity of illness and outcome in adult patients with primary varicella pneumonia, *Respiration; International Review of Thoracic Diseases* **69**(4), 330–334.

[13] Popara, M., Pendle, S., Sacks, L., Smego Jr, R.A., Mer, M. (2002). Varicella pneumonia in patients with HIV/AIDS, *International Journal of Infectious Diseases* **6**(1), 6–8.

[14] Tsokos, M., Zöllner, B. & Feucht, H.H. (2005). Fatal influenza A infection with *Staphylococcus aureus* superinfection in a 49-year-old woman presenting as sudden death, *International Journal of Legal Medicine* **119**, 40–43.

[15] Drescher, J., Zink, P., Verhagen, W., Flik, J. & Milbradt, H. (1987). Recent influenza virus A infections in forensic cases of sudden unexplained death, *Archives of Virology* **92**(1–2), 63–76.

[16] An, S.F., Gould, S., Keeling, J.W. & Fleming, K.A. (1993). Role of respiratory viral infection in SIDS: detection of viral nucleic acid by *in situ* hybridization, *Journal of Pathology* **171**(4), 271–278.

[17] Bajanowski, T., Wiegand, P., Cecchi, R., Pring-Akerblom, P., Adrian, T., Jorch, G. & Brinkmann, B. (1996). Detection and significance of adenoviruses in cases of sudden infant death, *Virchows Archieve* **428**(2), 113–118.

[18] Hoang, M.P., Ross, K.F., Dawson, D.B., Scheuermann, R.H. & Rogers, B.B. (1999). Human herpesvirus-6 and sudden death in infancy: report of a case and review of the literature, *Journal of Forensic Sciences* **44**(2), 432–437.

[19] Pfeiffer, H., Varchmin-Schultheiss, K. & Brinkmann, B. (2006). Sudden death in childhood due to varicella pneumonia: a forensic case report with clinical implications, *International Journal of Legal Medicine* **120**, 33–35.

[20] Dettmeyer, R., Sperhake, J.P., Müller, J. & Madea, B. (2008). Cytomegalovirus-induced pneumonia and myocarditis in three cases of suspected sudden infant death syndrome (SIDS): diagnosis by immunohistochemical techniques and molecularpathologic methods, *Forensic Science International* **174**(2–3), 229–233.

[21] Bajanowski, B., Rolf, G., Jorch, B. & Brinkmann, B. (2003). Detection of RNA viruses in sudden infant death (SID), *International Journal of Legal Medicine* **117**, 237–240.

[22] Fernandez-Rodriguez, A., Ballesteros, S., de Ory, F., Echevarria, J.E., Alvarez-Lafuente, R., Vallejo, G. & Gomez, J. (2006). Virological analysis in the

diagnosis of sudden 156 children death:, a medico-legal approach, *Forensic Science International* **161**, 8–14.

[23] Chan, E.D. & Welsh, C.H. (1995). Fulminant mycoplasma pneumoniae pneumonia, *The Western Journal of Medicine* **162**, 133–142.

[24] Daxboeck, F., Eisl, B., Burghuber, C., Memarsadeghi, M., Assadian, O. & Stanek, G. (2007). Fatal mycoplasma pneumoniae pneumonia in a previously healthy 18-year-old girl, *Wiener Klinische Wochenschrift* **119**(11–12), 379–384.

[25] Hamman, L. & Rich, A. (1935). Fulminating diffuse interstitial fibrosis of the lungs, *Transactions of the American Clinical and Climatological Association* **51**, 154–163.

[26] Vourlekis, J.S., Brown, K.K. & Schwarz, M.I. (2001). Acute interstitial pneumonitis: current understanding regarding diagnosis, pathogenesis, and natural history, *Seminars in Respiratory and Critical Care Medicine* **22**(4), 399–408.

[27] Vourlekis, J.S. (2004). Acute interstitial pneumonia, *Clinics in Chest Medicine* **25**(4), 739–747.

[28] Quefatieh, A., Stone, C.H., Di Giovine, B., Toews, G.B. & Hyzy, R.C. (2003). Low hospital mortality in patients with acute interstitial pneumonia, *Chest* **124**, 554–559.

[29] Turillazzi, E., Di Donato, S., Neri, M., Riezzo, I. & Fineschi, V. (2007). An immunohistochemical study in a fatal case of acute interstitial pneumonitis (Hamman–Rich syndrome) in a 15-year-old boy presenting as sudden death, *Forensic Science International* **173**, 73–77.

[30] De Bakey, M.E. (1954). A critical evaluation of the problem of thromboembolism: collective review, *Surgery, Gynecology and Obstetrics* **98**, 1–27.

[31] Steiner, I. (2007). Pulmonary embolism – temporal changes, *Cardiovascular Pathology* **16**, 248–251.

[32] Heit, J.A. (2006). The epidemiology of venous thromboembolism in the community: implications for prevention and management, *Journal of Thrombosis and Thrombolysis* **21**(1), 23–29.

[33] Woolf, N. (1998). *Pathology Basic and Systemic*. London, WB Saunders.

[34] Riede, U.N., Schaefer, H.E. & Wehner, H. (1998). *Allgemeine und Spezielle Pathologie*, 2nd Edition, Stuttgart, Georg Thieme Verlag.

[35] Kissane, J.M. (1985). *Anderson's Pathology*, 8th Edition, St. Louis, Mosby.

[36] Rubin, E. & Farber, J.L. (1994). *Pathology*, 2nd Edition, Philadelphia, Lippincott.

[37] Damjanov, I. (2002). *Pathology Secrets*. Philadelphia, Hanley & Belfus.

[38] Braunstein, H. (1987). *Pathology*, 2nd Edition, St. Louis, CV Mosby.

[39] Kumar, V., Abbas, A.K. & Fausto, N. (2005). *Robbins and Cotran Pathologic Basis of Disease*, 7th Edition, Elsevier Saunders.

[40] Meyer, J.R. (1926). Embolus pulmonar-caseosa, *Brazilian Journal of Medical and Biological Research* **2**, 301–303.

[41] Steiner, P.E. & Luschbaugh, C.C. (1941). Maternal pulmonary embolism by amniotic fluid, *Journal of the American Medical Association* **117**, 1245–1254.

[42] Stafford, I. & Sheffield, J. (2007). Amniotic fluid embolism, *Obstetrics and Gynecology Clinics of North America* **34**(3), 545–553.

[43] Davies, S. (2001). Amniotic fluid embolus: a review of the literature, *Canadian Journal of Anaesthesia* **48**, 88–98.

[44] Martin, R.W., Amniotic fluid embolism, *Clinical Obsetetrics and Gynecology* (1996). **39**(1) 101–106.

[45] Cunningham, G.F., Gant, N.F., Leveno K.J., Gelstrap L.C., Hawth J.C. & Wenstrom K.D. (eds). (2001). Obstetrical hemorrhage. in *Williams's Obstetrics*, New York, McGraw-Hill.

[46] Perozzi, K.J. & Englert, N.C. (2004). Amniotic fluid embolism. An obstetric emergency, *Critical Care Nurse* **24**, 56–61.

[47] Clark, S.L., Hankins, G.D.V., Dudley, D.A., Dildy, G.A. & Porter, T.F. (1995). Amniotic fluid embolism: analysis of the national registry, *American Journal of Obstetrics and Gynecology* **172**, 1158–1169.

[48] Lurie, S., Feinstein, M. & Mamet, Y. (2000). Disseminated intravascular coagulopathy in pregnancy: thorough comprehension of etiology and management reduces obstetricians' stress, *Archives of Gynecology and Obstetrics* **263**, 126–130.

[49] Robillard, J., Gauvin, F., Molinaro, G., Leduc, L., Adam, A. & Rivard, E. (2005). The syndrome of amniotic fluid embolism: a potential contribution of bradykinin, *American Journal of Obstetrics and Gynecology* **193**, 1508–1512.

[50] Benson, M.D., Kobayashi, H., Silver, R.K., Oi, H., Greenberger, P.A. & Terao, T. (2001). Immunologic studies in presumed amniotic fluid embolism *Obstetrics and Gynecology* **97**(4), 510–514.

[51] Benson, M.D. (2007). A hypothesis regarding complement activation and amniotic fluid embolism, *Medical Hypotheses* **68**(5), 1019–1025.

[52] Fletcher, S.J. & Parr, M.J.A. (2000). Amniotic fluid embolism: a case report and review, *Resuscitation* **43**, 141–146.

[53] Malhotra, P., Agarwal, R., Awasthi, A., Das, A. & Behera, D. (2007). Delayed presentation of amniotic fluid embolism: lessons from a case diagnosed at autopsy, *Respirology (Carlton, Vic)* **12**(1), 148–150.

[54] Devriendt, J., Machayekhi, S. & Staroukine, M. (1995). Amniotic fluid embolism: another case with non-cardiogenic pulmonary oedema, *Intensive Care Medicine* **21**, 698–699.

[55] Martin, S.R. & Foley, M.R. (2006). Intensive care in obstetrics: an evidence-based review, *American Journal of Obstetrics and Gynecology* **195**, 673–689.

[56] Fineschi, V., Gambassi, R., Gherardi, M. & Turillazzi, E. (1998). The diagnosis of amniotic fluid embolism: an immunohistochemical study for the quantification of pulmonary mast cell tryptase, *International Journal of Legal Medicine* **111**(5), 238–243.

[57] Farrar, S.C. & Gherman, R.B. (2001). Serum tryptase analysis in a woman with amniotic fluid embolism. A case report, *The Journal of Reproductive Medicine* **46**(10), 926–928.

[58] Nishio, H., Matsui, K., Miyazaki, T., Tamura, A., Iwata, M. & Suzuki K. (2002). A fatal case of amniotic fluid embolism with elevation of serum mast cell tryptase, *Forensic Science International* **126**, 53–56.

[59] Rainio, J. & Penttila, A. (2003). Amniotic fluid embolism as cause of death, in a car accident – a case report, *Forensic Science International* **137**, 231–234.

[60] Bouldin, M.J., Ross, L.A., Sumrall, C.D., Loustalot, F.V., Low, A.K. & Land, K.K. (2006). The effect of obesity surgery on obesity comorbidity, *The American Journal of the Medical Sciences* **331**, 183–193.

[61] Poirier, P., Giles, T.D., Bray, G.A., Hong, Y., Stern, J.S., Pi-Sunyer, F.X. & Eckel, R.H. (2006). Obesity and cardiovascular disease: pathophysiology, valuation, and effect of weight loss, *Circulation* **113**, 898–918.

[62] Oluseun, A., Sowemimo, S.M., Yood, J.C., Moore, J., Huang, M., Ross, R., McMillian, U., Ojo, P. & Randolph, B.R. (2007). Natural history of morbid obesity without surgical intervention, *Surgery for Obesity and Related Diseases* **3**, 73–77.

[63] Poirier, P., Martin, J., Marceau, P., Biron, S. & Marceau, S. (2004). Impact of bariatric surgery on cardiac structure, function and clinical manifestations in morbid obesity, *Expert Review of Cardiovascular Therapy* **2**, 193–201.

[64] Drenick, E.J. & Fisler, J.S. (1988). Sudden cardiac arrest in morbidly obese surgical patients unexplained after autopsy, *American Journal of Surgery* **155**(6), 720–726.

[65] Rössner, S., Lagerstrand, L., Persson, H.E. & Sachs, C. (1991). The sleep apnoea syndrome in obesity: risk of sudden death, *Journal of Internal Medicine* **230**(2), 135–141.

[66] Duflou, J., Virmani, R., Rabin, I., Burke, A., Farb, A. & Smialek, J. (1995). Sudden death as a result of heart disease in morbid obesity, *American Heart Journal* **130**(2), 306–313.

[67] Graham, J., Mosunjac, M., Hanzlick, R.L. & Mosunjac, M. (2007). Sickle cell lung disease and sudden death: a retrospective/prospective study of 21 autopsy cases and literature review, *The American Journal of Forensic Medicine and Pathology* **28**(2), 168–172.

[68] Thomas, A.N., Pattison, C. & Serjeant, G.R. (1982). Causes of death in sickle cell disease in Jamaica, *British Medical Journal* **285**, 633–635.

[69] Gray, A., Anionwu, E.N., Davis, S.C. & Brozovic M. (1991). Patterns of mortality in sickle cell disease in the United Kingdom, *Journal of Clinical Pathology* **44**, 459–463.

[70] Adedeji, M.O., Cespedes, J., Allen, K., Subramony C & Hughson, M.D. (2001). Pulmonary thrombotic arteriopathy in patients with sickle cell disease, *Archives of Pathology and Laboratory Medicine* **125**, 1436–1441.

[71] Dail, D.H. & Hammar, S.P. (1993). *Pulmonary Pathology*, 2nd Edition, New York, Springer-Verlag, 685–691.

[72] Knight, J., Murphy, T. & Browning, I. (1999). The lung in sickle cell disease, *Pediatric Pulmonology* **28**, 205–216.

[73] Manci, E.A., Culberson, D.E., Yang, Y.M., Gardner, T.M., Powell, R., Haynes Jr, J., Shah, A.K. & Mankad, V.N. (2003). Causes of death in sickle cell disease: an autopsy study, *British Journal of Haematology* **123**(2), 359–365.

[74] Platt, O.S., Brambilla, D.J., Rosse, W.F., Milner, P.F., Castro, O, Steinberg, M.H. & Klug, P.P.. The correct title is Martality in sickle cell disease. Life expectancy and risk factors for early death. (1994a). Mortality in sickle cell disease, *The New England journal of Medicine* **330**, 1639–1644.

[75] Darbari, D.S., Kple-Faget, P., Kwagyan, J., Rana, S., Gordeuk, V.R. & Castro, O. (2006). Circumstances of death in adult sickle cell disease patients, *American Journal of Hematology* **81**(11), 858–863.

[76] Diggs L.W. (1984). The sickle cell trait in relation to the training and assignment of duties in the armed forces: III. Hyposthenuria, hematuria, sudden death, rhabdomyolysis, and acute tubular necrosis, *Aviation, Space, and Environmental Medicine* **55**(5), 358–364.

[77] Sateriale, M. & Hartm P. (1985). Unexpected death in a black military recruit with sickle cell trait: case report, *Military Medicine* **150**(11), 602–605.

[78] Kerle, K.K. & Nishimura, K.D. (1996). Exertional collapse and sudden death associated with sickle cell trait, *Military Medicine* **161**(12), 766–767.

[79] Evans, P. & Murray, M.J. (1997). Sudden exertional death and sickle cell trait, *American Family Physician* **55**(3), 784.

[80] Wirthwein, D.P., Spotswood, S.D., Barnard, J.J. & Prahlow, J.A. (2001). Death due to microvascular occlusion in sickle-cell trait following physical exertion, *Journal of Forensic Sciences* **46**(2), 399–401.

[81] Bock, H., Seidl, S., Hausmann, R. & Betz, P. (2004). Sudden death due to a haemoglobin variant, *International Journal of Legal Medicine* **118**(2), 95–97.

[82] Channa Perera, S.D. & Pollanen, M.S. (2007). Sudden death due to sickle cell crisis during law enforcement restraint, *Journal of Forensic and Legal Medicine* **14**, 297–300.

[83] Mitchell, B.L. (2007). Sickle cell trait and sudden death – bringing it home, *Journal of the National Medical Association* **99**(3), 300–305.

[84] Thogmartin, J.R. (1998). Sudden death in police pursuit, *Journal of Forensic Sciences* **43**(6), 1228–1231.

[85] Romero Mestre, J.C., Hernández, A., Agramonte, O. & Hernández, P. (1997). Cardiovascular autonomic dysfunction in sickle cell anemia: a possible risk factor for sudden death?, *Clinical Autonomic Research* **7**(3), 121–125.

[86] Schütt, M. & Meier M. (2005). Sudden death in sickle cell trait: could coexistent diabetes play a role? *Medical Hypotheses* **64**(1), 217.

[87] Gerber, N. & Apseloff, G. (1993). Death from a morphine infusion during sickle cell crisis, *The Journal of Pediatrics* **123**, 322–325.

[88] Oppenheimer, E.H. & Esterly, J.R. (1971). Pulmonary changes in sickle cell disease, *The American Review of Respiratory Disease* **103**, 858.

[89] Haque, A.K., Gokhale, S., Rampy, B.A., Adegboyega, P., Duarte, A. & Saldana, M.J. (2002). Pulmonary hypertension in sickle cell hemoglobinopathy: a clinicopathological study of 20 cases, *Human Pathology* **33**, 1037–1042.

[90] DiMaio, S.M., DiMaio, V.J. & Kirkpatrick, J.B. (1980). Sudden, unexpected deaths due to primary intracranial neoplasms, *The American Journal of Forensic Medicine and Pathology* **1**, 29–45.

[91] Gezelius, C. & Eriksson, A. (1988). Neoplastic disease in a medicolegal autopsy material: a retrospective study in northern Sweden, *Zeitschrift fur Rechtsmedizin* **101**, 115–130.

[92] Luke, J.L. & Helpern, M. (1968). Sudden unexpected death from natural causes in young adults: a review of 275 consecutive autopsied cases, *Archives of Pathology* **85**, 10–17.

[93] Nestok, B.R., Goldstein, J.D. & Lipkovic, P. (1988). Splenic rupture as a cause of sudden death in undiagnosed chronic myelogenous leukaemia, *The American Journal of Forensic Medicine and Pathology* **9**(3), 241–245.

[94] de Fijter, C.W., Schuur, J., Potter van Loon, B.J., Kingma, W.P. & Schweitzer, M.J. (1996). Acute cardiorespiratory failure as presenting symptom of chronic lymphocytic leukaemia, *The Netherlands Journal of Medicine* **49**(1), 33–37.

[95] Beaubien, E.R., Wilson, T.W. & Satkunam, N. (1998). Sudden death in a patient with chronic lymphocytic leukaemia, *Canadian Medical Association Journal* **159**(9), 1123–1125.

[96] Ahmed, S., Siddiqui, A.K., Rossoff, L., Sison, C.P. & Rai, K.R. (2003). Pulmonary complications in chronic lymphocytic leukaemia, *Cancer* **98**(9), 1912–1917.

[97] Aragona, M. & Aragona F. (2000). Unexpected death by leukostasis and lung leukostatic tumors in acute myeloid leukemia. Study of four cases, *Minerva Medica* **91**(10), 229–237.

[98] Lascari, A.D., Pearce, J.M. & Swanson, H. (1997). Sudden death due to disseminated cryptococcosis in a child with leukemia in remission, *Southern Medical Journal* **90**(12), 1253–1254.

[99] Hitosugi, M., Fukui, K., Takatsu, A., Harada, T., Homori, M. & Kawano, K. (1998). An autopsy case of sudden death caused by untreated sepsis after complete remission of acute promyelocytic leukaemia, *Nihon Hoigaku Zasshi* **52**(6), 355–359.

[100] Tsuda, N., Oka, R., Kajino, H., Kajino, M. & Okuno, A. (2000). Sudden death of a patient in complete remission after anthracycline therapy for acute lymphoblastic leukaemia, *Pediatrics International* **42**(3), 319–321.

[101] Byard, R.W. (2004). *Sudden Death in Infancy and Childhood*, 2nd Edition, Cambridge, Cambridge University Press, 643.

[102] Whybourne, A., Zillman, M.A., Miliauskas, J. & Byard, R.W. (2001). Sudden and unexpected infant death due to occult lymphoblastic leukaemia, *Journal of Clinical Forensic Medicine* **8**(3), 160–162.

[103] Somers, G., Smith, C.R., Perrin, D.G., Wilson, G.J. & Taylor, G.P. (2006). Sudden unexpected death in infancy and childhood ue to undiagnosed neoplasia: an autopsy study, *The American Journal of Forensic Medicine and Pathology* **27**(1), 64–69.

[104] Drolet, B., Simard, C. & Roden, D.M. (2004). Unusual effects of a QT-prolonging drug, arsenic trioxide, on cardiac potassium currents, *Circulation* **109**(1), 26–29.

[105] Tattersall, R.B. & Gill, G.V. (1991). Unexplained sudden death of type 1 diabetic patients, *Diabetic Medicine* **8**, 49–58.

[106] Start, R.D., Barber, C., Kaschula, R.O. & Robinson, R.T. (2007). The 'dead in bed syndrome' – a cause of sudden death in Type 1 diabetes mellitus, *Histopathology*. **51**(6), 843–845.

[107] Weston, P.J. & Gill, G.V. (1999). Is undetected autonomic dysfunction responsible for sudden death in Type 1 diabetes mellitus? The dead in bed syndrome revisited, *Diabetic Medicine* **16**, 626–631.

[108] Suarez, G.A., Clark, V.M., Norell, J.E., Kottke, T.E., Callahanm M.J., O'Brien, P.C., Low. P.A. & Dyck, P.J. (2005). Sudden cardiac death in diabetes mellitus: risk factors in the Rochester diabetic neuropathy study, *Journal of Neurology, Neurosurgery, and Psychiatry* **76**(2), 240–245.

[109] Kahn, J.K., Sisson, C. & Vinik, A.I. (1988). Prediction of sudden cardiac death in diabetic autonomic neuropathy, *Journal of Nuclear Medicine* **29**, 1605–1606.

[110] Stevens, M., Dayanikli, F., Raffelm D., Allman, K., Standford, T., Feldman, E., Wieland, D., Corbett, J. & Schwaiger, M. (1988). Scintigraphic assessment of regionalized defects in myocardial sympathetic intervation and blood flow regulation in diabetic patients with autonomic neuropathy, *Journal of the American College of Cardiology* **31**, 1575–1584.

[111] Vinik, A.I. & Ziegler, D. (2007). Diabetic cardiovascular autonomic neuropathy, *Circulation* **115**(3), 387–397.

[112] Mazzone, T. (2007). Prevention of macrovascular disease in patients with diabetes mellitus: opportunities for intervention, *The American Journal of Medicine* **120**(9 Suppl 2), S26–S32.

[113] Movahed, M.R., Hashemzadeh, M. & Jamal, M. (2005a). Diabetes mellitus is a strong, independent risk for atrial fibrillation and flutter in addition to other cardiovascular disease, *International Journal of Cardiology* **105**(3), 315–318.

[114] Movahed, M.R., Hashemzadeh, M. & Jamal, M. (2005b). Increased prevalence of third-degree atrioventricular block in patients with type II diabetes mellitus, *Chest* **128**(4), 2611–2614.

[115] Giunti, S., Bruno, G., Lillaz, E., Gruden, G., Lolli, V., Chaturvedi, N., Fuller, J.H., Veglio, M. & Cavallo-Perin, P. (2007). Incidence and risk factors of prolonged QTc interval in type 1 diabetes: the EURODIAB Prospective Complications Study, *Diabetes Care* **30**(8), 2057–2063.

[116] Movahed, M.R., Hashemzadehm M., Jamal, M. (2007). Increased prevalence of ventricular fibrillation in patients with type 2 diabetes mellitus, *Heart and Vessels* **22**(4), 251–253.

[117] Tsai, A.W., Cushman, M., Rosamond, W.D., Heckbert, S.R., Polak, J.F. & Folsom, A.R. (2002). Cardiovascular risk factors and venous thromboembolism incidence: the longitudinal investigation of thromboembolism etiology, *Archives of Internal Medicine* **162**, 1182–1189.

[118] Movahed, M.R., Hashemzadeh, M. & Jamal, M. (2005c). The prevalence of pulmonary embolism and pulmonary hypertension in patients with Type II diabetes mellitus, *Chest* **128**, 3568–3571.

[119] Khuu, H.M., Robinson, C.A., Brissie, R.M. & Konrad, R.J. (1999). Postmortem diagnosis of unsuspected diabetes mellitus established by determination of decedent's hemoglobin A1c level, *Journal of Forensic Sciences* **44**(3), 643–646.

[120] Dillon, E.S., Dyer, W.W. & Smelo, L.S. (1940). Ketone acidosis in nondiabetic adults, *The Medical Clinics of North America* **24**, 1813–1822.

[121] Thomsen, J.L., Felby, S., Theilade, P., Nielsen, E. (1999). Alcoholic ketoacidosis as a cause of death in forensic cases, *Forensic Science International* **75**, 163–171.

[122] Pounder, D.J., Stevenson, R.J. & Taylor, K.K. (1998). Alcoholic ketoacidosis at autopsy, *Journal of Forensic Sciences* **43**(4), 812–816.

[123] Brinkmann, B., Fechner, G., Karger, B. & DuChesne, A. (1998). Ketoacidosis and lactic acidosis – frequent causes of death in chronic alcoholics?, *International Journal of Legal Medicine* **111**(3), 115–119.

[124] Iten, P.X. & Meier M. (2000). Beta-hydroxybutyric acid – an indicator for an alcoholic ketoacidosis as cause of death in deceased alcohol abusers, *Journal of Forensic Sciences* **45**(3), 624–632.

[125] Koch, B. (2006). Selected topics of hypoglycemia care, *Canadian Family Physician* **52**(4), 466–471.

[126] Collier, A., Mathews, D.M., Young, R.J. & Clarke, B.F. (1987). Transient atrial fibrillation precipitated by hypoglycaemia: two case reports, *Postgraduate Medical Journal* **63**, 895–897.

[127] Lindstrom, T., Jorfeldt, L., Tegler, L. & Arnquist, H.J. (1992). Hypoglycaemia and cardiac arrhythmias in patients with Type 2 diabetes, *Diabetic Medicine* **9**, 536–541.

[128] Baxter, M.A., Garewal, C., Jordan, R., Wright, A.D. & Nattrass, M. (1995). Hypoglycaemia and atrial fibrillation, *Postgraduate Medical Journal* **66**, 981.

[129] Heller, S.R. (2002). Abnormalities of the electrocardiogram during hypoglycaemia: the cause of the dead in bed syndrome?, *International Journal of Clinical Practice. Supplement* **129**, 27–32.

[130] Robinson, R.T.C.E., Harris, N.D., Ireland, R.H., Lee, S., Newman, C. & Heller, S.R. (2003). Mechanisms of abnormal cardiac repolarization during insulin-induced hypoglycemia, *Diabetes* **52**, 1469–1474.

[131] Denmark, L.N. (1993). The investigation of beta-hydroxybutyrate as a marker for sudden death due to hypoglycemia in alcoholics, *Forensic Science International* **62**(3), 225–232.

[132] Simson Jr, L.R. (1976). Thyrotoxicosis: postmortem diagnosis in an unexpected death, *Journal of Forensic Sciences* **21**, 831–832.

[133] Terndrup, T.E., Heisig, D.G. & Garceaum, J.P. (1990). Sudden death associated with undiagnosed Graves disease, *The Journal of Emergency Medicine* **8**, 553–555.

[134] Randall, B.B. (1992). Fatal hypokalemic thyrotoxic periodic paralysis presenting as the sudden, unexplained death of a Cambodian refugee, *The American Journal of Forensic Medicine and Pathology* **13**, 204–206.

[135] Guthrie, G.P., Hunsaker III, J.C., O'Connor, W.N. (1987). Sudden death in hypothyroidism, *The New England journal of Medicine* **317**(1291), 1–5.

[136] Edston, E. (1996). Three sudden deaths in men associated with undiagnosed chronic thyroiditis, *International Journal of Legal Medicine* **109**(2), 94–97.

[137] De Letter, E.A., Piette, M.H., Lambert, W.E. & De Leenheer, A.P. (2000). Medico-legal implications of hidden thyroid dysfunction: a study of two cases, *Medicine, Science, and the Law* **40**(3), 251–257.

[138] Vestergaard, V., Drostrup, D.H. & Thomsen, J.L. (2007). Sudden unexpected death associated with lymphocytic thyroiditis, *Medicine, Science, and the Law* **47**(2), 125–133.

[139] Siegler, R.W. (1998). Fatal heatstroke in a young woman with previously undiagnosed Hashimoto's thyroiditis, *Journal of Forensic Sciences* **43**(6), 1237–1240.

[140] Lorin De La Grandmaison, G., Izembart, M., Fornes, P., Paraire, F. (2003). Myocarditis associated with Hashimoto's disease, a case report, *International Journal of Legal Medicine* **117**(6), 361–364.

[141] Edston, E., Druid, H., Holmgren, P. & Öström, M. (2001). Postmortem measurements of thyroid hormones in blood and vitreous humor combined with histology, *The American Journal of Forensic Medicine and Pathology* **22**(1), 78–83.

[142] Sampson, H.A., Munoz-Furlong, A., Bock, S.A., Schmitt, C., Bass, R., Chowdhury, B.A., Decker, W.W., Furlong, T.J., Galli, S.J., Golden, D.B., Gruchalla, R.S., Harlor, Jr, A.D., Hepner, D.L., Howarth, M., Kaplan, A.P., Levy, J.H., Lewis, L.M., Lieberman, P.L., Metcalfe, D.D., Murphy, R., Pollart, S.M., Pumphrey, R.S., Rosenwasser, L.J., Simons, F.E., Wood, J.P. & Camargo, C.A., Symposium on the definition and management of anaphylaxis: summary report, *The Journal of Allergy and Clinical Immunology* (2005). **115**, 584–591.

[143] Pumphrey, R.S.H. & Roberts, I.S.D. (2000). Postmortem findings after fatal anaphylactic reactions, *Journal of Clinical Pathology* **53**, 273–276.

[144] Greenberger, P.A., Rotskoff, B.D. & Lifschultz, B. (2007). Fatal anaphylaxis: postmortem findings and associated comorbid diseases, *Annals of Allergy, Asthma and Immunology* **98**(3), 252–257.

[145] Yunginger, J.W., Nelson, D.R., Squillace, D.L., Jones, R.T., Holley, K.E., Hyma, B.A., Biedrzycki, L., Sweeney, K.G., Sturner, W.Q. & Schwartz, L.B. (1991). Laboratory investigation of deaths due to anaphylaxis, *Journal of Forensic Sciences* **36**, 857–865.

[146] Ansari, M.Q., Zamora, J.L. & Lipscomb, M.F. (1993). Postmortem diagnosis of acute anaphylaxis by serum tryptase analysis, *American Journal of Clinical Pathology* **99**, 101–103.

[147] Schwartz, L.B. (2006). Diagnostic value of tryptase in anaphylaxis and mastocytosis, *Immunology and Allergy Clinics of North America* **26**, 451–463.

[148] Parish W.E., Barrett, A.M., Coombs, R.R.A., Gunther, M. & Camps, F.E. (1960). Hypersensitivity to milk and sudden death in infancy, *Lancet* **II**, 1106–1110.

[149] Parish, W.E., Richards, C.B., France, N.E. & Coombs, R.R.A. (1964). Further investigations on the hypothesis that some cases of cot death are due to a modified anaphylactic reaction to cows' milk, *International Archives of Allergy and Applied Immunology* **24**, 215–243.

[150] Nishio, H. & Suzuki, K. (2004). Serum tryptase levels in sudden infant death syndrome in forensic autopsy cases, *Forensic Science International* **139**(1), 57–60.

[151] Platt, M.S., Yunginger, J.W., Sekula-Perlman, A., Irani, A.M.A., Smialek, J., Mirchandani, H.G. & Schiwartz, L.B. (1994b). Involvement of mast cells in sudden death syndrome, *The Journal of Allergy and Clinical Immunology* **94**, 250–256.

[152] Hagan, L.L., Goetz, D.W., Revercomb, C.H. & Garriott, J.G. (1998). Sudden infant death syndrome: a search for allergen hypersensitivity, *Annals of Allergy, Asthma and Immunology* **80**, 227–231.

[153] Edson, E., Gidlund, M., Wickman, H., Ribbing & van Hage-Hamsten, M. (1999). Increased mast cell tryptase in sudden infant death – anaphylaxis, hypoxia or artefact?, *Clinical and Experimental Allergy* **29**, 1648–1654.

[154] Buckley, M.G., Variend, S., Walls, A.F. (2001). Elevated serum concentrations of b-tryptase, but not a-tryptase, in sudden infant death syndrome (SIDS). An investigation of anaphylactic mechanisms, *Clinical and Experimental Allergy* **31**, 1696–1704.

[155] Varon, J., Chen, K., Sternbach, G.I. (1998). Rupert Waterhouse and Carl Friderichsen, adrenal apoplexy, *The Journal of Emergency Medicine* **16**, 643–647.

[156] Doherty, S. (2001). Fatal pneumococcal Waterhouse-Friderichsen syndrome, *Emergency Medicine (Fremantle, W.A.)* **13**(2), 237–239.

[157] Tsokos, M. (2003). Fatal Waterhouse-Friderichsen syndrome due to *Ewingella Americana* infection, *The American Journal of Forensic Medicine and Pathology* **24**(1), 41–44.

[158] Hamilton, D., Foweraker, H.J. & Gresham, G.A. (2004). Waterhouse-Friderichsen syndrome as a result of non-meningococcal infection, *Journal of Clinical Pathology* **57**(2), 208–209.

[159] Adem, P.V., Montgomery, C.P., Husain, A.N., Koogler, T.K., Arangelovich, V., Humilier, M., Boyle-Vavra, S. & Daum, R.S. (2005). *Staphylococcus aureus* sepsis and the Waterhouse-Friderichsen syndrome in children, *The New England journal of Medicine* **353**(12), 1245–1251.

[160] Tsokos, M. (2002). Postmortem measurement of serum procalcitonin concentration in Waterhouse-Friderichsen syndrome, *Virchows Archieve* **441**(6), 629–631.

[161] Tsokos, M., Reichelt, U., Nierhaus, A. & Puschel, K. (2001). Serum procalcitonin (PCT): a valuable biochemical parameter for the post-mortem diagnosis of sepsis, *International Journal of Legal Medicine* **114**, 237–243.

Further Reading

McGee, J.O.D., Isaacson, P.G. & Wright, N.A. (1992). *Oxford Textbook of Pathology*, Oxford University Press, Oxford.

Sampson, U.A., Muñoz-Furlong, A., Bock, A., *et al.* (2005). Symposium on the definition and of anaphylaxis, summary report, *The Journal of Allergy and Clinical Immunology* **115**, 584–591.

VITTORIO FINESCHI AND EMANUELA TURILLAZZI

Neuropsychological Assessment

Neuropsychological assessment refers to the application of standardized psychological measurement techniques to determine the relationship between brain impairment and its cognitive and behavioral concomitants (*see also* **Head Injury: Neuropsychological Assessment**). Forensic neuropsychological assessment is the use of such techniques in responding to official or legal questions. Forensic neuropsychologists are increasingly being called upon to provide opinions in a wide range of types of legal referrals. A recent count of citations of the word "neuropsychologist" in a database of court decisions yielded a large and growing count of cases involving neuropsychological consultation. Forensic issues were recently identified as the most-researched areas in the field of neuropsychology, with recent exponential growth in the number of studies and articles referencing forensic matters, such as malingering assessment (*see* **Malingering: Forensic Evaluations**).

Choice and Credentialing of the Neuropsychological Expert

The clinical neuropsychologist is typically a clinical psychologist who is also trained in the neurosciences and clinical neurology, and who has completed appropriate postdoctoral training and experience in the field of neuropsychology. Definitions of clinical neuropsychologists have been published by Division 40 (neuropsychology) of the American Psychological Association (APA) [1] and by the National Academy of Neuropsychology [2]. These definitions define clinical neuropsychologists as being licensed psychologists with appropriate graduate training in psychology, neuropsychology, and/or the neurosciences and at least two years of supervised postdoctoral experience. The Division 40 definition includes review by one's peers as a test of these competencies, and specifically references diplomate status as the "clearest evidence" that competence as a clinical neuropsychologist is met.

The choice of a forensic neuropsychological expert should take these issues of training and credentialing into account. One should also consider the neuropsychologist's experience with the disorder in question. For example, in cases involving closed head injuries, a neuropsychologist with experience in working with head injury patients in a rehabilitation setting may be preferred. The neuropsychologist's history as a forensic expert should also be taken into account. Some attorneys may prefer to choose neuropsychologists with a substantial history of testifying in court, while others may prefer experts who have little or no track record as an expert witness. In any event, a neuropsychologist who has consistently testified in prior cases for both the plaintiff and the defense, or the prosecution and the defense in criminal matters, is a plus. Consider also the neuropsychologist's capacity for teaching or educating the court and/or jury regarding the clinical issues that are apparent in the case. Reports should be written in an accessible, straightforward style with normal language and avoidance of highly technical terms as much as possible.

Admissibility of Neuropsychological Evidence

Properly credentialed neuropsychologists are able to serve as expert witnesses in most jurisdictions (*see also* **Expert Opinion: United States**). In some states, such as North Carolina and Georgia, there has been case law limiting neuropsychologists from expressing opinions regarding the medical causation of brain injuries. In Georgia, this issue was remedied with legislation. In most states, however, courts have consistently found that neuropsychologists are qualified to render opinions regarding the causation and pathology of brain insults. The Iowa Supreme Court case decision in *Hutchison v. American Family Mutual Insurance Company* provides an outline for the use of neuropsychological testimony in court, including that neuropsychologists are qualified to diagnose the general state of the brain and the causation of brain pathology. The *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (*see also* ***Daubert v. Merrell Dow Pharmaceuticals***) case sets out key principles for the admission of scientific evidence and expert testimony in courts: whether the technique can be tested (falsifiability), has the technique been published (peer review), is there an acceptable error rate, whether the techniques have gained general acceptance in the scientific field, and whether there

is a published guide or manual to govern the use of the instrument. In general, most of the widely used neuropsychological tests meet these standards; indeed, neuropsychology as a profession, because of its emphasis on empirical methods of assessment and long tradition of careful research, seems ideally suited to meet *Daubert* challenges in court.

The American Academy of Neurology [3] has produced official guidelines affirming the value of neuropsychological assessment in the practice of neurology. Professional psychological practice organizations, such as the National Academy of Neuropsychology and the American Academy of Clinical Neuropsychology, have published guidelines for performing forensic neuropsychological assessments. These guidelines are intended to describe the "most desirable and highest level of professional conduct", as opposed to other documents such as APA's *Ethical Principles of Psychologists and Code of Conduct*, which simply describe standards for competent and adequate conduct. These guidelines indicate that neuropsychologists working for a third party, such as an insurance company, attorney, or the court, must inform the examinee of the nature of the evaluation, the neuropsychologist's duty to remain objective, and the destination of the report that will be produced. The presence of a clinician–patient relationship must not be implied, but disavowed. The examination must be conducted in a manner adequate to answer the questions posed by the referral source. The measurement procedures must meet standards for psychometric accuracy, including the presence of acceptable reliability, validity, and appropriate normative standards. The potential for motivational bias must be addressed in the testing, and the report should reflect a reasoned, knowledgeable assessment of potential biasing issues in the assessment. Interpretations of test results that appear to deviate from normal levels must take into account other information as well, such as information about the patient's history, information from the direct observation of the patient, the functional abilities of the patient, and the nature of the potential neuropsychological syndromes that could account for the test findings.

The Uses and Utility of Neuropsychological Assessment

Neuropsychological assessment is often indicated when questions about cognition or brain–behavior relationships are raised during the course of a medical evaluation that are not explicable by history, mental status observations, radiological techniques, or laboratory techniques alone. The neuropsychologist is typically called upon when a complex set of symptoms and behaviors dictates in-depth measurement and observation of an individual's neurocognitive function and/or neurobehavioral characteristics. In the forensic realm, neuropsychological assessment is often sought when an allegation of brain injury or brain dysfunction is relevant to a given legal question.

The potential types of forensic referrals are many and varied. In the criminal forensic area, these include questions of trial competence in individuals with suspected neurocognitive impairment; criminal responsibility evaluations in defendants with suspected brain dysfunction or severe mental illness; evaluations to determine other criminal competencies, such as competence to waive one's Miranda rights or waive one's rights to appeal; evaluations to determine factors relevant to sentence mitigation; and evaluations to determine the appropriateness for transfer of a juvenile's case to an adult court. In the civil realm, a nonexhaustive list would include evaluations in personal injury cases in situations of alleged brain injuries, evaluations of injured employees in workers' compensation cases, alleged dementia or memory problems in disability applicants, and evaluation of civil competencies, such as evaluation of testamentary capacity or ability to manage one's affairs in the case of an allegedly demented older individual.

Neuropsychologists are often called upon to estimate whether a brain-injured individual has returned to a prior level of function, such as when confronted with the question of whether the patient can return to work. This is an issue of great interest in many medicolegal situations, relevant to the estimation of compensable impairment in workers' compensation injuries, or damage liability in personal injury cases. For example, after receiving a left frontal contusion, a professional editor may expect to encounter difficulty in tasks that tap verbal reasoning, whereas a skilled machinist might be just as debilitated from a comparable right-hemisphere lesion.

Examination of these referrals often hinges upon comparison of current test results with premorbid abilities. Premorbid function can be assessed informally by taking a careful history including prior educational achievement, occupations, daily activities, hobbies, and personal achievements. Both actuarial

and assessment techniques also exist that allow estimation of prior function. Prior intelligence can be predicted within a degree of error by mathematical regression techniques, utilizing demographic information such as educational level, occupational history, age, and region of the country. Reading ability is highly resistant to most acquired cognitive disorders, with the exception, of course, of alexia. Thus reading skill is a good marker of prior function. Other cognitive skills that can be assessed with the neuropsychological evaluation tend to remain constant, even with significant cortical insults. The most powerful techniques are those that mathematically combine both psychometric data with demographic estimation techniques (e.g., Wechsler Test of Adult Reading, Oklahoma Premorbid Intelligence Assessment).

Another important set of referral questions addresses the patient's level of residual functional capacity after recovery. Normal resumption of activities of daily living, socialization, and performance in work-like settings may not be possible depending on age, course of illness, severity of injury, and the presence of concomitant psychiatric illness. These are functional questions that may be best answered by a combination of assessment techniques, including administering formalized tests (see next section), observing the patient in demanding environments, interviewing the family, and taking a complete history of the illness.

Neuropsychological Assessment Instruments

The most recent version of a popular handbook for neuropsychologists reviews more than 160 tests which have found their way into common use among neuropsychologists Straus *et al.* [4], and these are only a small subset of the tests available commercially or in the research literature. Instead of attempting to present specific tests to the reader, categories and representative examples of neuropsychological tests will be described here, in order to convey a sense of the range and utility of neuropsychological testing.

Intelligence Tests in Neuropsychological Assessment

A comprehensive intelligence test, usually the Wechsler test appropriate to the age of the patient, often serves as the backbone of most neuropsychological evaluations. Comprehensive intelligence tests have the advantage of carefully developed norms and the fact that they tap many key neurocognitive abilities (e.g., verbal expression, processing speed, visual analytic skills). Because of the development of the standardization samples on these tests, it is also possible to account for the expected effects of such variables as educational level and ethnicity, if desired. A disadvantage is the fact that comprehensive intelligence quotient (IQ) scores and summary index scores tend to be resistant to changes caused by neurocognitive disorders. Partially, this is due to the fact that neurocognitive disorders rarely cause generalized dysfunction except in the more severe cases. Also, intelligence tests are not weighted heavily with executive function and memory tests; the tests are shown to be most sensitive to cortical dysfunction. Indeed, it is not unusual for an early to mid-stage Alzheimer's patient to be relatively debilitated due to memory loss, but score near-normal limits on most sections of the Wechsler Adult Intelligence Scale, or for a Pick's Disease (an illness involving deterioration of executive functioning and personality change due to brain degeneration) patient to exhibit grossly inappropriate behavior but to have an average or better IQ.

Folstein Mini-Mental State Examination (MMSE)

The Mini-Mental State Examination (MMSE) is a universally recognized neuropsychological test that is widely used by psychiatrists, psychologists, neurologists, and geriatric practitioners. This test taps several cognitive functions including orientation, word memory, naming, verbal comprehension, writing, and drawing. Benefits of the MMSE are the instant recognizability of the test and the fact that so many clinicians across differing disciplines can readily interpret its scores. Unfortunately, it is also short on items tapping executive function, and its memory section, as generally administered, is crude. This test is often supplanted with tests that have additional memory and executive items to allow greater sensitivity to senile dementia and a broader range of assessed functions (e.g., Modified MMSE; Dementia Rating Scale).

Halstead–Reitan Neuropsychological Test Battery

The Halstead–Reitan Neuropsychological Test Battery (HRNTB) contains many of the most commonly

used neuropsychological tests. It originated from the work of Ward Halstead, who in 1947 at the University of Chicago published his observations of several hundred case studies of patients who had frontal lobe damage. By using 10 scores, Halstead blindly distinguished patients with confirmed brain lesions from control subjects. Reitan [5], a student of Halstead, modified the battery in 1955 to identify lateralizing features of patient performances such as motor deficits expected in subtle stroke, the effect of temporal lobe epilepsy on memory, and the loss of abstraction ability associated with frontal damage. Reitan also modified the original battery to include tests that would accurately measure aphasia and variations of normal aging. An extensive norming project [6] has provided an excellent source of demographically corrected norms for the tests.

The Halstead Impairment Index is a global measure of brain dysfunction in neuropsychology backed by research and a wide degree of acceptance. Most normal subjects are able to pass 60–100% of the tests included in the Index. Patients who have moderate impairments may be within the normal range on only 30–60% of the tests, and those with severe dysfunction on less than 30%. From continued use of the battery over the decades, 3 of the original 10 scores were dropped because of questionable validity. The Index includes seven scores within the following five subtests:

1. Category Test

The Category Test is an abstract reasoning task consisting of 180 items. The patient is required to use mental flexibility and problem solving to form concepts, utilizing feedback from the examiner about the accuracy of their attempts.

2. Tactual Performance Test

The Tactual Performance Test is a test requiring the integration of multiple cognitive skills: spatial skills, spatial memory, dexterity, processing speed, and planning ability. The patient is blindfolded and placed before a board with cutouts into which blocks can be inserted. The patient is then asked to place the blocks in the cutouts, first using the dominant hand, then the nondominant hand, then both. Then the patient is asked to draw a representation of the board from memory. Since more than 40% of the brain's processing power is devoted to visual processing, removal of vision presumably cripples the efficiency with which the patient can approach the task. For the brain-injured patient, excessive time is often required to complete this task. The lateralized nature of the test allows some comparison of relative hemispheric efficiency.

3. Finger Tapping Test

Finger Tapping is a test of fine motor speed. Five consecutive 10 s trials are obtained with the dominant and nondominant index fingers, and then compared to norms. Injury to either region of motor cortex, as well as injuries affecting overall cortical efficiency, can result in degraded scores on this task.

4. Rhythm Test

The Rhythm Test originated from the Seashore Measures of Musical Talent test, in which the patient is asked to differentiate between 30 pairs of rhythmic patterns. These pairs are presented in rapid succession on a tape recorder, and the patient must distinguish whether they are the same or different.

5. Speech Sounds Perception Test

The Speech Sounds Perception Test involves 60 spoken nonsense words that are administered by audiotape. The patient is required to underline the corresponding printed response on an answer sheet, measuring verbal discrimination and sustained attention.

General Neuropsychological Deficit Scale

In addition to the seven scores provided by five index subtests, other subtests from the HRNTB are used to represent patients' performances on the General Neuropsychological Deficit Scale (GNDS). Subtests used to contribute to this scale, in addition to those in the Index, include the Lateral Dominance Examination, Grip Strength, the Sensory-Perceptual Examination, Tactile Form Recognition, the Trail Making Test Parts A and B, and the Aphasia Screening Test. The GNDS, much like the Impairment Index, provides a global impairment rating but takes into account 42 variables, thereby increasing reliability. Clinical judgment also allows the clinician to take into account variables in the interpretation of data related to the level and pattern of performance, pathognomonic signs, and laterality. In practice, though, the entire HRNTB is rarely given, with the neuropsychologist instead relying upon subtests of the battery scored with the help of the demographically corrected norms, described above.

Functions Measured in Neuropsychological Assessments

Sensorimotor Ability

Many neuropsychologists administer a partial neurological examination as part of their overall assessment. Depending on the referral questions, the sensorimotor assessment may include informal testing of olfactory function, visual fields, auditory perception, stereognosis (tactile appreciation of objects), and cerebellar function (e.g., heel-to-shin testing, alternating hand movements). As mentioned above, standardized testing of strength, motor speed (hand and foot tapping), and fine motor control are generally a component of the examination.

Attention and Concentration

After ensuring an adequate level of consciousness, assessment of attention provides fundamental information as to whether evaluation of other cognitive domains, such as intelligence, memory, or language, will be valid. Further, because the ability to focus and maintain attention is highly sensitive to many acute and ongoing conditions (e.g., alcohol withdrawal, intoxication, delirium), a stable, chronic process may be affected less than will a recent or changing one. At the most basic level, this domain measures the patient's ability to attend to incoming information without being distracted. Examples of simple attentional tasks include digit repetition or visual tapping span, requiring forward or backward sequencing of auditory or visual stimuli.

Sustained attention, known as *vigilance or continuous performance*, is another type of attention. An area of cognitive assessment which shows great promise for practitioners who work with attention deficit hyperactive disorder (ADHD) children and adults is continuous performance testing for sustained attention (e.g., Conners Continuous Performance Test and Test of Variables of Attention). Continuous performance tests are most often computer administered, and generally exist in either visual or auditory formats. In the visual modality, the child (or adult) is asked to sit before a computer screen and to respond to one type of stimulus while suppressing responses to another. The tests go on for several minutes, and the boring nature of the task tends to elicit omissions and variable response times from individuals with attentional deficits. While only limited success has been achieved with these tests as diagnostic instruments for ADHD, they are quite successful in measuring the effectiveness of psychostimulant medication on individuals known to have ADHD. As such, an assessment using these tests can allow the psychiatrist to avoid the use of costly and potentially risky psychostimulant medications among patients who are unresponsive to its effects. These tests are also widely used by neuropsychologists to assess the attentional abilities of brain-injured patients.

Learning and Memory

Compromise of memory is the most common patient complaint and referral question for neuropsychological testing. Milder memory problems, or problems with visual memory, may not be readily apparent on screening tasks such as the MMSE. Since memory is not a static or unitary process, careful assessment can help to characterize variations in performance, which can have diagnostic value. Patients who have anterograde amnesia, or faulty learning of new material or events after the onset of their disorder, have more difficulty consolidating their learning experiences into longer storage. Retrograde amnesia, or the inability to retrieve remote memories, is less prominent in organic memory loss, especially if the onset is sudden, as in head injury. However, certain chronic conditions and disease processes, such as Korsakoff's syndrome and Alzheimer's disease, may produce a more dense retrograde amnesia. Memory is often evaluated in the mental status examination by assessing orientation, current events, and recall of words or objects. Although screening is sometimes adequate to determine the presence of a gross dementia, full neuropsychological evaluation can detect whether the difficulty is with encoding, storage, or retrieval mechanisms, and whether the impairment is associated more with one modality than another (e.g., verbal or visual).

The most widely used global memory test is the Wechsler Memory Scale (WMS), now undergoing its most recent revision, which will result in the WMS–IV. These tests yield memory quotients similar to IQ scores (expected mean is 100; standard deviation is 15). It provides for evaluation of immediate memory, delayed memory, and working memory skills, in both visual and verbal modalities.

Language

Testing for aphasia is a necessary part of diagnosis and treatment planning in some developmental and learning disabilities, progressive disorders (e.g., dementia and tumor), or recovery from an acute injury such as stroke or head injury. Impairment of language on a gross level is often more noticeable because of the frequent need for clear communication. However, subtle language deficits can go undetected in informal conversation and occasionally in more formal assessment. At the very least, language screening should include tasks to measure quality of spontaneous speech, naming, comprehension, repetition, reading, writing, calculation, and left–right orientation. Naming, the most sensitive element of most underlying language disturbances, is often measured through confrontational naming tasks such as the Boston Naming Test, a 60-item test of picture identification. Well-studied aphasia screening batteries include the Boston Diagnostic Aphasia Examination and the Western Aphasia Battery. If the neuropsychological assessment confirms the presence of aphasia, a speech language pathology evaluation may be required to plan detailed treatment.

Executive Functioning

Executive functioning refers to cognition and personality skills that are integrative in nature, allowing the person to attend to salient stimuli while disregarding others, maintain situational awareness, plan, reason hypothetically, solve problems, self-monitor demands and emotional reactions; in short, executive function is composed of all meta-cognitive and emotional processes that serve to help the person adapt to the environment or reach an overarching goal. Though many of these functions are partly, if not primarily, subserved by the frontal lobes, their complex nature generally requires the entire brain working in concert for maximal success; thus the descriptive term *integrative functions* is also often used. Neuropsychological functioning under the rubric of executive tasks includes abstraction, problem solving, set generation and sequencing, ability to maintain or terminate behaviors, and ability to plan and organize. These functions are mediated by the frontal lobes. Personality characteristics having to do with judgment, social appropriateness, inhibition versus impulsivity, and motivation are also at least partially related to frontal activity. The most famous example of an organically driven personality change is that of Phineas Gage, a railroad worker in the 1880s, who sustained a traumatic injury to the frontal lobes when a tamping iron was blown through his head. This formerly docile, responsible worker became irascible, foul-mouthed, and disinhibited. As his reckless and menacing behavior continued, his coworkers described him as "no longer Gage".

Specific tasks that are frequently used to study the integrity of the frontal systems include the Wisconsin Card Sorting Test. This procedure provides an index of how well the examinee can formulate hypotheses and solve problems. Cards are matched by principles that are not stated directly, and the patient must shift cognitive sets as the rules for sorting change without forewarning. Other good measures of executive functioning include Controlled Oral Word Fluency, which requires verbal set generation (e.g., words beginning with certain letters), the Category Test (concept formation), and the Trail Making Test (measuring visual shifting and sequencing).

Visuospatial Functioning

Visuospatial functioning measures both the patient's ability to function within his or her environment with regard to recognition of objects and his or her construction and perception of spatial relations. Such abilities are important in daily activities that require the patient to recognize faces and remember geographic location and spatial orientation and in procedural learning tasks such as driving a car. Since patients with progressive neurologic disorders, such as Alzheimer's disease, and patients with right-hemisphere strokes, are particularly susceptible to deficits in this area, careful testing will be able to determine whether the patient is likely to encounter problems in getting lost or living independently. Figure copying, clock drawing, visual organization of parts to whole, facial recognition, and map orientation are samples of tasks used to assess this domain.

Malingering in Neuropsychological Assessment

The accuracy of neuropsychological assessment techniques is largely dependent upon the effort and/or truthfulness of the examinee. Malingering (*see also*

Malingering: Forensic Evaluations), or faking, neurocognitive problems is a key consideration in performing forensic neuropsychological assessments. It is estimated that approximately 30–50% of individuals undergoing forensic neuropsychological assessment intentionally underperform on cognitive testing, or otherwise seek to magnify their symptoms. Even young children have been shown in research studies to be able to successfully mislead skilled clinicians. The prospect that one out of every two or three medicolegal patients is feigning or exaggerating is a sobering one.

Patients often malinger physical disorders, pain, psychosis, posttraumatic stress disorder, severe depression, panic disorder, seizures, and memory deficits, to name a few feigned presentations. Disorders that are episodically manifested, such as panic attacks or seizures, represent a unique challenge to the clinician because of the relative lack of power of the mental status examination. In our experience, memory deficits are often feigned, presumably because it seems to many patients that loss of memory would be easy to both describe and feign on examination, though these expectations are, in fact, inaccurate.

The most powerful tool the clinician has in identifying malingered disorders is thorough knowledge of, and experience with, the disorder in question. It is a challenge for most patients to learn what symptoms to report to feign a given disorder, and an even greater challenge to convincingly produce facsimiles of symptoms upon mental status examination. For this reason, referrals to rule out malingering should preferably be sent to clinicians who are very experienced with the disorder in question. The patient suspected of feigning schizophrenia should be referred to a specialist in psychotic disorders, and the apparently head-injured patient to a clinician with extensive experience with patients who have experienced mild, moderate, and severe head injuries.

Detection of Malingered Deficits

Many patients present with malingered memory or reasoning deficits. Intellectual, neuropsychological, and other cognitive tests can be easily compromised by poor effort, and often such lack of effort is intentional. Without formal testing for malingering, low test scores or apparent dementia on mental status examination may inaccurately result in classifications of patients as demented or mentally retarded.

Cognitive malingering tests utilize several different strategies to detect feigned cognitive problems, and fortunately knowledge of these strategies does not always subvert the tests, even in clever and well-informed patients. One set of tests utilizes floor effect tests, that is, tests that appear to be difficult but, in fact, are nearly always successfully performed even by individuals with moderate cognitive impairment. The malingering patient often scores well below the expected level as compared to genuinely impaired patients, thereby allowing a probabilistic determination of malingering to be made.

Forced choice tests are another strategy, often combined with the strategy above. These tests utilize the fact that forcing the patient to choose between two dichotomous responses, one correct and the other incorrect, results in a known expected value if the patient has no ability in that area at all: 50%. If a patient scores significantly worse than he or she would by flipping a coin, then strong evidence of malingering is produced. In that circumstance, the only credible explanation is that the patient must have known the right answer, since the wrong one was so often chosen. When failed, these tests provide the strongest evidence of malingering. The technique is also adaptable to a wide variety of possibly malingered conditions, e.g., blindness, tactual imperception, and deafness. Unfortunately, very few malingers fail forced choice tests, presumably recognizing that a higher level of performance would be the norm even for a severely impaired individual.

An example of a widely used test that combines both forced choice and floor effect features is the Test of Memory Malingering (TOMM). This test is composed of 50 pictures, which are shown to the patient in repeated trials. The patient is tested by being given a dichotomous choice between each correct picture and a foil. The patient's absolute score yields a classification of likelihood of malingering. If the patient goes on to score below chance levels, a precise probability rating can be assigned demonstrating the certainty level of malingering beyond any reasonable question.

Other cognitive malingering tests utilize known principles of learning or test-taking to identify unusual patterns of test performance. For example, even individuals with severe memory impairment tend to respond with significantly more correct

answers when they are given recognition choices. Another principle is that, on average, the patient should score better on easier items than on more difficult ones, and that once a patient's level of ability has been exceeded, further answers should have no greater accuracy than chance. These and other atypical performance patterns offer the neuropsychologist further tools to identify malingering examinees.

The Minnesota Multiphasic Personality Inventory-2 (MMPI-2) is a widely used self-report personality inventory that contains a number of validity scales useful in identifying the individual malingering neuropsychological problems. These scales include the Infrequency Scale or *F* scale, the Infrequency–Psychopathology Scale or $F(p)$ scale, the Infrequency Back Scale or *Fb* scale, the *F* minus *K* Index, the sum of obvious – subtle differences, the Dissimulation-Revised Scale, and the Fake Bad Scale. These scales, and combinations thereof, have been shown to successfully identify individuals feigning psychiatric and neuropsychological dysfunction. The latter scale, the Fake Bad Scale, appears to have especial utility among persons alleging acquired physical debility and/or neurocognitive dysfunction.

Thus, the forensic neuropsychological assessment is not complete without an adequate assessment for malingering behavior. The assessment should include psychometric techniques for identifying underperformance and symptom exaggeration, as well as careful analysis of the examinee's presentation for indications of inconsistency. Such indications include notable discrepancies between test performances and observed abilities, discrepancies between claimed deficits and observed or reported functional activities, discrepancies between self-reported symptoms and known patterns of brain functioning, and discrepancies between self-reported deficiencies and symptoms reported by reliable collateral informants.

In sum, neuropsychological assessment, the application of standardized cognitive measurement techniques to brain–behavior relationships, has proved a useful and often determinative source of information in individuals with neurodevelopmental disorders and acquired brain dysfunction. Its validity and usefulness has been recognized by other disciplines and by the legal system. Neuropsychological assessment is a valuable resource, particularly in situations involving forensic questions about an individual's mental abilities, the nature/degree of neurocognitive injury, and one's neurobehavioral prognosis.

References

[1] American Psychological Association (1989). Division 40: definition of a clinical neuropsychologist, *The Clinical Neuropsychologist* **3**.

[2] National Academy of Neuropsychology (2001). *Definition of a Clinical Neuropsychologist: Official Position of the National Academy of Neuropsychology*. http://www.nanonline.org/PostitionPageLinks/Pages/DefinitionofaNeuropsychologist.aspx.

[3] American Academy of Neurology (1996). Assessment: neuropsychological testing of adults: considerations for neurologists, *Neurology* **47**, 592–599.

[4] Strauss, E., Sherman, E.M.S. & Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*, 3rd Edition, Oxford University Press.

[5] Reitan, R. & Wolfson, D. (1993). *The Halstead-Reitan Neuropsychological Test Battery: Theory and Clinical Application*, Neuropsychology Press.

[6] Heaton, R.K., Miller, S.W., Taylor, M.J. & Grant, I. (2004). *Revised Comprehensive Norms for an Expanded Halstead-Reitan Battery: Demographically Adjusted Neuropsychological Norms for African American and Caucasian Adults*, Psychological Assessment Resources.

Further Reading

American Academy of Clinical Neuropsychology (2007). American Academy of Clinical Neuropsychology (AACN) practice guidelines for neuropsychological assessment and consultation, *The Clinical Neuropsychologist* **21**, 209–231.

National Academy of Neuropsychology (2003). *Independent and Court-ordered Forensic Neuropsychological Examinations: Official Statement of the National Academy of Neuropsychology*. http://www.nanonline.org/PostitionPageLinks/Pages/Independent.aspx.)

JAMES S. WALKER

Neuropsychological Assessment: Child

Introduction

Child neuropsychological (NP) assessment involves the application of psychometric test methods in the evaluation of brain abnormalities in children. It shares

many of the same goals and principles involved in the NP assessment of adults (*see* **Neuropsychological Assessment**). However, there are many important differences. Developmentalists argue that a child should never be viewed as simply a scaled-down version of an adult. Likewise, an appropriate assessment of brain–behavior relationships in children cannot be based simply on scaled-down versions of assessment methods designed for use with adults. Not only do children differ in terms of the types of brain insult commonly experienced, they also differ with respect to the specificity of behavioral effects manifested, the pattern and course of (re)acquisition after injury, the modifying effects of ongoing developmental change, and the extent to which deficits sometimes can be "silent" until later developmental periods.

There have been a number of books devoted exclusively to the methods and unique issues involved in the NP assessment of children (e.g., [1, 2]). The focus of this article is admittedly selective and begins with a discussion of what *is* NP assessment and how it differs from other related diagnostic methods. Next, there will be an overview of different clinical applications and major test methods and approaches currently available. That will set the stage for a discussion of some key conceptual and practical issues in this area.

Contrast to Other Assessments

What defines a NP assessment and distinguishes it from simply an assessment of a child's mental abilities? In other words, how does it differ from a general psychological assessment?

Many of the standard instruments available for assessing children's abilities, such as the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV; [3]), can be used quite effectively in evaluating a child's NP functioning. This is especially true for younger children (up to about 5 or 6 years) for whom decrements in general intelligence are often among the chief manifestations of early childhood brain injury [4]. *Sensitivity to brain dysfunction* is certainly a necessary qualification for a test to be considered a NP measure. This is an empirical issue, based on whether and to what extent performance on the test allows a valid differentiation of normal children and those suffering from one or another type of brain abnormality.

However, most tests used in a general psychological assessment were never developed from a NP perspective, nor were they designed to facilitate various types of inferences regarding a child's NP functioning. For example, different patterns of performance on the WISC-IV may be suggestive of left versus right hemispheric brain dysfunction, but the test was not constructed in a way that readily lends itself to that interpretation. Similarly, although qualitative observations during test performance on the WISC-IV may offer insights regarding a child's executive functioning, it was not designed to isolate this important area of NP functioning. Although virtually any measure of a child's abilities can be used in making NP inferences, some measures are better than others in revealing specific aspects of NP functioning. Thus, another essential feature of NP assessment is that it *facilitates the conceptualization of a child's performance in terms of brain–behavior relationships*. Sensitivity to brain damage is not enough. A test must also be constructed and validated in a manner that permits inferences regarding specific areas of NP functioning.

The NP assessment also should be based on a sufficiently broad appraisal of a child's functioning. The interest is not simply on the child's language abilities, memory, or inhibitory controls – taken separately – but instead on the child's overall performance as reflected in a *functional profile.* This is why a proper assessment typically involves a battery of tests to assure sufficient breadth and depth of coverage. This will be discussed further later on in the section dealing with current approaches to assessment.

To recap, a NP assessment of a child produces a *functional profile derived from a battery of tests sensitive to brain dysfunction and organized in a manner that facilitates inferences regarding brain–behavior relationships.*

What about distinguishing a child NP assessment from other diagnostic methods evaluating neurological abnormalities in children? Specifically, how is it different from a pediatric neurological exam (PNE)? The precise features vary across practitioners, but a PNE typically consists of a review of the child's development, clinical history and presenting complaints as well as a direct examination of areas such as station and gait, basic motor and sensory-perceptual functions, cranial nerves, and reflex integrity. The findings of the exam often are used to judge whether other neurodiagnostic methods should be pursued.

It also includes at least a brief mental status exam (assessing such things as comprehension, naming skills, immediate recall, and so forth) although this is typically rather basic and informal. The child NP assessment, by contrast, is designed to provide a far more in-depth and standardized evaluation of mental processes. Accordingly, it would be more sensitive to neurological conditions having an impact on higher mental functioning. Indeed, this is why suspicious findings on the PNE often may prompt the physician to refer the child for a formal NP assessment. An important distinguishing feature also has to do with the *standardized* nature of the NP assessment – even as it pertains to the evaluation of areas of lower functioning such as motor and sensory-perceptual abilities. In a standardized test or procedure, the components of the exam are administered and scored in a strictly defined manner and evaluated according to specific norms. This permits the measurement of more subtle variations – as, for example, mild decrements in fine-motor speed or control. With this, the NP assessment sometimes may be sensitive to milder or earlier functional changes accompanying a disease process.

The NP assessment can be further differentiated from other neurodiagnostic methods such as brain imaging. Whereas computed tomography (CT) or magnetic resonance imaging (MRI) may identify and localize an underlying brain injury, the NP assessment would instead be used to help specify the particular functional deficit(s) associated with it. Each set of methods occupy opposite ends of a *structural–functional assessment continuum*, with tools such as CT and standard MRI focusing on brain *structure* whereas the NP assessment deals with brain *function*. An abnormal MRI finding may lead to an expectation of a particular functional impairment (as, for example, indications of an injury involving the left temporal lobe suggesting that the child's verbal comprehension and memory may be affected), although this is only an educated inference needing confirmation by actual functional assessment. The reverse is also true – namely, that a particular pattern of performance on NP testing may suggest a particular type or localization of a brain injury, although that too is only an inference based on current knowledge of brain–behavior relationships. In some cases, the findings may diverge, as when abnormalities on NP assessment are not accompanied by abnormalities on brain imaging and *vice versa*. This may occur in cases with milder injuries, or where the NP finding may relate to other organic variations (such as abnormal blood flow or metabolism) rather than to structural deviations, *per se* [5]. A comprehensive evaluation of the brain injured child often requires an integrated and multimethod approach incorporating both structural and functional assessments.

Applications

Child NP assessment may be used in a broad range of clinical applications. These include: (i) aiding in the detection of brain dysfunction, (ii) providing a specification of the neurobehavioral effects of a known brain injury, (iii) helping to identify specific underlying dimensions of dysfunction in particular handicaps, (iv) using assessment data to help formulate effective treatment strategies, (v) helping to assess the child's prognosis and risk for certain developmental outcomes, and (vi) conducting follow-up assessments of functional change over the course of development and in response to particular interventions [6]. The relative importance of these different applications depends, in part, on the particular clinical population under consideration. Broadly speaking, there are four different but overlapping clinical populations to whom NP assessment may be applied. These include children with: neurological disorders, systemic illness, psychiatric disorders, and learning disabilities.

Children with *neurological disorders* comprise a very broad and varied category. The possible conditions are many, and include children with genetic disorders affecting the brain (e.g., Turner's syndrome), structural abnormalities (e.g., agenesis of the corpus callosum, hydrocephalus), traumatic injuries, epilepsy, and a variety of neuropathological processes such as anoxic episodes, encephalitis, toxic encephalopathy (e.g., carbon monoxide or lead poisoning), metabolic disorders, demyelinating diseases, neuromuscular disorders, brain tumors, and, more rarely in children, cerebral vascular accidents. The NP assessment may serve to identify functional impairments which, taken together with other appropriate evaluations, may lead to the diagnosis of a particular neurological condition. For example, NP findings pointing to a decline in a child's cognitive functioning may prompt a broader neurological work-up that ultimately leads to the diagnosis

of a degenerative disorder, such as metachromatic leukodystrophy – a progressive demyelinating disease of the brain. Or, the presence of localizing signs on NP assessment may prompt a more thorough exploration of a specific seizure focus in a child with epilepsy. This could have implications for possible neurosurgical intervention in the child with intractable seizures. However, there is a very important point to emphasize here. In neither of these examples is the NP assessment *diagnostic* of the condition in question; rather the findings may be part of a broader series of evaluations from which a diagnosis is ultimately made. A particular set or pattern of results on NP assessment may correlate with, but, taken by themselves, should never be considered to be diagnostic of these or other neurological conditions.

Besides the issue of detection, an NP assessment may be used to obtain a detailed appraisal of *what* aspects of mental or behavioral functioning have been affected by a known or suspected neurological condition. This was explained before in contrasting what is unique about NP assessment in comparison to other methods used in evaluating the brain-disordered child. The NP assessment is needed to determine what aspects of functioning have been *impaired or spared* as a result of neurological conditions such as epilepsy, congenital abnormalities, toxic states, and so forth.

An area of intensive NP investigation, and one of particular interest within a forensic context, has to do with traumatic brain injury (TBI) in children. The interested reader may wish to refer to one of the published works devoted exclusively to the topic (e.g., [7]). Briefly, some of the key issues in this area have to do with conducting a thorough assessment that is sensitive to the affected areas of functioning, establishing a preinjury baseline (which is especially problematic in younger children) against which current deficits may be compared, and determining not only recovery but possible longer-range developmental impact – especially in terms of delayed effects or deficits that may be "silent" until challenged more as the child grows older. These are among the issues that contribute to the evaluation of TBI and other acquired conditions often being more complicated in children than in adults (more on this later).

Another broad population or category for general application has to do with children having *systemic illness*. This category consists of children having one or more pediatric diseases or conditions, not primarily neurological in nature, which can have a *potentially* adverse impact on central nervous system (CNS) functioning. Examples would include neonatal complications associated with prematurity or very low birth weight, defects in specific organ systems (congenital heart disease, pulmonary or renal dysfunction), infections, metabolic disorders, autoimmune disorders, and cancer. Here too the interested reader is referred to *volumes* dealing thoroughly with the topic (e.g., [8, 9]). What is import to appreciate is that the NP assessment in this general category has become a growing area of need. Advances in medical care over the past several decades have brought about a dramatic increase in the survival of children with serious illnesses. However, with the decrease in mortality has come a corresponding increase in morbidity – as the conditions (or their treatments) sometimes will have an adverse impact on the developing brain and the child's later functioning. The effects can be subtle, and manifested sometimes in the form of low-grade problems with attention or learning. There clearly is a need for careful follow-up in which the NP assessment can play an important role.

Children with *psychiatric disorders* make up another group for whom the NP assessment can be important. This may be viewed from two perspectives. First, children with brain dysfunction are at higher risk for the development (or exacerbation) of a psychiatric disorder. In some cases the relationship may be direct, as when a frontal lobe injury may give rise to impulsive behavior problems. In other cases the relationship is more indirect, whereby brain dysfunction may set the stage for other factors to come into play which, themselves, act to produce or aggravate an emotional or behavioral disturbance [10]. For example, an underlying brain injury may give rise to one or more learning disabilities which may render the child more likely to encounter frustration or failure in school. This, in turn, may result in any of a variety secondary problems ranging from anxiety, depression or social withdrawal to defiance and other acting-out behaviors. Early identification of the functional deficits and behavioral risks in the brain-impaired child can be an important step in limiting the development or progression of a psychiatric disorder.

The other perspective with respect to psychiatric disorders has to do with examining the disorders, themselves, in terms of possible underlying neurobiological and NP abnormalities. Whether or not there is a history of known brain injury or neurological disorder in the case of an individual

child, there is rapidly growing evidence that disorders such as autism, attention-deficit/hyperactivity disorder (ADHD), childhood schizophrenia and other psychotic disorders, as well as certain forms of depression and anxiety disorders are each associated with abnormalities in underlying neural mechanisms and NP functioning. For many children with ADHD, for example, disordered frontal-striatal functioning appears to play a key role in the deficits in behavioral inhibition and task persistence commonly seen. Such a perspective has an important bearing on how the condition is approached both in assessment and treatment planning. The interested reader is referred *Tramontana* and associates [11] for a more complete discussion dealing with the neuropsychology of child psychopathology.

Children with *learning disabilities* comprise yet another important population or category for application of NP assessment. Here the interest is not so much on identifying whether or not the child is brain impaired. Nor is there necessarily a question that the child is experiencing a significant handicap in academic learning – whether it has to do with reading, math, writing, and so forth. Indeed, the assessment often may begin with there already being ample documentation that the child is struggling in one or more of these areas. Rather, the emphasis is on using the NP assessment to better understand how and why the learning process may be breaking down. For example, reading can be impeded by any of a variety of underlying problems, including deficits in phonological processing, visual pattern recognition, phoneme-grapheme association, segmentation and sequencing, grammatical awareness, as well as verbal comprehension and memory. The breadth of coverage typically involved in the NP assessment may serve to highlight deficits in relevant underlying component skills. Also, a NP perspective can integrate isolated findings into coherent patterns relevant to brain-based models of learning disabilities. This, in turn, can help to distinguish different patterns or *subtypes* of learning disability having different implications for prognosis and intervention. Here too there are volumes dealing specifically with this topic to which the interested reader may wish to refer (e.g., [12]).

Methods and Approaches

There are a variety of different approaches to child NP assessment. A *fixed-battery* approach is one that aims to provide a comprehensive assessment of brain function using an invariant set of validated test procedures. The composition of the battery is not tailored to the presenting characteristics of the individual child or to the specific clinical questions to be addressed. Rather, the emphasis is on administering a well-defined set of designated tests. It is assumed that individual variability is captured reasonably well as long as the battery is constructed in a fashion that covers a broad range of functioning. An important advantage of this approach is that it provides a standard data base on which different clinical groups can be compared.

Probably the most commonly known example of a fixed-battery approach is the Halstead–Reitan neuropsychological battery (HRNB; see [13]). There are two different versions of the battery for children depending on the age of the child (5–8 years and 9–14). These methods, along with supplemental tests commonly used, were the dominant approach to child NP assessment for decades. Golden [14] introduced a children's revision of a then newly developed NP battery for adults, the Luria-Nebraska Neuropsychological Battery, extending it downward for children ranging from 8 to 12 years of age Luria–Nebraska Neuropsychological Battery-Children's Revision (LNNB-CR). A detailed description and review of these batteries is beyond the scope of this article. These can be found elsewhere [6, 15].

One of the chief limitations of both batteries is that neither was originally developed with children in mind. Areas such as language processing or memory and new learning – areas highly pertinent in the evaluation of children – receive little coverage in the basic HRNB. The same can be said with respect to the area of executive functioning and the LNNB-CR. Neither battery provides an assessment of a key domain such as attention. Thus, more than simply distinguishing brain damaged and normal children (which each battery can do about equally well), an important goal of NP assessment should be to describe how a child is functioning in key domains such as these (more on this in a moment).

At the other extreme are various *flexible approaches* to NP assessment. In *qualitative approaches,* the examiner is less concerned with using methods to quantify the extent of deficits. Instead the emphasis is on determining *how* an individual passes or fails a particular task. Here, for example, the examiner would follow-up with informal inquiries or tasks

that may clarify the faulty retrieval strategies that the child used in failing to recall an item correctly. On the basis of this, the examiner may then branch out and select other tasks to gain an evolving picture of what is wrong. *Process-oriented approaches* represent a hybrid of quantitative and qualitative methods. An example is the *Boston Process Approach* [16] that draws selectively from a core set of standardized tests for a quantified overview of general functioning. Depending on the initial results, and guided by clinical hypothesis-testing, the examiner uses various "satellite" tests, including improvised procedures, in order to pinpoint the precise nature of the individual's deficits. Thus, although there are shared components, much of the assessment varies from patient to patient. Another hybrid approach is reflected in the NEPSY-II, (from NEuroPSYological Investigation for Children), which combines a core assessment along with a flexible assessment of selective areas of interest [17]. All of the components of the exam, including those dealing with selective processes, are on the basis of psychometric test scores.

Most neuropsychologists are probably not purists with respect to one approach or set of methods. Whether implicitly or explicitly, many tend to adopt an *eclectic approach* to the NP assessment. Any of a variety of available tests may be selected to quantify the extent of deficit in various areas of functioning. There usually is at least an implicit outline of the relevant functions and abilities to be assessed routinely. The psychometric properties of the tests (adequacy of norms, validated discrimination of brain dysfunction), as well as how well they complement each other in a test battery, are important factors guiding test selection. Qualitative analysis and other flexible approaches may be used to add richness and individualization to the examination, although most neuropsychologists would view these as supplemental rather than primary aspects of the exam.

Table 1 provides an outline of *key domains* for child NP assessment. It is based both on robust factors derived from statistical studies of fixed test batteries as well as a consideration, on conceptual grounds, of key areas that a comprehensive assessment of a child ought to include [6]. In general, the evaluation should include specific measures spanning the following areas: motor functioning, sensory-perceptual abilities, language and spatial processes, attention, memory and new learning, and executive functioning. Also, whether performed by the neuropsychologist or made available from another source, the evaluation should incorporate findings from tests of general intelligence, academic achievement, as well as measures of behavioral adjustment and personality functioning.

Table 1 Key domains in child neuropsychological assessment[(a)]

Motor control
Sensory-perceptual abilities
Language
Spatial organization
Attention
Memory
Executive functions

[(a)] A comprehensive evaluation also would include an assessment of intelligence, academic achievement, and behavioral adjustment

The foregoing has to do with the breadth or scope of the NP assessment. It defines a *horizontal analysis* that is fixed or invariant with respect to the key domains to be assessed. The particular tests used would depend, in part, on the age of the child. They would also depend on the extent to which there are concerns regarding specific areas or subcomponents of functioning on the basis of the history or referral questions concerning the child. Thus, within each domain, there would be a more in-depth or *vertical analysis* of specific areas for which there would be a flexible selection of appropriately normed and validated tests. Table 2 illustrates such an assessment framework. As an example, the domain of language is broken down into subcomponents from which one or more areas might be assessed. The more relevant the domain is in the case of the individual child or population studied, the more thoroughly it would be assessed.

Lastly, Table 3 provides examples of specific tests corresponding to each of the key assessment domains noted above. The listing is only a limited sampling of the possible tests available. There is a helpful volume providing a compendium of these and other NP tests by Strauss *et al.* [18].

Conceptual and Practical Issues

Applying a developmental perspective in child NP assessment means more than simply assuring that the tests and measures are appropriate for the age of the

Table 2 Neuropsychological assessment framework

D1	D2	Language	D4	D5	...	Dx
SD1	SD1	Phonology	SD1	SD1		SD1
SD2	SD2	Repetition	SD2	SD2		SD2
SD3	SD3	Semantics	SD3	SD3		SD3
.	.	Comprehension	.	.		.
.	.	Naming	.	.		.
.	.	Formulation	.	.		.
SDy	SDy	Pragmatics	SDy	SDy		SDxy

D, domain; SD, subdomain

Table 3 Examples of test procedures

Domain	Function assessed
Motor control	
Finger Oscillation Test	Fine-motor speed
Grooved Pegboard	Manual dexterity
Sensory-perceptual	
Tactile Finger Recognition	Simple tactile discrimination
Fingertip Number Writing	Complex tactile perception
Seashore Rhythm Test	Auditory pattern recognition
Picture Completion	Visual closure
Language	
Peabody Picture Vocabulary Test	Word comprehension
Expressive Vocabulary Test	Confrontation naming
Token Test	Listening comprehension
Verbal Fluency	Controlled word association
Spatial	
Embedded Figures Test	Visual-spatial organization
Developmental Test of Visual-motor Integration	Visuoconstructive ability
Tactual Performance Test	Tactile-spatial organization/learning
Attention	
Continuous Performance Test	Vigilance and sustained attention
Trail Making Test	Rapid focusing/sequencing
Memory	
Digit Span	Verbal working memory
Benton Visual Retention Test	Visual reproduction
California Verbal Learning Test	List learning
Wide Range Assessment of Memory and Learning	General memory
Executive functioning	
Category Test	Conceptual reasoning
Wisconsin Card Sorting Test	Flexible problem-solving

child being assessed. It involves understanding how brain functions develop normally as well as under various pathological conditions. A child not only grows more competent with age, but also develops new and more efficient ways of completing tasks and solving problems. An injury or other brain abnormality seldom results in the obliteration of a function or ability. Rather, there often is some recovery of function or compensatory development, although the net effect may be that the function is executed less efficiently than would be true for a normal child of the same age and background. A

proper evaluation of a child's NP assessment must be guided by an understanding of concepts such as these.

There are many factors that can obscure or complicate the assessment of brain–behavior relationships in children. These include: (i) problems specifying the precise time of onset for various forms of brain pathology, (ii) the absence of a referential baseline of premorbid functioning in cases of early brain damage, (iii) the sometimes blurred distinction between neurodevelopmental anomalies and normal variations in the rate or pattern of acquisition of function, (iv) the extent to which deficits can sometimes be delayed or "silent" until later developmental periods, (v) the interacting effects of various nonneurological attributes within the child that may serve to compound or mitigate the effects of brain dysfunction, and (vi) the impact of environmental factors, including the family, in shaping the child's outcomes [6]. Although factors similar to these can play a role in the evaluation of adults, they generally are more prominent and complex in the case of the child.

There are practical issues as well. It is well known that an important challenge in working with children in this area has to do with maintaining behavioral compliance and sustained attention for what can be a fairly lengthy assessment. Also, the younger child may be unable to provide a reliable report of deficits and the circumstances in which they are problematic. Reports of parents and teachers often are crucial in providing accounts of the child's abilities outside of the testing situation. On the other hand, feigned impairments on the NP assessment are fairly uncommon in younger children, although distortions of one kind or another may invalidate the history and presenting complaints reported by parents. A careful appraisal of possible parental bias or embellishment in their report of the child's symptoms must be an important part of the overall assessment, especially in cases where litigation is involved.

Age-related constraints on the NP assessment are especially problematic in the evaluation of children during the first 3 or 4 years of life. The limited response repertoire of infants and toddlers, together with the degree to which their performance is state-dependent, place major constraints on formal NP assessment. This is why the bulk of practice in child neuropsychology has dealt mainly with children 5 years of age and older. (See [19], for a thorough discussion of the NP assessment in early childhood.) In a limited sense, however, the careful application of a NP perspective in the evaluation of standard developmental testing of very young children **is** a NP assessment – as long as it is embedded within a conceptual framework of developmental brain–behavior relationships as discussed above. Areas of potential developmental risk can be identified, on the basis of both known/suspected CNS insults coupled with direct observations of the child, which can then be highlighted for further, more formal NP assessment once the child reaches about 5 years of age.

Finally, there are unique challenges that arise in the evaluation of special populations. Included here are children with handicapping conditions such as sensory loss (blind and hearing impaired), physical deformities, motor disabilities, and language impairment. It is important to appreciate the impact of the handicap on the child's general functioning and adjustment. At the same time, special care must be taken to assure that the handicap does not overshadow and distort the picture of the child's functioning in nonimpaired areas. This is an area that calls upon specific skills, experience, and ingenuity on the part of the examiner. The same is true for working effectively with non-English speaking children and those with other cultural differences.

The foregoing provided a general overview of child NP assessment. Hopefully, with it, the reader has gained an understanding of how to conceptualize and contrast this type of assessment in comparison to other related evaluation methods for children. Major applications were discussed and there was a delineation of the various approaches and methods currently available in the field. Lastly, various conceptual and practical issues were highlighted – including the critical importance of a development framework in conducting and evaluating the child NP assessment, as well as the special challenges that present when evaluating various subgroups of children. The coverage was selective, but nonetheless should have given a general idea of the defining characteristics and current status of this field.

References

[1] Baron, I.S. (2004). *Neuropsychological Evaluation of the Child*, Oxford University Press, New York.

[2] Tramontana, M.G. & Hooper, S.R. (1988a). *Assessment Issues in Child Neuropsychology*, Plenum Press, New York.
[3] Wechsler, D. (2003). *Wechsler Intelligence Scale for Children – (WISC-IV)*, 4th Edition Psychological Corporation, Harcourt Assessment, San Antonio.
[4] Boll, T.J. & Barth, J.T. (1981). Neuropsychology of brain damage in children, in *Handbook of Clinical Neuropsychology*, S.B. Filskov & T.J. Boll, eds, John & Sons Wiley, New York, pp. 418–452.
[5] Hillary, F.G. & DeLuca, J. (2007). *Functional Neuroimaging in Clinical Populations*, Guilford Press, New York.
[6] Tramontana, M.G. & Hooper, S.R. (1988b). Child neuropsychological assessment: overview and current status, in *Assessment Issues in Child Neuropsychology*, M.G. Tramontana & S.R. Hooper, eds, Plenum Press, New York, pp. 3–38.
[7] Granacher, R.P. (2003). *Traumatic Brain Injury: Methods for Clinical and Forensic Neuropsychiatric Assessment*, CRC Press, New York.
[8] Hynd, G.W. & Willis, W.G. (1988). *Pediatric Neuropsychology*, Grune and Stratton, New York.
[9] Yeates, K.O., Ris, M.D. & Taylor, H.G. (2000). *Pediatric Neuropsychology: Research, Theory, and Practice*, Guilford Press, New York.
[10] Tramontana, M.G. (1983). Neuropsychological evaluation of children and adolescents with psychopathological disorders, in *Foundations of Clinical Neuropsychology*, C.J. Golden & P.J. Vincente, eds, Plenum Press, New York, pp. 309–340.
[11] Tramontana, M.G., Hooper, S.R., Watts-English, T., Ellison, T. & Bethea, T.C. (2008). Neuropsychology of child psychopathology, in *Handbook of Clinical Child Neuropsychology*, C.R. Reynolds & E. Fletcher-Janzen, eds, 3rd Edition, Springer Press, New York.
[12] Fletcher, J.M., Lyon, G.R., Fuchs, L.S. & Barnes, M.A. (2007). *Learning Disabilities: From Identification to Intervention*, Guilford Press, New York.
[13] Reitan, R.M. & Davison, L.A. (1974). *Clinical Neuropsychology: Current Status and Applications*, John Wiley & Sons, New York.
[14] Golden, C.J. (1981). The Luria-Nebraska Children's battery: theory and formulation, in *Neuropsychological Assessment and the School-age Child: Issues and Perspectives*, G.W. Hynd & J.E. Obrzut, eds, Grune and Stratton, New York, pp. 277–302.
[15] Reynolds, C.R. & Fletcher-Janzen, E. (eds) (2008). *Handbook of Clinical Child Neuropsychology*, 3rd Edition, Plenum Press, New York.
[16] Milberg, W.P., Hebben, N. & Kaplan, E. (1986). The Boston process approach to neuropsychological assessment, in *Neuropsychological Assessment of Neuropsychiatric Disorders*, I. Grant & K.M. Adams, eds, Oxford University Press, New York, pp. 65–86.
[17] Korkman, M., Kirk, U. & Kemp, S. (2007). *NEPSY*, 2nd Edition Psychological Corporation, Harcourt Assessment, San Antonio.
[18] Strauss, E., Sherman, E.M.S. & Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*, 3rd Edition, Oxford University Press, New York.
[19] Aylward, G.P. (1988). Infant and early childhood assessment, in *Assessment Issues in Child Neuropsychology*, M.G. Tramontana & S.R. Hooper, eds, Plenum Press, New York, pp. 225–248.

MICHAEL G. TRAMONTANA

Neuropsychological Assessment: Head Injury *see* Head Injury: Neuropsychological Assessment

Northwest Juvenile Project

The Northwestern Juvenile Project (NJP) is the first large-scale, prospective longitudinal study of alcohol, drug, and mental (ADM) disorders in juvenile detainees. The NJP is funded by a consortium of nine federal agencies and five private foundations. The sample includes a diverse sample of 1829 youth, aged 10–18 years at baseline, who were arrested and detained between 1995 and 1998 in Cook County (Chicago metropolitan area), Illinois. Initially funded to examine the ADM disorder service needs and service use of juvenile detainees, the aims of the NJP have expanded to include an examination of (i) changes in ADM disorders and the comorbidity of disorders over time; (ii) patterns of mental health service use over time; and (iii) pathways and patterns of drug use, violence, and risk behaviors for human immunodeficiency virus (HIV)/acquired immune deficiency syndrome (AIDS) and other sexually transmitted infections (STIs) over time. The NJP tracks and reinterviews these youth wherever they are living, whether they are back in their communities or incarcerated. The NJP also obtains records cross-validating self-reported data on criminal justice

involvement, and mental health and substance use service utilization from 16 correctional and service agencies.

Background to the Northwestern Juvenile Project

Reports issued by the Surgeon General [1] and the President's New Freedom Commission on Mental Health [2, 3] note that the mental health needs of juvenile detainees are largely underserved. Under the Eighth Amendment (barring cruel and unusual punishment) and the Fourteenth Amendment (right to substantive due process for youths in the juvenile justice system) of the US Constitution, juvenile detainees with serious mental disorders have a right to receive needed treatment as part of the state's obligation to provide needed medical care. Providing appropriate services, however, requires accurate and reliable data on the prevalence of mental disorders among juvenile detainees. Data are necessary for planning how to best utilize the finite resources of the juvenile justice and community mental health systems to meet the mental health needs of detained youth.

Despite the importance of epidemiological data on the mental health and substance abuse needs of juvenile detainees, the juvenile justice system has not had comprehensive, accurate, and reliable data upon which to guide policy decisions. Although a number of studies have examined the mental health and substance abuse needs of detained youth, estimates of rates across studies are discrepant. For example, estimates of the prevalence of affective disorder vary from 5% [4] to 72%; [5] substance use disorders vary from 20% [6] to 88%; [5] and psychosis varies from 16% [5] to 45% [6].

The variability in rates may result from discrepancies in the methods of prior studies in the following areas:

Sample Composition

The composition of samples substantially varied across prior studies. The racial and ethnic compositions of the samples in many of these studies are not representative of the national juvenile justice population; some studies did not even report the racial or ethnic composition of the sample. Females, an increasing proportion of juvenile detainees, were excluded entirely from some investigations.

Small Size

Because of the difficulties associated with the study of detained youth, many studies had small samples. Small samples make it difficult for studies to generate reliable rates, especially for more severe mental disorders with low base rates in the general population (i.e., 1–4%). Many of the studies sampled too few subjects to generate reliable rates even for the more common mental disorders [7]. Most studies did not have enough participants in key demographic subgroups to compare participants by gender, race and ethnicity, or age.

Measurement

Some studies relied on nonstandard or untested instruments; others failed to consider impaired functioning or reported data on only one category of disorder (e.g., substance use disorders, anxiety disorders, and personality disorders) rather than a spectrum of disorders. Even when multiple disorders were assessed, most of the studies did not examine patterns of comorbidity between mental and substance use disorders.

The Northwestern Juvenile Project

The NJP was designed to overcome the methodological limitations of prior studies. The NJP has three interrelated goals:

1. To assess how ADM disorders develop over time among detained youth.

The NJP assesses persistence and change in mental and substance use disorders (including remission and recurrence) patterns of comorbid disorders, and associated functional impairments and outcomes during critical points of development: adolescence, emerging adulthood, and young adulthood.

2. To investigate barriers, pathways, and patterns of service use.

The NJP assesses if and when youth who *need* services *receive* them and from which sectors: corrections, child welfare, education, general health, mental health, and informal services. The NJP can also evaluate how patterns of service use are associated with longitudinal outcomes.

3. To determine pathways and patterns of drug use, violence, and risk behaviors for and prevalence of HIV and other STIs.
The NJP examines the development of these risk behaviors among our subjects, focusing on gender differences, racial/ethnic differences, the antecedents of these risky behaviors, and how these behaviors are interrelated. The NJP will also provide the first estimate of the prevalence and incidence of HIV and other STIs in this population.

Methods

The sample, longitudinal design, and assessments used in the NJP address many of the limitations of prior studies of detained youth.

Sample

Participants in the NJP were a randomly selected sample of male and female youth who were arrested and subsequently detained at the Cook County Juvenile Temporary Detention Center (CCJTDC) between November 20, 1995 and June 14, 1998. As part of the study design, selected strata were oversampled to obtain enough data within key subgroups. The sample was stratified based on the following variables: age (10–13 years or $\geq$ 14 years), gender, race/ethnicity (African-American, non-Hispanic white, and Hispanic), and legal status (processed as a juvenile or an adult). Sample weights were used in statistical analyses so that the findings reflect CCJTDC's population rather than the stratified sample. The sample provides several advantages over prior studies, including:

1. Large sample size.
The final sample includes 1829 participants, a sample size large enough to allow for reliable estimates of uncommon disorders, comorbidity of disorders, and high-risk behaviors.

2. A large subsample of females.
Because females were oversampled, the NJP has a large enough subsample of females (657 females; 35.9% of the sample) to examine differences by gender, as well as differences among females on other key variables. Studying females is critical because females make up increasing proportions of juvenile arrestees (24%), juvenile detainees (19%), adult arrestees (22%), and adult detainees (21%) [8–11].

3. Racial/ethnic diversity.
The stratification of the sample also provides the NJP with adequate participants to examine racial/ethnic diversity within the sample, including 1005 African-Americans (54.9%), 524 Hispanics (28.7%), 296 non-Hispanic whites (16.2%), and four from other racial/ethnic groups (0.2%).

4. Wide age range.
The age range of the sample is 10–18 years old (mean, 14.9 years) at baseline. Youth aged 10–13 were oversampled to provide adequate numbers to examine age differences.

5. Diversity of criminal behavior.
Because the NJP sampled participants from *all* newly arrested youth entering detention, it represents a variety of youngsters who are sent to detention centers. In contrast, other studies focus on specific subpopulations of juvenile detainees, such as serious and violent offenders (i.e., those adjudicated (convicted) of a felony, misdemeanor weapons offense, or misdemeanor sexual assault).

6. Youth processed as adults.
The sample of the NJP includes youth that are automatically transferred to the adult system for processing. These youth are excluded from the jurisdiction of juvenile court based on their type of offense, criminal history, and/or age.

Longitudinal Design

Although cross-sectional studies can be useful to ascertain basic epidemiologic data, a developmental epidemiologic approach requires prospective longitudinal studies [12]. Longitudinal studies allows researchers to (i) examine the antecedents of disorders and other behavioral problems within a temporal context, thereby providing the basis for understanding causal mechanisms [13]; (ii) reveal pathways, changes *within* individuals, variation *among* individuals, and, most important, *sequences* in the development of disorders and other problem behaviors; and (iii) examine age-dependent change, independent of disorder [12, 13].

The NJP is currently funded to collect 9 waves of data spanning 14 years. The longitudinal design

allows for the examination of detailed data on mental disorders, substance use patterns, criminal behavior (including violence), use of mental health and other health services, and HIV/AIDS risk behaviors across three critical developmental periods: adolescence (ages 10–18 years), emerging adulthood (ages 18–25 years), and young adulthood (ages 25–30 years). All participants are tracked until located, and subsequently interviewed wherever they are found. Records cross-validating self-reported data (e.g., on arrests, incarceration history, health, and service use) are also obtained from 16 correctional and community service agencies.

Measurement

The NJP employs standardized diagnostic instruments that are sensitive to the varying developmental stage of our participants. The Diagnostic Interview Schedule (DISC) version 2.3 (based on revised third edition of the Diagnostic and Statistical Manual for Mental Disorders (DSM)) [14, 15], the most recent English and Spanish version then available, was used for the baseline assessments. Diagnostic assessments were changed in accordance with changes in participants' age and improvements in diagnostic technology. For subsequent interviews, the NJP administered version 4.0 of the DISC (based on *DSM-IV*), modified by its authors for use with young adults. The Diagnostic Interview Schedule, version IV (DIS-IV) (based on *DSM-IV*) was used to assess disorders not assessed, or not adequately assessed by the DISC 4.0, including substance use disorders, schizophrenia, cognitive impairment, and antisocial personality disorder (APD) [16]. Most of the samples aged 18 years or older in 2002, at which time the NJP stopped using diagnostic tools designed for children and adolescents and began administering the World Mental Health – Composite International Diagnostic Interview (WMH-CIDI). The WMH-CIDI assesses suicidality and the following *DSM-IV* disorders: depression, mania, panic, generalized anxiety, and posttraumatic stress [17]. The WMH-CIDI represents the state-of-the-art in structured diagnostic interviews, building on earlier versions of the CIDI (World Health Organization (CIDI version 2.1); University of Michigan (UM-CIDI); and Munich, Germany (M-CIDI)) and the DIS-IV [17]. The NJP continues to use the National Institute of Mental Health DIS-IV to assess (i) APD, because it is not included in the WMH-CIDI 2000; (ii) substance use disorders, because the WMH-CIDI 2000 classified drugs into "other" categories rather than identifying specific drugs abused; and (iii) schizophrenia, because the WMH-CIDI 2000 only screens for psychosis.

In addition to the assessment of mental disorders, the NJP assesses the use of mental health services, risk behaviors for HIV/AIDS and other STIs, and a variety of risk and protective factors:

- cognitive and behavioral risk factors for HIV/AIDS and other STIs (self-efficacy, perception of risk, behavioral information and skills, HIV/AIDS knowledge, normative support for HIV prevention, and attitudes and beliefs about HIV prevention);
- functional and cognitive impairments;
- adverse life events, victimization, trauma exposure, and experiences of loss and death;
- criminal and violent activity;
- adult social role performance (education, employment, finances, residential stability, living situation, and parenting);
- physical health;
- mortality (including the cause of death);
- quality of life;
- general attitudes and beliefs (self-esteem, self-efficacy, attitudes toward deviance, stages of change, religiosity, and future orientation);
- characteristics of the family of origin;
- marital and intimate relationships;
- deviant associations;
- social network and support;
- acculturation; and
- neighborhood and community characteristics.

Overview of Published Findings from the NJP

Published data from the NJP have been cited in the Surgeon General's Report on Children's Mental Health and is used by national advocacy groups and in reports to Congress. Analyses of data from the NJP are ongoing. To date, articles have been published in *Archives of General Psychiatry, American Journal of Public Health, Journal of the American Academy of Child and Adolescent Psychiatry, Journal of Consulting and Clinical Psychology, Pediatrics*, and *Psychiatric Services*. The following is a brief summary of key findings.

- **Prevalence of mental disorders.**

Almost three-quarters of females and two-thirds of males had one or more mental disorders. Substance use disorders, the most common type of disorder, affected over 50% of males and 46% of females. Females had significantly higher odds than males of having any disorder. Non-Hispanic whites had significantly higher odds than the African-Americans or Hispanics of having any disorder [18]. Nearly 93% of participants reported 1 or more traumas; 11% met criteria for post-traumatic stress disorder in the past year [19].

- **Comorbidity of ADM disorders.**

Comorbidity was common. Significantly more females (57%) than males (46%) had comorbid mental disorders. Participants with major mental disorder (i.e., major depression, mania, and psychosis) were significantly more likely to have a substance use disorder than those without major mental disorders [20]. Log-linear and latent class models were used to empirically identify the most common combinations of alcohol, marijuana, and other drug use disorders. Over 21% had 2 or more substance use disorders. The most prevalent combination was alcohol and marijuana. Four of the five participants with an alcohol disorder also had one or more drug use disorders [21].

- **Substance use.**

Self-reported drug use had a high level of veracity for the use of cannabis; among detainees with positive urinalysis results, 88% reported use in the past six months. Combining self-report and urinalysis reveals a *minimum* prevalence of 85% for any illicit drug use in the past 6 months [22].

- **Prevalence of HIV/AIDS risk behaviors.**

Baseline HIV/AIDS risk behavior data were collected from 800 participants when funding for this component became available. HIV/AIDS risk behaviors were prevalent, irrespective of gender, race/ethnicity, or age; 95% engaged in three or more risk behaviors, and over 60% engaged in 10 or more risk behaviors. All specific risk behaviors were more prevalent in our sample than the general population. Significantly more African-Americans than non-Hispanic whites had *sexual* risk behaviors; significantly more non-Hispanic whites than the African-Americans had *drug* risk behaviors [23].

- **Substance use disorder, major mental disorder, and HIV/AIDS risk behaviors.**

Substance use disorder significantly increased the odds of engaging in HIV/AIDS risk behaviors. Among youth who had comorbid major mental and substance use disorders, 58% engaged in unprotected sexual activity while drunk or high, compared to only 7% of youth with a major mental disorder but no comorbid substance use disorder. Among the youth with a substance use disorder, more than 63% engaged in 5 or more sexual risk behaviors [24].

- **Development and persistence of HIV/AIDS risk behaviors.**

The development of HIV/AIDS risk behaviors among participants between their baseline and three-year follow-up interviews was examined. Among males, high-risk sexual behaviors were significantly more prevalent at the follow-up than at baseline. Significantly more females engaged in other sexual behaviors at the follow-up than at baseline, including trading sex and drugs and recent unprotected vaginal sex. Patterns of development of HIV/AIDS risk behaviors differed by gender and race/ethnicity. More males than females developed sexual risk behaviors; more females than males developed drug risk behaviors. Significantly more non-Hispanic whites and Hispanics than African-Americans developed drug risk behaviors [25].

- **Development of antisocial personality disorder.**

On the basis of a subsample of 1112 detained youth who were adults at the time of the 3-year follow-up interview, nearly one-fifth of the male juvenile detainees later developed APD. Significantly more males than females developed APD; no differences were found by race/ethnicity. Conduct disorder (CD) diagnosis and the number of CD symptoms were significantly associated with developing modified-APD (M-APD; APD without the CD requirement). *Post hoc* analyses, however, suggested a threshold effect: participants with five or more CD symptoms were significantly more likely to develop M-APD than participants with fewer than five symptoms. Analysis also indicated that several other disorders were significantly associated with the development of M-APD, including dysthymia, alcohol use disorder, or generalized anxiety disorder. Although some disorders were strong predictors of APD, none were useful for identifying which youth would later develop M-APD [26].

- **Detecting and treating mental disorder.**

Institutional records on the provision of mental health services by the detention center and public health system were examined for up to 6 months after intake to detention. Among detainees who had major mental disorders and associated functional impairments, only 15% were treated in the detention center and even less (8%) received treatment in the community. Significantly more females than males with major mental disorders were identified and treated for mental disorders. Detection of need for mental health treatment and undergoing treatment were predicted, in part, by clinical, demographic, and legal variables. Specifically, the odds of being detected or treated was greater among the youth with a major mental disorder, treatment history, or suicidality reported at intake, and lower among racial/ethnic minorities, males, older detainees, and detainees transferred to adult court for legal processing [27].

- **Mortality.**

Seventy-seven participants (4%) died since the study began. Mortality rates of the 65 (4%) who died before March 2004 were analyzed; 96% of these youth died from homicide or legal intervention (e.g., killed by police). Standardized mortality is more than *4 times* the general population rates. Mortality among females is nearly *8 times* the general population rates. African-American males have the highest mortality; yet, they have the lowest mortality *ratio* because their mortality in the general population is high [28].

Conclusions

A substantial number of youth are involved in the juvenile justice system. Approximately two million youth are arrested each year and over 100 000 juveniles are in custody on any given day [29]. The public health and justice systems must have accurate and reliable data from which to guide public policy for youth entering the justice system. The NJP provides data that are critical to improving the outcomes of one of our nation's most needful population of youth.

References

[1] US Department of Health and Human Services (2000). *Report of the Surgeon General's Conference on Children's Mental Health: A National Action Agenda*, US Government Printing Office, Washington, D.C.

[2] The President's New Freedom Commission on Mental Health (2003). *Achieving the Promise: Transforming Mental Health Care in America* (Final Report), Report No: SMA-03-3832, Department of Health and Human Services, Rockville, MD.

[3] Hogan, M.F. (2003). New freedom commission report: the president's new freedom commission: recommendations to transform mental health care in America, *Psychiatric Services (Washington, D.C)* **54**, 1467–1474.

[4] McCabe, K.M., Lansing, A.E., Garland, A. & Hough, R. (2002). Gender differences in psychopathology, functional impairment, and familial risk factors among adjudicated delinquents, *Journal of the American Academy of Child and Adolescent Psychiatry* **41**, 860–867.

[5] Timmons-Mitchell, J., Brown, C., Schulz, S.C., Webster, S.E., Underwood, L.A. & Semple, W.E. (1997). Comparing the mental health needs of female and male incarcerated juvenile delinquents, *Behavioral Sciences and the Law* **15**, 195–202.

[6] Atkins, D.L., Pumariega, A.J., Rogers, K., Montgomery, L., Nybro, C., Jeffers, G. & Sease, F. (1999). Mental health and incarcerated youth, I: prevalence and nature of psychopathology, *Journal of Child and Family Studies* **8**, 193–204.

[7] Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edition, Lawrence Earlbaum Associates, Hillsdale.

[8] Snyder, H.N. (2005). *Juvenile Arrests 2003*, August. Report No.: NCJ209735, Office of Juvenile Justice and Delinquency Prevention, Washington, D.C.

[9] Sickmund, M., Sladky, T.J. & Kang, W. (2005). *Census of Juveniles in Residential Placement Databook*. Available from: http://www.ojjdp.ncjrs.org/ojstatbb/cjrp/, [cited 2007 March 1].

[10] Greenfeld, L.A. & Snell, T. (1999). *Women Offenders*, Report No.: NCJ 175688, US Department of Justice, Washington, D.C.

[11] US Department of Justice (2001). *Correctional Populations in the United States, 1998*, Report No: NCJ 192929, US Department of Justice, Washington, D.C.

[12] Angold, A. & Costello, E.J. (1991). Developing a developmental epidemiology, in *Rochester Symposium on Developmental Psychopathology*, D. Cicchetti & S.L. Toth, eds, University of Rochester Press, Rochester, pp. 75–96.

[13] Kessler, R.C. (1995). Epidemiology of psychiatric comorbidity, in *Textbook in Psychiatric Epidemiology*, M.T. Tsuang, M. Tohen & G.E.P. Zahner, eds, Wiley-Liss, New York, pp. 179–197.

[14] Shaffer, D., Fisher, P., Dulcan, M.K. & Davies, M. (1996). The NIMH diagnostic interview schedule for children version 2.3 (DISC-2.3): description, acceptability, prevalence rates, and performance in the MECA study, *Journal of the American Academy of Child and Adolescent Psychiatry* **35**, 865–877.

[15] Bravo, M., Woodbury-Farina, M., Canino, G.J. & Rubio-Stipec, M. (1993). The Spanish translation and cultural adaptation of the diagnostic interview schedule for

children (DISC) in puerto rico, *Culture, Medicine and Psychiatry* **17**, 329–344.
[16] Shaffer, D., Fisher, P., Lucas, C.P., Dulcan, M.K. & Schwab-Stone, M.E. (2000). NIMH diagnostic interview schedule for children version IV (NIMH DISC-IV): description, differences from previous versions, and reliability of some common diagnoses, *Journal of the American Academy of Child and Adolescent Psychiatry* **39**, 28–38.
[17] Kessler, R.C. & Üstün, T.B. (2004). The World Mental Health (WMH) survey initiative version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI), *International Jounral of Methods in Psychiatric Research* **13**, 93–121.
[18] Teplin, L.A., Abram, K.M., McClelland, G.M., Dulcan, M.K. & Mericle, A.A. (2002). Psychiatric disorders in youth in juvenile detention, *Archives of General Psychiatry* **59**, 1133–1143.
[19] Abram, K.M., Teplin, L.A., Charles, D.R., Longworth, S.L., McClelland, G.M. & Dulcan, M.K. (2004). Posttraumatic stress disorder and trauma in youth in juvenile detention, *Archives of General Psychiatry* **661**, 403–410.
[20] Abram, K.M., Teplin, L.A., McClelland, G.M. & Dulcan, M.K. (2003). Comorbid psychiatric disorders in youth in juvenile detention, *Archives of General Psychiatry* **60**, 1097–1108.
[21] McClelland, G.M., Elkington, K.S., Teplin, L.A. & Abram, K.M. (2004). Multiple substance use disorders in juvenile detainees, *Journal of the American Academy of Child and Adolescent Psychiatry* **43**, 1215–1224.
[22] McClelland, G.M., Teplin, L.A. & Abram, K.M. (2004). *Detection and Prevalence of Substance Use Among Juvenile Detainees*, June. Report No.: NCJ 203934, Office of Juvenile Justice and Delinquency Prevention, Washington, D.C.
[23] Teplin, L.A., Mericle, A.A., McClelland, G.M. & Abram, K.M. (2003). HIV and AIDS risk behaviors in juvenile detainees: implications for public health policy, *American Journal of Public Health* **93**, 906–912.
[24] Teplin, L.A., Elkington, K.S., McClelland, G.M., Mericle, A.A. & Washburn, J.J. (2005). Major mental disorders, substance use disorders, comorbidity, and HIV-AIDS risk behaviors in juvenile detainees, *Psychiatric Services (Washington, DC)* **56**, 823–828.
[25] Romero, E.G., Teplin, L.A., McClelland, G.M., Abram, K.M., Welty, L.J. & Washburn, J.J. (2007). A longitudinal study of the prevalence, development and persistence of HIV/sexually transmitted infection risk behaviors in delinquent youth: implications for health care in the community, *Pediatrics* **119**, e1126–e1141.
[26] Washburn, J.J., Romero, E.G., Welty, L.J., Abram, K.M., Teplin, L.A., McClelland, G.M. & Paskar, L.D. (2007). Development of antisocial personality disorder in detained youth: the predictive value of mental disorders, *Journal of Consulting and Clinical Psychology* **75**, 221–231.
[27] Teplin, L.A., Abram, K.M., McClelland, G.M., Washburn, J.J. & Pikus, A.K. (2005). Detecting mental disorder in juvenile detainees: who receives services, *American Journal of Public Health* **95**, 1773–1780.
[28] Teplin, L.A., McClelland, G.M., Abram, K.M. & Mileusnic, D. (2005). Early violent death among delinquent youth: a prospective longitudinal study, *Pediatrics* **115**, 1586–1593.
[29] Sickmund, M. (2004). *Juveniles in Corrections. Report*, Report No.: NCJ 202855, Office of Juvenile Justice and Delinquency Prevention, Washington, D.C.

JASON J. WASHBURN, LINDA A. TEPLIN AND KAREN M. ABRAM

Nuclear Forensics

Introduction

Nuclear forensics (nuclear forensic science/nuclear forensic analysis) is the branch of science that seeks to examine the nature, use, and origin of nuclear (fissile) and radioactive (nonfissile) materials. Nuclear forensic analysis produces a characteristic "signature" for the material and therefore may provide major evidence for nuclear attribution. It is most often applied to investigations involving the malevolent use of nuclear or radioactive materials and is of assistance in determining adherence to international safeguards against illicit trafficking [1].

Nuclear forensic analysis aims to provide legally admissible evidence that could lead to prosecution of the offenders involved with trafficking of the illicit material. Thus, it assists law enforcement agencies in the fight against illicit trafficking in nuclear and radioactive material in order to minimize the likelihood of such materials being used by terrorists as weapons in the form of radiological or nuclear (R or N) agents. Together with specific chemical and biological (C and B) agents, they form part of the CBRN agents, sometimes also called *weapons of mass destruction* (WMD). CBRN is a commonly used acronym which refers to chemical, biological, radiological or nuclear weapons that can be used to kill or injure large numbers of people, damage infrastructure, damage the economy or damage the

environment in general (*see also* **Chemical, Biological, Radiological, and Nuclear Investigations**; **Biological Agents**; **Chemical Warfare Agents** and **Bomb Scene Management**).

Nuclear attribution is the process of identifying sources of an interdicted nuclear or radioactive material in order to determine its origin, intended or original use; the routes of transfer between the point where legitimate control over the material was lost and the point where the material was intercepted; and the perpetrators responsible for the illicit dealings with the material [1].

A nuclear forensic investigation is conducted in conjunction with traditional forensic examinations of the illicit material. Such examinations might include analysis of explosive residues, biological samples, documents, ballistics, hairs, pollens, fibers, and other associated physical evidence.

The desired outcome of the nuclear forensic analysis is the appropriate prosecution of offenders in a court of law. Therefore, implementation of scientifically defensible sampling and analytical methods and the need to strictly observe chain of custody requirements are essential.

Radiological and Nuclear Agents

Examples of fissile and nonfissile radioactive materials that might be of interest in illegal trafficking are listed in Table 1. The classification of the radioactive materials (sources) follows the five categories as defined by the International Atomic Energy Agency (IAEA) [2]. Apart from nuclear materials, the radioactive sources in Radioactive source Categories 1, 2, and 3 are of specific concern. This is due to both, their high activities and their ubiquity. Devices containing sources such as caesium-137, cobalt-60, or americium-241 are used heavily in industrial and medical establishments worldwide, with access to the materials often poorly controlled. Depending on their nature and the radiation characteristics, terrorists may use such materials in the form of an improvised nuclear device (IND), a radiological dispersion device (RDD; e.g., a "dirty bomb"), or a radiological emission device (RED).

An IND is a crude nuclear weapon constructed from material such as weapons-grade uranium (20–90% U-235) or plutonium. Its explosion would have similarly disastrous consequences, although most likely on a smaller scale, than a modern, dedicated nuclear weapon. The high degree of technical expertise and the cost of building a functional IND make this scenario less likely to occur.

A possible alternative to the construction of an IND is the acquisition of a nuclear weapon through the black market or from rogue nations, or *via* theft from the arsenal of a vulnerable nuclear-weapon state. Nuclear weapons, depending on the yield, have the potential to destroy entire cities. Their impact would be physically, psychologically, and economically devastating.

An RDD disperses radioactive material into the environment, resulting in radioactive contamination of an area. Perhaps the most well-known form of the RDD is the "dirty bomb", in which a conventional high explosive is combined with a radioactive material. Alternatively, specific installations (such as a nuclear reactor, nuclear fuel manufacturing or reprocessing plant, and a radiation source manufacturing plant or a radioactive waste disposal facility) could be targeted as a means of releasing radioactive material and causing contamination of the environment. The extent of radioactive contamination would depend on the size of the explosion, weather conditions, the nature of the affected environment, and the amount, type, and form of radioactive materials involved.

An RED is a high-activity radioactive source that emits energy in the form of highly penetrating gamma or neutron radiation under concealment in a frequented public place. Unlike an IND or RDD, a RED is not designed to cause destruction and/or radioactive contamination of an area, but is aimed at exposing people to high levels of radiation. The existence of the device may not be realized until radiation exposure symptoms are observed among the affected individuals or the perpetrators make the existence of the device publicly known. The detrimental health impact of such a device on the exposed would vary, depending on the total dose received by an individual and is influenced by the duration of their exposure, the activity of the source, and the type and the energy of the radiation emitted from the source. Exposure to high doses of radiation, of the order of sieverts (i.e., hundreds of rems), would cause serious adverse health effects within a short time, and could lead to death [4].

The severity of consequences associated with the use of a device and the probability that a device will

Table 1 Categories of nuclear and other radioactive material[a]

Category	Type of material or device	Radioactive components
Unirradiated direct-use nuclear material	High enriched uranium (HEU)	>20% U-235
	Plutonium and mixed U-Pu oxides (MOX)	<80% Pu-238
	U-233	Separated isotope
Irradiated direct-use nuclear material	Irradiated nuclear fuel material	In irradiated nuclear fuel elements or in spent fuel reprocessing solutions
Alternative nuclear material	Americium-241 Neptunium-237	Separated element or present in irradiated nuclear material, in separated plutonium, or in mixtures of uranium and plutonium
Indirect use nuclear material	Depleted uranium (DU)	<0.7% U-235
	Natural uranium (NU)	0.7% U-235
	Low enriched uranium(LEU)	>0.7% U-235 and<20% U-235 (typically 3–5%) U-235
	Plutonium (Pu-238)	>80% Pu-238
	Thorium	Th-232
Radioactive source Category 1	Radioisotope thermoelectric generators	Pu-238, Cm-244 and Sr-90
	Irradiators/sterilizers	Co-60 and Cs-137
	Teletherapy source	Co-60 and Cs-137
Radioactive source Category 2	Industrial gamma radiography sources	I-192[b]
	High/medium dose rate brachytherapy sources	Co-60 and Cs-137
Radioactive source Category 3	Fixed industrial gauges Well logging gauges	Co-60, Cs-137, and Am-241
Radioactive source Category 4	Low dose rate brachytherapy sources Thickness/fill level gauges Portable gauges (e.g., moisture, density) Bone densitometers Static eliminators	Co-60, Cs-137, and Am-241
Radioactive source Category 5	Eye plaques, permanent implants	Sr-90[b]
	X-ray fluorescence devices	Fe-55[b]
	Electron capture devices	Ni-63[b]
	Mössbauer spectrometers	Co-57[b]
	Positron emission tomographs	Ge-68[b]
	Medical diagnostic sources	Short-lived radioisotopes, e.g., I-131
	Fire detectors	Am-241 and Pu-238

[a] Reproduced with permission from Ref. 3. © IAEA, 2006.
[b] Information added by the authors.

be used is directly linked to the type of the device (Figure 1).

The most detrimental consequences would result from the use of a technologically sophisticated device, such as an illegally obtained nuclear weapon or the use of an IND. However, the difficulties associated with procuring a nuclear weapon or constructing an IND markedly diminishes their probability of being used. Perpetrators would likely favor the use of a device that is less complicated

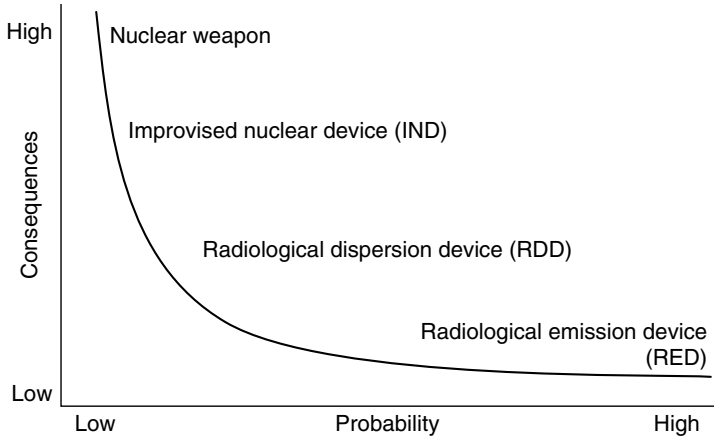


Figure 1 Relationship between the probability of the use of a given device and the detrimental consequences of its use. [Figure adapted from [5]]

and/or easier to procure, such as an RDD or a RED.

Nuclear Forensic Analysis

Nuclear forensic analysis involves physical characterization and subsequent forensic interpretation of nuclear or radioactive material [1, 3].

Characterization of the nuclear or radioactive material aims to determine the chemical and physical signature of a sample of the intercepted illicit material. A number of analytical tools may be utilized to determine the signature of the material being investigated. As in traditional forensic analysis, selection of a particular analytical method is often determined by the results obtained in the consecutive stages of sample analysis. These analytical methods can be grouped into three categories: bulk analysis tools, imaging tools, and microanalysis tools [3, 6].

Bulk analysis tools enable the determination of the elemental and isotopic composition of the material, including the presence and concentrations of trace constituents. They include chemical assays, high-resolution gamma spectrometry (HRGS), radiochemistry and radiation counting techniques, X-ray fluorescence (XRF) and X-ray diffraction (XRD) analysis, inductively coupled plasma–mass spectrometry (ICP–MS), thermal ionization mass spectrometry (TIMS), gas chromatography–mass spectrometry (GC–MS), and glow discharge–mass spectrometry (GD–MS).

Imaging tools are used to document the physical characteristics of the material such as size, shape, and topography. In addition, imaging tools allow the determination of the chemical composition across the material, such as identifying whether the sample is homo- or heterogeneous. Imaging methodologies include the use of visual inspection and photography, optical microscopy, scanning electron microscopy (SEM), and transmission electron microscopy (TEM).

Microanalysis tools are used to quantify individual constituents of heterogeneous samples *via* determination of elemental and isotopic composition. Examples of microanalysis tools include ICP–MS, TIMS, XRD, secondary ion mass spectrometry (SIMS), scanning electron microanalysis with energy dispersive sensor (SEM/EDS) or scanning electron microanalysis with wavelength dispersive sensor (SEM/WDS), and infrared (IR) spectrometry.

After the material has been characterized, the process of forensic interpretation begins. It involves matching analytical data to existing information about the origin of similar materials and the methods used in the production and processing of this material. In addition, potential similarities with previously investigated cases of illicit materials are examined. High-quality national and international databases of reference information, in addition to improved information sharing between nuclear forensic laboratories worldwide, promote a high degree of successful investigations.

Nuclear attribution is the culmination of the analysis and interpretation of nuclear and traditional forensic evidence. It is divided into two key areas: source attribution and route attribution [1]. Source attribution considers data on the investigated material, including physical characteristics and origin, in addition to the analysis of packaging and other collateral items, the point where the material was diverted from a legitimate pathway, and the potential for further supply of such material.

Route attribution focuses on the potential involvement of a black market trafficking network, the identities of traffickers, the throughput capability of the illicit network, the frequency of shipments, and the likely end-user applications.

To date, nuclear forensic analysis has been applied to investigate illicit dealings with nuclear or radioactive material with the potential for terrorist use [1, 7]. It also has been successfully applied to investigations of a criminal (nonterrorist) or an accidental nature. An example of the former is the 2006 political assassination of Alexander Litvinenko in London, United Kingdom, by targeted radiation poisoning with polonium-210 [8]. An example of the latter is the 1992 "Cold Fusion" experiment-related explosion in an electrochemistry laboratory at Menlo Park in California [1].

International Collaboration in Nuclear Forensic Science

Commencing in late 1980s, the nuclear forensic discipline has become established in an international arena as a primary technological tool employed to investigate intercepted smuggled or illicitly traded nuclear and radioactive material for the purpose of attribution. This has been possible due to an early international collaboration of nuclear forensic scientists, conducted predominantly within the framework of the 1996 chartered International Technical Working Group (ITWG) on combating nuclear smuggling and its subset, the 2004 chartered association of the International Nuclear Forensic Laboratories (INFL) [6, 9].

References

[1] Moody, K.J., Hutcheon, I.D. & Grant, P.M. (2005). *Nuclear Forensic Analysis*, Taylor & Francis, Boca Raton (Note: this is currently the primary textbook and a reference publication in the field of the Nuclear Forensic Analysis).

[2] IAEA (2005). *Categorization of Radioactive Sources (Safety Guide)*, *IAEA Safety Standard Series No. RS-G.1.9*, International Atomic Energy Agency, Vienna.

[3] IAEA (2006). *Nuclear Forensics Support*, *IAEA Nuclear Security Series No.2*, International Atomic Energy Agency, Vienna.

[4] Bevelacqua, J.J. (2004). *Basic Health Physics*, Wiley-Vch Verlag, Weinheim.

[5] Wilber, K.T. (2003). *Overview of Radiological/Nuclear Devices and Response*, JUSTNET Justice Technology Information Network, National Law Enforcement and Corrections Technology Centre, Rockville (accessed 8 Feb 2008) www.nlectc.org/training/nij2003/wilber.pdf.

[6] Mayer, K., Wallenius, M. & Ray, I. (2005). Nuclear forensics – a methodology providing clues on the origin of illicitly trafficked nuclear materials, *Analyst* **130**, 433–441.

[7] Wallenius, M., Mayer, K. & Ray, I. (2006). Nuclear forensic investigations: two case studies, *Forensic Science International* **156**(1), 55–62.

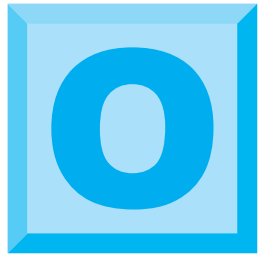
[8] Harrison, J., Leggett, R., Lloyd, D., Phipps, A. & Scott, B. (2007). Polonium-210 as a poison, *Journal of Radiological Protection* **27**, 17–40.

[9] Lawrence Livermore National Laboratory (2007). *Identifying the Sources of Stolen Nuclear Materials*, LLNL Science & Technology Review January/February 2007, 12–18.

J. George Koperski and Serena F. Abbondante

Nuclear Investigations ***see*** Chemical, Biological, Radiological, and Nuclear Investigations

Number Restoration ***see*** Serial Number Restoration: Firearm



Obliterations in Documents: Detection of *see* Alterations: Erasures and Obliterations of Documents

Odontology

Crime Scene and Recovery of Remains

Often the activity of an odontologist begins at the scene of crime, or where a body is found. All data useful for identification of an unidentified corpse or of an offender are therefore directly gathered by whoever has the knowledge to collect useful dental information. When dealing with charred (Figure 1), badly preserved, skeletonized bodies, some dental features, very important for individualization, could become very fragile or could be displaced from the mouth. During the recovery of a corpse, it should be mandatory to protect the oral area so as to preserve all useful information and to avoid the loss of teeth, bone fragments, or prosthetic devices. The oral region is therefore protected from any loss of data by using a fixative spray or means of mechanical protection; the scene is then accurately inspected to find any useful dental element or fragment (sometimes partially or totally charred).

In other crime scene scenarios, the odontologist may be useful for finding useful evidence that may help identify an offender by his lip prints or by his bitemarks on objects or food.

The Biological Profile

Teeth can provide useful information for biological profile reconstruction (*see* **Anthropology**), which is briefly summarized here.

Species

Knowledge of dental anatomy, when intact teeth are found, allows for an easy species attribution to



Figure 1 Surgical mask used to protect the oral region during corpse transportation

a trained eye (Figure 2) [1, 2]. When macroscopic morphology is not sufficient for a sure attribution of species or when studying small dental fragments, microscopic analysis of the teeth can sometimes solve the problem.

Microscopic observation of tooth enamel in fact can show the typical keyhole shape of the human enamel prisms.

Sex

The lack of reliable differences between sexes renders this approach redundant for sex determination for forensic purposes; so anthropological or genetic methods, through the extraction of genetic material from the pulp chamber, are to be preferred.

Race

Teeth can provide some indication about racial affiliation. Some characteristics are for example more frequent in Negroid subjects (as the presence of a diasthema between upper central incisors or an evident prognathism), in Mongoloid subjects (shovel-shaped teeth) (Figure 3) or in Caucasoids (Carabelli's cusp) [1, 2].

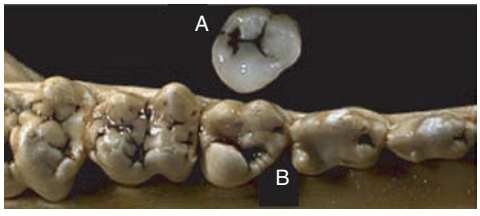


Figure 2 The very similar shape of a third human upper molar (A) and pig premolar (B)



Figure 3 Palatal view of upper dental arch where shovel-shaped incisors are evident

These dental features are not sufficient, however, for a definite race determination, and only add information to the study of the entire skeleton.

Age

Starting from intrauterine life till about 21 years of age (when complete maturation of the third molar occurs) teeth are a reliable indicator of biological age, correlated to chronological age [1–4]. The age of a fetus can be evaluated via odontological methods studying the mineralization of the cusps visible in a radiograph of the mandible (Figure 4).

Up to about 14 years of age, simple observation of the number of primary and permanent teeth allows for age estimation with a growing error starting from about three months (when aging subjects of a few months of age) to a maximum of 36 months (when aging subjects of about 14 years of age). By studying the maturation of teeth apices, it is possible to age subjects till about 21 years of age. Once dental development has ended it is possible to estimate the age of an adult via continuous tooth modification. Dental abrasion is usually used to estimate the age of ancient populations, but is not considered a valuable indicator for forensic purposes anymore. Other dental modifications such as root transparency, root reabsorption, migration of the periodontal junction, continuous apposition of dentin, and cementum are studied, categorized and measured to estimate adult subject age (Figure 5). Root transparency seems to be the variable most closely linked to aging.

There are also some more complex dental methods for estimating age, such as cementum annulation count, the quantification of aspartic acid in teeth or of artificial radiocarbon.

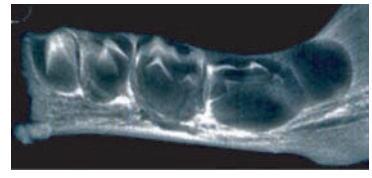


Figure 4 Mineralization of dental cusps in a radiograph of a fetus mandible. The presence of the cusps of the two deciduous molars denotes a fetal age of 38 weeks



Figure 5 Canine transilluminated section in which some features correlated to chronological age are visible: (A) root transparency, (B) apposition of dentin, (C) periodontal junction degeneration, and (D) occlusal abrasion

Aging the Living

Since teeth are a reliable indicator of chronological age, they are used to evaluate the age of individuals without valid documents, often to verify if they reached a specific age (for instance 14, 16, or 18). It is thus possible to establish with which probability a subject reached 18 years of age by studying the development of the third molar, to evaluate the age of a young subject via the ratio between the open apices and the length of the crowns or the age of an adult by observing the ratio between the area of a pulp chamber and the entire tooth.

A final remark should be made, before proceeding to the more common applications of forensic odontology (positive identification and bitemark analysis) concerning a frequently used sign in cases of suspected infanticide – the neonatal line. This is a defect in enamel (or dentine) formation occurring at birth which, in the case of enamel, shows up as a thin dense line visible microscopically which separates prenatal from postnatal enamel. The neonatal line is related to the physiological and temporary interruption of the continuous deposition of dentine and enamel and is an indicator of survival of a newborn for some hours after birth (when the body is decomposed and other markers of survival cannot be searched on soft tissue). Therefore, if the neonatal line is noticed on the dental cusps of a newborn skeleton this suggest that the baby was born alive and survived for some hours.

Positive Identification

Odontological identification is the restitution of identity to a decomposed, skeletonized, charred, or heavily traumatized corpse via oral and dental features [2–6]. Identification is probably the main activity of the odontologist as unidentified bodies are a reality in every country: giving back a name to a corpse is important for every culture either for moral, religious, criminal, or civil reasons. Teeth are the most resistant part of the human body, they are particular to each individual due to anatomical, pathological, and therapeutic features; they also are the only human skeletal/calcified part visible during the life of a subject; these considerations make teeth a good individualization tool. Some cultural modification as color alteration, nontherapeutic morphology modification, setting of stones or jewels in enamel, and nontherapeutic crowns, can provide information about a subject, sometimes these are geographical variants. Some habits or occupations (wind instrument players, pipe smokers, carpenters, or tailors who use teeth to hold metal objects) can leave distinctive marks making individualization much easier. Odontological individualization is based on the comparison between antemortem (AM) data (information about a missing person) and postmortem (PM) data (observation made on an unidentified corpse).

Antemortem Data

Odontological AM data include all information regarding teeth and the contiguous tissues, pertaining to a missing person who is believed to correspond to an unidentified corpse. In most industrialized countries such information is quite easy to find as most citizens have been visited at least once by a dentist who probably kept his or her patient's radiographs, casts or pictures. Precious individualization information can be found in dental files, in orthopantomographs (OPGs) in which it is possible to see all the teeth, in periapical radiographs with a better resolution than OPGs that allow the observation of even the smallest peculiarities, in lateral and anteroposterior skull radiographs which can show dental features (Figure 6), in every radiograph in which a part of the mouth is visible, in clinical pictures (often taken by dentists to document their works) and in normal portraits which show the teeth of the photographed subject.

All antemortem data are summarized in special forms (for instance the INTERPOL DVI form) devised to standardize data collection, to make the matching process between a missing person and an unidentified body easier, to avoid misunderstanding due to the use of different teeth nomenclatures or different scientific vocabularies.

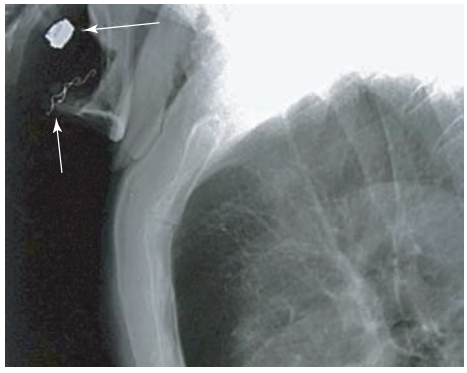


Figure 6 In this antemortem radiograph of a missing person, it is possible to see (white arrows) dental treatment useful for a positive individualization. It is thus possible to perform a positive odontological individualization starting from nondental radiographs

Postmortem Data

Odontological PM data includes all information gathered from an unidentified corpse and inserted into special forms similar to those used for AM-data collection. As unidentified bodies are usually charred, decomposed, skeletonized, or traumatized, oral structures can become very fragile and oral examination should be considered unrepeatable; therefore it is advisable to record and take pictures of every step of the examination procedures (especially if the case will be brought in a court of law). An elementary kit to perform a PM oral examination should include: oral mirrors, probes, scalpel, saws, brushes, sponges, ultraviolet light (to better detect aesthetic restorations) (Figure 7), a camera, radiographic instrumentation, and cast materials.

To collect all dental elements and dental works that could be displaced from the oral cavity a first inspection of the entire body could be necessary; some useful elements for individualization are often found in the pharynx, trachea, and esophagus or even in the bag used to carry the corpse. Especially in case of charred or mummified bodies, mouth opening could be a difficult operation. Temporomandibular joint disarticulation and removal of the entire mandible can be a solution; a quick way is to perform three cuts, using a "stryker" type saw, parallel to the occlusal plane: one passing through the nasal spine to remove the maxillary bone and two passing through the vertical parts of the mandible to remove the jaw (Figure 8).



Figure 7 Aesthetic fillings can be hard to find, especially if of good quality and if teeth are not perfectly clean. A UV light can enhance the shape of this kind of restoration



Figure 8 Maxillary bones and mandible after resection. Postmortem examination can be carried out accurately only if every dental surface is visible

Before taking the jaws from the body the type of occlusion should be recorded. Which teeth are present and which are not is then recorded (dental formula); if missing teeth are noticed it is important to chronologically place their loss by studying the bone remodeling process. Any kind of anomaly (in shape, size, and position) is recorded as well as the presence, type and location of every kind of oral disease. Very often dental individualization is based on the comparison of therapeutic features because they are usually well described in dental antemortem files (Figure 9). For each dental work, one must therefore carefully describe the type, location, and the applied technique and technology. Postmortem examination is then completed by taking dental impressions and radiographs.



Figure 9 It is possible to detect the presence of a filling even if lost. In this example the shape of the cavity on the broken premolar and the dark stain suggest the presence of an amalgam filling probably lost because of the trauma

Matching Process

During individualization processes all AM and PM data collected in the forms are compared to discover compatibility or incompatibility between a missing person and an unidentified body. Positive identification is commonly achieved via radiographic comparison, which nowadays is usually computer aided. In a radiographic comparison, for even the smallest anatomical, pathological, and therapeutic features, antemortem and postmortem radiographs are placed side by side, superimposed and eventually measured and quantified to establish if they come from the same subject (Figure 10).

Postmortem radiographs should be taken with the same technique and with the same projection used for the antemortem ones so that the results are reliable.



Figure 11 Postmortem dental profile superimposed to the picture of a missing person

Dental Superimposition

When dealing with missing persons who do not have clinical data dental superimposition is frequently a reliable identification technique (Figure 11) [5]. It is based on the superimposition of the teeth visible in a picture of a missing person to the dental cast of an unidentified corpse so as to study any incompatibility and correspondence.

Palatal Rugae

Palatal rugae are mucosal ridges on the hard part of the palate. They are morphologically peculiar in each individual and do not change shape for a long time during the life of an individual. When a superior dental impression is taken the palatal rugae configuration is recorded and then transferred onto the cast. All these characteristics make palatal rugae a suitable tool for individualization even with the limitation of being a soft tissue (think of modifications due to dehydration, decomposition, and carbonization). Morphological comparison of palatal rugae is usually performed by superimposition of scanned antemortem casts (provided by the missing person's dentist) and

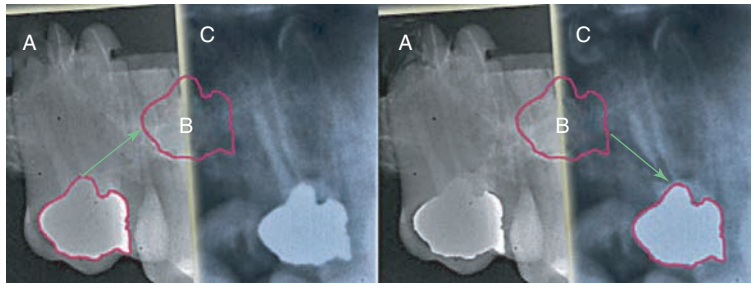


Figure 10 Postmortem radiograph (A) compared to antemortem radiograph (C). Filling postmortem outline (B) is superimposed to the antemortem one to increase comparison reliability

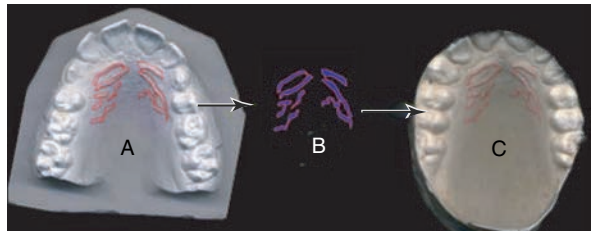


Figure 12 Antemortem (A) palatal rugae pattern (B) superimposed to postmortem cast (C). In this example of a positive identification case, more than 20 years had passed between the antemortem and the postmortem casts

of scanned postmortem casts of the corpse's palate (Figure 12).

Mass Disasters

The contribution of odontology in the identification of the victims of a mass disaster is often decisive. Subjects identified with odontological methods are often more than 80% of the victims [2, 3, 6, 7]. Therefore forensic odontologists should always be included in antemortem and postmortem identification teams. Identification methods applied in case of mass disaster are the same as those used in cases of a single unidentified body; special care is to be taken with respect to logistics, to the organization of the team work, and to the precise compilation of ante and postmortem forms.

Bitemark Analysis

A forensically useful bitemark is a physical alteration or a lesion showing some characteristic of the dental arches that produced it [3, 2, 8, 9]. Bitemarks can be found on a crime scene on objects or food, on the skin of a corpse or of an assaulted subject and can be used to identify who provoked them. The comparison between the dental status of a defendant and the bitemark is almost always done by a computer aided image superimposition: the occlusal pattern of the defendant's dental arches is superimposed to the bitemark image so as to analyze the similarity of the two patterns. Bitemarks are frequently found in sexual assault and the typical locations are arms, breasts, legs, and genitalia. The lesion is usually elliptic, with a central ecchymosis, and with tooth marks on its perimeter (Figure 13).

It is important to study a bitemark as soon as possible either on a living subject (because of morphological changes due to tissue healing), or on a corpse (to avoid postmortem alterations). First of all DNA swabs must be taken from the lesion as the aggressor could have left some genetic material on the skin. Then photographs are shot taking care to use a metric scale and to avoid any possible deformation or distortion (Figure 13). It is possible to produce a cast of the lesion by using dental impression materials (e.g., polyvinylsiloxane). The lesion can be excised from the corpse and if transilluminated can better show some important peculiarities. If the lesion is in a place which can be reached by the dentition of the victim it could be necessary to take his or her dental impression in order to exclude a self-inflicted bite. Some bitemarks can hardly be associated with a human bite because of the low-pressure applied (only a round ecchymosis is visible) (Figure 14), because of excessive pressure (tissue amputation), or if multiple, because bites are superimposed or partial or because they have been inflicted on a portion of skin which has an excessive curvature. Some objects with a tubular section can leave on the skin a mark



Figure 13 Typical elliptic bitemark found on the breast of a corpse: note the teeth pattern on its perimeter, the ecchymosis, and the presence of the American Board of Forensic Odontology (ABFO) ruler

that could erroneously be associated with the action of human dental arches.



Figure 14 Multiple bitemarks on a curved surface: it could be difficult to compare the dental pattern to a defendant's dental arch or even to associate this kind of lesions to a human bitemark

The defendant is usually subjected to a dental visit: missing teeth are recorded as well as any pathology, dental fracture, abnormal tooth position, rotation, or inclination; a dental cast is then produced from accurate impression materials so as to record even the smallest enamel peculiarity.

References

[1] Hillson, S. (1996). *Dental Anthropology*, Cambridge University Press, Cambridge.

[2] Stimson, P.G. & Mertz, C.A. (1997). *Forensic Dentistry*, CRC Press, Boca Raton.

[3] Bowers, C.M. & Bell, G. (1995). *Manual of Forensic Odontology*, American Board of Forensic Odontology.

[4] Whittaker, J. & MacDondal, D.G. (1989). *Color Atlas of Forensic Dentistry (Wolfe Medical Atlases)*, Mosby International Press.

[5] Clark, D. (1992). *Practical Forensic Odontology*, Butterworth Heinemann, London.

[6] Bowers, M.C. (2004). *Forensic Dental Evidence: An Investigator's Handbook*, Elsevier Academic Press.

[7] Cattaneo, C., De Angelis, D. & Grandi, M. (2006). *Mass Disasters in Forensic Anthropology and Medicine: Complementary Sciences From Recovery to Cause of Death*, A. Schmitt, E. Cunha & J. Pinheiro, eds, Humana Press, New York.

[8] Bowers, C.M. & Johansen, R. (2000). *Digital Analysis of Bitemark Evidence*, Academic Press.

[9] Dorion, R.J. (2005). *Bitemark Evidence*, Marcel Dekkeer, New York.

CRISTINA CATTANEO AND DANILO DE ANGELIS

Older Adults in Court *see* Elderly in Court

Opioids

Introduction

Natural and synthetic morphine derivatives possessing morphinelike actions are known as *opioids*. These include common drugs such as morphine, heroin, codeine, etc. and are collectively known as *opioids*. Other terms used to describe these drugs include narcotic analgesics and opiates.

After cannabis and stimulants, the most widely used illicit drugs are opioids. Legitimate opioids, such as morphine, codeine, oxycodone, and hydrocodone, are being consumed at an increasing rate. According to the International Narcotic Control Board, approximately 50 countries have increased their consumption of opioid analgesics by more than 100% during the last decade [1]. The increase is particularly evident in Europe and North America.

Globally, it is estimated that 16 million people take opioids, including 11 million who use heroin. In many countries, the majority of heavy drug users seeking treatment are primarily addicted to heroin; however, it is of increasing concern that legitimate opioids, such as oxycodone and methadone, are being diverted for illegal use.

The use of heroin is common in Asia (54%), Europe (25%), America (14%), and Oceania (6%) [1]. Evidence from national surveys and other data sources suggests that the prevalence of the use of heroin in general populations is relatively low (ranging from 0.2 to 2%). According to the USA 2005 National Survey on Drug Use and Health survey, 379 000 persons (0.2%) reported using heroin in the past year. It was estimated in the year 2000 that there were between 40 000 and 100 000 heroin users in Australia [2] comparable to Britain and other countries in Europe.

The worldwide production of heroin has more than doubled or even tripled since 1985. Scientific reports published by the United Nations Drug

Control Program [1] showed that there was a global increase in the production, transportation, and consumption of opioids, throughout the 1990s; however, in the last 5 years, the production of opium has steadily decreased. World events including the political changes in Afghanistan, the world's largest producer of heroin, led to significant decline in the supply and subsequent demand of the drug. In the golden triangle consisting of Laos, Myanmar, and Thailand, production has fallen by 85% from 1998 to 2006 [1]. Recent reports suggest that production is reaching the high levels of the 1990s.

The use of heroin, in particular, is causing widespread health and social problems in many countries. In Europe, heroin injectors who regularly consume large amounts of different drugs face a risk of death, which may be 20–30 times higher than non-drug users in the same age range. Since heroin is commonly used by injection, the health risks including that of human immunodeficiency virus (HIV) and hepatitis transmission are substantial.

The primary use of legal opioids is for the relief of pain. Depending on the severity of pain, different opioids can be used to diminish or relieve pain. Severe pain can be treated using fentanyl, hydromorphone, methadone, morphine, oxycodone, tramadol, and pethidine, while mild-to-moderate pain can be treated with codeine, dihydrocodeine, and dextropropoxyphene. Other opioids are used to induce or supplement anesthesia, such as fentanyl and the fentanyl analogues alfentanil, remifentanil, etc.

Some opioids (codeine, dihydrocodeine, and, to a lesser extent, pholcodine) can also be used as cough suppressants (antitussives). Methadone, buprenorphine, naltrexone, and naloxone can be used for the treatment of addiction to opioids.

Opioids can also be mixed with other pharmaceuticals (nonopioid drugs) to enhance analgesia (e.g., analgesic-antipyretic preparations and use with some phenothiazines). A summary of selected opioids and their uses are given in Table 1.

There are substantial region-to-region differences in the availability and usage of opioids. For example in 2004, the United States accounted for more than 99% of the global consumption of hydrocodone.

Sources of Opioids

Morphine and codeine are naturally occurring alkaloids extracted from the milky juice and stalks of the opium poppy, *Papaver somniferum*. The content of morphine in these plants can vary from as little as 5% to as high as 25%. Codeine is usually found in much smaller amounts (~0.2%) and the content of other naturally occurring opioids, such as thebaine (~0.1 – 0.3%), narcotine (~0.3%), narceine (~4 – 10%), and papaverine (~1%), can also vary depending upon the quality and origin of the plant [3]. Many semisynthetic opioid derivatives are made by relatively simple modifications of the morphine or thebaine molecule, e.g., heroin (diacetylmorphine) is a semisynthetic opioid produced from the acetylation of morphine. Other semisynthetic drugs produced from morphine or thebaine molecules include hydromorphone, oxycodone, and naloxone (Figure 1).

Pharmacology of Opioids

Effects on the Central Nervous System

Opioids exert their diverse pharmacological effects by binding to opioid receptors that are distributed in distinct patterns throughout the central nervous system (CNS) and peripheral nervous system. The major therapeutic use of opioids is the modulation or reduction in pain. When an opioid is administered therapeutically to patients suffering from pain, the sensation of pain is reduced and their distress becomes less intense. However, when an opioid is given to normal drug-free individuals, nausea and vomiting are more common symptoms along with drowsiness, lethargy, and a reduced physical activity. Patients may experience euphoria, and at increased doses side effects become more severe along with muscular rigidity and respiratory depression. Higher doses far in excess of those required to produce analgesia may lead to convulsions [3].

Respiratory Depression

Most opioids depress the respiratory center. The depression is a dose-dependent effect, which can be fatal. Opioids act to reduce the responsiveness of the brain stem respiratory centers to increased concentrations of arterial carbon dioxide. The subsequent effect of increasing concentrations of hydrogen ions as a result of dissociation of carbonic acid lowers blood pH that directly affects the pontine and medullary centers involved in regulating respiratory function.

Table 1 Selected opioids, the action at drug receptor, typical dose, half-life, and indication for use

Drug	Receptor action	Dose (mg)[a]	$T_{1/2}$ (h)[b]	Indication for use[c]
Naturally occurring				
Morphine	Agonist	5–20	2–3	Analgesia, moderate-to-severe pain, postoperative pain control
Codeine	Agonist	30–60	2–4	Analgesia, moderate-to-severe pain, cough suppressant
Semisynthetic				
Heroin	Agonist	5–10	0.03–0.05	Analgesia, moderate-to-severe pain, not available for therapeutic use
Hydromorphone	Agonist	2–4	1.5–4	Acute pain, chronic cancer pain
Naloxone	Antagonist	0.4–2	1–2	Used to reverse effects of opioids, such as heroin
Oxycodone	Agonist	5–20	4–5	Analgesia, relief of chronic pain
Synthetic				
Buprenorphine	Partial agonist	–	2–3	Analgesia, maintenance drug used to treat heroin dependence
Fentanyl	Agonist	0.05–0.2	1–6	Analgesia, strong pain, used in anesthesia as adjuvant
Methadone	Agonist	5–100	10–25	Analgesia, relief of chronic pain, maintenance drug used to treat heroin dependence
Pethidine	Agonist	50–100	3–10	Analgesia
Propoxyphene	Agonist	50–150	8–24	Analgesia
Tramadol	Agonist	50–400	4–8	Analgesia

[a] Common recommended therapeutic dose given orally
[b] Time for blood concentration to halve
[c] Most common medical applications

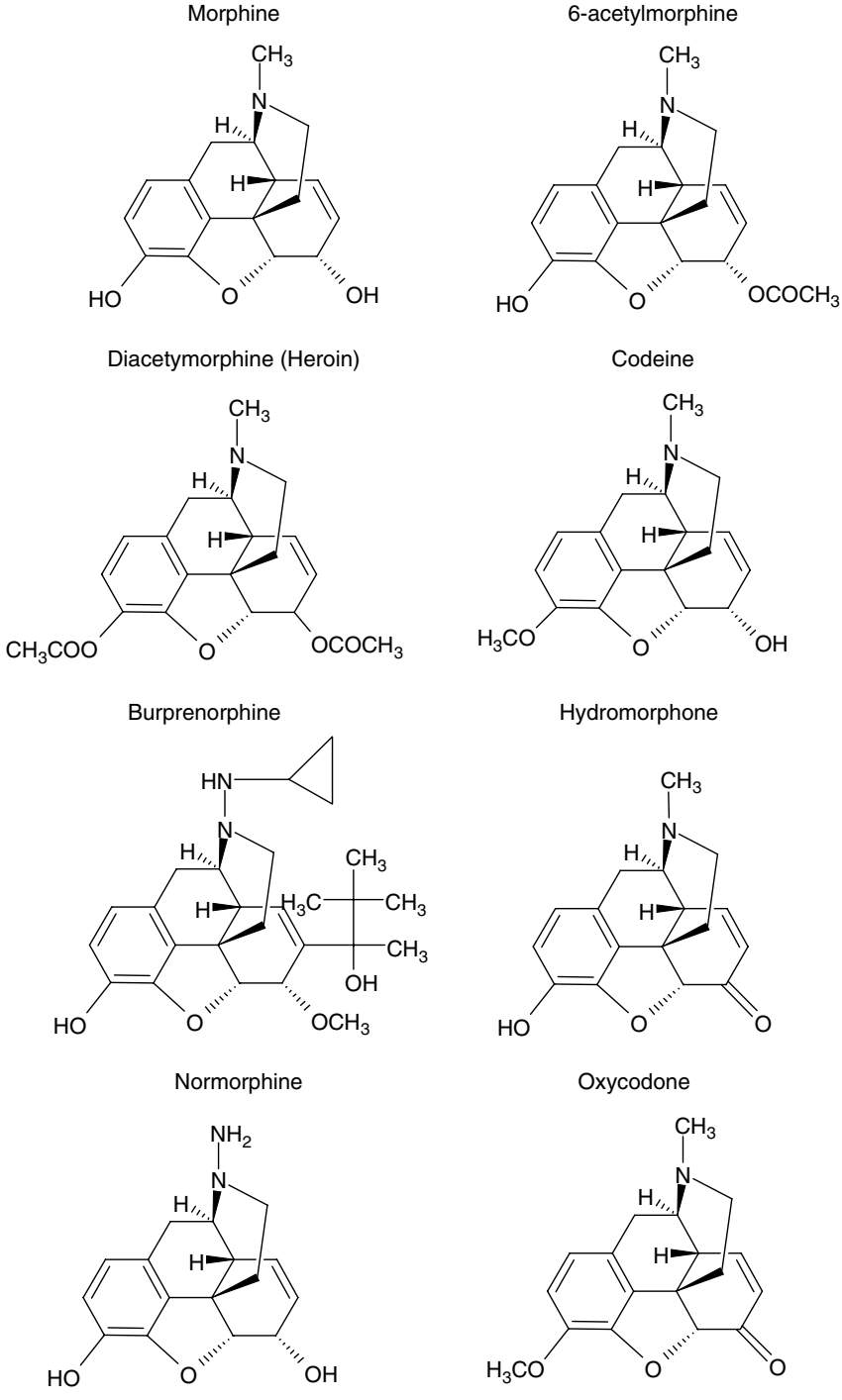


Figure 1 Morphine and selected structural derivatives of morphine

This has the effect of reducing the rate of breathing and with toxic amounts of morphine the rate may fall to 3 or 4 breaths per min instead of 12–16. Codeine is much less toxic than morphine and even in large doses does not appear to have the strong CNS depression characteristics of morphine.

Suppression of Cough Reflex

Opioids depress the cough reflex by a direct effect on the cough center in the medulla. There seems to be no relationship between the analgesic, respiratory-depressant action of opioids, and cough suppression; however, some opioids are more effective in depressing the cough reflex than others. Codeine and, to a greater extent, pholcodine are more effective than morphine in depressing the cough reflex [4, 5].

Nausea and Vomiting

Nausea and vomiting occur as a result of opioids affecting the chemoreceptor trigger zone in the brain. These effects are common initially but usually disappear with repeated administration. All clinically useful opioids produce some degree of nausea and vomiting [3]. Antiemetic drugs such as haloperidol and prochlorperazine are commonly used to treat vomiting in clinical situations. Metoclopramide is often useful if nausea and vomiting persist [5].

Effects on the Gastrointestinal Tract

Opioids cause varying effects on the gastrointestinal tract depending on the dose and the drug species. For example, morphine causes a marked increase in tone and a reduced motility, resulting in constipation, while codeine is less constipating than morphine and may be used to relieve abdominal pain. The emptying of gastric contents can also be delayed, which can further retard the absorption of drugs, a property shared by most opioids [3].

Tolerance and Dependence

Tolerance is a condition whereby after repeated administration, a given dose of a drug produces a decreased effect and increasingly larger doses must be taken to obtain the pharmacological effects observed with the original dose.

It is the level of tolerance that is of particular importance in heroin addicts. When the use of opioid becomes more regular, the intensity of response diminishes from one injection to the next. To produce the same response or as before, the dose of the drug has to be increased. Thus, in the course of days or weeks, tolerance may build up until many times a lethal dose can eventually be tolerated.

In humans, for example, an initial dose of 100–200 mg of morphine would be sufficient to cause profound sedation, respiratory depression, anoxia, and death; however, tolerant subjects can handle this and more. Tolerance diminishes rapidly (few days) after withdrawal so that a previously tolerated dose may prove fatal [4]. Tolerance to other opioids can also occur depending on the potency of the drug.

Dependence is a condition that has developed as a result of repeated drug administration and/or abuse. It is characterized by an overwhelming need to continue taking the drug, or one with a similar pharmacological property. Opioid analgesics, such as heroin, are abused for their euphoriant effects and dependence develops rapidly with regular use. Prolonged administration of heroin or morphine in a clinical environment is less likely to produce dependence as a result of therapeutic use and the dose of the drug can be reduced as the underlying pain decreases [6].

The abrupt withdrawal of opioids from persons physically dependent on them precipitates a withdrawal syndrome, the severity of which depends on the individual, the drug used, the size and frequency of the dose, and the duration of drug use. Opioid analgesics with some antagonist activity, such as buprenorphine, butorphanol, or pentazocine, may also precipitate withdrawal symptoms in patients who are dependent on opioid narcotics.

The onset and duration of withdrawal symptoms also vary according to the duration of action of the specific drug. Withdrawal symptoms may be terminated by a suitable dose of morphine or another opioid. Methadone is currently the most widely used pharmacotherapeutic agent for maintenance treatment of heroin addicts [7]. Methadone is effective in the suppression of withdrawal symptoms and in the reduction or elimination of an addict's compulsion to take heroin. The major objective of the methadone maintenance program is to achieve long-lasting stabilization of the user's drug dependence by providing

methadone indefinitely in doses large enough to produce a level of cross-tolerance that is sufficient to diminish the effects of ordinary doses of heroin.

Serious Adverse Effects/Toxic Reactions

Overuse of opioids can lead to depression of the CNS, slowing down breathing that may result in postural asphyxia or cardiorespiratory arrest. Intravenous drug abuse of heroin can lead to a number of other life-threatening side effects and/or complications. These can include neurological disorders, low blood pressure, swelling of the brain, stroke, death of vascular tissue, i.e., necrotizing angiitis, and nerve damage.

Intravenous drug users show a high susceptibility to infections and other diseases due to the suppression of their immune systems. The manifestation of infectious diseases in intravenous drug users plays a critical role in life-threatening complications for addicts. The factors below contribute to their diminished immune function. Factors such as needle sharing, unsafe sexual practices, under nourishment, poor hygiene, devastation of skin barrier, and the injection of unfamiliar substances into the body all increase the risk of acquired immune deficiency syndrome (AIDS), viral hepatitis, pneumonia, and tuberculosis. Infectious complications as a result of all these factors include endocarditis and septicemia, viral hepatitis, liver cirrhosis, meningitis, and tetanus [8].

Disposition of Opioids

Absorption and Metabolism

The extent of absorption depends on the type of opioid and the route of administration. When given orally, some opioids, particularly morphine and heroin, are removed from the portal circulation very efficiently by the liver and are metabolized extensively so that the amount reaching the systemic circulation is considerably less than the amount absorbed into the portal vein. This effect is known as *first-pass metabolism*. As a consequence the available dose is only a proportion of the dose taken. Oral administration of morphine is one-sixth as effective as parenteral administration. In contrast, codeine is well absorbed and quite active when given orally. In fact, codeine is approximately two-thirds as effective orally as parenterally [9]. Other opioids that are well absorbed orally include oxycodone and pethidine (meperidine).

Some opioids are also available as sustained release preparations (i.e., slow or controlled release). This means that a patient suffering from chronic pain may only have to take a tablet once or twice per day, which results in stable blood concentrations as a consequence of the drug's more predictable pharmacokinetics. Methadone and oxycodone rectal suppositories are useful alternatives to oral sustained release morphine preparations.

Opioids undergo extensive metabolism in humans. There are three main biotransformation pathways that have been established for the metabolism of opioids: hydrolysis, glucuronidation, and oxidation.

For example, following intravenous injection, heroin is rapidly converted (within seconds) to 6-acetylmorphine (6-AM), which is subsequently hydrolyzed (within minutes) to morphine. The conversion of heroin to 6-AM occurs both as a result of enzymes and spontaneous hydrolysis [10].

The majority of opioid metabolism in humans occurs in the liver through a process called *glucuronidation*; this occurs mainly in the liver and to a lesser extent in the intestine and kidneys [11, 12]. For example, morphine is conjugated primarily to morphine-3-glucuronide (M3G) and to a lesser extent the biologically active morphine-6-glucuronide (M6G) (Figure 2).

Excretion

Most opioids are excreted in the urine usually within the first 24 h following administration. The concentrations of conjugated metabolites usually exceed those of the parent drug and can be present for longer periods of time [13]. As with most pharmacokinetic parameters, there is great individual variation in the elimination half-life. This value is greatly prolonged in renal failure resulting in drug accumulation. Some opioids with long half-lives such as methadone can accumulate in tissues. Such accumulation can be toxic or even fatal if doses are not appropriately administered and the drug not given enough time to be cleared by the body. This can lead to prolonged respiratory depression and pronounced enterohepatic recirculation (excretion into the gastrointestinal system through bile followed by reabsorption) [14].

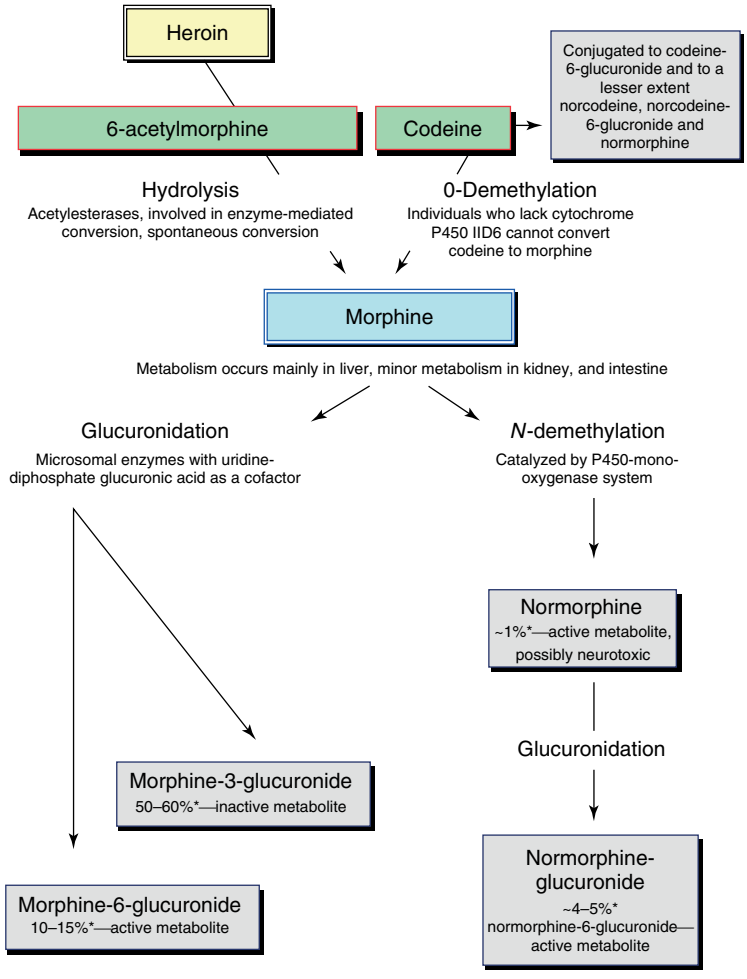


Figure 2 Fate of heroin, morphine, and codeine in humans (*urinary excretion, percent of dose)

Death from the Use of Opioid

The mechanism of death in opioid users is often uncertain, but it is likely to be multifactorial. For example, the use of heroin results in a wide range of adverse effects due to a variety of pharmacological and physiological responses to heroin, hypersensitivity reactions to the cutting agents or contaminants, and diseases associated with intravenous use. The largest numbers of deaths have been attributed to an acute reaction, whereby death occurs shortly after injection [15, 16]. Three overdose syndromes are recognized: death from profound respiratory depression, death from arrhythmia and cardiac arrest, and death as a consequence of severe pulmonary edema [17].

Death may also occur indirectly as a complication of unconsciousness. This is caused by a nonfatal dose, leading to airways obstruction in a setting of diminished respiratory function. Respiratory disease can also reduce the ability of person to tolerate a dose of opioid.

Fentanyl, methadone, oxycodone, hydrocodone, and morphine are the most common legal opioids known to cause death if misused [18].

Drug Interactions

Drugs such as alcohol, barbiturates, and benzodiazepines enhance the depressant effects of opioids on the CNS [19].

Opioid users frequently take benzodiazepines, e.g., diazepam, nitrazepam, etc., to reduce anxiety and to minimize the unpleasantness of any withdrawal symptoms. Drug combinations are therefore potentially serious complications in opioid users, particularly for heroin users. Data from a review of deaths in Victoria [20] show that benzodiazepines were the most prevalent drug group (~55%) and alcohol was present in ~30% of all heroin-related deaths. Amphetamines were also present in many cases (~13%) (*see also* **Benzodiazepines**).

Drugs such as the sedating antihistamines can prolong morphine metabolism, leading to increased respiratory depression. Phenothiazines, including promethazine and chlorpromazine, are also known to potentiate the effects of opioids by interfering with the metabolism of morphine [21].

One of the problems confounding the forensic toxicologist and pathologist is the relevance and interpretation of drug concentrations in different specimens following the use of opioids.

Interpretation of Toxicological Data; Difficulties for the Expert

The interpretation of postmortem tissue results is dependent upon number of considerations. These include chronic or acute use of the drug; concentrations of drug may be higher in those who are regular users, allowing some assessment of toxicity.

The route of administration is also important. Oral ingestion of opioids means that some of the efficacy of the dose is removed by the liver before the drug enters the blood stream (first-pass metabolism).

Another factor that will affect the interpretation of toxicology results includes the incidence of disease, i.e., AIDS, hepatitis, etc., presence of liver and kidney dysfunction as well as other immunological suppressive diseases that may affect metabolism and excretion of opioids and their metabolites. The accumulation of opioids may result in adverse effects and even death. Concurrent natural disease may predispose individuals to cardiovascular collapse and even unpredictable responses such as convulsions[3].

The presence of other drugs, the possibility of antagonism for one drug by another, and more often the additive or synergistic effects produced by the interaction of two or more depressant drugs must all be considered.

Tissue distribution and redistribution of opioids and its metabolites are also important in assessing postmortem contributions of opioids to adverse effects and death. The statistical data on body distribution and redistribution studies from fatalities associated with heroin, morphine, and other opioids use are invaluable in the evaluation of future toxicological findings (*see* **Postmortem Toxicology: Artifacts**).

Methods of Analysis

Initial Testing

Most clinical and forensic laboratories test for the presence of the class of opioids in the initial phase of their investigations. These immunoassays are effective for the more common opioids, such as morphine, codeine, and heroin metabolites. This technique using commercial kits enables all members of the class to be detected; however, owing to their differing immunoreactivities, the sensitivity to different opioids and their metabolites will vary, and for some may be quite poor. This applies particularly to the more synthetic opioids, e.g., oxycodone, buprenorphine, and methadone. Hence, more specific immunoassays are required to detect these drugs.

Initial tests are commonly conducted in urine since the concentration of the opioids or its metabolite is often much higher than blood. However, immunoassays designed for blood/plasma, or even oral fluid, are commercially available. An example of this is the enzyme-linked immunosorbent assay (ELISA).

When a class test is positive, further (confirmation) tests are required to detect the specific opioid (or its metabolite) that is causing this positive response (*see* **Toxicology: Initial Testing**).

Confirmation Testing

The definitive confirmation method in forensic toxicology is mass spectrometry. This can be gas chromatography/mass spectrometry (GC-MS) or liquid chromatography/mass spectrometry (LC-MS). liquid chromatography/tandem mass spectrometry (LC-MS/MS) or tandem LC-MS methods are now dominating the measurement for this class of drugs due

to its very high sensitivity and specificity and require very little or no chemical modification to permit chromatographic analysis.

Depending on the specimen, the metabolic pathway of the target opioid (or metabolite) will vary. When determining heroin, most if not all of the drug is metabolized to 6-AM and ultimately morphine, which are the main target metabolites, whereas for buprenorphine, the parent drug and metabolite norbuprenorphine are measured.

References

[1] International Narcotics Control Board, United Nations (2007). *Annual Report International Narcotics Control Board 2006*.

[2] Hall, W.D., Ross, J.E., Lynskey, M.T., Law, M.G. & Degenhardt, L.J. (2000). How many dependent heroin users are there in Australia? *The Medical Journal of Australia* **173**(10), 528–531.

[3] Jaffe, J.H. & Martin, W.R. (1985). *Goodman and Gilman's The Pharmacological Basis of Therapeutics*, 7th Edition, A.G. Gilman *et al.*, eds, Macmillan, New York, pp. 495–531.

[4] Rang, H.P. (1987). in *Pharmacology*, 2nd Edition, H.P. Rang & M.M. Dale, eds, Churchill Livingstone, Edinburgh, pp. 547–567.

[5] Badewitz-Dodd, L.H. (1994). *The MIMS Annual*, Australian Edition, Intercontinental Medical Statistics (Australasia), Crows Nest, pp. 302–333.

[6] Victorian Drug Usage Advisory Committee. Analgesic Guidelines Sub-Committee (1988). *Analgesic Guidelines*, [Prepared by the Analgesic Guidelines Sub-Committee Victorian Drug Usage Advisory Committee], 1st Edition, Victorian Medical Postgraduate Foundation, Toorak, on behalf of the Victorian Drug Usage Advisory Committee.

[7] Peachey, J.E. (1986). The role of drugs in the treatment of opioid addicts, *The Medical Journal of Australia* **145**(8), 395–399.

[8] Janssen, W., Trubner, K. & Puschel, K. (1989). Death caused by drug addiction: a review of the experiences in Hamburg and the situation in the Federal Republic of Germany in comparison with the literature, *Forensic Science International* **43**(3), 223–237.

[9] Reynolds, J.E.F. (1999). *Martindale: the Complete Drug Reference*, 30th Edition, Pharmaceutical Press, London, pp. 1065–1098.

[10] Yonemitsu, K. & Pounder, D.J. (1992). Postmortem toxico-kinetics of co-proxamol, *International Journal of Legal Medicine* **104**(6), 347–353.

[11] Regnard, C.F. & Twycross, R.G. (1984). Metabolism of narcotics, *British Medical Journal (Clinical Research Ed)* **288**(6420), 860.

[12] McQuay, H.J., Moore, R.A., Hand, C.W. & Sear, J.W. (1987). Potency of oral morphine, *Lancet* **2**(8573), 1458–1459.

[13] Schneider, J.J., Ravenscroft, P.J., Cavenagh, J.D., Brown, A.M. & Bradley, J.P. (1992). Plasma morphine-3-glucuronide, morphine-6-glucuronide and morphine concentrations in patients receiving long-term epidural morphine, *British Journal of Clinical Pharmacology* **34**(5), 431–433.

[14] Hanks, G.W., Hoskin, P.J., Aherne, G.W., Chapman, D., Turner, P. & Poulain, P. (1988). Enterohepatic circulation of morphine, *Lancet* **1**(8583), 469.

[15] Helpern, M. (1972). Fatalities from narcotic addiction in New York City. Incidence, circumstances, and pathologic findings, *Human Pathology* **3**(1), 13–21.

[16] Baden, M. (1993). Investigations of death from drug abuse, in *Spitz and Fisher's Medicolegal Investigation of Death: Guidelines for the Application of Pathology to Crime Investigation*, 3rd Edition, W.U. Spitz, ed, with a foreword by R. Clark, C.C. Thomas, Springfield, pp. 527–555.

[17] Cotran, R.S., Kumar, V. & Robbins, S.L. (1989). *Robbins and Cotran Pathologic Basis of Disease*, 4th Edition, V. Kumar, A.K. Abbas & N. Fausto, eds, with illustrations by J.A. Perkins, Saunders, Philadelphia, pp. 497–498.

[18] Stout, P.R. & Farrell, L.J. (2002). Opioids – effects on human performance and behavior, *Forensic Science Review* **15**, 29.

[19] Iwamoto, E.T., Fudala, P.J. & Mundy, W.R. (1987). *Toxicology of CNS Depressants*, I.K. Ho, ed, CRC Press, Boca Raton, pp. 145–196.

[20] Gerostamoulos, J., Staikos, V. & Drummer, O.H. (2001). Heroin-related deaths in Victoria: a review of cases for 1997 and 1998, *Drug and Alcohol Dependence* **61**(2), 123–127.

[21] Keeri-Szanto, M. (1974). The mode of action of promethazine in potentiating narcotic drugs, *British Journal of Anaesthesia* **46**(12), 918–924.

DIMITRI GEROSTAMOULOS

Oral Fluid Toxicology

Introduction

Over the last few decades, oral fluid has been evaluated as a diagnostic aid in medicine for determining oral and systemic disease markers as well as for monitoring the presence of numerous drugs, narcotics, and

hormones. The easy, rapid, and noninvasive nature of collection and the relationship between oral fluid and plasma levels make oral fluid a valuable clinical tool and an alternative specimen where blood sampling could be difficult to perform e.g., in children, in those with poor venous access, and in anxious subjects. Furthermore, it may provide a cost-effective approach for the screening of large populations. However, unlike plasma, the composition of oral fluid varies widely both intra- and interindividually. The result of any oral fluid analysis is influenced by many different factors, including the sampling protocol, which must be performed under standardized conditions.

Oral fluid analysis for drugs was first used almost 30 years ago for the purpose of therapeutic drug monitoring. At that time, it was already known that it is possible to predict the free fraction of a circulating drug by analysis of the corresponding oral fluid sample. Since abuse of drugs is now a widespread problem across society, its negative impact on performance and safety cannot be overstated. New strategies for drug testing should offer an effective solution to persistent problems such as the potential for sample adulteration and substitution and concerns over issues of individual privacy. These have resulted in the use of oral fluid for drugs of abuse testing in the workplace. Other areas of interest include testing of intoxicated drivers, monitoring illicit drug use in drug treatment, and oral fluid testing for drug detection in schools and in transportation and insurance industries. This article discusses developments in the field of collection devices, pharmacokinetics of common drugs of abuse in oral fluid, and recent technological advances and guidelines that increase the actual use of oral fluid testing in a legal context and its acceptance by the criminal justice system. Oral fluid testing is a reliable, new technology that overcomes many of the problems of older methods for drug detection.

Physiology of Oral Fluid

Oral fluid originates from three pairs of major salivary glands (parotid, submandibularis, and sublingualis) (Figure 1), a great number of minor salivary glands, the oral mucosa, and gingival crevices. The mixture of gingival crevicular fluid and mucosal transudate is referred to in the literature as “saliva”, “whole saliva”, and “oral fluid”. In addition to these fluids, oral fluid contains a certain amount of cell-debris arising from the epithelial cells of the mouth, together with food residues. Water is the major constituent (99%). Other components include mineral salts, proteins such as mucins (lipoproteins related to lubrification), and enzymes for digestion. The total protein concentration is less than 1% of that of plasma, but almost all of the organic compounds of plasma may be detected in oral fluid in trace amounts. The most important functions of oral fluid are as follows: (i) to moisten the mucous membranes of the upper aerodigestive tract in order to facilitate speech and solubilize food to ease swallowing; (ii) maintenance of oral health by controlling the bacterial flora of the mouth, and establishing defense and killing mechanisms; and (iii) to supply enzymes for food digestion, hormones, and other pharmacologically active compounds.

The detailed morphology of salivary glands has been reviewed by several authors [1–3]. The glandular tissue comprises acinar or tubular cells, specialized groups of cells arranged as endpieces surrounding a small central lumen. A narrow intercalated duct leads from the secretory endpieces to the striated ducts, which in turn drain into the secretory ducts to form a single main secretory duct that drains into the oral cavity.

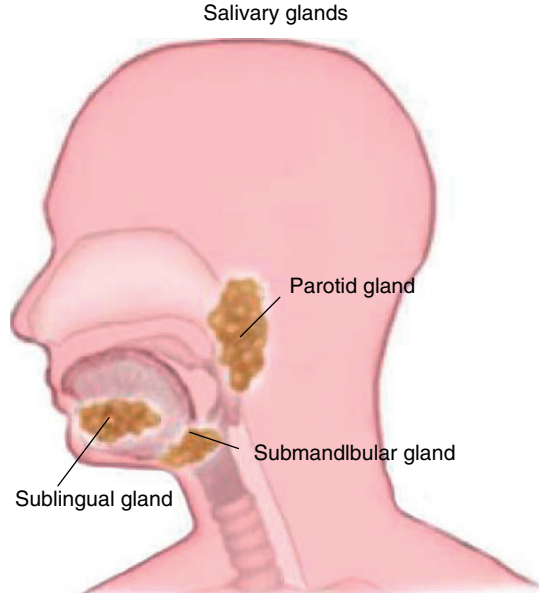


Figure 1 The three major salivary glands

Saliva contains the usual electrolytes of the body fluids. The formation of a primary saliva, isotonic compared to plasma, is situated in the endpieces of the salivary glands, and depends on the active transport of one or more of the principal ions (Na^+, Cl^-, K^+, and HCO_3^-) from the interstitial fluid to the acinar cells and the lumen. Water enters the lumen by osmosis. As this initial fluid moves down the ductal system of the salivary gland, both an active reabsorption of Na^+ and an active secretion of K^+ occur. The resulting saliva becomes increasingly hypotonic in the ductal system, the osmolality depending on the salivary flow (Figure 2). Generally, as the salivary flow increases, higher concentrations of Na^+ and Cl^-, and lower concentrations of K^+ are obtained in the final saliva. The excretion patterns of HCO_3^- are extremely variable among different glands. Usually, the bicarbonate concentration increases when the salivary flow increases, resulting in a higher salivary pH. The resting pH is about 6.8. Increasing the salivary flow results in a higher osmolality and a pH that approaches the pH of plasma or even slightly higher [4].

Under healthy conditions, adults approximately produce 500–1500 ml oral fluid per day. Salivary secretion is a reflex response controlled by both parasympathetic and sympathetic nerves. In addition, the fluids secreted by the various glands are considerably different from each other. The glandula parotis produces a serous fluid i.e., devoid of mucin, the glandula submandibularis a sero-mucous secrete, while the glandula sublingualis secretes a mucous saliva. Moreover, every type of salivary gland is stimulated to another degree by different stimuli, thus contributing differently to the total salivary production. Several factors are important such as the moment of the day (circadian rhythm), sex, age, nutritional or emotional state, and the type of the

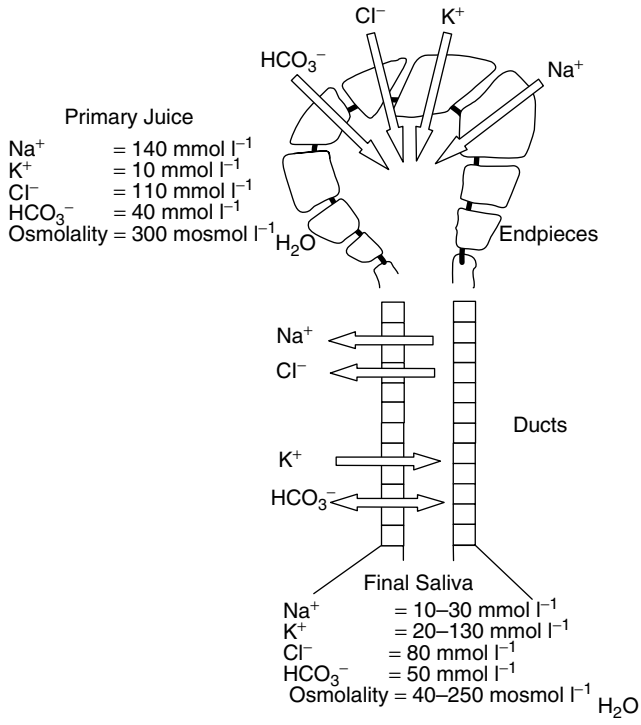


Figure 2 Hypothetical model of electrolyte transport to explain the two-stage formation of saliva. The isotonic primary fluid is produced in the secretory endpieces and is postulated to be either Cl^- rich or HCO_3^- rich. As the primary fluid passes down the gland duct system, Na^+ and Cl^- are reabsorbed while K^+ is secreted actively at somewhat lower rates. The result is the formation of a hypotonic final saliva, usually poor in NaCl and rich in $KHCO_3$ [Reproduced with permission from Ref. 4. © Central Police University Press, 1999.]

salivation stimulus [5]. Taste and olfactory stimuli, mechanical stimulation (chewing), pain, pregnancy-related hormonal changes, aggression, and sympathomimetic and parasympathomimetic drugs increase the salivary flow rate. Menopause-related hormonal changes, stress, antiadrenergic, and anticholinergic drugs decrease the salivary flow rate. An overview of the influence of different therapeutic drugs on salivation has been published recently by Aps and Martens [5]. For example, a typical β-adrenergic drug causes a viscous, protein and mucin rich, secrete. This type of saliva has a foamy appearance, while the volume produced is rather low. The cholinergic (parasympathetic) drug pilocarpine can be used, in some limited cases, to help patients complaining of dry mouth. Variations in salivary flow can also be affected, reversibly or irreversibly, by numerous physiological and pathological factors. This is the major difference between oral fluid and serum, in which the concentrations of the various components can only vary between narrow border values.

Mechanisms of Drug Transfer in Oral Fluid

Salivary glands have a high blood flow. Before any drug circulating in plasma can be discharged into the salivary duct, it must pass through the capillary wall, the basal membrane, and the membrane of the glandular epithelial cells, which is the rate-determining step. Different mechanisms are thought to occur: passive diffusion through the membrane, active processes against a concentration gradient, and filtration through pores in the membrane. Passive diffusion is by far the most common mechanism of drug transport, and is limited to nonprotein bound, nonionized molecules, with a molecular weight of less than 500 Da and a certain degree of lipophilicity. Oral fluid contains predominantly the parent drug because of the higher lipid solubility and, therefore, higher potential for passive diffusion.

In addition to the physicochemical properties of the drug (pK_a, lipid solubility, molecular weight, and spatial configuration) and the degree of plasma protein binding, this transport process is influenced by the pH of both media. Mathematical models have been developed for prediction of the saliva-to-plasma (*S/P*) drug concentration ratios for both acidic and basic drugs [6]. Especially for drugs with a pK_a close to the pH of saliva e.g., cocaine and most opiates, the degree of ionization changes drastically with small changes in pH, which is reflected in the *S/P* ratio. Calculated *S/P* ratios assuming a resting pH of 6.8 have been listed for a large number of substances [7]. The influence of salivary pH on the *S/P* ratio of many drugs is perhaps the reason why experimentally determined *S/P* ratios are different from the theoretical values. This phenomenon is also directly associated with the importance of the salivary flow and the collection protocol (with or without stimulation). Since only nonionized drugs can cross biological membranes and the pH of saliva is usually lower than the plasma pH (7.4), basic drugs usually concentrate in oral fluid due to ion-trapping. As such, oral fluid concentrations are much higher for these drugs than the corresponding plasma concentration [8–10]. In addition, in a controlled study with 3,4-methylenedioxy-*N*-methylamphetamine (MDMA), salivary pH appeared to be lowered by the actual intake of the drug itself, which explained to some extent the high *S/P* ratios observed, exceeding by far the theoretical calculated value [8, 11].

In contrast, there is very little partitioning of Δ^9-tetrahydrocannabinol (THC), the active ingredient of cannabis, between plasma and oral fluid. Since THC is a weak acid with a pK_a of about 9.5 and substantially plasma protein-bound, the calculated *S/P* ratio is 0.1. In practice, it has been demonstrated that shortly after drug use by smoking, oral ingestion, or nasal insufflation (e.g., heroin, methamphetamine, marijuana, and cocaine), contamination of the oral cavity can lead to dramatically elevated oral fluid concentrations of the parent drug, clearly inflated relative to concomitant blood levels. Only after a period of 2 h the oral fluid levels more approximately reflect blood levels. For THC, similarity in time profiles between oral fluid and plasma occur. This is possibly due to reserves of THC deposited in the oral mucosa that are leached out with time [12].

Collection of Oral Fluid

Any potential application for oral fluid testing needs a thorough understanding of the chosen collection method in order to interpret test results. In addition, the choice of a collection protocol should not only depend on the ease-of-use, but also the analytical considerations have to be taken into account. Common

methods of oral fluid collection are spitting, draining, suction, and collection on various types of absorbent swabs.

Since drug concentrations can be decreased when increasing the salivary flow, it might be advantageous to collect oral fluid without stimulation. Spitting itself is usually a sufficient stimulus to elicit a flow but the sample volume is often insufficient and the intra- and intersubject variability is large. Moreover, it is not a well-accepted and hygienic procedure for donor and acceptor and it is time-consuming in subjects suffering from dry mouth caused by stress, smoking, and the use of amphetamines and other psychotropic drugs that pharmacologically reduce salivation. In these circumstances, the flow can be stimulated mechanically (by placing paraffin wax, Teflon™, rubber bands, or chewing gum in the mouth) or chemically (lemon drops or citric acid crystals) to obtain a cleaner, more abundant specimen. A mechanical stimulus stimulates a flow of approximately 1–3 ml min^{-1}; citric acid stimulation may produce flows from 5 to 10 ml min^{-1}. It is also suggested that stimulation might decrease the intersubject variability [13]. Stimulation of salivary flow is not without drawbacks though. Different studies have confirmed that stimulation of oral fluid (citric acid) reduces drug concentration [10, 14–16]: two- to threefold for codeine and methamphetamine and fivefold for cocaine. These changes are attributed to a dilution effect by the increased output of oral fluid as well as a possible pH effect.

To allow the use of oral fluid testing for drug screening of large populations, a variety of devices have been marketed over the years. They promote an easy, quick, and reproducible collection and a cleaner specimen that is more suitable for analysis. As a general rule, they consist of a sorbent material that becomes saturated in the mouth of the donor, and after removal, the oral fluid is recovered by centrifugation or by applying pressure. Commercial devices include Omni-Sal® (Cozart Biosciences Ltd., Abingdon, UK), Salivette® (Sarstedt AG, Rommelsdorf, Germany), Intercept® (OraSure Technologies, Bethlehem, PA, USA), Finger Collector® (Avitar Technologies, Inc, Canton, MA, USA), ORALscreen™ (Avitar Technologies, Inc,), and Quantisal™ (Immunalysis Corporation, Pomona, CA, USA). Significant differences are reported in percent recovery from the sorbent material for various drugs and a large variability in the ultimately measured oral fluid concentrations [16]. For some devices, e.g., the Quantisal device, excellent recoveries were obtained [17], whereas for other devices, e.g., Salivette, a substantial sequestration on the cotton roll could be observed [18–20]. A typical example is the observation that THC remains bound to the sampling device and that an organic solvent is needed to release it [21–23]. A modification of the sampling procedure for the Intercept® collector (Figure 3), consisting of the addition of 2 ml of methanol to the elution buffer, resulted in complete recovery of THC over a large concentration range [22]. In addition, various pretreatment methods of the spitted samples can lead to significant differences in concentrations between collection protocols [23, 24]. The most recent study on this topic studied the recovery using three different collection devices supplied by Cozart, Immunalysis, and Microgenics [25]. Drugs studied were THC, benzodiazepines, methamphetamine, and morphine. Of the three systems studied, only the Cozart product gave acceptable recovery of THC from drug-spiked oral fluid.

Significant differences in analyte stability were observed for some compounds, e.g., THC and 6-AM (6-acetylmorphine), depending on the storage conditions and the sampling device used [16, 26–28]. In general, the presence of a stabilizing buffer guarantees the stability of the analytes for longer periods of time. The addition of buffer results in difficulties in estimating the actual volume of oral fluid collected [16, 25, 29]. It has been shown that for the Intercept® device, reliable quantitative results can still be obtained when applying a gravimetric



Figure 3 Example of an oral fluid collection device: Intercept®. After gently wiping the collector pad between gum and cheek for approximately 2 min (as a kind of toothbrush), the device is placed in the supplied vial, which contains a stabilizing buffer solution, and sealed. After centrifugation in the laboratory, the recovered fluid can be transferred and analyzed

determination of the amount of oral fluid [19]. An additional problem is that, even when using a collection device, the amount of oral fluid collected can be inadequate for analysis [20].

Pharmacokinetics and Interpretation

A large number of key publications contain valuable concentration kinetics data for oral fluid after controlled administration of the substance to volunteers. An excellent review on the pharmacokinetics of drugs in oral fluid has been recently published by Drummer [30].

Analytical Procedures

A comprehensive review of the analytical methods from 2000 to 2006 for the analysis of drugs of abuse in oral fluid, has recently been submitted [31]. A correct procedure for forensic toxicological analysis involves two different methods. In general, after application of an immunoanalysis technique, a chromatographic method is performed: the first allows for a preliminary monitoring of a large number of samples in a reduced period of time, while the second step provides the required specificity for confirmation. For oral fluid, the amount of matrix collected is smaller when compared to urine and blood and the target range of concentrations is generally lower than in the corresponding urine sample.

The assays routinely used for screening of blood and hair, have a high sensitivity and cross-reactivity for the parent drug and are therefore also advantageous for the analysis of drugs in oral fluid. Several microtiter plate enzyme immunoassays (EIA) have been evaluated for the screening of oral fluid specimens; they only need a limited amount of specimen [25, 28, 29, 32–41]. Erroneous quantitative results might be caused by eating or drinking shortly before sampling or by the collection technique itself, e.g., interference of citric acid or the cotton roll [12, 16]. Commercially available EIA kits are designed for use with specific collection devices. However, it was shown that the application of certain EIA kits was not restricted to the use with one single type of collection method through the validation of sensitive and specific assays [25, 39–41].

As for the confirmation analysis, oral fluid can be extracted and analyzed in the same manner as other biological fluids, such as blood. Although gas chromatography (tandem) mass spectrometry GC-MS(MS) is regarded as the standard technique for drug testing in oral fluid [21, 42], the use of liquid chromatography (tandem) mass spectrometry (LC-MS(MS)) for the analysis of this sample matrix has steadily increased [43–50]. Analytical sensitivity is a significant concern when conducting oral fluid drug testing. Decreased salivary secretion is a side effect of many drugs including stimulants yielding small volumes of oral fluid available for analysis. In addition, metabolite concentrations can be lower than parent drug concentrations, further elevating sensitivity concerns [30].

As oral fluid contains considerably less proteins and lipids compared to blood or plasma, it was suggested that this allows an easier and faster analysis [51]. However, in view of the higher protein, amino acid, and especially mucin content of oral fluid relative to the urine matrix [52], the development of fast and easy sample preparation methods could be hampered [51, 53]. In addition to the endogenous substances present in oral fluid, specific collection devices mostly include a preservation buffer containing other compounds such as stabilizing salts, nonionic surfactants, and antibacterial agents. Their presence can significantly impact precision and accuracy of subsequent LC-MS-MS measurements through a phenomenon called *ion suppression* or *ion enhancement* [54].

Cannabinoids

Cannabis is the collective term for the psychoactive substances of the *Cannabis sativa* plant and one of the most frequently used illicit drugs in the Western world. THC (Figure 4), the primary psychoactive analyte, is found in the plant's flowering or fruity tops, leaves, and resin. Cannabis is usually smoked but can also be ingested in preparations such as tea and cake. The composition of the cannabinoids in oral fluid samples remains a topic of interest for several research groups. For several years, the only compounds detected in oral fluid after smoking of cannabis were THC, cannabidiol, and cannabinol. As mentioned above, THC has a calculated *S/P* ratio of approximately 0.1 [7]. The Substance Abuse and Mental Health Service Administration (SAMHSA) proposes a screening cutoff of 4 ng ml^{-1} for THC as the target analyte for the initial screen and 4 ng ml^{-1}

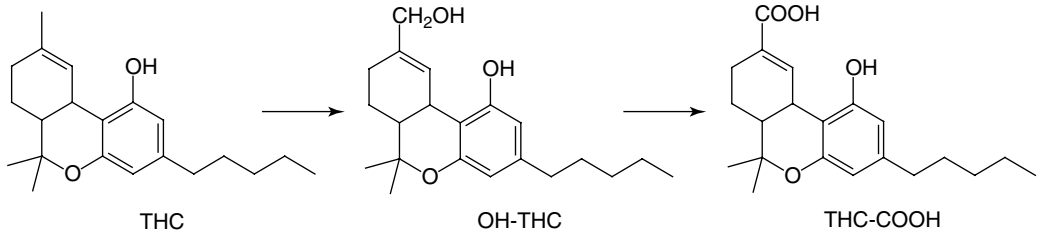


Figure 4 Major metabolic route for Δ^9-tetrahydrocannabinol (THC)

of THC in the confirmation analysis [55]. However, a recently published document by Standards Australia refers to target concentrations of 25 and 10 ng ml^{-1} for screening and confirmation analysis, respectively [56] (Table 1).

Since 2000, five studies investigating the pharmacokinetics of THC in this matrix have been published [12, 48, 57–59]. Substantial differences in THC levels were observed depending on the sampling method. In the studies by Niedbala *et al.* and Laloup *et al.*, oral fluid samples from volunteers having smoked a "joint" were collected with the Intercept® device and analyzed using GC-MS/MS (limit of quantification (LOQ): 0.2–0.25 ng ml^{-1}) or LC-MS/MS (LOQ: 0.1 ng ml^{-1}) [48, 57–59]. Average peak concentrations ranged from 31 to 596 ng ml^{-1} 15–20 min after smoking. Interestingly, doubling the THC dose and removal of tobacco did not result in a significant increase in the observed THC concentrations in oral fluid. A typical concentration profile for the detection of THC in these oral fluid samples is shown in Figure 5.

Collection of oral fluid by spitting after citric acid stimulation yielded much higher THC concentrations

Table 1 Cutoff concentrations (ng ml^{-1}) for screening and confirmation for each drug in undiluted oral fluid[a]

	SAMHSA[b]		Australian standard[c]	
Drug	Initial test	Confirmatory test	Initial test	Confirmatory test
Cocaine	20		50	
Cocaine		8		25
BE		8		25
EME		–		25
Opiates	40		50	
Morphine		40		25
Codeine		40		25
6-AM		4		10
Amphetamines	50		50	
Amphetamine		50		25
Methamphetamine		50		25
MDMA		50		25
MDA		50		25
MDEA		50		–
Cannabis	4		25	
THC		4		10

[a] Reproduced with permission from Reference [56]. © Standards Australia, 2006
[b] SAMHSA-proposed cutoff concentrations for workplace drug testing (Draft 4) [55]
[c] Approved target concentrations by Standards Australia for the detection and quantification of drugs in oral fluid, intended for workplace, medico-legal and court-directed issues [56]

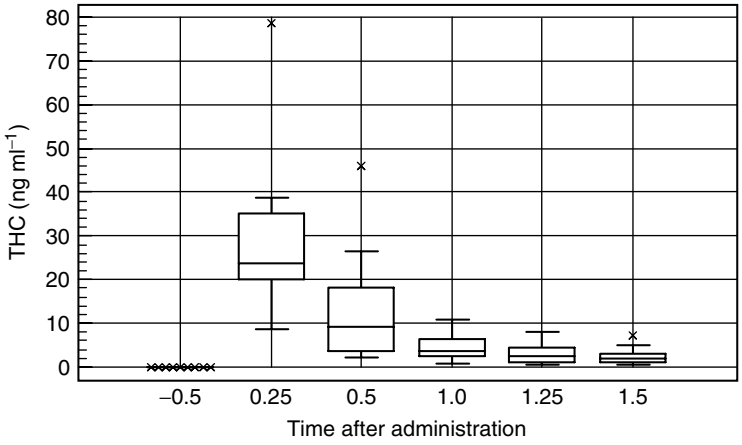


Figure 5 Box- and whisker plots of THC levels in preserved oral fluid samples from nine healthy volunteers following smoking of a single marijuana cigarette. Oral fluid samples were taken with Intercept® 0.5 h prior to smoking and at 0.25, 0.5, 1, 1.25, and 1.5 h after smoking. Concentrations plotted on the *Y*-axis are expressed as ng ml^{-1}. The central box represents the values from the lower to upper quartile (25–75 percentile). The middle line represents the median. The horizontal line extends from the minimum to the maximum value, excluding "outside" (not present) and "far out" values (cross marker) that are displayed as separate points [Reproduced from Ref. 48. © Elsevier, 2005.]

relative to the former collection method [12]. As discussed previously, this is probably due to the fact that THC was only partially recovered from the Intercept® sampling device [22]. In citric acid stimulated oral fluid, peak concentrations of 864 and 4167 ng ml^{-1} were observed 0.2 h after smoking of a 'joint' containing 1.75% (15.8 mg) and 3.55% (33.8 mg) THC, respectively. In addition, a close relationship between oral fluid and plasma concentrations of THC was observed, possibly due to reserves of THC deposited in the oral mucosa that is leached out with time. This was confirmed in a study by Ramaekers *et al.* [60], where a strong and linear relation between THC in serum and oral fluid was obtained after single doses of 250 and 500 μg kg^{-1} THC by smoking.

Irrespective of the collection method, the concentrations of THC in oral fluid appear to decline in a biphasic manner. After cannabis administration, peak concentrations observed in the first oral fluid specimen are attributed to initial contamination of oral fluid during smoking. The elevated THC concentrations in oral fluid due to this initial direct contamination dissipate rapidly (within 30 min) followed by a second phase of decline of oral fluid-sequestered THC in a similar manner as plasma THC. Estimates of the initial half-life of THC show a mean value of 21 min. Estimates of the terminal half-life of THC show a mean of 122 min [59]. Oral fluid concentrations generally declined below 1 ng ml^{-1} after 12–16 h.

After the oral consumption of 20–25 mg THC in the form of brownies, much lower peak oral fluid concentrations of about 4 ng ml^{-1} at 1–2 h were observed [57]. Concentrations declined rapidly and were below the limit of detection (LOD) of the GC-MS/MS by 16 h.

One additional issue in the use of oral fluid for drug screening is the possibility of passive contamination. Indeed, two recent studies demonstrated the presence of detectable THC concentrations in oral fluid from volunteers exposed passively to cannabis smoke in an unventilated room (area 36 m^2) [59] or van (15.3 m^2) [58]. Concentrations ranged from 7 to 26 ng ml^{-1} (mean 13 ng ml^{-1}) and from 4.5 to 7.5 ng ml^{-1} (mean 6.2 ng ml^{-1}), respectively. THC concentrations declined rapidly and became negative within 30–45 min. However, oral fluid collector devices exposed to circulating air inside the van but not to oral fluid were also highly contaminated with THC. Further analysis revealed that, when the subjects were removed from the space where passive exposure occurred prior to sample collection, the THC concentrations in oral fluid specimens were only minimal (0.0–1.2 ng ml^{-1}). Considering a SAMHSA cutoff of 4.0 ng ml^{-1}, passively exposed subjects are

considered to be negative by avoiding the environmental contamination.

Moore *et al.* and Day *et al.* recently demonstrated the presence of the principal urinary metabolite, 11-nor-9-carboxy-Δ^9-tetrahydrocannabinol (THC-COOH) in the oral fluid matrix [61, 62]. To approach the required sensitivity, oral fluid samples were analyzed using GC-MS/MS [61] or two-dimensional GC-MS [62]. Day *et al.* reported an LOQ of 10 pg ml^{-1} and concentrations up to 240 pg ml^{-1} present in the oral fluid specimens, previously identified as positive for THC. In a study involving a frequent marijuana user, THC-COOH could be detected in all samples collected at 15 min until 8 h after smoking (range 51–134 pg ml^{-1}) [62]. Interestingly, the last sample was the highest in concentration, which suggests that the increase may be reflective of deposition of this metabolite in blood. Therefore, THC-COOH could be a potential long-term marker of marijuana use since it is unlikely to come from oral contamination as is the case for parent THC. In another report of these authors, the same GC/GC-MS method was used to study the contribution of THC-COOH to enzyme-linked immunosorbent assay (ELISA) commonly used for oral fluid screening [63]. They have shown that the inclusion of this metabolite in the confirmation profile for cannabinoids in oral fluid increases the confirmation rate of the immunoassay result by at least 9.7% and minimizes the argument for passive contamination of the oral cavity.

Amphetamines

Amphetamine, methamphetamine, and the designer amphetamines are synthetic stimulants, with similarities to some naturally occurring weak stimulants such as ephedrine and pseudoephedrine (*Ephedra* species). The β-phenylisopropylamine backbone is the structural basis for many of the sympathomimetic amines. Substitutions on the nitrogen and the ring system account for most of the structural variations and differences in stimulating and euphoric effects. MDMA ('ecstasy') is increasingly used for its desired feelings of emotional closeness and sensory pleasure in the context of large, all-night dance parties where the drug is regarded as safe by the party visitors, and where multiple doses are consumed during the night [64, 65]. The calculated theoretical *S/P* ratios for amphetamine, methamphetamine, and MDMA are 2.2, 4.0, and 3.9, respectively [66]. However, the experimental values obtained always exceeded the theoretical value, possibly due to the effect of the salivary pH as discussed previously.

Proposed SAMHSA screening cutoff values are 50 ng ml^{-1} for *d*-methamphetamine as the target analyte and confirmation cutoff values are 50 ng ml^{-1} for amphetamine, methamphetamine, MDMA, MDA (3,4-methylenedioxy-*N*-amphetamine), and MDEA (3,4-methylenedioxy-*N*-ethylamphetamine). In addition, when methamphetamine is present in the specimen, it must also contain amphetamine at a concentration $\geq$ LOD. Standards Australia established similar screening cutoff values; however, the confirmation cutoffs were lowered to 25 ng ml^{-1} for amphetamine, methamphetamine, MDMA, and MDA (Table 1). In two controlled studies [8, 9] where healthy volunteers were administered a single dose of MDMA, the patterns of oral fluid and plasma MDMA concentration–time profiles agreed well. However, the intersubject variability was substantially higher for oral fluid [8, 9]. Peak MDMA concentrations of 1728–6510 ng ml^{-1} at 1.5 h after administration of 100 mg MDMA were observed, declining to a mean concentration of 126 ng ml^{-1} at 24 h [8]. A lower mean peak concentration of 1215 $\pm$ 944 ng ml^{-1} was noted by Samyn *et al.* 2 h after administration of 75 mg MDMA [9]. Levels of MDMA averaged 526 $\pm$ 372 ng ml^{-1} at 5 h after administration. Oral fluid samples were collected by spitting, without any kind of stimulation, and the salivary pH was determined for MDMA users and control subjects in one study [8]. MDMA levels exceeded the plasma concentrations by, on average, a 10-fold at peak concentrations declining to a lower more constant value (4–7) after 5–24 h. MDA, the metabolite of MDMA, was also present in these oral fluid samples, with concentrations representing about 4–5% of the concentration of oral fluid MDMA, as was also observed for plasma in this study [8].

One study documents the presence of MDMA and MDA in oral fluid samples collected with the Intercept® device [54]. Specimens were collected 1.5 and 5.5 h after the administration of a single dose of 75 or 100 mg MDMA. Analysis using LC-MS/MS (LOQ: 2.0 ng ml^{-1}) revealed a median MDMA concentration of 448 and 316 ng ml^{-1} after 1.5 and 5.5 h, respectively, in the preserved oral fluid samples. The corresponding MDA concentrations remained quite low, with a median concentration of 7.7 and 16.5 ng ml^{-1} respectively.

There is only one recent report dealing with methamphetamine and amphetamine pharmacokinetics in oral fluid after a controlled administration of methamphetamine [10]. Oral fluid specimens were collected using citric acid candy stimulation or Salivette® cotton swabs treated with or without citric acid after a single 10- or 20-mg dose of sustained release methamphetamine. Significant higher concentrations were obtained in oral fluid samples collected with neutral cotton swabs in comparison with those obtained from citric acid-treated swabs and after citric acid candy stimulation. Following citric acid candy stimulation, mean peak methamphetamine concentrations were 106 ± 25 and 192 ± 121 ng ml^{-1}, respectively, occurring 4–8 h and 2–12 h postdose. All citric acid candy stimulated oral fluid specimens collected at 24 h after drug administration were above the LOQ of the GC-MS method for methamphetamine (2.5 ng ml^{-1}). When using the SAMHSA-proposed cutoff values, only one of 13 individuals tested positive 24 h after a single methamphetamine dose. Oral fluid methamphetamine concentrations were double than those observed in plasma. However, a considerable within- and between-subject variability contributed to the poor observed correlation between plasma and oral fluid methamphetamine concentrations. Amphetamine, as the metabolite of methamphetamine, was not always detected in oral fluid specimens following administration of methamphetamine. When present, concentrations were about one-tenth that of methamphetamine.

Cocaine

Cocaine is an alkaloid obtained by extraction of the leaves of *Erythroxylon coca*, a plant that grows at 1000–2000 m elevations in the Andean mountains under warm tropical conditions. Cocaine has been used by the Peruvian Indians for centuries for the well being and increased endurance it produces after chewing the leaves of the plant. In the late nineteenth century, it was used medically as an anesthetic. It was not until the late 1960s that the recreational use of cocaine became a significant social problem in the Western world, especially in the US [67–69]. Cocaine is sold on the street in two forms: the hydrochloride (HCl) salt and the free base form "crack". Street purity ranges from 1 to over 90%, with cocaine HCl mainly used for intravenous injection and nasal insufflation, and the free base used for smoking (volatilization with a glass pipe and heat).

The metabolism of cocaine has been thoroughly studied (Figure 6) [67, 70, 71]. The main primary metabolites of cocaine are benzoylecgonine (BE), formed by chemical hydrolysis and enzymatic hydrolysis by a liver methylesterase, and ecgonine methyl ester (EME), formed by enzymatic hydrolysis by liver carboxylesterases and plasma butyrylcholinesterase. Both inactive metabolites are further metabolized to ecgonine [72]. Norcocaine is a minor metabolite in humans.

Cocaine is a highly lipophilic, basic drug and has a pK_a of 8.6. The theoretical *S/P* ratio calculated

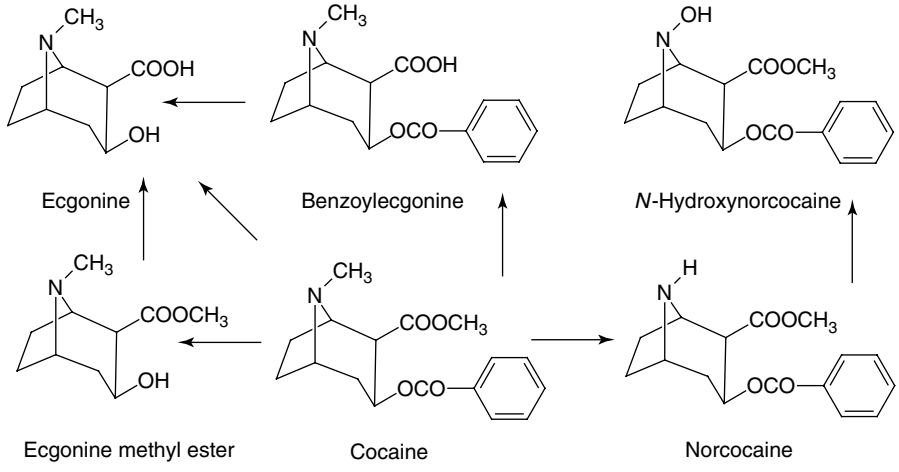


Figure 6 Main metabolic pathways of cocaine

for a salivary pH of 6.8 is 3.8 [7]. When assuming a pH in the range of 6.8–7.8 (depending on the nature of stimulation), *S/P* ratios of 3.82–0.44 were calculated. SAMHSA-proposed oral fluid cutoff values are 20 ng ml^{-1} for cocaine metabolites for screening assays and 8 ng ml^{-1} for cocaine or BE in confirmation analyses. Higher cutoff concentrations for cocaine and metabolites of 50 and 25 ng ml^{-1} for screening and confirmation, respectively, were adopted by the Australian standard (Table 1).

The Cone group published the most complete study on cocaine excretion in saliva in 1997 [73]. Cocaine appeared in saliva rapidly following different administration routes. Contamination of the oral cavity after smoking and sniffing was variable but significant during the first hours after administration of a single low dose of cocaine, resulting in high *S/P* ratios. BE and EME usually appeared in oral fluid between 0.08 and 1 h after administration. Typically, cocaine concentrations exceeded metabolite concentrations over the first 2 h, rapidly declining, sometimes below metabolite concentrations within 4–6 h. Anhydroecgonine methylester (AEME) is a pyrolysis product that also appears in oral fluid in high concentrations; it can be used as a marker of cocaine smoking.

Following repeated oral cocaine administration, the obtained pharmacokinetic parameters were generally comparable to those reported from acute dosing studies [74].

A study on the disposition and elimination of cocaine in self-reported users of cocaine (0.1–2 g smoked cocaine daily) showed cocaine oral fluid concentrations on average 2.4-fold higher than those found in plasma [75]. These findings suggest a prolonged elimination of cocaine in active street users than in occasional users, though the half-life of its main metabolite BE remains similar.

Opiates

The opiates are a large family of drugs with a mechanism of action and pharmacological activities related to morphine. Potent opiates such as morphine are particularly important in medicine due to their ability to relieve moderate to severe pain. Codeine is one of the most widely used opiates, and is easily available in some countries in low dose over-the-counter preparations with weak analgesic properties and as a cough suppressant. A number of synthetic opiates have been developed that resemble morphine when viewed as a 3D image on the receptor, although at first glance they do appear quite different.

Heroin is widely available as an illicit drug, and is produced in clandestine laboratories by synthetic acetylation of morphine from crude extracts of the exudates from the opium poppy (*Papaver somniferum*). Heroin is more lipophilic and crosses the blood–brain barrier more rapidly than morphine resulting in more intense euphoric effects. Heroin is injected intravenously (usual dose 10–15 mg), snorted or inhaled (by smoking the free base or inhaling the vapors of heated powder). Intravenous injection of heroin is common and is the fastest way to deliver the drug to the brain. Peak heroin plasma concentrations are obtained within 2 min, and are no longer detectable after 30 min. Heroin is almost immediately metabolized to 6-AM, which is detectable at 1 min and peaks at 5 min and which is rapidly transformed to morphine (Figure 7).

Heroin has a pK_a of 7.6, which results in variable *S/P* ratios, highly depending on the salivary pH. Morphine is less lipid soluble than heroin. It has a pK_a of 8.1 and is for about 65% bound to plasma proteins. The theoretical *S/P* ratio, assuming a salivary pH of 6.8, is 1.22. Since codeine has a pK_a of 8.2 and is only bound to plasma proteins to a minor extent, it has a higher theoretical *S/P* ratio of 3.32.

As for cocaine, heroin, and its metabolites 6-AM and morphine are detected in oral fluid after administration of heroin by different routes; concentrations are depending on the route of administration and on the salivary pH. For opiates, their levels in oral fluid appear to mimic blood levels suggesting that comparable results between these matrices can be achieved [76, 77]. The proposed SAMHSA cutoffs are 40 ng ml^{-1} for screening and confirmation, except for 6-AM, for which a 10-fold lower confirmation cutoff value was suggested. Laboratories are also permitted to initial test all specimens for 6-AM using a 4 ng ml^{-1} cutoff. Standards Australia recommends a screening target concentrations of 50 ng ml^{-1} with confirmation at 25 ng ml^{-1} for morphine and codeine and 10 ng ml^{-1} for 6-AM (Table 1).

The number of studies concerning heroin are limited. Heroin is detected within minutes in oral fluid following nasal insufflation (snorting). A 12-mg

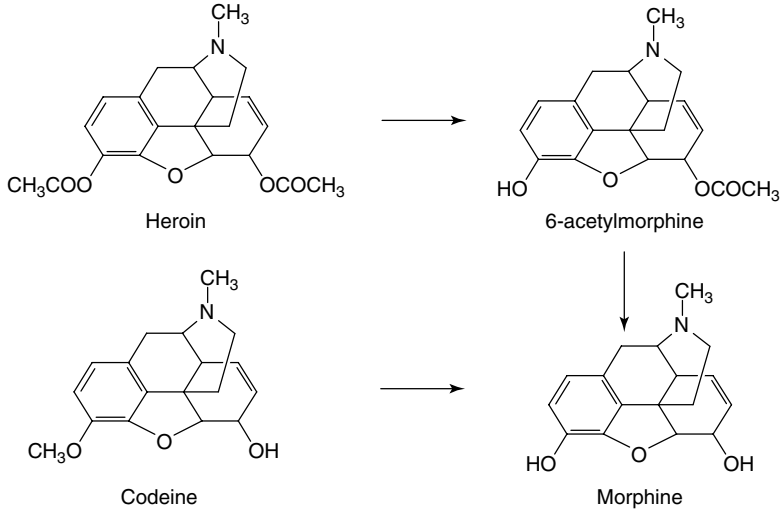


Figure 7 Major metabolic route for heroin

dose resulted in a peak heroin concentration of 300 ng ml^{-1} in oral fluid and it was detectable for about 1 h (LOD: 1.0 ng ml^{-1}). 6-AM and morphine were rapidly detected in oral fluid peaking at 60 and 25 ng ml^{-1} at 10 and 60 min, respectively [78]. Two subjects given 2.6–10.5 mg smoked doses revealed peak concentrations of heroin greater than 3000 ng ml^{-1}, but less than 10 ng ml^{-1} beyond 60–120 min. Much lower peak concentrations were detected after i.v. injection of heroin: up to 30 ng ml^{-1} after 3–12 mg doses. The mean *S/P* ratio was approximately 1 [76].

Intramuscular morphine sulfate at doses of 10 and 20 mg resulted in a peak concentration of 11 and 38 ng ml^{-1} in two subjects, respectively, using a stimulated collection process [79]. At 24 h, concentrations were barely detectable.

Codeine presence in oral fluid has been studied in four studies [15, 37, 80, 81]. Screening/confirmation cutoffs of 20/20 [80] and 30/40 ng ml^{-1} [37] have been proposed based on two studies with volunteers given 60 and 120 mg codeine.

Substantially different pharmacokinetic parameters were obtained for codeine with different collection methods [15]. After stimulation with citric acid, mean peak concentrations of 556 ± 301, 639 ± 64, and 1599 ± 241 ng ml^{-1} were observed following 30 mg liquid codeine phosphate and 60 and 120 mg codeine sulfate capsules, respectively [15, 81]. Although a high intra- and intersubject variability was observed, the time course of codeine concentration in oral fluid was similar to that seen in plasma [81]. Generally, codeine concentrations were higher in oral fluid than in plasma and codeine could be detected 5–9 h longer in oral fluid than in plasma when the LOQ of the GC-MS method (2.5 ng ml^{-1}) was used as cutoff value. Norcodeine was also detected in small amounts in oral fluid.

Dihydrocodeine is a weak narcotic analgesic and used as an antitussive [82]. Following a single 60-mg dose, peak oral fluid concentrations ranged from 423 to 1421 ng ml^{-1}, reached 2–4 h after dose. The concentrations of this compound in oral fluid samples from 20 chronic users (0.4–2.7 g daily) within 24 h after the last dose ranged up to 66 154 ng ml^{-1}. Samples were collected with the Salivette®. Dihydrocodeine was detectable until about 24 h after the last use with an LOD of 5 ng ml^{-1}.

Finally, one study was conducted to determine the concentrations of morphine in oral fluid and urine following the consumption of commercially available poppy seeds [83]. Oral fluid samples from volunteers who ingested in addition to one poppy seed bagel as many possible poppy seeds as possible (amounts ranging from 9.82 to 20.82 g), tested positive up to 1 h after ingestion with a concentration greater than the proposed SAMHSA confirmation cutoff for morphine.

Benzodiazepines

The benzodiazepines are a large class of prescribed drugs, used widely for different medical conditions such as the treatment of insomnia (mostly short-acting analogues), the treatment of anxiety-related conditions (mostly longer acting analogues), as muscle relaxants, and as anticonvulsants [84–86]. Some of them are abused, either by themselves or in combination with other drugs, particularly the narcotic analgesics (opiates). Their ability to suppress withdrawal symptoms and to boost the effects of heroin have made them a favored drug type among the drug-using population [87, 88]. They are also abused by cocaine users to increase the seizure threshold [84].

The benzodiazepine structure is based on the diazepine ring fused to benzene, with the nitrogens usually in the 1,4 positions. Substitution leads to varying degrees of metabolic stability, pharmacological potency, and lipid solubility (Figure 8). The 7-nitro substituted 1,4-benzodiazepines, such as flunitrazepam, are often abused. Somewhat newer analogues are based on the diazolo- and triazolo-ring types and include alprazolam and triazolam, very potent benzodiazepines requiring much lower doses. Zolpidem is an imidazopyridine derivative, with a chemical structure unrelated to benzodiazepines.

In general, *S/P* ratios below 1 are expected due to the extensive plasma protein binding of benzodiazepines.

Until now, five recent studies have examined the presence of benzodiazepines in oral fluid. Samyn *et al.* reported peak concentrations of flunitrazepam (0.29–0.58 ng ml^{-1}) and its metabolite 7-aminoflunitrazepam (0.94–3.05 ng ml^{-1}) at 1–4.5 h following the administration of 1 mg flunitrazepam to four subjects [89]. Owing to the extremely low concentrations of flunitrazepam and its metabolite in oral fluid, GC-MS in chemical ionization mode was used (LOQ: 0.10 and 0.15 ng ml^{-1}, respectively). Sodium fluoride preservative was required to detect flunitrazepam in the oral fluid specimens. These authors proposed a cutoff value of approximately 0.25 ng ml^{-1} of 7-aminoflunitrazepam in oral fluid to detect use of flunitrazepam within the last 6 h.

Following the ingestion of a single oral dose of 2.5 mg lorazepam to one volunteer, a peak concentration of 18 ng ml^{-1} was observed in oral fluid after 15 min of intake. Oral fluid samples were still positive for this compound after 8 h (0.3 ng ml^{-1}) [90].

Peak concentrations of midazolam in oral fluid samples following a single 2 mg bolus of the drug to three patients ranged from 1.42 to 13.98 ng ml^{-1} after 10–40 h [20]. However, substantially higher concentrations of midazolam in these specimens were obtained after continuous infusion of the drug.

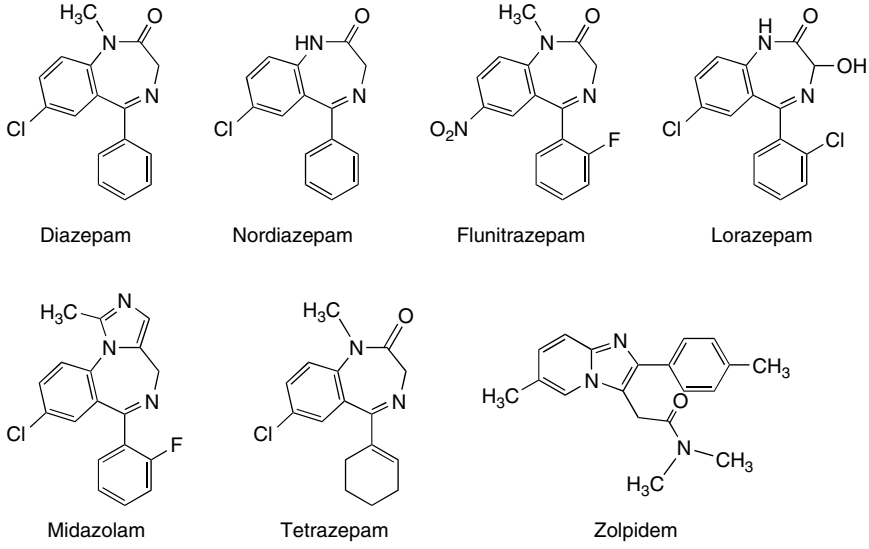


Figure 8 Structures of selected benzodiazepines and zolpidem

Oral administration of a single 50-mg dose of tetrazepam to one volunteer produced a peak concentration of 6.6 ng ml^{-1} after 150 min, which is approximately the time of peak plasma concentration [91]. Oral fluid tested positive for tetrazepam over 8.6 h with an LOQ of the LC-MS/MS method of 0.03 ng ml^{-1}.

After oral administration of 10 mg zolpidem to two subjects, peak concentrations of this drug were obtained after 150 min (53.5 ng ml^{-1}) and 180 min (75.7 ng ml^{-1}), respectively [92]. Oral fluid still tested positive for zolpidem at 8 h (9–15 ng ml^{-1}).

As suggested, an *S/P* ratio lower than 1 was obtained for the studied benzodiazepines [90–92].

Applications

Oral fluid drug testing has been used as an alternative matrix to blood and urine in clinical and forensic toxicology [93]. Owing to its particular advantages, interest in the use of oral fluid testing for drugs in the workplace [29, 94] and at the roadside [23, 95–97] is increasing rapidly. The use of oral fluid in drug treatment facilities [98] and for therapeutic drug monitoring [99–101] has also been reported.

Workplace Drug Testing

It is clear that laboratory-based analytical methods to measure drugs in oral fluid are robust and well developed. Many published studies have examined the known or potential rate of error of these tests (sensitivity, specificity, positive predictive value, negative predictive value, overall accuracy). There is certainly an acceptance of the accuracy and reliability of oral fluid testing within the scientific community. Unfortunately, there are no formal laboratory certification programs yet established for drug testing in oral fluid [102]. In April 2004, SAMHSA has published in the Federal Register a Proposed Rule for incorporating oral fluid in its federal workplace drug-testing program. This was followed recently by an Australian standard that is also intended for workplace drug testing (Table 1). In the United Kingdom, the UK Workplace Drug-testing forum is currently working on guidelines for oral fluid specimen analy-sis. However, to date, no published document is available. Alltrix, the largest specialist oral fluid drug-testing laboratory in Europe is using its own cutoff levels, which are based on the levels proposed by the manufacturer of the immunoassays (OraSure Technologies) for the screening assays (Clarke, J. Altrix Healthcare, *Personal Communication*, January 2007). All samples are collected using the Intercept® device.

The results of the first large-scale database on oral fluid testing in private industry were published by Cone *et al.* [29]. A total of 77 000 specimens were screened by the Intercept® immunoassay at manufacturer's recommended cutoff values for five drug categories (marijuana, cocaine, opiates, phencyclidine, and amphetamines). Presumptive positive specimens were confirmed by GC-MS. An overall positive rate of 5.06% was noted. The confirmed positive specimens consisted primarily of THC and cocaine (85.75%). In comparison to the corresponding urine samples, the prevalence of positives for cocaine and amphetamines was approximately 60% higher in oral fluid. Of the 48 morphine-positive specimens, 32 specimens were also positive for 6-AM. Considering the short detection window for this unique metabolite in urine (approximately 3.3 h following single dose administration of heroin), it is concluded that oral fluid is superior to urine in the confirmation of recent use of heroin [29, 103]. After a single administration of THC by inhalation, there was a substantially higher rate of detection of positive oral fluid specimens over the first 8 h in comparison to the corresponding urine testing [57]. This unique characteristic is of great importance when one considers the short time course of marijuana's effect on performance. In general, oral fluid is considered as the only other body fluid revealing recent drug use and therefore related to possible impairment, whereas urine testing is clearly more useful to detect past use of the drug.

Both SAMHSA and Standards Australia have also established guidelines for the collection and storage of specimens. For SAMHSA, the suggested minimum amount of specimen to be collected is 2 ml, which can be divided into 1.5 ml for the primary specimen and 0.5 ml for retesting. SAMHSA is presently recommending oral fluid sampling by spitting into a neat tube, whereas for Standards Australia specimens either collected by spitting or through the use of a device are accepted. However, to overcome the problems of adsorption of certain drugs to the devices (as discussed previously), each manufacturer is required to document recovery of all target analytes from the devices. In addition, some oral fluid specimen collection schemes involve collection on an absorbent swab

with subsequent transfer into a buffer, an issue that can cause difficulties for the interpretation of quantitative results. Another problem is that many devices cause either a mechanical or chemical stimulation of oral fluid flow. As already discussed above, this process affects the pH of oral fluid, and therefore the concentrations of drugs and metabolites in the oral fluid sample.

Owing to a lack of resolution of several of the above mentioned important research questions, SAMHSA has proposed a conservative approach to the addition of oral fluid testing to the different drug-testing matrices. In its proposed rules, SAMHSA noted that less is known about the pharmacokinetics and disposition of drugs in oral fluid as compared to urine. However, several authors have investigated these pharmacokinetics, and, therefore, it is important to note that although it may not be possible to answer some of these questions definitively, qualified experts may nonetheless be able to provide sufficient information to assist in its decision-making. The new SAMHSA guidelines also recommend collecting urine specimens simultaneously with oral fluid, because of the potential of passive contamination for THC, as discussed previously. However, the recent demonstration by Moore *et al.* and Day *et al.* of the presence of THC-COOH in the oral fluid matrix may eliminate the necessity of the additional urine sampling [61, 62].

Substance Abuse Monitoring

Drug testing of patients in treatment is generally practiced to provide objective assessment of patients' progress. Many treatment programs require individuals to detoxify from the offending drug(s) and remain drug-free throughout treatment. Other programs utilize substitution therapy in which the patient is switched from an illicit drug, e.g., heroin, to an acceptable substitute, e.g., methadone or buprenorphine, which is administered by treatment personnel. In most of the programs, patients' progress is monitored by urinalysis. However, interest in oral fluid as an alternative drug-testing tool has grown in this area. Oral fluid testing appears to be a good indicator of the validity and reliability of drug user's self-report data [101].

Dihydrocodeine, codeine, morphine, 6-AM, heroin, cocaine, BE, methadone, amphetamine, MDA, MDMA, MDEA were measured by GC-MS in oral fluid of donors being monitored in a drug misuse treatment program as part of the investigators' effort to evaluate the Cozart® RapiScan screening device and the Cozart® Microplate EIA [33–36, 104].

The concordance of oral fluid and urine opiate results for participants in drug withdrawal therapy was assessed [105]. Oral fluid specimens were collected with a Clin Rep® device from Recipe and analyzed by GC-MS for dihydrocodeine, codeine, 6-AM, and morphine. The authors found a good correlation between opiate results in both matrices, in specimens collected every 2–4 days over a 3-month period. They concluded that oral fluid was appropriate for analysis of drugs of abuse. Niedbala *et al.* [28] collected oral fluid and urine specimens from known opiate abusers in a drug treatment center and found an agreement of more than 90%, applying a cutoff of 10 ng ml^{-1} for one or more opiates in the collected oral fluid and a 2000 ng ml^{-1} cutoff for interpretation of the urine results. This was confirmed by Bennett *et al.* [98] comparing the accuracy of on-site testing of oral fluid with urine on-site testing in 157 drug-dependant persons and concluded that oral fluid testing is as accurate as urinalysis in detecting the presence of opiates and methadone.

Recently, Dams *et al.* [106] analyzed oral fluid specimens from 16 pregnant opiate-dependent women. Specimens were collected with the Salivette® oral fluid collection device. 6-AM, heroin, and morphine were the major opiates detected with median concentrations of 5.2, 2.3, and 7.5 ng ml^{-1}, respectively. Detection of 6-AM and/or heroin in 84% of positive oral fluid specimens proved heroin usage, rather than licit codeine or morphine, or ingestion of opiate-containing foodstuffs. In contrast, only 37 urine specimens collected the same day in the same subjects as the oral fluid samples were positive at a 300 ng ml^{-1} cutoff concentration of opiates in urine. Cocaine and BE had median concentrations of 6.4 and 3.4 ng ml^{-1}. The major analytes identified were cocaine and BE, detected in approximately 98% of all cocaine positive specimens. In contrast, only 22 urine specimens were positive for cocaine metabolites at a 300 ng ml^{-1} cutoff. These authors conclude that oral fluid is a promising alternative matrix to monitor opiate and cocaine use in drug-testing programs.

Acetylcodeine has also been reported as a marker of illicit heroin abuse in oral fluid samples [107]. A significant correlation was found between this new

marker and 6-AM ($r = 0.95$) in samples collected from patients attending a substance abuse clinic.

Oral fluid has also been examined as an alternative matrix for the therapeutic monitoring of methadone [108, 109]. A poor correlation between oral fluid and serum was obtained when considering total methadone concentrations [108]. However, a good correlation was found between both matrices when the enantiomeric ratios of methadone were taken into account. Methadone and its major metabolite, 2-ethylidine-1,5-dimethyl-3,3-diphenylpyrrolidine (EDDP), were examined in oral fluid and plasma as a function of salivary pH [110]. Oral fluid methadone ranged from 120 to 3460 ng ml^{-1} with *S/P* ratios from 0.6 to 7.2. For EDDP the ranges were 40–100 ng ml^{-1} for oral fluid, with an *S/P* ratio of 0.2–1.8. An inverse correlation was found between the methadone concentration and the oral fluid pH.

Roadside Drug Testing

While there is only limited knowledge of the prevalence of drugs other than alcohol in road traffic, it appears that drugged driving is a significant problem worldwide [111–113]. Since the late 1990s, several western countries have, in analogy to alcohol, introduced *per se* legislation for driving under the influence of drugs (DUID). The policy is in whole or in part based on the detection of any amount of illicit drug in blood or urine of the driver. However, a positive urine test is by no means an indication that the subject was "under the influence" at the time of sampling. The presence of certain illicit drugs or their metabolites in urine of potentially impaired drivers is merely evidence of relatively recent exposure, except for certain drugs such as cannabis. The urinary metabolite of THC can be detected in urine for days, or even weeks, after the last use in regular users. When the drug is detected in blood, there is a higher probability that the subject is experiencing pharmacological effects at the time of sampling. A recently published controlled study has approached the relatively obscure relation between THC concentration in serum and driver impairment [60]. It was concluded that a serum THC concentration of 2–5 ng ml^{-1} was associated to the limit for impairment in performance tests measuring skills related to driving.

Oral fluid is probably the only other body fluid that might parallel blood in some regards and therefore may be related to behavioral performance [4, 114]. Using receiver operating characteristic (ROC) analysis of data from 139 individuals suspected of driving under the influence of cannabis, Laloup *et al.* [115] calculated an optimal cutoff value of 1.2 ng ml^{-1} THC in preserved oral fluid to predict a positive plasma result (LOQ 0.5 ng ml^{-1}) (sensitivity: 94.7% and specificity: 92.0%). Several studies have shown that a reasonably good correlation can be found between the presence or absence of drugs in oral fluid and in blood [8, 23, 116, 117]. Moreover, oral fluid appears to be superior to urine in correlating with serum analytical data and impairment symptoms of drivers under the influence of drugs of abuse [95]. Epidemiological data from drivers (randomly stopped) and accident-involved drivers, respectively, have been collected by analysis of oral fluid samples in the context of the European project IMMORTAL (Impaired Motorists, Methods of Roadside Testing, and Assessment for Licensing) [97, 118].

Since many years, police officers involved in road safety have expressed the need for a rapid and reliable drug test that can be applied at the roadside. The on-site urine drug tests and the first generation of the oral fluid tests were thoroughly evaluated in the DG VII EC project ROSITA (Roadside Testing Assessment) [112]. At that time, three on-site tests for oral fluid screening were available: Drugwipe® (Securetec, Germany), RapiScan® (Cozart Bioscience, UK), and Oralscreen® (Avitar, US). A clear majority of the participants preferred oral fluid as the matrix of choice for on-site testing. The possibility of collecting oral fluid by nonmedical personnel and without the need for special facilities (i.e., sanitary van) was a distinct advantage. However, police officers and researchers encountered problems related to insufficient sample volume, the viscosity of the samples and insufficient sensitivity of the analytical methods. The on-site oral fluid screenings, particularly for testing of THC, required significant improvement [96].

As has been mentioned earlier in this article, steady progress has been made in the last five years in the field of sample collection and laboratory-based immunoassays and confirmation techniques. Most of the newer generation on-site devices offer the possibility to collect enough sample volume (sometimes diluted with buffer) that can be used

for confirmation analysis in the laboratory. In 2002, a compilation of the analytical advantages and test performances of out-of-lab drug-testing devices was published [119]. Several authors have compared the use of some of these on-site devices with laboratory results, including Cozart® RapiScan [37, 38, 104, 120–122], Drugwipe® [23, 123, 124], Toxiquick® [125], Dräger® DrugTest [115], Oratect® [126], and Oraline® [127]. However, the absence of internationally accepted cutoff values for oral fluid hampers the comparison of on-site test results with chromatographic results.

In an evaluation of six on-site test devices using spiked oral fluid samples [128, 129], it appeared that theoretical cutoff values for a specific analyte are considerably different across devices. The ability to accurately and reliably detect cocaine and amphetamine was dependent on the individual device. Most devices performed well for the detection of methamphetamine and opiates, but all performed poorly for the detection of THC. Two reasons appear to be apparent: (i) the devices target the wrong analyte (THC-COOH instead of THC) and (ii) the cutoff concentrations are too high considering the low concentrations of THC generally present in oral fluid.

In late 2003, the Rosita-2 project was started, involving six European countries and five US states [111]. Nine devices were evaluated: American Biomedica Oralstat, Branan Medical Oratect, Cozart Bioscience RapiScan (only in the United States), Dräger/Orasure DrugTest/Uplink, Lifepoint Impact, Securetec Drugwipe, Sun Biomedical Oraline, Ultimed Salivascreen, and Varian OraLab. For six devices, the number of failed test runs exceeded 25%, and their evaluation was stopped. During the study, two devices were withdrawn from the market: the Dräger DrugTest and Lifepoint Impact. At the end of the study, based on the analytical evaluation (comparison to reference methods in oral fluid and/or blood), no device was considered reliable enough to be recommended for roadside screening of drivers. The sensitivity of the on-site screening devices for cannabis and benzodiazepines needs to be improved dramatically. For example, the performance of the Dräger DrugTest to screen for THC in oral fluid was evaluated using oral fluid and plasma as reference samples [115]. Whereas the specificity of the test was high (93–100%), a sensitivity of only 50% was calculated, indicating a high number of false-negative test results. Despite these drawbacks, in a few countries, legislation was passed that allows the potential use of oral fluid as a matrix for screening or confirmation [96].

In the state of Victoria in Australia, legislation introduced in December 2003 allows police to perform random roadside oral fluid testing for cannabis, methamphetamine, and MDMA [130]. Police officers use the Drugwipe II® device (Securetec) to do the initial screening by wiping the tongue, when the driver is still in his car. If this test shows a positive result, for cannabis, methamphetamine, or MDMA, an oral fluid sample is requested so the police can do an immediate screening on-site with the Cozart RapiScan system for the same drug classes. When this second test is positive for either drug, the oral fluid sample is confirmed in the laboratory for amphetamines and THC using predefined cutoff values, and the driver can take the split sample with him. From the 13 176 drug tests performed at the roadside, 313 positive cases were selected for GC-MS analysis (2.4%). A high number of positive methamphetamine ($n = 269$; median: 1136 ng ml^{-1}) and MDMA ($n = 118$; median: 2724 ng ml^{-1}) findings are reported, and a relatively low number of THC positives ($n = 87$; median: 81 ng ml^{-1}) [131]. It would be interesting to analyze the data further with respect to THC levels in those cases where both oral fluid tests were positive for THC, and to know the prevalence of THC-only positives. Based on the results of Rosita-2, the relatively good performance of the test devices to pick up amphetamines should reflect in a high number of methamphetamine positive tests and the low sensitivity to detect THC in oral fluid, should lead to a serious underestimation of the number of cannabis positives. However, the aim of this type of random roadside drug testing is to have a deterrent effect on drugged drivers and ultimately decrease the accident risk. One major risk is the possibility that drivers will realize that they often test negative after having used cannabis.

Conclusion

Oral fluid appears to sufficiently meet the requirements to be added to workplace drug-testing laboratory-based programs. There are adequate methods available for screening and confirmation in the laboratory and the relevant drugs and metabolites have been identified. External proficiency testing is

being developed and cutoff concentrations have been proposed by scientific organizations. Appropriate certification of laboratories worldwide and introduction of legal provisions in several countries is the next step. In different environments such as drug treatment facilities, oral fluid has become an interesting alternative matrix for drug testing. Advantages of oral fluid testing include noninvasive sample collection, and the ability to collect the sample under observation reducing the opportunity for adulteration. The potential to estimate circulating drug concentrations and therefore closely associated to 'being under the influence', make it the matrix of choice for police officers and law enforcement agencies for screening of intoxicated drivers. However, the search for an on-site screening method that can provide acceptable accuracies for the detection of cannabinoids in oral fluid remains a major hurdle. Additional considerations that have received attention are the variability in the volume of sample collected and the drug recovery from the many different specimen collection systems on the market.

References

[1] Mandel, I.D. (1993). A contemporary view of salivary research, *Critical Reviews in Oral Biology and Medicine* **4**, 599.

[2] Veerman, E.C.I., van den Keijbus, P.A.M., Vissink, A. & van Nieuw Amerongen, A. (1996). Human glandular salivas: their separate collection and analysis, *European Journal of Oral Sciences* **104**, 346.

[3] Young, J.A. & Van Lennep, E.W. (1978). *The Morphology of Salivary Glands*. Academic Press, London.

[4] Samyn, N., Verstraete, A., van Haeren, C. & Kintz, P. (1999). Analysis of drugs of abuse in saliva, *Forensic Science Review* **11**, 1.

[5] Aps, J.K. & Martens, L.C. (2005). Review: the physiology of saliva and transfer of drugs into saliva, *Forensic Science International* **150**, 119.

[6] Mucklow, J.C., Bending, M.R., Kahn, G.C. & Dollery, C.T. (1978). Drug concentration in saliva, *Clinical Pharmacology and Therapeutics* **24**, 563.

[7] Idowu, O.R. & Caddy, B. (1982). A review of the use of saliva in the forensic detection of drugs and other chemicals, *Journal of the Forensic Science Society* **22**, 123.

[8] Navarro, M., Pichini, S., Farre, M., Ortuno, J., Roset, P.N., Segura, J. & de la, T.R. (2001). Usefulness of saliva for measurement of 3,4-methylenedioxymethamphetamine and its metabolites: correlation with plasma drug concentrations and effect of salivary pH, *Clinical Chemistry* **47**, 1788.

[9] Samyn, N., De Boeck, G., Wood, M., Lamers, C.T., de Waard, D., Brookhuis, K.A., Verstraete, A.G. & Riedel, W.J. (2002). Plasma, oral fluid and sweat wipe ecstasy concentrations in controlled and real life conditions, *Forensic Science International* **128**, 90.

[10] Schepers, R.J., Oyler, J.M., Joseph Jr, R.E., Cone, E.J., Moolchan, E.T. & Huestis, M.A. (2003). Methamphetamine and amphetamine pharmacokinetics in oral fluid and plasma after controlled oral methamphetamine administration to human volunteers, *Clinical Chemistry* **49**, 121.

[11] Pichini, S. (2005). *Distribution of 3,4-methylenedioxymethamphetamine (MDMA) in Non Conventional Matrices and its Applications in Clinical Toxicology*. Doctoral Thesis, Universitat Autonoma de Barcelona, Spain.

[12] Huestis, M.A. & Cone, E.J. (2004). Relationship of delta 9-tetrahydrocannabinol concentrations in oral fluid and plasma after controlled administration of smoked cannabis, *Journal of Analytical Toxicology* **28**, 394.

[13] Choo, R.E. & Huestis, M.A. (2004). Oral fluid as a diagnostic tool, *Clinical Chemistry and Laboratory Medicine* **42**, 1273.

[14] Kato, K., Hillsgrove, M., Weinhold, L., Gorelick, D.A., Darwin, W.D. & Cone, E.J. (1993). Cocaine and metabolite excretion in saliva under stimulated and nonstimulated conditions, *Journal of Analytical Toxicology* **17**, 338.

[15] O'Neal, C.L., Crouch, D.J., Rollins, D.E. & Fatah, A.A. (2000). The effects of collection methods on oral fluid codeine concentrations, *Journal of Analytical Toxicology* **24**, 536.

[16] Crouch, D.J. (2005). Oral fluid collection: the neglected variable in oral fluid testing, *Forensic Science International* **150**, 165.

[17] Quintela, O., Crouch, D.J. & Andrenyak, D.M. (2006). Recovery of drugs of abuse from the Immunalysis Quantisal oral fluid collection device, *Journal of Analytical Toxicology* **30**, 614.

[18] Campora, P., Bermejo, A.M., Tabernero, M.J. & Fernandez, P. (2003). Quantitation of cocaine and its major metabolites in human saliva using gas chromatography-positive chemical ionization-mass spectrometry (GC-PCI-MS), *Journal of Analytical Toxicology* **27**, 270.

[19] Campora, P., Bermejo, A.M., Tabernero, M.J. & Fernandez, P. (2006). Use of gas chromatography/mass spectrometry with positive chemical ionization for the determination of opiates in human oral fluid, *Rapid Communications in Mass Spectrometry* **20**, 1288.

[20] Quintela, O., Cruz, A., Concheiro, M., de Castro, A. & Lopez-Rivadulla, M. (2004). A sensitive, rapid and specific determination of midazolam in human plasma and saliva by liquid chromatography/electrospray mass spectrometry, *Rapid Communications in Mass Spectrometry* **18**, 2976.

[21] Gunnar, T., Ariniemi, K. & Lillsunde, P. (2005). Validated toxicological determination of 30 drugs of

abuse as optimized derivatives in oral fluid by long column fast gas chromatography/electron impact mass spectrometry, *Journal of Mass Spectrometry* **40**, 739.

[22] Kauert, G.F., Iwersen-Bergmann, S. & Toennes, S.W. (2006). Assay of delta9-tetrahydrocannabinol (THC) in oral fluid-evaluation of the OraSure oral specimen collection device, *Journal of Analytical Toxicology* **30**, 274.

[23] Samyn, N., De Boeck, G. & Verstraete, A.G. (2002). The use of oral fluid and sweat wipes for the detection of drugs of abuse in drivers, *Journal of Forensic Sciences* **47**, 1380.

[24] Teixeira, H., Proenca, P., Verstraete, A., Corte-Real, F. & Vieira, D.N. (2005). Analysis of delta9-tetrahydrocannabinol in oral fluid samples using solid-phase extraction and high-performance liquid chromatography-electrospray ionization mass spectrometry, *Forensic Science International* **150**, 205.

[25] Dickson, S., Park, A., Nolan, S., Kenworthy, S., Nicholson, C., Midgley, J., Pinfold, R. & Hampton, S. (2006). The recovery of illicit drugs from oral fluid sampling devices, *Forensic Science International* **165**, 78.

[26] Concheiro, M., de Castro, A., Quintela, O., Lopez-Rivadulla, M. & Cruz, A. (2005). Determination of MDMA, MDA, MDEA and MBDB in oral fluid using high performance liquid chromatography with native fluorescence detection, *Forensic Science International* **150**, 221.

[27] Moore, C., Vincent, M., Rana, S., Coulter, C., Agrawal, A. & Soares, J. (2006). Stability of delta(9)-tetrahydrocannabinol (THC) in oral fluid using the Quantisal™ collection device, *Forensic Science International* **164**, 126.

[28] Niedbala, R.S., Kardos, K., Waga, J., Fritch, D., Yeager, L., Doddamane, S. & Schoener, E. (2001). Laboratory analysis of remotely collected oral fluid specimens for opiates by immunoassay, *Journal of Analytical Toxicology* **25**, 310.

[29] Cone, E.J., Presley, L., Lehrer, M., Seiter, W., Smith, M., Kardos, K.W., Fritch, D., Salamone, S. & Niedbala, R.S. (2002). Oral fluid testing for drugs of abuse: positive prevalence rates by Intercept immunoassay screening and GC-MS-MS confirmation and suggested cutoff concentrations, *Journal of Analytical Toxicology* **26**, 541.

[30] Drummer, O.H. (2005). Review: pharmacokinetics of illicit drugs in oral fluid, *Forensic Science International* **150**, 133.

[31] Laloup, M., De Boeck, G. & Samyn, N. Unconventional samples and alternative matrices, in *Handbook of Analytical Separations*, M.J. Bogusz, ed, Elsevier, Amsterdam, Vol. 2, (2008).

[32] Clarke, J. & Wilson, J.F. (2005). Proficiency testing (external quality assessment) of drug detection in oral fluid, *Forensic Science International* **150**, 161.

[33] Cooper, G., Wilson, L., Reid, C., Baldwin, D., Hand, C. & Spiehler, V. (2005). Comparison of GC-MS and EIA results for the analysis of methadone in oral fluid, *Journal of Forensic Sciences* **50**, 928.

[34] Cooper, G., Wilson, L., Reid, C., Hand, C. & Spiehler, V. (2006). Validation of the Cozart((R)) amphetamine microplate EIA for the analysis of amphetamines in oral fluid, *Forensic Science International* **159**, 104.

[35] Cooper, G., Wilson, L., Reid, C., Baldwin, D., Hand, C. & Spiehler, V. (2005). Validation of the Cozart microplate EIA for analysis of opiates in oral fluid, *Forensic Science International* **154**, 240.

[36] Cooper, G., Wilson, L., Reid, C., Baldwin, D., Hand, C. & Spieher, V. (2004). Validation of the Cozart microplate EIA for cocaine and metabolites in oral fluid, *Journal of Analytical Toxicology* **28**, 498.

[37] Kacinko, S.L., Barnes, A.J., Kim, I., Moolchan, E.T., Wilson, L., Cooper, G.A., Reid, C., Baldwin, D., Hand, C.W. & Huestis, M.A. (2004). Performance characteristics of the Cozart RapiScan Oral Fluid Drug Testing System for opiates in comparison to ELISA and GC/MS following controlled codeine administration, *Forensic Science International* **141**, 41.

[38] Kolbrich, E.A., Kim, I., Barnes, A.J., Moolchan, E.T., Wilson, L., Cooper, G.A., Reid, C., Baldwin, D., Hand, C.W. & Huestis, M.A. (2003). Cozart RapiScan Oral Fluid Drug Testing System: an evaluation of sensitivity, specificity, and efficiency for cocaine detection compared with ELISA and GC-MS following controlled cocaine administration, *Journal of Analytical Toxicology* **27**, 407.

[39] Laloup, M., Tilman, G., Maes, V., De Boeck, G., Wallemacq, P., Ramaekers, J. & Samyn, N. (2005). Validation of an ELISA-based screening assay for the detection of amphetamine, MDMA and MDA in blood and oral fluid, *Forensic Science International* **153**, 29.

[40] Niedbala, R.S., Kardos, K., Fries, T., Cannon, A. & Davis, A. (2001). Immunoassay for detection of cocaine/metabolites in oral fluids, *Journal of Analytical Toxicology* **25**, 62.

[41] Kim, I., Barnes, A.J., Schepers, R., Moolchan, E.T., Wilson, L., Cooper, G., Reid, C., Hand, C. & Huestis, M.A. (2003). Sensitivity and specificity of the Cozart microplate EIA cocaine oral fluid at proposed screening and confirmation cutoffs, *Clinical Chemistry* **49**, 1498.

[42] Cognard, E., Bouchonnet, S. & Staub, C. (2006). Validation of a gas chromatography-ion trap tandem mass spectrometry for simultaneous analyse of cocaine and its metabolites in saliva, *Journal of Pharmaceutical and Biomedical Analysis* **41**, 925.

[43] Oiestad, E.L., Johansen, U. & Christophersen, S. (2006). Drug screening of preserved oral fluid by liquid chromatography-tandem mass spectrometry, *Clinical Chemistry* **53**, 300–309.

[44] Clauwaert, K., Decaestecker, T., Mortier, K., Lambert, W., Deforce, D., Van Peteghem, C. & Van Bocxlaer, J. (2004). The determination of cocaine, benzoylecgonine, and cocaethylene in small-volume oral

fluid samples by liquid chromatography-quadrupole-time-of-flight mass spectrometry, *Journal of Analytical Toxicology* **28**, 655.

[45] Concheiro, M., de Castro, A., Quintela, O., Cruz, A. & Lopez-Rivadulla, M. (2004). Development and validation of a method for the quantitation of delta9tetrahydrocannabinol in oral fluid by liquid chromatography electrospray-mass-spectrometry, *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences* **810**, 319.

[46] Dams, R., Murphy, C.M., Choo, R.E., Lambert, W.E., De Leenheer, A.P. & Huestis, M.A. (2003). LC-atmospheric pressure chemical ionization-MS/ MS analysis of multiple illicit drugs, methadone, and their metabolites in oral fluid following protein precipitation, *Analytical Chemistry* **75**, 798.

[47] Kintz, P., Villain, M., Concheiro, M. & Cirimele, V. (2005). Screening and confirmatory method for benzodiazepines and hypnotics in oral fluid by LC-MS/MS, *Forensic Science International* **150**, 213.

[48] Laloup, M., Ramirez Fernandez, M.M., Wood, M., De Boeck, G., Henquet, C., Maes, V. & Samyn, N. (2005). Quantitative analysis of delta9-tetrahydrocannabinol in preserved oral fluid by liquid chromatography-tandem mass spectrometry, *Journal of Chromatography A* **1082**, 15.

[49] Mortier, K.A., Maudens, K.E., Lambert, W.E., Clauwaert, K.M., Van Bocxlaer, J.F., Deforce, D.L., Van Peteghem, C.H. & De Leenheer, A.P. (2002). Simultaneous, quantitative determination of opiates, amphetamines, cocaine and benzoylecgonine in oral fluid by liquid chromatography quadrupole-time-of-flight mass spectrometry, *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences* **779**, 321.

[50] Wood, M., De Boeck, G., Samyn, N., Morris, M., Cooper, D.P., Maes, R.A. & de Bruijn, E.A. (2003). Development of a rapid and sensitive method for the quantitation of amphetamines in human plasma and oral fluid by LC-MS-MS, *Journal of Analytical Toxicology* **27**, 78.

[51] Mortier, K.A., Clauwaert, K.M., Lambert, W.E., Van Bocxlaer, J.F., Van den Eeckhout, E.G., Van Peteghem, C.H. & De Leenheer, A.P. (2001). Pitfalls associated with liquid chromatography/electrospray tandem mass spectrometry in quantitative bioanalysis of drugs of abuse in saliva, *Rapid Communications in Mass Spectrometry* **15**, 1773.

[52] Kidwell, D.A., Holland, J.C. & Athanaselis, S. (1998). Testing for drugs of abuse in saliva and sweat, *Journal of Chromatography. B, Biomedical Sciences and Applications* **713**, 111.

[53] Dams, R., Huestis, M.A., Lambert, W.E. & Murphy, C.M. (2003). Matrix effect in bio-analysis of illicit drugs with LC-MS/MS: influence of ionization type, sample preparation, and biofluid, *Journal of the American Society for Mass Spectrometry* **14**, 1290.

[54] Wood, M., Laloup, M., Ramirez Fernandez, M.M., Jenkins, K.M., Young, M.S., Ramaekers, J.G., De Boeck, G. & Samyn, N. (2005). Quantitative analysis of multiple illicit drugs in preserved oral fluid by solid-phase extraction and liquid chromatography-tandem mass spectrometry, *Forensic Science International* **150**, 227.

[55] Department of Health and Human Services Substance Abuse and Mental Health Services Administration (2004). *Proposed Revisions to Mandatory Guidelines for Federal Workplace Drug Testing Programs*, Vol. 69, FR 19673.

[56] Standards Australia *Australian Standard® – Procedures for Specimen Collection and the Detection and Quantitation of Drugs in Oral Fluid*, AS4760-2006, available from www.standards.org.au.

[57] Niedbala, R.S., Kardos, K.W., Fritch, D.F., Kardos, S., Fries, T., Waga, J., Robb, J. & Cone, E.J. (2001). Detection of marijuana use by oral fluid and urine analysis following single-dose administration of smoked and oral marijuana, *Journal of Analytical Toxicology* **25**, 289.

[58] Niedbala, R.S., Kardos, K.W., Fritch, D.F., Kunsman, K.P., Blum, K.A., Newland, G.A., Waga, J., Kurtz, L., Bronsgeest, M. & Cone, E.J. (2005). Passive cannabis smoke exposure and oral fluid testing. II. Two studies of extreme cannabis smoke exposure in a motor vehicle, *Journal of Analytical Toxicology* **29**, 607.

[59] Niedbala, S., Kardos, K., Salamone, S., Fritch, D., Bronsgeest, M. & Cone, E.J. (2004). Passive cannabis smoke exposure and oral fluid testing, *Journal of Analytical Toxicology* **28**, 546.

[60] Ramaekers, J.G., Moeller, M.R., van Ruitenbeek, P., Theunissen, E.L., Schneider, E. & Kauert, G. (2006). Cognition and motor control as a function of delta9-THC concentration in serum and oral fluid: limits of impairment, *Drug and Alcohol Dependence* **85**, 114.

[61] Day, D., Kuntz, D.J., Feldman, M. & Presley, L. (2006). Detection of THCA in oral fluid by GC-MS-MS, *Journal of Analytical Toxicology* **30**, 645.

[62] Moore, C., Coulter, C., Rana, S., Vincent, M. & Soares, J. (2006). Analytical procedure for the determination of the marijuana metabolite 11-nor-delta9-tetrahydrocannabinol-9-carboxylic acid in oral fluid specimens, *Journal of Analytical Toxicology* **30**, 409.

[63] Moore, C., Ross, W., Coulter, C., Adams, L., Rana, S., Vincent, M. & Soares, J. (2006). Detection of the marijuana metabolite 11-nor-delta9-tetrahydrocannabinol-9-carboxylic acid in oral fluid specimens and its contribution to positive results in screening assays, *Journal of Analytical Toxicology* **30**, 413.

[64] Banken, J.A. (2004). Drug abuse trends among youth in the United States, *Annals of the New York Academy of Sciences* **1025**, 465.

[65] Degenhardt, L., Copeland, J. & Dillon, P. (2005). Recent trends in the use of "club drugs": an Australian review, *Substance Use and Misuse* **40**, 1241.

[66] de la Torre, R., Farre, M., Navarro, M., Pacifici, R., Zuccaro, P. & Pichini, S. (2004). Clinical pharmacokinetics of amfetamine and related substances: monitoring in conventional and non-conventional matrices, *Clinical Pharmacokinetics* **43**, 157.

[67] Baselt, R.C. (ed) (2000). *Disposition of Toxic Drugs and Chemicals in Man*, Chemical Toxicology Institute, Foster City.

[68] Benowitz, N.L. (1993). Clinical pharmacology and toxicology of cocaine, *Pharmacology and Toxicology* **72**, 3.

[69] Isenschmid, D.S. (2002). Cocaine – effects on human behavior and performance, *Forensic Science Review* **14**, 133.

[70] Brzezinski, M.R., Abraham, T.L., Stone, C.L., Dean, R.A. & Bosron, W.F. (1994). Purification and characterization of a human liver cocaine carboxylesterase that catalyzes the production of benzoylecgonine and the formation of cocaethylene from alcohol and cocaine, *Biochemical Pharmacology* **48**, 1747.

[71] Inaba, T., Stewart, D.J. & Kalow, W. (1978). Metabolism of cocaine in man, *Clinical Pharmacology and Therapeutics* **23**, 547.

[72] Blaho, K., Logan, B., Winbery, S., Park, L. & Schwilke, E. (2000). Blood cocaine and metabolite concentrations, clinical findings, and outcome of patients presenting to an ED, *The American Journal of Emergency Medicine* **18**, 593.

[73] Cone, E.J., Oyler, J. & Darwin, W.D. (1997). Cocaine disposition in saliva following intravenous, intranasal, and smoked administration, *Journal of Analytical Toxicology* **21**, 465.

[74] Jufer, R.A., Wstadik, A., Walsh, S.L., Levine, B.S. & Cone, E.J. (2000). Elimination of cocaine and metabolites in plasma, saliva, and urine following repeated oral administration to human volunteers, *Journal of Analytical Toxicology* **24**, 467.

[75] Moolchan, E.T., Cone, E.J., Wstadik, A., Huestis, M.A. & Preston, K.L. (2000). Cocaine and metabolite elimination patterns in chronic cocaine users during cessation: plasma and saliva analysis, *Journal of Analytical Toxicology* **24**, 458.

[76] Jenkins, A.J., Oyler, J.M. & Cone, E.J. (1995). Comparison of heroin and cocaine concentrations in saliva with concentrations in blood and plasma, *Journal of Analytical Toxicology* **19**, 359.

[77] O'Neal, C.L., Crouch, D.J., Rollins, D.E., Fatah, A. & Cheever, M.L. (1999). Correlation of saliva codeine concentrations with plasma concentrations after oral codeine administration, *Journal of Analytical Toxicology* **23**, 452.

[78] Wang, W.L., Darwin, W.D. & Cone, E.J. (1994). Simultaneous assay of cocaine, heroin and metabolites in hair, plasma, saliva and urine by gas chromatography-mass spectrometry, *Journal of Chromatography. B, Biomedical Applications* **660**, 279.

[79] Cone, E.J. (1990). Testing human hair for drugs of abuse. I. Individual dose and time profiles of morphine and codeine in plasma, saliva, urine, and beard compared to drug-induced effects on pupils and behavior, *Journal of Analytical Toxicology* **14**, 1.

[80] Barnes, A.J., Kim, I., Schepers, R., Moolchan, E.T., Wilson, L., Cooper, G., Reid, C., Hand, C. & Huestis, M.A. (2003). Sensitivity, specificity, and efficiency in detecting opiates in oral fluid with the Cozart Opiate Microplate EIA and GC-MS following controlled codeine administration, *Journal of Analytical Toxicology* **27**, 402.

[81] Kim, I., Barnes, A.J., Oyler, J.M., Schepers, R., Joseph Jr, R.E., Cone, E.J., Lafko, D., Moolchan, E.T. & Huestis, M.A. (2002). Plasma and oral fluid pharmacokinetics and pharmacodynamics after oral codeine administration, *Clinical Chemistry* **48**, 1486.

[82] Skopp, G., Potsch, L., Klinder, K., Richter, B., Aderjan, R. & Mattern, R. (2001). Saliva testing after single and chronic administration of dihydrocodeine, *International Journal of Legal Medicine* **114**, 133.

[83] Rohrig, T.P. & Moore, C. (2003). The determination of morphine in urine and oral fluid following ingestion of poppy seeds, *Journal of Analytical Toxicology* **27**, 449.

[84] Drummer, O.H. (ed) (2001). Benzodiazepines, in *The Forensic Pharmacology of Drugs of Abuse*, Arnold, London, pp. 103–175.

[85] Greenblatt, D.J. (1992). Pharmacology of benzodiazepine hypnotics, *The Journal of Clinical Psychiatry* **53**, 7.

[86] Shader, R.I. & Greenblatt, D.J. (1993). Use of benzodiaepines in anxiety disorders, *The New England Journal of Medicine* **328**, 1398.

[87] Barnas, C., Rossmann, M., Roessler, H., Riemer, Y. & Fleischhacker, W.W. (1992). Benzodiazepines and other psychotropic drugs abused by patients in a methadone maintenance program: familiarity and preference, *Journal of Clinical Psychopharmacology* **12**, 397.

[88] San, L., Tato, J., Torrens, M., Castillo, C., Farre, M. & Cami, J. (1993). Flunitrazepam consumption among heroin addicts admitted for in-patient detoxification, *Drug and Alcohol Dependence* **32**, 281.

[89] Samyn, N., De Boeck, G., Cirimele, V., Verstraete, A. & Kintz, P. (2002). Detection of flunitrazepam and 7-aminoflunitrazepam in oral fluid after controlled administration of rohypnol, *Journal of Analytical Toxicology* **26**, 211.

[90] Kintz, P., Villain, M., Cirimele, V., Pepin, G. & Ludes, B. (2004). Windows of detection of lorazepam in urine, oral fluid and hair, with a special focus on drug-facilitated crimes, *Forensic Science International* **145**, 131.

[91] Concheiro, M., Villain, M., Bouchet, S., Ludes, B., Lopez-Rivadulla, M. & Kintz, P. (2005). Windows of detection of tetrazepam in urine, oral fluid, beard, and hair, with a special focus on drug-facilitated crimes, *Therapeutic Drug Monitoring* **27**, 565.

[92] Kintz, P., Villain, M. & Ludes, B. (2004). Testing for zolpidem in oral fluid by liquid chromatography-tandem mass spectrometry, *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences* **811**, 59.

[93] Kintz, P. & Samyn, N. (2002). Use of alternative specimens: drugs of abuse in saliva and doping agents in hair, *Therapeutic Drug Monitoring* **24**, 239.

[94] Caplan, Y.H. & Goldberger, B.A. (2001). Alternative specimens for workplace drug testing, *Journal of Analytical Toxicology* **25**, 396.

[95] Toennes, S.W., Kauert, G.F., Steinmeyer, S. & Moeller, M.R. (2005). Driving under the influence of drugs – evaluation of analytical data of drugs in oral fluid, serum and urine, and correlation with impairment symptoms, *Forensic Science International* **152**, 149.

[96] Verstraete, A.G. (2005). Oral fluid testing for driving under the influence of drugs: history, recent progress and remaining challenges, *Forensic Science International* **150**, 143.

[97] Wylie, F.M., Torrance, H., Seymour, A., Buttress, S. & Oliver, J.S. (2005). Drugs in oral fluid Part II. Investigation of drugs in drivers, *Forensic Science International* **150**, 199.

[98] Bennett, G.A., Davies, E. & Thomas, P. (2003). Is oral fluid analysis as accurate as urinalysis in detecting drug use in a treatment setting? *Drug and Alcohol Dependence* **72**, 265.

[99] Lintz, W., Beier, H. & Gerloff, J. (1999). Bioavailability of tramadol after i.m. injection in comparison to i.v. infusion, *International Journal of Clinical Pharmacology and Therapeutics* **37**, 175.

[100] Moolchan, E.T., Umbricht, A. & Epstein, D. (2001). Therapeutic drug monitoring in methadone maintenance: choosing a matrix, *Journal of Addictive Diseases: The Official Journal of the ASAM, American Society of Addiction Medicine* **20**, 55.

[101] Neale, J. & Robertson, M. (2003). Comparisons of self-report data and oral fluid testing in detecting drug use amongst new treatment clients, *Drug and Alcohol Dependence* **71**, 57.

[102] Kadehjian, L. (2005). Legal issues in oral fluid testing, *Forensic Science International* **150**, 151.

[103] Presley, L., Lehrer, M., Seiter, W., Hahn, D., Rowland, B., Smith, M., Kardos, K.W., Fritch, D., Salamone, S., Niedbala, R.S. & Cone, E.J. (2003). High prevalence of 6-acetylmorphine in morphine-positive oral fluid specimens, *Forensic Science International* **133**, 22.

[104] Cooper, G., Wilson, L., Reid, C., Main, L. & Hand, C. (2005). Evaluation of the Cozart RapiScan drug test system for opiates and cocaine in oral fluid, *Forensic Science International* **150**, 239.

[105] Speckl, I.M., Hallbach, J., Guder, W.G., Meyer, L.V. & Zilker, T. (1999). Opiate detection in saliva and urine–a prospective comparison by gas chromatography-mass spectrometry, *Journal of Toxicology. Clinical Toxicology* **37**, 441.

[106] Dams, R., Choo, R.E., Lambert, W.E., Jones, H. & Huestis, M.A. (2007). Oral fluid as an alternative matrix to monitor opiate and cocaine use in substance-abuse treatment patients, *Drug and Alcohol Dependence*, **87**, 258–267.

[107] Phillips, S.G. & Allen, K.R. (2006). Acetylcodeine as a marker of illicit heroin abuse in oral fluid samples, *Journal of Analytical Toxicology* **30**, 370.

[108] Ortelli, D., Rudaz, S., Chevalley, A.F., Mino, A., Deglon, J.J., Balant, L. & Veuthey, J.L. (2000). Enantioselective analysis of methadone in saliva by liquid chromatography-mass spectrometry, *Journal of Chromatography A* **871**, 163.

[109] Rosas, M.E., Preston, K.L., Epstein, D.H., Moolchan, E.T. & Wainer, I.W. (2003). Quantitative determination of the enantiomers of methadone and its metabolite (EDDP) in human saliva by enantioselective liquid chromatography with mass spectrometric detection, *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences* **796**, 355.

[110] Bermejo, A.M., Lucas, A.C. & Tabernero, M.J. (2000). Saliva/plasma ratio of methadone and EDDP, *Journal of Analytical Toxicology* **24**, 70.

[111] Verstraete, A.G. & Raes, E.E. (eds) (2006). *Rosita-2 Project. Final Report*, Academia Press, Gent.

[112] Verstraete, A.G. (ed) (2001). *Rosita: Roadside Testing Assessment*, Rosita Consortium, Gent.

[113] Walsh, J.M., de Gier, J.J., Christopherson, A.S. & Verstraete, A.G. (2004). Drugs and driving, *Traffic Injury Prevention* **5**, 241.

[114] Verstraete, A.G. (2004). Detection times of drugs of abuse in blood, urine, and oral fluid, *Therapeutic Drug Monitoring* **26**, 200.

[115] Laloup, M., Ramirez Fernandez, M.M., Wood, M., De Boeck, G., Maes, V. & Samyn, N. (2006). Correlation of delta9-tetrahydrocannabinol concentrations determined by LC-MS-MS in oral fluid and plasma from impaired drivers and evaluation of the on-site Dräger DrugTest®, *Forensic Science International* **161**, 175.

[116] Toennes, S.W., Steinmeyer, S., Maurer, H.J., Moeller, M.R. & Kauert, G.F. (2005). Screening for drugs of abuse in oral fluid–correlation of analysis results with serum in forensic cases, *Journal of Analytical Toxicology* **29**, 22.

[117] Verstraete, A.G. & Puddu, M. (2001). Evaluation of different roadside drug tests, in *Rosita: Roadside Testing Assessment*, A.G. Verstraete, ed, Rosita Consortium, Gent, pp. 167–232.

[118] Bernhoft, I.M., Steentoft, A., Johansen, S.S., Klitgaard, N.A., Larsen, L.B. & Hansen, L.B. (2005). Drugs in injured drivers in Denmark, *Forensic Science International* **150**, 181.

[119] Jenkins, A.J. & Goldberger, B.A. (eds) (2002). *On-site Drug Testing*, Humana Press, Totawa.

[120] Jehanli, A., Brannan, S., Moore, L. & Spiehler, V.R. (2001). Blind trials of an onsite saliva drug test for

marijuana and opiates, *Journal of Forensic Sciences* **46**, 1214.
[121] Moore, L., Wicks, J., Spiehler, V. & Holgate, R. (2001). Gas chromatography-mass spectrometry confirmation of Cozart RapiScan saliva methadone and opiates tests, *Journal of Analytical Toxicology* **25**, 520.
[122] De Giovanni, N., Fucci, N., Chiarotti, M. & Scarlata, S. (2002). Cozart Rapiscan system: our experience with saliva tests, *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences* **773**, 1.
[123] Kintz, P., Bernhard, W., Villain, M., Gasser, M., Aebi, B. & Cirimele, V. (2005). Detection of cannabis use in drivers with the drugwipe device and by GC-MS after Intercept device collection, *Journal of Analytical Toxicology* **29**, 724.
[124] Samyn, N. & van Haeren, C. (2000). On-site testing of saliva and sweat with Drugwipe and determination of concentrations of drugs of abuse in saliva, plasma and urine of suspected users, *International Journal of Legal Medicine* **113**, 150.
[125] Biermann, T., Schwarze, B., Zedler, B. & Betz, P. (2004). On-site testing of illicit drugs: the use of the drug-testing device "Toxiquick", *Forensic Science International* **143**, 21.
[126] Wong, R.C., Tran, M. & Tung, J.K. (2005). Oral fluid drug tests: effects of adulterants and foodstuffs, *Forensic Science International* **150**, 175.
[127] Cirimele, V., Villain, M., Mura, P., Bernard, M. & Kintz, P. (2006). Oral fluid testing for cannabis: on-site OraLine IV s.a.t. device versus GC/MS, *Forensic Science International* **161**, 180.
[128] Crouch, D.J., Walsh, J.M., Flegel, R., Cangianelli, L., Baudys, J. & Atkins, R. (2005). An evaluation of selected oral fluid point-of-collection drug-testing devices, *Journal of Analytical Toxicology* **29**, 244.
[129] Walsh, J.M., Flegel, R., Crouch, D.J., Cangianelli, L. & Baudys, J. (2003). An evaluation of rapid point-of-collection oral fluid drug-testing devices, *Journal of Analytical Toxicology* **27**, 429.
[130] Parliament of Victoria (2003). Road Safety (drug driving) act 2003, 111/2003.
[131] Drummer, O.H., Gerostamoulos, D. & Chu, M. (2006). Drugs in oral fluid in randomly selected drivers. *Presented at the 44th TIAFT meeting*, August 26th-September 1st 2006, Ljubljana.

Related Articles

Amphetamine
Benzodiazepines
Confirmation Testing: Toxicology
Cannabis
Cocaine
Drug-Impaired Driving
Opioids
Toxicology: Initial Testing

LALOUP MARLEEN, SAMYN NELE AND DE BOECK GERT

Outpatient Commitment *see* Civil Commitment

Outreach: Mental Health *see* Disaster Mental Health



Packaging and Transport

General Issues

Anticontamination and Continuity Measures

Where an investigation involves more than one scene (a location, vehicle, or subject), no person should recover or package items from more than one of them. All subjects, victims and suspects, should be kept apart and transported in different vehicles using different officers.

Whenever possible, packaging material should be taken to the location where evidence is found and packaged *in situ*. Vital evidence may be lost if evidence is transported to a different location for packaging. Every item of evidence must be packaged separately in a sealed container, bag or box. The package must be sealed with tape to prevent tampering with the evidence. While normal adhesive tape is permissible in some countries, some jurisdictions disapprove of the practice and require tamperproof evidence tape be used to seal evidence.

Staples, pins, and paperclips must *never* be used to seal forensic packaging or attach evidence vouchers to the packaging. They will puncture the packaging and it could be alleged that cross contamination of hair, fibers, microscopic pieces of glass and paint could have occurred through any holes caused.

While contamination and cross contamination have always been issues for any evidence types in forensic science, the increasing sensitivity of DNA profiling have made anticontamination measures even more important.

Storage and Transport

All storage of exhibits must be in secure facilities. The basic requirements for exhibit storage are a secure store or locker, depending on the size of agency. If secured by key, there should only be one available to personnel and agencies should have protocols governing entry and documentation of such entry. If entry is by keypad or swipe card, this must be linked to a computer that records entry and exit.

A suitable freezer and refrigerator are also required for biological and other exhibits and access should be *via* a secure exhibit storage facility.

Some larger crime labs and agencies have adopted the use of radiofrequency identification (RFID) tags to track the location of exhibits.

Items from one subject (suspect or victim) or scene must *never* be stored or transported in the same box or large paper sack containing items from any other subject or the scene they are suspected of involvement in. This will prevent any accidental cross contamination.

Packaging of Exhibits

Clothing, Footwear, Bedding, and Similar Items

Dry items, such as clothing, bedding and footwear, should be placed in separate paper bags. These bags are also referred to as *Kraft* bags in some

areas. The top of the bag should be folded over about 1 in. (2.5 cm) and then folded over the same distance again. The bag should then be completely sealed, using 2-in. (5 cm) wide adhesive tape or, if not acceptable in the jurisdiction, an approved tape. Where protocols dictate, it may be necessary to sign across the seal, half on the tape and half on the bag. Where bags have no chain of custody details printed on them, the relevant evidence voucher (*see* **Crime Scene Documentation**) should then be attached using an approved tape. Where the item is too large to fit in standard paper bags, it should be carefully wrapped in one or more sheets of brown paper, fully sealed to prevent contamination of the item. As an extra preventative measure, a second layer of wrapping may be employed.

In the United Kingdom, advice from forensic laboratories is that dry items must *never* be placed in polyethylene bags as it is possible that a buildup of moisture could degrade any biological or trace evidence. However, advice in other countries may vary and it may be permissible to ensure items are completely dry before placing them in polyethylene bags.

Wet or Damp Items

With the exception of footwear, damp or wet items recovered from scenes or subjects should be placed in individual polyethylene bags. Items of footwear must *never* be placed in polyethylene bags as they will become moldy.

Wherever possible, polyethylene bags with self tamperproof seals should be used as a normal adhesive tape will lose adhesion and compromise the integrity of the exhibit.

Never seal or store damp or wet items in polyethylene bags without freezing at the earliest opportunity. Frozen items must not be allowed to thaw during transportation to the laboratory. When folding, personnel should avoid folding the item in such a manner that could affect bloodstain patterns.

Where facilities, such as forensic drying rooms and cabinets are available, the items should be air dried prior to being repackaged as per the guidelines for dry items. The air-drying process should not include heating to accelerate the drying time. The items should be hung separately and a large sheet of paper placed below to collect any debris that may fall off during the drying process.

The polyethylene bag should then be packaged and labeled as an exhibit.

Where no forensic drying cabinets are available, extreme care should be taken when choosing a drying location. Careful consideration should be given to the possibility of infection from the person shedding the blood, not only to prevent any accidental contamination or interference with the evidence but also where the item is bloodstained.

When dry, the items should be carefully moved using appropriate anticontamination and protective measures. The item should be carefully folded as necessary and placed in a paper bag and sealed and labeled accordingly. If the item is bloodstained, in addition to protective clothing and gloves, personnel dealing with the clothing should wear a protective mask and goggles to prevent any possible infection from airborne dried blood particles.

The sheet of paper placed below the item should be packaged separately and treated as a new exhibit. The clothing, polyethylene bag, and paper sheet should all be sent to the laboratory for examination. In this way, all debris that may have detached from the exhibit can be examined if necessary.

Weapons, Tools, and Miscellaneous Items

There are many flat-pack forensic boxes of many different sizes available. They vary from small boxes through some similar to pizza boxes to others capable of containing a rifle or possibly a baseball bat.

Weapons and tools should be placed a suitable flat-pack cardboard box. Ensure that adhering debris is protected by paper folded around the tip and shaft. The implement should then be tied securely in place with string to prevent movement. This is achieved by piercing holes in the box close to the item and threading string through them. The knots should be on the inside of the box, to prevent accidental or deliberate interference and the holes created to thread the string through must be sealed with an approved tape. The box should be fully sealed around the edges, with an approved tape. Where protocols dictate, each piece of tape should be signed, half on the tape and half on the box to prevent any allegations of tampering. Some evidence boxes come with a card insert with a grid of prepunched holes, to allow items to be secured without the need to pierce the box.

The relevant evidence voucher (*see* **Crime Scene Documentation**) should then be attached using an approved tape.

The box should always be stored and transported flat, with the securing string to the bottom. The box should be labeled to indicate which way up the box should remain and also if it contains a sharp object and any health hazards such as bodily fluids.

Wherever possible, rigid knife tubes should be used to package knives and other sharp instruments. These are two plastic tubes, open at one end, which fit inside each other by means of a spiral groove and hold the item in position at either end. The tube is then sealed in the same way as a box and an evidence voucher attached.

Firearms should be made safe by an authorized officer, under the supervision of crime scene investigators (CSIs). Where protocols dictate, an appropriate certificate should be attached to the outside of the exhibit. A suitable box should be used, sealed, and labeled as above. Ammunition and magazines should be packaged separately.

A magazine should be packaged in a box as above. Dependent on local protocols, ammunition may need to be removed from the magazine or chamber of a revolver. It is good practice to mark a revolver to show the relative position of the chambers at the time of recovery.

Ammunition, either projectiles or shell casings, spent or otherwise, should be packaged in small plastic pots padded with tissue paper to prevent damage. The pots should be individually sealed and the appropriate evidence voucher attached.

A recent suggestion for the packaging of spent shell casings is to insert a dry swab into the open end and place the casing in a plastic urine pot. Those pots with a conical base mean that only the smallest part of the rim is in contact with anything. The swab is then broken at the lip of the pot allowing the casing to be held in position. This method prevents possible damage to the casing and any fingerprints on it.

Footwear Impressions

Where a footwear or tire impression has been cast, they too are placed in a cardboard box. The cast is placed soil side up in the box. The soil should not be removed as this provides a sample from exactly where the offender stood or the vehicle was driven. The cast is tied into the box, sealed, and labeled as above.

If a footwear impression in dust has been recovered using an electrostatic lifting apparatus (ESLA), it is also packaged in a box. The foil is placed black side up in the box and taped into position in each corner. If the impression was recovered from a movable object, such as paper strewn across the floor, this too is taped into a box. The boxes should be sealed and labeled as above.

Glass Exemplars

Whenever glass has been broken at a crime scene, *exemplars*, also known as *control samples*, should be taken and packaged in a cardboard box. The box should be sealed and labeled as above.

Storage. Store all the above "right way up" in a dry exhibit store.

Transport. No special requirements other than keep "right way up".

Dry Miscellaneous Traces

Paint flakes, exemplars or unknown source, powders or other debris, should be placed in folded paper, druggists' wraps. The wrap should then be sealed in a polyethylene bag, and the appropriate evidence voucher completed and attached.

Damp or Wet Miscellaneous Traces

Items such as oils and greases should be submitted in the original container if available. If not, they should be recovered into a suitable container, sealed in a polyethylene bag and the appropriate evidence voucher completed and attached.

Storage. Store all the above "right way up" in a dry exhibit store.

Transport. No special requirements other than keep "right way up".

Fiber Tapings

Each recovered taping should be sealed in a polyethylene bag and the appropriate evidence voucher completed and attached.

Storage. They should be stored in a dry environment.

Transport. No special requirements.

Accelerants or Volatile Substances

Liquid Samples – Packaging and Storage. These must be packaged in a nylon bag, tied, and sealed at the top with a swan neck. This bag should then be placed in a polyethylene bag, again tied and sealed with a swan neck. An evidence voucher must be attached using string (*see* **Crime Scene Documentation**). Do *not* use an approved tape to attach an evidence voucher to the packaging. Containers with rubber or plastic seals and plastic vials should *not* be used. They should be stored in a dry environment, isolated from other items.

Transport. When transporting, the sample bags should be placed in a rigid outer container, and suitably padded to prevent impact damage.

Clothing. Clothing should be searched prior to submission. It should be loosely packaged in a nylon bag, tied, and sealed at the top with a swan neck. This is achieved by twisting the top of the bag tightly until a long "spout" is made. This is then tied around itself and then the knot is tied with string to secure it. No form of adhesive tape should be used to secure the knot. This bag should then be placed in a polyethylene bag, again tied, and sealed with a swan neck. An evidence voucher must be attached using string (*see* **Crime Scene Documentation**). Do *not* use any type of adhesive tape to attach an evidence voucher to the packaging. They should be stored in a dry environment, isolated from other items.

Transport. When transporting, the sample bags should be placed in a rigid outer container, and suitably padded to prevent impact damage.

Fragile Items from Fire Scenes. Fragile items should be packed in a sealed sturdy box or tin.

Transport. The items should be carefully transported. They should be transported personally to the laboratory.

Accelerant Samples from Fire Scenes. These must be packaged in a nylon bag, tied, and sealed at the top with a swan neck. This bag should then be placed in a polyethylene bag, again tied and sealed with a swan neck. An evidence voucher must be attached using string (*see* **Crime Scene Documentation**). Do *not* use an adhesive tape to attach an evidence voucher to the packaging. Leave a quantity of air above the sample, often known as a *headspace*. Unused glass jars with well-fitting caps can be substituted for nylon bags.

Where methylated spirits are suspected, a glass bottle with a well-fitting cap should be used. Double wrapping in nylon bags may be used as an alternative.

Samples should be submitted as soon as possible.

In some cases, items suspected of containing accelerants will have sharp or broken edges or protrusions. Any sharp items should be placed in a rigid box, sealed in a nylon bag. The packaging should be marked to clearly indicate that a sharp item is contained within.

Biological Samples

Swabs. Swabs as an evidence type are the most complex as regards packaging. What is accepted practice in some jurisdictions is frowned upon in others. Reference must be made to local protocols.

In many US states, swabs are air dried. They may be placed into sample envelopes to dry prior to submission. Some jurisdictions use a small box with an insert designed to allow the swab to be held in position away from the box and air dried in that. The box is sealed and the details documented on an evidence voucher.

In some jurisdictions worldwide, swabs from subjects and also crime scenes are frozen as soon as possible and submitted frozen. In others, swabs from subjects are frozen but scene stains are air dried.

In England and Wales, where the majority of swabs are frozen, standard medical swabs are used for crime scenes, including victims and suspects. They are sealed in the tube provided, placed in a polyethylene bag and the appropriate evidence voucher attached.

Cigarette Ends. They can be placed in individual polyethylene bags, envelopes or paper bags, sealed, and the evidence voucher attached. If dry they are stored in a dry environment. If wet they are frozen.

Storage. Frozen samples should remain frozen. Air-dried samples should be stored in a dry environment.

Transport. Frozen samples should remain frozen in transit. Air-dried samples should remain in a dry environment.

Condoms. Condoms should be knotted to prevent the loss of any liquid inside. They should be placed in a rigid plastic pot. The pot should be placed in a polyethylene bag, sealed, and the appropriate evidence voucher attached. The exhibit should be frozen as soon as possible and remain frozen in storage and transit.

ALLAN MATHIESON SCOTT

Paint

Introduction

Paint can be defined as a coating applied to a surface to provide decorative, protective, or other properties. In the context of forensic trace evidence, paint can be encountered in a range of situations, from traffic accidents and criminal damage, through burglaries and art fraud, to murder and terrorist activity.

"Essentially, a paint consists of a pigment dispersed in a resinous binder, reduced to an acceptable application viscosity with solvent. One or more additives may be incorporated to modify one or more of the paint's film properties, application, or storage characteristics [1]."

Paint is a common material in everyday life: buildings, vehicles, furniture, tools, and even fashion accessories can be painted. However, each of these uses is different; there are very many colors and shades of paint available; the composition of paints can vary widely; each additional layer of paint adds to a paint flake's individuality. Traces of paint can be transferred from one item to another, or to a person in contact with a painted item, and this can have major forensic significance.

Forensic Aspects

The forensic characterization of a paint usually involves microscopic examination allied to some chemical analysis of the organic and/or inorganic components.

Forensic paint analyses and comparisons are typically distinguished by a small sample size that precludes the application of many standard industrial paint analysis procedures or protocols. The forensic paint examiner must address concerns, such as the issues of a case or investigation, sample size, complexity and condition, environmental effects, and collection methods. These factors require that the forensic paint examiner choose test methods, sample preparation schemes, test sequence, and degree of sample alteration and consumption that are appropriate for each specific case.

Forensic science is concerned with generating information, which can be used to support or refute propositions made within the legal context, principally in courts of law. It is important at the outset, therefore, to identify the nature of the customer's proposition and any reasonable alternatives in some detail, to enable the correct approach to be followed to obtain all the relevant data.

Paint investigations are mostly requested in relation to incidents where cars have been involved (e.g., hit and run accidents, a raid on a bank where a car has been used to crash through a window, or a murder where a body has been carried in a car) or to incidents where paint has been transferred to or from a tool or another painted object (e.g., burglaries or assaults).

Normally, the request will be to determine if there is a link between a microtrace of paint found on one object/person and another painted object. In hit and run cases, the request may be to determine from paint microtraces which model of car, and from what year, has been involved in the incident.

Two stages can be distinguished in many investigations:

- In the *investigative stage,* the customer is interested in clues which help him to determine which avenues of investigation appear to be most fruitful. Speed of the response is normally more important in this phase than the strength of evidence.
- In the *evidential stage,* a suspect has been found and procedures have been started to determine if the suspect should be prosecuted in court. In this phase, strength of the evidence is normally more important than the speed of response.

During the paint investigation, the emphasis of the request may shift from speed of response to strength of evidence. Normally, from the request it can be deduced what the current phase of the investigation is and what the priorities of the customer are [2].

In the majority of cases involving paint, the requirement is for the laboratory to compare samples of paint to see if they could have a common origin. This will mean the application of a range of techniques from low-power microscopy to more specialized analytical methods.

Other cases require that the paint samples submitted to the laboratory are examined to assist an investigation where there is as yet no comparison sample. Such cases include traffic accidents where automotive paint databases are searched for paint matching traces recovered from an injured person's clothing to help identify the vehicle involved.

Paint Composition

In the past, it was easy to define paint. Paint in its simplest form was composed of four main types of ingredient described as follows:

1. Resin (also referred to as *polymer* or *binder*)

This is essentially the "glue" that binds the pigments and additives in place and dries to a solid "film". The polymeric binder (or resin) provides the continuous matrix for the paint film and gives the paint its adhesive quality, durability, flexibility, and chemical resistance. Sometimes a modifying resin is also used with the primary resin.

2. Pigment

Finely divided particles, which are dispersed throughout the liquid paint and give the dried paint film: color, opacity (hiding power), film reinforcement and functionality, gloss, and permeability.

3. Solvent

The solvent is the volatile portion, which is used to provide application-related properties including flow and dry time. Often this "solvent" is primarily water. Several "solvents" may be used in a coating. Upon drying of the paint film, the solvent is evaporated and eventually lost. Proper drying of the paint film is essential to achieve desired film properties.

4. Additives

Additional components, which lend added properties to the finished film and/or make application and manufacturing easier. They also contribute to gloss, viscosity, and other properties.

Today, with the advent of high technology, paint is still essentially as described above; however, certain types of coatings – such as most powder coatings – are 100% solid and do not contain solvent [3].

Paint Application

Different paints can be applied to surfaces by several means: for example, by brush and roller, by manual or robotic spraying, by dipping or as a powder coat [4].

Architectural or domestic paints are most often applied by *brush,* at least in the case of the paint types submitted for forensic examination. Manual application will typically lead to more variation in thickness and in homogeneity than more controlled industrial application procedures. This may mean that a greater number of control or known samples will need to be taken from the surface to ensure that the examiner has a representative sample for comparison.

Other decorative finishes featuring paint can also be found on items like furniture, or on products like fire extinguishers and domestic appliances, but these are encountered less frequently in case work.

Where paint is *sprayed*, either from an aerosol can or using a compressed air or pump paint sprayer, layer thickness and homogeneity may also vary. There may also be a contribution from paint that was previously sprayed using the same equipment (for example, traces of mica or aluminum particles from an "effect" paint carried into a solid paint). Stray paint droplets ("overspray") can drift during spraying onto nearby surfaces, which are not being painted, resulting in microscopic droplets of a different color on the surface, which individualize the paint further.

Robotic application, for example, in the motor industry, gives very reproducible results, with even layers of a consistent thickness. On a continuous production line, the color of the applied paint can change to meet the requirements of the product: the undercoat color of cars can change to suit the topcoat color, and during the change there may be a variation from the front to the back of a single vehicle.

Dipping a vehicle body into an electrostatic paint bath is a technique, which allows paint to penetrate into cavities that would otherwise be inaccessible, and is used for applying anticorrosion paint. Again, the results are consistent in terms of layer thickness and anticorrosion coatings may be very thin.

New technologies and environmental, health, and economic considerations have meant that water-borne paints, paint with less solvent, and powder coating (where no solvent is present) have become more common. The phasing-out of pigments like lead chromate was also a feature of paint development in the late twentieth century.

Transfer Mechanisms

Dry paint is normally transferred by direct contact. As paint is designed to bind strongly to a substrate, unless the paint is old and flaking, it will require some force to remove it. Scene of crime examiners will be familiar with the difficulty of removing samples from surfaces like automobiles. For this reason, paint may not transfer unless sufficient force is involved. In a minor traffic accident where there is a glancing blow at a shallow angle, only a smear of paint may be transferred, or none may be transferred at all. If a more severe impact occurs, one or more layers may be transferred, in one or both directions. There may also be smearing and mixing of the layers, indicating impact. The more layers transferred, the stronger the evidential value.

In a burglary, where a window is forced open with a tool such as a jemmy or pry bar, paint may be transferred from the window to the tool and from the tool to the window, while there may also be other evidence types like toolmark impressions, fingerprints, fibers, or DNA involved [5, 6].

Transferred smears of paint can indicate forceful contact, as in a collision between two cars, but the examiner should be cautious in arriving at a conclusion as to the color of the offending vehicle based only on smears of paint. Paint smeared thinly onto a light colored surface can appear a lighter color than the paint from which it came and the reverse may be true with paint smeared onto a dark surface. With effect paints or paints with a clear topcoat, if only the top layer (the *clearcoat*) is smeared, it may appear white, though it is usually colorless.

Wet paint can be transferred by direct contact with a painted surface, by splashing or spraying. A vandal painting or spraying graffiti could get paint splashes or droplets on their shoes, skin, or clothing; an intruder's clothing in contact with a newly painted surface could transfer paint to the clothing.

Loose paint particles can flake from a surface and be recovered at a scene. If this is a possibility, the examiner should consider whether there might be physical matches (also called *physical fits* or *fracture matches*) between pieces of recovered paint and a painted item. The author has encountered a case where a white-painted metal bar was used to break into motor cars. Flakes of paint were left inside one of the cars. Some of this recovered paint was found to fit precisely back onto the metal bar so as to show that it had come from this item.

On a larger scale, pieces of automobile body filler left at accident scenes can also provide scope for physical matches as well as for paint layer examination.

Persistence

In contrast with glass, there has been little published on the persistence of transferred paint. Published clothing surveys for paint describe work by Lau *et al.* [7] and Pearson *et al.* [8].

Other Paints

Other types of paint may be encountered from time to time. One such paint is anticlimb paint, which is designed to remain slippery and not to dry, to deter burglars or intruders from climbing on buildings. These paints may have additional components added as markers in the paint to help identify offenders when caught.

French scientists have studied the paint from weapons rockets to try to identify the types of rockets used in terrorist attacks [9].

Layer Structure

It is a combination of the layers and colors, which give paint flakes their individuality. It is easier to discriminate colored paints microscopically than black or white paints because they have a range of shades of color. The greater the number of layers, the more distinctive the paint will be. If a paint has many layers, the examiner may decide that no further examination is needed, other than microscopy, as the paint is already sufficiently characterized. This is more likely to be the case with an architectural paint, particularly from an old building, but it could also apply to a vehicle that has been repaired and repainted. Laboratory policy and an examiner's own judgment will determine whether chemical analysis, in addition to microscopic comparison, is required.

Where fewer layers are present, the examination should include some analysis of the paint. In most forensic laboratories, this means that Fourier transform infrared (FTIR) spectroscopy follows comparison microscopy. The availability of other instrumentation, the judgment of the examiner and laboratory policy will again provide guidance as to the tests conducted on a sample.

Paint Recovery

Paint flakes may have been recovered by a scene of crime examiner and submitted to the laboratory, or an item may have to be examined in the laboratory to recover paint from it. Paint traces can be very small ($<1\,mm^2$) so that appropriate anticontamination measures must be adopted throughout the examination.

The combination of a searching microscope and a fine forceps will allow even very small paint flakes to be recovered. This method will also allow the precise location of recovered flakes to be recorded, unlike brushing or shaking where the paint could have come from any part of an item or garment.

Paint may be recovered from tools or clothing under a searching microscope (magnification of 10× to 100× or so) or loose flakes can be recovered from clothing by brushing the surface or shaking the garment over a large sheet of paper or a special collection receptacle (e.g., a stainless steel funnel. See Figure 1).

Although the collection of paint at scenes using adhesive tape can be convenient for scene of crime examiners, the submission of paint flakes on tape lifts is not recommended as it can be difficult to remove the paint from the adhesive and the presence of adhesive traces can cause difficulties in interpreting infrared spectra. The best method is to remove the paint down to the underlying surface with a clean scalpel blade and place the paint flakes in a suitable



Figure 1 Brushing paint from clothing into a stainless steel funnel

container like a self-seal (e.g., Ziploc) plastic bag, which can be labeled.

Laboratory Examination. A possible stage-wise scheme for comparison of paint samples is shown below. The aim is to discriminate the paints using the techniques available. If at any stage significant differences are noted the examination may be stopped and the paints reported as different.

Sample description
Examine physical characteristics: potential physical match?
Evaluate sample for analytical approach: sufficient quantity? sequence of tests?
Comparison microscopy
FTIR analysis of selected layer(s)
Additional analysis: e.g., elemental analysis (scanning electron microscopy with energy-dispersive X-ray spectrometry (SEM–EDX); X-ray fluorescence spectroscopy (XRF)), fluorescence microscopy, pyrolysis gas chromatography–mass spectrometry (PGC–MS), and microspectrophotometry (MSP).

Microscopic Examination

Following the recovery of any paint, the first examination step is normally comparison microscopy (See Figure 2).

The human eye is capable of distinguishing over 100 000 different colors [10] and microscopy is an extremely powerful tool for the paint examiner. The subsequent laboratory analysis will depend on the instruments available.

Samples of control (*known or "K"*) and transferred (*questioned or "Q"*) paint are mounted on microscope slides and placed on the stages of a comparison microscope. The top and bottom surfaces of a paint flake can be viewed at a magnification of 250× using incident light. Cross sections of the paints can be prepared and mounted on the same slides as the surface samples.

If multiple layers are present they can be recorded in notes, diagrammatically or photographically. Layers of similar colors (e.g., shades of white) may be examined using UV light: any fluorescence may help discriminate the layers.

Cross sections of multilayer paint flakes can be prepared using a microtome, with or without embedding, or by the technique of "thin peels" where the edge of the flake is shaved with a scalpel blade to remove a sample containing all the layers.

"Effect" paints can contain aluminum flakes, mica (for a pearlescent appearance) or other particulate components. A feature of metallic paints is the form of the aluminum flakes, which should be noted; these can be rounded ("dollar") or with indented edges ("cornflake"). Effect paints may have one or both types of aluminum, with or without mica, or interference pigments (for example, mica coated with metallic oxides) to give color, which varies with coating thickness, illumination angle, or viewing angle.



Figure 2 FT-IR microscope and bench

Other microscopic examinations (involving polarized light microscopy, cathodoluminescence, or high-power microscopy) can also be used to characterize paint components.

Analysis

Following microscopy, the normal procedure is for an analytical scheme, which allows both the organic and inorganic components of a paint to be examined,which can cope with small samples, and which, as far as possible, uses nondestructive techniques [11].

Fourier Transform Infrared (FTIR) Spectroscopy. Most modern forensic laboratories use FTIR spectroscopy to analyze paint flakes, which are indistinguishable by comparison microscopy. This technique has the advantage of being fast, nondestructive, requiring little sample preparation and, when the microscope attachment is used, being able to cope with small sample sizes (See Figures 3 and 4).



Figure 3 Comparison microscope

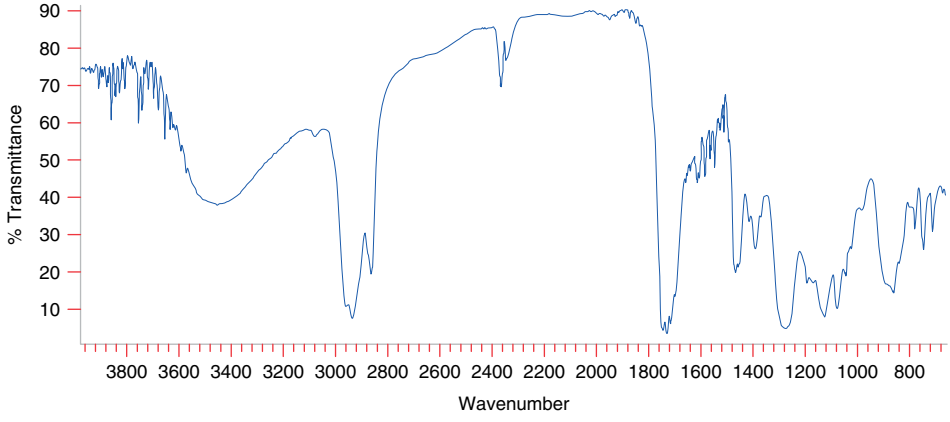


Figure 4 Infrared spectrum of paint

The mercury cadmium telluride (MCT) detector currently fitted to many FTIR microscopes has a wave number range from 4000 to 650 cm^{-1}, allowing the organic and some inorganic components to be detected. Wide-band MCT detectors are also available, providing a range of 4000–450 cm^{-1} but with lower sensitivity.

FTIR microscopes generally have an FTIR bench associated with the microscope, with a deuterated triglycine sulfate (DTGS) detector. If sufficient sample is available it may be possible to use this detector, giving a range of 4000–200 cm^{-1}, allowing the examiner to obtain more information about the inorganic components of the paint. This may be particularly useful when looking at vehicle undercoats.

Infrared (IR) spectroscopy deals with the infrared region of the electromagnetic spectrum. It covers a range of techniques, the most common being a form of absorption spectroscopy. It can be used to identify compounds or investigate sample composition. Infrared spectroscopy exploits the fact that molecules have specific frequencies at which they rotate or vibrate corresponding to discrete energy levels (vibrational modes).

The infrared spectrum of a sample is collected by passing a beam of infrared light through the sample. Examination of the transmitted light reveals how much energy was absorbed at each wavelength. This can be done with a monochromatic beam, which changes in wavelength over time, or by using a Fourier transform instrument to measure all wavelengths at once (FTIR). From this, a transmittance or absorbance spectrum can be produced, showing at which IR wavelengths the sample absorbs. Analysis of these absorption characteristics reveals details about the molecular structure of the sample. Current models of instrument are normally Fourier-transform infra red spectrometers which can collect a spectrum in seconds rather than minutes required for an older grating instrument. A further refinement, the addition of a microscope, means that very small samples can be analyzed with little sample preparation. Spectra can be presented in either transmittance or absorbance mode.

A detailed account of this technique is given by Beveridge *et al.* [12].

Elemental Analysis

Further information on a paint can be obtained by elemental analysis. Because of the small size and the condition of the samples typically encountered in forensic science laboratories, the examiner may be limited in what tests can be performed.

SEM–EDX is a technique used quite widely in forensic laboratories for elemental analysis. The technique is nondestructive and can cope with small samples. Some sample preparation is normally required (usually mounting and coating with carbon to make the surface conductive), but if a paint flake has been embedded and cross sectioned, only a very small additional amount is required for SEM.

If a low-vacuum or environmental SEM is available, no sample preparation may be needed [13, 14].

XRF and micro-XRF are techniques, which may also be applied to obtain elemental information. These require no sample coating but require larger samples than SEM.

SEM and XRF analysis can be done as a comparative test (semiquantitatively) or quantitatively by using either standards of known elemental composition or using the instrument's software to calculate percentage composition.

The scanning electron microscope is widely used as an analytical tool in forensic laboratories. The main advantage of this instrument is that multidimensional information may be gathered: chemical visualization by means of back-scattered electrons (BSE) and X-ray mapping, topographical information, detection of secondary electrons (SE), and qualitative and quantitative elemental composition collected by the energy-dispersive spectrometer (EDS). It is a nondestructive technique requiring some sample preparation but can be applied to samples of $<1\ mm^2$.

XRF is the emission of characteristic "secondary" (or fluorescent) X-rays from a material that has been excited by bombarding with high-energy X-rays or γ rays. The phenomenon is widely used for elemental analysis. Instruments can either use energy-dispersive X-ray (EDS or EDX) or wavelength dispersive X-ray (WDX) detectors. Little sample preparation is required and both qualitative and quantitative analysis can be performed.

Pyrolysis Gas Chromatography. PGC, usually coupled with PGC–MS, can be used to analyze organic material such as paint, plastic, or rubber. Although it is a destructive technique, it can be used on quite small samples. If suitable databases are available, it can also be used as an investigative tool to identify unknown samples such as flakes of car paint or plastic at an accident scene.

Some work has also been done on characterizing automotive clearcoats and coatings on plastic headlight lenses using PGC–MS.

Without the mass spectrometer, however, PGC alone may add little additional information because of the difficulty in identifying peaks in a chromatogram with certainty [15].

PGC–MS or PyGC–MS is a method of chemical analysis in which the sample is heated to decomposition (pyrolysis) to produce smaller molecules that are separated by GC and detected using MS. Although it is a destructive technique, it is useful in identifying and comparing polymeric materials like plastics, rubber, or individual layers of paint.

Microspectrophotometry. Microspectrophotometry using visible light and ultraviolet light (UV-MSP) can also be used as a comparative technique to provide objective information about whether two paints can be distinguished on the basis of their color. The MSP can produce a graphical printout of the color spectrum or can process the results into tristimulus values or color coordinates.

In certain situations, two paints may be visually indistinguishable even though the color has been achieved in different ways (e.g., by using different pigments). This phenomenon (metamerism) may be identified by MSP.

Though MSP is often used for fibers, it can also be used with paint using transmitted or incident light. Transmitted light is preferred as it eliminates any problems associated with sample preparation, surface features and illumination angle, although the thickness of the cross sections from paints being compared must be the same.

For color measurement and comparison, paint sections of approximately 3-μm thickness are prepared, while for the examination of UV-absorbers in clearcoats, thicker sections (20 μm) are required [16–18].

Hand-held spectrophotometers are available for objective color measurement. While they may have applications in collecting data for inclusion in databases, their use in the laboratory is limited because of their inability to measure the small samples (<3-mm diameter) usually seen in casework.

MSP is the technique used to measure the color spectrum of a microscopic sample. Depending on the type of sample it can be operated in the transmission mode (e.g., for fibers) or the reflectance mode (for opaque samples). Some instruments can also measure in the ultraviolet range.

Raman spectroscopy

Raman spectroscopy is a vibrational spectroscopy technique that is becoming more common in forensic laboratories. Raman spectrometers are becoming smaller, less costly, and more user-friendly. Raman spectroscopy can provide complementary information to that provided by infrared spectroscopy and can be useful in comparing and identifying pigments, extenders, and other components [19].

Raman spectroscopy is used to study vibrational, rotational, and other low-frequency modes in a system. It relies on inelastic scattering (Raman scattering) of monochromatic light, usually from a laser in the visible, near infrared, or near ultraviolet range. Infrared spectroscopy yields similar, but complementary information. In the past, photomultiplier tubes were the detectors of choice for dispersive Raman setups, which resulted in long acquisition times. Modern Fourier Transform Raman instruments have shorter acquisition times. In paint examination, Raman spectroscopy is a useful technique for analyzing pigments.

Other Analytical Techniques

Some laboratories may have other instruments available to them, which can be used for paint analysis. One example is *inductively coupled plasma mass spectrometry* (ICP-MS), possibly with laser ablation (LA-ICP-MS). Although it is a destructive technique, the amount of sample consumed in analysis is very small, and it can provide information on the elemental composition of paints, though currently it is more usually associated with glass analysis. These instruments are becoming smaller and more affordable and should become more common in forensic laboratories in the future [20].

Another analytical method being developed is *laser-induced breakdown spectroscopy* (LIBS), which may, in the future, provide a fast and relatively inexpensive alternative to LA-ICP-MS [21].

Investigative Examination

Most of the techniques and procedures referred to above relate principally to a comparison of a recovered sample of paint with a control or reference sample.

Cases also arise where only a recovered paint sample is submitted and the examiner is requested to establish its origin or provide the investigator with some other information to assist the enquiry.

Cases of this type include the identification of automotive paint to establish what type of vehicle it came from. Much work has been done in this area by expert paint working groups and databases of automotive paints from Europe, North America, Japan, and other countries have been built up over the years.

It may be possible to identify the source of a flake of automotive paint recovered from the clothing of a traffic accident victim by examining the color, layer structure, and chemical composition of the recovered paint layers. At best, the make, model, year, and manufacturing plant could be identified. In other cases, useful information on the source can still be provided to the investigator; for example, the likely manufacturer of the vehicle, the approximate age and the range of models to which this particular paint was applied.

These results depend on having the full original layer structure present. If the vehicle has been repainted or if only certain layers are present, the information available will be reduced, or it may not be possible to provide any useful information at all.

Both the European Network of Forensic Science Institutes (ENFSI) European Collection of Automobile Paints (EUCAP) [22] and Scientific Working Group on Materials Analysis (SWGMAT) Paint Data Query (PDQ) [12] search procedures can provide information on automotive paint. It should be remembered that not all car paints are in the databases and certain categories of paint (e.g., commercial vehicles and motorcycles) are not currently included.

Significance and Reporting

Following laboratory examination, the examiner's next step is to report the results. One approach to assessing the significance of results is to review the information available before starting any laboratory work. The examiner may need to make some enquiries as to what happened, what the prosecution and defense positions are, whether all relevant material has been submitted, etc. The examiner may then develop an examination plan and also formulate a set of expectations and alternative propositions. For example, the examiner may decide that, in a certain set of circumstances, there should be transfer of a large amount of multilayer paint between two items. If the result of the examination shows this to be the case, then the evidence supports this particular version of events; if the results do not match the expectations, then the results may favor

an alternative version of events, perhaps favoring the defense hypothesis; in some situations, the results may not be clear-cut and may add nothing to the existing knowledge about what really happened. This approach will depend on whether the legal system in the particular jurisdiction will allow it.

Reporting results should be done in a descriptive, factual way, bearing in mind that the reader is unlikely to have scientific training and – if the evidence is presented in court – those hearing the evidence presented may not have the written version in front of them but may have to rely on hearing the oral evidence only.

There have been publications on assessing the significance of paint evidence using Bayes' Theorem and work on this aspect of reporting physical evidence is continuing [23–25].

In order to assess the significance of a result, it may be necessary to refer to collections of paint color frequency data or national vehicle populations, or information may have to be collected to address specific issues in particular cases.

In general, the higher the number of layers present in a paint flake, the more individual a paint will be, and the more significance will be attached to a positive result. Two-way transfer of paint flakes (e.g., between two vehicles or between a tool and a point of entry) will be more significant than one-way transfer, but the assessment of the significance becomes more complex because the two-way transfers are not independent events. Further discussion on this area is beyond the scope of this article.

Reporting paint results from clothing will require knowledge of transfer and persistence studies, to reflect the prevalence of paint flakes on the clothing of relevant populations. Studies by Lau *et al.* [7] and Pearson *et al.* [8], for example, have documented the frequency of finding paint flakes and glass on clothing from dry cleaners and on students' clothing, and are valuable starting points.

Further Information

An excellent source of up-to-date information on forensic examination of paint is the Best Practice Guidelines produced by forensic expert groups such as the ENFSI Paint and Glass Working Group or its North American equivalent, the SWGMAT Paint Group. Details of these can be found on the ENFSI and Federal Bureau of Investigation (FBI) web sites www.enfsi.eu and www.fbi.gov.

In addition to these expert working groups, Interpol organize a triennial International Forensic Science Symposium. Part of this meeting is a review of all relevant papers published in the preceding three years, which is a very valuable resource on current work. Further information can be found on the Interpol web site www.interpol.int.

Periodic reviews of the current literature are also published in the scientific journals (for example, Ryland *et al.* [26]).

Case Histories

Green River Murders. In 1982, five women were murdered and left in or near the Green River, Washington, USA. Over the next several years, more bodies were found in the same area. In many cases, months or even years had passed since the victim's disappearance, and all that was found were skeletal remains. Eventually, 49 victims were listed as victims of the Green River Killer. The murders remained unsolved for nearly two decades.

In 1987, police took saliva samples and other evidence from a man called Gary Ridgway but did not find anything that linked him to the murders.

By 1991, the killings stopped and the case was dormant. Then in 2001, with DNA testing technology becoming more widespread, police began reexamining some of the evidence they had collected during the 1980s. They discovered a DNA match between Ridgway and semen taken from four of the victims.

Police arrested Ridgway and charged him with murder. He appeared to have been absent or off duty from work on every known occasion when a victim had disappeared.

His attorney planned to make the argument that the DNA evidence proved nothing more than that he had been a customer of the women. Prosecutors were hesitant about basing their case nearly exclusively on the DNA evidence. Police turned to forensic scientist Skip Palenik in an effort to find additional evidence to link Ridgway with the victims.

Ridgway had a job painting trucks during the time when the murders occurred. Paint samples had been collected from his car, clothes, and work locker and also from the victims. Several thousand small paint samples associated with Ridgway were analyzed by FTIR microscopy. When Palenik found a sample

with a matching infrared spectrum, he examined it using high magnification polarized light microscopy to identify the pigment, and X-ray spectroscopy to determine the elemental composition. While there were a number of initial spectral matches, none of the samples held up all the way through the analysis process.

"I had completed my assignment but I had a hunch that there was a lot more to this aspect of the case than met the eye," Palenik said. "I asked the prosecutors to provide me with Ridgway's and the victims' clothing so I could take a much closer look."

Palenik used reusable cassette-type filters connected to a vacuum cleaner, and collected particles with sizes in the 20–100-μm range, only about one hundredth the size of the particles he had examined in the first phase of his analysis. Spray painting guns, such as those used by Ridgway in his job, generate tiny spheres of paint in the micron range that are so light they float in the air rather than fastening themselves to the object being painted.

Palenik collected tiny particles and then, viewing them through a powerful optical microscope, picked them out with tungsten needles and placed them on microscope slides for analysis.

The fact that no further sample preparation was required for IR microscopy was essential because it would have been difficult or impossible with samples this small. It was possible to generate usable IR spectra from particles as small as $10\,\mu m \times 10\,\mu m$.

Palenik said, "As the Green River investigation demonstrates, the smaller the particle you can analyze the better chance you have to solve the case. This investigation also illustrates another important advantage of this instrument. It can save huge amounts of time by automating the process of checking a sample against a reference library."

He found hundreds of these tiny spheres in many colors on the clothes of six of the victims, including two that had been linked to Ridgway by DNA evidence. Most of the samples were of Imron paint, a very rare type that was used extensively in the paint shop where Ridgway worked at Kenworth Trucks. Working with Dupont, who manufactured the paint, he was able to tie the samples to paint that Ridgway had been using around the time of the murders.

On the basis of this evidence, prosecutors charged Ridgway with further murders. In November 2003, shortly before the trial was to begin, Ridgway confessed to murdering 48 women. He said that these were only the ones he could remember out of a total that he estimated at 70.

Ridgway's attorney was quoted in the press as saying that the paint evidence was crucial in his client's change of heart. "When this paint business came up, here you have something other than DNA – it's a particular paint you can link to Gary – and you start saying 'Well, here are seven dead women and they all can be linked to Gary one way or another,' Savage said." "What are the odds of that happening by accident?" [27].

Murder of Lord Mountbatten. In Europe, paint has also provided crucial evidence in high profile court cases. One of the first major cases in which the Irish Forensic Science Laboratory in Dublin was involved was the murder in 1979 of Lord Mountbatten, a cousin of Britain's Queen Elizabeth II, by the illegal Irish Republican Army (IRA).

He was going fishing in his green, wooden boat, *Shadow V*, when an explosion ripped through the vessel, killing him and three others and badly injuring three more.

That morning, before the explosion, two men were stopped by a police checkpoint more than 100 km away and questioned. One denied putting any bomb on the boat – but at this stage there had been no explosion and the police did not know what he was talking about. The second man, Thomas McMahon, was a known IRA man and explosives expert.

When news came of the explosion on Lord Mountbatten's boat, the two suspects were detained and their clothing and samples from their car were taken.

At the laboratory, flakes of two-layer green paint were recovered from McMahon's jacket and boots and from the passenger seat of the men's car. This paint matched the green paint recovered from the bodies and taken from the boat.

In addition, grains of sand and traces of nitroglycerin recovered from the suspects added to the trace evidence in the case against them. McMahon was convicted of murder and sentenced to life imprisonment [28].

Graffiti. Graffiti is a widespread phenomenon arising, for example, from simple vandalism or from the actions of racist or activist groups. The financial cost

of removal can be substantial and it is treated very seriously by transport police.

Allegation of criminal damage involving one suspect and two submitted control scrapings from a control box at the side of a London underground station.

A suspect was arrested running from a scene where the spraying of graffiti had been witnessed. A hooded top, seized soon after arrest, was submitted to the laboratory and microscopically examined. It was found to have a moderate concentration of paint balls on the outside surface, the majority of which were located on the front of the right sleeve and cuff. Both control paints were red and were optically indistinguishable from each other, but were chemically different when tested using FTIR. Red paint was the only color found on the surface of the hooded top. When tested microscopically, this red paint was indistinguishable from both of the red control paint samples. Both the control paints and these recovered paint balls contained red and brown particles when examined with transmitted, polarized light and did not fluoresce when observed under light passed through blue and ultraviolet filters. The recovered paint was indistinguishable from one of the submitted controls when analyzed by FTIR. The finding of red paint, indistinguishable from one of the submitted control samples on a hooded top, was deemed to provide *moderately strong scientific support* for the proposition that the wearer of the top had been in the vicinity of this graffiti while it was being sprayed [29].

References

[1] Woodbridge, R. (1991). *Principles of Paint Formulation*, Taylor and Francis.

[2] EUROPEAN PAINT GROUP (2006). *Guidelines for Best Practice in Forensic Paint Examination*, ENFSI.

[3] Brezinski, DR (1994). *SciQuest Paint Tutorial CD*, Consolidated Research, Kingsford, MI, USA, Vol. 1.

[4] Bentley, J. (2001). Composition, manufacture and use of paint, (Ch 7), *Forensic Examination of Glass and Paint*, B. Caddy, ed, Taylor and Francis, New York, London, pp. 123–141.

[5] Buzzini, P., Massonnet, G., Birrer, S., Egli, N., Mazzella, W.D. & Fortini, A. (2005). Survey of crowbar and household paints in burglary cases – population studies, transfer and interpretation, *Forensic Science International* **152**, 221–234.

[6] Buzzini, P., Massonnet, G. & Mizrahi, S. (2003). Interpretation of household paint transfer: a burglary case with foreign paint traces found on the blade of a crowbar, *Forensic Science International* **136**(Suppl 1), 355–356.

[7] Lau, L., Callowhill, B.C., Conners, N., Foster, K., Gorvers, R.J., Ohashi, K.N., Sumner, A.M. & Wong, H. (1997). The frequency of occurrence of paint and glass on the clothing of high school students, *Journal of Canadian Society of Forensic Science* **30**, 233–240.

[8] Pearson, E.F., May, R.W. & Dabbs, M.D.G. (1971). Glass and paint fragments found in men's outer clothing, *Journal of Forensic Sciences* **16**, 283–300.

[9] Helstroffer, S., Espanet, B. & Milet, S. (2003). Class identification of rockets types by paint analysis. A new way? *Forensic Science International* **136**(Suppl 1), 353–354.

[10] Boynton, R.M. (1979). *Human Color Vision*, Holt, Rheinhart and Winston, New York.

[11] ASTM E1610-02 (2008). *Standard Guide for Forensic Paint Analysis and Comparison*, American Society for Testing & Materials, Philadelphia.

[12] Beveridge, A., Fung, T. & MacDougall, D. (2001). Use of infrared spectroscopy for the characterization of paint fragments (Ch. 10), in *Forensic Examination of Glass and Paint*, B. Caddy, ed, Taylor and Francis, New York, London, pp. 183–241.

[13] Henson, M.L. & Jergovich, T.A. (2001). Scanning electron microscopy and energy dispersive X-ray spectrometry (SEM/EDS) for the forensic examination of paints and coatings (Ch 11), in *Forensic Examination of Glass and Paint*, B. Caddy, ed, Taylor and Francis, New York, London, pp. 243–272.

[14] ENFSI (2007). *ENFSI Best Practice Guide: SEM/EDS and Paint Analysis*, ENFSI.

[15] Challinor, J.M. (2001). Pyrolysis techniques for the characterization and discrimination of paint (Ch 9), in *Forensic Examination of Glass and Paint*, B. Caddy, ed, Taylor and Francis, New York, London, pp. 165–183.

[16] ENFSI (2007). *ENFSI Standard Guide for Microspectrophotometry and Colour Measurement in Forensic Paint Analysis*, ENFSI.

[17] Stoecklein, W. (2001). The role of colour and microscopic techniques for the characterization of paint fragments (Ch 8), in *Forensic Examination of Glass and Paint*, B. Caddy, ed, Taylor and Francis, New York, London, pp. 143–163.

[18] Cousins, D.R. (1989). The use of microspectrophotometry in the examination of paints, *Forensic Science Review* **1**, 141.

[19] Buzzini, P., Massonnet, G. & Monard-Sermier, F. (2006). The micro Raman analysis of paint evidence in criminalistics: case studies, *Journal of Raman Spectroscopy* **37**(9), 922–931.

[20] Hobbs, A.L. & Almirall, J.R. (2003). Trace elemental analysis of automotive paints by laser ablation inductively-coupled plasma mass spectrometry, *Analytical and Bioanalytical Chemistry* **376**, 1265.

[21] Almirall, JR., Umpierrez, U., Castro, W., Gornushkin, I. & Winefordner, J. (2005). *Forensic elemental analysis*

of materials by laser induced breakdown spectroscopy (LIBS), Proc SPIE, v. 5778, pp. 657–666.
[22] Piotrowski, G. (1999). European collection of automotive paints (EUCAP), *Problems of Forensic Sciences* **V224**, 19.
[23] Willis, S., McCullough, J. & McDermott, S. (2001). The interpretation of paint evidence, (Ch 12), in *Forensic Examination of Glass and Paint*, B., Caddy, ed, Taylor and Francis, New York, London, pp. 273–287.
[24] Seccombe, A.L. (2001). *Discrimination and Evidential Value of Vehicle Paint in Forensic Casework*, Master of Science thesis, University of Auckland, New Zealand.
[25] Edmondstone, G., Hellmann, J., Legate, K., Vardy, G.L. & Lindsay, E. (2004). An assessment of the evidential value of automotive paint comparisons, *Canadian Society of Forensic Science* **37**, 147.
[26] Ryland, S.G., Jergovich, T.A. & Kirkbride, K.P. (2006). Current trends in forensic paint examination, *Forensic Science Review* **18**, 97.
[27] Moving IR spectroscopy down to the micron level puts serial killer behind bars, Perkin Elmer case study, (2005.
[28] O'Connor, N. (2001). *Cracking Crime*, O'Brien Press, Dublin.
[29] Marsh, L. (2007). Some call it art: case studies investigating the spraying of illegal graffiti in the UK, *NIJ/FBI Trace Evidence Symposium*, Clearwater Beach, August 2007, http://projects.nfstc.org/trace/.

JOHN MCCULLOUGH

Paint: Interpretation

Introduction

Paint may be used as an investigative or an evaluative tool. Here, the focus is on interpretation, which is fundamental for paint and other transfer evidence (*see* **Paint**). When comparing two paint samples with a given analytical sequence, two main outcomes are possible. They are the following:

- The two samples can be differentiated.
- The two samples are not differentiated, they are analytically indistinguishable.
- There is a physical fit (rare).

In the first scenario, given that the reference paint is representative of the source material, a common origin can be excluded. In the second scenario, the examinations reveal a concordance in properties, but are these properties common or rare?

The authors believe that the forensic scientist is in the best position to give an opinion on the value of this concordance. In general, paint interpretation is based on the frequency of occurrence of the measured characteristics. Very common paint types have less evidential value than very rare paints. An estimate of the frequency of occurrence of the different characteristics measured for each paint type would be an invaluable tool.

Even if the literature on paint interpretation is rather sparse in comparison to other trace evidence like fibers or glass, there are surveys or information that can help the scientist to evaluate the strength of evidence. The main ones are the following:

Frequency figures:
- population studies and frequency distribution;
- information on batch variations; and
- databases/reference collections.

Others:
- expert opinion;
- paint found at random on clothing and surveys on paint traces;
- transfer and persistence; and
- a likelihood ratio approach.

Paint experts also rely on data concerning the discrimination power of the technique, or the sequence of techniques, used to present their findings. The discriminating power is calculated according to the formula from [1] and is usually mentioned in population studies.

Automotive Paints

Introduction

In the automotive industry, the application of the different paint layers to the metallic automotive body follows several steps. The process stages for original equipment manufacturer (OEM) paints can be simplified as follows [2]:

- *Metal pretreatment*: degreasing, passivation, phosphatage layer.
- *Electro dip coating*: application of the primer layer by electro deposition.

Table 1 Automotive paints, population studies

Authors	Countries	Number of samples	Methods	Number of nondifferentiated pairs
Gothard [3]	Australia	500	Optical: color, layer sequence Chemical tests Analytical: infrared spectroscopy, emission spectrography, pyrolysis GC	2
Ryland and Kopec [4]	USA	200	Optical: color, layer sequence Chemical tests Analytical: infrared spectroscopy, emission spectrography, pyrolysis GC, X-ray fluorescence, NAA	0
Gothard and maynard [5]	Australia	500	Optical: color, layer sequence Chemical tests Analytical: infrared spectroscopy, pyrolysis GC	3
Massonnet [6]	Switzerland	124 Gray metallic	Optical: color, layer sequence Analytical: infrared spectroscopy, MSP	4
Reeza Alwi and Kuppuswamy [7]	Malaysia	100	Optical: color, layer sequence	0
Edmonstone *et al.* [8]	Canada	260	Optical: color Analytical: infrared spectroscopy	2 (1 with the analysis of all layers)

- *Fillers*: application of the primer surface layer.
- *Finish or topcoat*: usually applied in one layer for solid paints and in two layers for metallic or effect finishes (first the effect layer is applied then a clearcoat).

Other original layers can be added to this sequence like protection against chipping by stone, sealing layer, or other extra layer to increase the quality of the final product. During this process, nonoriginal or repair layers can also be added at any step if paint defects are observed.

Automotive paint traces are mostly found in road accident and are transferred to the recipient surface when contact occurs. Paint can be transferred to vehicles, pedestrians, or fixed objects. Paint may be the only evidence in hit-and-run cases. Paint smears or fragments may also be found on objects or bodies transported in cars.

Population Studies

Several authors published comprehensive studies concerning the differentiation of paints collected at random on vehicles in different countries [3–8]. A summary of these surveys is presented in Table 1. These researches provide details on the distribution of automotive paints using a chosen analytical sequence.

Some general conclusions may be drawn from these population studies as follows:

- Optical examinations alone are very discriminating. According to Ryland and Kopec [4], 87% of their sample set is discriminated on the basis of color and layer sequence.
- All the authors agree that the layer sequence is the most significant point of comparison. Figure 1 shows the frequency distributions of the paint samples according to their number of layers in the different surveys. Gothard [3] states that "a large number of layers agreeing with regards to color, thickness, and layer sequence can be taken as proof of common origin without further examination". According to Ryland and Kopec [4] "... the probability of two paint chips originating from different sources is extremely remote

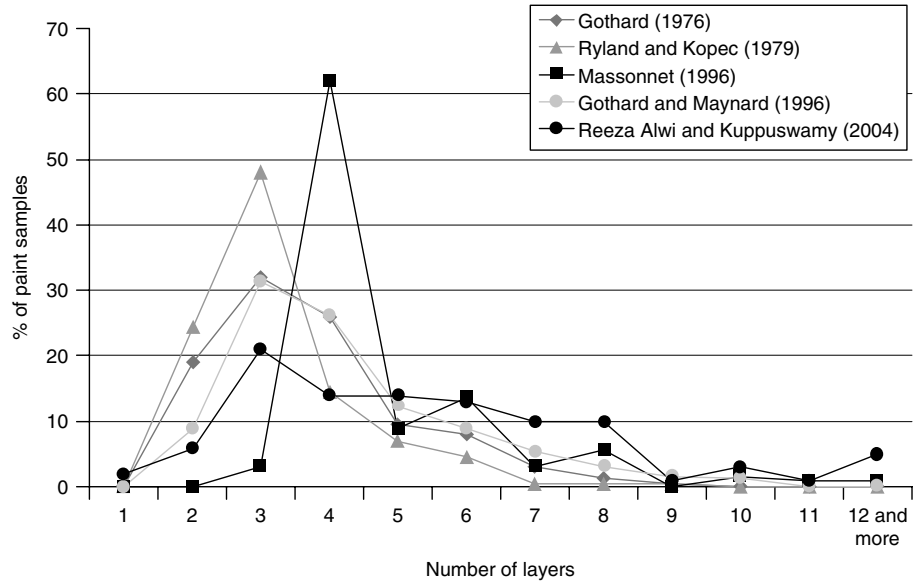


Figure 1 Frequency distribution of the paint samples in the different surveys according to their number of layers

when they have numerous layers (six or more) consistent in color, tint, type of finish, and layer thickness".

- When a complete sequence of analysis (optical and analytical) is applied, non-OEM or refinished paints are easily differentiated. The number of nondifferentiated pairs is low (between zero and four) and these samples are all OEM paints belonging to vehicles of the same make, model, color, and approximate year of production.
- Edmonstone *et al.* [8] compared the topcoat only (microscopy and chemical composition). Only two sample pairs were indistinguishable: they belonged to vehicles of the same make, model, and production plant. The first pair belonged to vehicles from 1995 and 1997, and the second pair had been produced the same year (1989). When all layers where compared and analyzed using FTIR, samples from the first pair were distinguished, and the second pair remained nondifferentiated.

Globally, these surveys show that if two paints with numerous nonoriginal layers are indistinguishable using a state-of-the-art analytical sequence, there are generally few doubts that they have a common origin [2].

Vehicle Topcoat Colors: Frequency Distribution

Several authors worked on the frequency distribution of vehicles' topcoat color [9–14]. Vehicles in circulation were counted according mainly to their topcoat color, but sometimes also considering their make, model, or year of production.

The main outcomes of these surveys are the following:

- Color categories chosen by the authors are more detailed than the one used in the official statistics and can include the difference between effect and solid paints. The frequencies obtained are lower than in the official register. The most common color group depends on the country considered and has a frequency of occurrence between 20 and 25%.
- When combining detailed color group and car make, the most common combination has a frequency of occurrence of less than 5% (between 3.2 and 4.8% depending on the survey). Thus, the occurrence of vehicle from particular make and color on the road can be considered from a statistical point of view as an unusual event.
- Globally, there is a rather good accordance between the data observed (color and make) and

the corresponding numbers in the official statistics of vehicles in a given country or geographic area.

Tippett [9] counted the number of cars from a specific make and color in 100 groups of 100 cars. Two car makes were chosen, one being the most common (10.8%, frequency of occurrence) and the other one being rare (1%). He found that the data obtained follow a Poisson distribution and thus that they can be statistically predicted. With this model, it is possible to predict the number of times a particular colored model is seen on the road in a given time frame depending on the density of traffic.

McDermott *et al.* [14] collected and used the data obtained in a likelihood ratio framework for the interpretation of automotive paint evidence.

Batch Variations

The conclusion that was drawn from population studies is that if two multilayer paint systems are nondifferentiated with a complete sequence of analyses, they are original OEM paints from vehicle of the same make and model and approximate year of production.

The next question is, are we able to distinguished OEM paints or, according to Stoecklein and Palenik [15]. "In how many vehicles produced by one manufacturer and painted in the same color, can the paint be distinguished?"

Different batches from three of the most frequent colors applied to vehicles produced between 1989 and 1994 were analyzed by [15]. Their conclusions are that by using a broad spectrum of analytical techniques (microscopy, fourier transform infrared spectroscopy (FTIR), Microspectrophotometry (MSP), Raman, and pyrolysis gas chromatography mass spectrometry (Py-GC-MS)), it is possible to discriminate between OEM topcoat of different batches. Within one batch only a limited number of automobiles can be coated. After being in use, only few of these cars have coatings, which still show the same physical and chemical features. Therefore, one can assume that by analyzing OEM paints, the expected weight of the evidence will be strong or very strong [15].

To evaluate the correspondence of OEM multilayer paint systems, the following facts must be considered [2]:

- One batch of paint (topcoat) can be used to paint 8–3,500 vehicles.
- About 40% of the cars are locally repainted at the production plant.
- Change in batches can also occur for the other layers.
- In circulation, paints are differently affected by weathering and can be repainted after damage.

When using a comprehensive range of methods, in particular quantitative analysis instead of only qualitative, it is possible to differentiate between batches as well as between different conditions of weathering [15, 16]. Thus, the number of vehicles on the street with exactly the same OEM layer sequence is very low. Stoecklein [2] concludes that "there is very little likelihood of a second vehicle, with an identical characteristic profile of its coating material, being on the street at the time of the accident, in the vicinity of the scene of crime. The evidential value of such result is to be regarded as very high".

Databases/Reference Collections

Two main databases of automotive paints are available to police forces. Paint data query (PDQ) is mainly used in Canada and in the United States [17, 18]. The European collection of automotive paint (EUCAP) is used by the European laboratories [2]. Japanese police forces also have their own database for Asian cars. These collections are continuously updated.

The aim of these databases is to provide information to the police about cars (make, model, color, and production years) based on the analysis of paint traces found in hit-and-run accidents. Infrared spectra of thousands of samples are recorded for each different paint layer.

These databases can also provide information concerning the commonness/rarity of different paint types. For example, a clearcoat based on acrylic melamine and styrene will be present on several different vehicles and thus will have a rather low evidential value. On the other hand, a very special paint corresponding to only one vehicle will have a higher evidential value. Care should, however, be taken as the frequency distributions obtained in such databases do not represent the population of vehicles in circulation in a specific country but the variations observed between the different car make and color present in the database.

More generally, there should be reference collections available for each of the techniques used in

Table 2 Conclusions used by paint examiners for paint exchange scenarios [21]. Reprint with permission of Thomson Publishing Service

	Conclusion (%)				
Scenario	Slight support	Support	Strong support	Very strong support	Conclusive
A	**77.4**	19.4	1.6	1.6	0
B	15.3	**64.5**	15.3	4.8	0
C	4.0	**68.5**	23.4	4.0	0
D	0.8	8.9	**51.6**	34.7	4.0
E	0.8	21.8	**57.3**	19.4	0.8
F	0.8	8.1	35.5	**46.8**	8.9
G	0.8	19.4	**44.4**	31.5	4.0
H	0.8	1.6	10.5	**51.6**	35.5

A, one layer transferred in one direction;
B, one layer transferred in each direction;
C, multilayer manufacturer's finish transferred in one direction;
D, multilayer manufacturer's finish transferred in each direction;
E, multilayer manufacturer's finish transferred in one direction and one layer transferred in the other direction;
F, multilayer nonmanufacturer's finish transferred in one direction and one layer transferred in the other direction;
G, multilayer nonmanufacturer's finish transferred in one direction;
H, multilayer nonmanufacturer's finish transferred in each direction.

the different laboratories. A few surveys can also be found in the literature. For example, using Raman spectroscopy it was determined that 59% of the 27 light red automotive paint samples analyzed contain a specific red pigment (CI PR254). As this pigment is very commonly used in red paints, the evidential value of such finding will be low. On the other hand, a very rare combination of pigments will have a higher evidential value [19].

Expert Opinion

McDermott and Willis [20] circulated a questionnaire to 235 paint examiners. Different hypothetical automotive paint transfer scenarios were presented and the respondents were requested to use a scale of conclusions ranging from slight support to conclusive. Table 2 illustrates the hypothetical transfer scenarios and the answer given by the paint experts [21].

This survey gives an insight into the value placed on such evidence by forensic scientists working in the field of paint analysis. This survey globally shows that a multilayer repaint system has more evidential value than a multilayer OEM paint system and that a multilayer paint has more evidential value than a single layer paint. Scientists also give more evidential value to cross transfer compared to a single direction transfer if the same type of paint is considered [20].

Household Paints

Introduction

Paints covering surfaces other than vehicles are commonly called *household paints* [21]. The interpretation of this kind of paint evidence is quite different from the interpretation of vehicle paint. Vehicle paint follows a definite sequence of color, number, and thickness of layers, which are specified by the manufacturer. Household paint is more often encountered in single layer, but is considered to provide strong evidence due to the large range of paints available.

Household paint traces encountered in forensic cases can originate from buildings (architectural paints), tools, spray cans, etc. In a burglary, paint traces can be transferred on the tool used to force a door, and if the tool is covered by a paint layer, paint particles from the tool can be possibly found on the door. In graffiti case, the paint can be compared with a reference spray can, but microscopic wet aerosol paint droplets can be transferred on the clothes and on the hands of the offender and could provide highly significant evidence. In these two examples, three kinds of paints (building, tools, spray) and four kinds of supports (door, tool, clothes, hands) are encountered. To evaluate the evidential value of these paint traces, the forensic expert will need information about the frequency of the analytical characteristics of the

type of paint encountered, and about the frequency of pounding randomly paint traces on the kind of support on which the traces were found.

Population Studies

Several authors have studied the discriminating power of an analytical sequence on different populations of household samples [22–29]. Table 3 shows a summary of these surveys.

These researches show that certain binder types are more commonly encountered. They also demonstrate the high potential of common analytical sequences to discriminate between household paint samples, even for samples of the same color and binder type.

Table 3 Household paints, population studies

Authors	Samples	Methods	Commonest binders/ extenders combinations		Number of subgroups (maximum number of items per group)	Discriminating power of the whole sequence
Tippett *et al.* [22]	200 paint samples from buildings	Microscopy	–			0.999
		Solvent tests				
		Spectrography				
		Py-GC-MS				
May and Porter [23]	31 household gloss paints (11 white, 10 red and 10 green)	Microscopy	ALK	80%	8 (3)	0.93 (white)
		Emission spectrography	ALK + PUR	10%	10 (1)	1 (red)
		Solvent tests	ALK + fire retardant	10%	10 (1)	1 (green)
		FTIR				
		X-ray diffraction				
		Py-GC-MS				
		Py-FTIR				
Laing *et al.* [24]	169 household gloss paints (divided in 12 color categories)	Microscopy	–			0.89 to 1
		MSP				
Castle *et al.* [25]	100 case-openers	MSP	ALK	61%	–	–
		Py-GC-MS	ALK + STY	8%		
			EPOXY	8%		
			ALK + VINYL TOL	6%		
			ACR	6%		
Buzzini and Massonnet [26]	40 green spray paints	FTIR	ALK OPH + NCL	28%	31 (6)	0.98
		Raman	ACR	23%		
			ALK OPH + $BaSO_4$	15%		

(*continued overleaf*)

Table 3 (*continued*)

Authors	Samples	Methods	Commonest binders/ extenders combinations		Number of subgroups (maximum number of items per group)	Discriminating power of the whole sequence
Govaert and Bernard [27]	51 red spray paints	Microscopy	ALK + NCL	45%	37 (4)	0.988
		FTIR	ALK	24%		
		X-ray fluorescence				
Buzzini *et al.* [28]	41 blue crowbars	FTIR	ALK OPH + NCL	71%		
			ALK OPH + $CaCO_3$	17%		
			ALK OPH	7%		
Gosse *et al.* [29]	38 black spray paints	FTIR	ALK OPH +NCL	42%	34 (2)	0.994
		X-ray fluorescence	ALK OPH	21%		
		Py-GC-MS	ACR	13%		
			ALK OPH +NCL + STY	11%		
Gosse *et al.* [29]	38 black spray paints	FTIR	ALK OPH +NCL	42%	34 (2)	0.994
		X-ray fluorescence	ALK OPH	21%		
		Py-GC-MS	ACR	13%		
			ALK OPH +NCL + STY	11%		

ACR, acrylic; ALK, alkyd; ALK OPH, orthophthalic alkyd; NCL, nitrocellulose; PUR, polyurethane; STY, styrene; VINYL TOL, vinyl toluene

For example, all the red spray paints analyzed by [27] are alkyd based, but these samples can be classified into 37 different groups. The discriminating power of this sequence (microscopy, FTIR, and X-Ray fluorescence) is 0.988.

Batch Variations

As for vehicle paints, if two household paints are indistinguishable, the question of batch to batch variation is asked. Could this paint trace come from any other batch of the same brand, or is it possible to exclude other batches and to link the trace to a given batch?

Inkster *et al.* [32] compared 14 batches of the same architectural acrylic white paint by microscopy, microspectrophotometry, FTIR, micro-XRF and Py-GC-MS. Only one batch could be differentiated from the other. In this study, only qualitative analysis was used. On the other hand, [33] observed significant differences between two pairs of batches, using statistical analysis of the results by principal component analysis (PCA). These results suggest the high potential of chemometric tools or semiquantitative analysis to differentiate between closely related spectra.

Databases/Reference Collections

Databases for household paint are less developed than for vehicle paints. However, some specific databases exist. For example, the European database of spray paint contains the infrared spectrum of 209 spray paints. This database can be used to identify an unknown spray paint trace, but also to evaluate the frequency of the infrared characteristics of a given spray paint. The French database of black

spray contains the infrared spectrum of 38 black spray paints with information about the Py-GC-MS results [29].

Each laboratory has the opportunity to create their own databases, for example, containing traces found in caseworks or various reference samples.

Paint Found at Random on Clothing and Surveys on Paint Abrasion Traces

Introduction

When paint cannot be unambiguously attributed to criminal activity, it is then necessary to assess if it could be present as background. Population studies demonstrate the potential of analytical methods, but cannot be used to evaluate the frequency of paint traces found on different kind of supports. To do so, it is necessary to collect and analyze traces, and not control samples. Indeed, background paint may not have the same origin (i.e., vehicle, household, tools) as samples. These backgrounds or so-called random paint studies are more complicated to carry out, because of the difficulty of pinding samples and the time needed to collect and count them. These studies often contain information about the frequency of the color, the size, and number of layers of the traces. Analytical results are less often available.

Paint Found at Random on Clothing

The two available studies about paint found at random on clothing show contradictory results. Lau *et al.* [31] found that on the majority of searched items no paint traces are recovered. If traces are recovered, it is rare to find more than one fragment. On the other hand, in the study of [30] four or more fragments are recovered on 80% of the items. This difference, also observed for glass, as the authors looked at both paint and glass sample, could be explained by the size of the fragments: for 52% of the recovered fragments by [30] the size is less than 0.3 mm. Lau *et al.* [31] indicate that the majority of the recovered paint fragments were smaller than 1×1 mm, but no information is available about the minimum size considered. In addition, it must be noted that Pearson *et al.* looked at particles recovered in pockets and cuffs, where one might expect to find more particles, because of longer retention time. Table 4 shows a summary of the main results obtained in both studies.

Table 4 Paint fragments found at random on clothing

Authors	Items searched	Commonest colors of paint traces		Number of paint fragments on each item
Pearson *et al.* [30]	100 suits	Red	29%	0 : 3%
		Green	23%	1 : 8%
		Cream	14%	2 : 3%
		Blue	12%	3 : 6%
				4 and more : 80%
Lau *et al.* [31]	216 upper body garments	Yellow	31%	0 : 86%
		Pink	18%	1 : 12%
		Red	15%	2 : 1%
				3 : 1%
	213 lower body garments	Yellow	32%	0 : 88%
		Pink	22%	1 : 10%
		White	8%	2 : 2%
				3 : 0.5%
	164 pairs of footwear	White	27%	0 : 77%
		Red	21%	1 : 10%
		Black	15%	2 : 4%
		Yellow	13%	3 : 2%
				4 and more : 5%

Table 5 Survey of abrasion paint traces

Authors	Items searched	Commonest colors of paint traces		Number of types of paint traces on each item
McDermott *et al.* [14]	1000 vehicles	–		0 : 91% one and more : 9%
Ranzi *et al.* [34]	147 foreign traces found on damaged cars	Gray	26%	–
		Red	22%	
		Blue	17%	
		Green	6%	
		White	6%	
		Yellow	6%	
		Black	5%	
Buzzini *et al.* [28]	207 crowbars	White	61%	0 : 34%
		Red	18%	1 : 50%
		Green	7%	2 : 13%
		Brown	5%	3 : 2% 4 : 1%

Surveys on Paint Traces

Studies on paint found on clothing concern paint fragments, which have to be differentiated from paint abrasion traces. A few studies on paint abrasion traces are available for cars [14] and tools [28, 34], see Table 5.

Transfer and Persistence

Again, when paint cannot be unambiguously attributed to criminal activity, knowledge of transfer and persistence phenomena becomes important to assess the value of paint traces.[a] The expert should be able to evaluate if the quantity of paint recovered is consistent with the alleged contact. However, transfer experiments are difficult to perform, especially with vehicle paints. For this reason, the number of studies involving transfer experiments is low.

Generally, a strong contact is necessary to remove paint from its support (metal, wood, etc.). This can be caused by an impact or an abrasion. The quantity of paint transferred depends on the force applied, the condition of the paint, the duration of the contact, and the nature of the object in contact with the paint. This concerns dry paint, but sometimes wet paint could be transferred, for example, in cases involving spray paint, where wet paint could be transferred on the hands and on the clothing of the suspect [21]. Moreover, the persistence of wet paint is expected to be very good, because after drying the paint is strongly attached to the recipient.

Krausher [35] studied the transfer of droplets of paint to clothing during the use of aerosol paint. The density of droplets on the clothing of a sprayer is high, especially in areas exposed to the paint cloud. The size of the paint droplets is generally about 10–40 microns. Owing to their small size, the persistence of the paint droplets is lower than that for large wet paint transfer. The droplets are easily removed from the surface by washing. Marin *et al.* [36] have also studied the transfer of paint droplet from spray paint and obtained quite similar results.

Buzzini *et al.* [28] studied the cross transfer of paint between a crowbar and a painted wooden surface. In every simulated contact, a reciprocal transfer of paint was observed. A correlation between the quantity of household paint transferred in one direction and the quantity of tool paint transferred in the other direction was observed.

A Likelihood Ratio Approach Within a Hierarchy of Propositions

Bayes theorem shows how the evidence influences the probabilities associated with two alternative hypotheses. Three different levels of hypothesis can be considered: source level, activity level, and offense level [37]. For example, in a case where paint traces are recovered on a tool and compared with the paint of a forced door, the hypothesis could be as follows:

Source level

H_p: the traces found on the tool come from the paint of the forced door;
H_d: the traces found on the tool do not come from the paint of the forced door and are present as background.

Activity level

H_p: the tool has been used to force the door;
H_d: another tool has been used to force the door.

Offense level

In most cases, the information to consider is neither known by the forensic scientist (e.g., modus operandi of the suspect), nor in his/her field of expertise. It is therefore very rare for paint specialists to assess likelihood ratios at the offense level, such as follows:

H_p: the suspect has forced the door using that tool;
H_d: the suspect has not forced the door.

The assessment of the source level depends on analytical information obtained during examination [38]. The frequency of the analytical characteristics of the evidence is the important factor of the likelihood ratio. This frequency is evaluated in a given population. The choice of the relevant population is crucial, because a wrong population could drastically change the value of the frequency. In transfer evidence, the relevant population is often traces found on a given support. For example, in the case of paint traces recovered on a tool, the relevant population could be paint traces recovered on a population of tools seized by the police on suspects. Paint traces studies give the appropriate information to the evaluation of the frequency. However, the number of population studies of paint traces is quite low and in some cases the corresponding trace population study is not available. In this situation, general population studies could be used. Generally, the role of the expert is to make an inventory of all the information found in the literature, and to estimate the frequency of the paint trace on the basis of these studies.

The assessment of the activity level is more complex. In addition to the frequency parameter, transfer and background parameters have to be taken into account. Transfer parameters concern the probability that paint traces were transferred to the given support, persist, and are recovered. Compared to glass or fiber evidence, the phenomenon of paint transfer is not fully understood, especially the transfer of vehicle paint. Background parameter concerns the probability of finding such a paint trace at random on this given support. Background parameter could be estimated on the basis of population studies of traces.

The number of studies about Bayesian interpretation of paint evidence is limited. McDermott *et al.* [14] and Willis *et al.* [21] calculated the likelihood ratio for various transfer scenarios for automotive paints and [28] used the Bayesian approach to evaluate the value of a cross transfer of paint between a tool and a forced door.

Conclusion

The ability of the expert to estimate the frequency of a specific paint type will rely mainly on experience and familiarity with the relevant literature as well as on access to representative databases. Other information like paint found at random or other expert opinion is also a great help to the interpretation of some cases.

Background data are available in the literature to help the scientist to assess the evidential value of his/her paint examination (automotive and household). Databases are very important to provide frequency of occurrence on different paint categories. If possible, these databases should be local to take into account the population of paint in a given geographic area. More work is needed principally on traces population studies and on paint interpretation using a likelihood ratio approach.

End Notes

[a.] If the paint is unambiguously attributed to the criminal activity, the probability of the evidence

being transferred, having persisted and being recovered, is one under the prosecution's proposition (i.e. Vehicle A*** hit vehicle B). The probability of the evidence given the defense proposition (i.e.Vehicle A*** did not hit vehicle B) is the frequency of the observed characteristics in the suspect population. It is certain that the evidence is not background.

References

[1] Smalldon, K. & Moffat, A. (1973). The calculation of discriminating power for a series of correlated attributes, *Journal of the Forensic Science Society* **13**, 291–295.

[2] Stoecklein, W. (1992). Die Verkehrunfallflucht: Kriminaltechnische Möglichkeiten der Aufklärung am Beispiel Autolack, *Schriftenreihe der Polizei-Führungsakademie* **1**, 36–59.

[3] Gothard, J.A. (1976). Evaluation of automobile paint flakes as evidence, *Journal of Forensic Sciences* **21**, 636–641.

[4] Ryland, S.G. & Kopec, R.J. (1979). The evidential value of automobile paint chips, *Journal of Forensic Sciences* **24**, 140–147.

[5] Gothard, J. & Maynard, P. (1996). Evidential value of automotive paint, *Proceedings 13th International Symposium of the ANZFSS*, Sydney.

[6] Massonnet, G. (1996). *Les peintures automobiles en criminalistique*, PhD thesis, University of Lausanne, Law faculty and School of forensic sciences, Switzerland.

[7] Reeza, A.A. & Kuppuswamy, R. (2004). Studies on the layer structure of paint flakes collected from motor vehicles in Kuala Lumpur, Malaysia, *Journal of Forensic Identification* **54**, 645–652.

[8] Edmondstone, G., Hellman, J., Legate, K., Vardy, G.L. & Lindsay, E. (2004). An assessment of the evidential value of automotive paint comparisons, *Canadian Society of Forensic Science Journal* **37**, 147–153.

[9] Tippett, C.F. (1964). Car distribution statistics and the hit-and-run driver, *Medicine, Science and the Law* **4**, 91–97.

[10] Ryland, S.G., Kopec, R.J. & Somerville, P.N. (1981). The evidential value of automobile paint. Part II: frequency of occurrence of topcoat colors, *Journal of Forensic Sciences* **26**, 64–74.

[11] Buckle, J., Fung, T. & Ohashi, K. (1987). Automotive topcoat colours: occurrence frequencies in Canada, *Canadian Society of Forensic Science Journal* **204**, 45–56.

[12] Volpé, G.G., Stone, H.S., Rioux, J.M. & Murphy, K.J. (1988). Vehicle topcoat colour and manufacturer: frequency distribution and evidential significance, *Canadian Society of Forensic Science Journal* **21**, 11–18.

[13] Stone, H.S., Murphy, K.J., Rioux, J.M. & Stuart, A.W. (1991). Vehicle topcoat colour and manufacturer: frequency distribution and evidential significance, part II, *Canadian Society of Forensic Science Journal* **24**, 175–185.

[14] McDermott, S.D., Willis, S.M. & McCullough, J.P. (1999). The evidential value of paint. Part II: a Bayesian approach, *Journal of Forensic Sciences* **44**, 263–269.

[15] Stoecklein, W. & Palenik, C. (1998). Forensic analysis of automotive paints: evidential value and the batch problem, *Proceeding of the 4th European Paint Group Meeting*, Paris.

[16] Stoecklein, W. & Fujiwara, H. (1999). The examination of UV-absorbers in 2-coat metallic and non-metallic automotive paints, *Science and Justice* **39**, 188–195.

[17] Cartwright, N. & Rodgers, P. (1976). A proposed data base for the identification of automotive paint, *Canadian Society of Forensic Science Journal* **9**, 145–154.

[18] Cartwright, N., Cartwright, L., Norman, E., Cameron, R., MacDougall, D. & Clark, W. (1982). A computerized system for the identification of suspect vehicles involved in hit and run accidents, *Canadian Society of Forensic Science Journal* **15**, 105–115.

[19] Massonnet, G. & Stoecklein, W. (1999). Identification of organic pigments in coatings: application to red automotive topcoats. Part III: Raman spectroscopy (NIR FT-Raman), *Science and Justice* **39**, 181–187.

[20] McDermott, S.D. & Willis, S.M. (1997). A survey of the evidential value of paint transfer evidence, *Journal of Forensic Sciences* **42**, 1012–1018.

[21] Willis, S., McCullough, J. & McDermott, S. (2001). The interpretation of paint evidence, in *Forensic Examination of Glass and Paint*, B. Caddy, ed, Taylor and Francis, London and New York, pp. 273–287.

[22] Tippett, C.F., Emerson, V.J., Fereday, M.J., Lawton, F., Richardson, A., Jones, L.T. & Lampert, S.M. (1968). The evidential value of the comparison of paint flakes from sources other than vehicles, *Journal of the Forensic Science Society* **8**, 61–65.

[23] May, R.W. & Porter, J. (1975). An evaluation of common methods of paint analysis, *Journal of Forensic Science Society* **15**, 137–146.

[24] Laing, D.K., Dudley, R.J., Home, J.M. & Isaacs, M.D.J. (1982). The discrimination of small fragments of household gloss paint by microspectrophotometry, *Forensic Science International* **20**, 191–200.

[25] Castle, D.A., Curry, C.J. & Russell, L.W. (1984). A survey of case-openers, *Forensic Science International* **24**, 285–294.

[26] Buzzini, P. & Massonnet, G. (2004). A market study of green spray paints by Fourier transform infrared (FTIR) and Raman spectroscopy, *Science and Justice* **44**(3), 123–131.

[27] Govaert, F. & Bernard, M. (2004). Discriminating red spray paints by optical microscopy, Fourier transform infrared spectroscopy and X-ray fluorescence, *Forensic Science International* **140**, 61–70.

[28] Buzzini, P., Massonnet, G., Birrer, S., Egli, N.M., Mazzella, W. & Fortini, A. (2005). Survey of crowbar and household paints in burglary cases – population studies, transfer and interpretation, *Forensic Science International* **152**, 221–234.
[29] Gosse, R., Milet, S. & Espanet, B. (2005). Discrimination of Black Spray Paints, *Proceedings of the European Paint and Glass Group Meeting*, Berlin.
[30] Pearson, E.F., May, R.W. & Dabbs, M.D.G. (1971). Glass and paint fragments found in men's outer clothing – report of a survey, *Journal of Forensic Sciences* **16**(3), 283–300.
[31] Lau, L., Beveridge, A.D., Callowhill, B.C., Conners, N., Foster, K., Groves, R.J., Sumner, A.M. & Wong, H. (1997). The frequency of occurence of paint and glass on the clothing of high school students, *Canadian Society of Forensic Science Journal* **30**(4), 233–240.
[32] Inkster, J., Maynard, P., Roux, C. & Fergusson, B. (2006). Intrasample vs intersample variability in architectural paint, *Fourth European Academy of Forensic Science Meeting*, Helsinki.
[33] Bell, S.E., Fido, L.A., Speers, J. & Armstrong, W.J. (2005). Rapid forensic analysis and identification of "lilac" architectural finishes using Raman spectroscopy, *Applied Spectroscopy* **59**(1), 100–108.
[34] Ranzi, R., Antonetti, G., Buzzini, P. & Massonnet, G. (2004). Population study of foreign traces recovered on bodies of damaged cars, *Proceedings of the 10th European Paint and Glass Group Meeting*, Prague.
[35] Krausher, C.D.J. (1994). Characteristics of aerosol paint transfer and dispersal, *Canadian Society of Forensic Science Journal* **27**(3), 125–142.
[36] Marin, D., Berger, N., Buzzini, P. & Massonnet, G. (2004). Transfer, detection and in situ Raman analysis of spray paint traces on clothes, *Proceedings of the 10th European Paint and Glass Group Meeting*, Prague.
[37] Cook, R., Evett, I.W., Jackson, G., Jones, P.J. & Lambert, J.A. (1998). A hierarchy of propositions: deciding which level to address in casework, *Science and Justice* **38**(4), 231–239.
[38] Aitken, C.G.G. & Taroni, F. (2004). *Statistics and the Evaluation of Evidence for Forensic Scientists*, John Wiley & Sons, Chichester.

GENEVIÈVE MASSONNET AND FLORENCE MONNARD

Palmprints *see* Friction Ridge Examination (Fingerprints): Interpretation of

Palynology

Introduction

One of the newer techniques now being used more frequently is forensic palynology: the collection and examination of pollen and spores associated with crime scenes, other types of illegal activities, or terrorism. The value of using pollen and spores (collectively called *palynomorphs*) as forensic tools relies on four important aspects.

First, many types of pollen and spore-producing plants disperse vast quantities of these palynomorphs into the air that are carried by air currents and eventually fall to the ground in a thin coating called the *pollen rain*. In some regions, the amount of pollen and spores dispersed is so great that exposed land and water surfaces turn yellow from the pollen rain. Although not a precise measurement of the surrounding vegetation, and thus by inference the climate of the area, the pollen rain in each region of the world is nevertheless a snapshot of that area and becomes a "pollen print" that can be used to assist in identifying the region.

Second, pollen and spores are microscopic in size, invisible to the naked eye, and can become trapped on almost any type of surface. This means that at any geographical location, pollen or spores from plants in that region, or more specifically, the pollen and spores from a specific crime scene can become evidence that will link a suspect or some object with the region or crime scene.

Third, there are nearly one-half million different plant species that produce either pollen or spores. Fortunately, each of these species produces pollen or spores that can be identified as coming from the parent plant; however, often differences in the pollen and spores of closely related species or even related genera may appear so similar that precise identification can only be achieved through detailed studies using the resolution capabilities of a scanning electron microscope (SEM) or transmission electron microscope (TEM).

Fourth, most pollen and spores are highly resistant to destruction or decay. This means that pollen and spore evidence from a region or crime scene can remain intact for years, hundreds of years, or even thousands and millions of years. This means that if crime scene evidence is handled correctly and

stored safely, years or decades later the trapped pollen and spores can still be recovered and used to assist investigators.

Even though the first reported use of pollen to help solve a crime occurred nearly 50 years ago, very few attempts to use this forensic tool occurred in the meantime, and even then the number of times it was used has been rare. Only recently, during this century, the number of applications and attempts to use pollen and spore evidence for forensic use has been increasing. As the number of times that pollen and spores are successfully used as evidence in forensic cases increases, so does the range of its potential applications. Recently solved criminal cases demonstrate that the forensic use of pollen and spores can be applied to cases of forgery, production and distribution of illegal drugs, assaults, robbery, rapes, homicide, genocide, terrorism, arson, hit and run crimes, counterfeiting of currency, identifying the origin of fake prescription drugs such as Viagra, and various other types of criminal activity. Pollen and spore evidence has been used to resolve a variety of civil cases involving forged documents, fake antiques, authentication of paintings by master artists, removal of artifacts from historic or archaeological sites, illegal poaching of animals or fish, and cases involving illegal pollution of the environment [1].

Successful use of pollen and spores in forensic applications may depend on the palynologist's knowledge and experience. One must know and understand the plant ecology and plant communities in and around the crime scene area or the region where pollen samples were collected. A sound understanding of palynology including knowledge of the production, dispersal patterns, preservation potential, and identification of pollen and spores is essential to understand and interpret the expected total palynomorph assemblage recovered in forensic samples. As with most types of forensic evidence timing is critical. The sooner the palynologist is called to investigate a crime scene, the more likely will be the potential value of the collected samples. Because pollen and spores are small, light, easily recycled, and present in abundant amounts in the atmosphere, a crime scene can quickly become contaminated by recycled palynomorphs accidentally carried into the area on the clothing or shoes of crime scene investigators, deposited out of the atmosphere, or from other noncrime scene sources if the forensic pollen sample collection is delayed or collected improperly. Under ideal circumstances, and with proper and timely collection of forensic pollen samples, one might even be able to determine the season of the year when a crime was committed and occasionally how long ago a crime was committed. Much of that information pertains to knowing the pollination cycles of various plants and being able to check for the presence of cytoplasm and the innermost wall of pollen grains, called the *intine*, which is composed of cellulose and degrades fairly rapidly after pollen grains are dispersed [2].

Pollen Types

Some of the most useful types of pollen and spores for forensics are the wind-pollinated types. This group includes the spore-producing plants such as fungi, ferns, and mosses as well as a wide range of pollen types produced by the gymnosperms (nonflowering seed-bearing plants such as pines, cedars, and spruce), and a significant number of angiosperms (flowering seed-bearing plants such as aspen, elms, and chestnuts). Because wind pollination is an inefficient method of dispersion, these plants must produce vast quantities of pollen or spores that are usually lightweight and are aerodynamically designed to travel easily in air currents. The enormity of pollen production in many of the wind-pollinated (anemophilous) plants is exemplified by statistics such as the following: a single shoot of marijuana (*Cannabis*) produces about 500 million pollen grains, one dock (*Rumex*) plant produces about 400 million pollen grains, a single panicle of sorghum (*Sorghum*) disperses 100 million pollen grains, and just one male strobilus on a branch of a lodgepole pine (*Pinus contorta*) produces over 600 000 pollen grains. In addition to these examples, many of the other wind-pollinated plants such as ragweed, grasses, some species of eucalyptus, oaks, hickory, walnut, birch, alder, and elms produce between 10 000 and 100 000 pollen grains per anther (the part of a flower that produces and contains pollen and is usually borne on a stalk). Other anemophilous plants, some of which are low pollen producers, still produce more than 10 times the amount of pollen per anther and flower than almost all species of insect-pollinated plants. The amounts of pollen dispersed annually by the wind-pollinated plants is so vast that their pollen can be found in almost every environment in the world and

the distribution of those pollen types in the pollen rain (total pollen deposited annually at any given location) of each region gives each locale its own pollen print. As such, these pollen types are the most common ones found in the fossil pollen record of a region and are also the most common types found in forensic pollen samples [3].

Some flowering plants live completely submerged in water, release their pollen underwater, and then rely on the pollen to float to the surface or ride water currents in an effort to accomplish fertilization. This method of transport, like the wind, is an inefficient method of pollination; therefore, like wind-pollinated plants, submerged plants produce high levels of pollen. These types of pollen could be found on the clothing, in the lungs, or possibly in the stomach of individuals who drown or were thrown into lakes or streams after being killed. Nevertheless, these types of pollen are often of little potential value for forensic work because they decay very easily and are difficult to recover without accidentally destroying them in the process.

The largest group of flowering plants is the insect or animal-pollinated types (entomophilous). This group depends on the transport of their pollen grains from the anther (male portion) of one flower to the stigma (female portion) of another by some type of insect (bee, wasp, beetle, moth, mosquito, and ant) or by some type of animal (hummingbirds, lizards, nectar-feeding bats, or other small mammals). The pollen grains produced by these entomophilous plants are generally ornate, have a surface covered with sticky lipids and waxes so that they attach easily to insects and mammal hairs, and most of these pollen types have a strong, thick outer wall (exine) that protects them from abrasion during transport and from rapid changes in humidity [4].

Because of the pollination efficiency of entomophilous plants, pollen productivity per anther and flower is much less than that in wind-pollinated plants. In maples (*Acer*), for example, each anther often contains no more than 1000 pollen grains and in flax flowers (*Linum*) each anther may contain as few as 100 pollen grains. In spite of the low pollen production in most entomophilous plants, they can often provide some of the most useful forensic clues. Because these pollen types are often large and heavy, and have a sticky surface, they are rarely cast adrift in wind currents and thus are rarely found in the normal pollen rain of a region. This means that it would be extremely rare to find these types of pollen grains in the natural deposits of an area and thus they would rarely be an important type found in a region's pollen print. These attributes are both good and bad. They are good because if any of these types of pollen grains are found on objects at a crime scene or in another type of forensic sample, it generally means that the object or sample came in direct contact with the flowers or perhaps the leaves of the parent plant. This becomes an advantage because it often means that one can confidently conclude that an item or person was associated with a crime scene or some other specific locale where those parent plants and pollen types are found. It also provides a high degree of confidence that the pollen in the forensic sample belongs with the sample and that the pollen was not an atmospheric "contaminant". The downside of entomophilous pollen types is that so little pollen is produced by each plant that the chances of those pollen grains being transferred from the plant to some foreign object or person is often reduced.

Interpretation of Pollen Data

Understanding the rules that govern pollen production and dispersion are essential factors, which must be considered before evaluating forensic samples. Other important factors that will affect forensic pollen samples include how rapidly different types of pollen and spores settle out of the atmosphere (sinking speed), how well various types of pollen and spores remain preserved once they are deposited, and what types of clues indicate that pollen grains may have been recycled after they were deposited.

How rapidly airborne pollen and spores sink to the surface will determines how much and which species of pollen and spores will actually becomes part of the pollen rain and thus the pollen print of a given locale or region [5]. For example, marijuana, alder, juniper, and birch pollen are very small and very light pollen grains that have a sinking speed of about 1 or $2\,\mathrm{cm\,s^{-1}}$. This means that finding a few of these pollen grains in a forensic sample does not necessarily mean these plants are actually growing at the sampling spot. Instead, it might mean that the pollen resulted from long distance transport from sources many miles away. Before being able to determine this possibility, one would need to calculate how much pollen from these airborne types occurred in a forensic sample, how strong and from which direction the

prevailing winds are in the sampled region, and what the total pollen concentration value (amount of total pollen deposited and subsequently preserved in one unit of deposit such as pollen per cubic centimeter or pollen per gram) is for the pollen rain of the sampled region. On the other hand, if one were to find anemophilous pollen types such as maize (*Zea mays*), wheat (*Triticum*) spruce (*Picea*), Douglas fir (*Pseudotsuga*), or fir (*Abies*) pollen in a forensic sample, it generally means that either someone may have carried those pollen grains to the sample site or those source plants were growing very close to the sampled area. The reason one could make that assumption is because the sinking speeds for these pollen types are very fast causing them to fall to the surface at a rate of $6-12\,\mathrm{cm\,s^{-1}}$, which is 4–6 times faster than the lighter ones. Each pollen type has its own sinking speed, which will cause it to be dispersed either very close to the parent plant or scattered over a wide area. In addition to sinking speeds, pollen grains of different sizes and mass will be more or less subject to be scoured out of the air currents when they are hit by raindrops or when they hit objects of various sizes, such as twigs, leaves, or various manmade objects. All of these factors must be carefully calculated for each locale in order to determine the potential, or expected, composition of the pollen rain in a given region. Once those data have been determined, then the pollen print for that region can be used for comparison against forensic samples that are suspected to have come from the same region. Depending upon how well the comparison of pollen spectra match, the palynologist may conclude that there is, or is not, a valid relationship between both samples. One way this is done is by using the likelihood ratio, which is based on using Bayes' theorem and considers how well the pollen spectra match and what are the probabilities of finding similar pollen assemblages in other locations [6].

Pollen degradation can also become an important factor in some types of samples. Pollen samples that might be only days or weeks old usually do not suffer much pollen destruction; however, pollen samples associated with buried objects or buried bodies might become degraded or in some cases completely destroyed by a variety of conditions. However, the potential of pollen destruction caused by germination is not a factor. When examining forensic pollen samples associated with buried objects, the palynologist must consider how much, and which types, of pollen or spores may have been destroyed. Depending upon a variety of conditions, the total pollen and spore spectrum of a sample might be altered due to the differential pollen preservation and degradation. Those potential changes in the overall pollen spectrum of a sample must be considered when trying to interpret the overall pollen assemblage for a sample [7].

Some of the important factors, which will affect pollen and spores and can, cause the total loss of certain taxa include the degrading effects caused by microbial activity, high soil pH and Eh (oxidation/reduction potential as judged by whether on not it is an oxygen rich or oxygen poor environment) that can cause mechanical breakdown of soils, various types of soil movement such as solifluction (a type of mass wasting where waterlogged sediments slowly move downslope over impermeable material), frequent changes in the level of soil moisture, and the inherent strength and durability of the pollen grain's outer wall. A number of previous studies have demonstrated that some very durable types such the spores of ferns and lycopods and the pollen from grasses, pigweed, amaranths, composites, oaks, and pine often remain preserved even in fairly harsh environments where most other pollen and spore types are totally destroyed by oxidization or have become so broken and degraded that they are no longer recognizable. Being able to recognize the levels and types of pollen grain and spore damage and destruction in forensic samples are critical aspects that must be recognized and understood before trying to interpret the overall pollen data. Although a few deposited pollen grains might occasionally burst open or rarely a spore might germinate, these potential problems are not of major concern.

Pollen recycling is another aspect that one must recognize when examining pollen samples. Understanding the vegetational environments where recycling occurs and being able to recognize the severity of recycling become important interpretive clues when examining pollen samples. Sometimes ancient fossil pollen and spores will erode from deposits thousands or millions of years old and will be recycled into contemporary samples. Such recycling might prove extremely useful because different fossil pollen or spore types tend to be deposited, and later eroded out from different aged sediments. An example of this type of clue became the key evidence leading to the conviction of a murder suspect who could be placed at the crime scene because of

a unique, 20-million-year old Miocene-age hickory pollen grain that eroded out of a sediment outcrop, became recycled, and then got trapped as part of the dirt (*see* **Soil: Forensic Analysis**) in the tread of the suspect's boots [8]. In some cases, such as this example, the recycled fossil pollen provided the precise clue needed for identifying the exact location where the crime was committed and where the victim's body was buried. In other cases, recycled pollen from fairly recent deposits might combine with modern pollen in ways that might mislead some palynologists into making an incorrect interpretation.

Depending upon what types of sediments have yielded recycled pollen grains and depending on how badly the contemporary fossil pollen has been degraded, those recycled pollen types may, or may not be easily separated from the actual modern pollen rain at a given locale. If the recycled pollen grains cannot be distinguished from the modern pollen rain, the addition of recycled pollen may mask the true identity of an actual locale by producing a combined pollen spectrum quite different from the expected pollen print for that locale. Nevertheless, most skilled forensic palynologists with years of experience would not mistake most types of recycled pollen for contemporary pollen deposited by the local pollen rain.

There are several techniques palynologists have used to try to correctly identify fossilized recycled pollen in modern pollen samples. O'Rourke [9] tried to separate recently deposited pollen from recycled pollen by staining them with basic fuchsine. She found that the different layers in a pollen wall stain differently depending on their molecular structure. Modern pollen, she found, still contained the innermost cellulose layer in the pollen wall called the *intine*, which stained a light pink as opposed to the outer layers in the pollen wall that stained dark red. During her subsequent study of the pollen spectra from modern samples, she considered only the pollen still containing an intine as being part of the normal pollen rain; all other pollen grains that stained dark red were considered to be recycled pollen. That technique is useful in some situations but not others. For example; the destruction of a fresh pollen grain's cytoplasm and intine layer can occur very rapidly in warm and moist environments where oxidation rates are rapid. In cold and dry regions fresh pollen grains are slower to lose their cytoplasm and intine. Unless one knows for certain what environmental conditions exist and how rapidly fresh pollen lose these aspects, this technique should be considered fairly unreliable. Likewise, the degree to which a pollen grain's wall will absorb stain, such as basic fuchsine or Safranin-O, and thus will appear as being lighter or darker depends on a number of factors including the thickness or the pollen wall, the percentage of the pollen wall that is composed of cellulose and protein molecules, and the amount of sporopollenin (a very durable substance composed of long-chains of carbon-based molecules that are similar to structure to carotenoids) present. Sporopollenin is a group of highly resistant organic molecules that are unique to various types of palynomorphs, including pollen and spores. In pollen and spores this molecule does not absorb stain but forms a latticework in the pollen and spore walls that help gives them shape, durability, and structure. As pollen and spores slowly degrade, the cellulose and protein molecules in the walls break down into compounds that no longer absorb stains. Eventually, degraded pollen and spores have very little cellulose and protein left and yet may maintain their shape due to the durable sporopollenin latticework remaining in their walls. When pollen and spores reach this point of degradation most will become very lightly stained or almost completely transparent even when immersed in pure stain. Because different pollen and spore taxa have different wall thicknesses and different proportions of cellulose, protein, and sporopollenin molecules in their outer walls, staining alone is not a reliable indicator of recycling. Different degrees of staining often reflect different levels of decay and recycling, but those differences might also reflect the innate differences found in the outer walls of various taxa of pollen or spores (Figure 1).

Another technique sometimes used to try to identify recycled or reworked pollen from modern examples is fluorescence microscopy [10]. The theory behind using both staining and fluorescence is the belief that recycled pollen and spores have different depositional histories than recently released pollen assemblage and thus will make the recycled grains appear as being different. What makes fluorescence perhaps a more useful technique for determining recycled pollen and spores is that pollen, spores, and other types of organic materials in different stages of preservation, or in different stages of carbonization, will emit light of varying intensity and wavelengths [11]. These differences can be detected under

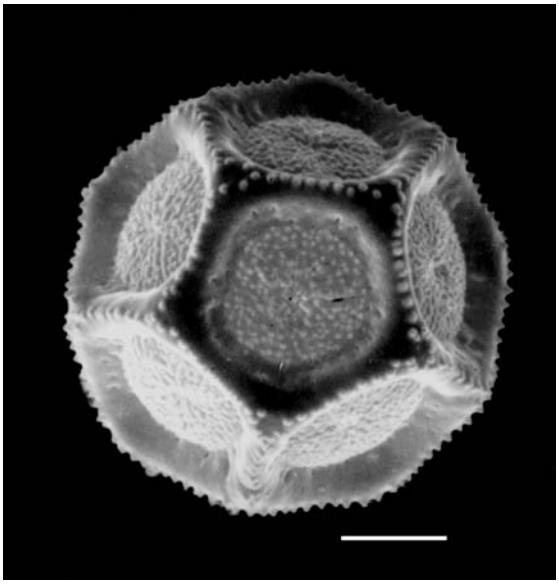


Figure 1 This is an SEM micrograph of *Alternanthera philoxeroides* (K. von Martius) A. Grisebach. This plant is in the Amaranthaceae plant family and the common name is alligator weed. The bar scale is 5 μm long [Courtesy of Gretchen D. Jones, Ph.D., & Ester F. Wilson, USDA-ARS, APMRU.]

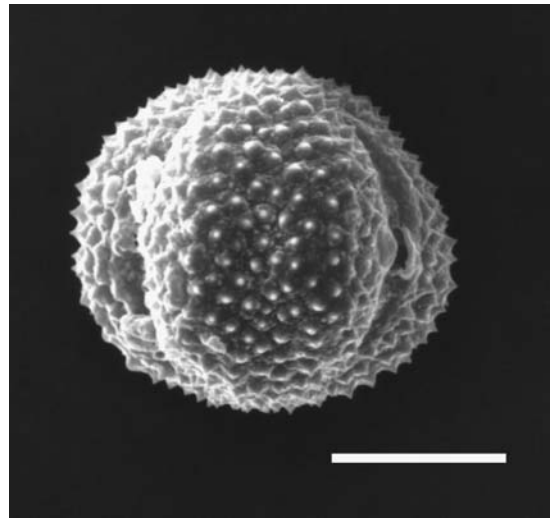


Figure 2 This is an SEM micrograph of *Artemisia californica* (Less). This plant is in the Arteraceae plant family and the common name is California sagebrush. The bar scale is 10 μm long [Courtesy of Gretchen D. Jones, Ph.D., & Ester F. Wilson, USDA-ARS, APMRU.]

ultraviolet light during fluorescence studies because those differences will give various pollen and spores different color hues.

Although differential staining and fluorescence microscopy has proven useful for some types of pollen studies, neither has helped us very much in forensic applications, thus there are reservations concerning their applicability in most forensic pollen work. In a laboratory experiment we conducted we "spiked" 18 000-year-old peat sediments collected from a subarctic environment with modern pollen from tropical plant taxa. Those combined samples were then processed using standard laboratory techniques for pollen recovery. After staining the recovered pollen and spores with Safranin-O, we found that only a few (less than 10%) of the added tropical pollen grains could be recognized strictly by differences in either staining or their fluorescence. In a similar study conducted in Arizona by Shellhorn *et al.* [12] they found that fluorescence microscopy did not help them separate modern surface pollen taxa from fossil pollen types known to be recycled from 20 000-year old deposits of the Wilcox Playa.

In forensic work we have found that a better guide to recycling seems to be the overall condition of the individual pollen grains and spores in a sample. Pollen samples recovered from crime scene locations in deserts, steppes, and semiarid regions where vegetation is usually minimal and where high winds are frequent, usually contain a wide variety of examples of degraded pollen and spores. The degrading process from those regions often produces grass pollen that is cracked, broken, crumpled, or shredded; pine and other types of bisaccate (conifer pollen types with a pair of air bladders to help keep them aloft after dispersion) pollen grains that usually have one or both bladders detached or broken and often are reduced to broken fragments of the main body of the pollen grain; and broken or fragmented parts of fragile pollen types such as the polyads (groups of individual pollen grains united into one large grain) of acacia (*Acacia*), and the inaperturate (pollen grains that lack any type of aperture) pollen grains produced by sedges (Cyperaceae), popular (*Populus*), cypress (*Cupressus*), and junipers (*Juniperus*) (Figures 1–10). The overall species composition of a pollen sample, the specific pollen types and the condition of the pollen in a sample can usually provide essential

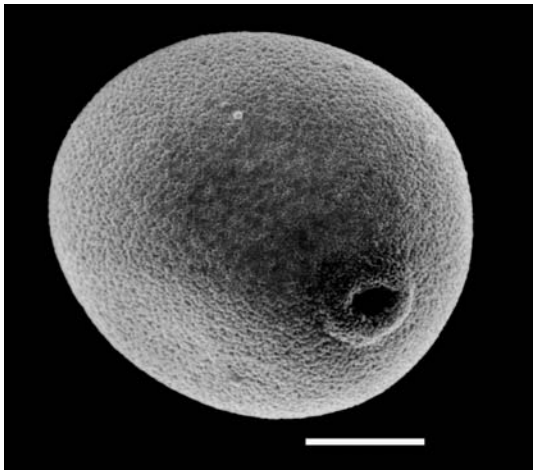


Figure 3 This is an SEM micrograph of *Arundinaria gigantea* (T. Walter) G. H. Muhlenberg. This plant is in the Poaceae plant family and the common name is giant southern cane. The bar scale is 10 μm long [Courtesy of Gretchen D. Jones, Ph.D., & Ester F. Wilson, USDA-ARS, APMRU.]

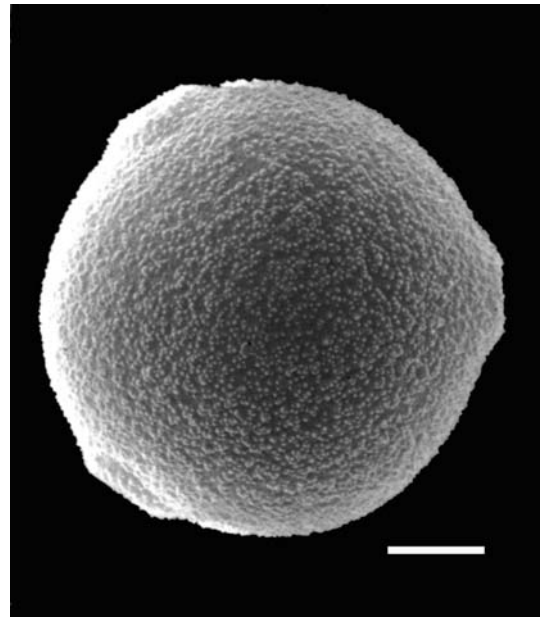


Figure 4 This is an SEM micrograph of *Cannabis sativa* (C. Linnaeus). This plant is in the Cannabaceae plant family and the common name is marijuana. The bar scale is 5 μm long [Courtesy of Gretchen D. Jones, Ph.D., & Ester F. Wilson, USDA-ARS, APMRU.]

clues about the ecology of a region where the sample originated as well as the degree and frequency of the recycling process. Once familiar with the types of potential degradation and overall appearance of pollen spectra from various types of environments, one can almost immediately determine, within broad parameters, whether the sample comes from an arid, semiarid, temperate, tropical, or arctic type of environment.

Collection and Extraction of Forensic Pollen Samples

Collection and extraction of forensic pollen and spore assemblages must be done with great care. If possible, the palynologist should be given access to a crime scene before other investigators arrive and begin collecting their samples. As a forensic specialist I realize that nearly "everyone" working a crime scene believes that they must be the first to visit the area. Nevertheless, the pollen and spore composition of the crime scene is extremely fragile and can easily be inadvertently altered, removed, or contaminated by the action of other forensic and crime scene investigators who also arrive to complete their own investigations. In addition, the use of improper collection techniques and/or accidental contamination of collected materials may render a sample useless for forensic pollen use and will provide a basis for the dismissal of forensic pollen evidence in court (*see* **Error Rates in Forensic Methods**). Whenever possible, forensic pollen samples should be collected by a trained forensic palynologist or someone immanently familiar with proper collection protocol. This type of training generally comes from working with established professional forensic palynologists or gaining the basics from attending short courses on proper collection procedures. Collection by these types of specialists will ensure that the samples are collected correctly, that they remain contamination-free through all stages of storage, and that the samples will be processed and analyzed correctly. When forensic palynologists are present, others members of the crime scene investigation team should follow prescribed protocol to ensure that all forensic pollen samples are collected properly and that the samples remain contamination-free. In all cases, as with other forensic samples, it is essential to keep detailed and accurate records and pictures

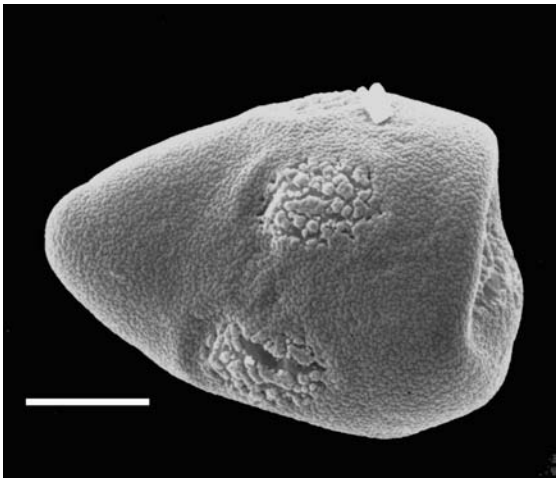


Figure 5 This is an SEM micrograph of *Carex microdonta* (J. Torrey & W. J. Hooker). This plant is in the Cyperaceae plant family and the common name is little-toothed Caric-sedge. The bar scale is 10 μm long [Courtesy of Gretchen D. Jones, Ph.D., & Ester F. Wilson, USDA-ARS, APMRU.]

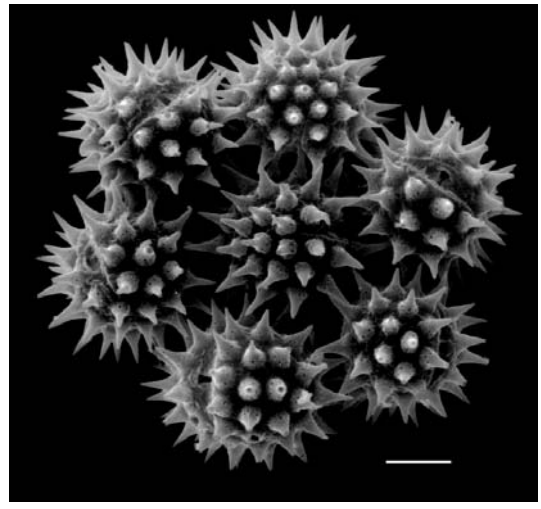


Figure 7 This is an SEM micrograph of a cluster of seven pollen grains stuck together of *Helianthus annuus* (C. Linnaeus). These are in the Asteraceae plant family and the common name is common sunflower. The bar scale is 10 μm long [Courtesy of Gretchen D. Jones, Ph.D., & Ester F. Wilson, USDA-ARS, APMRU.]

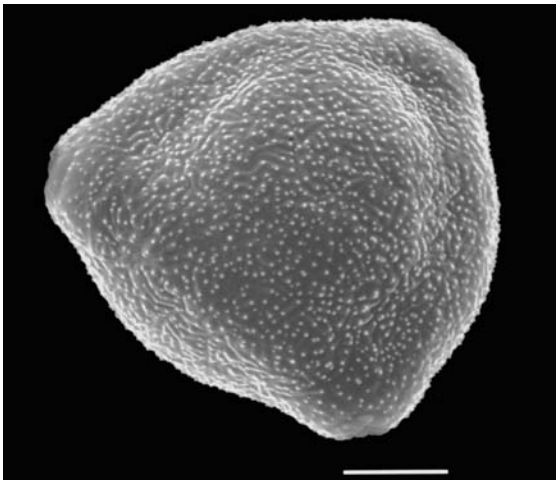


Figure 6 This is an SEM micrograph of *Corylus americana* (Walter). This plant is in the Corylaceae plant family and the common name is American hazelnut. The bar scale is 5 μm long [Courtesy of Gretchen D. Jones, Ph.D., & Ester F. Wilson, USDA-ARS, APMRU.]

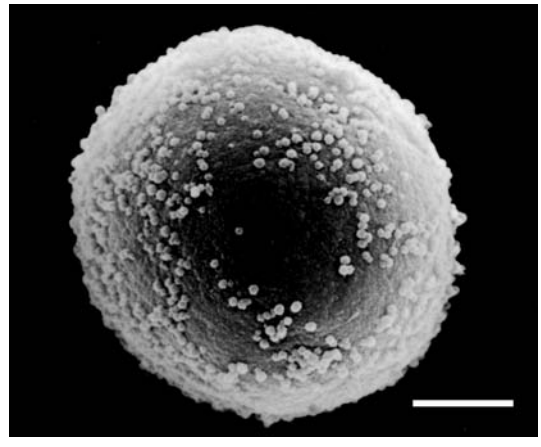


Figure 8 This is an SEM micrograph of *Juniperus virginiana* (C. Linnaeus). This plant is in the Cupressaceae plant family and the common name is Virginia red cedar. The bar scale is 5 μm long [Courtesy of Gretchen D. Jones, Ph.D., & Ester F. Wilson, USDA-ARS, APMRU.]

of how and where each sample was collected, what happened to each sample from the time of collection until it is analyzed, and the security of all pollen evidence until reports are written or statements are presented in court. If any hint of contamination, either natural or unintentional can be implied or proven, then doubt can be cast upon the resulting interpretations. Herein lays one of the greatest problems facing

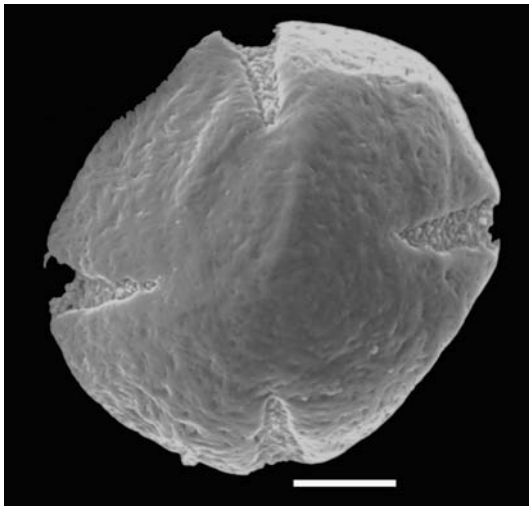


Figure 9 This is an SEM micrograph of *Melia azedarach* (C. Linnaeus). This plant is in the Meliaceae plant family and the common name is Chinaberry tree. The bar scale is 10 μm long [Courtesy of Gretchen D. Jones, Ph.D., & Ester F. Wilson, USDA-ARS, APMRU.]

forensic palynology. Because there are so few forensic palynologists, most crime scene pollen samples are collected by other people. Although most forensic personnel and other crime scene investigators can be trained to collect samples using proper protocol, and even though a crime scene photographer might take ample photographs, there are still other, subtle observations that only a trained forensic palynologist might notice as being important. Often the composition of the plant associations at and near the crime scene (*see* **Botany**), damage to individual plants, or evidence that someone brushed against a bush, or the presence of some exotic or unusual outcrop nearby each might become a vital clue when trying to understand and interpret the collected forensic pollen evidence from the crime scene or match pollen evidence collected from some suspect at a later time. These are reasons why, under ideal circumstances, the forensic palynologist should be present and should collect the samples.

An essential part of any forensic pollen investigation is the collection of control (also called *comparator*) samples. Control samples are samples of surface dirt, dust, fibers, or other materials at, near, or directly associated with a crime scene. The control samples are essential for any forensic pollen study because their pollen spectra provide a "baseline" of pollen information about the "expected pollen assemblage" from a specific object or the pollen print from a given

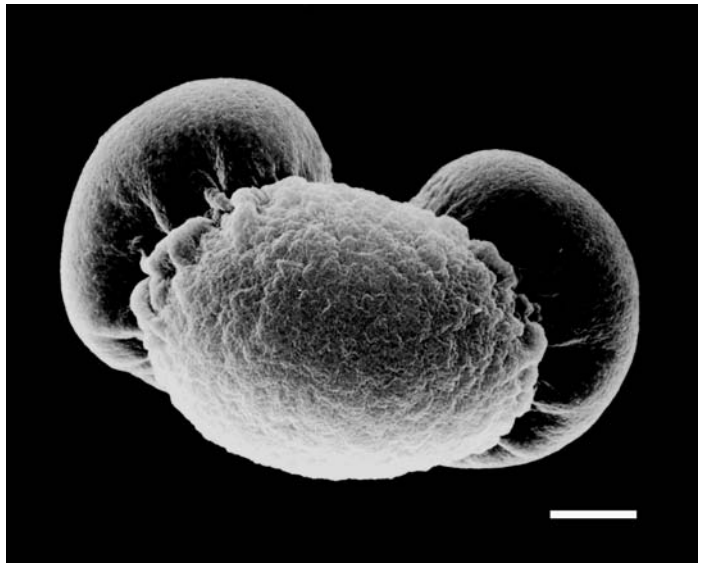


Figure 10 This is an SEM micrograph of *Pinus echinata* (P. Miller). This plant is in the Pinaceae plant family and the common name is short-leaf pine. The bar scale is 10 μm long [Courtesy of Gretchen D. Jones, Ph.D., & Ester F. Wilson, USDA-ARS, APMRU.]

crime scene or locale. Once the baseline of expected pollen data are determined from the various control samples, then the pollen recovered from a suspect or from forensic specimens thought to be associated with the object or crime scene can be compared against the control data to see if both match. One example involved a pollen control sample consisting of surface dirt collected from the ground where a sexual attack occurred. A forensic pollen study of the woman's soiled clothing matched the pollen types found in the control sample, confirming that she had been attacked at that location and had struggled with her assailant. Later, a suspect was identified and a search warrant revealed a soiled shirt and pants containing a pollen spectrum matching the control sample collected from the location where the sexual attack occurred. The pollen evidence alone did not confirm the suspect's guilt, but it did confirm he had been at that precise location, which was an area he said he had never visited.

Knowing how many control samples to collect from the scene of a crime is difficult. The more control samples one collects, and examines, the more pollen information one has about the locale where a crime was committed. Because the pollen spectrum in each control sample may vary slightly in reference to pollen types and percentages of each type, their combined spectra offer a potential range of pollen variation that can be expected for the actual crime scene. These types of data make matching the pollen results found on objects or suspects thought to be associated with a crime scene either convincing or obviously not similar. Not having enough or not having adequate control samples from a crime scene prevent a forensic palynologist from knowing what pollen types, and in what percentages, he or she should "expect" to find at the crime scene location. Without that knowledge, it becomes difficult to argue either for or against the confirmed association of the pollen assemblage found on a suspect's shoes, or car, or clothing, with the pollen types found at the actual crime scene.

Often a forensic palynologist has only a very small amount of material available for analysis. A few examples that I have examined in the past include the dust collected from the top of a table, a tissue wipe of the surface of a sandal, a piece of crumpled wire from some type of electrical device, a piece of electrical tape 3 cm long, one shoelace, a tiny piece of torn tissue no larger than a postage stamp, a sticky tape of dust collected from the dash of an vehicle, small tuffs of human hair from a body, the remains of a few bugs caught on the windscreen of a car, moths, honeybees, and a few pollen grains vacuumed from an item of clothing. Sometimes, when I have been fortunate, I could examine pollen trapped in a few specks of dirt trapped in the tread of a shoe or dirt caught in the carpet of a car. Once, the pollen recovered in the tiny bits of "fuzz" trapped in the bottom of the pockets of a pair of corduroy pants helped solve a murder case. There is almost no limit to what types of samples one can collect for potential use in forensic pollen studies. Nevertheless, often these types of samples create two major challenges for the forensic palynologist. First, there will rarely be enough material in a sample to try different extraction techniques. Second, there is rarely enough material to conduct a second test if something goes wrong during the initial pollen extraction and recovery process (e.g., centrifuge tube breaks, a beaker spills, a sample is dropped or accidentally mixed, a microscope slide breaks, etc.). For these reasons, the pollen extraction and recovery process must be conducted with the utmost care. In addition, in an effort to remove and concentrate the durable pollen and spores in forensic samples, it is generally necessary to destroy all of the nonpollen components. This means a sample's matrix will not be available later for other types of forensic testing. Studies of the DNA, trace elements, hairs and fibers present, or particles of dirt and sand in a forensic pollen sample must be conducted before pollen extraction occurs. Such studies, however, must be completed carefully and must not increase the risk of potential pollen contamination

Almost anything can be tested for forensic pollen. Dirt and dust trapped in almost any object often contain pollen and spores from the locale where the object originated. The following case illustrates this point. In New Zealand a man robbed a store and escaped on a motorcycle. Police gave chase and almost caught the thief, but at the last minute the thief abandoned his motorcycle and ran up a muddy hill and escaped into a wooded area. The next day a man reported that his motorcycle had been stolen the day before. When the claimant arrived to recover his motorcycle, the police realized he closely resembled the suspected thief. Armed with a search warrant, the police recovered a pair of muddy boots owned by the suspect. When asked about the mud, the suspect said that the mud came from the farm where he

worked and denied that he has ever been in the area where his motorcycle had been abandoned. A forensic palynologist collected a series of control soil samples from areas on the farm where the suspect worked and examined them to obtain a pollen print of the farm region. Next, the palynologist collected a series of control soil samples from the muddy hill where the motorcycle had been abandoned and examined them for their pollen contents. Finally, after examining the pollen trapped in the mud on the suspect's boots, the palynologist could confidently say that the pollen assemblage matched the pollen print from the muddy hill, and not those of the control samples from the farm. This type of forensic sample discrimination between two or more different locales is one of the common ways in which pollen evidence is being used to aid in criminal investigations (*see* **Soil: Forensic Analysis**).

Almost any item of clothing becomes an ideal trap for pollen. Clothing made from wool, linen, and cotton are among the best pollen traps, while garments made of leather, nylon, rayon, and other types of materials often make good potential pollen traps. In a murder case from Australia a man killed his girlfriend and then drove her car 50 km to a remote coastal park where he then hid her body under some wattle (*Acacia*) brushes that had been planted to stabilize coastal dunes. After returning home he parked her car at her home and then thoroughly washed his clothing. Later, when he became a suspect, police collected items of his clothing and gave them to a forensic palynologist to examine. In spite of the suspect's attempt to thoroughly wash and clean his clothing, a few pollen grains from two different species of wattle bushes remained trapped on his clothing. Although one species of wattle plants were quite common in that region of Australia, the second species was not native and was imported to help stabilize coastal dunes. The imported species occurred only in the coastal park region but grew close to other native species of wattles. The trapped pollen from both species of wattle bushes had already been found in the victim's car and now both were also found on the suspect's clothing. Eventually, it emerged that the suspect had driven his estranged wife's car to the coastal crime scene where he hid her body under wattle bushes and also brushed against other species of native wattle bushes growing in the car parking area near the coast. Although the pollen evidence alone did not provide proof that the suspect was the murderer, it did place him and his wife's car at the crime scene. In spite of his vigorous denial that he had not used his wife's car and had never visited to the coastal park, the pollen evidence placed him at the crime scene and helped to convict the suspect [13].

In a recent murder case, a young teenage girl, believed to be hitch hiking or a runaway, was killed in 1979 and left abandoned in a field with no identification. Her fingerprints were not on file and she was killed before DNA studies became routine in forensic work. Her identity remains a mystery even today and the murderer is still at large. However, nearly 30 years after her death, her clothing was thoroughly examined for traces of pollen and spores. Time had not degraded any of the pollen and spore evidence and wisely all of her clothing had been sealed in sterile containers and stored in a contamination-free environment for decades. On the basis of the forensic pollen studies in 2006, her identity could not be determined, but the pollen assemblage found in the pockets of her clothing and the inner lining of her jacket strongly suggested that she had very recently visited or had lived in a region more than 2000 miles west of the place where she was murdered. Among the pollen evidence recovered from her clothing were pollen grains from she oaks (*Casuarina*), which is a tree native to Australia and now grows only in restricted coastal areas of frost-free southern California and in similar regions in Southern Florida. These trees have never grown in New York because of winter freezes. In addition, it is highly improbable that even a single pollen grain from a she oak tree could have traveled 2000 miles from its source to the region of New York where the body was discovered. Other pollen from inside the murdered girl's pockets included species of spruce and birch that are common in the mountain flora of California, but do not grow in the region of New York where the body was discovered.

At the time of her death, authorities believed she was a local girl and thus searched only for missing person cases in the immediate vicinity of the crime scene. What made this case (one of the author's) interesting was how a rare combination of pollen can convincingly point to the victim's direct association with a distant region on the West Coast of the United States even though she had been murdered near the East Coast. This case also illustrates that if forensic samples are stored properly, decades later they can

still reveal their pollen evidence [14]. At the time of this writing authorities are now searching the missing person records for the West Coast of the United States for the years of the late 1970s and early 1980s.

Vehicle air filters can also become good sources of forensic pollen evidence. Pollen evidence from vehicle air filters have been used to reconstruct the regions where the vehicle had traveled. A case from the 1980s, that illustrates this point occurred on the island of Oahu in Hawaii. A stolen van was used in a bank robbery in Honolulu and then abandoned only a few miles away. A forensic pollen study of the air filter in the van contained not only pollen types typically found in the city of Honolulu, but also unusual tropical pollen and spore types found only in the nearby Koolau Mountains in the center of the island of Oahu. Using the forensic pollen and spore evidence as intelligence information rather than direct evidence, Honolulu police began interviewing residents in the Koolau Mountain region. A grocery store worker remembered the van visiting his store on several occasions and later seeing the van parked near a small cabin in the area. That information led to the capture of the bank robbers who were hiding in the cabin until they could safely leave the island.

Human and other types of animal hair are ideal pollen traps for pollen from given locales [15]. Therefore, when a person or animal travels to another region and their hair is examined fairly soon after arrival, the analysis of the pollen still trapped in the hair would provide a pollen print of the previous, not the current location. This type of use for pollen data was applied to a murder case in Texas where it was hoped that forensic pollen studies of hair samples would provide a key to where five murdered women had lived. Hair samples collected from the five bodies of the unidentified women found buried in a series of shallow graves were examined for pollen in hopes that it might provide a clue as to where the women had lived before being murdered. Soil samples collected from the graves and from the surface soil in nearby areas of the pine forest provided the control samples and the "expected" pollen prints from that region. The recovered pollen assemblages from the women' hair were then compared to the pollen prints of the crime scene to find major pollen differences. As in the case with the murdered teenage girl, it was hoped that the forensic pollen samples might provide suggestions regarding where each of the women may have lived before being killed. Authorities hoped that such information could help them search in the appropriate geographical regions for the identity of the missing women. Unfortunately, the recovered forensic pollen samples from the women's hair closely matched the pollen prints of the crime scene. This led to the suggestion that either the hair samples had become so contaminated by local pollen from the pine forest and from the soils in the grave site that it obscured any original potential pollen present, or perhaps all of the women came from areas of East Texas. Many areas in East Texas have pine forests and each area produces a fairly similar pollen print making it difficult to distinguish them apart. As of this time, those answers have not yet been resolved.

Like human hair, animal hairs are also excellent pollen traps. In New Zealand a rustler stole 300 sheep from a ranch and weeks later tried to sell them at a livestock auction. Although the sheep had no markings on them, the auctioneer became suspicious because he knew the seller was selling more sheep than the pastureland on his small ranch could support. The police were called in and they impounded the sheep. The original owner believed they "might" be his sheep, but had no proof so the police sheared a small patch of wool from the backs of several impounded sheep and sent them to a forensic palynologist for study. Control pollen samples were collected from sheep in both areas and also the surface soils in various locations on both the seller's and original owner's pasturelands. The pollen prints from each control sample was compared with the pollen assemblage recovered from the sheep's wool, which closely matched the control samples from the original owner's pasturelands. The thief complained that the pollen trapped in the sheep's wool had no relationship to pollen found in the surface soils of the two pastures. Countering this argument, the palynologist pointed out two important facts. First, the natural pollen rain in both regions would deposit airborne pollen on any surface in the pastures including the backs of sheep, the ground, or on the surface of several ponds and thus the pollen prints would indeed be similar. Second, there was a much different vegetational assemblage in and around each of the two pastures, which therefore produced very different types of pollen prints, as was noted in the pollen profiles of the various control samples. Thus, the pollen evidence proved critical in showing a clear discrimination between the expected pollen

rain and pollen prints of each of the two different pasturelands. As a result, the sheep were returned to the original owner and the rustler was convicted and sent to prison [16].

A shipment of 500 g of cocaine hydrochloride was seized in New York City. A portion was sent to a forensic palynologist for analysis. In similar situations, previous examinations of illegal marijuana (*Cannabis*) seized in New Zealand had revealed two important clues. First, the pollen print recovered from the seized sample contained not only marijuana pollen but it also contained many pollen types indicating the plants had been grown in Asia and not at any location in New Zealand. Second, pollen studies of marijuana samples seized in different drug raids carried out in widely dispersed regions of New Zealand revealed that all the samples contained very similar pollen spectra. This indicated that the illegal marijuana seized in all the raids came from the same, large imported shipment originating in Asia. It also suggested that the marijuana was being distributed through a single distribution network. A similar type of reconstruction was attempted for the seized cocaine from New York. After processing the sample, the recovered pollen assemblage provided three important clues about the cocaine. First, some of the trapped pollen came from tropical plants that typically grow in regions of Bolivia and Colombia where coca plants are grown commercially. Those pollen types linked the origins of the cocaine to those regions of tropical South America. A second group of pollen grains recovered from the cocaine came from subalpine conifers including Canada hemlock (*Tsuga canadensis*) and jack pines (*Pinus banksiana*). These two species of conifers do not normally grow in the same habitat, but both are found growing together in only limited regions of Eastern Canada and the Northeastern US. Therefore, the occurrence of pollen from both of these conifer types in the same sample suggests the cocaine was smuggled into North America someplace in Eastern Canada or the Northeastern United States. Once in North America, the cocaine was apparently opened, and exposed to airborne pollen types while it was being "cut" with powered sugar to increase its value. If that had occurred in any other region, then both conifer pollen types would not have been present in the same cocaine sample. Finally, the remaining pollen found in the cocaine came from weeds and plants (composites, grasses, birch, goosefoot and pigweed, etc.) commonly found growing in vacant lots in the urban slums of New York City. It is suspected that when the cocaine reached New York City it was again opened and cut further before being packaged for distribution on the street. It was during that time that the cocaine was again exposed to airborne pollen, which then became trapped in the sample [17].

A European manufacturing company shipped a number of crates filled with expensive machinery to a company in Asia. The ship carrying the cargo stopped at a number of ports between Europe and its final destination in Asia. When the crates were opened at their destination in Asia, the machinery was gone and had been replaced by bags of soil of approximately the same weight. Since the ship had stopped at a number of ports it was unknown where the machinery had been stolen and replaced with dirt. An analysis of the dirt revealed a wide variety of pollen and spores that closely matched the pollen composition found in soil samples collected from the port area in Capetown, South Africa. Although the pollen analysis did not solve the question of who stole the machinery, it did suggest that the switch had occurred in Capetown, which was one of the more than a dozen potential ports where the theft could have occurred. Thus, the pollen identified the location and therefore narrowed the search for the missing machinery to only one location between the ship's departure in Europe and final arrival in Asia [16].

Pollen evidence has also been useful in helping to resolve details about events in the past (*see* **Length Measurement**; **Archaeology**; **Mass Grave Investigation**. In a recent example, a mass grave containing the remains of 32 young males was discovered in 1994, in Magdeburg, Germany. An examination of the bodies revealed that they had been shot, but the unanswered question was who did it. Bits of clothing and other evidence found with the bodies did not provide conclusive proof about who killed them. Two possibilities emerged. Some believed they were partisans who were captured and killed by the Nazi Gestapo in the spring of 1945 before the area was overrun by advancing Soviet troops. Others believed the Soviet Secret Police captured and then killed a number of Russian soldiers who refused to kill local German citizens who rioted in late summer of 1953 against Soviet rule. The nasal cavities of seven skulls were rinsed and the material was examined for trapped pollen. The pollen analysis revealed high amounts of pollen from plants that

pollinate in the late summer, not plants that are spring pollinators. Comparisons of pollen types found in the dirt at the burial site confirmed that high levels of pollen from late summer blooming plants found in the nasal cavities must have been inhaled just prior to death and had not come from pollen in the dirt of the burial site. Thus, the pollen study confirmed that the victims probably died in late June or July, not during the spring and therefore were probably Russian soldiers killed by the Soviet Secret Police in the late summer of 1953 [18].

Summary

Although the use and application of forensic pollen studies are relatively new in many areas of the world, there is growing evidence that its use has a bright future. As more and more crime enforcement agencies become aware of the potential value of forensic pollen work, the need for trained personnel in the field will increase and will provide the emphasis needed to encourage the establishment of training centers for these specialists. Some countries and some law enforcement agencies are already aware of the value of pollen studies as a forensic technique and therefore conduct these studies routinely. In other countries pollen evidence is rarely collected at crime scenes and there appears to be little interest in using pollen for its potential forensic value. Perhaps one way to draw attention to this underutilized forensic tool is to briefly outline some of the important ways in which pollen evidence has already proven useful in helping to solve criminal and civil cases. Some of the crimes that pollen evidence has already helped to solve include instances involving homicide, terrorism, genocide, bombings, forgery, theft, rape, arson, counterfeiting, manufacturing and distribution of illegal drugs, assault, cases of hit and run, poaching, and identity theft.

Potentially, pollen evidence could be used to help resolve a wide range of circumstances associated with crime scene investigations. For example, pollen evidence could help relate a suspect to the scene of a crime, confirm that some item was associated with a crime scene or that some item left at a crime scene belonged to some suspect. Other types of information derived from pollen evidence might include proving or disproving a suspect's alibi, it might help narrow down a list of potential suspects to only one or two individuals, it might help track the travel route of some item or suspect, or provide the geographical source of some item. In other circumstances, pollen evidence has proven useful in helping authorities find human remains and clandestine graves, determined the season or sometimes even the general time of death, and confirmed the illegal poaching of wild animals or the adulteration of commercial foods.

References

[1] Mildenhall, D., Wiltshire, P.E. & Bryant, V.M. (2006). Forensic palynology: why do it and how it works, *Forensic Science International* **163**, 163–172.

[2] Milne, L.A., Bryant, V.M., Mildenhall, D.C. & Coyle, H.M. eds, (2005). Forensic palynology, in *Forensic Botany. Principles and Applications to Criminal Casework*, CRC Press LLC, Boca Raton, FL, pp. 217–252.

[3] Faegri, K., Kaland, P.E., Krzywinski, K., Faegri, K. & Iversen, J. eds (1989). *Textbook of Pollen Analysis*, 4th Edition, John Wiley and Sons, Chichester, p. 328.

[4] Woodhouse, R.P. (1935). *Pollen Grains*, McGraw Hill, New York, p. 574.

[5] Jackson, S.T. & Lyford, M.E. (1999). Pollen dispersal models in Quaternary plant ecology: assumptions, parameters, and prescriptions, *The Botanical Review* **65**(1), 39–75.

[6] Horrocks, M. & Walsh, K.A. (1998). Forensic palynology: assessing the value of the evidence, *Review of Palaeobotany and Palynology* **103**, 69–74.

[7] Wiltshire, P.E. (2006). Consideration of some taphonomic variables of relevance to forensic palynological investigations in the United Kingdom, *Forensic Science International* **163**, 173–182.

[8] Erdtman, G. (1969). *Handbook of Palynology*, Hafner Publishing Co., New York, p. 486.

[9] O'Rourke, M. (1990). Pollen reentrainment:contributions to the pollen rain in an arid environment, *Grana* **29**, 147–152.

[10] Traverse, A. (2007). *Paleopalynology*, 2nd Edition, Springer, Dordrecht, p. 814.

[11] Hunt, C.O., Rushworth, G. & Dykes, A.P. (2007). UV-fluorescence microscopy and the coherence of pollen assemblages in environmental archaeology and Quaternary geology, *Journal of Archaeological Science* **34**, 562–571.

[12] Shellhorn, S.J., Hull, H.M., & Martin, P.S. (1964). Detection of fresh and fossil pollen with fluorochromes, *Nature* **202**, 315–316.

[13] Milne, L.A. (2005). *A Grain of Truth: How Pollen Brought a Murderer to Justice*, Reed New Holland, Sydney, p. 175.

[14] Associated Press (2006). http://www.iht.com/articles/ap/2006/10/03/america/NA_GEN_US_Cold_Case_Pollen.php.

[15] Wiltshire, P.E. (2006a). Hair as a source of forensic evidence in murder investigations, *Forensic Science International* **163**, 241–248.
[16] Bryant, V.M. & Mildenhall, D.C. (1998). Forensic palynology: a new way to catch crooks, in *New Developments in Palynomorph Sampling, Extraction and Analysis*, *American Association of Stratigraphic Palynologists, Contributions Series*, V.M. Bryant & J.H. Wrenn, eds, American Association of Stratigraphic Palynologists, Foundation, Dallas, Vol. 33, pp. 145–155.
[17] Stanley, E.A. (1992). Application of palynology to establish the provenance and travel history of illicit drugs, *Microscope* **40**, 149–152.
[18] Szibor, R., Schubert, C., Schöning, R., Kruse, D. & Wendt, U. (1998). Pollen analysis reveals murder season, *Nature* **395**(6701), 449–450.

VAUGHN M. BRYANT

Paper Analysis

Introduction

Writing and printing papers are largely composed of bleached botanical fiber and mineral fillers to which can be added fluorescent whitening agents (FWAs) to enhance the whiteness of the paper or organic dyes to make papers of different colors. The purpose of the fiber is to form a bright contrasting flat surface to which darker inks, pigments, and toners can be applied. Fillers (usually in the form of finely ground calcium carbonate or various mineral clays) are also added to increase sheet opacity and improve the printing or writing surface. Modern printing papers (and much less commonly writing papers) may also have a coating, comprising calcium carbonate and/or clay, and/or titanium dioxide mixed with an organic polymer binder on one, or more frequently both sides of the sheet.

This article deals mainly with A4 and A3 copy papers, bond papers, and standard US Letter "A" paper, as these papers are now the most common hardcopy medium for commercial, legal, and personal communication. The principles, practices, and tools discussed below can be applied to all samples of paper, paperboard, ("cardboard" in common parlance) and fiber-based packaging materials.

Ever since monetary, commercial, and legal transactions have been committed to paper they have been subject to fabrication and fraud. Sometimes spurious documents remain in currency undetected but at other times, for various reasons, they fall under suspicion. A few examples among many are the fabrication of whole documents which are backdated, fabrication of company documents, minutes and receipts, fabrication of wills, fabrication of certificates and academic qualifications, and the manufacture of spurious historical documents.

Often fraudulent documents are exposed by discrepancies or inconsistencies in their written content (i.e., structure of signatures, dates appended, poorly composed letterheads, logos, etc.) and sometimes by the chemistry of the inks, pigments, and toners employed in their manufacture. For instance, the printing technologies with which the ink is applied, or the chemistry of the ink or toner itself can demonstrate that the document in question is either inconsistent with the purported date, or the period in which an undated document is represented to have been written or printed. These areas fall within the realm of the document examiner, ink and printing expert, and forensic chemist.

Another useful technique in exposing fraud is the field of "paper and fiber analysis" in which the paper on which the document is written can be characterized and differentiated from properly validated samples. Paper properties, both physical and botanical, although outwardly similar in appearance to the layman, can be, and often are, distinctly individual to a particular hemisphere, country, company, and even papermaking machine.

In the course of any investigation, the paper is first characterized for a range of physical properties which include basis weight or mass per unit area (commonly called *grammage* in countries using the metric system of measurement), caliper (or apparent thickness), sheet density, filler content, filler type, paper chemistry, direction of cut, etc. After these initial tests are completed, a small but representative portion of the paper can be mechanically broken down and dispersed into a dilute aqueous suspension of individual fibers. A few milliliters of this dilute suspension is placed on a standard microscope slide, dried, stained, and then examined under low magnification ($\times 40$ or $\times$ 100) to determine paper or fiber "furnish". Fiber furnish is basically the particular combination of pulp types, or the "pulp recipe",

a papermaker, or individual paper mill uses to create a particular grade of paper.

"Fiber furnish analysis" can be a powerful tool. In essence, it is the identification of the various pulp types found in a particular sheet of paper and the determination of the ratio in which they occur. The technique can be further refined by identifying the various plant species from which the fibers derive. This latter identification, in combination with the other factors, is very useful in determining the possible origin (or, conversely, rendering that it came from a particular country or source highly improbable) with regard to hemisphere and region, and sometimes even country, company, and mill, depending on the depth of industrial knowledge of the forensic analyst. As in most aspects of forensic analysis, great care should be exercised when drawing conclusions and some common pitfalls will be detailed later.

Even if the origin of a sample of paper cannot be established, it is often just as important to determine whether the paper used in a document is consistent with similar material emanating from the same government body, public company, or private premises at the time the document was alleged to have been produced. Also, if there is a belief that a particular page has subsequently been appended to the front or back, or inserted within the body of a document, the page in question can be analyzed and compared with those adjoining. Paper examination can also be useful in such circumstances when, for example, narrowing down possible origins of a ransom note. If the paper on which the note is written is of an unusual type, and it cannot be differentiated from the remains of an open ream or loose sheets found at the suspect's premises, these findings could add weight to circumstantial evidence against the suspect.

It is fortunate for forensic science that technology and manufacturing processes are in a constant state of flux as commercial companies are either trying to improve their product or make changes to it to maintain a competitive edge. These changes, and knowledge of when they were made, are an invaluable tool in determining the legitimacy or otherwise of certain articles. The same is true within the paper industry and knowing the manufacturing history within various geographic regions, countries, and companies within these countries, is of great advantage to the forensic examiner.

To an industry outsider, however, it is almost a matter of impossibility to keep track of all these changes so it is very useful for the analyst to maintain strong links with the key personnel within a country or locality's paper industry. It is also important for the paper analyst to regularly examine locally produced, "off the shelf" paper products in order to keep familiar with these changes. When purchasing photocopy or bond papers for examination, it is always useful to retain the ream wrapper (outer packaging) as it also contains much useful information together with a dozen or so sheets of the content as a growing archive. Apart from obvious general information such as country of origin, company, brand, grade, grammage, etc., wrappers often carry a discretely positioned inkjet printed company code that can inform the manufacturer when, and at what mill and on what machine, the paper was manufactured.

Most paper companies retain a history of when various manufacturing changes take place and, for their own protection, usually retain (often for many years) a swatch of "out-turn" paper sheets sampled from every commercial machine roll of paper produced. In the case of criminal investigations, most companies will willingly assist in providing samples for comparison with police evidence.

As a considerable quantity of paper is traded internationally and individual grades from a single paper machine can be packaged in a dozen, or more ream wraps for large volume purchasers (so-called home brand papers) the archive kept by each analyst can become very valuable in determining that a particular sheet of paper could have been produced, for example, in an Indonesian paper mill and purchased at a large retail office stationary chain in the United States in a given year. A year later the same US stationary chain might well be purchasing its "home brand" copy paper from a mill in China, or Thailand, or Finland, depending on the market price fluctuations.

Some Notes of Papermaking and Paper Testing

Paper is made by a formula of raw materials, set by individual paper mills, depending on the desired performance qualities required in the end product. The inclusion of certain raw materials is also governed by availability and price. As already stated, paper is largely composed of botanical fiber, be it softwood (gymnosperm) fiber, hardwood (angiosperm) fiber, Gramineae fiber (grass fiber such as that derived from cereal straw, sugar cane, or bamboo), or other

plant materials such as cotton, cotton linter fiber, or (rarely) fiber extracted from more exotic sources, such as the paper mulberry tree. The common factor between the vast majority of botanical fibers used in the papermaking process is that they are derived from the modified cell walls of elongated structural cells (i.e., fibers). Adventitiously, smaller quantities of other nonfibrous types of plant cells, such as vessel cells and ray cells will also be included with the fibers. Broadly speaking, all plant cells used in papermaking are largely composed of cellulose. In the living plant (and in some types of paper), these cellulose structures are impregnated with, and tightly bonded to adjoining cells with two organic bipolymers known as *hemicellulose* and *lignin*.

Fibers are extracted or freed (prepared) from the plant matrix by various chemical and mechanical means. These processes, particularly the chemical processes, are designed to liberate individual fibers from the plant material by dissolving or making soluble the lignin component that bonds the cellulose and hemicellulose materials together. Chemical pulps, as the name implies, are prepared by "cooking" the raw plant material in a chemical cocktail (usually weak acids or strong alkalis) at elevated temperature and pressure to facilitate the transformation of lignin into water-soluble products that can be separated from the fibers by washing. Often 50% or more of the original plant material is lost during this procedure but a very strong and flexible fiber, ideal for later use in writing and printing papers, is produced. In the case of most printing and writing papers (excluding newsprint and many cheaper magazine and catalog papers – see below) the freed fiber or pulp has to be further chemically treated by bleaching to remove the last traces of lignin giving the desired whiteness.

In contrast, heat in combination with various mechanical processes is also employed by the paper industry to separate the fibers from the wood matrix. These "thermomechanical" processes leave most of the lignin *in situ* around and within the fiber. The lignin can be made almost colorless in a second processing stage using oxidizing chemicals, such as hydrogen peroxide, however, this change is not permanent and the paper will start to show the yellow–brown color of unoxidized lignin after a period of days to months depending on the conditions of storage. Therefore, pulps produced via these latter thermomechanical methods are usually not used in high-quality writing or printing papers, but form the basis of many ephemeral products, such as newspaper and cheaper magazines and leaflet/catalog grade papers as they tend to form a weak sheet and "yellow" rapidly in sunlight. The main advantage of such papers is their lower cost of production stemming from the fact that almost twice the quantity of mechanical grade paper can be made from each tonne of wood, compared to chemical or "woodfree" papers. The most prevalent of these pulping processes, and their importance to the paper and fiber analyst, will be gone into more detail later. Combinations of smaller amounts of chemical processing with heat and mechanical energy are also used to produce the so-called chemimechanical and chemithermomechanical pulps and papers.

Physical Properties of Paper

As already mentioned, when characterizing a sheet of paper, it is not only the composition of the sheet that is important but also the paper's physical properties. For evidence to hold up in a legal enquiry, it must be demonstrated that, as far as physically possible, the paper from a document has been tested scientifically, and according to the accepted standard test methods. Like most other manufacturing industries, the paper industry globally has a recognized system of standard pulp and paper test methods that are overseen and continually reviewed by a number of industry-related organizations; a selection of these being the Scandinavian Pulp and Paper Association (SCAN), the US Technical Association of the Pulp and Paper Industry (TAPPI), and the Australian Pulp and Paper Industry Technical Association (APPITA). Many of these methods refer back to fundamental ISO standard methods.

Industry methods usually specify the minimum number of tests and the minimum area or number of sheets from which these tests must be taken (e.g., for grammage of photocopy paper, 10 individual measurements must be derived from 10 individual A4 sheets selected from specified positions throughout the ream). A common handicap for the paper and fiber analyst, however, is that there is usually only a very limited amount of sample to work with (often a single page) and the police or investigating body are often loathe to let it out of their hands, leave alone have part of it irretrievably destroyed for the purposes of testing. In these cases, the analyst can only stick to

the standard as close as possible, taking note of where deviations from the method occur, and the reasons why, for later inclusion in their report. Physical test results derived from limited sample should be treated as indicative only.

Further information in this article is presented in three sections; Some Basic Paper Properties, Fiber Analysis Using Light Microscopy & Fibre Sources for the Paper Industry.

Some Basic Paper Properties

For standard test methods for measuring the physical and chemical properties of paper and paperboard samples, readers are recommended to contact the national standard setting organizations in their respective countries. For convenience, some standards specified by TAPPI (North America) are provided in the Reference Section at the end of this article.

Moisture Absorbent Properties of Paper/Paper Conditioning Prior to Testing

One important property of paper is its moisture content. All commonly encountered paper sheets absorb between 5 and 12% of moisture from air depending on the ambient relative humidity and an equilibrium exists between water absorbed and the vapor in the atmosphere in contact with the paper, so water is an integral part of the sheet [1].

The amount of moisture a paper carries can significantly affect properties such as sheet weight, thickness, tensile and tearing strength, and optical properties such as brightness and opacity. Therefore, for results generated by one laboratory, to be comparable with those generated by another, paper and paperboard products are tested under strictly controlled atmospheric conditions. The International Organization for Standardization (ISO) standard environmental conditions for paper testing are as follows: relative humidity $50 \pm 2\%$ and temperature $23 \pm 1^{\circ}C$.

Most paper strength properties show hysteresis behavior with respect to moisture content, so it is important that papers are dried in an atmosphere at 25% relative humidity at $25^{\circ}C$ before being allowed to equilibrate under the atmospheric conditions specified by ISO prior to testing. "Conditioning" as this equilibrium is called, usually takes 24 h and sometimes longer if the paper or board sample is coated, exceptionally compact or thick.

Anisotropic Properties of Paper

Commercially made papers are anisotropic materials, meaning that almost all physical properties differ depending on the direction in which the measurement is made within the plane of the sheet. Papermakers often refer to paper as having a "grain" that will generally run along the longer axis of the rectangular sheet (long grain) or less commonly along the perpendicular shorter axis of the sheet (short grain).

The anisotropy of the sheet results from the process by which paper is manufactured. A dilute suspension of pulp fibers in water together with any mineral filler particles and chemical such as sizing agents, filler retention aids, and FWAs are dewatered on a revolving fine mesh forming fabric (commonly called a *wire*) moving at speeds that can range from $100\,m\,min^{-1}$ on machines making heavier basis weight paperboards to $2000\,m\,min^{-1}$ for lighter weight papers, such as newsprint. The continuous wire rotates around two drive metal cylinders that rotate about an axis that runs from the back of the machine to the front – the so-called cross direction (CD) of the machine. The axis that is perpendicular to the CD that runs longitudinally parallel to the length of the machine is called the *machine direction* (MD) – this is the direction in which the paper is formed. The fluid dynamics that applies to the fiber suspension as it impacts the forming fabric causes the majority of the fibers to align in the "MD". Some directional-dependant paper properties are tensile strength (greater in the MD), stretch (greater in the CD), tearing strength (greater is the CD), and paper stiffness (greater in the MD).

When reels of office papers are slit and cut into A4 or quarto sheets, the long direction of the sheet is usually aligned in the MD. Occasionally, however, certain photocopy and bond papers are found where the MD of the paper is orientated across the short dimension of the A4 sheet. "Direction of cut" is usually decided on the basis of which cut will minimize the "trim" (paper that must be deliberately cut from the front and back edges of the machine reel in order to guarantee an acceptably straight smooth edge on every sheet). Given that the cross directional width of the forming fabric (the "deckle width") is fixed at the time the machine is built, the minimum trim wasted will be achieved when either the MD or the CD dimension of the cut sheet is an integral factor

of the trim width of the machine. "Direction of cut" is sometimes useful in distinguishing a sheet made on one machine from a very similar sheet made on another machine (perhaps in a different mill) with a different deckle width.

For the sake of completeness, it should be noted that the remaining axis of the paper sheet that is orthogonal to both the MD and the CD is referred to as the *Z direction* (ZD).

Paper has many physical properties that are routinely measured by the manufacturer against either internal specifications or those specified by the mill's customers. The suite of tests is chosen depending on the ultimate end use. For the purpose of characterizing photocopy, bond and security papers, the forensic scientist would initially be interested in the following:

- sheet dimensions (metric, American or Imperial);
- grammage (basis weight);
- caliper (thickness);
- sheet density (calculated from sheet dimensions, grammage, and caliper);
- basic paper chemistry (is the sheet acid sized with a pH between 3.5 and 4.5, neutral sized with a pH between 6 and 6.5 or "alkaline" sized with a pH between 7.0 and 7.5 – "permanent paper")
- filler content and type;
- ISO brightness, opacity, and CIE color coordinates (L* a* and b*); and
- fluorescent properties under UV light (FWA content).

Photocopy and Bond Paper – Standard Paper Dimensions

The nominal dimensions of an A4 sheet are 297 mm $\times$ 210 mm $(L \times W)$ and an A3 sheet, 420 mm $\times$ 297 mm. These dimensions are governed by the ratio $1 : \sqrt{2}$ or $1 : 1.414$ such that sheets of diminishing sizes in the A series (and the B series) decrease by 50% in area in going from A1 to A2 and from A3 to A4, so there is zero wastage when the paper is cut to standard sizes.

Using this principle, two A5 sheets cut from a "long grain" A4 sheet will be 210 mm long $\times$148.5 mm wide and will be "cross grain". To elaborate on this discussion is beyond the scope of this article but an excellent explanation titled "International standard paper sizes" is given at the Internet site http://www.cl.cam.ac.uk/~mgk25/iso-paper.html. The purpose of discussing sheet dimensions here is to state that, for various reasons, sheets are not always cut exactly to these dimensions, and because a typical writing paper can expand and contract by up to 4% in the CD and 1.5% in the MD as the moisture content goes from 12 to 5%, the length and width of the conditioned sheet must be accurately remeasured after conditioning.

Grammage (Basis Weight)

In countries using the metric system of measurement, a basic property of paper is "grammage" or the weight of the sheet in grams per square meter (gsm). General use photocopy paper is usually manufactured to a grammage specification of 80 gsm but cheaper A4 "draft papers" are often made to lighter grammages (e.g., 65 and 70 gsm). Higher quality specialty printing and bond papers can have specific grammages in the 90–120 gsm range [2].

Continental North America runs to a different system. The dimensions of a standard US Letter "A" Paper (commonly used in North American homes and offices) is 11 in. $\times$ 8.5 in. or 280 mm $\times$ 216 mm. The basis weight of US Letter paper can range from 16 lb to 24 pounds (lb).[a] An explanation of the American paper basis weight system is also given at the above quoted website but a simpler explanation can be found under "Weights & Sizes of Paper" at http://www.inkjetart.com/weight.html.

Calculation of Paper Grammage for an A4 sheet – An Example. Area of an A4 sheet = 62 370 mm^2 (i.e., length by width in millimeters, assuming the dimensions of the conditioned sheet are 210 mm $\times$ 297 mm). This is equivalent to 0.06237 m^2.

Factor (No. of A4 sheets per square meter) = 1 m^2/0.06237 m^2 = 16.03

For this example, the conditioned weight of a single A4 page was found to be 4.97 g.

Therefore 4.97 g $\times$ 16.03 = 79.7 gsm.

Some notes on Grammage. Grammage with regard to copy and bond paper will not give the forensic investigator much information in itself as all paper manufacturers strive to produce paper within a certain grammage specification (e.g., copy paper with a nominal grammage of 80 gsm may have a product/manufacturing specification "Aim" grammage of 79.5 gsm with an "upper control limit"

of 80.7 gsm and a "lower control limit" of 78.5 gsm. Often manufacturers will endeavor to run the machine in the lower portion of the product range but keep within specification. A small saving in raw materials over each production run can result in large cost savings annually. Copy and bond paper is sold by area, not by weight – each ream containing normally 500 sheets (31.2 m^2 of paper), but sometimes 750 or 1000 sheets in "economy reams". On modern paper machines, grammage between sheets should not vary by >2%.

The importance of grammage is in the later calculation of sheet density. Every paper machine is unique. Despite, in this instance, copy paper being made to a nominal grammage of 80 gsm, depending on the pulp type, the amount of mechanical work done to the pulp (refining), filler content, calendering, and other machine parameters, thickness, and sheet density may vary significantly. If the authenticity of a page from a particular document is under suspicion, and its caliper and density differ significantly from measurements on other pages in the document, the discrepancy often warrants further investigation.

Sheet Thickness – Single or Bulking Thickness

Often a document examiner has only the document at hand to examine and, as already mentioned, the document may constitute a single sheet or even a fragment of a single sheet. Bulking thickness, determined on a stack of 10 sheets, is the preferred method when sufficient sample is available. Determination of thickness on a single sheet should be treated as a sighter measurement only. Determination of thickness, like grammage, should be done on fully conditioned sheets. Single sheet thickness or caliper is measured using a specialized "dead weight" paper micrometer fitted with a wide diameter foot (16 mm) to compensate for the paper's inherent surface variability. The deadweight pressure between the two measuring platens must be 100 kPa ± 10. The weighted foot should rest on the paper's surface for at least 2 s and no greater than 5 s before a reading is taken. For writing and printing papers, as with most other papers, thickness is recorded in micrometers. Readings should be evenly spaced over the CD of the sheet as this is the direction of greater variability. "Apparent thickness" is what is actually measured, not true average thickness (refer to *TAPPI test method*) [3].

Apparent Sheet Density

Sheet or paper density can be determined by calculation using the grammage and thickness data obtained. Paper density is usually quoted as kilogram per cubic meter. There are many different ways to calculate the sheet density. One format is given below.

Consider a stack of sheets cut into squares measuring 1 m × 1 m. If the sheets are neatly stacked one atop another until they reach a height of 1 m they would take up a volume of 1 m^3.

Calculation. Each sheet being a meter square would weigh 79.7 g or 0.0797 kg. (see previous calculation). For the purpose of this example, 20 individual thickness measurements for the sample gave an average caliper of 104 μm.

The apparent volume of each sheet is therefore $1 \times 1 \times 0.00104\ m^3$.

Therefore apparent density = mass per unit apparent volume = $0.0797\ kg/0.00104\ m^3 = 766.3\ kg\ m^{-3}$

Wire and Felt Marks

One final and potentially useful characteristic for confirming that two samples of paper came from the same ream when all the other analyses point to this being highly likely is the characterization of the subtle indentations left on the surface of many office, printing and writing papers by the forming fabric (or "machine wire"), and/or the drying felts on the paper machine. These markings can be made visible by very low angle illumination (typically 5–°10) and image capture using a low magnification (typically × 10) microscope, followed by contrast enhancement using image analysis software.

If a fast Fourier transform is performed on the enhanced image, and then all sections of the transformed pattern are deleted and the pattern "reverse transformed", a useful image of the particular weave pattern used in making the forming fabric is often obtained. If the weave pattern obtained from the evidentiary document (e.g., a ransom note) is identical to the weave pattern from samples of paper taken from the suspect's home, or office, and all other features of the sample are identical, this is fairly strong evidence that the paper sheets are from the same ream.

If, alternatively, the weave patterns are not identical, caution must be exercised. Nonidentity of the weave patterns does not necessarily mean that the

sheets are not from the same ream as five to six different machine rolls (made at different times on the same machine when the machine was "clothed" using forming fabrics of different weaves) may have been cut simultaneously to produce the ream of cut sheets. In a case where the first sheet from the ream gives a different pattern, it would be necessary to examine up to six sequential sheets from the ream in order to determine whether all the rolls that were sliced and cut to produce the ream have the same nonidentical weave pattern. In the event that all the six weave patterns from the sequential sheets taken from the ream are identical, but different to the evidentiary sample, then it can be concluded that the evidentiary sample is not from that source. The way in which common A4 (or US Letter) sheets are cut and compiled into reams involves cutting and sheeting between four and six machine rolls simultaneously in the same sequence.

To take a specific example, if mill personnel had taken six machine rolls from the reel store of which five rolls were made between October and December when the machine was clothed with a forming fabric with the weave pattern "A" and one roll was made in, say, January (following a holiday maintenance shut when the old forming fabric was replaced with a new forming fabric having a weave pattern "B"), then the sheets in the ream when subjected to image analysis above would show the sequence . . . A,A,B,A,A,A,A,A,B,A,A,A,A,A,B,A,A, . . . and so on, repeating this pattern with one sheet of pattern B separated by five sheets of pattern A down through the ream. If the identity of the mill that produced the paper is known and accessible, consultation with the mill staff will be a useful way of confirming how the reams from that mill are normally made up and whether or not there were changes made to the machine fabrics during the period in question.

An Introduction to Paper Chemistry

Major chemical properties of paper are also important in characterizing samples of paper. One of the quickest and simplest chemical tests is the pH of the sizing agent used when the paper was manufactured. A "pH pen" (a felt-tipped pen filled with an aqueous solution of pH indicator dye, phenol red) will produce a yellow spot on older "acid sized" papers and a violet spot on more modern "alkaline sized" sheets (i.e., 7.0–7.5 pH). Although convenient, the results of this simple test can be difficult to interpret if the paper has been sized with a modified "neutral" rosin size with a pH between 6.0 and 6.5 when a grayish color results. More certain data is obtained using TAPPI methods T509 and T435, or American Society for Testing and Materials ASTM D778.

Most commercial office, printing and writing grade papers contain mineral fillers in quantities up to 25% of the mass of the paper. Most alkaline sized papers (paper sized with the synthetic chemicals alkyl ketene dimer (AKD) or alkenyl succinyl anhydride (ASA)) will usually contain calcium carbonate as a filler, whereas the most common filler in acid sized (using emulsions made from natural softwood rosin soaps precipitated onto the fibers using aluminum sulfate) and neutral sized (using chemically modified rosin soaps and aluminum sulfate) papers is kaolin. Talc, calcium sulfate, barium sulfate, and titanium dioxide are also used as fillers but much less commonly. Paper made from de-inked recycled fiber will usually contain a mixture of many mineral pigments.

If the paper contains calcium carbonate, bubbles of carbon dioxide will form if a drop of dilute hydrochloric acid is placed on the surface. The amount of calcium carbonate present can be determined by ashing a portion of the sheet in a muffle furnace at 520°C and taking note of the weight of the ash obtained. The same sample is then further heated to 925°C to drive off carbon dioxide after which it is again weighed. The calculation of how much carbonate (as equivalent $CaCO_3$) was present in the original sample can be determined by simple stoichiometry. The mineral content of the residual ash can then be further analyzed using electron dispersive spectroscopy. Access to a scanning electron microscope can be invaluable in fully characterizing papers.

The presence of fluorescent whitening agents (FWAs), less correctly called *optical brightening agents* (OBAs) can be established by examining the paper sample under a UV light source in the 350–390 nm wavelength range. The paper will fluoresce if FWA's have been used during manufacture. More definitive results can be obtained by measuring ISO Brightness using one of the many light and colour measuring instruments now available. This is achieved by measuring the paper's brightness with the UV filter absent, and then present, in the incident light beam. By subtracting the former value from the latter, a measure of fluorescence can be obtained.

Most office papers also contain cooked starch to impart strength to the surface of the sheet. Spreading a few drops of dilute iodine solution in aqueous potassium iodide on the surface will confirm the presence of starch by the appearance of a characteristic but transient blue-black color.

Fiber Analysis Using Light Microscopy

Fiber Analysis

As previously mentioned, "fiber analysis" is the technique of breaking down a representative portion of a sheet of paper into its individual fibrous and nonfibrous cellular components, dispersing the resultant dilute suspension on a microscope slide and then drying and staining the fiber for microscopic analysis [4–9].

The cellular components are examined from two perspectives; firstly, to determine pulp types (and the proportion of each type in the total furnish) and secondly, the identification of the various wood or nonwood species represented in the furnish. Species from which the fibers derive can be identified by their morphology (i.e., shape and structure) and pulp types can be determined by the way individual fibers react to, or take up, chemical stain. The reaction to a particular stain is very much influenced by the chemical processes employed in producing or treating the raw plant material to produce the particular papermaking pulp.

It is impossible to cover the subject of fiber analysis in detail within this article but a full and broad understanding of the techniques can be gained from the above listed the standard methods sighted in the Reference Section. Distinguishing fibers from different trees, for example, is not a skill that can be picked up immediately and it often takes many years of gaining experience examining fibers from well-documented sources before a fiber analyst can feel fully competent in his or her abilities.

The fiber analyst's main tool is a binocular compound light microscope fitted with a mechanical stage and Abbe condenser capable of making observations at $\times 40$, $\times 100$, and $\times 400$. Higher magnification is of no advantage. One eyepiece should be fitted with a crosshair. A multistation tally counter, similar to that used to count blood cells, is also required.

Briefly, in the course of an analysis, the investigator would prepare two or more representative slides. After staining and the application of a cover slip, a few minutes should be allowed before observing to allow let the stain develop. It is recommended that the investigator first scan the slide at $\times 40$ to gain a general feel for the pulp types and species represented. Only then should they increase the magnification to $\times 100$ to make a formal analysis.

The analysis would start by the investigator moving to the top left or right hand corner of the cover slip, and moving vertically down the slide – by means of moving the mechanical stage – record each fiber passed over by the crosshair, taking note of its pulp type and whether it is a softwood, hardwood, or nonwood fiber. Each fiber and its pulp category must be immediately registered on the multistation tally counter and the traverse continued. When the investigator reaches the bottom of the cover slip the slide should be moved a few mm across and the traverse continued in the opposite direction until the cover slip has been fully examined. This process is continued until at least 800 fibers are counted. During or after the count, the individual species encountered should be noted for later inclusion in the investigator's report, if relevant. Sometimes, in species identification, $\times 400$ is used to clarify species by allowing closer examination of microstructure (e.g., hardwood vessel elements or softwood crossray pitting).

Fiber or Pulp Weighting

It will be obvious to the observer that not all fiber types are the same with regard to length and cross-sectional area and, if the proportion of each fiber type is to be reported as a weighted percent of the total, a weight factor must be applied. A list of generic weight factors for various pulp types and species is given in the right hand column of Table 1 – but it is recommended that each fiber analyst further refine these factors by calibrating their own observations against manufactured "dummy" furnish samples of known composition. It is also recommended that the analyst regularly cross-check their observations against those of other analysts. Between trained analysts, the weighted results obtained for each fiber type should not vary by more then 3%.

An example of a typical fiber analysis report compiled by the author is given in Figure 1. In this particular instant, the report pertains to a customer complaint where the printing company thought they were using paper from a local manufacturer but

Table 1 Stain reaction and weight factor for various fiber types

Pulp type	Norval–Wilson stain resultant color	Graff C stain resultant color	Fiber weighting
Unbleached chemical hardwood (kraft and soda pulp)	Dark or navy blue through to pink	Dark or navy blue through to pink	0.4
Bleached chemical hardwood (kraft and soda pulp)	Dark or navy blue through to pink	Dark or navy blue through to pink	0.4
Unbleached sulfite softwood pulp	Light brown	Usually yellow	1.0
Bleached sulfite softwood pulp	Lavender	Purplish blue	0.9
Unbleached kraft softwood pulp	Brown	Yellow to brownish yellow	1.0
Bleached kraft softwood pulp	Bluish gray	Blue to gray	0.9
Semichemical hardwood pulp	Khaki green to yellow brown	Brownish orange	0.6
Thermomechanical softwood pulp	Golden yellow	Bright yellow	1.7
Bleached chemi-thermomechanical softwood pulp	Copper brown	–	1.6
Cereal straw	Green to blue	Greenish yellow to greenish blue	0.4
Rag/linen/cotton	Pink	Red to pink	1.0
Jute/manila	–	Yellow orange	0.6
Esparto	–	Blue	0.5

were actually using an imported grade. Details of the various parties involved have been changed in order to maintain commercial confidentiality. An example of how fiber is weighted and calculated as a percent of the total is shown in the table included in the Figure 1 report. The reader will also note that the weight factors used by the analyst differ from the generic weight factors. These have been refined through trial and error to correlate with the author's own observations.

Fiber Sources for the Paper Industry

Common fiber sources used in the paper industry are categorized as wood fibers and nonwood fibers.

Wood Fibers – Two Types. *Gymnosperms (i.e., Wood Fibers from Coniferous (Cone Bearing) Trees).* Some examples of families of coniferous trees used commercially are pine, fir, spruce, cedar, larch, and hemlock. Conifers are often referred to as *softwoods*. Softwoods produce "long-fibered" pulp and are ideal where high tearing strength is required in the paper (e.g., paper sacks and bags). Addition of small percentages of long-fiber pulps also provide wet strength to office, printing, and writing papers (that are made mainly from short hardwood fibers (see below)) so that manufacture and converting operations can be carried out at higher line speeds. Depending on the species, average fiber length can range from 2 to 7 mm but most species fall within the 2–4 mm range with a wide distribution of lengths present within almost all the samples of fiber. Softwood fibers are called *tracheids*. Their function is to provide both mechanical support and the conduction of fluids.

Angiosperms (Flowering Plants). Angiosperms produce flowers and fruit and have a more complex and specialized vascular structure within their woody tissue. In the case of trees, angiosperms include both deciduous and evergreen examples. Angiosperm trees are generally referred to as *hardwoods* because of the normally higher density of their woody tissue compared to woods derived from most gymnosperm trees. Some examples of angiosperm tree families are elm, maple, alder, birch, poplar, eucalyptus, acacia, dipterocarps, and other tropical hardwoods.

Hardwoods produce "short-fiber" pulps. Average fiber length ranges from 0.7–1.5 mm depending on the species, with again a wide distribution of lengths being present in almost all samples observed. Hardwood fibers have a narrow diameter (typically 10–50 μm), compared with softwoods fibers and lack pits in their walls. Hardwood fibers (referred to as



CURRAWONG PAPER TEST LABORATORIES

Test Report

1st November 2007

To:	Mr. Kevin Kraft	Austral Paper Co. 6 Ream Road Papercut, Victoria, Australia, 3999
From:	John Murphy	Currawong Paper Test Laboratories
cc:	Mr. Peter Parchment	Austral Paper Co.

Subject: Fibre Analysis Report

Emporium Sales Catalogue

An Emporium Pre-Christmas Sale brochure (valid from Wednesday Dec. 8th until Sunday Dec. 19th, 2007) which manifested unacceptable levels of print mottle has been subjected to examination by optical microscopy in order to determine whether the paper on which it was printed is Australian product, or otherwise. The results of this investigation are given below.

Fiber Class	Chemical Hardwood*	Chemical Softwood#	Semi-Chem. Hardwood$	Mechanical Softwood**	Total
Fiber Count	12	328	444	61	845
Factor	×0.33	×1.0	×0.6	×1.8	
Weighted Factor	4	328	266	110	708
%	<1 (Trace)	46	38	16	~100

* Chemical Hardwood Component – Birch (Exclusively)
\# Chemical Softwood Component – Scot's pine (in the majority) & Norway Spruce (in the minority).
$ Semi-chem. Hardwood Component – Aspen (Exclusively)
** Mechanical Softwood Component (TMP) – Norway Spruce

Fibre Analysis according to AS/NZS 1301-451rp using Norval-Wilson stain.
Date analysed – 31.10.2007

Conclusion: The paper used to produce this brochure is not Australian made. The mix of species noted in the paper's furnish point strongly towards a European origin.

John Murphy
Research Scientist - Currawong Paper Test Laboratories

Figure 1 A typical fiber analysis report recommercial claim

libriform fibers) are ideal for producing printing and writing papers as they form a very flat, uniform, smooth surface.

Nonwoods (Angiosperms that Include Grasses and Shrubby Plants). The stems from wheat, rice, sugar cane, bamboo, papyrus, kenaf, sisal, jute, hemp, flax and the bolls, and the seed coat (linters) from cotton are all used to make paper in countries that lack adequate forest resources.

For the purposes of papermaking, nonwood fibers can be broadly broken down into "bast fibers" (from the outer tissues of the stems of sisal, jute, hemp, and flax plants) and Graminaceae (grass) fibers. Graminaceous fibers can be further broken down into cereal straw fibers (e.g., wheat, oats, and rice) and noncereal stems (e.g., reed, bamboo, and sugarcane (bagasse)).

Cotton, collected from the cotton boll and seedpod coverings of the cotton plant, falls into a category of its own. Cotton as a papermaking fiber is very much in decline, except for use in specialty writing papers, surgical tissues, and laboratory filter papers.

An excellent resource for the fiber analyst is a publication from the *Springer Series in Wood Science* titled *Fiber Atlas, Identification of Papermaking Fibers* by Marja-Sisko Ilvessalo-Pfäffli ([10]/ISBN 3-540-55392-4). This publication gives a thorough description of fiber morphology for species identification.

Common Pulp Types

The four most common pulp types employed in the paper industry today are chemical hardwood pulp (kraft or sulfite pulping process), chemical softwood pulp (kraft or sulfite pulping process), semichemical hardwood pulp (pulping processes using mechanical and heat energy with small quantities of chemicals), and thermomechanical softwood pulp (pulping processes using a combination of heat and mechanical energy and no chemicals (other than water)).

Other pulp types less commonly used in developed countries in recent years are "soda pulp" (chemical) and "groundwood" (mechanical).

1. Kraft process (also known as the *sulfate process*). Woodchips are cooked using NaOH and Na_2S in water at 170°C and 8 atm. This process takes place in alkaline conditions and produces very strong pulps used for both packaging and writing papers (kraft is German for strength). Extensive chemical processing is followed by minimal mechanical processing. If the pulp is going to be used to produce photocopy/printing/writing papers pulping is followed by a multistage bleaching process that can be based on chlorine, chlorine dioxide (elemental chlorine free (ECF)), or ozone and hydrogen peroxide (totally chlorine free (TCF)) that dissolve the last traces of brown-colored lignin.
2. In sulfite cooking the "cooking liquor" a mixture of water, calcium, or magnesium bisulfite (referred to as *combined* SO_2) and free SO_2 in excess. The excess SO_2 reacts with water to form sulfurous acid and therefore the cooking is acidic in nature. This process is also conducted at 170°C and pressure in excess of 8 atm and produces a very light-colored pulp with high opacity that is relatively low in strength compared to kraft pulp. Extensive chemical processing is followed by minimal mechanical processing. If the pulp is going to be used to produce photocopy, printing, or writing papers, it is followed by a similar multistage bleaching process to that used to bleach kraft pulps.
3. The semichemical method involves treatment of woodchips with lower quantities of either sodium sulfite (at pH 7–9) or sodium hydroxide at pH 12 followed by considerable mechanical processing. Cheaper to produce than full chemical pulps, bleached chemi-thermomechanical hardwood pulps (BCTMP) are sometimes added to lower quality writing papers to increase opacity and bulk.
4. To make thermomechanical pulps (TMP) softwood chips are pretreated with steam and then broken down into fiber by extensive mechanical processing. Comparatively the pulp is low cost to produce and is used in ephemeral products such as newspapers, magazines, junk mail, low cost paper back books, and writing pads. TMP has comparatively low-strength properties and readily yellows when exposed to sunlight and other light sources containing UV wavelengths.
5. Soda pulp is produced by cooking wood chips using sodium hydroxide in water as the sole chemical agent. As with sulfate and sulfite pulps, lignin removal is achieved at elevated temperatures and pressures.

6. Groundwood pulping is achieved by pressing whole logs against a heavily burred rotating grindstone.

Fiber Staining

The most common microscopy stain used for plant fiber identification is Norval–Wilson stain. Another useful stain is Graff C. Norval–Wilson and Graff C stain termed as *differential* stains. They are capable of staining different classes of fiber with different diagnostic colors depending on the chemical or mechanical processes utilized to produce them as shown in Table 1.

Softwood Identification

As stated earlier, softwood fibers are called *tracheids* and provide both mechanical support for a plant and also act as a conduit for the conduction of fluids to all the living parts of the plant. Fluids are transferred from fiber to fiber via "pits" in the tracheid wall. These pits act as valves to prevent the flow of liquids reversing. Basically there are two types of pitting. "Bordered pits", which convey fluid from tracheid to tracheid in an axial direction, and "crossray" pits, which convey fluids throughout the plant in a transverse direction. The structure of crossray pitting is particularly useful in identifying the species. "Spiral thickenings" can also help to identify species, as in the case of the commonly utilized softwood fiber from the Douglas fir (*Pseudotsuga menziesii*). See *Fiber Atlas, Identification of Papermaking Fibers,* referred to in section 2.3, or a similar publication.

Hardwood Identification

Hardwoods, and in fact all angiosperms, use fibers for structural support and use an auxiliary specialized vascular system for the conduction of fluids. These conduits, called *vessels*, are made up of smaller units or "vessel elements" which act in a similar fashion to man-made water conduit pipes laid end to end. These vessel elements are separated during the pulping process but remain mixed with the fibers in the pulp after the pulping and bleaching process. Fortunately for the fiber microscopist, vessel elements have very distinctive morphology depending on genus, and can be readily utilized to identify the various species of angiosperm present in a paper's furnish. See *Fiber Atlas, Identification of Papermaking Fibers* referred to in **Fiber Sources for the Paper Industry** section, or similar publication.

Comments on Indigenous and Exotic Species

Tree and other plant species can help identify the continent and sometimes country of origin of a paper sample. This, however, is becoming increasingly more difficult as plantations of various exotic species are becoming evermore prominent in paper-producing countries far from the regions in which they occur as natives. For example Spain, Portugal, South Africa, South America (and now, Southeast Asia) grow substantial plantation stands of eucalyptus for papermaking. This genus was once only indigenous to Australia and parts of New Guinea. One species, *Eucalyptus deglupta* is also native to the Celebes and the Philippines. It is worth noting that some species that were formerly classified within the genus *Eucalyptus* have recently been separated into a new genus, *Corymbia* by taxonomists, while retaining their specific names. *Corymbia maculate maculate* (spotted gum) and *Corymbia maculate citriodera* (lemon-scented gum) are the only two Corymbia species that may rarely find their way into commercially produced papers and neither species is cultivated in plantations as a papermaking resource.

The Monterey pine (*Pinus radiata*), originally native to a very confined region of the west coast of the United States, is now grown widely in Australia, New Zealand, South Africa, South America, and countries surrounding the Mediterranean. Outside its country of origin, it is a very fast growing tree and an ideal wood pulp resource. The same is true for Douglas fir, which can be found in plantation in Britain, Portugal, central Europe, and New Zealand.

Further Notes on Fiber Furnish

Identification of paper can sometimes be further refined by measuring the proportion, or percent, of different fiber types in a paper furnish (provided the paper manufacturing specifications are known). An example is given below.

Although bleached hardwood is the ideal raw material for the manufacture of photocopy paper, paper manufacturers in the southern hemisphere and tropics add a small portion of bleached softwood

to improve wet strength and speed of manufacture, and to also add dry strength for converting and cutting process (i.e., roll to ream). Until recent decades, there was insufficient capacity for producing bleached softwood fiber in the southern hemisphere (although Brazil, Chile, and New Zealand now have substantial output) and in countries such as Australia, for instance, that still import all their bleached softwood requirements, this material is an expensive impost. Most bleached softwood pulp arriving in Australia comes directly from New Zealand, Canada, and the United States but ever-increasing amounts are sourced from other regions such as Brazil and Chile depending on quality, availability, and price. As the imported softwood pulp is comparatively expensive, machine operators try to keep this component of the furnish down to 10 or 15% of the total fiber furnish (for an 80 gsm sheet). Indonesia, on more modern paper machines, can run with softwood concentrations as low as 1–3%. Information such as this can be very useful to the fiber analyst in identifying likely country of origin.

Recycled Fiber

A point of caution when examining paper furnish is to be certain that the paper does not contain recycled de-inked bleached fiber as this will greatly obscure the country of origin and could cause the analyst to reach quite misleading conclusions. To determine whether a paper contains recycled fiber or only virgin stock, it is good practice to first examine the intact sheet under $\times 10$ magnification using a stereo (binocular) microscope to pick up any signs of debris on the surface. The debris could be in the form of ink particles, toner particles, unbleached fiber or minute globules of adhesives (called *stickies*). Valid conclusions from such an examination will be made increasingly difficult if the sample has become soiled or dirty during the suspected or alleged criminal activity.

Acknowledgment

The author would like to gratefully acknowledge the assistance, guidance, and technical input of Dr Warwick D. Raverty, CSIRO Forest Biosciences, Bayview Avenue Clayton Victoria 3168, Australia.

End Notes

[a.] Approximate conversion from lbs (pounds per 100 square feet of paper) to gsm:

- 16 lb ~ 60.2 gsm
- 18 lb ~ 67.9 gsm
- 20 lb ~ 75.2 gsm
- 24 lb ~ 90.3 gsm.

References

[1] TAPPI Method T402 2003. *Standard Conditioning & Testing Atmospheres for Paper, Board & Pulp Handsheets and Related Product*.
[2] TAPPI Method T410 2002. *Grammage of Paper & Paperboard – Weight per Unit Area*.
[3] TAPPI Method T411 2005. *Thickness (Caliper) of Paper, Paperboard & Combined Board*.
[4] ISO Method 9184/1-3 1990. *Paper, Board & Pulps – Fiber Furnish Analysis*.
[5] ISO Method 9184/2 1990. *Paper, Board & Pulps – Fiber Furnish Analysis Part 2: Staining Guide*.
[6] TAPPI Method T401 2003. *Fiber Analysis of Paper & Paperboard*.
[7] TAPPI Method T263 2002. *Identification of Wood & Fibers from Conifers*.
[8] TAPPI Method T259 2005. *Species Identification of Non-wood Plant Fibers*.
[9] ASTM D1030 1999. *Standard Test Method for Fiber Analysis of Paper & Paperboard*.
[10] Ilvessalo-Pfäffli, M.-S. (1995). *Fiber Atlas, Identification of Papermaking Fibers Springer Series in Wood Science*, ISBN 3-540-55392-4.

Further Reading

McDonald, R.G. & Franklin, J.N. (eds) (1969). *The Pulping of Wood*, McGraw Hill Book Company, Vol. 1, Library of Congress Catalog Card Number 68-20994.
Browning, B.L. (1977). *Analysis of Paper'*, 2nd Edition, Marcel Dekker, New York, 1977 ISBN 0-8247-6408-0.

JOHN P. MURPHY

Paraphilia ***see*** Sex Offenders: Treatment of

Parental Alienation

History of Parental Alienation

In the context of divorce, the pathological alignment of a parent and a child resulting in the child's rejection of the alienated parent was originally described by Wallerstein and Kelly (1977, 1980). Gardner (1987) [3] later introduced the term, *parental alienation syndrome*, to describe a diagnosable disorder occurring in the context of separation and divorce. Although Gardner contributed a good deal by describing the features of parental alienation, his work has been criticized because of his use of the word, "syndrome". Almost all mental health professionals who work with children of divorce agree that the phenomenon occurs, that is, children of high-conflict divorces sometimes gravitate to one side of the conflict and view that parent as totally good and the other parent as totally bad, all without a good cause. However, many professionals who work in this area do not refer to this phenomenon as a syndrome. Thus far, parental alienation syndrome or disorder has not been adopted into the *Diagnostic and Statistical Manual of Mental Disorders.* Most professionals in the family law setting – including judges, lawyers, and custody evaluators – simply refer to *parental alienation* and do not use the term, *parental alienation syndrome*.

Role of the Evaluator

Often the family law court is faced with a situation where a child is rejecting a parent. Sometimes the child is refusing to spend time with the parent, and sometimes the child is alleging that the parent is emotionally, physically, or sexually abusive. The rejection could be mild ("I just don't want to spend a lot of time with him.") or sizable ("I never want to see him again.")

There are three main reasons for a child's rejection of a parent: (i) The accused parent is indeed abusive and the child's rejection is appropriate and understandable. This is *not* considered parental alienation. (ii) The aligned parent (or another person) is responsible for the child's rejection of the other parent. This may result in parental alienation. (iii) The child is rejecting the alienated parent for the child's own inappropriate reason, despite the efforts of the nonalienated parent to support the relationship between the child and the alienated parent. This may also be called *parental alienation*.

Faced with the situation of a child who is rejecting a parent, the reason for the rejection may be obvious to the court. For example, the hated parent may have a long history of abusing the child, documented by child protective service investigations and/or admissions of abuse by the parent. At other times, the court may be convinced that the aligned parent is responsible for the parental alienation, for example, if there is clear proof to that effect. In these cases – when the cause of the child's rejection is obvious to the court – the court may not order an expert to evaluate the family for parental alienation. The court will base its conclusions and orders on these certain, proved reasons for the rejection.

Much of the time, however, the court is not certain. The aligned parent's argument that the alienated parent is abusive is persuasive, as is the alienated parent's assertion that the aligned parent is behind the alienation. The court may decide that an evaluator should interview the child, in the hope that the result of the interview will provide enough information to resolve the issue. The child interviewer often will inform the court about the content of the interview and provide an opinion regarding the believability of the child's statements. Sometimes, the court believes that the issues are too complex for a child interview to elucidate the truth and orders a full custody evaluation.

In some jurisdictions, there are other possibilities for the court to secure the truth. In Los Angeles, for example, the court can order a limited or "solution-focused evaluation". This involves a court evaluator's speaking to both parents and the child individually, observing the parent–child relationships, interviewing collaterals, and reviewing records, all in the short time frame of one morning. In the afternoon, the evaluator testifies to his or her opinion.

Occasionally, the alienated parent and attorney believe that they can persuade the judge by retaining an expert to opine after performing a limited investigation. For example, the expert would review records, interview collaterals, or interview the alienated parent and then, based on this limited information, render an opinion. Sometimes, the expert's opinions are helpful for clarifying the issue, despite the opposing side's argument that the retained expert's opinion is biased.

Experts performing these tasks need to limit their opinions as they have not performed an entire custody evaluation (*see* **Visitation Rights**).

Aligned Parents' Contributions to Parental Alienation

An aligned parent's behaviors may have contributed to the child's stance of refusing visitation. The aligned parent may have repeatedly expressed extremely negative views to the child such as, "She never wanted you", "He is a horrible mean man", or "He is an adulterer and caused the divorce"! Alternatively, the aligned parent's criticism may not be so overt. The criticism may merely contain innuendos that the alienated parent is dangerous: "Call me immediately if he tries to hurt you."

The aligned parent may contribute in other ways to the child's refusal of the alienated parent. For example, the aligned parent may tell the child that the other parent did not call, when the parent had. The aligned parent might discard letters from the other parent to the child. The aligned parent may not inform the alienated parent about school events important to the child, such as back-to-school nights and school performances. The rejected parent's parenting and personality flaws may be exaggerated and continuously discussed with the child. The aligned parent may praise the child ("That's right. Stand up for yourself.") when the child criticizes the other parent, and scorn the child when the child compliments the rejected parent ("You don't know what you're talking about. He's not smart at all.").

Sometimes the parent may cause the child's alienation from the other parent, but has not done so on purpose. For instance, the mother may be an anxious person who is unaware that she is communicating her anxiety to the child. She may start crying or be noticeably worried when the child leaves for visitation, but insist that the child should go and "have a good time". Similarly, a child may be so concerned about the depressed parent that the child feels that he or she needs to provide emotional support to that parent instead of going on the visitation with the other parent. The continuous drumbeat of negative statements, whether direct or indirect, and actions by whatever method may cause the child to adopt the aligned parent's negative feelings and reject the other parent.

Child's Contribution to Parental Alienation

Despite a parent's best efforts to support the other parent's relationship with the child, sometimes the child alienates an otherwise appropriate parent for no good reason.

Oppositional Child

The child may be oppositional and defiant and reject the parent who does not succumb to the child's demands. He or she may choose the parent who is least restrictive and vehemently reject the rule-setting parent.

Worried Child

The child may be worried after the departure and loss of the nonresidential parent, so the child becomes fearful that he or she is also going to lose the remaining residential parent as well. As a result, the child experiences an attachment to the residential parent that is greatly exaggerated, and fears separation from that parent.

Stubborn Child

Although the child has a good attachment to both parents, the child may be very upset that they have separated and divorced. That is, the child is upset (for example, sad, angry, resentful, worried) about the situation and doesn't want to participate in the process. The child expresses his or her feelings by objecting vehemently and stubbornly to the visitation, even though ordinarily he or she enjoys being with the other parent.

Child Escaping Conflict

Finally, there is a common psychological mechanism – cognitive dissonance – through which the child's affections can become extremely polarized. Specifically, the child's intense like of one parent and dislike of the other becomes his or her way of resolving the psychological tension that he or she experiences. For example, if the mother and father have been actively and visibly fighting with each other, the child would experience cognitive dissonance when trying to have affection for both of them

at the same time. The child is unable to reconcile two dissonant thoughts, "My mother is right" and "My father is right". The dissonance creates a tension in the child's mind, which is resolved by believing that he or she loves one parent and hates the other.

Other Alienators' Contributions to Parental Alienation

At times, the alienating party is a person other than one of the parents. That is, despite one parent's actively supporting the relationship between the other parent and the child, another person is actively indoctrinating the child against the other parent. The alienator could be a grandparent, stepparent, girlfriend, boyfriend, or even a sibling. The reasons for the alienation are diverse: a grandparent may be seeking revenge against the parent who left his or her child; a stepparent may be worried that the child will see the parent as a "real parent" and the stepparent as some unimportant figure; or perhaps a sibling may be trying to punish the parent who "caused" the divorce.

Court Orders for Treatment of Parental Alienation

Family court orders for a fractured parent/child relationship should match the reason for the relationship problems.

If the parent is truly abusive and the child's rejection is justified, the family court may order the parent to participate in parenting classes or enact other orders aimed at rehabilitating the abusive parent. The court will likely order any visitation to be monitored until the parent shows that he or she will behave appropriately. Therapy for the child should include the following components: supporting the child for having been abused ("What a difficult thing that you went through. You did not deserve to be abused."); teaching the child how to stay safe in the presence of the parent ("If he ever hits you again, call me immediately."); and education ("Not all authority figures are abusive."). Conjoint therapy involving the child and the abusive parent may help them reestablishing a healthy and mutually satisfying relationship.

If an alienating parent is found to be a source for the child's rejection of an appropriate parent, the court's approach varies. If the alienation is severe, the court may order that the child's time with the alienating parent be limited and monitored. If the child is refusing to live with the alienated parent and the court fears that the child may run away or hurt the alienated parent, the court may order the child to live with a third party until therapy corrects the situation. At times, in extreme circumstances (suicidality, running away, poor school performance, or delinquent behavior), the court may order that the child be placed in a residential facility.

In cases where the alienation is not very severe, the court may order therapy without removing the child from the alienating parent's custody. The ordered therapy may include "reunification therapy". The role of the therapist is to aid in building a better relationship with the alienated parent ("Tell me about a good time you had with your father."). Occasionally, the reunification therapist may choose to bring the alienating parent into therapy. The alienating parent may also be ordered into individual therapy. The individual therapy for the alienating parent should include education about the damaging effects of parental alienation on the child and the consequences of alienating behavior within the legal system ("The judge will never allow unmonitored visitation if you continue criticizing the mother in front of the child."). The therapist should monitor for alienating behavior and express praise when the alienating behavior is improving ("I spoke to the monitor and she says that you are doing a great job at not criticizing the father in front of the child."). The therapist should also express disappointment if the alienating behavior continues ("The letter you wrote your child stating 'your dad is a jerk' was inappropriate. I know you can do better."). The individual therapist should speak with the conjoint therapist and with others, so that alienating behavior can be promptly addressed. Both the conjoint therapist and the individual therapist may be asked to provide updates to the court regarding progress so that custody and visitation orders can be modified. An individual therapist for the child may be ordered by the court as well.

If the child is alienating a parent even though the favored parent is supporting the alienated parent's relationship, the focus of the therapy should be to stop this oppositional behavior. The child should be confronted with his manipulative ways and told that the behavior should stop. He should not be rewarded for refusing visitation; he should be disciplined. If the child cannot be forced to have visitation, his

privileges should be withdrawn. Behavior such as making false allegations should be similarly punished. The individual and conjoint therapists may be ordered to accomplish these treatment goals. Residential treatment may be ordered in special circumstances.

One of the important treatment goals in circumstances of alienation is for the child to learn proper conflict management skills. Prior to terminating therapy, the child should experience that it is unacceptable to lie, vilify, and reject those with who they are in conflict, especially loved ones.

References

[1] Bernet, W. & Ash, D. (2007). *Children of Divorce: A Practical Guide for Parents, Therapists, Attorneys, and Judges*, 2nd Edition, Kreiger, Malabar.
[2] Garrity, C.B. & Baris, M.A. (1994). *Caught in the Middle: Protecting the Children of High-Conflict Divorce*, Lexington Books, New York.
[3] Gardner, R. (1992). *The Parental Alienation Syndrome*. Creative Therapeutics, Cresskill.
[4] Gardner, R. (2001). *Therapeutic Interventions for Children with Parental Alienation Syndrome*. Creative Therapeutics, Cresskill.
[5] Gardner, R., Sauber, S.R. & Lorandos, D. (2006). *The International Handbook of Parental Alienation Syndrome: Conceptual, Clinical and Legal Considerations*, Charles C. Thomas, Springfield.

JOSEPH KENAN AND WILLIAM BERNET

Parental Rights and Prerogatives

Introduction: Parent, Child, and State

Parental possession and exercise of rights concerning their children must be understood in context. The rights of parents, the rights of children, and the interests of the state form a triangle. Laws affecting the parent–child relationship reflect the three legs of this triangle. The sources of such laws include the United States Constitution and the legislative, executive, and judicial branches of the federal and of state governments. Traditionally the states, as opposed to the federal government, are the primary source of law concerning families and there is considerable variation in statutes, regulations, and case law (court decisions) among states. Therefore, a forensic expert should not assume that, for example, laws defining sexual abuse or a parental duty to provide health care are exactly the same in every state; rather, the expert should take steps to inform him or herself about the law of the state where the expert is asked to work. Experts should also be aware that a state constitution may provide parents or children with rights beyond those guaranteed under the federal constitution, and that these rights will be enforced by that state's courts.

The parent–child relationship itself is the source of both parental rights and duties. A parent has the right to custody and companionship of his or her child – and a corresponding duty to provide the child with care and support. Generally, under the United States (federal) Constitution, a parent has a fundamental right to raise a child as the parent sees fit. The parent may exercise rights and prerogatives (and the law will defer to the parent) up to the point where there is actual harm or a substantial risk of harm to the child. At that point the state can intervene and impose restrictions upon the parent and child relationship.[a]

Thus, for example, the state can make education compulsory for children between certain ages (typically 6 to 16). It can enforce this requirement by prosecuting at criminal law parents who do not send their child to school. It can also act through the juvenile justice system against children of school age who are truant – whether or not the parent knows of or condones the truancy. However, the parent has a constitutional right to choose for the child's education a public or private school, or alternative types of education such as homeschooling.[b] The child does not have the legal right to choose how or where he or she will be educated. However, the state's interest in educating children and the child's interest in being educated are enforced by state curriculum and record-keeping regulations, which nonpublic schools, including homeschools, must satisfy.[c]

Similarly, parents have the right to give or deny their children permission to engage in paid labor. However, the state has power to limit the types of work, hours, and conditions under which children can be employed.[d] And although traditionally parents were entitled to their children's wages, state and federal laws now impose trustee obligations on parents

of child entertainers, athletes, models, etc. to manage and preserve the earnings for the benefit of the child.[e]

Parents have a "privilege" to use "reasonable, nonexcessive force" to discipline their children, including using physical restraint and corporal punishment. "Privilege" means that the parent has a right to do something to the child (confinement, infliction of pain) that would be otherwise be unlawful. The parent may delegate the privilege to a third party (a baby-sitter, a private school teacher) but cannot authorize any action that would exceed the scope of parent's own privilege. The privilege only applies to "reasonable, nonexcessive force". The state can act to prosecute criminally a person whose punishment of the child exceeds the scope of that privilege, can find the child abused under dependency court protection, or permit the minor through a legal representative to sue the person at civil law (in tort) for battery.[f]

The state's authority to do all this derives from two sources: the police power and the state's role as *parens patriae.* Under police power, the state can act to protect the child from harm and to protect the community from danger caused by the parent and possibly the child: for example, through criminal laws against child abuse and the juvenile justice system for children who commit acts that would be crimes if the minors were adults. *Parens patriae* means the state is acting as a parent. Under this source of authority, the state can use law to encourage and help parents carry out their duties of care and support, and to provide for and protect children whose parents are unable or unwilling to do so. Citing the special vulnerability of minors, states invoke *parens patriae* to justify individual laws such as special curfews for minors,[g] as well as dependency court jurisdiction over abused, neglected, or abandoned children.[h]

Juvenile Justice System

Juvenile justice courts generally have jurisdiction over three categories of children: those who have committed offenses which would be crimes if committed by adults (delinquents); those who have committed offenses that were against the law only because of the status of being a child (status offenders); and those who have been abandoned, abused, or neglected by their parents or guardians (dependent or neglected children). The juvenile justice system is a civil rather than a criminal one, and uses a distinctive terminology reflecting its original nonpunitive, rehabilitative purpose.[i] In the adult criminal justice system, the state files charges alleging that the defendant committed a crime; the case title reads "State (or Commonwealth) versus Name, Defendant". By contrast, in a juvenile delinquency, status offense, or dependency case, the state files a petition "in the matter of [first name and initial], a minor". The subject of the petition is referred to as *the minor*, rather than *the defendant,* and the petition contains not "charges" but reasons why the minor should be taken under the court's jurisdiction: the minor is a delinquent, the minor is a status offender, the minor is dependent or neglected. The court does not find the minor "guilty" or "not guilty" in delinquency or status offender cases, but rather sustains or denies the petition for jurisdiction. The hearing at which the court does this is called *an adjudication* rather than a trial. If the petition is sustained (that is, the court finds jurisdiction over the minor), the next step is "disposition" rather than "sentencing" of the minor. If the petition is denied, the case is dismissed. Similarly, in dependency cases the court does not find the minor's parent or parents "guilty" of abuse or neglect – rather it does or does not find the minor to be abused, neglected, or dependent. Only if the court sustains the dependency petition can it proceed to disposition and impose restrictions on the parent–child relationship up to and including termination of parental rights.

Delinquency

Under state codes, "delinquent" acts generally include the full range of behaviors considered criminal for adults, from minor crimes (misdemeanors) such as petty theft or disturbing the peace, to major crimes (felonies) such as robbery, rape, and murder.

The US Supreme Court has ruled that children who are the subject of delinquency petitions have the fifth amendment constitutional right against self-incrimination (also known as *the right to remain silent*) and the sixth amendment right to the assistance of an attorney (at state expense if the child is indigent)[j] Moreover, the state has to prove each element of the alleged delinquency beyond a reasonable doubt – the same standard of proof required in adult criminal cases.[k]

Minors alleged to be delinquent have most of the same rights to due process as are provided to adults in criminal proceedings, with two exceptions. Minors do not have the right to bail, nor the right to a jury trial.[l]

Dispositions in delinquency cases may range from "home on probation" to placement in "boot camps" or "group homes" to confinement in secure state youth facilities barely distinguishable from adult prisons. In most states, the maximum age for persons over whom the juvenile court has jurisdiction is 18 years. The age of jurisdiction refers to the age of the juvenile at the time the offense was committed. In the great majority of delinquency cases, the disposition imposed will end when the minor turns 18. However, for extremely serious offenses (such as murder) a juvenile may remain under the court's authority until age 21 or even 25, serving the years after age 18 in the adult prison system.[m]

Every state's juvenile justice system provides for some minors to be tried in adult court. Some state laws permit the district attorney to file charges directly in adult court when minors above a stated age are accused of a serious crime. Most states begin the case in juvenile court and hold a "fitness" or "waiver" hearing to determine whether the minor should be tried as an adult. "Fitness" refers to the minor's fitness to remain in the juvenile justice system: 'waiver" to the procedure under which the juvenile court waives jurisdiction over the minor in favor of transfer to the adult criminal court. Minimum age at which a minor can be transferred to or tried in the adult criminal court varies from state to state (from 10 to 16 years).[n] An adjudication of guilt in adult court exposes the minor to the same penalties for crime established by statute for adults, with the sole exception of the death penalty. The United States Supreme Court has declared it unconstitutional to execute an individual for a crime committed before his or her 18th birthday.[o]

In delinquency cases, parents' rights are minimal. The minor, not the parent, is the subject of the proceeding and his or her rights are those at stake. The minor, not the parent, has the right to assert or waive the 5th amendment right to remain silent, the amendment right to counsel, and the minor alone can choose to waive trial and enter a plea.[p] State laws typically give the parent a right to notice that a delinquency petition has been filed concerning the child, and the right to be present at critical stages of the case, such as a detention hearing or adjudication. However, the parent is not a party in the delinquency case and does not have a right to be heard by the court. The state can call the parent as a witness against the minor, however, and if the parent refuses to testify, the court can impose contempt penalties.

Because the *minor* has a constitutional right to counsel in a delinquency case, one will be provided at public expense if he or she cannot afford one. Of course, a parent may hire a private attorney to represent the child's interests; but whether private or public defender, ethically the attorney is bound to follow the minor client's instructions and serve the client's interests. The parent can urge the child to follow a particular course of action (for example, enter a guilty plea), but the attorney's advice to the minor client must be independent of the parent's wishes. Neither the parent nor the attorney can waive the minor's constitutional rights. Moreover, the attorney cannot reveal attorney–client communications to the parent, and with good reason. In the great majority of states, there is no parent–child privilege, so the parent can be forced to disclose confidential communications made by the child.[q]

Status Offenses

Depending upon the particular state code, "status offenses" may include behaviors such as truancy from school, running away from home, being habitually disobedient or "beyond the control of parents", or violating a local curfew law. These behaviors may be combined under a single label such as "children in need of supervision (CHINS)", "juveniles in need of supervision (JINS)", or "persons in need of supervision (PINS)".

States vary in the degree of procedural protections provided to minors who are the subject of status offender petitions.[r] Since status offenses are not "crimes", children in these proceedings may not be entitled to the 5th amendment right against self-incrimination. The US Supreme Court has not yet ruled on this matter. Under most state laws, children in status offense cases have the right to notice of the allegations against them, the right to a hearing, and right to the assistance of counsel. The standard of proof is only *preponderance of evidence*: that is, to take jurisdiction over the minor the judge need only find it *more likely than not* that the allegations have been proved. Federal and state laws prohibit "secure confinement" (placement in a locked facility such as a juvenile hall) of status offenders except in cases where a minor has been held in contempt for violation of a valid court order.[s]

Parents are likely to have greater involvement in status offense cases. The parent may have contacted the police or the juvenile court asking for help in controlling a child who is habitually disobedient or "incorrigible". In such a case, the parent will likely be the state's primary, if not sole, witness at the adjudication hearing. If the petition is sustained, the judge will try to order a disposition that will reinforce parental authority (often probation with specific conditions that the minor must obey her parents, avoid certain "bad influence" peers, submit to drug tests, etc.). If the minor violates these conditions, the parent can notify the court and the state can file contempt charges as discussed above.

However, the state is not required to file a petition just because the parent wants state reinforcement of his or her authority. The probation officer or prosecuting attorney may decide, rather than filing a petition, to refer the parent and child to family counseling or other services on a voluntary basis. Similarly, the state can file status offense petitions for truancy and curfew violations regardless of the parent's wishes, and may take action against a parent who has condoned the child's unlawful behavior. Once the court has sustained the petition, it can order the disposition it believes is best for the child, even if the parent opposes it.[t] If the court orders placement of a status offender in a foster home, group residence, or other out-of-home program, the parents may be required to pay the costs, as part of their duty to support their child.

Dependency Cases

Every state provides for protective action to be taken where a child is abused, neglected, abandoned by parents or guardians or otherwise "dependent" upon the state.[u] In some states, the "dependency" cases will be heard by the juvenile court: in others by the family court that also hears cases of divorce and child custody. Unlike family law custody cases, however, dependency proceedings are not conflicts between two private parties; rather, the state petitions the court to take jurisdiction over a child as "dependent" because of the acts or failure to act of parents or guardians.[v]

The degree of state intervention into the parent–child relationship through the dependency court can vary greatly. It may be temporary, as when life-saving surgery for a child is authorized by the court over parental objection yet care and custody of the child are returned to the parents immediately after the medical procedure.[w] It may last for months, as one or both parents try to regain custody of the child through compliance with a "reunification plan" designed to remove or alleviate the conditions that caused harm or put the child at risk.[x] It may be permanent, as where all parental rights are terminated, severing the legal relationship between parent and child. (Termination means total parental loss of rights to custody, visitation, and communication. The child becomes eligible for adoption by new parents, or for permanent placement with a legal guardian or foster family where adoption is not possible or appropriate.)[y]

Under the 14th Amendment to the United States Constitution, parents whose rights to care and custody are at risk in dependency proceedings must receive due process of law. Parents are entitled to notice of the allegations supporting the dependency petition and have the right to appear and be heard at the detention hearing (to decide where the child should live pending adjudication), the adjudication, the disposition hearing, and any hearings to review the status of the case. A dependency petition can be sustained at adjudication on a mere preponderance of evidence, because taking jurisdiction over the child is only a temporary deprivation of parental custody. Parental rights cannot be *terminated* without a finding based on clear and convincing evidence.[z] Nevertheless, the United States Supreme Court ruled that indigent parents are not always entitled to the assistance of counsel at public expense before termination of their rights: judges should consider parents' request for court-appointed free counsel on a case-by-case basis.[aa] In practice, most states do provide counsel for indigent parents in cases where termination is proposed.[bb]

The minors who are alleged to be "dependent" have an interest in reunification with their parent or guardian if this is possible without "detriment" or serious harm. If reunification is not possible, however, the child has an interest in permanent placement with legal guardians or adoptive parents who can and will provide appropriate care. Thus both the attorney representing the petitioner state and the attorney(s) for the parent(s) will argue that their client's proposed action also will benefit the child! The dependency court has the power to appoint counsel for the minor, however, and this is

increasingly becoming routine.[cc] As in delinquency cases, a minor's attorney, whether paid by the state or retained privately, is ethically required to represent the minor's interests. When the minor is old enough to consult with the attorney and reason out a position, the attorney must inform the court of the minor's wishes. When the minor is too young to do this (in some cases the minor can be an infant), the attorney must act as a *guardian ad litem* and recommend to the court what the attorney believes is in the minor's best interests.[dd] Therefore, minor's counsel may support parents' position or the state's recommendation – or advocate a different disposition from that proposed by either of the other parties.

End Notes

[a.] It is cardinal with us that the custody, care and nurture of the child resides first in the parents, whose primary function and freedom include preparation for obligations the state can neither supply nor hinder . . . And it is in recognition of this that [cited] decisions have respected the private realm of family life which the state cannot enter. But the family itself is not beyond regulation in the public interest . . ." *Prince v. Massachusetts*, 321 U.S. 158 (1944).

[b.] Parents Authority over the Education of Their Children: *Pierce v. Society of Sisters*, 268 U.S. 510 (1925) (establishing "the parent's authority to provide religious and secular schooling and the child's right to receive it, as against the state's requirement of attendance at public school").

[c.] Child's wishes not considered: *Wisconsin v. Yoder*, 406 U.S. 205 (1972) (holding that state compulsory education law is unconstitutional as requiring Amish parents to send their children to school for 2 years of high school . . . without regard to the wishes of the child).

[d.] State enforcement of child labor laws: *Prince v. Massachusetts*, 321 U.S. 158 (1944) (upholding enforcement of child labor law to prohibit 9 year old Jehovah's Witness from distributing religious publication on street).

[e.] For example, California Family Code Section 6750 authorizes the superior court to approve or disapprove minors' contracts "for the provision of artistic or creative services". Based on the "Coogan Law" first enacted in 1939, California Family Code Sections 6752 and 6753 require a portion of the minors' earnings to be set aside in trust until the minor reaches the age of majority (18).

[f.] Corporal punishment of children: Philip Jr, G. (1990). *Spare the Child: The Religious Roots of Punishment and the Psychological Impact of Physical Abuse*, Alfred Knopf.

[g.] Constitutionality of curfew laws for minors: Katherine, H.F. (1995). Children, curfews and the constitution, *Washington University Law Quarterly* **73**, 1315.

[h.] State power to establish dependency court system: Institute of Judicial Administration and the American Bar Association (IJA/ABA) (1977). Standards Relating to Child Abuse and Neglect.

[i.] History of juvenile court: Douglas, E.A. (2003). *A Very Special Place in Life: The History of Juvenile Justice in Missouri*, Missouri Juvenile Justice Association.

[j.] Due process rights of minors in delinquency cases: *In re Gault*, 387 U.S. 1 (1967) (children in delinquency proceedings are entitled to due process protections under the 14th Amendment to the Constitution, including the 5th amendment rights against self-incrimination, and the right to counsel).

[k.] Standard of proof in delinquency cases: *In re Winship*, 397 U.S. 358 (1970) (each and every element of the state's case in delinquency proceeding must be proved beyond a reasonable doubt).

[l.] No right to jury trial or to bail: *McKeiver v. Pennsylvania*, 403 U.S. 528 (1971) (the due process clause of the 14th Amendment does not assure the right to a jury trial in the adjudicative stage of a delinquency proceeding). *Schall v. Martin*, 467 U.S. 253 (1984) ("pre-adjudication" detention in locked juvenile facility of minors charged with delinquency did not violate their right to due process).

[m.] Juvenile court dispositions: Edward, H. (1996). *No Matter how Loud I Shout: A Year in the Life of Juvenile Court*.

[n.] Waiver laws and practices: Howard, S. & Melissa, S. (1995) *Juvenile Offenders and Victims. A National Report*, pp. 85–89, 154–156.

[o.] No death penalty for offense committed before age 18:
Roper v. Simmons, 543 U.S. 551 (2005).

[p.] Minor exercises own rights:
Katherine, H.F (1996). The Ethics of empowerment: rethinking the role of lawyers in interviewing and counseling the child client, *Fordham Law Review* **64**, 1655.

[q.] Role of attorney for minor in juvenile delinquency cases:
Institute of Judicial Administration and American Bar Association (IJA/ABA), Standards Related to Counsel for Private Parties (1976).
Jan C.C. (1980). Ethical issues in representing juvenile clients: a review of the IJA-ABA standards on representing private parties, *New Mexico Law Review* **10**, 255.
[r.] Due process rights afford status offenders:
Erin, M.S. (1992). In a child's best interests: juvenile status offenders deserve procedural due process, *Law and Inequality Journal* **10**(253).
[s.] Use of the contempt power to confine status offenders:
Jan, C. & Nancy, W. (1981). Incarcerating Status Offenders: Attempts to Circumvent the Juvenile Justice and Delinquency Prevention Act, *Harvard Civil Rights-Civil Liberties Law Review* **16**, 41.
[t.] Status Offender dispositions:
Status Offenders National Council of Juvenile and Family Court Judges (1990). A new approach to runaway, truant, substance abusing and beyond control children, *Juvenile and Family Court Journal* **41**, 5.
[u.] Definitions of dependency:
American Bar Association and National Council of Juvenile and Family Court Judges (1995). Resource Guidelines: Improving Court Practice in Child Abuse and Neglect Cases.
[v.] Distinction between dependency and family law cases:
Leonard, P.E. (1987)The relationship of family and juvenile courts in child abuse cases, *Santa Clara Law Review* **27**, 244–245.
[w.] Temporary authorization of medical procedures:
Joseph, G. (1977). Medical care for the child at risk: on state supervision of parental autonomy, *Yale Law Journal* **86**, 645. For a discussion of the balance between affording parents the right to limit medical attention to spiritual care *versus* homicide, see Commonwealth v. Twitchell, 617 N.E.2d 609 (Mass. 1993).
[x.] Reunification efforts:
Robert F.K (2000). Family preservation and reunification in child protection cases: effectives, best practices and implications for legal representation, judicial practice and public policy, *Family Law Quarterly* **34**, 359.
[y.] Susan, V.M. (2000). Extending non-exclusive parenting and the right to protection for older foster children: creating third options in permanency planning, *Buffalo Law Review* **49**, 835.
[z.] Preponderance standard for jurisdiction: clear and convincing standard for termination of parental rights: *Santosky v. Kramer*, 455 U.S. 745 (1982).
[aa.] Indigent parent not entitled to court-appointed counsel in every dependency matter; judge must decide on case-by-case basis:
Lassiter v. Dept. of Social Services, 452 U.S. 18 (1981).
[bb.] Parent's lawyer in child abuse cases:
Bruce, A.B. (1996). Ethical issues in representation of parents in child welfare cases, *Fordham Law Review* **64**, 1621.
[cc.] Martin, G. (2006). How children's lawyers serve state interests, *Nevada Law Journal* **6**, 805.
[dd.] Role of counsel for minor in dependency court:
Special Issue (1996). Ethical issues in the representation of children, *Fordham Law Review* **64**, 1281.
Ann, M.H., The Child's Attorney (1993, American Bar Association).

Related Articles

Children: as Defendants

Parental Alienation

JAN COSTELLO, JOSEPH KENAN, AND CHRISTOPHER THOMPSON

Parenting: Assessment of Capacity

Child maltreatment (abuse and neglect) is a serious problem that often surfaces with reports of suspected abuse being made to Child Protective Services (CPS) or law enforcement. In cases where serious maltreatment has occurred or where risk cannot be contained, children may be removed from the home by court order and placed in custody of the state. In such cases, there is often involvement of psychologists along with other professionals providing information to the courts about the individuals in the family, their intervention needs, whether maltreatment

likely occurred, and suggested ways to deal with risk factors.

In the United States, only in the past 40–50 years have there been national laws regarding the maltreatment of children and widespread state mandated reporting laws. Some of the first interventions for protection of children in this country were offshoots of efforts to control cruelty to animals [1, 2]. The hesitancy to establish state authority in this area was based on notions that children were possessions, that children were undeveloped morally and must be strongly disciplined in order to attain civility, and that family matters were best left to families to work out internally. It was only in the mid-nineteenth century that children came to be regarded as more than chattel, and it was as late as the seventeenth century before parents were prohibited from killing their children. Even then, they were not required to provide basic necessities for their offsprings [3, 4].

A paper by Kempe, a physician, in 1962 introduced the term *battered child syndrome* and urged physicians to inquire about etiology of broken bones, bruises, and soft tissue damage when children were presented for medical care [5]. Kempe's one-year study had found that 39% of the identified battered children suffered permanent brain damage or death. State mandatory reporting laws for physicians followed publication of the paper, and other professionals were later included. In line with this rising awareness of abuse, in1974, the Child Abuse Prevention and Treatment Act, the first national legislation directly addressing child maltreatment, was passed and defined abuse and neglect as the physical or mental injury, sexual abuse or exploitation, negligent treatment, or maltreatment of a child under the age of 18, or the age specified by the child protection law of the state in question, by a person who was responsible for the child's welfare. This federal law required states to adopt similar definitions in order to get child welfare funds [6]. Government reports indicated in 2004 more than 872000 children and youth were found to have been abused or neglected on the basis of investigation of an estimated 3 million referrals. For 1387 children, the maltreatment was fatal. Another half million had been removed and were living in foster care; of these, 117463 would not ever return to parents but were awaiting adoption by someone else [7].

Maltreatment Effects

Physical Effects

Efforts to study the effects of maltreatment on children are made more difficult because of the frequent co-occurrence of other aversive factors, such as family stress, dysfunctional patterns of family interaction, parental psychopathology, child psychopathology, and chaotic home environment. In addition, the age at which abuse or neglect occurs seems to make a difference in sequelae for children [8]. For instance, younger children are at more risk for death from their injuries than are older children because of greater vulnerability to physical damage, lack of ability to seek medical care for themselves, and lack of ability to inform others about risk factors [9].

Even when death does not occur, results can be very tragic as in cases of shaken baby or trauma to the head. Acceleration, deceleration, and twisting of the brain within the cranial vault, produced by shaking or throwing the child, can lead to life threatening emergencies such as hemorrhage and brain swelling [10]. These conditions can result in mental retardation, speech and language delay, and learning disorders. Some research shows significantly lower cognitive skills in abused children even when there is no evidence of head trauma. Some of these deficits may be attributed to physical neglect, malnutrition, or exposure to toxic substances and other factors, which complicate efforts to identify the particular effects of physical abuse [11].

Psychological Effects

Physically abused children have been found to show increased noncompliance, more aggression, deficits in social skills and peer relationships, less empathic tendencies, and worse adjustment to school compared to control groups [8]. Condi [12] listed several factors found to be associated with physical abuse of younger children: developmental delay, anxious attachment to parents, withdrawal or apathy, and hyperarousal. Older children with abuse history have been found to have more problems with social cognition, self-esteem, conduct problems, and substance abuse. Some studies have found over one-third of physically abused children meet criteria for posttraumatic stress disorder (PTSD) with about 10% meeting the criteria as long as 2 years after initial diagnosis [13].

Diagnosis of PTSD is also often justified when children have been sexually abused, and more than one-third continue to meet the criteria for the diagnosis as adults. These children and their grown-up fellow survivors are more likely than their peers to be depressed and to attempt suicide, to have sexual behavior problems, and to engage in self-mutilation [14]. Fortunately, even with the harsh sequelae for some survivors, literature reviews indicate that from 25 to 40% of sexually abused children do not show psychological symptoms in the short term. Of those having initial symptoms, more than half later show improvement [15].

Children who have been neglected show increased behavior problems, aggression, and school maladjustment. In childhood, they are often anxiously attached to parent figures with some studies showing up to two-thirds of neglected one-year olds lacking secure attachment. They often become angry and unaffectionate but highly dependent on caregivers; they tend to lack persistence, to be noncompliant and negativistic, and to be more unpopular than nonneglected peers. They are at greater risk than peers for academic failure and for dropping out of school [16].

Protective Factors

Several protective factors associated with less severe sequelae from sexual abuse were reported by Berlinger and Elliott [14] and include: less serious and less frequent or single occurrences of abuse; perpetrator not a central person in child's life; child informing parent who was then protective and supportive; child having a close bond with supportive parent; child having stable and mature family with limited dysfunctional patterns; child not being subject to lengthy court testimony or repeated interviews; and child being provided stress inoculation and support if facing testimony. In a study of 369 sexually abused children, Conte and Schuerman [17] found the one variable with the greatest predictive power for impact on the child was a poorly functioning family.

For maltreatment in general, research in the field suggests the long-term effects of abuse are ameliorated by the child having a loving, supportive adult who provides a distinct model of a caring child–adult relationship and offers the child a positive way to form a self-concept. Another factor found to be predictive of less harmful adjustment is successful participation in psychotherapy to enhance emotional stability and maturity as well as to facilitate the integration of the maltreatment into a coherent view of self (rather than becoming dissociative or fragmented) [18, 19]. In addition to these environmental or situational factors, internal factors such as flexible coping style (suggestive of emotional maturity), resistance to self-blame, at least average cognitive abilities, self-confidence, and active coping have been found to be helpful [14, 20].

Factors Associated with Maltreatment

Theories of Causation

As the problem of maltreatment has come to be seen as more complex, the models of explanation have likewise shifted from single causes to more multifactorial models which take into account potentiating and protective factors as well as complex interactional factors (information processing, stress management, developmental stage of the child, environmental factors, and interpersonal skills). A critical decision in development of a theory of causation (and its application for treatment approaches) is whether to see abuse and normal parenting as dichotomous or as points on a continuum. Those who hold a dichotomous view see parental maltreatment as the result of permanent "defect" (mental retardation and character disorder) while those who see a continuum of behavior patterns tend to look for parenting "deficiencies" (amenable to change) or "disruption" (changeable social or interpersonal forces interfering with appropriate parenting on a situational basis). A blended approach sees "differences" or mismatches between parent and child (needs, abilities, temperaments, and interactions) which to some degree can be ameliorated. The position one takes regarding the nature of maltreatment shapes the approach to evaluation and intervention [21].

Sociobiological Model. Child abuse is not limited only to humans, and studies show similarities of factors for human and nonhuman parents. A study covering 12 years of records for monkeys kept in a primate lab at the University of Colorado found abuse involving 28% of the infants; abuse was defined as physical abuse or neglect resulting in a medical problem. Of the abused infants, 62% died. Data supported hypotheses that abuse resulted from learned

aggression (intergenerational transmission of parenting patterns), aberrations of attachment, social alienation, and overcrowding. The fact that causative factors are similar to those for humans but incidence rates are different in this colony compared to statistics with human studies suggests factors such as cognitive beliefs and/or social norms may function in a protective manner [22].

Belsky [23] pointed out that from a biological perspective, interests of parent and child are not always the same and may even be in direct conflict for success even if not for survival. This may be particularly evident in situations of limited resources (such as poverty or single-parent household) coupled with excessive demands or competition (large families, unplanned pregnancies, ill or handicapped children). This biological/evolutionary model may also explain the higher occurrence of maltreatment of stepchildren or by young mothers (who are more capable of bearing additional children than are older mothers) or on young children (who are not reproductively mature and thus more expendable).

Intergenerational Transmission. This explanatory model is largely based on the recognition that abused children when grown-up are more likely to be abusive to their children than are parents who were not abused or neglected in childhood. But review of research finds intergenerational transmission rates of about $30 \pm 5\%$ meaning that two-thirds of abused children do not grow up to be abusive (at least during the brief windows of time on which research studies focus) [23, 24]. These figures are compared with self-reported figures obtained in epidemiological studies in which incidence rates in the general population are about 1% [25].

Information Processing Theory. Crittenden [26] presented a cognitive model with four elements which are involved in decisions and ultimately lead to behavior. These elements are perception (affected by biases); interpretation (affected by schemas and attributions); response selection (limited by repertoire and influenced by expectations); and implementation of behavior (affected by available resources and hierarchy of conflicting demands). Parents who become neglectful or abusive may be influenced by thinking errors, such as bias against a child, filters which block out or minimize significant positive information, faulty attributions such as parental helplessness or meanness of the child, or role reversing choices based on expecting children to be responsible for the wellbeing of parents.

Other elaborations on this model have been provided by Dix [27] and Azar [21, 28]. Azar's model posits that in the situation of a parent with bad schema for understanding child behavior combined with high risk personality factors (poor empathy, low flexibility, high distress, impulsivity), the abuse unfolds in four stages: (i) parent has unrealistic standard; (ii) child fails to meet standard; (iii) parent misattributes negative intent to the child or blames self when intervention does not change child's behavior; and (iv) parent overreacts and punishes excessively.

Presented below are factors which have been found across reviews of research in the field of child maltreatment. Most of the literature relates to neglect or physical abuse rather than sexual abuse although these categories are by no means mutually exclusive. However, there are also some significant differences between sexual abuse and other forms of maltreatment. For instance, although females (mothers) more often than males are identified with physical abuse or neglect, males by far outnumber females as reported perpetrators of sexual abuse. Also, physical abuse and neglect are almost always perpetrated by parent or other primary care giver while only about one-third of sexual abuse perpetration is by parent or primary care giver [17]. Sexual abuse has other unique dynamics as well such as family enmeshment and deviant sexual arousal patterns or practices of offenders [29].

Contextual Factors Associated with Maltreatment

Many specific contextual factors have been found to be predictive of abuse or neglect [16, 23,30]. For instance, poverty and scarcity of resources have consistently been predictive of higher rates of neglect and physical abuse. Living in a neighborhood with poor social fabric, no community pride, poorly organized resources, or an atmosphere of violence and aggression increases likelihood of maltreatment. Belsky [23] noted that cultures in which corporal punishment was rare had low rates of abuse. Chaotic home environment, occurrence of domestic violence, social isolation of the family, or presence of physically or emotionally demanding child care situations tend to increase risk of abuse.

Parental Factors Associated with Maltreatment

Identified parent factors are more numerous than contextual factors. Some are biological in nature (young age of parents and strongly aroused physical reactivity to stressful stimuli), whereas others are psychological. The neurotic triad of depression, anxiety, and hostility increases risk. Epidemiological studies [25] have found depressed parents have four times the risk for being abusive as nondepressed parents although Belsky [23] stated the association is curvilinear for abuse (greatest for moderate levels of depression) and linear for neglect (greatest for highest levels of depression). Maltreating parents also tend to be impulsive and to have poor self-esteem and low ego strength. The Minnesota Mother–Child Project, which followed at-risk first-time mothers, found emotional stability was the greatest single predictor of good caretaking *versus* maltreatment [31].

Behavioral patterns of parents differentiate maltreating from nonabusive parents. The maltreating individuals tend to be strict disciplinarians (abuse group) or inconsistent (abuse and neglect groups). They tend to use physical punishment or coercive discipline in preference to other forms of guidance and thereby place themselves at greater risk; if they are also more physiologically reactive and psychologically distressed, physical punishment can get out of hand and quickly become violent aggression. They are less effective at managing problems with children. They tend to be less responsive to needs and emotions of children, to have fewer play items for children in the home, and to engage in less communication and physical interaction with their children. Substance abuse is a strong predictor which was found in epidemiological studies to increase risk for neglect fourfold [25]. They tend to isolate themselves, to be more transient than nonabusers, and to remove themselves from social supports that are available to them.

Cognitive factors differentiating maltreating and other parents include more emotion-based than reason-based problem solving. The maltreating parents have less cognitive flexibility and more rigid thinking even though attention, distractibility, and verbal fluency are not found to be significantly different. These parents often have a negative attitude toward their children and tend to see negative behaviors of children as internally caused and stable, whereas positive behaviors are seen as externally caused (perhaps attributed to the parent) but not stable [13]. As a result, abusive parents tend to develop a negative, adversarial schema that leads to misinterpretation; neutral or ambiguous behaviors are seen as examples of misbehavior or even malice on the part of the children. Neglectful parents also have a negative schema but tend to believe that relationships will not be fulfilling. Finally, the internal model for parenting behaviors is often a deficient one as many maltreating parents have themselves been reared by neglectful or abusive parents [23, 30].

Treatment Interventions with Parents

Funding from the Child Abuse and Prevention Treatment Act of 1974 [6] led to many well-developed intervention projects, which were then evaluated in efforts to discover what was effective. Daro [32] reviewed results from 19 demonstration projects and found relapse rates ranging from 20% (sexual abuse) to 66% (neglect). She noted family therapy, individual problem-solving training, and group therapy tended to be associated with more successful results. She also reported on 89 programs involving over 3200 families [33], which found the best outcomes were associated with intensive services initially and a long maintenance period. Better results were also found with use of lay services (volunteers in the homes of target families) and use of groups for the parents (group therapy and classes). Interventions shorter than six months or longer than 18 months were less likely to show progress. Results suggested intervention should be comprehensive in addressing concrete (housing and medical services) and interpersonal needs (social support and faulty interaction patterns), but the most effectively spent money was for prevention.

Other research [34] found in reviewing control study research that a multifaceted, behavioral approach was most effective in reducing posttreatment relapse rates (10% in treated *vs.* 21% in control groups); but also found brief treatment did not work well. Home-based services with practical skills training and with family therapy showed vastly better rates for children remaining in home at one-year follow-up (74% with such services *vs.* 45% without). One example (Project 12-Ways) of this intensive, comprehensive model has been well described by Lutzker [35–37].

Legal Context for Considering Parenting Evaluations

The legal context for most child protection matters is domestic court or family court although criminal courts may become involved where evidence of maltreatment is sufficient to lead to an indictment and criminal prosecution. Domestic courts are not constitutionally established but arise from English common law where chauncery courts or common courts addressed marital disputes and matters concerning children. Although hesitant to become involved in matters of the family until there is breakdown of the family unit and/or harm to a child, the courts have a significant societal interest to take action when natural parent and child bonds are no longer sufficiently effective to assure the best interests of the child are being addressed by the family [3].

Even though parents are under statutory obligation to provide basic needs (physical, medical, and educational), they are mostly free to make decisions about residence, family structure, discipline, religious practices, social activities, financial arrangements, and even educational choices as long as basic needs are addressed and peace and order accomplished. Balanced against this deference to parents and privacy of the family are legal precepts recognizing the rights of children (protection of their best interests in disputed matters) and the obligation of the state to intervene when children need protection. This intervention is based on the concept of *parens patriae*, which regards the state as the protector of those citizens unable to care for themselves [38]. With origins in Roman law incorporated into English common law in the eleventh century, the concept extends the state's interest into the family and other matters when children (or disabled individuals) are unable to be guardians of their own interests.

Below are some of the legal cases from US courts, which help define the balance of interests in child-dependency cases.

Griswold v. State of Connecticutt (1965) [39]: prohibited undue government intrusion into "zone of privacy for family implied by Bill of Rights and 14th Amendment".

Parham v. J.L. & J.R. (1979) [40]: allowed parents to maintain substantial role in decision making for child unless abuse, neglect, or evidence contradicting assumption of parent acting in best interest of child.

Ingraham v. Wright (1977) [41]: corporal punishment not cruel and unusual for child; 8th Amendment protection only for convicted citizens and does not apply to children.

Deshaney v. Winnebago County Department of Social Services (1989) [42]: state is not liable for failure to protect child from injury by parents; 14th Amendment designed to protect citizens from state, not insure state protects individuals from each other.

Santosky v. Kramer (1982) [43]: threshold for permanent removal of children must be clear and convincing evidence of abuse; due process must be afforded to parent.

South Carolina Department of Social Services v. the Father and the Mother; In the Interest of the Child (1988) [44]: court affirmed religious freedom for beliefs but regulation of behavior by laws in case where parents claimed religious right for corporal punishment which left bruises on child.

State v. Evans (1992) [45]: unmarried father not living in home guilty of criminal neglect when child died of malnutrition; father did not assure child received care.

In re Glenn G. (1992) [46]: mother neglectful for failing to protect children from abusing, battering father.

E.C. v. District of Columbia (1991) [47] and *Egly v. Blackford County Department of Public Welfare* (1992) [48]: mental illness or mental retardation alone not grounds for termination of parental rights but effects on child and ability to handle needs of child must be considered.

In *Matter of Joshua O.* (1996) [49]: state must present clear and convincing evidence parent is currently and will remain incapable of caring for child if seeking termination of parental rights.

Assessment Issues

Psychologists and other mental health professionals often are asked to become involved in assessment tasks in cases where child maltreatment has occurred or is deemed to be a significant risk. Evaluators are asked to provide information about risk factors and prognosis for change, risks and benefits for parent–child contact or reunification, services needed, and ability to use services if provided. Reports may be used for treatment planning, case planning, or termination of parental rights suits. Because these cases

are always either court cases or have the potential to become such, evaluators are admonished to approach them carefully as the reports may have lasting effects on the lives of the people involved [12, 50]. These evaluations blend elements of clinical assessments and forensic evaluations and are conceptually related most closely to custody evaluations [4, 51]. The following discussion includes the literature from child custody evaluations because of this close relationship and because more research has been conducted on custody evaluations than on care and protection evaluations.

Factors to be Considered in Parenting Evaluations

Lack of Definition of Parenting. Evaluators seeking to conduct assessments within this area of high scrutiny where there is high likelihood of reports being used in an adversarial legal context are handicapped by lack of universal standards or behavioral criteria regarding minimal parenting practices. In addition, there is scarcity of appropriate tools for assessment of parenting skills. In a context of cultural pluralism and social class differences, evaluators provide information so that decisions can be made about "good enough" parenting, a concept arising from a British pediatrician-turned-psychoanalyst, D.W. Winnicott. Winnicott believed that child resilience made up for a host of parental limitations and mistakes.

Conceptual Approach. Grisso [52] proposed a model, which has become well accepted for exploration of various capacities and resulting reports in forensic matters. He suggested five components. "Functional" assessment explores a person's skills and deficits and often uses clinical as well as specialized forensic assessment instruments; it measures what a person is able to do. The "causal" component explains why these deficits occur; sometimes there is a connection between the deficits and a mental health condition. Thirdly, the "interactive" component, too often neglected in forensic reports, applies the functional limitations to the particular forensic context; it states or predicts what a person is specifically able or not able to do in terms of expected behavior in a particular required task. The other two components, "judgmental" and "dispositional", belong not to the evaluator but to the decision maker to determine if the person's abilities are adequate for the task and how the case proceeds. Such an approach to parenting evaluations does not stop with description of skills and deficits of the parent but relates these factors to specific effects the parent's behavior has or is likely to have on the particular child [4].

Common Elements of Parenting. Hoghughi's conceptual model has been used to simplify the myriad of parenting tasks to three elements: care, control, and development. Care refers to provision of physical needs as well as emotional factors; control refers to guidance and discipline; and development refers to those resources and conditions necessary for a child to mature according to developmental milestones. In order for parents to be successful with these tasks, they must have knowledge, motivation, resources, and opportunity (access) [53]. Another simplification has resulted in a model of two basic tasks: protection and care of the child for the purpose of socialization. The parent must know the child and attend to limits and discipline [54].

Factors Associated with Adequate and Inadequate Parenting. White [50] proposed flexibility as the hallmark of good parenting and suggested it was evident in the sustained pursuit of provision of the child's needs through accommodation to the changing environment and the changing child. Some factors which researchers have found to be identified with adequate or good parenting include being supportive and responsive or 'joining' the child, providing instruction and guidance, providing nurture and showing affection through talk and touch, having clear boundaries between parent and child, and maintaining a mutual positive emotional attachment. In contrast, features that have been associated with inadequate parenting or maltreatment include being controlling or power oriented with the child, being hostile or rejecting, being unresponsive or detached, using inconsistent or physical discipline, having high stress and poor coping skills, being emotionally immature or impulsive, having unrealistic expectations of the child, having substance abuse problems, and having poor frustration tolerance or hyperarousal tendencies [2, 55, 56].

Structuring the Assessment and Report

Guidelines. Guidelines provided by the American Psychological Association (APA) offer help in conducting parenting assessments [57]. Principles of

good practice begin by letting the referral question direct the scope of the evaluation, which then develops through the use of multiple methods to gain multiple sources of information. The parent must be informed of the nature and purpose of the evaluation and the limitations on confidentiality. The examiner is cautioned not to make inappropriate interpretations of test data and to tie any recommendations to potential welfare for the child in the matter.

Structure. The assessment process usually involves face to face contact as well as testing, extensive review of records, and use of collateral sources for obtaining information (extended family members, teachers, day care workers, medical personnel, neighbors, and friends). The process usually is structured into three areas: evaluation of the parent, evaluation of the child, and interaction between parent and child [12, 54]. The parent evaluation explores ability of the person to care for self and to care for the child, significant problems (such as substance abuse, mental illness, serious physical limitations, and behavior problems), problem solving and stress management skills, parenting skills and knowledge, access to resources, and empathy or bonding to the child. The child's evaluation includes exploration of progress with developmental milestones and any serious limitations, social and peer relationships, school adjustment, temperament, bonding with the parent, and physical and emotional health.

Observation of parent-child interaction is strongly recommended in this type evaluation although it is not always possible, particularly when termination of parental rights is the case at hand. A report about parent–child interaction should include comments about mutual attachment and boundaries, enmeshment and use of coercion, complementarity or reciprocity of interactions, parental responsiveness, and "goodness of fit" with temperament and needs of the child. Some suggest adding a fourth area for consideration of systemic factors (social context, environmental risk factors, and extended family issues) [50].

Problems Associated with Parenting Evaluations

Difficulties with Testing. Research published about custody evaluation methods led Brodzinsky [58] to criticize indiscriminate use of tests which results partly because psychologists are trained to test and therefore do it and partly because there is financial incentive to produce billable services. There was also note that clinical tests were developed for other purposes and often lacked research applying to the particular legal issue. Although an aura of science is associated with use of tests, use for purposes unsupported by research is deceptive. Finally, there was criticism of poor interpretation practices and not basing interpretation on clear research.

Others have pointed out a major difficulty involved in testing parents in this type evaluation is the response bias problem or socially desirable responding [59, 60]. Parents, usually wanting to present self in the most favorable light in order to impress the decision makers, will often resort to minimizing any difficulties and exaggerating reports of desirable traits and behaviors. Carr [61] reported base rates of compromised validity in these evaluations ranged from 20 to 60%, discouraged the use of tests without validity scales for measuring such response bias, and strongly discouraged adoption of separate validity norms for parents in this type evaluation. The Minnesota Multiphasic Personality Inventory 2 (MMPI-2), Personality Assessment Inventory (PAI), and Child Abuse Potential Inventory (CAPI) were noted to have good validity scales (*see also* **Psychological Testing**). Two other frequently used tests were noted to have deficiencies in the detection of bias: Parenting Stress Inventory (PSI) which has a poorly discriminative measure for defensiveness and Adult–Adolescent Parenting Inventory (AAPI-2) which has no validity scales. The CAPI has a validity scale specifically designed for use with this population. Whether the positive response bias is the result only of the evaluation context or is representative of an actual personality trait of this type parent is a question raised by the author.

Heilbrun [62] made three other suggestions for use of psychological test instruments in a forensic context. First, the instrument should be commercially available with a manual for administration and scoring and should have been reviewed in major publications. The psychometric properties should show reliability measures of $r \geqslant 0.80$. Finally, the test should produce data relevant to the legal issue or construct. There should be available appropriate research published in peer reviewed journals, which provides validity for using the instrument for the particular purpose of the evaluation. Added to these criteria would be an expectation the assessment instrument could pass Daubert [63] criteria if challenged in court.

Problems Associated with Parent–Child Observation. With a single event of limited duration, it is easy to place too much emphasis on what is briefly observed and to overinterpret the behaviors. The observer, who may be unaware of personal bias, may make inaccurate interpretations of behaviors as well; the parent–child dyad may have particular, unique patterns or rituals as well as subculture patterns of interaction. The observer must also be cautious about assuming the observed interactions are representative of the way the parent and child would normally behave if not in the unique situation which undoubtedly makes them anxious. In addition, the child may feel torn between loyalty to parent and perceived necessary alliance with foster parents and child protection workers who control life for the time being. It should also be noted that play activity interactions do not necessarily represent level of bonding, nor do displays of affection or failures to display affection [50, 59, 64].

There is, in addition, a problem with perceived expertise when none exists. Starr [65] found that untrained undergraduate students were better at identifying parent–child dyads with a history of abuse than were seasoned professionals in a parent–child observation task even though neither group was correct by more than chance. Examiner bias (personal prejudice), confirmatory bias (selective attention to information supporting favored hypothesis and disregard for data challenging it), and overuse of experience-based schemata (expectancy and overconfidence leading to premature conclusions) pose substantial threats to accurate perception and good decision making in these observations.

Suggestions for improving parent–child observations and the usefulness of resulting data have been offered in articles and books [56, 66, 67] Some basic recommendations include informing the participants and preparing them for the interaction in order to minimize anxiety, holding the event in a naturalistic setting and providing ample time for getting comfortable with the task, providing opportunities to observe free play as well as chores or structured activity, and minimizing observer interaction during the exercise. Whenever possible, it is useful to observe the child with another caregiver to get a sense of what is unique about interaction with the parent. It is also highly recommended the observer have a system for observing and recording actual behaviors rather than only making conclusory notes that provide little more than overall impressions. One tool which helps is the Keys to Interactive Parenting Scale (KIPS), which is provided by its author along with training and certification for its use [68].

Criticisms of Reports. Budd and colleagues [59, 69, 70] have taken the lead in researching parenting evaluations in child maltreatment cases although their work is largely confined to looking at practices in Cook County (Chicago), Illinois. They have noted several limitations of reports in these cases: failure to clarify and/or note in report the referral question; failure to document informed consent and discussion of limitations of confidentiality; failure to address the believability of the results; and failure to comment on limitations of the methods used in the evaluation and the conclusions provided. Oberlander [71] added a criticism that reports (and conclusions) relied heavily on personal clinical experience and prior courtroom experiences of the evaluator and lacked ties to research findings and professional literature. She also noted reports often failed to protect the rights of parents who were potential defendants as cases developed. Jacobsen *et al.* [72] noted that reports failed to address minimal rather than optimal level of parenting and failure to address cultural practices in discussion of parenting abilities and deficits.

Recommended improvements for parenting assessment reports include the following: clarifying the referral question; articulating parental strengths and available resources as well as problems and risks; using valid and reliable assessment tools but taking a conservative approach by acknowledging limitations of assessment tools and current knowledge of outcomes of differing parenting practices; focusing on direct assessment of parenting skills, preferably in a home setting; maintaining open inquiry which continues to pursue and incorporate information from diverse sources; addressing issues of cultural context and potential observer bias and misinterpretations; and provide information for the decision makers rather than addressing ultimate issue question as psychologists (and other evaluators) have no specialized knowledge or training qualifying them for such conclusions [59, 72].

Current Practices in Assessment of Parenting

Most of this information is taken from research conducted about parenting assessment in the context

of custody evaluations as there is more published literature about these evaluations. Bow [73] reviewed results of five major published surveys of custody evaluators. These studies were originally published between 1986 and 2004 and covered in excess of 500 evaluators in the United States and Canada. In conducting the evaluations, most time was spent interviewing the parents with testing, review of records, and parent-child observations taking time in descending order. The MMPI-2 and the Millon Clinical Multiaxial Inventory (MCMI) were the most widely used tests with trends across time showing consistent use of Rorschach Inkblots (about 40%), declining use of intelligence tests, and increasing use of Parenting Stress Inventory (PSI) and Parent–Child Relationship Inventory (PCRI).

A more recent survey by Bow [74] of 89 psychologists who conducted custody evaluations (averaging one-third of their practice) indicated the primary use of testing was to rule out presence of psychopathology; other primary uses were assessment of personality functioning and analysis of parental strengths and weaknesses. The MMPI-2, typically used by 91% of the respondents, was the most widely used test instrument followed by the MCMI (typically used by 58%; no version specified) and the PSI (27%). The Personality Assessment Inventory (PAI) was typically used by 18% of the group. The Child Abuse Potential Inventory (CAPI), not designed for use in such evaluations, and PCRI were listed without percentage of use by respondents who as a group failed to have any clear opinion regarding whether these two instruments met Daubert [63] criteria for admissibility in court (*see* **Psychological Testing**).

Budd's report [69] on Illinois parenting evaluations conducted between 1995 and 1997 contained a subset of 124 evaluations likely conducted by psychologists but for a variety of referral questions. Projective personality tests were used in 81% of these evaluations, cognitive tests in 73%, and objective personality tests in 64%. Parenting questionnaires were used in only 4%, and parent–child observation was present in only 17%.

Summary

There are no specific forensic assessment instruments which have been developed for child protection cases, and there is no consensus as to what constitutes minimally adequate parenting, much less how to measure such a level. The role of the examiner is to collect and interpret data and to present the results to decision makers who combine those results with other factors such as legal standards, case laws, and community mores and practices. General principles of assessment apply such as clarifying. The referral question, informing the examinee of the nature of the evaluation and limits of confidentiality, using appropriate assessment methods and acknowledging limitations of the tools, and use of multiple sources of information with particular attention to written records and collateral source material. Use of clinical personality tests, parenting inventories, and parent–child observation is recommended although there is caution about using clinical assessment instruments for forensic purposes [75].

Other Resources

Professional organizations have published guidelines offering recommendations about conducting assessment in parent–child forensic contexts [57, 76–79]. White [50] has an excellent literature review; she and Otto and Edens [4] offer critical information about specific test instruments. Condie [2, 12], Gould and Martindale [56] and Dyer [80] provide very useful resources for exploring issues of parenting assessment. Finally, Heilbrun [81] presents an example of a report on a forensic parenting evaluation (*see also* **Parental Rights and Prerogatives**; **Visitation Rights**; **Capacity Assessment**; **Psychological Testing**).

References

[1] Kalichman, S. (1999). *Mandated Reporting of Suspected Child Abuse: Ethics, Law, and Policy*, 2nd Edition, American Psychological Association, Washington, DC.

[2] Condie, L.O. & Condie, D. (2007). Termination of Parental Rights, in *Forensic Psychology: Emerging Topics and Expanding Roles*, A.M. Goldstein, ed, John Wiley & Sons, Hoboken, pp. 294–330.

[3] Hess, K. & Brinson, P. (1999). Mediating Domestic Law Issues, in *The Handbook of Forensic Psychology*, A. Hess & I. Weiner, eds, Wiley, New York, pp. 63–103.

[4] Otto, R. & Edens, J. (2003). Parenting Capacity, in *Evaluating Competencies: Forensic Assessments and Instruments*, 2nd Edition, T. Grisso, ed, Kluwer, New York, pp. 229–307.

[5] Kempe, C., Silverman, F., Steele, B., Droegemueller, W. & Silver, H. (1962). The battered child

syndrome, *Journal of American Medical Association* **187**, 105–112.
[6] Child Abuse Prevention Act. Pub L No 93–247, US Code 93rd Congress (1974).
[7] Child Welfare League of America (2007). National Fact Sheet Retrieved 10/12/07 From http://www.cwla.org/advocacy/nationalfactsheet07.htm.
[8] Lamphear, V. (1985). The impact of maltreatment on children's psychosocial adjustment: A review of the research, *Child Abuse & Neglect* **9**, 251–263.
[9] American Academy of Pediatrics Committee on Child Abuse and Neglect (1993). Shaken baby syndrome: inflicted cerebral trauma, *Pediatrics*, **92**, 872–875.
[10] Duhaime, A., Alario, A., Lewander, W., Schut, L., Sutton, L., Seidl, T., Nudelman, S., Budenz, D., Hertle, R., Tsiaras, W. & Loporchio, S. (1992). Head injury in very young children: mechanisms, injury types, and ophthalmologic findings in 100 hospital patients younger than 2 years of age, *Pediatrics* **90**, 179–185.
[11] Miller, L. (1999). Child abuse brain injury: clinical, neuropsychological, and forensic considerations, *Journal of Cognitive Rehabilitation* **17**, 10–19.
[12] Condie, L. (2003). *Parenting Evaluations for the Court: Care and Protection Matters*, Kluwer, New York.
[13] Kolko, D. (2002). Child Physical Abuse, in *The APSAC Handbook on Child Maltreatment*, 2nd Edition, J.E.B. Myers, L. Berlinger, J. Briere, T. Hendrix, C. Jenny & T. Reid, eds, Sage, Thousand Oaks, pp. 21–54.
[14] Berlinger, L. & Elliott, D. (2002). Sexual Abuse of Children, in *The APSAC Handbook on Child Maltreatment*, 2nd Edition, J.E.B. Myers, L. Berlinger, J. Briere, T. Hendrix, C. Jenny & T. Reid, eds, Sage, Thousand Oaks, pp. 55–78.
[15] Finkelhor, D. (2002). Introduction, in *The APSAC Handbook on Child Maltreatment*, 2nd Edition, J.E.B. Myers, L. Berlinger, J. Briere, T. Hendrix, C. Jenny & T. Reid, eds, Sage, Thousand Oaks, pp. xi–xvi.
[16] Erickson, M. & England, B. (2002). Child Neglect, in *The APSAC Handbook on Child Maltreatment*, 2nd Edition, J.E.B. Myers, L. Berlinger, J. Briere, T. Hendrix, C. Jenny & T. Reid, eds, Sage, Thousand Oaks, pp. 3–20.
[17] Conte, J. & Schuerman, J. (1987). Factors associated with an increased impact of child sexual abuse, *Child Abuse & Neglect* **11**, 201–211.
[18] Myers, J. (1992). *Legal Issues in Child Abuse and Neglect*, Sage, Newbury Park.
[19] Edgeworth, J. & Carr, A. (2000). Child Abuse, in *What Works for Children and Adolescents? A Critical Review of Psychological Interventions with Children, Adolescents, and Their Families*, A. Carr, ed, Routledge, New York, pp. 17–48.
[20] Garbarino, J., Kostelny, K. & Dubrow, N. (1991). What children can tell us about living in danger, *American Psychologist* **46**, 376–383.
[21] Azar, S., Povilaitis, T., Lauretti, A. & Pouquette, C. (1998). The Current State of Ecological Theories in Intrafamilial Child Maltreatment, in *Handbook of Child Abuse Research and Treatment*, J. Lutzker, ed, Plenum, New York, pp. 3–30.
[22] Caine, N. & Reite, M. (1983). Infant Abuse in Captive Pig-Tailed Macaques: Relevance to Human Child Abuse, in *Child Abuse: The Non-Human Data*, M. Reite & N. Caine, eds, Alan Liss, New York, pp. 19–27.
[23] Belsky, J. (1993). Etiology of child maltreatment a developmental-ecological analysis, *Psychological Bulletin* **114**, 413–434.
[24] Widom, C. (1989). Does violence beget violence? A critical examination of the literature, *Psychological Bulletin* **106**, 3–28.
[25] Chaffin, M., Kelleher, K. & Hollenberg, J. (1996). Onset of physical abuse and neglect: psychiatric, substance abuse, and social risk factors from prospective community data, *Child Abuse & Neglect* **20**, 191–203.
[26] Crittenden, P. (1993). An information processing perspective on the behavior of neglectful parents, *Criminal Justice and Behavior* **20**, 27–48.
[27] Dix, T. (1991). The affective organization of parenting: adaptive and maladaptive processes, *Psychological Bulletin* **110**, 3–25.
[28] Azar, S. (1997). A Cognitive Behavioral Approach to Understanding and Treating Parents Who Physically Abuse Their Children, in *Child Abuse: New Directions in Prevention Across the Lifespan*, A. Wolfe, R. Mc.Mahon & R. Peters, eds, Sage: Thousand Oaks, pp. 79–101.
[29] Friedrich, W. (1990). *Psychotherapy of Sexually Abused Children and Their Families*, Norton, New York.
[30] Milner, J. & Dopke, C. (1997). Child Physical Abuse: Review of Offender Characteristics, in *Child Abuse: New Directions in Prevention Across the Lifespan*, D. Wolfe, R. Mc.Mahon & R. Peters, eds, Sage, Thousand Oaks, pp. 27–54.
[31] Piante, R., England, B. & Erickson, M. (1989). The Antecedents of Maltreatment: Results of the Mother-Child Interaction Research Project, in *Child Maltreatment: Theory and Research on the Causes and Consequences of Child Abuse and Neglect*, D. Cicchetti & V. Carlson, eds, University Press, Cambridge, pp. 203–253.
[32] Daro, D. (1998). *Confronting Child Abuse: Research for Effective Program Design*, Free Press, New York.
[33] Cohn, A. & Daro, D. (1987). Is treatment too late? what ten years of evaluation results tell us, *Child Abuse & Neglect* **11**, 433–442.
[34] Oates, K. & Bross, D. (1995). What have we learned about treating child physical abuse? a literature review of the last decade, *Child Abuse & Neglect* **19**, 463–473.
[35] Lutzker, J. (1984). Project 12-Ways: Treating Child Abuse and Neglect From an Ecobehavioral Perspective, in *Parent Training*, R. Dangel, & R. Polster, eds, Guilford New York, pp. 260–297.
[36] Wesch, D. & Lutzker, J. (1991). A comprehensive 5 year evaluation of Project 12-Ways: an ecological program for treating and preventing child abuse and neglect, *Journal of Family Violence* **6**, 17–35.

[37] Lutzker, J. & Campbell, R. (1994). *Ecobehavioral Family Interventions in Developmental Disabilities*. Brooks/Cole, Pacific Grove.

[38] Garner, B. (ed) (1999). *Black's Law Dictionary*, 7th Edition, West Group, St. Paul.

[39] Griswold v. Connecticutt, 381 U.S. 479 (1965).

[40] Parham v. J.L. and J.R., 422 U.S. 584 (1979).

[41] Ingraham v. Wright, 430 U.S. 651 (1977).

[42] DeShaney v. Winnebago County Department of Social Services, 486 U.S. 549 (1990).

[43] Santosky v. Kramer, 455 U.S. 745 (1982).

[44] South Carolina Department of Social Services v. The Father and the Mother; In the Interest of the Child, 294 S.C. 518, 366 S.E. 2d 40 (App.1988).

[45] State v. Evans, 171 Wis.2d 471, 492 N.W. 2d 141 (1992).

[46] In re Glenn G., 154 Misc. 2d 677, 587 NY.S. 2d 464 (N.Y.Fam.Ct. 1992).

[47] E.C. v. District of Columbia, 589 A. 2d 1245 (D.C.App. 1991).

[48] Egly v. Blackford County Department of Public Welfare, 592 N.E. 2d 1232 (Ind. 1992).

[49] Matter of Joshua O, 641 N.Y.S.2d 475 (1996).

[50] White, A. *Assessment of Parenting Capacity: Literature Review*, NSW Department of Community Services, Ashfield, New South Wales (Australia) Retrievable from http://www.community.nsw.gov.au.

[51] Halikias, W. (1994). Forensic family evaluations: a comprehensive model for professional practice, *Journal of Clinical Psychology* **50**, 951–964.

[52] Grisso, T. (2003). *Evaluating Competencies: Forensic Assessments and Instruments*. 2nd Edition, Kluwer, New York.

[53] Reder P., Duncan S. & Lucey C. (2003). What Principles Guide Parenting Assessments? in *Studies in the Assessment of Parenting* P. Reder, S. Duncan & C. Lucey, eds, Brunner- Routledge, New York, pp. 3–26.

[54] Ayoub C. & Kinscherff R. (2006). Forensic Assessment of Parenting in Child Abuse and Neglect Cases, in *Forensic Mental Health Assessment of Children and Adolescents*, S.N. Sparta & G.P. Koocher, eds, Oxford, New York, pp. 350–335.

[55] Basic Behavioral Science Task Force of the National Advisory Mental Health Council (1996). Basic behavioral science research for mental health: family processes and social networks, *American Psychologist* **51**, 622–630.

[56] Gould, J.W. & Martindale, D.A. (2007). *The Art and Science of Child Custody Evaluations*, Guilford, New York.

[57] Committee on Professional Practice and Standards, American Psychology Association Board of Professional Affairs (1999). Guidelines for psychological evaluations in child protection matters, *American Psychologist* **54**, 586–593.

[58] Brodzinsky, P. (1993). On the use and misuse of psychological testing in child custody evaluations, *Professional Psychology: Research and Practice* **24**, 213–219.

[59] Budd, K.S. & Holdsworth, M.J. (1996). Issues in clinical assessment of minimal parenting competence, *Journal of Clinical Child Psychology* **25**, 2–14.

[60] Andrews, P. & Meyer, R. (2003). Marlowe-Crowne personal desirability scale and short form C: forensic norms, *Journal of Clinical Psychology* **59**, 483–492.

[61] Carr, G.D., Moretti, M.H. & Cue, B.J.M. (2005). Evaluating parenting capacity: validity problems with the MMPI-2, PAI, CAPI, and ratings of child adjustment, *Professional Psychology: Research and Practice* **36**, 188–196.

[62] Heilbrun, K. (1995). Child custody evaluation: critically assessing mental health experts and psychological tests, *Family Law Quarterly* **29**, 63–78.

[63] Daubert v. Merrell Dow Pharmaceutical, Inc., 509 U.S. 579 (1993).

[64] Milchman, M.S. (2000). Mental health experts' common error in assessing bonding in guardianship cases, *The Journal of Psychiatry and Law* **28**, 351–378.

[65] Starr, R. (1987). Clinical judgment of abuse-proneness based on parent-child interactions, *Child Abuse & Neglect* **11**, 87–92.

[66] Martindale, D.A. & Gould, J.W. (2004). The forensic model: ethics and scientific methodology applied to custody evaluation, *Journal of Child Custody* **1**(2), 1–22.

[67] Acklin, M.W. & Cho-Stutler, L. (2006). The science and art of parent-child observation in child custody evaluation, *Journal of Forensic Psychology* **6**, 51–62.

[68] Comfort, M. & Gordon, P.R. (2006). Keys to interactive parenting scale (KIPS): a practical assessment of parenting behavior, *NHSA Dialogue: A Research-to-Practice Journal for the Early Intervention Field* **9**, 22–48.

[69] Budd, K.S., Poindexter, L.M., Felix, E.D. & Naik-Polan, A. (2001). Clinical assessment of parents in child protection cases: an empirical analysis, *Law and Human Behavior* **25**, 93–108.

[70] Budd, K.S., Felix, E.D., Sweet, S.C., Saul, A. & Carleton, R.A. (2006). Evaluating parents in child protection decisions: an innovative court-based clinic model, *Professional Psychology: Research and Practice* **37**, 666–675.

[71] Oberlander, L. (1995). Psychological issues in child sexual abuse evaluations: a survey of forensic mental health professionals, *Child Abuse & Neglect* **19**, 474–489.

[72] Jacobsen, T., Miller, L. & Kirkwood, K. (1997). Assessing parenting competency in individuals with severe mental illness: A comprehensive service, *Journal of Mental Health Administration* **24**, 189–199.

[73] Bow, J.M. (2006). Review of empirical research on child custody practice, *Journal of Child Custody* **3**, 23–50.

[74] Bow, J.M., Gould, J.W., Flens, J.R. & Greenhut, D. (2006). Testing in child custody evaluations – selection, usage, and Daubert admissibility: a survey of psychologists, *Journal of Forensic Psychology Practice* **6**, 17–38.

[75] Archer, R.P. (2006). *Forensic Use of Clinical Assessment Instruments*, Lawrence Erlbaum, Mahwah.

[76] Committee on Ethical Guidelines for Forensic Psychologists (1991). Specialty guidelines for forensic psychologists, *Law and Human Behavior* **15**, 655–665.
[77] American Psychological Association (1994). Guidelines for child custody evaluations in divorce proceedings, *American Psychologist* **49**, 677–680.
[78] Association of Family and Conciliation Courts (2006). *Model Standards of Practice for Child Custody Evaluation*. Retrieved 2/05/06 from http://www.afccnet.org/resources/resources_model_child.asp.
[79] American Psychological Association (2002). Ethical principles of psychologists and code of conduct, *American Psychologist* **57**, 1060–1073.
[80] Dyer, F.J. (1999). *Psychological Consultation in Parental Rights Cases*, Guilford, New York.
[81] Heilbrun, K., Marczuk, G.R. & DeMatto, D. (2002). *Forensic Mental Health Assessment: A Casebook*, Oxford, New York.

PAUL ANDREWS

Parricide *see* Battered Child Syndrome

Particles: Form

Forensic trace evidence includes small fragments of material (particles) that can be used to assist with an investigation into crimes and accidents [1]. Particles are ubiquitous – cosmetics, pollen, fibers, glass, and minerals to name a few. There are usually many thousands of particles virtually on all exposed surfaces. They can originate from distant sources and be transported through air, sea, or the surfaces of animals, including people. Particles adhere to vehicles, tools, clothes, and objects creating a signature of particles with an almost infinite variety. The potential to relate where something was manufactured, where it has been, and who could have been associated with it is sometimes possible with critical examination. Forensic scientists analyze particles using microscopy and instrumental analysis to characterize, identify, and associate materials; to determine provenance, and link people, places and objects to assist the judicial system (*see* **Microscopy: Light Microscopes**).

Particles with considerable diversity of form can originate from botanical materials, such as pollen, spores, diatoms, seeds, plant hairs, grass including *Cannabis sativa* L., leaves, needles, rootlets, and phytoliths. Botanical materials are cultivated, altered, and consumed by man and animals as food. From Wonder Bread to Wheaties, from the corn in corn dogs, to the seeds in feces, small particles abound in the plant kingdom. Animals also have or produce an array of particles such as hairs, feathers, eggs, skin cells, teeth, scales, and processed body parts. Natural inorganic particles are ubiquitous in a multitude of types from minerals and rocks, microfossils, deposits evaporated from water, and cosmic particles such as micrometeorites and tektites.

Traces formed by the interaction between man and the environment have increased the range of particles. Botanically derived particles can originate from foodstuffs and vomit. Man-made particles can be derived from the breakdown and disintegration of building materials such as concrete and wallboard. Vehicles contribute metals, fibers, paint, polymers, tire rubber, glass, lubricants, and fuel. Clothing materials often shed fibers. Wood and paper products can form chips, dust, and pulp fibers. Man-made particles can be formed from metals, gunshot residue (GSR), explosives, tapes, putties, corrosion, drugs, fertilizers, insecticides, poisons, toxins, and cosmetics. Particles can be composites from multiple sources such as dust, paint chips, chemicals, and combustion products.

Particle size and shape are fundamental properties that can provide important information in identification, comparison, and provenance. The Particle Atlas [2, 3] uses an ingenious six-digit code and binary sum to classify particles in the process of identifying unknowns. There are six basic classification characteristics in the code: transparency, color, isotropy/anisotropy, refractive index, shape (first characteristic), and shape (second characteristic). The system uses either a "0" or a "1" to signify the absence (0) or presence (1) of the six different particle characteristics. Specifically, the fifth and sixth digits are used to describe a particle's shape or form:

1 1 – elongated and flattened (ribbon, blade, and lath)
1 0 – flattened but not elongated (plate or tablet)

0 1 – elongated but not flattened (needle or rod)
0 0 – neither elongated nor flattened (equant particle).

Particles can exist in crystalline or noncrystalline form. The study and characterization of crystals is known as *crystallography*. Solid particles form into one of six "crystal systems". A crystal system is a category of space groups, which characterize symmetry of structures in three dimensions in three directions, having a discrete class of point groups. The six crystal systems are cubic (isometric), hexagonal, tetragonal, orthorhombic, monoclinic, and triclinic. Crystalline particles can form on any scale, from large particles visible to the naked eye, to submicron crystalline particles that can only be examined microscopically. Cubic crystals are in the shape of a perfect cube. The crystallographic axes used in this system are of equal length and are mutually perpendicular, occurring at right angles to one another. A grain of common table salt is a good example of a cubic crystal. Examples of minerals which crystallize in the cubic system are halite, magnetite, and garnet. Minerals of this system tend to produce crystals of equidimensional or equant habit. A crystal form is a collection of equivalent crystal faces related to each other by mineral symmetry [4].

Trace evidence particles are collected and initially examined by stereomicroscopy at low magnification from 4× to 100×. During this screening process, particle form is often the first characteristic that is recognized. Forms, shapes, structures, angles, etc., between things can be obviously different. The same is true in the examination of particles in trace evidence, even on the microscopic scale.

Pollen

The study of pollen, spores, and other acid resistant microscopic plant particles is called *palynology* and is highly useful in paleoecology, paleontology, archaeology, and forensics. Pollen grains are microscopic in size (10–100 μm), occur abundantly in soil and dust on any exposed surface, are resistant to decay, and as different from each other as the plants that produce them. Pollen assemblages found in soil and water bodies at a crime scene often reflect the vegetation and environment of the area. Pollen grains come in a wide variety of shapes, sizes, and surface markings (Figure 1). The pollen grain has three walls

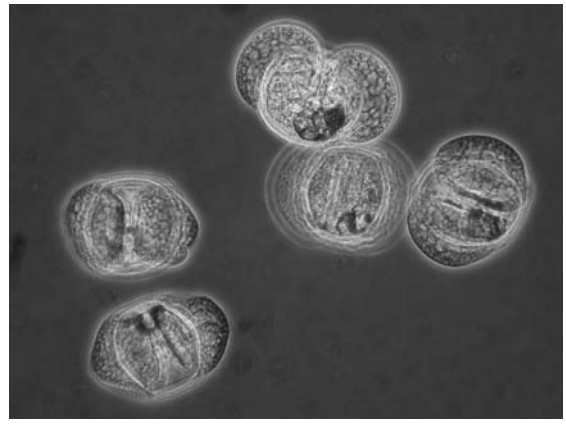


Figure 1 Pollen from *Pinus taeda*, loblolly pine. Phase contrast, original magnification 400×

or layers. The outer wall is most often sculptured, covered with any combination of spines (echinae), pores, ridges (rugulae), or bumps (scabrae) that help in identifying the grain. The middle layer contains the enzymes necessary for plant recognition in pollination. The third layer is a bottom wall and sometimes has columns rising up to the outer wall. Palynologists have classified pollen using two principal morphological features: pori (pores) and colpi (furrows). Pollen grains are divided into groups by the number, position, and shape of their pori and colpi. Pollen grains having only pores are called *porate*, those with only colpi are called *colpate*, and those with both are called *colporate*. The number of pores and colpi is denoted by attaching prefixes such as mono-, si-, terta-, and penta-. Further divisions are made by an examination of the fine structure and patterns observed on the outer layer. Descriptive terminology such as gemmate, pilate, regulate, scabrate, etc., are used [1] (*see* **Palynology**).

Diatoms

Diatoms are unicellular algae that live in any water body, large or small, from fresh to salt and brackish types. A characteristic feature of diatom cells is that they are encased within a unique cell wall made of silica (hydrated silicon dioxide) called a *frustule* or *test* (Figure 2). These frustules show a wide diversity in form, some very beautiful and ornate, usually consisting of two asymmetrical sides. The identification

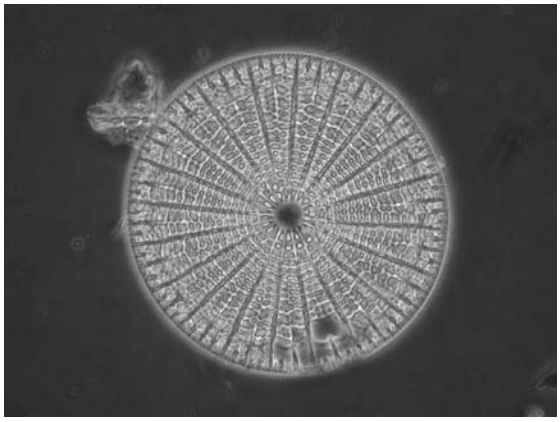


Figure 2 *Arachnoidiscus,* a diatom from San Pedro, California. Phase contrast, original magnification 400×

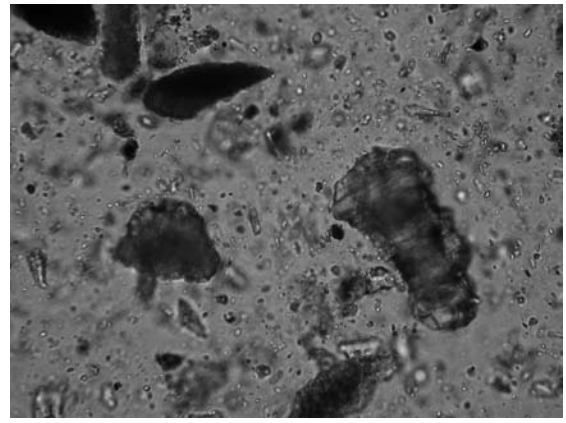


Figure 3 Malachite. An artist's paint pigment, naturally ground. Original magnification 400×

of diatoms is based on the morphology of the test. Two major groups of diatoms are generally recognized: the centric diatoms exhibiting radial symmetry (symmetry about a point) and the pennate diatoms with bilaterally symmetrical (symmetry about a line). Another particle with some similarities to diatoms are radiolarian, single-celled plankton in the kingdom Protista, which secrete silicate exoskeletons. They have lots of spines radiating outward. Seafloor deposits that are formed from radiolarian are called *radiolarian ooze*. This ooze forms a sedimentary rock known as *chert*. Decomposition and decay of diatoms may form diatomaceous earth, which can be mined and processed for use in a variety of products including paint, insulation, abrasives, and filtering agents (*see* **Diatoms**).

Paint Pigment

Paint pigment particles range from ground minerals used since antiquity to processed precipitated types used in the automotive industry. Pigments are often crystalline and tend to influence their external shape. Lead white forms as 1–50 μm hexagonal platelets. Zinc white forms as less than 2-μm rounded particles, many with spiked arms that look like children's jacks using 1000× oil immersion microscopy. Some crystalline pigments are too finely divided to show crystallinity. Titanium dioxide pigment, probably one of the most common, is less than 1 μm in size and is subrounded in shape. A few pigments are noncrystalline glasses and show conchoidal glassy fracture such as cobalt blue and gamboge. The shape of a particle tells how it was prepared, such as pigment formed by precipitation from a solution or vapor as in zinc oxide. Others are simply ground minerals like quartz, anhydrite, and malachite (Figure 3). The particle size of pigments is usually helpful in distinguishing between synthetic and natural varieties. Synthetic zinc white, titanium oxide, cadmium red, and lamp black are all up to about 1 μm in size. Natural quartz, gypsum, azurite, and ultramarine are much coarser, usually 5–10 μm (*see* **Paint: Interpretation**).

Paint chip particles may show a layer structure in cross section, originating from successive paint applications over time. The layered paint chip may also originate from an automobile having anywhere from three to six original paint layers. Paint chips having embedded reflective glass beads originate from highway fog marker/centerline paint. Still other paint particles may contain remnants of underlying surfaces such as plaster or wood. Microscopic rounded paint beads formed during aerosol deposition via spray painting have been found as associative physical evidence in many homicides including the Green River murders in Washington State.

Mineral and Rock Fragments

Minerals in the size range of fine sand are often examined in forensic soil cases. In the microscopic examination of mineral form, the shape of grains

can be described in two categories: thin sections (a mineralogical specimen that has been glued to a microscope slide and ground and polished down to approximately 25 μm, thus allowing examination by transmitted polarized light microscopy) and individual detrital grains mounted in immersion liquids. Grains with well-formed crystal faces are termed *euhedral*. Grains without crystal faces are called *anhedral*. In glassy rocks where crystallization abruptly stops, minerals may show crystallites and microlites [5]. The term *crystal habit* describes the favored growth pattern of the crystals of a mineral. A crystals habit may show little relation to the form of a single, perfect crystal of the same mineral, which would be classified according to crystal system. Crystal habit is often useful in identification. Mineralogists and microchemists use many descriptive terms to describe form such as blocky or equant, tabular, lamellar, flaky, micaceous, elongated, columnar, prismatic, bladed, acicular, platy, scaly, granular, radiating, foliated, felted, and fibrous. Many minerals have certain crystallographic planes where chemical bonding is weaker. These planes of weakness along which a mineral may break are called *cleavage planes*. When a mineral is broken or crushed it may also break along the fracture surfaces, which are unrelated to the mineral's crystal structure. One common crystal observed in soil derived from igneous and metamorphic rocks is zircon. Euhedral zircon crystals, forming in the tetragonal crystal system, are quite distinctive (Figure 4). The surface textures of

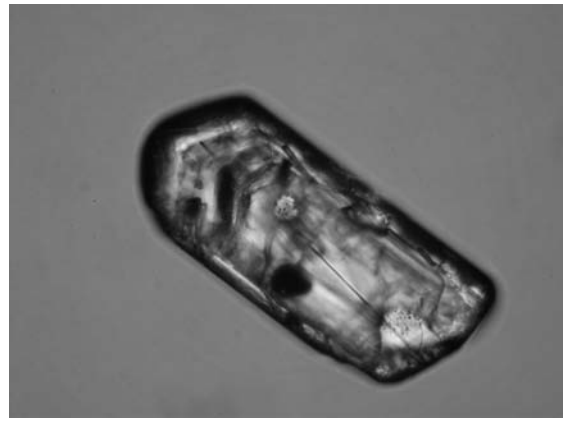


Figure 4 A crystalline zircon crystal. Original magnification 200×

sand grain and microfeatures observed by scanning electron microscopy can be used to link them to specific environments and geologic events (Figure 5) [6] (*see* **Soil: Forensic Analysis**).

Glass

The American Society of Testing Materials (ASTM) describes glass as an inorganic product of fusion that has cooled to a rigid condition without crystallizing [1]. Variables in physical properties of glass include thickness, curvature, manufacture marks, presence of mirror glass, and recognition of finely crushed

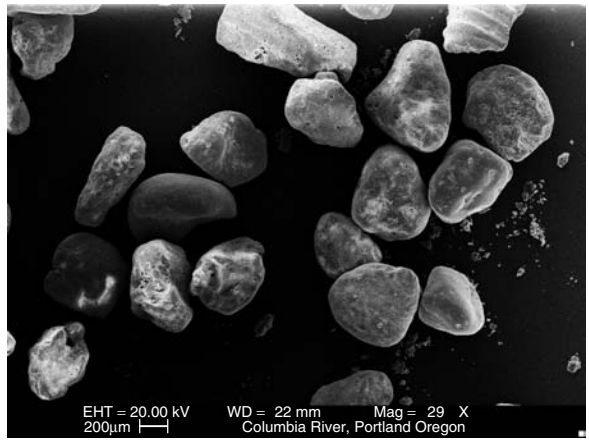


Figure 5 SEM photomicrograph of beach sand along the Columbia River near Portland, Oregon. Original magnification 29×



Figure 6 Glass particle showing conchodial fracture. Original magnification 40×

powdered glass originating from impact damage. Glass particles when broken show sharp, jagged edges known as *conchoidal fractures* (Figure 6). Tempered glass particles form "diced" pieces. Glass particles that have been annealed by coming into contact with a hot surface, such as a halogen lamp filament often have a molten smooth, beaded texture. Other glass particles such as pumice that originate from volcanism can show flow banding (Figure 7), glassy beads or glassy coatings on other mineral surfaces. Some glass particles have what are known as *hackle marks* on the edges, which are useful in interpreting the side of the glass where the force was applied when it broke. It have been reported that glass particles recovered as forensic evidence average from 0.35 to 1 mm in size, whereas fragments >5 mm are most likely to be easily lost [7] (*see* **Glass Evidence: Bayesian Approach to**).

Wood

Particles of wood can be identified by their relatively soft, porous, and fibrous texture. If the particle is large enough, the differentiation between softwood and hardwood can be made on the presence or absence of resin canals and vessel elements. The particle may be rough surfaced and torn, suggesting chainsaw wood chips or fine and powdery in form such as sawdust. Some particles may have attached bark and accessory decay particles. Softwood and hardwood fibers found in paper products can be identified microscopically based on fiber anatomy. Softwood fibers, called *tracheids*, are distinguished by bordered pits along their length (Figure 8). Ray cross-field pitting locations have very specific pits that are classified into groups: fenestriform, pinoid, taxodioid, cupressoid, and piceoid, characteristic of softwood species. Hardwood species are identified by

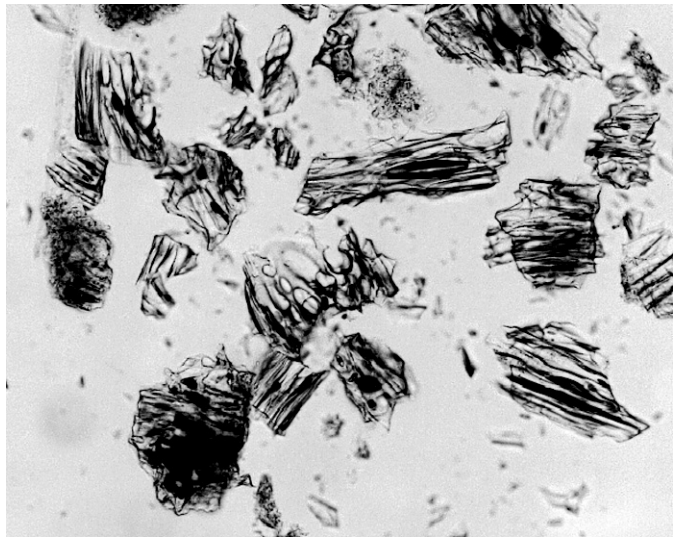


Figure 7 Pumice

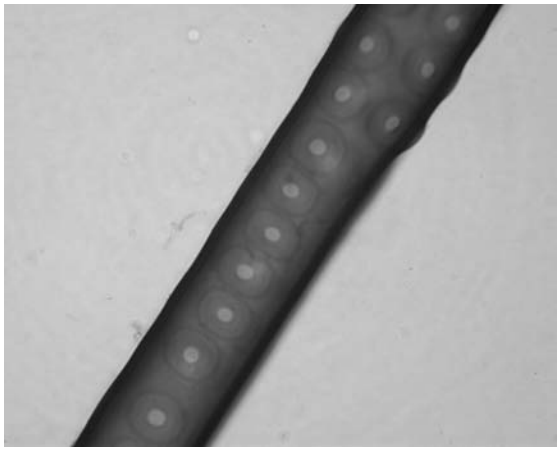


Figure 8 Softwood fiber from *Pinus Ponderosa*, Ponderosa Pine showing bordered pits. Original magnification 200×

their unique water-conducting structures called *vessel elements*. These tube-shaped structures separate out into discrete particles when the wood is pulped (*see* **Wood**).

Fibers

Fibers are often defined as particles having a high length to width ratio. Fibers are classified into groups: manufactured (nylon, polyester, etc.), natural plant (cotton, jute, hemp, etc.), and inorganic (asbestos and glass fibers). Physical characteristics of manufactured fibers include diameter, cross-sectional shape, crimp, striations, and damage. The cross-sectional shape of nonround fibers is often patented and very unique. Many types are extruded through a spinneret, a device having a series of engineered holes that a molten polymer travels through. The fibers leaving the spinneret have a cross-sectional shape matching the holes of the spinneret. Fibers such as nylon can have cross-sectional shapes ranging from round, delta, trilobal, and multichanneled just to name a few (Figure 9). Bicomponent fibers are comprised of two polymers of different chemical and/or physical properties extruded from the same spinneret with both polymers within the same filament. Most commercially available bicomponent fibers are configured in a sheath/core, side by side, or eccentric sheath/core arrangement. Natural plant fibers may be encountered

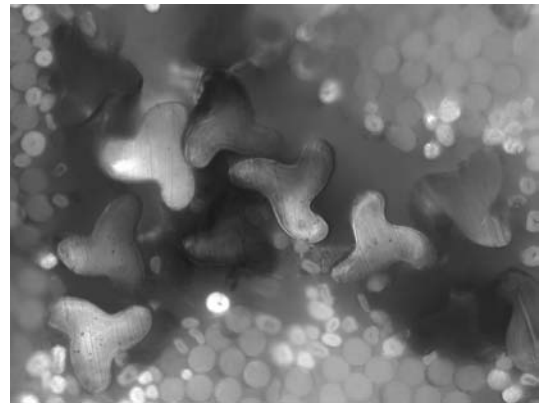


Figure 9 Trilobal nylon fibers encased in cotton backing fibers. Original magnification 200×

as the technical fiber (cordage, sacks, mats, etc.) or as individual cells (fabric and paper). Relative cell wall and lumen (canal in central shaft of fiber) thickness, cell length, and presence of surface markings, crystals, and twists are important characteristics to observe [8]. Mineral fibers in the asbestiform group range from chrysotile, having a soft, wavy texture, to the long, straight, inflexible fibers of amosite. Glass fibers can be placed into three catagories: fiberglass (continuous and noncontinuous), mineral wool (rock wool and slag wool having exotic twisty blob-like shapes), and refractory ceramic fibers that have exotic shapes, but are thinner in diameter compared to the others (*see* **Fibers**).

Hairs

Human hairs can be distinguished from other animal hairs by examining features such as the scale pattern, medulla, root, length, and shaft configurations (Figure 10) [5]. Specific terms used to describe human hair shaft form are straight, arced, wavy, curly, twisted, tightly coiled, and crimped. The cross-sectional shape in human hair can vary from rounded, oval, triangular, to flattened. The shaft can appear as buckled, convoluted, undulating, and split, [5]. The somatic origin of human hairs can often be determined by morphology (*see* **Hair: Microscopic Analysis**; **Hair: Animal**).

Many other trace evidence particles can present a variety of forms. Smokeless powder (commonly

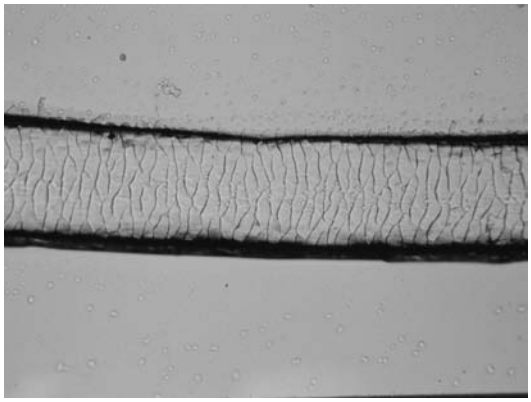


Figure 10 Caucasian head hair showing an imbricate scale pattern typical of man. Original magnification 200×

called *gunpowder*) can be manufactured in ball, flattened ball, disk, and tubular shapes. Databases have been developed, in which these morphological criteria are used to distinguish the different brands. GSR analysis relies on the identification of lead, barium, and antimony residues originating from the primer cap of cartridges. One criterion for identification is the minute (often less than 5 μm), rounded particle form characteristic of GSR.

References

[1] American Society for Testing and Materials (1965). *Standard Definitions of Terms Relating to Glass Products*, ASTM Standards, ASTM, Philadelphia, Part 13, p. 145.

[2] Mahaney, W.C. (2002). *Atlas of Sand Grain Surface Textures and Applications*, Oxford University Press.

[3] McCrone, W.C. & Delly, J.G. (1992). *The Particle Atlas*, 2nd Edition, Vol. 4, Ann Arbor Science.

[4] McCrone, W.C. (1982). The microscopical identification of artists' pigments, *Journal of the International Institute for Conservation* **7**(1 & 2), 11–34.

[5] Forensic Science Communications (2005). *Forensic Human Hair Examination Guidelines*, Scientific Working Group on Materials Analysis (SWGMAT).

[6] Heinrich, E.W.M. (1965). *Microscopic Identification of Minerals*, McGraw-Hill.

[7] Almirall, J.R. & Trejos, T. (2006). Advances in the forensic analysis of glass fragments with a focus on refractive index and elemental analysis, *Forensic Science Review* **18**(2) 74–95.

[8] Forensic Science Communications (1999). *Forensic Fiber Examination Guidelines*, Scientific Working Group on Materials Analysis (SWGMAT).

Further Reading

Bisbing, R.E. & Schneck, W.M. (2006). Particle analysis in forensic science, *Forensic Science Review* **18**(2) 119–144.

McCrone *McCrone Atlas of Microscopic Particles*, http://www.mccroneatlas.com.

Neese, W.D. (2000). *Introduction to Mineralogy*, Oxford University Press.

WILLIAM SCHNECK

Paternity *see* Missing Persons and Paternity: DNA

Pattern: Fire *see* Fire: Dynamics and Pattern Production

Patty Hearst *see* Stockholm Syndrome

Peak Height: DNA

Peak Height

The height of a peak on an electrophoretogram is measured in relative fluorescent units (rfu) and is a measure of the intensity of light detected by a DNA analyzer as sample is passed through a column. It is roughly proportional to the amount of DNA present, and is expected to be within a

certain range of values for standard analyses. For example, using the protocol validated by Applied Biosystems Inc. for the SGM Plus kit, 1–2.5 ng of input DNA will give peaks with heights in the range 150–5000 rfu.

As with any technology for measurement, there is a baseline "noise" that is recorded in the absence of sample, and is related to the sensitivity of the detection machine. There is some debate about the threshold value above which a peak can be declared as a "real" peak that represents a piece of DNA, as opposed to chance occurrence of noise of sufficient intensity to appear as a peak. Different laboratories use different thresholds, decisions that may be based on objective principle, scientific validation, or rule-of-thumb experience.

Two alleles of a heterozygotic locus will give heights that are roughly equal in height, and for the same amount of DNA under the same conditions, a homozygote would produce one peak that is roughly twice the height of the heterozygote peaks. Generally, alleles of smaller size (fewer repeats) (*see also* **Short Tandem Repeats**) amplify better than larger ones and so will have a slightly higher peak. Owing to variation in the analytical chemistry, this correlation is not absolute, but two peaks of a heterozygote are almost always more than 60% in proportion, smaller : larger or larger : smaller, and such peaks are said to be balanced. The loci that have been designed to give smaller (molecule length) PCR products, e.g., D3 *versus* D2 in the SGM Plus kit, will also often produce higher peaks than for the larger loci. This effect may be particularly obvious if the sample of DNA is partially degraded and is due to an increased chance of larger DNA molecules being degraded (and so unavailable for PCR amplification) rather than smaller molecules.

Peak heights may be used to interpret the DNA being analyzed. For example, when two peaks being compared are not above the 60% ratio, this is called peak height imbalance. When imbalance is seen, it is usually an indication that the peaks in question come from DNA that is partly or entirely from different contributing sources, i.e., there is a mixture of contributors. This information is used in conjunction with the number of peaks seen at each locus of a profile to identify the presence of a mixture. When a mixture of DNA sources exists, peaks will be seen that are roughly in proportion to their relative contributing amounts. Where alleles are common to more than one contributor, the rfu for each allele will be added to give a combined higher peak. Looking at the peak heights of a profile at each locus and all loci together, it may be possible to interpret all or part of at least one of the contributors. This is true especially for what are termed major/minor mixtures where there is a much greater proportion of one source compared with another. The major profile will be more completely decipherable than the minor one since some of the minor peaks may be masked by overlapping with peaks from shared alleles of the major profile. Mixtures of DNA in less extreme proportions are much more difficult to interpret objectively, assisted only when there is a reasonable expectation of DNA from one of the sources.

For a discussion of the use of peaks in mixture interpretation, see Gill *et al.* [1] and Clayton *et al.* [2]. Peak values in LCN analysis are discussed in Whitaker *et al.* [3].

References

[1] Gill, P., Sparkes, R., Pinchin, R., Clayton, T., Whitaker, J. & Buckleton, J. (1998). Interpreting simple STR mixtures using allele peak areas, *Forensic Science International* **91**, 41–53.

[2] Clayton, T.M., Whitaker, J.P., Sparkes, R. & Gill, P. (1998). Analysis and interpretation of mixed forensic stains using DNA STR profiling, *Forensic Science International* **91**, 55–70.

[3] Whitaker, J.P., Cotton, E.A. & Gill, P. (2001). A comparison of the characteristics of profiles produced with the AMPFlSTR SGM Plus multiplex system for both standard and low copy number (LCN) STR DNA analysis, *Forensic Science International* **123**, 215–223.

Further Reading

Buckleton, J., Triggs C.M. & Walsh, S.J. (2005). *Forensic DNA Evidence Interpretation*, CRC Press.

Butler, J.M. (2005). *Forensic DNA Typing*, 2nd Edition, Elsevier.

SCOTT BADER

Pedophilia *see* Child Sexual Abuse

Peer Review as Affecting Opinion Evidence

Obtaining the Approval of One's Peers in Science

Peer review is a term that, in scientific, publishing, and other circles, has a number of different meanings. The principal definition, however, describes a method whereby the appropriateness of research project results, as well as of articles, papers, and books based thereon, have been judged as worthy contributions to the scientific or professional field in which they belong. In some academic fields, peer review is referred to as *refereeing* [1–4]. No matter what the process or the end result is, the material under review is a document.

The purposes of peer review are to determine which research proposals will receive grants or monetary support to allow the research to go forth; whether an article of book is worthy to be published; and to evaluate the comparative ranking of researchers, scientists, or academics for advancement, remuneration, professional self-improvement purposes, as well as for comparative ranking among colleagues. Many journals publish guidelines for reviewers and these will usually explain the journal's purpose in using peer review. Not all materials published in peer reviewed journals are actually peer reviewed. For example, letters and correspondence may be peer reviewed in some journals but not in others.

The customary peer review process is applied to matters that include, for example, the dissemination of research results, an article or book on a method or topic. It involves the scrutiny of an impartial panel of several experts (usually two or three) on the same topic with which the research or publication deals. The process takes place prior to, and as a condition of acceptance and publication. Sometimes, peer review occurs by comments in the professional forum (bulletins, publications, or at meetings) made *after* and as a result of a project's publication.

There are many reasons for subjecting a project to an examination by such a panel. Primary among them is to determine whether the methodology followed by the researcher is suitable to support the conclusion. The vast majority of published work is not the final say on a topic, but a contribution to the knowledge in that field.

The possible results of peer review are;

1. outright rejection;
2. suggestions that the paper would be suitable if rewritten with identified problems corrected;
3. accepted with minor amendment; and
4. accepted as is.

Rejection may be for presentational, appropriateness to the readership of the journal, or other reasons not connected with the scientific content.

The end result of peer review does not, ordinarily, carry with it the implication that the approved work came to a correct conclusion, but means, instead, that it is sufficiently worthy of being exposed to the broader professional discipline for judgment, further study, examination, comment, or falsification (**Falsifiability Theory**). This is a frequent source of misunderstanding and even misrepresentation in courts.

In some cases, peer review proceeds anonymously, with reviewers having no prior knowledge of the identity of the authors. This is intended as a protection against bias, although it is sometimes possible to make a good guess at the source just from the content of the article.

Most frequently, the peer review process follows these steps.

Step 1. The project's methodology or an article's conclusions is submitted to a panel of experts, who may or may not know who designed the research project or wrote the article intended for publication and who may also not know who the other reviewers are.

Step 2. The reviewers make comments on the project's worth to a project evaluator or publisher's representative and make a recommendation as to whether the project rests on sound methodological applications of appropriate scientific principles, or whether the conclusions drawn from a research project are deemed to contribute to the body of knowledge and literature in the field. These reviewers recommend approval, rejection, or suggest changes, mandatory, or optional.

Step 3a. If the recommendations of at least two of a panel of three reviewers approve of the

project's methodology or the soundness of a paper, it is typically approved for acceptance and/or publication.

Step 3b. If recommendations for changes are made by the reviewers, these comments may be transmitted to the original researcher/author with a request to consider a redesign of the project or a rewrite of the paper. The originator of the project are, most often, not be told who the reviewers were.

The system is not necessarily democratic, with managers of the reviewing process or editors appointed by the publisher retaining the privilege to disregard the opinions of reviewers, appointing new reviewers, or deciding contrary to reviewers' recommendations.

Peer Review in Some Forensic Sciences

In a few forensic disciplines, peer review may have a totally different meaning, one that is more akin to verifying the accuracy of a particular examination result. The scrutiny is provided by having the prior result obtained during an examination of evidence verified by other experts of more advanced or equal stature in the profession.

The purpose of such verification is to determine whether the original conclusion is deemed accurate and trustworthy.[a] The verifier independently retraces the steps of an examiner to see whether the same result is reached by other skilled examiners.

Sometimes, the verifier may not know who the individual is who has reached the original result – a process sometimes referred to as *blind verification*. If the verifier is not aware that an examination was originally done by someone else, the process may be called *double-blind verification*.

Criticism of Peer Review

To accept a publication or a report based on a novel methodology after submission to peer review does not signify the reviewers agree with its conclusions, but rather whether it is deemed worthy to be submitted to the scrutiny of the field. Many studies that have been published are roundly criticized by others even after or as a result of their acceptance or publication. Others may cause the originators to modify or withdraw their original conclusions as a result of input received from reviewers or from others in the field that have become acquainted with the project.

In the traditional publishing business, the peer review process is also extremely cumbersome and time consuming. Most hard-copy peer reviewed journals work on a schedule involving many months of copy evaluation, submission to reviewers, returning comments to the original researcher, and receiving rewrites or revisions, leading ultimately to approval, copy preparation and publication. By the time publication occurs, the conclusions may already have been altered or, in some cases, shown to be in error.

The Internet is providing an alternative to the traditional peer review process by the existence of blogs, online-only publications, and an almost instantaneous exchange of views on any topic at virtually no cost through interactive science publishing. The Internet also provides a readier tool for identifying plagiarism before publication.

Also, because most novel techniques are developed by pioneers in a field, there is an inherent bias against conclusions reached that may go counter to those of peer reviewers who are considered the "established power structure" in a discipline. These individuals may well have a vested interest in safeguarding their own prior positions on the same topic or the positions which they have spent careers supporting or defending, to prevent them from being replaced by newer approaches. At the other end of the spectrum, some reviewers, in evaluating the worth of a project, may tend to have a bias in favor of conclusions that are in agreement with their own expressed views.

The inefficiency in identifying fraudulent conduct is also seen as a serious drawback of the peer review process. Most professions as well as science publications can offer examples of cases in their own areas wherein researchers who had come to be respected were thereafter exposed as having created fraudulent data or falsified the outcome of research to support their conclusions.

Peer Review in American Law

In American legal circles, the words "peer review" have taken on essentially the first one of the approaches to peer review described herein, as a

result of the momentous United States Supreme Court decision in *Daubert v. Merrell Dow Pharmaceuticals* [5] (*see* ***Daubert v. Merrell Dow Pharmaceuticals***), a court opinion that forever changed the legal landscape on how the decision on the admissibility of expert opinion testimony is to be made (**Expert Opinion: United States**).

Peer review, in *Daubert*, came to be lauded as a desirable, though not an essential, factor whereby the soundness of a scientific methodology utilized by particular experts can be judged. As a result, some forensic disciplines that consider peer review as synonymous with "verification of a result obtained in an analysis" have met some resistance in courts in having their conclusions deemed as satisfying a *Daubert*-style of peer review.

Thus, in *United States v. Mitchell*, the court, in examining whether "fingerprint identification" satisfied the peer review factor of *Daubert*, recognized that the type of verification required of latent-print examiners in the analysis, comparison, evaluation and verification ACE-V methodology might "not be peer review in its best form, but, on balance, the peer review factor does favor admissibility" [6]. The overwhelming majority of court decisions of appellate tribunals have reached similar decisions in allowing the "V" in the ACE-V methodology to satisfy the *Daubert* "peer review" factor.[b]

There remain critics of those forensic practices involved in individualization who would like to see the courts reject the admissibility of examination techniques wherein the "scientific" model of peer review is not strictly followed. Defenders of those practices respond that inasmuch as the existence of a peer reviewed literature guarantees neither accuracy nor reliability of an examination result, to impose such a high standard on the verification process in forensic science is probably unwarranted in light of the precise language in the United States Supreme Court's *Daubert* decision itself.[c]

End Notes

[a.] For example: In the Latent Friction Ridge Comparison discipline, verification of the results obtained by one examiner's individualization is required in some circumstances and recommended in others. See the Scientific Working Group of Friction Ridge Analysis, Study and Technology (SWGFAST), in its guidelines and standards at www.swgfast.org. Several other forensic disciplines use the ACE-V methodology in the comparison sciences, among them forensic document examiners, firearm and toolmark examiners.

[b.] The ACE-V process is described in **Friction Ridge Skin: Comparison and Identification**.

[c.] In *Mitchell, supra* [6] at 244, the court explained: "*Daubert* does not require that a party who proffers expert testimony carry the burden of proving to the judge that the expert's assessment of the situation is correct. As long as the expert's scientific testimony rests on 'good grounds, based on what is known,'" the testimony should be admitted and tested by the adversary system.

References

[1] On peer review generally, Shatz, D. (2004). *Peer Review: A Critical Inquiry*, Rowman & Littlefield, Lanham.

[2] Hames, I. (2007). *Peer Review and Manuscript Management in Scientific Journals: Guidelines for Good Practice*, Black well publishing.

[3] Wagner, E., Godlee, F & Jefferson, T. (2002). *How to Survive Peer Review* BMJ Books, London.

[4] Wiegers, K.E. (2001). *Peer Reviews in Software: A Practical Guide*, Addison-Wesley, Boston.

[5] Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).

[6] United States v. Mitchell, 365 F.3d 215, at 239 (3rd Cir. 2004), cert. denied 543 U.S. 974 (2004).

ANDRE MOENSSENS

Pen and Writing Instruments *see* Writing Instruments and Printing Devices

Persistence: Trace *see* Trace Evidence: Transfer, Persistence, and Value

Person Identification ***see*** Elderly in Court

Personality Testing ***see*** Psychological Testing

Persuasibility of Children ***see*** Children: Suggestibility of

Persuasion ***see*** Deception: Truth Serum

Pharmacodynamics ***see*** Alcohol: Interaction with Other Drugs

Pharmacogenomics

Introduction

In 1865, Gregor Mendel discovered the basis of inheritance and genetics. Pharmacogenetics and its related science pharmacogenomics emerged as our understanding of genetics advanced, in particular, that certain enzymes coded by the genetic blueprint process both endogenous and exogenous substances (such as drugs). This understanding is likely to develop further in years to come and may even allow personalization of drug treatment – personalized medicine, based on a person's genetic profile [1, 2]. Thus, pharmacogenomics is likely to increasingly contribute to clinical and forensic practices in the future.

Pharmacogenomics has already been useful to optimize treatment of lung cancer by identifying genotypes likely to predispose to toxicity and poor outcome [3]. Patients with homozygous *UGT1A1*28* are more likely to develop neutropenia, whereas patients with *GSTP1 1105 G/A* or *G/G* genotypes may show partial response. Another study identified the combination of five-gene biomarkers (*DUSP6, MMD, STAT1, ER BB3*, and *LCK*) as an independent predictor of relapse-free and overall survival [4].

Pharmacogenomics and pharmacogenetics are currently used interchangeably; however, pharmacogenetics is readily defined as the study of the genetic effect, e.g., single-nucleotide polymorphism (SNP), on an individual's ability to metabolize a drug or other substance, pharmacgenomics is concerned with the whole-genome effect on drug metabolism and efficacy.

The emerging clinical applications of pharmacogenetics/pharmacogenomics may be directly verified by the use (and approval by various regulatory agencies) of genotyping methodologies/platforms, the frequent inclusion of this topic in scientific and clinical meetings, and the availability (in 2007) of a pharmacogenomic survey program offered by the College of American Pathologists.

Pharmacogenomics serves as an "adjunct" to other tests and scientific practices. For example, genotyping can be conducted as an adjunct to the autopsy (molecular autopsy) and complements the current forensic applications such as DNA fingerprinting.

Thus, this article presents pharmacogenomics as a complementing discipline to enhance drug therapy and serve as an adjunct to forensic pathology/toxicology. Forensic applications include an algorithm for the use of pharmacogenomics in forensic pathology/toxicology, with emphasis on interrelationships to circumstances and drug use/abuse history, scene investigation, and autopsy findings.

On the opposite end of the spectrum, the use of molecular/genetic testing in postmortem forensic science has included DNA fingerprinting for identity testing. Recently, pharmacogenomics as molecular autopsy has been used for the assessment of genetic contribution to drug toxicity in postmortem forensic toxicology. In common with the other applications

of clinical and scientific findings in forensic science, the findings might add to the understanding of disease mechanisms and optimization of treatment including drug therapy. Thus, the use of pharmacogenomics in forensic toxicology may add to the understanding of drug toxicity due to genetically predisposed impaired drug metabolism, and in so doing thus provide better interpretation and indirectly enabling the emerging personalized medicine.

Principles of Pharmacogenetics/Pharmacogenomics

According to the central dogma of molecular biology, the genetic code of DNA is passed, through transcription, onto *mRNA*. The information in *mRNA* is passed, through translation, in protein synthesis. These proteins may be drug-metabolizing enzymes, transporters, and receptors. As a result, DNA genetic variations (*see also* **Polymorphism: Genetic**) determine the enzyme activity, transporters, and receptor sensitivity.

For drug-metabolizing enzyme genes, the lack of and the presence of genetic variations results in normal, deficient, or higher enzyme activities. Genetic variations might include SNPs, deletion, duplications, and other variations. The polygenic determinants of drug response are illustrated by Figure 1. By comparing an individual with two wild-type alleles (normal or extensive metabolizer *see also* **Phenotype**) on the left to an individual with two variant alleles (a poor metabolizer) on the right, the genetic variations would result in lower enzymes activity and elevated plasma concentration *versus* time curve (calculated as area under curve, AUC), with corresponding increased toxicity and decreased receptor sensitivity and efficacy. The heterozygous individual in the middle with one variant allele (an

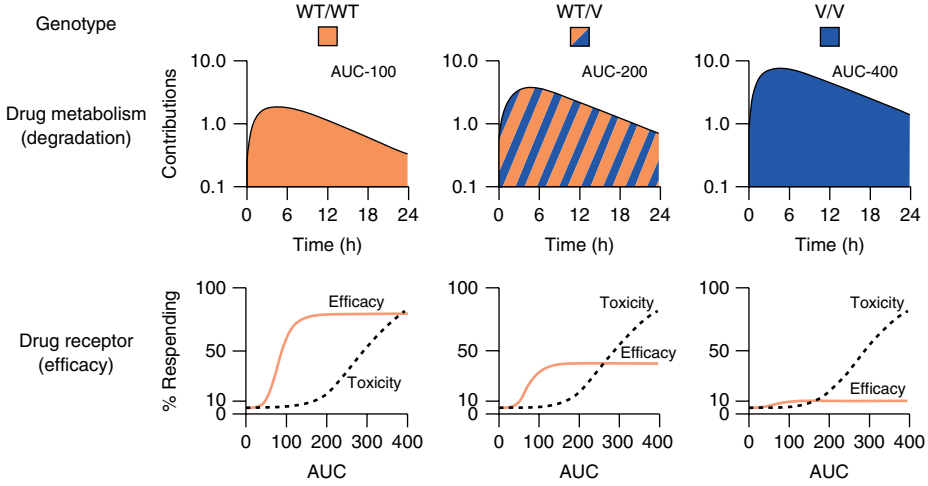


Figure 1 Polygenic determinants of drug response [With permission from publisher of Ref. 5, © Massachusetts Medical Society, 2003.]

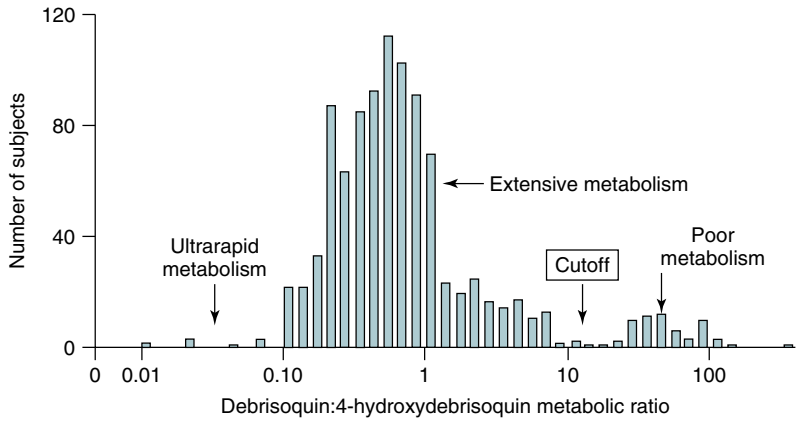


Figure 2 Pharmacogenetics of CYP 2D6 [Reprinted by permission of Macmillan Publishers Ltd: Clinical Pharmacology & Therapeutics [6], ©1992.]

intermediate metabolizer) exhibits AUC, toxicity, and efficacy intermediate between those of the extensive and poor metabolizers. With a possible combination of nine metabolism and receptor genotypes, the therapeutic index (separation of therapeutic and toxic responses) could range from 13 to 0.125.

Further, individuals with multiple copies of the genes are ultrarapid metabolizers (UMs). An example of this is shown in Figure 2, which shows the debrisoquin metabolic ratios (MRs) of these phenotypes [6].

Race and ethnicity can affect the prevalence of "poor" metabolizers, but they are usually relatively uncommon. For example, there are variations in the prevalence of one of the more common P450 enzymes, CYP2D6, in African-Americans and the Chinese. Patients who are genotypically slow metabolizers may, for example, require more tramadol (analgesic) than other "normal" subjects, possibly explaining the variable response to pain management in some patients.

Genetic variation can also make a person more susceptible to disease [7].

Tests and Methodologies

The majority of the current testing is primarily based on pharmacogenetics (PGx). The top 10 tests in 2005 as assessed by the American Association of clinical Chemists (AACC) are listed in Table 1 (in descending order).

Within the list, CYP and other phase II enzymes (conjugating enzymes) such as uridine diphosphate glucuronosyltransferase 1A1 (UGT1A1) accounted for the majority of drug/substrate metabolism for drugs approved in the United States, about 75%

Table 1 Top 10 pharmacogenomics tests

Abbreviation	Name and function
CYP 2D6	Cytochrome P450 2D6, phase I drug-metabolizing enzyme
TPMT	Thiopurine *S*-methyltransferase, phase II drug-metabolizing enzyme
CYP 2C9	Cytochrome P450 2C9, phase I drug-metabolizing enzyme
CYP 2C19	Cytochrome P450 2C19, phase I drug-metabolizing enzyme
NAT	*N*-acetyltransferase, phase II drug-metabolizing enzyme
CYP 3A5	Cytochrome P450 3A5, phase I drug-metabolizing enzyme
UGT1A1	Uridine diphosphate glucuronosyltransferase 1A1, phase II drug-metabolizing enzyme
MDR1	Multidrug resistance (P-glycoprotein), drug protein transporter
CYP 2B6	Cytochrome P450 2B6, phase I drug-metabolizing enzyme
MTHFR	Methylenetetrahydrofolate(CH_2THF) reductase converts CH_2THF to 5-methyltetrahydrofolate

involving CYP 3A4 and cytochrome P450 2D6 (CYP 2D6) enzymes. Testing is complicated since there are more than 160 alterations for *CYP 2D6* genes alone. Assay problems can include allelic drop-out, intra-allelic recombination, the need for specific assays not affected by pseudogenes *CYP 2D7* and *CYP 2D8*, and to address gene conversion of *CYP 2D6* from *CYP 2D7*. Consequently, genetic testing is not yet widespread.

The more common pharmacogenomic tests are readily performed either by home-brew assay or a commercially available test or platform. The approaches included nonamplification, e.g., fluorescent *in situ* hybridization (FISH), target and signal amplification methods including endpoint polymerase chain reaction (PCR) detection, allele-specific primers, length analysis using restriction fragment length polymorphism (RFLP) and oligonucleotide ligation assay (OLA), real-time PCR, signal amplifications, and new methods including solid-phase microarrays and fluorescent-based bead assay (liquid mircoarray). The manufacturers and the status of FDA approvals are listed in Table 2.

Some laboratories performing genotyping have adapted the PCR liquid bead-based detection. The choice of the platform and assays seems to reflect the ease of "home-brew" assay development and the lower cost of the instrument and reagents. Nevertheless, there are relative advantages and disadvantages of the various testing kits available. These need to be understood before having a test conducted.

Clinical Applications

The clinical applications of pharmacogenomics are classified according to the drug group, specialties, and diseases and include opioids, pain management, nicotine addiction, HIV treatment, immunosuppressants, and thiopurine *S*-methyltransferase (TPMT) for acute lymphoblastic leukemia, and psychiatry. Polymorphisms of the neurotransmitter transporters (serotonin, norepinephrine (noradrenaline), dopamine and P-glycoprotein, and serotonin transporter show little effect on serotonin active antidepressants selective serotonin reuptake inhibitor (SSRI) response, suggesting that the current scientific literature showing transporter genotypes is not yet contributory to predictive therapy [9]. There was evidence to support the use of genotype-based dosing for drug transporters such as P-glycoprotein (ABCB1) and OATP-C (SLC21A6) and other CYP enzyme genes. Other important and emerging areas include cancer, cardiovascular disorders, and hematology (Table 3)

It would be important to recognize the role of therapeutic drug monitoring (TDM) and toxicology

Table 2 Methodologies for pharmacogenetics testing[a]

Method	Company	FDA cleared or approved
Sequencing[b]	Abbott (Abbott Park, IL)	Yes
Real-time PCR	Applied Biosystems (Foster City, CA)	–
PCR arrays	Autogenomics (Carlsbad, CA)	yes
Sequencing[b]	Bayer Healthcare (Tarrytown, NY)	Yes
Pyrosequencing	Biotage AB (Uppsala, Sweden)	–
Real-time PCR	Celera Diagnostics (Alemeda, CA)	–
Real-time, allele-specific PCR	DxS Genotyping (Manchester, UK)	–
PCR	Gentris (Morrisville, NC)	yes
User-developed PCR arrays	Nanogen (San Diego, CA)	–
Nanoparticles	Nanosphere (Northbrook, IL)	yes
PCR arrays	Roche Diagnostics (Indianapolis, IN)	Yes
Invader assay	Thirdwave Technologies (Madison, WI)	Yes
PCR bead-based detection	Tm Biosciences Corp. (Toronto, ON)	–
FISH	Vysis (Des Plaines, IL)[c]	Yes

PCR, polymerase chain reaction; FISH, florescent *in situ* hybridization

[a] Reproduced with permission from Ref. 8. © AACC, 2006

[b] Sequencing for HIV drug resistance

[c] Vysis is now Abbott Molecular Diagnostics

Table 3 Examples of associations between drug response and genetic variants[a]

Drug	Variable clinical effect	Genes with associated variants	Possible mechanism
Azathioprine and mercaptopurine	Bone marrow aplasia	TPMT	Hypofunctional alleles
	Reduced therapeutic effect at standard doses		Wild-type alleles
Some antidepressants and β-blockers	Increase side-effect risk	CYP 2D6	Hypofunctional alleles
	Decreased efficacy		Gene duplication
Omeprazole	*Helicobacter pylori* cure rate	CYP 2C19	Hypofunctional alleles
Irinotecan	Neutropenia	UGT1A1	Decreased expression due to regulatory polymorphism
HIV protease inhibitors	Central nervous system levels	MDR1	Altered P-glycoprotein function
β-Blockers	Blood pressure lowering and heart rate slowing	ADRB1	Altered receptor function or number
Diuretics	Blood pressure lowering	ADD1	Altered cytoskeletal function by adducin variants
Warfarin	Anticoagulation	VKOrC1	Variant haplotypes in regulatory regions leading to variable expression
		CYP 2C9	Coding region variants causing reduced *S*-warfarin clearance
Abacavir	Immunologic reactions	HLA variants	Altered immunologic responses
QT-prolonging antiarrhythmics	Drug-induced arrhythmia	Ion-channel genes	Exposure of subclinical reduction in repolarizing currents by drugs
General anesthetics	Malignant hyperthermia	RYR1	Anesthetic-induced increased release of sarcoplasmic reticulum calcium by mutant channels
Inhaled steroids	Bronchodilation	CHCR1	Unknown
HMG-CoA reductase inhibitors (statins)	Low-density lipoprotein	HMGCR	Altered HMG-CoA reductase activity

[a] Reproduced with permission from Ref. 10. © ACP, 2006

more generally as global phenotypic indexes including contributing pharmacokinetic, pharmacodynamic, drug–drug interaction, and other environmental factors develop. Thus, pharmacogenomic biomarkers might be readily characterized as an adjunct to enable the practice of personalized medicine. To update on these applications, a summary of recent examples would include pharmacogenomics for alcoholism, psychiatric disorders, and opiates (opioids).

Alcohol

Alcoholism is a complex psychiatric disorder with high heritability (50–60%) and with a lifetime prevalence of alcohol dependence of 20% in men and 8% in women in the United States. Alcoholics may be categorized as follows: type 1 – later onset with feelings of anxiety, guilt, and high harm avoidance and type 2 – early age of onset, usually men, manifesting as impulsive and antisocial behavior, and associated with low levels of brain serotonin [11].

Genetic variations of alcohol metabolizing genes affect drinking behavior and hence decrease alcoholism. Genes of neurotransmitter pathways "reward pathway" (serotonin, dopamine, gamma amino-butyric acid (GABA), glutamate, and β-endorphin) and the behavioral stress response system (corticotrophin-releasing factor and neuropeptide Y) may constitute therapeutic targets. By using the type 1/2 systems, progress is being made to understand the pharmacogenomic response to current pharmacotherapy. Further, alcohol inheritability may be affected by genetic variations for a number of enzymes and other factors such as those affecting cognitive function, stress/anxiety response, and opioid function [12].

Psychiatric Disorders

There are a number of schizophrenia-associated genes associated with affective disorders.

A number of P450 enzymes are involved in metabolizing antipsychotic drugs including CYP 2D6, CYP 1A2, and CYP3A4 (Table 4). Since CYP 1A2 is inducible, individuals with *CYP1A2* variants and some SNP combinations (haplotypes), in the 5′-regulatory regions, may demonstrate variable response [13].

Genetic variability leading to clinically significant changes has been shown for a number of antidepressants and antipsychotic drugs [14–17]. This has led to a lower incidence of adverse effects and has optimized the response to the drug therapy.

Opioids

The best example of an opioid response being affected by genetic variability is codeine [18].

Table 4 CYPs with major roles n the *in vivo* clearance of antipsychotic agents[(a)]

CYP	Antipsychotic drug	Altered drug substrates and inhibitors that may be used in psychotic patients	Inducers	Numbers of allelic variants
1A2	Clozapine	Omeprazole	Omeprazole	24 plus wild-type (also 9 predicted haplotypes)
	Olanzapine		Cigarette smoke Barbecued meats	
2D6	Risperidone Chlorpromazine Thioridazine	Dextromethorphan Codeine Imipramine Nortriptyline Paroxetine	None	94 plus wild type
3A4	Ziprasidone Quetiapine Aripiprazole Haloperidol	Erythromycin Diltiazem Ciclosporine Ethinyl estradiol	Rifampicin Carbamazepine Phenytoin Dexamethasone	38 plus wild type

[(a)] Reproduced from Ref. 13. © Pharmaceutical Press, 2006

Codeine is metabolized CYP 2D6 to its active metabolite morphine. UMs may suffer exaggerated and toxic opioidergic effects and poor metabolizers may experience reduced pain relief [19]. Even though P-glycoprotein is an opioid transporter, *ABCB1* genotypes' influence on opioid pharmacodynamics and dosage requirements were highly variable [20].

Forensic Applications

Pharmacogenomics is best treated as an adjunct to forensic pathology (*see also* **Toxicology: Forensic Applications of**), complementing information including autopsy findings, case history including medication, and scene investigation. The term *molecular autopsy* has been used to represent the application of genetic testing to assist the more conventional autopsy examination.

Case review initially assesses the likelihood of a genetic factor explaining a certain event. These factors include the presence of acute or chronic toxicity, autopsy findings (*see also* **Autopsy**), sample collection sites, postmortem intervals, coadministered drugs, case/medical/medication histories, scene investigations, and possible intent. As the case review continues with developing toxicological findings, elevated drug concentrations and identification of known and unknown drug/metabolites that might have interacted are prime criteria for case selection for pharmacogenomics (*see also* **Postmortem Biochemical Examinations** and **Toxicology: Forensic Applications of**). Further, the postmortem intervals are also considered in case selection and in interpretation. Once the case is selected, whole-blood samples are then transferred (*see also* **Toxicology: Analysis**), with chain of custody, for pharmacogenomics testing by the molecular and pharmacogenomics laboratory.

Currently, the testing platform is based on Pyrosequencing™ and include the following: *CYP 2D6*2-*8, CYP 2C9*2*3, CYP 2C19*2-*4, CYP 3A4*1B*, and *CYP 3A5*3.* The forensic applications are illustrated by the opioids (*see also* **Opioids**): methadone, oxycodone, and fentanyl.

UMs will produce larger amounts of morphine from codeine, and this is more likely to cause toxicity as well as increase the likelihood of overdose to breast-feeding infants [21].

Methadone response is affected by the activity of CYP3A4 and CYP2B6 and to a lesser extent CYP 2D6. ABCB1 accounted for minor pharmacokinetic variability. Other studies showed that *CYP* genes did not affect methadone metabolism.

*Case History 1. The decedent was a 41-year-old female, 6 month in her pregnancy. She had heart murmur and rheumatoid arthritis treated with methadone. Further, amitriptyline was prescribed for her depression. She celebrated New Year' Eve with her husband. On the following morning, she was found dead in her living room. Scene investigation revealed her ingestion of nine 50-mg tablets of amitriptyline within 17 days and two to three 95-mg dose of methadone. Several years before, she had attempted suicide by drug ingestion. Toxicological analysis of iliac blood showed the following drugs and concentrations in milligrams per liter: methadone, 0.7; amitriptyline, 1.5; nortriptyline, 2.2; diazepam, 0.19; and N-desmethyl diazepam, 0.13. The elevated antidepressant concentrations of iliac, peripheral blood, would not be due to postmortem drug redistribution and more attributable to acute drug ingestion. Molecular autopsy by pharmacogenomics showed that she was homozygous for CYP 2D6*4, corresponding to a poor metabolizer phenotype. This would result in the lack of hydroxylation of methadone, amitriptyline, and nortriptyline, thus resulting in elevated parent drug concentrations. Death certification was as follows: cause, mixed drug toxicity, and manner, accident [22].*

Further, cytochrome P450 2B6 (CYP 2B6) poor metabolizers are more likely to develop QT elongation. In addition, higher doses are needed for individuals with two copies of the wild-type haplotype and lower doses for AGCTT haplotype. Thus, haplotyping may be used to individualize methadone therapy.

Together, the genetic contribution to drug metabolism impairment may be interpreted as a gene dose effect, resulting in more pronounced drug toxicity. Such an effect might be readily demonstrated as a pharmacogenomics converging continuum and is helpful to understand the effect in the living.

Similarly, oxycodone response is affected by the activity of CYP 2D6 isozyme as Case Report 2 illustrates.

*Case History 2. A 49-year-old male with a history of alcoholism and chronic lower back pain was prescribed OxyContin™ and Percocet™. He also had depression, posttraumatic stress disorder, and attempted suicide once. The scene investigation revealed only 12 of the 60 oxycodone pills that were obtained from the day before. Last seen by his roommate in the morning, the decedent was found unresponsive later that afternoon. Toxicological analysis of subclavian blood showed methadone, 0.44, mg l^{-1}. Testing showed that he was CYP 2D6*4 homozygous, corresponding to a poor metabolizer phenotype. Autopsy also showed that he had hepatic cirrhosis and atherosclerotic heart disease. Given the short postmortem interval and subclavian blood source, the elevated oxycodone was not due to postmortem drug redistribution but might be due to poor metabolizer phenotype and hepatic cirrhosis, both contributing to impaired drug metabolism. Death was certified as follows: cause of death, oxycodone overdose, and manner of death, accident [23].*

Similarly, fentanyl response is affected by the activity of a number of enzyme types as Case Report 3 illustrates.

Other Applications

Pharmacogenomics including proteomic, RNA interference, and other molecular and functional biomarkers are being increasingly used in drug discovery and development as encouraged by governmental agencies and scientific and professional organizations. Pharmacogenomics has been used to enhance patient's safety in therapies with approved drugs. Thus, the use of pharmacogenomics is regarded as an adjunct for optimizing drug therapy and, in forensic pathology/toxicology, for providing a molecular autopsy. Thus, TDM and toxicology would be helpful as global indexes. This would add to the understanding of potential genetic contribution to metabolism of approved drugs such as methadone and oxycodone, thus enabling and improving the practice of antemortem drug therapy, a tangible benefit for family members of decedents.

*Case History 3. A 44-year-old white female complained about her knee pain and was treated with Duragesic™ fentanyl patches. She appeared to be "goofy" and went to bed. She was found dead 24 h later. The decedent was a drug abuser with psychiatric history. Previously, she cut her arm to obtain drugs and also expressed suicidal ideation. Toxicological analysis of subclavian blood showed the following drugs and concentrations in milligrams per liter: fentanyl, 0.019; norfentanyl, 0.008; cyclobenzaprine, 0.16; tramadol, 0.06, diphenhydramine, 0.08; citalopram, 0.22; and olanzapine, positive. The MR of fentanyl/norfentanyl was 2.5. Molecular autopsy by pharmacogenomics showed that she was heterozygous for CYP 3A4*1B and CYP 3A5*3, different from the majority of Caucasians CYP 3A4, WT, and CYP 3A5*3 HM. In this study with limited number of fentanyl cases, the MR of this case is lower than the MRs of majority of the cases with high-fentanyl concentrations. Together, these findings suggested that for the first time, cytochrome P450 3A5 (CYP 3A5) comediated with CYP 3A4 the metabolism of fentanyl to norfentanyl. Death certification for the above case was as follows: cause of death, mixed drug toxicity, and manner of death, accident [24].*

The experience may be helpful to apply pharmacogenomics in the possible emerging practice of personalized justice. One such emerging application would be applying pharmacogenomic biomarkers for the interpretation of possible "side effect/behavior/impaired performance" of the drivers arrested in cases involving driving under the influence of drugs (*see also* **Drug-Impaired Driving** and **Behavioral Toxicology**). By using pharmacogenomics, the driver's impaired driving performance may be partially explained on the basis of genetics-predisposed impairment of drug metabolism and therefore accumulation. This might result in driving

impairment. Similar to the use of DNA fingerprinting in identity testing, applying pharmacogenomics and other molecular biomarkers in the future as adjunct biomarkers in the above context may constitute a rational approach in understanding and the deliberation of the driver's liability, thus offering the possibility of personalized justice.

Further, with the availability of proficiency survey program by the College of American Pathologists in 2007 and quality assurance/control from commercial sources, the clinical adaptations will soon be readily achieved by clinical laboratories. National Academy of Clinical Biochemistry (NACB) guidelines would certainly pave the way. Challenges remain for adequate reimbursement, clinical interpretation, ethical guidelines, and education of the patients and healthcare professionals. Thus, Gregor Mendel's "little trick – long story" would indeed continue in the form of new articles as a result of the human genome project and the rapid advances in molecular biology and pharmacogenomics.

References

[1] Wong, S.H.Y. (2007). Pharmacogenomic and personalized medicine for drug addiction and toxicology – towards personalized justice? *11th Asian Pacific Congress of Clinical Biochemistry*, October 2007, Beijing.

[2] Wong, S.H.Y., Linder, M.W. & Valdes Jr, R. (eds) (2006). *Pharmacogenomics and Proteomics – Enabling the Practice of Personalized Medicine*, AACC Press, Washington, DC, pp. 1–386.

[3] Pillot, G.A., Read, W.L., Hennenfent, K.L., Marsh, S., Gao, F., Viswanathan, A., Cummings, K., McLeod, H.L. & Govindan, R. (2006). A phase II study of irinotecan and carboplatin in advanced non-small cell lung cancer with pharmacogenomic analysis: final report, *Journal of Thoracic Oncology* **1**(9), 972–978.

[4] Chen, H.-Y., Yu, S.-L., Chen, C.-H., Chang, G.-C., Chen, C.-Y., Yuan, A., Cheng, C.-L., Wang, C.-H., Terng, H.-J., Kao, S.-F., Chan, W.-K., Li, H.-N., Liu, C.-C., Singh, S., Chen, W.-J., Chen, J.J.W. & Yang, P.-C. (2007). A five-gene signature and clinical outcome in non-small-cell lung cancer, *The New England Journal of Medicine* **356**, 11–20.

[5] Evans, W.E. & McLeod, H.L. (2003). Pharmacogenomics – drug disposition, drug targets and side effects, *The New England Journal of Medicine* **348**, 538–549.

[6] Bertillsson, L., Lou, Y.Q., Du, Y.-L., Liu, Y., Kuang, T.-Y., Liao, X.-M., Wang, K.Y., Reviriego, J., Iselius, L. & Sjoqvist, F. (1992). Pronounced differences between Chinese and Swedish populations in the polymorphic hydroxylations of debrisoquin and S-mephenytoin, *Clinical Pharmacology and Therapeutics* **51**, 388–397.

[7] Paul, N.W. & Fangerau, H. (2006). Why should we bother? Ethical and social issues in individualized medicine, *Current Cancer Drug Targets* **7**(12), 1721–1727.

[8] Payne, D. (2006). Pharmacogenetic testing: how to choose a method to analyze genetic changes, *Clinical Laboratory News* **7**, 14–16.

[9] Kirchheiner, J., Gründemann, D. & Schömig, E. (2006). Contribution of allelic variations in transporters to the phenotype of drug response, *Journal of Psychopharmacology* **20**,(Suppl 4), 27–32.

[10] Roden, D.M., Altman, R.B., Benowitz, N.L., Flockhart, D.A., Giacomini, K.M., Johnson, J.A., Krauss, R.M., McLeod, H.L., Ratain, M.J., Relling, M.V., Ring, H.Z., Shuldiner, A.R., Weinshilboum, R.M. & Weiss, S.T. The Pharmacogenetics Research Network (2006). Pharmacogenomics: challenges and opportunities (review), *Annals of Internal Medicine* **145**, 749–757.

[11] Enoch, M.A. (2003). Pharmacogenomics of alcohol response and addiction, *American Journal of Pharmacogenomics* **3**(4), 217–232.

[12] Oroszi, G. & Goldman, D. (2004). Alcoholism: genes and mechanisms, *Pharmacogenomics* **5**(8), 1037–1048.

[13] Murray, M. (2006). Role of CYP pharmacogenetics and drug-drug interactions in the efficacy and safety of atypical and other antipsychotic agents, *The Journal of Pharmacy and Pharmacology* **58**(7), 871–885.

[14] Kirchheiner, J., Brøsen, K., Dahl, M.L., Gram, L.F., Kasper, S., Roots, I., Sjoqvist, F., Spina, E. & Brockmoller, J. (2001). CYP2D6 and CYP2C19 genotype-based dose recommendations for antidepressants: a first step towards subpopulation-specific dosages, *Acta Psychiatrica Scandinavica* **104**, 173–192.

[15] Dorado, P., Berecz, R., Peñas-Lledó, E.M., Cáceres, M.C. & Llerena, A. (2006). Clinical implications of CYP2D6 genetic polymorphism during treatment with antipsychotic drugs, *Current Drug Targets* **7**(12), 1671–1680.

[16] Lin, Y.-C., Ellingrod, V.L., Bishop, J.R. & Miller, D.D. (2006). The relationship between P-glycoprotein (PGP) polymorphisms and response to olanzapine treatment in schizophrenia, *Therapeutic Drug Monitoring* **28**, 668–672.

[17] Baumann, P., Barbe, R., Vabre-Bogdalova, A., Garran, E., Crettol, S. & Eap, C. (2006). Epileptiform seizure after sertraline treatment in an adolescent experiencing obsessive-compulsive disorder and presenting a rare pharmacogenetic status, *Journal of Clinical Psychopharmacology* **26**, 679–681.

[18] Kirchheiner, J., Schmidt, H., Tzvetkov, M., Keulen, J.T., Lötsch, J., Roots, I. & Brockmöller, J. (2007). Pharmacokinetics of codeine and its metabolite morphine in ultra-rapid metabolizers due to CYP2D6 duplication, *The Pharmacogenomics Journal* **7**(4), 257–265.

[19] Gasche, Y., Daali, Y., Marc Fathi, M., Chiappe, A., Cottini, S., Dayer, P. & Desmeules, J. (2004). Codeine intoxication associated with ultrarapid CYP2D6 metabolism, *The New England Journal of Medicine* **351**, 2827–2831.
[20] Somogyi, A.A., Barratt, D.T. & Coller, J.K. (2007). Pharmacogenetics of opioids, *Clinical Pharmacology and Therapeutics* **81**, 429–444.
[21] Koren, G., Cairns, J., Chitayat, D., Gaedigk, A. & Leeder, S.J. (2006). Pharmacogenetics of morphine poisoning in a breastfed neonate of a codeine-prescribed mother, *Lancet* **368**, 704.
[22] Wong, S.H., Wagner, M.A., Jentzen, J.M., Schur, B.C., Bjerke, J., Gock, S.B. & Chang, C.C. (2003). Pharmcogenomics as an aspect of molecular autopsy for forensic pathology/toxicology: does genotyping CYP2D6 serve as an adjunct for certifying methadone toxicity? *Journal of Forensic Sciences* **48**, 1406–1415.
[23] Jannetto, P.J., Wong, S.H., Gock, S.B., Laleli-Sahin, E., Schur, B.C. & Jentzen, J.M. (2002). Pharmacogenomics as molecular autopsy for postmortem forensic toxicology: genotyping cytochrome P450 2D6 for oxycodone cases, *Journal of Analytical Toxicology* **26**, 438–447.
[24] Jin, M., Gock, S.B., Jannetto, P.J., Jentzen, J.M. & Wong, S.H. (2005). Pharmacogenomics as molecular autopsy for forensic toxicology: genotyping cytochrome P450 3A4*1B and 3A5*3 for 25 fentanyl cases, *Journal of Analytical Toxicology* **29**, 590–598.

Further Reading

Abrahams, E., Ginsburg, G.S. & Silver, M. (2005). The personalized medicine coalition: goals and strategies, *American Journal of Pharmacogenomics* **5**(6), 345–355.
International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921.
Jicinio, J. & Wong, M.-L. (2002). *Pharmacogenomics*, Wiley-VCH, Weinheim, pp. 1–559.
Linder, M.W., Prough, R.A. & Valdes Jr, R. (1997). Pharmacogenetics: a laboratory tool for optimizing therapeutic efficiency (review), *Clinical Chemistry* **43**(2), 254–266.
Tsai, K.Y., Tsao, H. (2007). Primer on the human genome, *Journal of the American Academy of Dermatology* **56**(5), 719–735.
Venter, J.C., Adams, M.D., Myers, E.W., *et al.* (2001). The sequence of human genome, *Science* **291**, 1304–1351.
Weber, W.W. (1997). *Pharmacogenetics*, Oxford University Press, Oxford, pp. 1–344.
Weinshilboum, R. (2003). Inheritance and drug response, *The New England Journal of Medicine* **348**, 529–537.
http://www.nacb.org/lmpg/LMPG_Pharmacogenetics.pdf. 2007.

STEVEN H. Y. WONG

Pharmacokinetics *see* Alcohol: Interaction with Other Drugs

Phenotype

Introduction

The genomics and bioinformatics era has enabled the recent development of systems and methods for the derivation of human phenotype from the genotyping of crime stain DNA evidence. This is accomplished through an empirical process of inference, either directly through genotypes for the functionally relevant genetic positions (loci) or indirectly through an appreciation of the genetic heritage of the donor. Forensic scientists are accustomed to using DNA sequence polymorphisms as identifiers. On the basis of the frequency of the identifier, we can statistically link individuals with samples associated with criminal investigations. Microsatellites, such as the short tandem repeats (STRs) employed for the Federal Bureau of Investigation (FBI's) combined DNA index system (CODIS) have been most commonly used for this purpose because they are multiallelic. That is to say, they have many alleles (varieties per location, or locus) and so relatively few loci are necessary to produce sequence signatures. Microsatellites are not chosen from the human genome based on their human ancestry or phenotype information, since this type of information would render them more or less powerful as identification tools among different human subpopulations. Thus, they are relatively useless for inferring phenotype. If we find that a suspect or a database entry matches the STR profile, we can extract probative value from the profile, otherwise, we have traditionally employed nongenetic investigative processes designed to produce the suspects necessary to achieve this objective.

For many cases, the investigative process begins with an attempt to ascribe characteristics or features to the perpetrator that can lead to identification. If a human eyewitness were available, we would query the witness about physical appearance – what the suspect was wearing, the suspects "race" and, more

specifically, we would obtain estimates and ranges for the basic anthropometric phenotypes such as skin color, eye color, height, etc. In the best-case scenario, we might obtain the basic physical descriptors or features such as those that might be present on a work identification card for the suspect, and, in a suboptimal scenario, we might obtain descriptors that are inaccurate or misleading. Unfortunately, the best-case scenario with human eyewitnesses is rarely achieved; unreliable information is sometimes provided by design (perhaps the witness has a motive to deceive) or by circumstance such as might be the case if the witness observed from afar on a darkly lit street. Assuming that the integrity of the witness is intact, and the witness had a clear view of a suspect, we are still left with fundamental problems created by the subjective nature of the human experience. Human testimony is generally not objective, almost always unstandardized, and never formally quantitative. Physical descriptions provided by one witness may or may not comport with those from another – not necessarily because one is wrong, but because one witness' opinion on what it is to look "dark" or "Asian", for example, may be different from the other. These basic problems translate into a fundamental defect associated with the extraction of phenotype data from human eyewitnesses – namely, that the testimony is not falsifiable, meaning that an investigator cannot verify its accuracy through independent examination. Nonetheless, there is a reason investigators consider themselves lucky to have access to eyewitness testimony. An investigation is much like a classification problem, where one attempts to make the most specific classification possible (the ascription of identity). Given the number of variables, and the fact that the number of incorrect classifications far outnumbers those that are correct (one) the classification problem is complex and we need classification features in order to delimit the likely possibilities. Here, data from witness testimony is analogous to a classification feature, and information theory (e.g., Bayes theorem) teaches us that even suboptimally informative (as opposed to uninformative or misleading) "features" are better than no features at all [1, 2]. Clearly, it would be far better to extract physical information on the donor using empirical techniques that lend themselves to the scientific method – techniques based on the observations of unperturbed nature, producing objective, quantitative and falsifiable data that can be communicated logically to others using standardized terms. It happens that if DNA is available from a crime scene, we now have access to new methods meeting these criteria for the inference of certain aspects of physical appearance (human traits or phenotypes).

The recent completion of the first human genome draft has provided a foundation for the development of empirical processes by which a partial physical portrait of a DNA donor can be constructed. In this article, we discuss certain overt phenotypes that are highly heritable, which are required if we hope to draw connections between DNA sequences and trait values. For example, parents with dark skin tend to have children with similarly dark skin and blue-eyed parents tend not to have brown-eyed offspring, suggesting that if we knew which and how genetic regions controlled or were informative for skin and eye color, we could predict these phenotypes simply through genetic observation. In fact, since most anthropometric (comparative human) phenotypes are determined by inherited gene sequences, much of a crime-scene DNA donor's physical information is imbedded in their DNA – we have only to figure out how to extract it. The inheritance of some phenotypes is relatively simple, and this information can be extracted with the investment of time and money. For others of more complex inheritance, the information is not currently extractable and may never be with existing technology.

There are two basic methods for inferring phenotypes from DNA. If we desire to infer phenotype from DNA using *direct methods*, we need to understand at least the major elements of the genetic architecture of the phenotype – that is, which genes and gene variants (as well as environmental factors) underlie variable expression of the phenotype. These genes are called *phenotypically active loci*, and their variants *phenotypically active variants*. With the direct method, we relate genotypes to specific phenotypes through our understanding of the genetic architecture, or if we do not fully understand the genetic architecture, through an empirical process based on prior experience and the use of databases. Of course, at least some of the genetic architecture must be understood or captured by our databases for the direct method to be possible. Even for "relatively simple" phenotypes, the expression tends to be extraordinarily complex, involving multiple variants per gene, sometimes more than one gene and sometimes even environmental factors. As basic genetics research

powered to identify these genes and factors must involve many hundreds, even thousands or tens of thousands of subjects (depending on the phenotype, the number of loci, and their frequency in the relevant populations), arriving at the requisite level of understanding is very expensive and time consuming. As a result, we can so far only use direct methods with a couple of the overt human phenotypes that have so far proven amenable – hair color, iris color and, possibly soon, skin color as well (at least within some human populations).

Another method of inferring phenotype from DNA is based not on any understanding of the genetic architecture of the phenotype, but based on the recognition that expression of the phenotype is correlated with certain elements of human population structure. This is called the *indirect method* of DNA-based phenotype inference. Indirect methods are based on an appreciation of individual genomic ancestry in terms of admixture. We choose the term *admixture*, rather than mixture, because the distribution of ancestry within populations and individuals tends toward amalgamation and the preservation of structure rather than dilutive blending and homogenization (by analogy, the colors yellow and red combining to produce an amalgamation of yellow and red, rather than orange). The correlation between elements of ancestry and phenotype within populations of individuals allow for an inference of the latter on an individual-by-individual basis. If skin color, for example, is systematically and quantitatively darker among individuals with increasing levels of African genetic ancestry, and if we have a good database of individuals of varying African ancestry and their skin color measurements, we can infer the skin color of a crime-scene donor from a precise genomic ancestry estimate of the donor's African admixture. The use of databases makes the process both objective and empirical in that the inference is based strictly on observation (as opposed to based on a model such as a model of the genetic architecture). As we describe, in this article, depending on the phenotype and its distribution among the world's various populations, we can employ the empirical method with reference databases to not only infer trait value but also to do so quantitatively and with predefined levels of confidence (based on the likelihood that the inference is correct). In what follows, we first describe the currently available methods for indirectly inferring phenotype from DNA, then those for direct inference. We then discuss some of the first cases that have employed these methods with success and close with a brief discussion on the issues associated with the penetration of these new methods into the modern forensics investigative process. Most of our discussion here is necessarily brief and significantly more details can be found in [3] which represents the first text to discuss this topic in depth.

Indirect Method of Phenotype Inference

The indirect method of phenotype inference relies on the empirical process, which itself is driven by observation. With this method, we rely on observations of correlations between phenotypes and ancestry, rather than direct relationships with gene sequences. This focuses our attention to certain phenotypes unevenly distributed among the world's various populations, which arose either because they conferred selective advantage among our ancestors in certain geographical regions or through sampling and genetic drift. Reproductive barriers such as oceans, geographical extremes, and assortative mating helped to solidify a global amalgamation of populations as and after our ancestors expanded out of Africa some 200 000 years ago (reviewed in [4, 5]).

The main phenotypes distributed as a function of genetic ancestry and of interest to the forensic investigator are those that are overt, such as skin color, certain facial features, hair/iris color, stature, etc. To infer them accurately, we thus need an accurate tool with which to measure an individual's ancestry, and we need to measure this ancestry (and phenotype) among large population of individuals. Polymorphism of the Y and mtDNA chromosomes are the gold standard for reconstructing human population histories and measuring the apportionment of genetic diversity among the world's populations (Figure 1), but these chromosomes are uniparental (e.g., the Y is inherited by males from their father, who inherited it from his father, and his father, etc.) and as such, they are of little use to the forensic scientist. For example, from measures of Y chromosome mixes among various worldwide populations, we might note that the distribution of higher eumelanin index values (darker skin colors, Figure 2) is correlated with certain Y-haplogroups such as E that dominates in present-day Africa, or the L of South Asia and that lighter eumelanin index values correlated with those

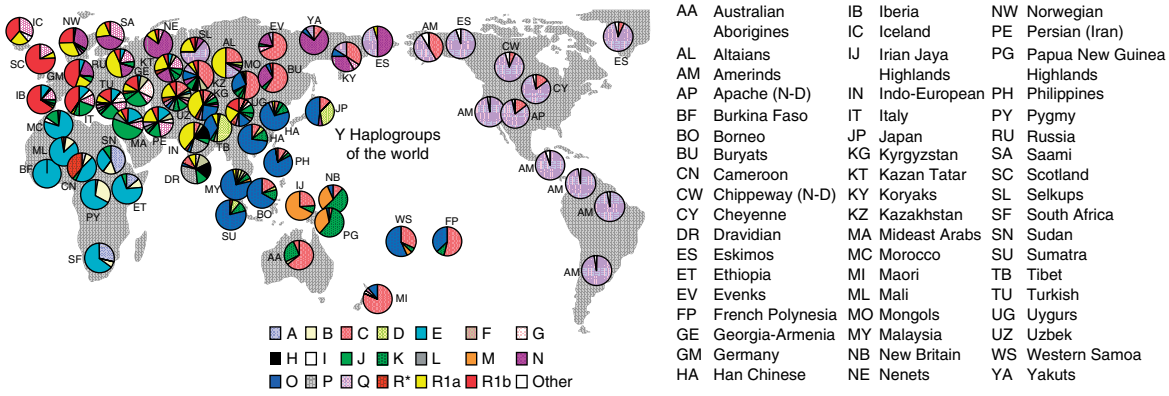


Figure 1 Global apportionment of Y haplogroup diversity. Pie charts illustrate the proportion of haplogroups as identified in the legend at the bottom. Populations are identified with a two-letter code/defined by the legend to the right [Reproduced with permission from J. D. McDonald. © 2004.]

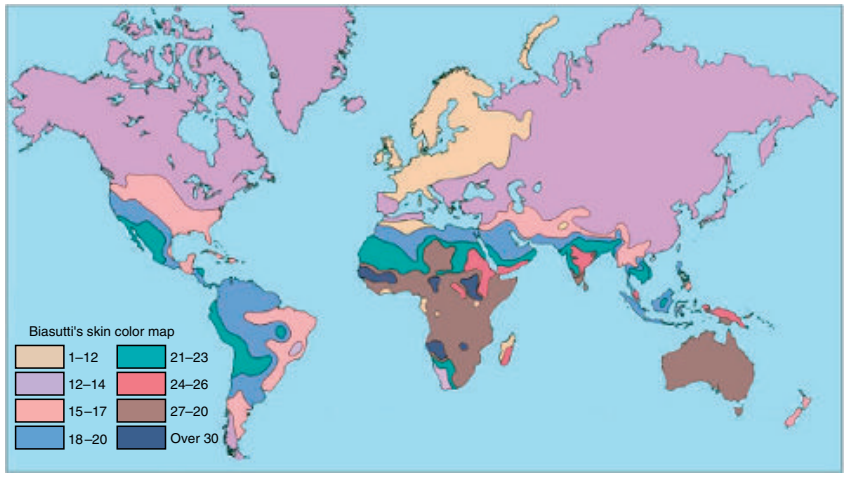


Figure 2 Geographical distribution of skin melanin levels. Higher values correspond to darker colors [Reproduced with permission from Cengage Learning © 2000.]

such as R or I shared among Europeans (compare Figures 1 and 2). Armed with this data, we might predict light European-like skin color for 28% of African Americans based on Y haplogroup sequence alone because no fewer than 28% of African Americans in the United States have a European Y chromosome as a result of recent admixture [6]; [7]; [3]. Indeed, the history of many other populations involves the admixture of those parental populations within which our anthropometric phenotype differences evolved and the ancestry for any one individual is best considered as a unique point along a sliding scale of admixture among these populations. In the field, we call this "sliding scale" a multidimensional continuum of admixture, based on the graphical methods we use to display the results (Figure 3). Uniparental chromosome polymorphisms are not useful for predicting phenotype because they do not tell us enough about the ancestry of any given individual, which is a complex function of all of the ancestors not just their patrilineal or matrilineal lines.

To indirectly infer a phenotype from the DNA of an individual, we must first have a method by which to quantitatively estimate individual genetic

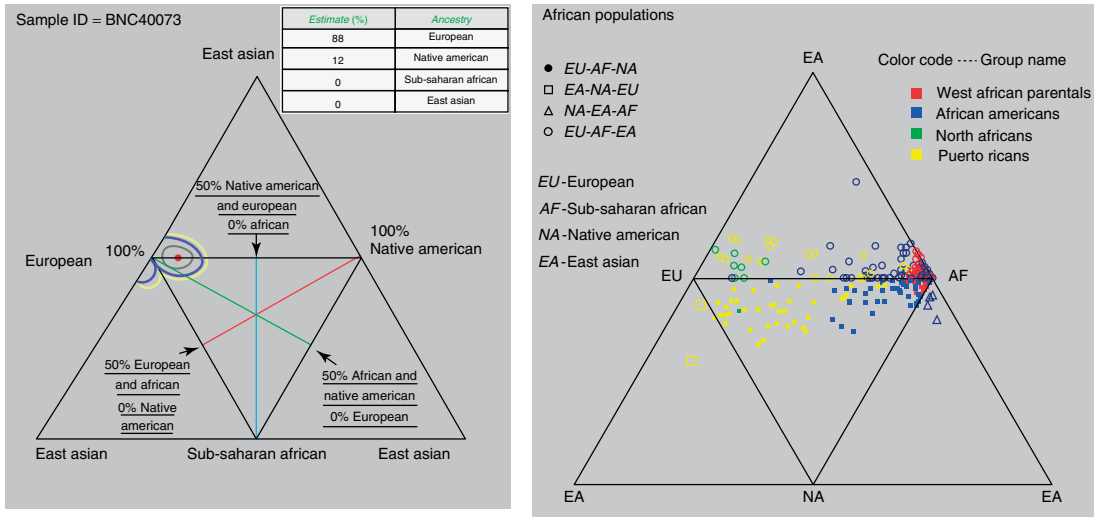


Figure 3 Individual genomic ancestry estimates portrayed with a tetrahedron plot. (a) The most likely estimate (MLE) for an individual is shown with a spot (red) which corresponds to specific percentages shown in the upper right-hand box. The percentages are obtained in four-dimensional space with an algorithm, but can be displayed on a two-dimensional piece of paper using this plot diagram; projecting the MLE spot perpendicularly on the each of the three axes within any of the four subtriangles (arrows) gives the corresponding percentage. The closer the spot is to the triangle vertices (labeled European, sub-Saharan African, East Asian, or Native American), the higher the percentage admixture corresponding to that type of ancestry. (b) Plot of numerous MLEs for individuals of color-coded self-described ancestry (legend upper right) obtained using the 176 AIM panel described in the text. Though continental Africans show predominantly African admixture, this African-American sample shows considerable European admixture and the Puerto Rican sample shows even more

ancestry admixture, contributed by all the ancestors of an individual (e.g., an individual might register with 80% "European" and 20% "African" admixture or some other mix). To do this, we measure ancestry informative markers (AIMs) on all 23 chromosomes of an individual. Secondly, we need to derive our understanding of the relationship between phenotype and ancestry from the same type of data – individual phenotype measurements and individual genomic ancestry admixture estimates. Note that we are therefore working on the level of the individual (within and/or between populations) rather than on the level of the population and so rather than using uniparental polymorphisms we must use AIMs distributed among all 23 chromosomes. Though binning or assigning an individual to a single population (such as "Caucasian") is conceptually easier than estimating the admixture of ancestry, it is not suitable for this purpose for the same reason we cannot use inferences of ancestry from uniparental chromosome polymorphisms. Specifically, we need to know what percentage of an individual's karyotype is derived from the ancestors of a given population, if we hope to compute the probability that the individual expresses a phenotype that is highly characteristic of that population.

AIMs are polymorphisms showing significant allele frequency differences between human populations. For example, an AIM may come in two "flavors" – a G allele and an A allele, with G rarely found in European or Eurasian populations but commonly found among various African populations. Most AIMs are single nucleotide polymorphisms (SNPs) such as this, with only two alleles, and SNP–AIMs constitute about 0.04% of the 2 million or so SNPs in the human genome. To construct an admixture panel, the most informative AIMs are selected from those that have been databased via the human genome (or similar other) project and the allele frequencies/frequency differentials among our founding parental populations is inferred through analysis of modern-day representative descendants. Using the AIM genotypes of any one individual, we invert the frequency of these alleles in various

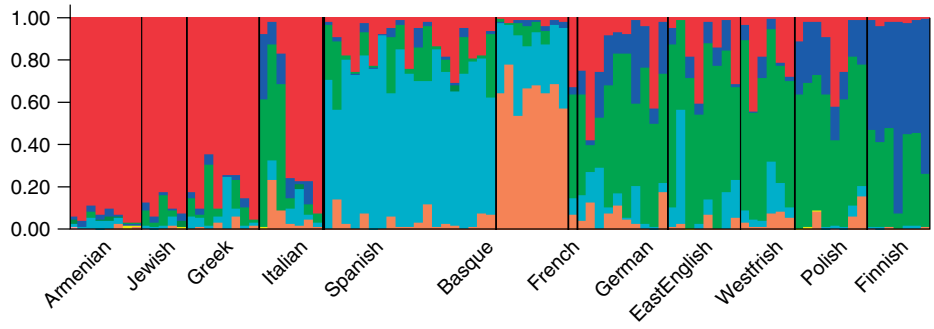


Figure 4 Individual genomic ancestry admixture estimates among Europeans with respect to a five-population European model. The 1346 European AIM panel described in the text was used to generate the data. For each individual, the proportional ancestry derived from these five parental European populations is represented with a bar, using colors coding for each ancestry type and the scaling on the left. Markers and genotype data were derived from the work published by [11] [Reproduced from Ref. 11. © Elsevier, 2007.]

populations in order to calculate the likelihood that the individual's ancestry was derived proportionally from these populations. The likelihood for all possible proportions is calculated, we select the best as the most likely estimate (MLE), and select those within two–tenfold likelihood values fall within the confidence intervals for the MLE (Figure 3a). If the admixture panel is to address the global population, worldwide population models are usually chosen on the basis of hypothesis-free clustering results (e.g., see [9]). One panel that has been well characterized incorporates a four-population model; this panel (developed in the laboratory of the author) was the first to be extensively characterized and applied for forensic cases ([3, 10]; discussed further below). Since individuals derived mainly from the Eurasian continent share one element of ancestry, we can arbitrarily name this element "European" or "Eurasian", and so on for the other three elements. Other more complex global models are possible but whatever the model, the choice of nomenclature is arbitrary, and usually based on either the modern-day origin of the parental representatives or geographical origin of the parental population (estimated from paleoarcheological, linguistics, and/or uniparental chromosome analyses). Mathematical methods allow for an accommodation of uncertainty with regard to the estimation of parental allele frequencies from modern-day representatives, as well as other pertinent variables, and the reader can refer to [3] for more details. For this discussion, here, suffice it to say that whatever the chosen population model, as long as it is based on hypothesis-free clustering results, an individual's admixture proportions are a function of genetic distance and thus potentially relevant for indirectly predicting phenotype. As of this article, three panels are available to forensic professionals through a company located in Sarasota Florida (DNAPrint genomics, Inc., see www.dnaprint.net or www.ancestrybydna.com) – the aforementioned 176-AIM continental panel (European, African, Indigenous American, and East Asian; [3, 10]), a 320-AIM Eurasian panel (Northern European, Southeastern European, Middle Eastern, and South Asian; [3]) and a 1476-AIM European panel (Figure 4; [3]).

Discerning whether particular elements of individual ancestry are correlated with certain phenotypes is accomplished through databases and regression analyses, where we plot one variable (such as a trait value) against another (such as percent admixture for a given ancestry type) and note whether or not there is a statistical dependence of the former on the latter. Regression analyses have shown correlations for skin eumelanin content and individual African ancestry levels using a 30-AIM ([12], African-American population) and the aforementioned 176-AIM panels ([10], Puerto Rican Afro-Caribbean population; Figure 5). For example, we could conclude from the regression analyses in Figure 5 that an individual with >75% African admixture determined using the 176-AIM panel is most likely to have an M (eumelanin content) value above 40 and an individual with less than 50% African ancestry is most likely to have an M value less than 40. With larger databases, we could provide not only an expected M value by

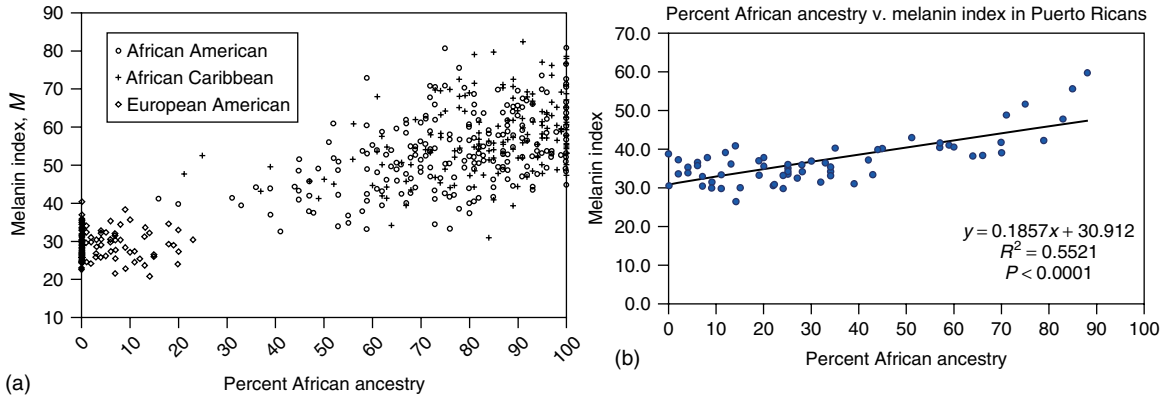


Figure 5 Regression of eumelanin value (M) from skin measurement on African individual genomic ancestry estimates in a population of (a) African Americans, European Americans, and Afro-Caribbean samples, obtained using a 30-AIM panel [12] and (b) Puerto Ricans, obtained using the 176-AIM panel described in the text [3, 10]. Each spot represents the point estimate of African admixture for an individual. Higher M values correspond to darker skin colors (higher concentration of eumelanin per unit skin area)

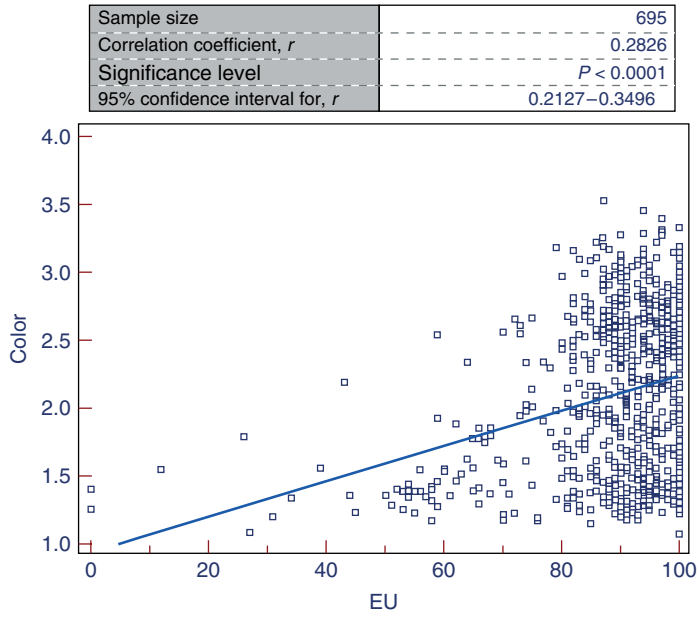


Figure 6 Regression of iris eumelanin scores from digital photographs on European individual genomic ancestry estimates in a population predominantly of self-described "Caucasians". Each spot represents a point estimate of European admixture for an individual. Higher color scores correspond to lighter colors (less eumelanin)

taking an average but we could also quantify the reliability of this value with confidence intervals. A correlation has also been demonstrated between iris color and European ancestry [13]; we can see from Figure 6 that while individuals with high "European" admixture have light (color $>$ 2.2) and darker iris color (color $<$ 2.2), individuals with substantial non-European admixture (20% or greater) almost always have darker iris colors. Thus, a crime-scene DNA sample determined to have been deposited by an

individual with 40% East Asian and 60% European admixture, or 50% African and 50% European admixture can be inferred to have an iris color on the darker end of the range observed in the human population. With larger databases, we could even quantify the likelihood that this type of conclusion is wrong. For example, if we see that the conclusion is true for 999 of 1000 individuals, we could estimate that the conclusion is correct with 99.9% certainty and that only 1 in 1000 such conclusions would be wrong. As of this article, the indirect method can only be used for these two phenotypes but through the construction and use of admixture databases, we may soon learn of others. For example, if we construct an admixture database of individuals, which included the self-described "ethnicity" and a digital photograph for each entry, and then query the database with a particular admixture profile (± a reasonable range), we might get a return of several entries like that shown in Figure 7. From these returns, we could discern whether the individuals are more likely to refer to themselves as belonging to one particular population than another, and how reliable such an inference can be expected to be. Using software biometric tools, we might learn that the average distance between the eyes (for example) is significantly different from that of a random collection of individuals, or a group of individuals with a different type of admixture profile. Some investigators may be able to use collections of digital photographs corresponding to an admixture profile to mentally compile a composite sketch, though the future promises to provide software for identifying all of the phenotypes statistically characteristic of the profile (compared to randomly specified samples) as well as the ranges of trait values, we can expect algorithms to construct an *in silico* "rendering" of the suspect similar to that provided by a human eyewitness.

A very important point that needs to be highlighted is that if our admixture methods are insensitive or inaccurate, if there are problems with the population model or parental representatives we have built our test on, or if there are glitches with the database software, there will be error. However, assuming that the size of the database is adequate, this error is expected to result in an increase rather than a decrease in entropy of the system [3]. The increase in entropy leads to a loss of information and an inability to recognize correlations in regression analyses (false negatives) rather than an ability to recognize false correlations (false positives). For example, an admixture assay based on only a few AIMs covering only

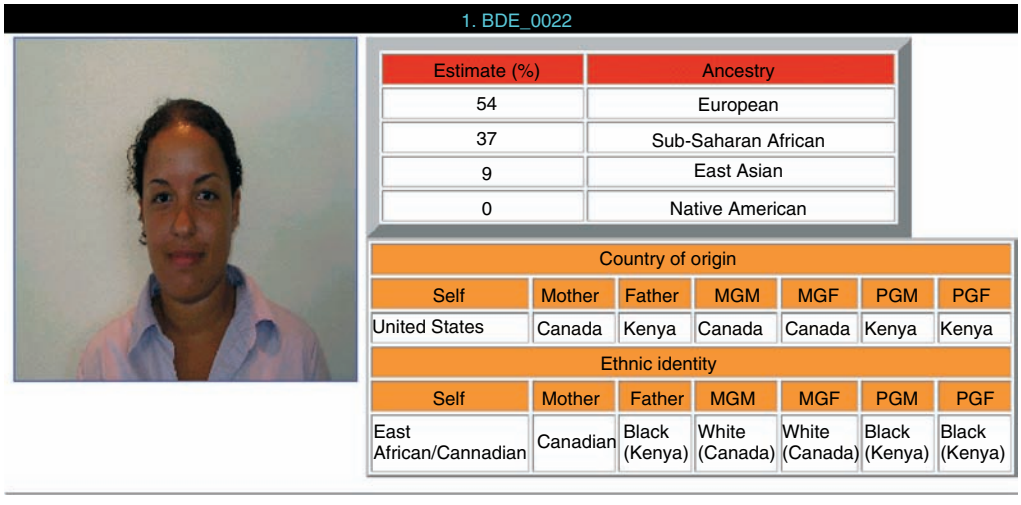


Figure 7 Example of an admixture database entry. Entries in this particular database (www.dnawitness.net) included a digital photograph taken under standardized conditions, country of origin, that of their mother, father, and their maternal grandmother (MGM) and paternal grandmother (PGM) as well as maternal grandfather (MGF) and paternal grandfather (PGF). Similarly, the self-reported "ethnic identity" is provided by each subject. The laboratory that administers this database took the photograph and determined the admixture profile with respect to a global four-population model using the 176-AIM panel discussed in the text. The database had 4700 entries as of August 2007

one chromosome would produce a large standard error in admixture estimates and the imprecise estimates may be so scattered about their true values in a regression plot that the relationship between trait value and ancestry is concealed (e.g., in Figure 5b, imagine the spots so far scattered above and below the line at low as well as high African admixture values that the sum of squared distance to a line indicating no correlation is similar to that for one indicating a good correlation). False positives can only obtain if false positive correlations are produced between phenotypes and elements of ancestry, which equates to the creation of false pattern, but random error that we would expect from a poorly performing assay would, by definition, obfuscate pattern and increase the entropy. It is possible to create false pattern from an assay that produces estimates with biased error, but with admixture/phenotype databases even these tend to result in a decrease rather than an increase in entropy. For example, consider an assay that has a tendency to erroneously estimate low levels of African admixture for South Asians (but not for other "European" populations such as Middle Eastern or Continental Europeans). It might seem that this error could lead to the mistaken conclusion that low levels of African admixture for individuals of primarily European ancestry is correlated with very dark skin color, even though most Continental Europeans with low levels of African admixture exhibit lighter skin colors (e.g., Figure 5). However, note that all of the Europeans would be present in the database, and using the high European/low African profile to query the database would result in a jumble of various light and dark skin color phenotypes rather than the shade that would be observed, were the error not present. Thus, we have an increase in entropy over which we would expect without the error, leading to an inability to make an accurate inference and false negative (type I) error rather than the creation of a false positive. Nonetheless, admixture panels that produce biased estimates are easily identified if they are properly validated and characterized [3]. A properly constructed and validated admixture panel should exhibit single-digit percentage accuracy (total error from statistical imprecision and bias combined). The laboratory of the author operates such a validated admixture/phenotype database ($n = 4700$ samples as of August 2007) and has used it in several homicide cases, but as we discuss later, so far only to indirectly infer skin color and iris color. Finding other phenotypes for which the indirect method will be useful in the future and the development of more advanced data mining and composite sketch software tools will require significantly greater investment.

Direct Method

Though they can be correlated with phenotypes, ancestry and markers of ancestry do not cause phenotypes – genes do. In contrast with the indirect method, the direct method relies on measurements of the actual genes underlying the phenotype of interest. Though far more satisfying in a theoretical sense, since we expect to be able to do a better job of predicting trait value (more accuracy and tighter confidence intervals), the direct method is rarely practicable because it requires an understanding of the dominant aspects of the genetic architecture of a phenotype, and acquiring this understanding is a very expensive and time-consuming endeavor. To date, research on the genetic basis for only human iris color and one aspect of hair color has been productive enough to enable direct phenotyping from DNA. The objective with the direct method is to define polymorphisms associated strongly enough with the phenotype that they are predictive for that phenotype with good sensitivity and specificity. Note that hundreds of good gene/phenotype associations have been described in the literature over the past decade or so (mostly for clinical phenotypes), but few of them are sufficiently strong or detailed enough within each gene to enable "genetic classification". Given the complexity of human phenotypes, useful polymorphisms are likely to be found for phenotypes where expression is controlled by one to a few genes at most, and will almost always be useful in the context of diplotypes (diploid pairs of haplotypes, where a haplotype is a chromosomal string of SNP alleles). Which diplotypes are associated with which trait values would be best determined through an empirical process of database construction and query, as with the indirect method, except we query with diplotypes rather than admixture profiles.

Iris Color

The best example of an effective direct system for inferring a human phenotype is human iris color. Linkage screens and association scans have shown

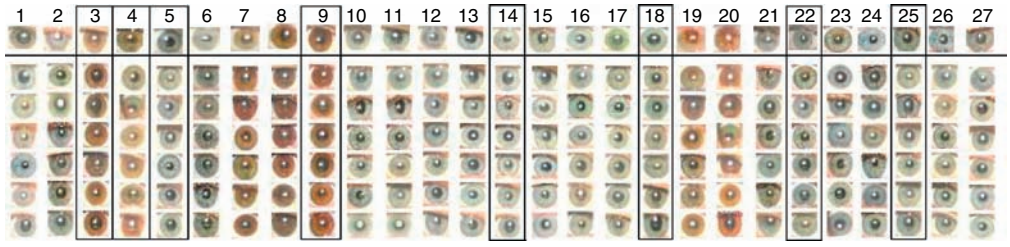


Figure 8 Examples of iris color inference enabled with the 33 marker/iris color database described in the text. Iris color is inferred using the average color exhibited by samples in the database with matching diplotypes and an interval is provided around this point estimate using the range of colors exhibited or a default range, whichever is larger. This inferred range is then used to query the database and all of the irides falling within the range are presented. This typically produces tens or hundreds of irides of similar overall color from a distance (determined by eumelanin content) though of different pattern, depending on the color and database size. Shown here are returns for 27 test subjects. Six representatives of the inferred iris color score range are provided for each of the 27 test irides below the line and the actual color of the test iris is shown above the line

that variable human iris color in humans is primarily determined by polymorphism in the Oculocutaneous Albinism 2 (OCA2) gene [13–15]. OCA2 was first discovered by researchers studying the human albino phenotype as a locus with mutations that affected iris melanin production but not skin melanin production (as well as other mutations that affect both). Recently, [13] built upon earlier reports [14–17] to identify 33 OCA2 SNPs associated with digitally quantified iris color independent from their ancestry information. As it is with many phenotypes (e.g., Figure 5), ancestry was itself correlated with the phenotype (Figure 6) and correcting the associations with respect to population structure was crucial for demonstrating that they were *bona fide* (that is, that the association was with iris color, not an element of population structure that is itself correlated with iris color; [3]). Though each of the 33 SNPs were marginally (independently) associated with iris color, none were very useful on their own as iris color classification features. However, when assembled into diplotypes, the alleles for these 33 SNPs were highly predictive for the overall eumelanin content of the iris; among 1100 diplotypes from individuals of European descent, there existed 96% concordance of iris colors among those samples with the same diplotypes [3, 13]. To predict the iris color of a given sample, this team thus built a database of phenotyped diplotypes, and then queried the database much as we do with the indirect method of phenotyping – though, in this case, with test diplotypes rather than admixture profiles. The return for this query provides an average iris color and range of iris colors as a point estimate and range of inferred color for the test diplotype. The results were satisfying; a validation sample provided a 96% accuracy rate (the inferred or predicted iris color for the unknown iris fell within the predicted range 96% of the time; Figure 8) and demonstrated that the method was capable of pinpointing the overall eumelanin content of the iris and often the particular shade, though not the pattern of iris pigmentation [13].

Particularly interesting from this work on iris color is what it teaches us to expect for other phenotypes. The predictive power of these SNPs required a consideration within the context of diplotypes. Not only were SNPs unable to provide predictive power on their own, haplotypes were equally insufficient and even diplotypes composed using smaller numbers of SNPs than these 33 (such as hap-tag SNPs) were insufficient to achieve good prediction results [13]. This illustrates the apparent historical and mechanistic complexity of even this relatively simple (predominantly single-gene) phenotype. A by-product of this complexity is the need for a massive database in order to handle most test samples; with so many SNPs part of the equation, the number of diplotypes in the human population is very large and the chances that a test sample from a crime scene would have a match in the database at its current size ($n =$ 1100, Summer, 2007) is about 10%. Nonetheless, the inferences for these 10% are highly accurate and based on this utility, the iris color diplotype database system has been developed as a forensic service by DNAPrint genomics under the trade name Retinome™ (DNAPrint genomics is the laboratory within which the author conducted much of the work

described here and throughout this article). So far the method has been applied to several homicide cases; however, due to the currently small size of the database, many detectives desiring to use Retinome™ have been unable to do so because their crime-scene samples did not have a match in the database. For this method of predicting iris color to have more of a broad impact on the investigative process, the size of the database will have to be increased substantially.

In spite of its limitations, the Retinome™ system was the first system for the direct inference of a complex human phenotype. Since its introduction on the lecture circuit in 2004, other OCA2 systems involving additional single nucleotide polymorphisms (SNPs) were described [23], and eventually, a single SNP was discovered in 2008 [24–26] that was so powerfully associated with the light/dark iris color dichotomy that it was posited to represent no less than a founder mutation – an evolutionarily instrumental mutation that not only explains but also represents the historical genesis of the crude light versus dark dichotomy extant throughout the world today. The strength of the association (nearly perfect, with respect to the dichotomy), functional studies with cultured melanocytes, the universality of the association (the C allele associated with lighter colors in individuals from around the world), and the location of the SNP in an important regulatory region of the OCA2 gene all combined to suggest that this hypothesis is indeed true. Although the SNP is only useful for crude predictions (lighter versus darker colors), the founder status of the mutation meant that prediction was suddenly possible for all samples – not merely a small fraction. Indeed, it is likely that some of the previously described SNPs were useful as components of a predictive tool through linkage with this founder mutation. However, others are likely to represent important pieces of the final solution for predicting precise colors and shades with economy and practicality. For example, one of the 33 Retinome™ SNPs (rs1800407) was demonstrated to be a penetrance modifier of the newly discovered founder mutation, and a minimalist set of OCA2 polymorphisms for the "ultimate" in forensics utility (precise color, perhaps pattern as well, for all crime-scene samples) will likely require a composite diplotype system involving the founder mutation and some number of the previously described SNPs in [3] and [23]. This "ultimate" system has not yet been developed, though integration of the founder mutation with those of the Retinome™ system in a newly available forensic service called Retinome™ 2.0 (DNAPrint genomics, Inc., Sarasota, Florida) represents a step toward this goal.

Hair Color

Hair color cannot yet be comprehensively predicted from pigmentation gene genotypes. The iris color polymorphisms just described are not associated with human hair color and association and linkage scans have so far been relatively fruitless in identifying other useful associations ([3]; Zhu G. and Martin N., Queensland Institute of Medical Research, Brisbane AU, personal communication, and T. Frudakis unpublished results). This may be due to the fact that unlike the crudest aspect of iris color – the quantity of eumelanin in the iris – the genetic basis for hair color is a function of significant locus heterogeneity and complex historical origin. However, it appears that red color may be an exception. Valverde *et al.*, [27] was the first to identify MC1R associations with pheomelanogenic red hair color (RHC), and subsequently, several other authors have extended these results to identify what are today called the *RHC* phenotype alleles (all SNPs, Box *et al.*, [28]; Duffy *et al.*, [29]; Smith *et al.*, [30]; Palmer *et al.*, [31]; Box *et al.*, [32]; Bastiaens *et al.*, [33]; Bastiaens *et al.*, [34]; Kennedy *et al.*, [35]; Flanagan *et al.*, [36] and reviewed by Sturm, [37]). The associations are sufficiently strong to enable good predictive power – with odds ratios ranging from 2.3 to over 100 (reviewed in Sturm, 2002; [3]). The United Kingdom's forensic science service (FSS) has condensed the major MC1R red-hair polymorphisms into a 12-marker test that is sold to the forensics community. This test is generally only useful in cases of homozygosity; individuals who are homozygous for any of these mutations, or heterozygous for any two separate mutations (called *compound heterozygotes*) are accurately predicted to be redheaded (accuracy = 96%, from a test with $n = 48$ subjects) and those without a mutation (homozygous wild type) are almost always not redheaded (accuracy = 100%, from a test with $n = 35$ subjects; [18]). Approximately 84% of redheads are detectable by these criteria. Predictive ability is lower in the case of simple heterozygotes (only one mutation present, in the heterozygous state), which is by definition the state in which most of these RHC SNP alleles are expected to be found, in which case 88% are not redheaded while 12% are redheaded ($n = 33$). Indeed,

we might expect that each of these mutations will have varying influences on the phenotype depending on the MC1R context within which it is found, and application of diplotype databases as described for iris color may be helpful in teasing more predictive power.

Other Phenotypes

Databases are not yet available for the direct inference of skin color, but this may change soon. [12] and Bonilla *et al.*, [38] used a process called *admixture mapping* to identify variants in the OCA2, MATP, ASIP, and TYR pigmentation genes associated with skin color in a manner that is independent from their ancestry information content. None of these appear to be strong enough to enable accurate prediction, but [19] convincingly described additional variants of the SLC24A5 gene that underlie additional variable skin pigmentation in humans (as well as other vertebrates). Diplotypes involving SNPs in these five genes may be sufficient for accurate inference of skin color within the context of the database systems we have been discussing, though such as database has not yet been constructed. Promising markers for human stature have also been identified in the RUNX2 gene [20] and RANK gene [21], but its unclear whether these two genes are sufficient for predicting this phenotype (especially since environment is likely to play such a role) and RUNX2/RANK variant databases for this phenotype have also not yet been constructed.

Ethical/Procedural Issues and Case Studies. Were the STR profile for each of the world's human inhabitants deposited into an international forensic DNA database, there would be no need to infer phenotype from crime-scene DNA since we would always be able to achieve a database match. Owing to ethical and procedural concerns, this is unlikely to come to pass in the foreseeable future. The ethical issues surrounding the inference of phenotype are extensively covered in [3], but for our purposes here, we can simply note that there is not a fundamental difference between learning about phenotype from human eyewitnesses versus DNA, except when DNA is available, it is more likely to provide reliable and falsifiable information. At some point, ethicists that decry the use of DNA for predicting phenotype as part of the investigative process will have to choose between the rights of DNA donors to remain anonymous and uncharacterized and the rights of future victims of these donors not to be future victims. Phenotype profiles may cause the inclusion of innocent individuals into suspect pools, which would cause inconvenience to these individuals but not increase the likelihood of a false conviction (their CODIS profile must still match that of the crime-scene sample, the probability of which is not reasonably a function of whether or not they are tested using the current state-of-the-art CODIS assays, and is irrespective of their phenotype). It could be argued that many such individuals would or should become suspects for other reasons – such as a life of crime, proximity to the crime scene and/or relation to the victim – and that information about phenotype merely hones attention to a subset of these individuals. Indeed, many investigators consider these other reasons adequate priors (bases) on their own for defining who is and is not of interest in an investigation and honing attention to a subset of them with DNA-based phenotype information could reduce unnecessary inconvenience for many not fitting the phenotype "profile". Notwithstanding, weighted against the death of innocent victims that could be caused by not applying intelligence provided by DNA-based methods, inconvenience caused to innocent individuals subsequently proven not to match the CODIS profile of a given crime-scene sample pales toward insignificant in comparison. Indeed, we seem to have made the decision to use physical information to shape investigations already – as it pertains to human eyewitness testimony; even with all of its pitfalls, its subjectivity, and poor performance rate, eyewitness testimony currently represents one of the most important cornerstones of the investigative process. The use of DNA-based methods promise only to improve this performance using DNA as an additional and/or alternative source for information, at least for cases where DNA is available. Even so, ethicists have complained that DNA-based phenotyping methods are akin to "racial profiling", but the difference between using physical information gleaned from a crime-scene specimen and "racial profiling" is the application of falsifiable science rather than prejudice. For example, focusing on an individual based on information extracted from a crime scene is an exercise based on evidence and data and the conclusions are falsifiable by other laboratories. In contrast, focusing on an individual based on a belief

that individuals of "group X" are more likely to be criminals – lacking specific data derived from a crime scene – is based more on prejudice and subjectivity than the scientific method and generally speaking, such correlations are not adequately powerful in a predictive sense to constitute meaningful priors when determining the likelihood of an individual's involvement in a crime.

The methods described in this article have been applied to numerous criminal investigations. The first application of the indirect method was for the Louisiana Multiagency Homicide Task Force Investigation (Louisiana Serial Killer Case) in the spring of 2003 [3, 22]. Investigators had adequate DNA from various rape/murder scenes throughout the state, but without a CODIS match, they were forced to target their investigation based on two eyewitness accounts that subsequently proved irrelevant to the case. Over a year passed with the task force looking for a "Caucasian" male (not only with standard investigative practices but also with DNA sampling dragnets) until a 73-AIM version of the DNAPrint 176-AIM genomic ancestry panel described in this article (which was also sold under the trade name DNAWITNESS™ as version 2.0) was applied and indicated that the donor was an individual of primarily African ancestry (85%) with a small amount of Indigenous American admixture (15%). In this particular case, the lack of European admixture was used to infer a relatively dark skin shade with respect to the average African-American in the United States (Figure 5a). The investigation was refocused with this data and within a couple of months, the newly refocused investigation lead to an ex-con in the area fitting this profile whose CODIS STR profile was subsequently matched to the crime scene. Were it not for the application of the phenotyping methods, the investigation would have continued on its misdirected path, and others would likely have been raped and murdered [22].

Another murder case in Napa California was initially focused on Hispanic suspects based on an eyewitness account. Napa detectives applied the 176-AIM panel discussed in this article and obtained a continental admixture result that was found from DNAPrint genomics' DNAWITNESS 2.5 database to be consistent with individuals of Continental European as well as Middle Eastern descent – not Hispanic. They then applied the DNAPrint Eurasian panel of 320 AIMs and the RETINOME™ iris color panel described earlier in this article and learned from database searches that the individual was most likely of Northern European ancestry and of light-colored irides. The investigation was focused appropriately, and the perpetrator of the crime in this particular case (who was of Northern European descent with blue eyes) was eventually identified and linked to the crime scene via the CODIS profile. As with human eyewitness data, DNA-based phenotype information does not always result in quick arrests of course, and of the hundred or so investigations that have employed the methods described herein, the majority remain open. However, even for these cases, investigators have saved money and time that would have been spent investigating individuals with phenotypes very different from the crime-scene DNA donor.

To advance the field, the SNP and AIM associations and databases currently available need to be amplified so that an inference for every sample can be obtained, and more research is needed so that additional phenotypes can be considered. In the near future, this may be an uphill battle, at least in the United States. For example, many US grant-funding agencies have not embraced much of the work described in this article – particularly those associated with indirect methods of phenotype inference and it seems most likely that the task of expanding this field will likely to be left to private enterprise or public laboratories outside the United States. Indeed, the work described in this article was funded with private capital, and though commercial demand for the products has not so far economically justified the investment, the work has at least provided some public service and could prove just as useful for other more commercially lucrative areas of research such as in drug development (where constructing a portrait of an ideal patient for a given drug could have a significant impact on the likelihood of clinical trial success). The institutional resistance in the United States toward the type of work discussed in this article is interesting and deserves some thought here. Some argue that the US justice system is not built for efficiency and is biased toward the interests of the accused. If true, this may explain why US budgets for solving crimes are often limited such that US investigators have difficulty funding basic CODIS processing of their crime-scene samples (which should always be done first since identity testing provides probative, not merely presumptive results). For example, backlogs of a year or more

and anecdotal reports of rape kits stacked to the ceiling awaiting funding for CODIS processing are not uncommon in the United States. For agencies experiencing such backlogs, budget allocation for phenotyping is likely to remain de-prioritized and until phenotyping methods have had more time to penetrate the field through continued demonstration of utility. Until the problems underlying the CODIS backlogs have been solved, the application of DNA-based phenotyping methods is likely to continue on a case-by-case basis, with emphasis on high-profile cases that investigators are under unusual pressure to quickly solve (such as serial homicide cases).

Acknowledgments

I would like to thank all of the volunteers who provided their informed consent to be part of our forensic databases here at DNAPrint genomics, Inc.

End Notes

[a.] Though, since all loci carry some ancestry information, calculations on the statistical certainty of such a match requires the use of appropriate population databases since they are based on allele frequencies that vary subtly from population to population.

References

[1] Theodoridis, S. & Koutroumbas, K. (1999). *Pattern Recognition*, Academic Press Publishers, London.

[2] Robert, C. (2001). *The Bayesian Choice*, 2nd Edition, Springer-Verlag New York Inc., New York.

[3] Frudakis, T., Terravainen, T. & Thomas, M. (2007). Multilocus OCA2 genotypes specify human iris colors, *Human Genetics*.

[4] Cavalli-Sforza, L. & Bodmer, W. (1999). *The Genetics of Human Populations*, Dover Publications, Mineola, NY.

[5] Jobling, M., Hurles, M. & Tyler-Smith, C. (2004). *Human Evolutionary Genetics. Origins, Peoples and Disease*, Garland publishing, New York, NY.

[6] Lind, J., Hutcheson-Dilks, H., Williams, S., Moore, J., Essex, M., Ruiz-Pesini, E., Wallace, D., Tishkoff, S., O'Brien, S. & Smith, M. (2007). Elevated male European and female African contributions to the genomes of African American individuals, *Human Genetics* **120**(5), 713–722.

[7] Chakraborty, R. (1986). Gene admixture in human populations: models and predictions *Yearbook of Physical Anthropology* **29**, 1–43.

[8] Jurmain, R., Nelson, H., Kilgore, L. & Trevathan, W. (2000). *Introduction to Physical Anthropology*, Wadsworth/Thomson Learning, Belmont, CA.

[9] Rosenberg, N., Pritchard, J., Weber, J., Cann, H., Kidd, K., Zhivotovsky, L. & Feldman, M. (2002). Genetic structure of human populations, *Science* **298**, 2381–2385.

[10] Halder, I., Shriver, M., Thomas, M., Fernandez, J. & Frudakis, T. (2007). A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: utility and applications, *Human Mutation* **29**(5), 648–658.

[11] Bauchet, M., McEvoy, B., Pearson, L., Quillen, E., Sarkisian, T., Hovhannesyan, K., Deka, R., Bradley, D. & Shriver, M. (2007). Measuring European population stratification using microarray genotype data, *American Journal of Human Genetics* In press.

[12] Shriver, M., Parra, E., Dios, S., Bonilla, C., Norton, H., Jovel, C., Pfaff, C., Jones, C., Massac, A., Cameron, N., Baron, A., Jackson, T., Argyropoulos, G., Jin, L., Hoggart, C., McKeigue, P. & Kittles, R. (2003). Skin pigmentation, biogeographical ancestry and admixture mapping, *Human Genetics* **112**(4), 387–399.

[13] Frudakis, T. (2007). *Molecular Photofitting: Predicting Phenotype and Ancestry from DNA*, Elsevier Academic Press Publishers, Burlington, MA.

[14] Zhu, G., Evans, D., Duffy, D., Montgomery, G., Medland, S., Gillespie, N., Ewen, K., Jewell, M., Liew, Y., Hayward, N., Sturm, R., Trent, J. & Martin, N. (2004). A genome scan for eye color in 502 twin families: most variation is due to a QTL on chromosome 15q, *Twin Research* **7**, 197–210.

[15] Frudakis, T., Thomas, M., Gaskin, Z., Venkateswarlu, K., Chandra, S., Ginjupalli, S., Gunturi, S., Natrajan, S., Ponnuswamy, V. & Ponnuswamy, K. (2003). Sequences associated with human iris pigmentation, *Genetics* **165**, 2071–2083.

[16] Rebbeck, T., Kanetsky, P., Walker, A., Holmes, R., Halpern, A., Schuchter, L. *et al.* (2002). P gene as an inherited biomarker of human eye color, *Cancer Epidemiology Biomarkers and Prevention* **11**(8), 782–784.

[17] Sturm, R. & Frudakis, T. (2004). Eye colour: portals into pigmentation genes and ancestry, *Trends in Genetics* **20**(8), 327–332.

[18] Grimes, E., Noake, P., Dixon, L. & Urquhart, A. (2001). Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype, *Forensic Science International* **122**, 124–129.

[19] Lamason, R., Mohideen, M., Mest, J., Wong, A., Norton, H., Aros, M., Jurynec, M., Mao, X., Humphreville, V., Humbert, J., Sinha, S., Moore, J., Jagadeeswaran, P., Zhao, W., Ning, G., Makalowska, I., McKeigue, P., O'donnell, D., Kittles, R., Parra, E., Mangini, N., Grunwald, D., Shriver, M., Canfield, V. & Cheng, K. (2005). SLC24A5, a putative cation exchanger,

affects pigmentation in zebrafish and humans, *Science* **310**(5755), 1782–1786.
[20] Ermakov, S., Malkin, I., Kobyliansky, E. & Livshits, G. (2005). Variation in femoral length is associated with polymorphisms in RUNX2 gene, *Bone* **38**(2), 199–205.
[21] Chen, Y., Xiong, D.H., Yang, T.L., Yang, F., Jiang, H., Zhang, F., Shen, H., Xiao, P., Recker, R.R. & Deng, H.W. (2007). Variations in RANK gene are associated with adult height in Caucasians, *American Journal of Human Biology* **19**(4), 559–565.
[22] Stanley, S. (2006). *An Invisible Man: The Hunt for a Serial Killer Who Got Away with a Decade of Murder*, Berkley Books, New York.
[23] Duffy, D.L., Montgomery, G.W., Chen, W., Zhao, Z.Z., Le, L., James, M.R., Hayward, N.K., Martin, N.G. & Sturm, R.A. (2007). A three-single-nucleotide polymorphism haplotype in intron 1 of OCA2 explains most human eye-color variation, *American Journal of Human Genetics* **80**(2), 241–252.
[24] Eiberg, H., Troelsen, J., Nielsen, M. Mikkelsen M. (2008). Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression, *Human Genetics* **123**(2), 177–187.
[25] Sturm, R.A., Duffy, D.L., Zhao, Z.Z., Leite, F.P., Stark, M.S., Hayward, N.K. & Martin, N.G., Montgomery, G.W. (2008). A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color, *American Journal of Human Genetics* **82**(2), 424–431.
[26] Kayser, M., Liu, F., Janssens, A.C., Rivadeneira, F., Lao, O., van Duijn, K., Vermeulen, M., Arp, P., Jhamai, M.M., van Ijcken, W.F., den Dunnen, J.T., Heath, S., Zelenika, D., Despriet, D.D., Klaver, C.C., Vingerling, J.R., de Jong, P.T., Hofman, A., Aulchenko, Y.S., Uitterlinden, A.G., Oostra, B.A. & van Duijn, C.M. (2008). Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene, *American Journal of Human Genetics* **82**(2), 411–423.
[27] Valverde, P., Healy, E., Jackson, I., Rees, J., Thody, A. (1995). Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans, *Nature Genetics* **11**, 328–330.
[28] Box, N., Wyeth, J., O'Gorman, I., Martin, N. & Sturm, R. (1997). Characterization of melanocyte stimulating hormone variant alleles in twins with red hair, *Humman Molecular Genetics* **6**, 1891–1897.
[29] Duffy, D., Box, N., Chen, W., Palmer, J., Montgomery, G., James, M. *et al.*, (2004). Interactive effects of MC1R and OCA2 on melanoma risk phenotypes, *Human Molecular Genetics* **13**(4), 447–461.
[30] Smith, R., Healy, E., Siddiqui, S., Flanagan, N., Steijlen, P., Rosdahl, I. *et al.*, (1998). Melanocortin 1 receptor variants in an Irish population, *The Journal of Investigative Dermatology* **111**, 119–122.
[31] Palmer, J., Duffy, D., Box, N., Aitken, J., O'Gorman, L., Green, A., *et al.*, (2000). Melanocortin-1 receptor polymorphisms and risk of melanoma: Is the association explained solely by pigmentation phenotype?, *American Journal of Human Genetics* **66**(1), 176–186.
[32] Box, N., Duffy, D., Irving, R., Russell, A., Chen, W., Griffyths, L. *et al.*, (2001). Melanocortin-1 receptor genotype is a risk factor for basal and squamous cell carcinoma, *The Journal of Investigative Dermatology* **116**, 224–229.
[33] Bastiaens, M., ter Huurne, J., Kielich, C., Gruis, N., Westendorp, R., Vermeer, B. *et al.*, (2001a). The melanocortin-1-receptor gene is the major freckle gene, *Human Molecular Genetics* **10**(16), 1701–1708.
[34] Bastiaens, M., ter Huurne, J., Kielich, C., Gruis, N., Wetendorp, R., Vermeer, B. *et al.*, (2001b). Melanocortin-1 receptor gene variants determine the risk of nonmelanoma skin cancer independently of fair skin and red hair, *American Journal Human Genetics* **68**(4), 884–894.
[35] Kennedy, C., ter Huurne, J., Berkhout, M., Gruis, N., Bastiaens, M., Bergman, W. *et al.*, (2001). Melanocortin 1 receptor (MC1R) gene variants are associated with an increased risk for cutaneous melanoma which is largely independent of skin type and hair color, *The Journal of Investigative Dermatology* **117**(2), 294–300.
[36] Flanagan, N., Healy, E., Ray, A., Philips, S., Todd, C., Jackson, I. (2000). Pleitotropic effects of the melanocortin 1 receptor (MC1R) gene on human pigmentation, *Human Molecular Genetics* **9**, 2531–2537.
[37] Sturm, R. (2002). Skin colour and skin cancer – MC1R, the genetic link, *Melanoma Research* **12**(5), 405–416.
[38] Bonilla, C. Parra, E., Pfaff, C., Dios, S., Marshall, J., Hamman, R. *et al.*, (2004). Admixture in the Hispanics of the San Luis Valley, Colorado and its implications for complex trait gene mapping *Annals of Human Genetics* **68**, 139–153.

TONY FRUDAKIS

Phosphatase: Acid *see* Acid Phosphatase

Photography: Length Measurement *see* Length Measurement

Photography: Marks, Impressions, and Documents

Introduction

Photography plays a pivotal role in criminalistics by providing several key functions including nondestructive methods of detection, recording, preservation, and enhancement of physical evidence. Standard operating procedures for most forms of physical evidence require the evidence to be photographed either at the crime scene or in the forensic laboratory. In many cases involving crime scene investigation, the photographs of physical evidence obtained at the scene become the only form in which that evidence exists. The preservation of perishable forms of physical evidence places a higher emphasis on photographic methods to retain the evidence long after the scene has changed or the physical evidence has perished.

Some forms of crime scene photographs also under go a process of comparative analysis made directly from the photographic source. Evidence such as bloodstain patterns, footwear impressions, fingermarks, toolmarks, and tire impressions all require a forensic examination made directly from the images. These photographs become the primary source [1] of the forensic analysis, and the value of the physical evidence is a function of the quality and accuracy of the photographic evidence.

The application of photography certainly covers a broad range of activities and some specific taxonomy needs to be developed when attempting to model photography's purpose in the forensic sciences [2]. This section discusses the central technical aspects of photography when used in criminalistics. Information regarding basic camera operation or techniques for simplistic recording of items found at a crime scene are not provided in this section. It will, however, examine the aspects of photographing evidence that utilize photographs in the analysis of evidence by forensic criminalists. Aspects including accurate recording methods, specialized techniques, and optical enhancement of evidence are discussed.

Currently, photography can be divided into two separate forms: silver halide (film) photography and digital photography. Both the forms of photography are currently practiced in the forensic science domain and are considered the same in this section. The light-sensitive recording mechanisms are the most obvious differences between each technology; however, camera operation remains relatively the same (with some exceptions).

Digital imaging is a different concept that involves using digitized images on a computer platform and a distinction is made between digital photography and digital imaging. Digital photography is the practice of photography using a digital camera, while digital imaging is the processing or alteration of digital images using image editing software programs or the digitization of a photograph using peripheral computer equipment such as a scanner.

The practice of digital photography and digital imaging are naturally implicit. The application of image editing software (i.e., Adobe Photoshop™) is also an integral element within the overall digital photography workflow. It is important to recognize, however, that the improvement capacity of images using image editing software does not substitute for consistent quality photography techniques at the capture stage.

This section presents information regarding the optical enhancement of evidence in three separate components: (i) maintaining the image integrity, (ii) optical enhancement, and (iii) digital imaging enhancement.

Image Integrity

The debate regarding the legitimatization of digital images in law enforcement and forensic science has been an exhaustive process by several law enforcement working parties, committees, and organizations over the years [3–5]. Legal organizations have generally accepted the new technology and have adopted their own forms of operating procedures when using digital photography. Like all modes of physical evidence, continuity of the evidence is a critical component of the management of the evidence integrity. Photography should be not different to any other mode of evidence and sound procedural policy regarding forensic photography practices should be developed and maintained throughout all the forensic organizations. It is, however, not the function of this section to discuss those procedures.

Image integrity also refers to the ability of the image to represent the aspects of the photographic evidence accurately, reliably, and truthfully. The integrity of the photography evidence is a paramount consideration in criminalistics. The following attributes are considered in relation to image integrity for the photography of physical evidence that requires analysis directly from those photographs. More simplistic applications of photography to record the subjects *in situ* may not require such strict technical parameters.

Dimensional Integrity

The reproduction of three-dimensional subjects into accurate two-dimensional photographs is fraught with difficulty by the very nature of this dimensional transition. The need for accurate dimensional representation is, however, a consideration when examining evidence such as fingermarks, footwear impressions, toolmarks, and bloodstain pattern evidence. Conveniently, these evidence forms are mostly represented as two-dimensional (or very close to it) and photography can reproduce a fair representation in this dimensional exchange (two-dimensions into two-dimensions). Accurate reproduction in this situation is not, however, automatic and requires some technical considerations to exclude the aspects of image distortion. Figure 1 shows four photographs taken of the same subject, a wire frame. It demonstrates how different the same object may be represented in a photograph and the importance of accurate photography. The image in Figure 1(a) has correctly maintained the subject's dimensional integrity, while those in Figure 1(b–d) show keystoning or optical distortion.

Forensic photographs that are used in comparative analysis must maintain the integrity of the subject's dimensional aspects. Images that do not represent the subject's dimensions and shape accurately are considered to be distorted. Image distortion is caused by several visual conditions including: (i) the camera viewpoint in relation to the subject (perspective distortion), (ii) the optical characteristics of the lens (curvilinear distortion), (iii) the dimensional stability of the recording material (film), and (iv) incorrect resizing of images in digital imaging software.

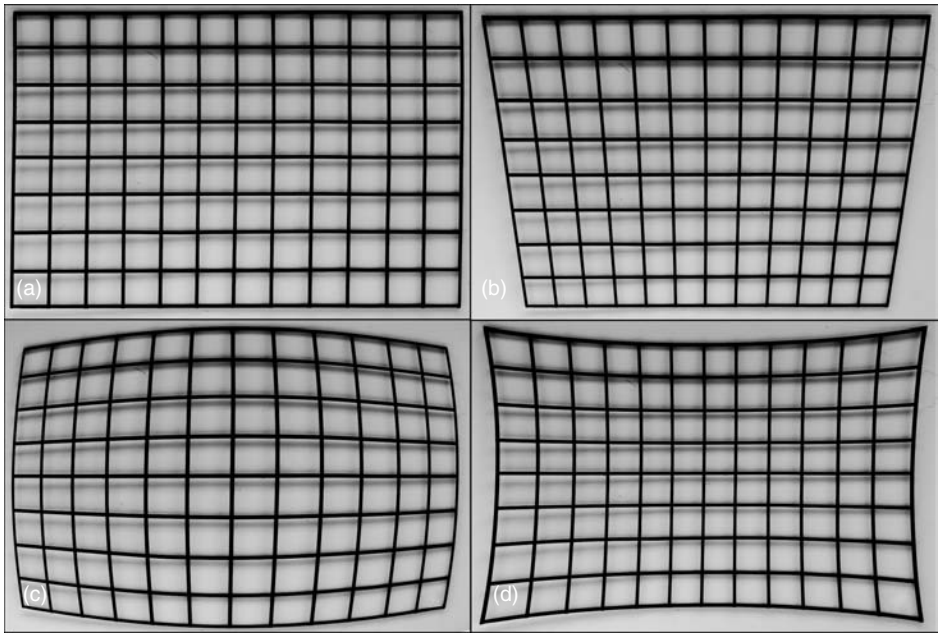


Figure 1 Four photographs of the same object displaying different representations of image dimensional integrity: (a) correctly photographed displaying good image dimensional integrity; (b) image displaying perspective distortion or keystoning; (c) image displaying barrel distortion; and (d) image displaying pincushion distortion

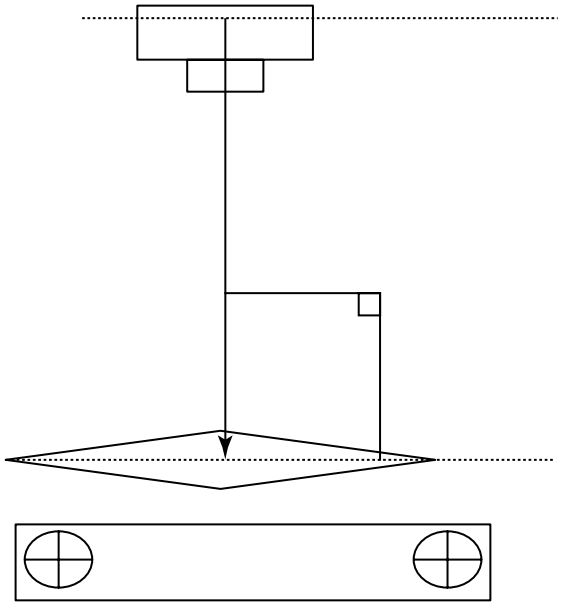


Figure 2 Camera position in relation to the subject. The lens axis is perpendicular to the subject and film planes while the film plane and subject planes are parallel

There are two essential requirements for the successful photography of physical evidence while maintaining the dimensional integrity of the evidence: (i) position the camera correctly to avoid perspective distortion (keystoning) and (ii) always use a distortion-free lens (Hard return). The camera viewpoint must be directly over the subject, which means the lens axis is perpendicular to the subject plane and the camera focal plane and subject plane are parallel [6]. Figure 2 illustrates the camera position in relation to the subject to avoid introducing perspective distortion while maintaining the integrity of the subject's dimensional aspects.

Several lens designs suffer from an optic aberration called *curvilinear distortion* [7], which will also corrupt the integrity of the image dimensional qualities. Curvilinear distortion is often described by the visual result that appears in the image and is referred to as either *barrel distortion* or *pincushion distortion* [8, 9] (see Figure 1). These effects are indicative of their descriptions and are caused by variations of image magnification across the field [1]. Curvilinear distortion can be minimized by using a symmetrical or quasi-symmetrical lens design. Lenses that are significantly asymmetrical in design, such as telephoto and retrofocus lenses, tend to suffer appreciably from curvilinear distortion [8]. Zoom lens are also prone to distortion aberration and generally display pincushion distortion at longer focal lengths and barrel distortion at shorter focal lengths [8, 10]. Lens designs that are distortion free are called *orthoscopic lenses* [8, 11]. A critical aspect of photographing physical evidence and maintaining the dimensional integrity of the evidence is to select a lens that is free from curvilinear distortion.

Macro lenses are designed for photography at shorter working distances than infinity (∞) or for a magnification range of 0.1–1.0 times [12]. They are also either symmetrical or quasi-symmetrical in design and are considered as highly corrected for curvilinear distortion aberration. All photography of physical evidence that requires a high standard of image dimensional integrity should be taken with a macro lens and never with a telephoto, wide angle retrofocus, or zoom lens. This aspect of forensic photography quality is imperative for evidence such as fingermark impressions, footwear impressions, physical fit evidence, toolmark evidence, and other forms of evidence requiring the examination of photographs. This aspect is naturally not as important for general recording of evidence such as *in situ* images at crime scenes. Zoom lenses are very useful for general crime scene photography but not for more critical forensic photography.

Some digital imaging software provides the facility to correct lens distortion. Even though lens distortion may be corrected using digital imaging software such as Adobe Photoshop™ [13, 14], digital correction of this image artifact should only be applied when it is absolutely necessary and should not replace the standard photography practices that avoid lens distortion. Prevention is better than the cure and it is certainly more preferable to avoid curvilinear distortion by using a lens that is considered as distortion free or orthoscopic in the first instance.

Representation of Scale

The incorporation of linear scales into physical evidence photographs is an essential practice when recording evidence. There are various linear scales available for different types of evidence. Linear scales are used as a reference to the size of the subjects photograph and this reference is used to enlarge the photographs to predetermined magnifications for

comparative analysis examination (1 : 1 for footwear impressions, 5 : 1 for fingerprint impressions, etc.). Linear scales are also used to calibrate the scale of digital images when using image analysis applications. Calibration is achieved when the amount of pixels are counted across a known length present in the image. Each pixel then represents a lineal value and several mathematical functions may be applied to the image.

Important considerations when incorporating linear scales include the following: (i) use an appropriate sized scale for the type of evidence photographed, (ii) make sure the scale does not hide the features of the evidence, and (iii) the position of the linear scale must be parallel to the subject and at the same height. Some linear scales also incorporate circular references to detect any perspective distortion that may have resulted from poor photography technique. Digital correction of perspective distortion may also be conducted using Photoshop™ and using the circular references as a standard.

Image Quality

The quality of the image is naturally an important aspect of recording evidence that will undergo further examination and analysis. Image quality for forensic purposes may be defined by several parameters including image sharpness or resolution, detail or clarity, image contrast, color fidelity and dynamic range, and dimensional integrity. Table 1 provides a list of those parameters and also describes what influences these considerations of image quality.

The maintenance of image quality needs to be carefully considered and embedded into the forensic practitioners working practice. Image quality forms a critical basis for forensic photography and is an essential requirement before further optical and digital enhancement techniques may be applied. Digital enhancement is not a correction for poorly executed forensic photography. Evidence enhancement techniques build on the foundation of sound image quality and permit further visualization and the application of forensic evidence.

Table 1 Image quality parameters and their causes

Quality parameter	Cause
Image sharpness and resolution	i) Lens focus ii) Depth of field iii) Camera shake and shutter speeds iv) Quality of optics
Detail and clarity	i) Image sharpness ii) Shadow and highlight detail controlled by camera exposure iii) Lighting technique iv) Color fidelity and tonal range v) Digital workflow
Image contrast	i) Lighting quality ii) Lighting ratios iii) Subject properties (and treatment) iv) Level adjustments (digital photography) or film processing (film photography)
Color fidelity and dynamic range	i) Color temperature balance between lighting and capture source (white balance for digital photography, film, and filter combinations for film photography) ii) Digital photography; color space, bit depth, file applications, and gamma settings iii) Color management and calibration of computer monitor
Image dimensional integrity	i) Camera viewpoint ii) Optics iii) Application of linear scales iv) Resizing digital images

Evidence Enhancement

The most common rationale to enhance physical evidence is to improve the visibility of the evidence. Contrast between the evidence and background can become problematic when evidence is found on items that are of similar color or tonal value. Increasing the difference of brightness values between the evidence and the background provides an increase of contrast. Evidence deposited onto material with complex patterns can also lower the visibility of the evidence and cause problems with interpretation of the evidence. An optical enhancement example is found in the two crime-scene photographs (Figure 3).

The photograph in Figure 3(a) is a bloodstained footwear impression made on a black ceramic tile found at a crime scene. This image displays very little visibility and contrast between the bloodstain impression and the black tile background. Figure 3(b) is the same footwear impression that has undergone chemical and optical enhancement to increase the contrast and improve the visibility of the evidence. The bloodstain footwear impression was first treated with a blood reagent called *Hungarian Red* and photographed using a monochromatic light source of 530 nm. A Wratten 25 barrier filter was attached to the camera lens.

Evidence enhancement may be achieved by using the following methods:

- chemical treatment (development or staining);
- optical enhancement;
- specialized lighting techniques; and
- digital imaging enhancement.

Enhancement may be conducted using a single method or a combination of methods as demonstrated in Figure 3(b). Chemical treatment to enhance or develop physical evidence is a practice that is well established in forensic science. The treatment may independently increase contrast due to chemical staining or it may combine physically with the deposited material and alter the spectral qualities of the material. Optical or photographic enhancement involves various techniques utilizing the optical properties of the evidence material and its relationship to the radiation source and spectral sensitivity

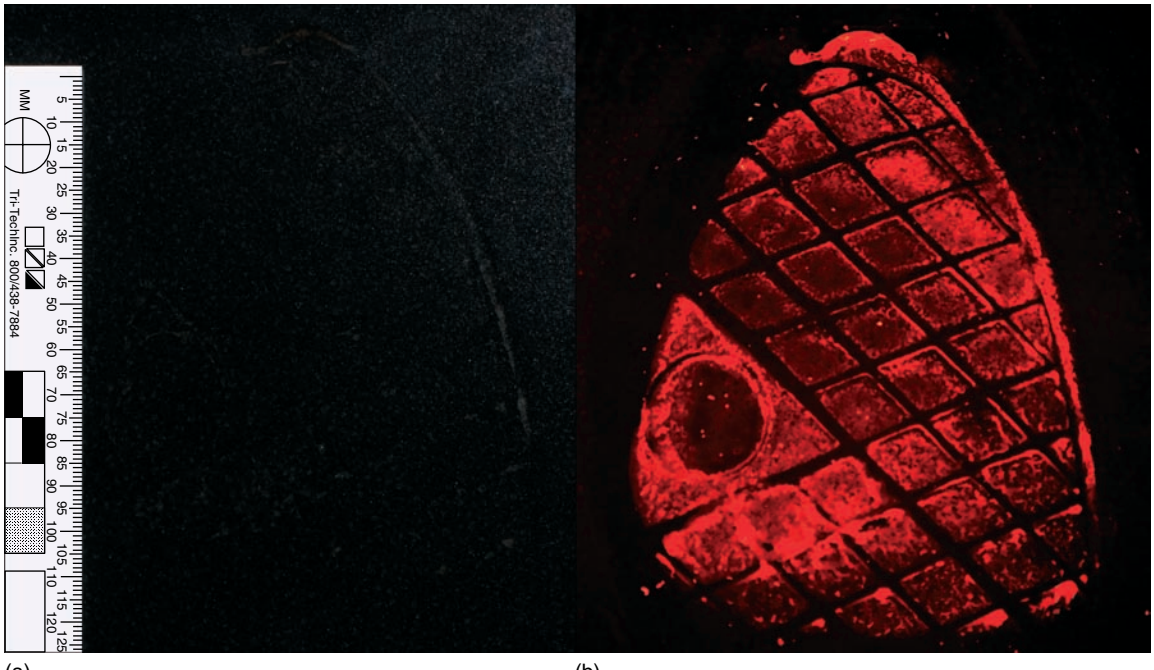


Figure 3 (a) Footwear impression in blood on a black ceramic tile. (b) Same impression treated with Hungarian Red and photographed with a 530 nm monochromatic light source and a Wratten 25 barrier filter on the lens

of the capture device. Digital imaging enhancement also utilizes digital imaging editing software such as Adobe Photoshop™. Optical and digital imaging enhancement are discussed in the next section.

Optical Enhancement

Chemical treatment of physical evidence (i.e., blood or fingermarks) will either stain the material to increase the contrast or make the material react to a light source of specific spectral output. Details regarding the various chemical treatments are discussed elsewhere in this section, however, the reaction of chemically treated specimens to certain wavelengths of light provides the basis of optical enhancement.

Optical enhancement of physical evidence provides the visualization of latent evidence or provides improved visibility by increasing the contrast between the evidence and its background. There are three essential elements that form the foundation for optical enhancement of physical evidence. These elements include: (i) the specific spectral distribution of the light source illuminating the specimen, (ii) the specimen's response when illuminated by the light source, and (iii) the spectral sensitivity of the capture device (film or digital sensor). Each component requires careful consideration when performing optical enhancement of physical evidence.

Spectral Distribution of Light Sources

The approximate spectral sensitivity range of human vision is between 400 and 700 nm, with higher sensitivity within certain ranges depending on the level of illumination. When light enters the eye and forms an image on the retina, nerve cells called *photoreceptors* [15] convert the light energy into a signal that is translated by the mind. There are two types of photoreceptors within the retina called *rods* and *cones* and these receptors react to different levels of illumination. Malacara [16] suggests there are approximately 100 million rods and 5 million cones. The cone receptors are used for bright light conditions and have a sensitivity peak at 555 nm, while the rod receptors provide vision in low light (night vision) and have a sensitivity peak at 505 nm [16–18].

Most digital cameras capture color by using a color filter array positioned over the camera's sensor. This array consists of a mosaic of red, green, and blue (RGB) filters known as a *Bayer filter*. The Bayer filter has twice the number of green filters as red or blue to produce color images that relate to the spectral sensitivity of vision [19].

The spectral output of a light source is called its *spectral distribution* and may be measured using a spectrophotometer. Optical enhancement techniques use various different light sources including natural sources such as direct sunlight and open shade, and artificial light sources such as incandescent tungsten, tungsten halogen, electronic flash, xenon arc lamps, and others.

White light containing a mixture of several colors is considered as a polychromatic light source. Polychromatic light is suitable for most general photography applications and essential for color photography. However, optical enhancement techniques often require a light source with a narrow spectrum or monochromatic light sources [20]. Monochromatic light sources are produced by filtering polychromatic light through an optical filter or filters. Lasers are another source of monochromatic light [21].

Specialized forensic lighting equipment such as the Rofin Polilight™ provides monochromatic light for a range of forensic applications including optical enhancement. These units use a high-intensity xeon light source with inbuilt interference filters to produce monochromatic light with bandwidth of ≈40 nm for most settings. The inbuilt interference filters are also tunable by adjusting the angle of the filter in relation to the transmitting light through the filter. A tuning rate of 30 nm may be achieved with a 45° tilt and the shift is always downtuned toward a shorter wavelength. Figure 4 provides details of the spectral distribution of the Polilight on the 590-nm setting. The monochromatic nature of the light output can be seen by the narrow shape of the spectrum.

Spectral Response of Specimen

Optical enhancement operates on the premise that specimens record in a particular way when they are irradiated by specific wavelengths from a monochromatic light source or transmitted through an optical filter over the camera lens. Optical filtering of the illuminating light to produce a monochromatic source isolates the response from the specimen. These effects can be achieved with polychromatic light sources; however, due to the broad nature of polychromatic

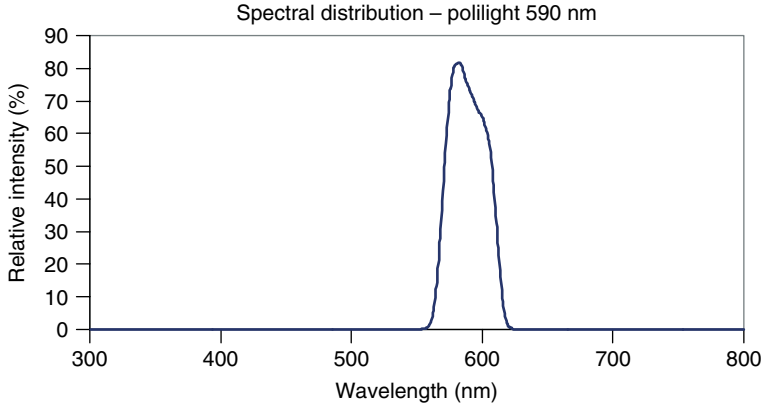


Figure 4 Monochromatic spectra from a Polilight™ forensic light source (setting 590 nm)

lighting, the effects are not seen or are masked by the other wavelengths.

Optical filters placed over the camera lens can also alter the properties of light transmitting through the optical system. Optical filters produce specific spectral responses from the specimen. The application of optical filters on the camera lens allows optical enhancement to be conducted in standard white light conditions, unlike using monochromatic light sources, which will need to be conducted in a darkened room. However, the monochromatic forensic light sources available today provide a narrower spectrum that is difficult to achieve when using a single optical filter over the lens.

Optical enhancement may produce four different conditions or responses from the specimen when illuminated by monochromatic lighting:

- the specimen may *absorb* the light and become darker;
- the specimen may *reflect* the light and become *lighter*;
- the specimen may *transmit* the light and become *transparent*; and
- the specimen may *luminesce* and become *fluorescent*.

These responses to certain wavelengths of monochromatic light may be referred to as *modes of optical enhancement lighting*. These modes may be described as (i) absorption mode, (ii) reflection mode, (iii) transmission mode, and (iv) photoluminescence mode. The purpose of optical enhancement is to make the specimen visible or introduce greater visibility by enhancing the contrast between the specimen and background. When monochromatic light sources are used in the optical enhancement of evidence, natural color rendition of the specimen is not possible due to the specific spectral range of the light used. Strong color casts are produced when using monochromatic light with color photography and this effect should be avoided because it often appears like a photographic error. Digital images photographed in color should be converted to a grayscale image or the color desaturated to – 100%.

Absorption and Reflection Modes. Absorption and reflection modes are chromatic effects that provide an increase of contrast between the specimen and its background. These modes operate on the principle of selective absorption or selective reflection of a colored specimen when illuminated with a specific monochromatic light source (or transmitted via an optical filter). When a monochromatic light source illuminates a color specimen, the specimen will either darken or become lighter in tone depending on the relationship between the color of the light source and the color of the specimen.

Absorption mode is a method of increasing the contrast between the specimen and the background by darkening the tonal value of the specimen or background. Absorption occurs when a colored specimen is illuminated by a monochromatic light that is an opposite color to that of the specimen [20]. The specimen darkens due to its selective absorption properties

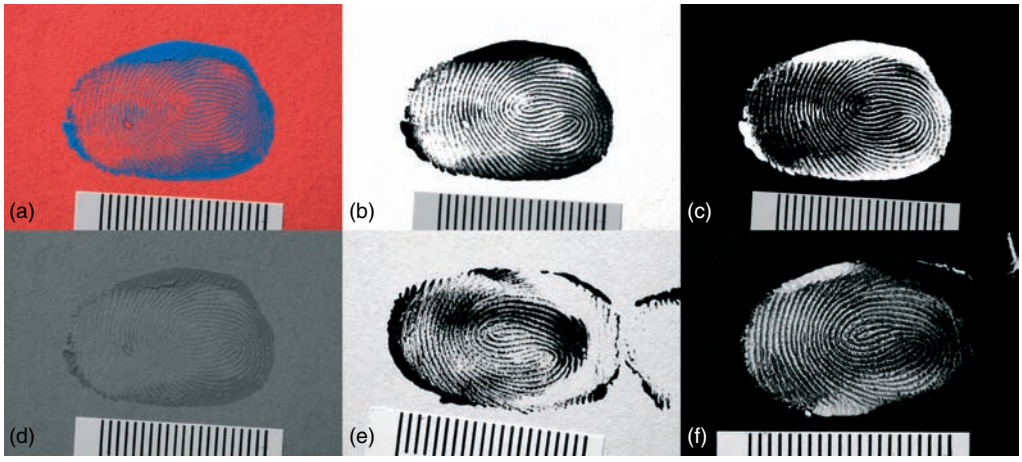


Figure 5 (a) Color image of a fingermark made in cyan-colored paint on a red background. (b) Absorption mode on a red background. (c) Reflective mode on a red background. (d) Grayscale image without enhancement. (e) Absorption mode on white background. (f) Reflective mode on black background

and this mode is most commonly used to increase the contrast.

Reflection mode is also a method of increasing the contrast, however, it lightens the specimen or background by selective reflection. Refection is also a chromatic condition and requires a colored specimen and a monochromatic light source of similar color to the specimen. Figure 5 is a photograph of a fingermark made in cyan colored paint found on a red document. Figure 5(a) is the color image using standard white light (electronic flash), while Figure 5(d) is the same image converted to a grayscale image. Figure 5(b) uses a red monochromatic light (650 nm central bandwidth) and darkens the fingermark due to the absorption of the red illumination. Figure 5(c) uses a blue monochromatic light (450 nm central bandwidth) and lightens the fingermark's tonal value. This series of images also demonstrates an ideal situation for absorption and reflective modes. The color of the background is opposite to that of the specimens. Although this ideal condition is rarely observed in reality, the examples provide an optimum result for this technique.

The general rule for selecting the color of the monochromatic light source for absorption or reflection mode is based on the selective absorption and selective reflection properties of the specimen [20]. It suggests that the color of the light source lightens its own color and darkens its opposing color. Figure 8(a) represents two models of color synthesis that are regularly used in photography and digital imaging. The two models are "additive" and "subtractive" color synthesis. Additive color synthesis uses primary colors such as red (R), green (G), and blue (B). In an ideal condition, if equal quantities of red, green, and blue light are added together, their combination will produce white light [8]. Hence, RGB primary colors are considered as additive and each color ideally represents one-third of the white light (visual) spectrum. The reproduction of color using the three components of red, green, and blue is called *trichromacy* and relates to the function of color vision known as the *Young* and *Helmholtz theory* [8, 15–17].

The subtractive color synthesis model is based on the subtraction of opposing primary colors (called *complementary*) from white light. Therefore, if green is taken out of white light, the resultant color will be magenta. Magenta light is considered as without the green component of white light due to its subtraction. Primary colors may be subtracted from the white light by using the subtractive color synthesis model comprising of the complementary colors, e.g., cyan (C), magenta (M), and yellow (Y). When an ideal complementary filter (cyan, magenta, or yellow) is placed over a white light source, one-third of the spectrum is subtracted (red, green, or blue) while two-thirds is transmitted. Therefore, subtractive color synthesis is a combination of the two remaining primary colors and makes up two-thirds of the

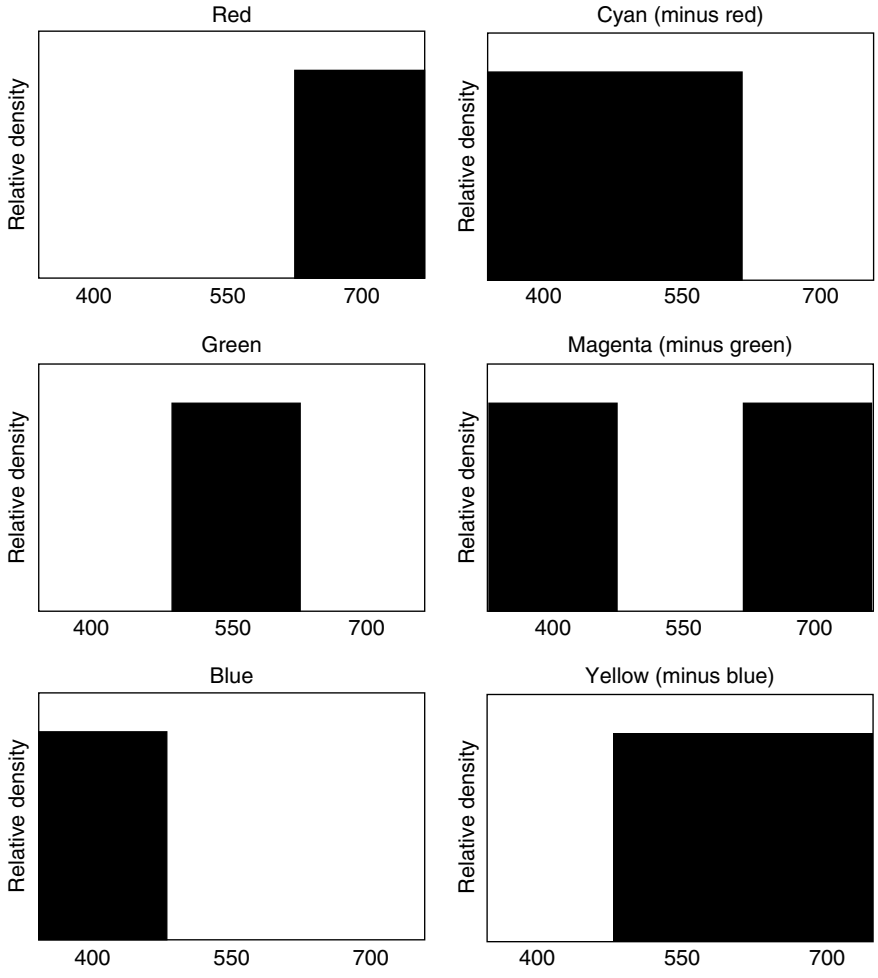


Figure 6 A representation of spectral output from ideal conditions for additive and subtractive color synthesis. The graphs on the left illustrate that each primary color (RGB) produces approximately one-third of the visual spectrum. However, the subtractive model comprising of CMY colors produce two-thirds of the white light spectrum

visual spectrum. Figure 6 illustrates the relationship between additive and subtractive color synthesis models. Each color in the additive color model (RGB) represents one-third of the visual spectrum, while each color in the subtractive model (CMY) make up to two-thirds of the white light.

Additive and subtractive models of color synthesis are also the basis of color film technology and digital imaging color modes [22, 23]. If you examine an image on a color negative film (minus the integral layer which is the amber base color of the film base) you will see that the color image is represented in its negative colors. That is, the reds will be cyan, greens will be magenta, and blues will be yellow. This effect can also be seen if you invert a color digital image. Figure 7 provides an insight into how these two modes of color synthesis are used in photography. The colors rendered in the inverted or negative image clearly show the relationship between additive and subtractive colors in photography.

Digital imaging also uses the same color modeling with RGB color and CMYK color modes. The "K" is an additional channel and represents black. RGB color is the standard for most digital photography applications, while CMYK is a color space often

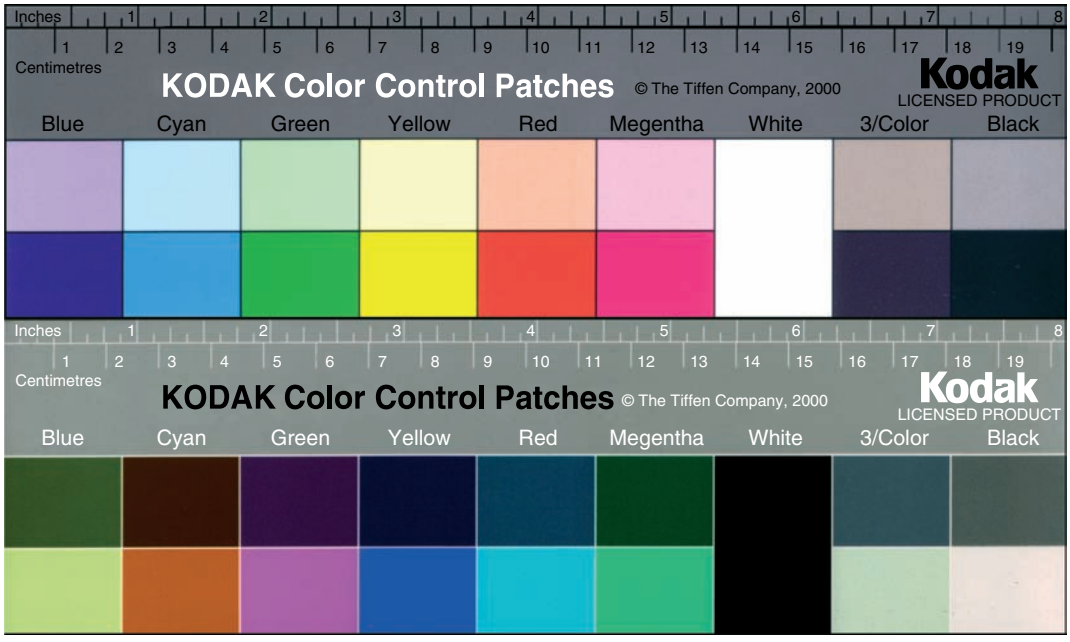


Figure 7 Positive and negative color images displaying the relationship between additive and subtract color synthesis in photography

used for photography used in graphic reproduction (reproduction of images in print media).

Figure 8(a) further displays the relationship between additive and subtractive color models and provides a guide for the selection of monochromatic light for absorption or reflective modes. Each triangle represents a color synthesis model, RGB or additive and CMY (without the K) or subtractive, to produce a color wheel or more appropriately a color star. Colors on the opposing points of the star (e.g., red and cyan) represent opposite colors, while the colors positioned on points next to the additive primary colors are their composites. The opposing colors will provide optimum absorption for specimens and selective monochromatic light sources or optical filters and will darken the specimen. The two adjacent colors next to each opposing color will also darken the specimen.

This diagram forms the basis for absorption mode and the specimens opposing color should be selected as the color of the monochrome light source or optical filter. While absorption mode is usually more effective to enhance contrast than reflective mode, selecting the same color monochrome light source as the color of the specimen will lighten the tonal value of the specimen. Color wheels as illustrated in Figure 8(a) are for photographic applications and are calculated based on the additive and subtractive models of color synthesis of light [22]. Color wheels that are designed for pigments and used in painting and graphic design are not the same and are not suited for optical enhancement or photography.

The color of the background is also a factor to consider when increasing the contrast using absorption and reflective modes. Neutral backgrounds such as white, gray, or black will display little change when illuminated by monochromatic light due to the achromatic (without color) character of these tones. However, if the background is colored, then the same absorption or reflection principles will occur to the background as it does to the specimen. Backgrounds, which are the opposite color to the specimen, have the potential to produce the highest degree of contrast when applying absorption or reflective modes. White backgrounds also work well with absorption mode because the specimen can be darkened against a light background. Black backgrounds work best with reflective mode by making the specimen lighter to enhance contrast. Figure 5(e) demonstrates the results of absorption mode with a white background, while

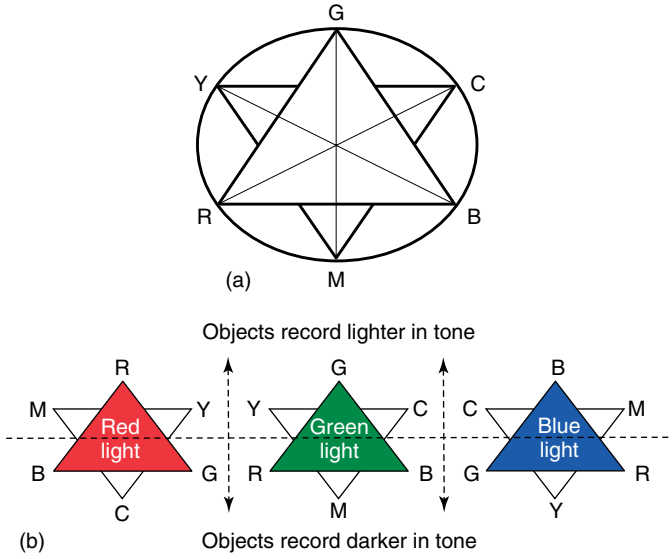


Figure 8 (a) Additive and subtractive color synthesis models demonstrating color opposites; (b) Additive and subtractive color synthesis models demonstrating the tonal changes when using monochromatic light sources. The object colors positioned above the horizontal line will record in a lighter tonal value, while the object colors indicated below the line will record darker in tone due to the absorption of the light

Figure 5(f) illustrates the effect of reflection mode on a black background.

As previously mentioned, absorption and reflective modes may also be achieved by using optical filters placed on the camera's lens and photographed using standard white light. This technique has a long history with black and white film photography. Optical filters, known as *contrast filters*, provide a similar effect as monochromatic lighting, except that the effect is influenced and controlled by the transmission properties of the optical filter with white light [24–26]. Filter factors must also be applied to the camera exposure to compensate for the loss of light caused by the absorption of light by the filter. The application of photographic filters may also be achieved digitally with Adobe Photoshop™. Digital imaging enhancement is discussed in the following section.

Transmission Mode. While an increase of contrast is the objective for absorption and reflective modes of optical enhancement, transmission mode has a different function. The function of transmission mode is to reveal evidence that may be obscured or not visible [27]. It uses the transmission properties of selective wavelengths of light incorporation with the spectral characteristics of the specimen and the spectral sensitivity of the recording medium. This enhancement mode allows the specific wavelengths of light to transmit through thin specimens of specific spectral qualities and renders the specimen transparent [6, 26–29]. When the specimen becomes transparent, items beneath the specimen can become visible. An example is shown in Figure 9 where the obliterated writing on a document has become visible due to the transparent properties of the ink covering the text.

Photoluminescence Mode. When specimens possessing certain properties are illuminated by specific wavelengths of monochromatic light, they may absorb the excitation light and then reemit light of a different wavelength. This effect is know as *luminescence* [29, 30]. There are two different results caused by luminescence: (i) fluorescence that emits light at a longer wavelength while the excitation source is illuminating the specimen and (ii) phosphorescence that emits light and continues to do so for a period of time after the excitation source ceases to illuminate the specimen [29, 31, 32]. The material painted onto watch dials that illuminate in the dark is an example of phosphorescence.

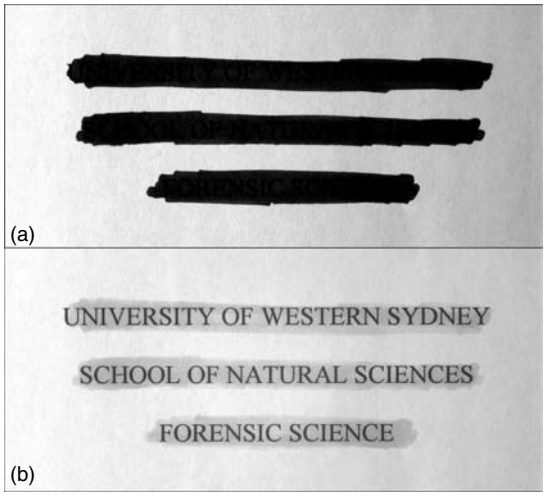


Figure 9 Transmission mode. The ink obliterating the text has become transparent and allowing the visualization of the text beneath the ink

There are also other forms of luminescence such as bioluminescence that is found in several organisms (e.g., glowworms) and chemiluminescence that is caused by a chemical reaction (e.g., luminol) [29]. Bioluminescence and chemiluminescence are caused without the presence of an excitation light source. Lennard and Stoilovic [31] suggest that luminescence caused by the absorption of light is called *photoluminescence* and this type of luminescence is often used in optical enhancement. Bioluminescence, chemiluminescence, and photoluminescence effects can all be photographed; however, due to the low level of light emission, these procedures must be conducted in a darkened environment [29].

Photoluminescence mode enables the specimen to fluoresce and the emitted light from the specimen can be recorded by a camera. Recording specimens using this type of luminescence is also called *fluorescence photography* [29, 32, 33]. This mode requires consideration of all aspects of the optical enhancement triangle including (i) the specific spectral output of the excitation light source, (ii) a photoluminescent response (fluorescence) from the specimen, and (iii) the spectral sensitivity of the recording device (film or digital sensor).

The selection of the specific excitation spectra will depend on the photoluminescent properties of the specimen material. Forensic evidence specimens such as seminal fluid, saliva, and urine fluoresce naturally when illuminated with certain excitation light sources [30, 33]. This naturally occurring phenomena is referred as *autofluorescence* [1, 29] and optical enhancement may be achieved without chemical treatment in these cases and is nondestructive. Other forms of evidence that do not naturally fluoresce (like blood) may be induced to fluoresce by chemical treatment with a fluorochrome [29]. This process is called the *secondary fluorescence* [1, 29] and treatments such as Rhodamine 6G, Indanedione, Diazofluorenone (DFO), and Hungarian Red are some of the common reagents used in posttreatment for photoluminescence mode [34].

Photoluminescence mode is often an alternative to other optical enhancement methods when the background does not suit the absorption or reflective modes. An example may be found in blood evidence. Absorption mode is generally the preferred method for blood evidence using a monochromatic light source at 415 nm, the maximum absorption of dried blood [33]. However, if the blood evidence is deposited onto a dark background, absorption mode will not provide an increase of contrast. Instead, it will darken the blood against a dark background, resulting in very little contrast. Figure 3(b) provides an example of photoluminescence mode for blood evidence.

Excitation ranges for photoluminescence mode vary depending on the autofluorescent characteristics of the specimen or the reagent used for posttreatment secondary fluorescent photography. Most forensic applications use spectral ranges in the lower order including ultraviolet (UV), violet, blue, and green region wavelengths. The resultant emitting light from photoluminescence mode is always longer in wavelength than the excitation source used. Therefore, most photoluminescence applications provide an emission light within the visual spectrum (400–700 nm). Some document examination applications are the exception where infrared luminescence methods produce an emission source further into the invisible spectrum. This method is generally outside the standard photography range and is conducted using specialized instrumentation called a *video spectral comparator* (VSC).

The emission light source produced by the photoluminescence is usually quite weak in brightness value and as previously mentioned requires photography in a darkened environment. Barrier filters are also applied to the camera lens to improve the contrast

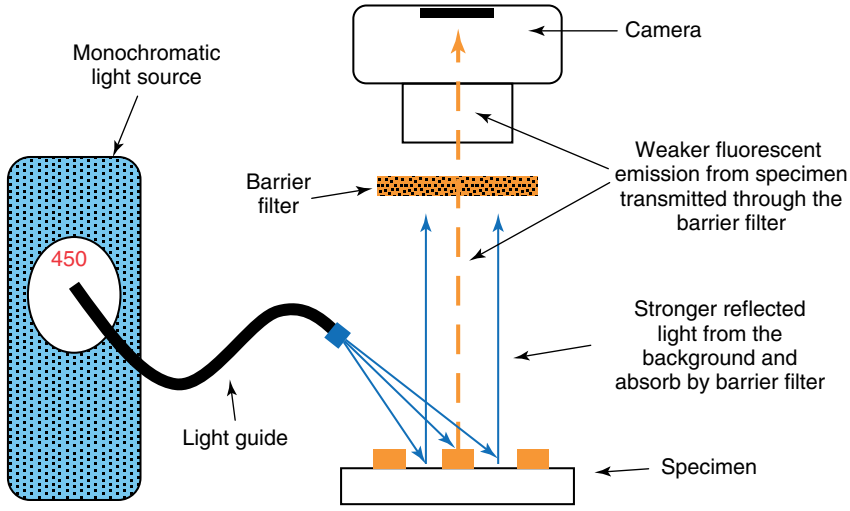


Figure 10 Photoluminescence mode displaying the application of a monochromatic light source, a specimen that fluoresces and a barrier filter over the camera lens

between the specimen and background. The rule for applying barrier filters for photoluminescence mode is to use a filter of the same color as the emission source. This provides maximum transmission of the fluorescence and will also darken any opposing background reflected light. For example, when a blue-colored light source is used to produce an orange-colored fluorescence, an orange barrier filter transmits the orange colored emission and darkens the reflected blue light of the background.

Photoluminescence mode requires the following parameters: (i) an excitation source suitable for the specimen material, (ii) a specimen that provides fluorescence (naturally or induced), (iii) a barrier filter to improve contrast, and (iv) the appropriate spectral sensitivity of the recording device (usually within the visual spectrum). Figure 10 illustrates the requirements of photoluminescence mode. Figure 11 is a fingerprint dusted with a green fluorescent fingerprint powder and photographed using a 540 nm excitation source. Figure 3(b) is a bloody footwear impression treated with Hungarian Red reagent excited by 530 nm excitation source from a Polilight and a Wratten 25 (red) barrier filter on the camera lens.

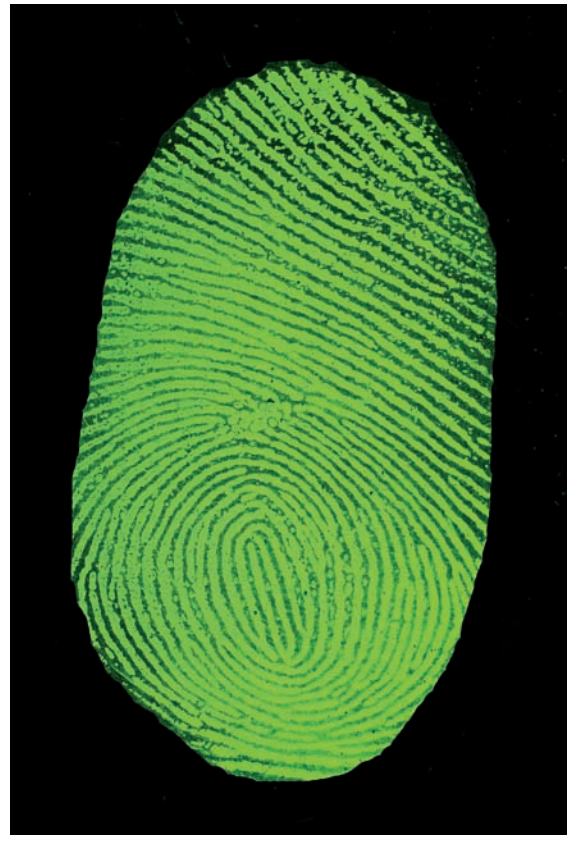


Figure 11 Fingermark using photoluminescence mode

Spectral Sensitivity of Capture Device

The spectral sensitivity of the recording mechanism (digital sensor or film) is the spectral range that the

device or material is able to record. Spectral sensitivity is an essential consideration when conducting optical enhancement methods. The recording device must be capable of recording in the spectrum used in the enhancement method.

The spectral sensitivity of film may be examined using wedge spectrograms. Wedge spectrograms are graphs that plot the spectral response of the film. These graphs are produced by exposing film to a dispersed light that has passed through a diffraction grating and then a neutral density wedge, which is placed on the film's surface.

Tripack color films have a spectral sensitivity range within the visual spectrum (400–700 nm). These films are not sensitive in the infrared region and an UV absorption layer is situated before any of the RGB sensitivity layers which prohibits any recording in the UV region. Generally, color films are not used in optical enhancement techniques. They produce a color cast when using monochromatic lighting or using a colored optical filter and are not sensitive outside the visual spectrum.

Film emulsion is the suspension of silver halides in gelatin to form a light sensitive emulsion. It is naturally sensitive only to UV and blue wavelengths and early film remained only sensitivity in these regions. In 1873, Vogel discovered that film emulsions can be made sensitive to blue and green spectra by adding a dye to the emulsion [8]. This process of dye sensitization was later refined and emulsions spectral sensitivity were extended into the red and infrared spectra. Black and white films are regularly used in optical enhancement and are still available in various spectral sensitivities. There are five classes of black and white films with different spectral sensitivity ranges. These classes are as follows:

- blue sensitive – only sensitive to UV and blue wavelengths;
- orthochromatic – sensitive through to the green region;
- panchromatic – sensitive through to red or covering the visual spectrum up to ≈650 nm;
- extended red sensitivity – extends further into the red spectrum up to ≈750 nm; and
- infrared – sensitive into the infrared region up to ≈900 nm.

For a more detailed examination of the film's spectral sensitivity, consult wedge spectrograms for each film type. Film manufactures readily publish this information with the product description.

The spectral sensitivity of digital cameras is more difficult to obtain from most of the camera manufacturers. There are generally three different types of digital sensors, also called *semiconductors*, used in digital cameras:

- charged coupled device (CCD);
- complementary metal oxide semiconductor (CMOS); and
- Foveon.

Single-shot digital cameras using CCD or CMOS sensors use a Bayer filter, which is a mosaic of RGB filters placed over the pixels. Foveon sensors, however, use a system of three different layers of semiconductors stacked together to form a multilayer sensor. These sensors design replicates that of the tripack color film, which consists of three different color sensitive layers (RGB). Each layer is suspended in a silicon wafer and the longer wavelengths are able to penetrate through the previous layer/s [35]. The advantage of Foveon sensors is that they do not require demosaicing interpolation like systems using a Bayer filter and the physical size of the sensor is not divided by each color channel (therefore, increasing sensor resolution).

Digital sensors have spectral sensitivity characteristics like all the light sensitive devices. CCD and CMOS sensor's spectral sensitivity works very differently to that of silver halide film. While film is naturally sensitive to UV and blue light and is made sensitive to other spectral regions using dye sensitization, digital sensors are more sensitive to the infrared (IR) region with very little sensitivity in the blue and UV spectral regions. The lower sensitivity in the blue region is the cause of higher noise levels found in the blue channel. Digital sensors are monochromatic and color is produced by processing of the image using trichromacy (combination of RGB). To improve chromatic aberration effects on the optics, most digital cameras also place an infrared blocking (absorbing) filter over the sensor to limit the spectral sensitivity to the visual spectrum. The spectral sensitivity of digital cameras is therefore a combination of (i) the spectral response of the monochromatic semiconductor, (ii) the absorption properties of the IR blocking filter, and (iii) the transmission properties of the Bayer filter.

Various digital cameras are available for forensic photography. Excluding the standard low-end consumer compact digital cameras, cameras suitable for forensic photography include: standard digital single-lens reflex (SLR) (pro and proconsumer models), digital SLR cameras with the IR blocking filter removed (modified for infrared photography) and scientific-based monochrome digital cameras. These scientific types of digital cameras provide various ranges of spectral sensitivity and generally publish the response data. Standard SLR camera manufacturers rarely publish the spectral sensitivity data and are generally only considered for applications within the visual spectrum (400–700 nm), although some cameras will perform slightly outside this range.

Specialized Lighting Techniques

Photography readily exploits light and its interaction with the surfaces of objects. The term *lighting* is used in photography as a way of describing the "visual effect" that results when this interaction of surfaces and light energy manifests. Some lighting techniques provide aesthetic elements to a photograph, while some others may enhance aspects of the physical evidence. Lighting may provide a visual representation of shape, line, form, subject texture, detail, aerial perspective, and many more visual aspects [35]. The topic of photographic lighting is an expansive consideration and cannot be fully justified in this section of the text. However, four specialized lighting techniques used for the enhancement of physical evidence are discussed in this section, which are as follows:

- axial illumination;
- near-axial illumination;
- diffuse reflection method; and
- oblique lighting.

Axial Illumination

Axial illumination is the light that illuminates the specimen from the camera's optical axis [36]. This form of lighting produces no shadow visible by the camera and is a form of shadowless lighting. Owing to its directionality, it may be used successfully to photograph specimens in cavities when getting light into small crevices is difficult. Axial illumination is used to detect topographical variances found on flat metal surfaces including coins or medals. It is also a technique that may also be used to photograph untreated fingermarks found on metal or reflective surfaces.

Axial illumination is a similar technique as epi-illumination used in photomicroscopy [18, 36]. This form of illumination is produced by using a collimated (parallel) light source directed onto a thin semitransparent (semisilvered) mirror beam splitter positioned at a 45° to the lens axis [36, 37]. Thin semitransparent mirrors are more suitable than plane glass because the mirror surface increases the illumination efficiency and reduces the double-image effect, which is more visible on plane glass beam splitters. A slight double image results due to the separation between the two surfaces of the beam splitter (front and back). Thicker beam splitters produce a greater shift between each image that makes this artifact more obvious. Other ambient light must also be carefully controlled to avoid extraneous light reflecting off the beam splitter's reflective surfaces. In particular, reflected light transmitting through the beam splitter and illuminating other objects within the room can produce a reflection on the near side of the beam splitter and cause a reflection of the object within the image space. A light absorber such as black velvet cloth should be used to absorb this light (see Figure 12). Figure 12 illustrates the components of axial illumination.

Near-Axial Illumination

Near-axial lighting produces a similar lighting effect as axial illumination. The light source is positioned close to the lens axis [36] producing a light quality that is void of form, texture, and shadow. The intent of near-axial lighting is to illuminate an object or scene evenly. Several portable light sources, including portable flash, may be used to produce this form of illumination. Ring light attachments (see Figure 13) attach to the camera lens and can also produce a diffuse near-axial illumination [36].

Diffuse Reflection Method

When light reflects off a surface, the properties of the surface will affect the type of reflection produced. There are two types of reflected light possible: diffuse reflection and direct reflection [8, 17, 38]. Direct reflection (also referred to as *specular reflection*) results from light reflected off polished surfaces.

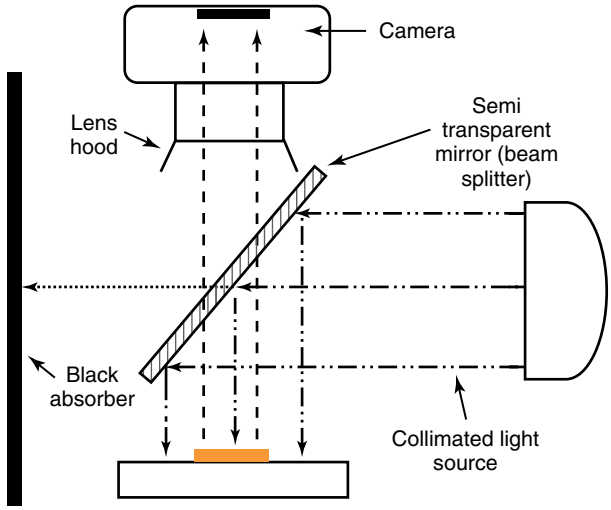


Figure 12 Axial illumination lighting diagram



Figure 13 A ring flash attachment on a digital SLR camera used for near-axial illumination

Direct reflection follows the first law of reflection which states that the angle of incidence equals the angle of reflection ($i = r$) [8, 12, 17, 22].

Diffuse reflection is the scattering of reflected light caused by a matt surface. The uneven nature of matt surfaces reflect the incident light in different directions by maintaining the first law of refection ($i = r$). The effect produces reflected light in multiple directions causing a diffuse effect (see Figure 14b). Diffuse reflection may also be considered as partly diffuse or totally diffuse [22]. Totally diffuse reflected light is perfectly diffuse meaning the spread of reflected light forms evenly across the entire surface. Totally diffuse reflections obey Lambert's law and the surface is considered to be a *Lambertian surface* [8, 12, 22].

Diffuse reflection lighting method exploits the outcomes of both direct and diffuse reflection. Figure 15 is a photograph of an untreated (undeveloped) fingermark on a highly polished flat surface, which is a computer hard drive. This subject produces two different surfaces: the highly polished surface of the hard drive and the matt surface of the fatty deposit of the fingermark. When incident light is applied to both surfaces simultaneously, a direct reflection will be produced off the hard drive and a diffuse reflection from the fingermark.

The relationship between the position of the camera lens, the position of the incident light source, and the position of the specimen are all critical aspects of this technique. Their positions are also interrelated and careful consideration is essential. The camera should be positioned perpendicular (90°)

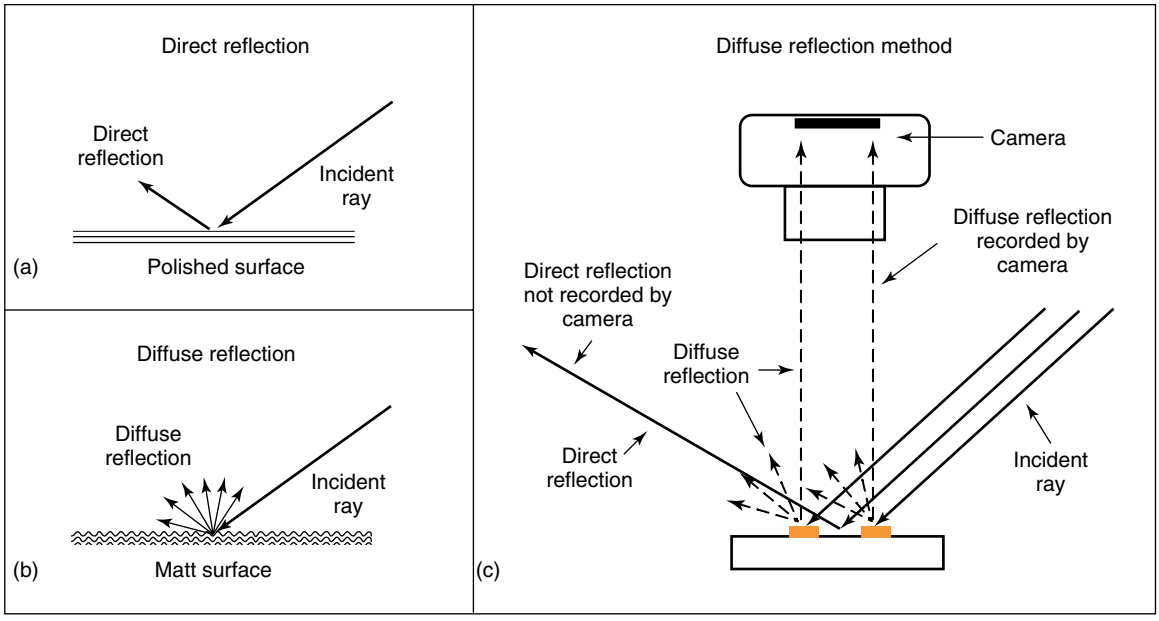


Figure 14 (a) Direct reflection from a polished surface ($i = r$), (b) diffused reflection from a matt surface, and (c) lighting, illustrating diffuse reflection method seen in Figure 15



Figure 15 Untreated fingerprint found on a highly polished computer hard drive and photographed using diffuse reflection method

to the subject plane. The incident light source is positioned at ≈45° angle from the specimen, which will also result in a ≈45° angle from the lens axis (see Figure 14c). The visualization of the fingermark is achieved due to the differences between the reflected light values from the diffuse light reflected from the fingermark and the direct light from the polished surface. Owing to the law of reflection ($i = r$), ≈100% of the direct reflection will reflect away from the camera lens and not record in the camera. This will result in the polished surface having no light entering the camera and producing a dark toned surfaced in the photograph. The diffuse reflection emanating from the fatty deposit of the fingermark will direct some reflected light into the camera lens and record in the photograph. The result is illustrated in Figure 15 whereby the fingermark records a light tone against a dark background. Diffuse reflection method is a simple and effective method of recording fingermarks on polished surfaces without having to treat or develop the mark.

Oblique Lighting

Oblique lighting (also referred to as *side lighting*) is one of the most valuable lighting techniques used

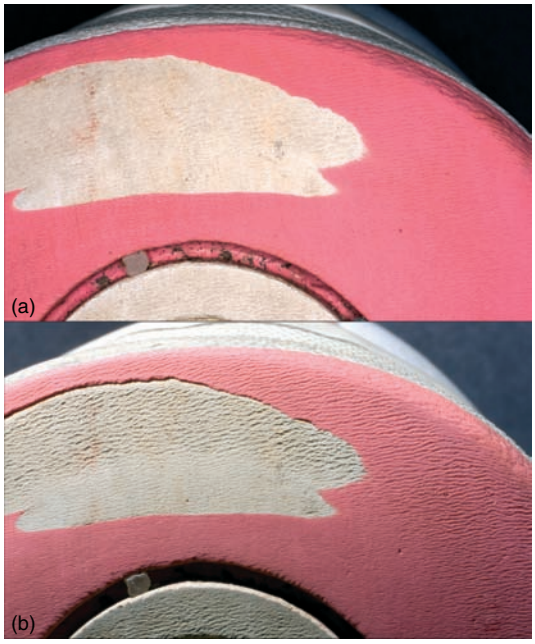


Figure 16 (a) Photograph depicting the heel of a shoe using near-axial illumination (ring flash). (b) Same shoe using oblique lighting. Schallamach patterning becomes more visible with oblique lighting

when photographing mark evidence. Oblique lighting may provide a significant increase in detail and provide the visualization of individual identification characteristics that may be crucial to the examination. Oblique lighting is achieved by positioning the light source at an oblique or low angle to the specimen. The lighting direction runs across the surface of the specimen and essentially increases the local contrast or variations of highlight and shadow on the surface of the specimen. This effect increases the topographic detail and provides further visual information to examine.

Figure 16 provides an example of the differences between oblique and near-axial illumination. The Schallamach patterning found on the heel of a shoe is greatly emphasized by oblique or side lighting. Figure 16(a) was photographed using a near-axial portable flash, while Figure 16(b) uses oblique lighting. Oblique lighting has increased the local contrast on the surface of the heel making the Schallamach pattern more visible.

Digital Imaging Enhancement

Digital enhancement of physical evidence is the application of computer software to make adjustments to digital images of evidence. Digital enhancement provides several similarities to the techniques used in digital photography and optical enhancement. The most significant advantage digital imaging has provided forensic science is its ability to make fine adjustments to image parameters such as contrast, color, sharpness, image noise, and many other adjustments. Items in images can also be quantified by counting, measuring, and making mathematical calculations using image analysis applications. The advantages of digital imaging for forensic science are expansive. This section examines two common practices of digital enhancement using such controls as contrast and the chromatic modification of specimens.

Digital enhancement of physical evidence should not be considered as a process of correcting poorly executed digital photography. Digital enhancement of evidence requires the input of quality digital photography and the enhancement techniques provide further tools for the forensic scientist. Concepts of quality include sharpness, image dimensional integrity, lighting, exposure, and dynamic range.

Digital images record brightness linearly unlike film which responds to light nonlinearly [39]. Film's nonlinear response to light may be seen in the "S"-shaped characteristic curves that plot their exposure and density parameters. Obtaining quality digital images with a higher dynamic range is preferable when performing digital enhancement. Essentially, the more inherent information the digital image contains, the more control image enhancement offers. Best practice suggests digital images should be captured (camera or scanner) using a 16 bit (or a 14 bit for some cameras) and RAW images are better processed using Photoshop processing appliances such as Camera Raw™ than in-camera processed JPEGs.

Compression applications such as the JPEG captured and processed in-camera at an 8 bit offers significant less information and enhancement control. Digital enhancement processing in Photoshop should also be carried out using the 16 bit per channel for improved results [13, 19]. There is a significant difference in pixel depth between 8 and 16 bit per channel modes. The 8 bit per channel provides 256 variations per channel that results in approximately 16 million possibilities for RGB files. While the

16 bit per channel images contain 65 536 variations for each channel and results in billions of color and tonal representation possibilities.

Contrast Adjustment

Digital imaging can provide adjustments to image contrast with significantly more control and ease than more traditional film-based technology. Visualization of evidence is predominately made possible by the contrast between the evidence and its background. Fingermark evidence is a good example and is most reliant on contrast between the friction ridge detail and the substrate background. Like optical enhancement, digital imaging enhancement using contrast adjustments in Photoshop is an effective method of improving the detail and visualization of evidence.

There are various methods in Adobe Photoshop™ that allow contrast adjustment. The most suitable method is using the "curves" function located in the "Image > Adjustments > Curves" menu. In Photoshop versions greater than CS3 the exposure histogram is also embedded into the curve graph. Figure 17 is an example of how curves may adjust to the contrast of the specimen. The image is of a fingermark on a glass window and developed in black powder. To avoid specular reflections, the fingermark was the photograph using brightfield illumination that combined a small sheet of white Perspex situated behind the fingermark and a portable flash positioned behind the mark was used as the illumination source. The original image displayed high levels of detail but with low contrast and the contrast was enhanced using the "curves" adjustment. The inserts situated in each image provide details of the before and after curve adjustments.

The first graph shows a straight line (curve) diagonally across the graph which represents the linear nature of digital images. The gradient of the curve influences the degree of image contrast. The greater the gradient (more steeper the curve) the less variation there is between certain sections of the shadow and highlight regions resulting in more image contrast. Lowering the gradient (flatter curve) reduces the contrast. The second insert found in Figure 17 shows an increase in the gradient of the curve that has resulted in the increase of contrast. The shadow and highlight regions were also "clipped" slightly which also added to the contrast improvement. Contrast enhancement adjustments using "curves" may be conducted in each channel separately or as an alpha channel indicated as "RGB".

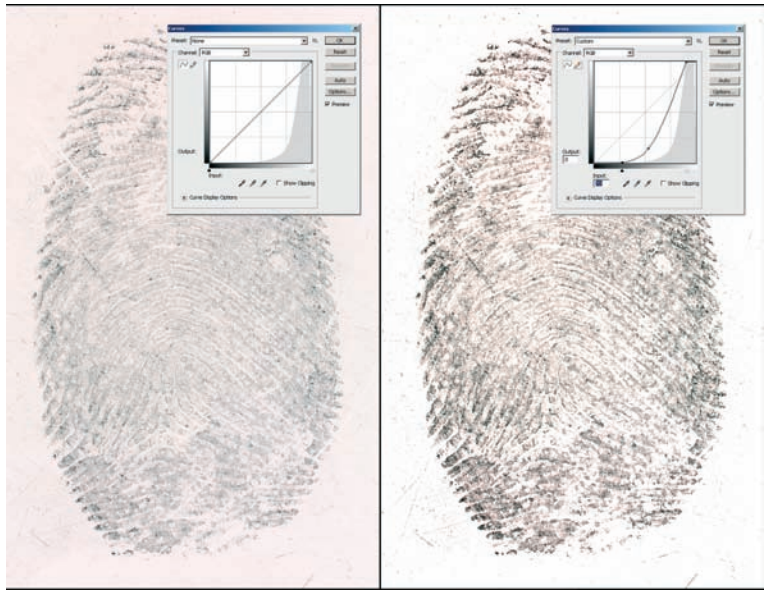


Figure 17 Before and after fingermarks when contrast is enhanced using Photoshop curves

Chromatic and Tonal Modification

Adjustments to the spectral response of color may also be modified in digital imaging software in a similar way to optical enhancement techniques. When working with Adobe Photoshop™ there are always several different methods to obtain the same or similar result. Changing the chromatic aspects of specimens can also be achieved using several Photoshop functions; however, the "Black & White" adjustment is most likely the simplest. The "Black & White" adjustment automatically converts the image to a monochrome image while maintaining the file in RGB mode, so that modification of the colors, now tones, can be adjusted. As previously explained, when using optical enhancement techniques using monochromatic lighting or colored optical filters over the camera, the image should be converted to a monochrome image to avoid color casts.

The "Black & White" function can be found on Photoshop versions higher than the CS3 and work on the principles of black and white contrast filters (see optical enhancement). It may be found on the Photoshop menu "Image > Adjustment > Black&White". Like contrast filters used on the camera with monochrome film, the "Black & White" function alters the brightness of tonal values by altering the spectral reflective and absorption properties. Logically, the digital imaging application is altering the spectral response artificially by adjusting the brightness values of each color and its resultant tonal value in monochrome (unlike optical enhancement). The "Black & White" function provides individual darkening or lightening of the following colors; red, green, blue, cyan, magenta, and yellow. These are the colors previously mentioned in the additive and subtractive models. There are three options when operating this function: (i) the custom option which uses slider bars to modify the tonal value of each color, (ii) use preset filters in the drop-down box, and (iii) develop your own presets. The custom function offers more control over the tonal values of each color and is considered more suitable for evidence enhancement. The image must be in RGB mode to operate "Black & White" and it does not operate in grayscale or CMYK modes. As previously mentioned, working with 16 bit per channel images provides a distinctive advantage for this method due the more subtle changes possible between tonal adjustments.

Figure 18 provides and example of how the "Black & White" adjustments function can enhance evidence by selective contrast variations of tones due to their original color. The blue ink of the stamp was lightened by moving the blue slider bar to the right to increase its brightness value. The red ink of the handwriting was darkened slightly by moving the red slider bar left or lowering the brightness value of the red color. The result is a remarkable difference in contrast difference between the blue and red inks and the blue stamp has almost disappeared. The signature can now be examined without the interference of the overlaying blue stamp.

The Black & White function is also useful for removing colored backgrounds that interfere with developed fingermarks. Figure 19 is a ninhydrin

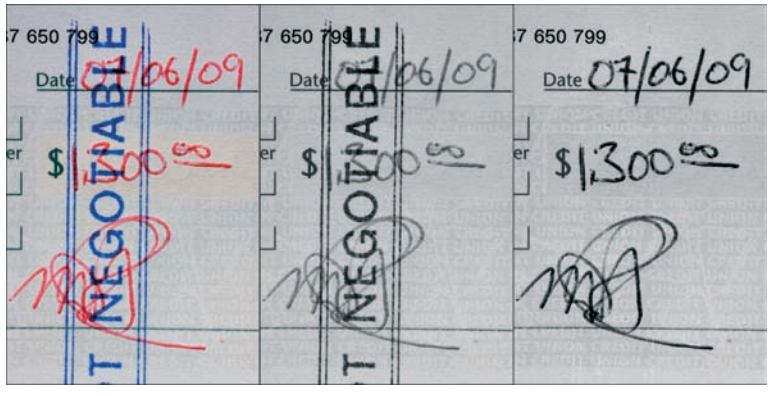


Figure 18 The handwriting on the bank check was written in red ink. The blue stamp has obliterated components of the signature and check amount. The Black & White adjustment has lightened the blue stamp and darkened the writing slightly to increase the contrast between the two inks (stamp and pen)

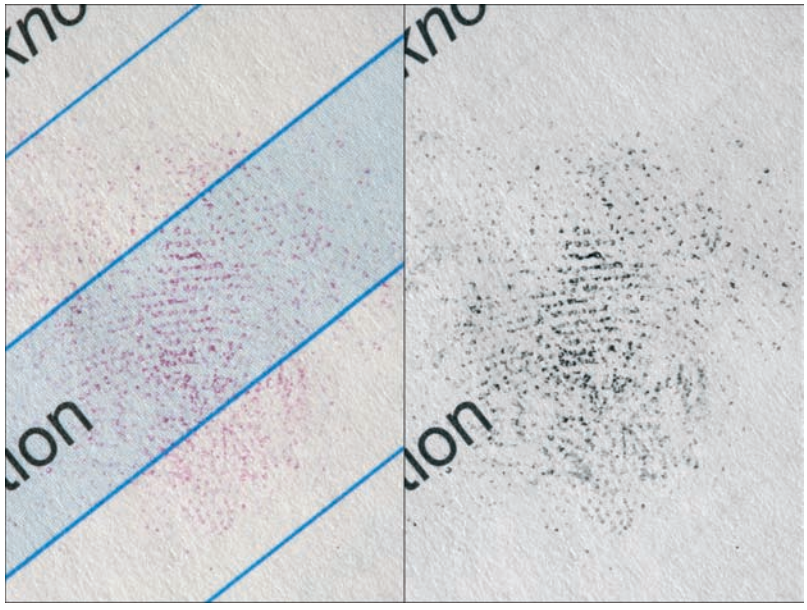


Figure 19 (a) Color photograph of a ninhydrin developed fingermark; (b) The same fingermark with the blue lines and blue area of the document removed and the ninhydrin fingermark darkened

developed fingermark on a document with strong blue lines and blocked sections of blue. The blue lines and blue background were removed using the "Black & White" filter application and the magenta color of the ninhydrin print darkened to enhance the contrast between the fingermark and the background. This process takes approximately 20–30 s to complete once the image is loaded into Photoshop™, making it a highly efficient method of evidence enhancement.

References

[1] Porter, G. (2004). Specialised photography and imaging, in *The Practice of Crime Scene Investigation*, J. Horswell, ed, CRC Press, pp. 139–159.

[2] Porter, G. (2007). Visual culture in forensic science, *Australian Journal of Forensic Sciences* **39**(2), 81–91.

[3] House of Lords, Select Committee on Science and Technology (1998). *Digital Images as Evidence: Evidence*, The Stationary Office, London.

[4] House of Lords, Select Committee on Science and Technology (1998). *Digital Images as Evidence: Report*, The Stationary Office, London.

[5] SMANZFL (2004). *Australasian Guidelines for Digital Imaging Processes*, Version 2, National Institute of Forensic Science.

[6] Robinson, E.M. (2007). *Crime Scene Photography*, Academic Press.

[7] Langford, M. (1998). *Advanced Photography*, 6th Edition, Focal Press.

[8] Jacobson, R.E., Ray, S.F., Attridge, G.G. & Axford, N.R. (2000). *The Manual of Photography: Photographic and Digital Imaging*, 9th Edition, Focal Press.

[9] Taylor, J.T. (2005). *The Optics of Photography and Photographic Lenses*, Elibron Classics.

[10] Canon, E.F. (2003). *Lens Works III: The Eyes of EOS*, Canon.

[11] Ray, S.F. (1995). *Applied Photographic Optics*, 2nd Edition, Focal Press.

[12] Ray, S.F. (1992). *The Photographic Lens*, 2nd Edition, Focal Press.

[13] Reis, G. (2007). *Photoshop CS3 for Forensic Professionals: A Complete Digital Imaging Course for Investigators*, Sybex.

[14] Baron, C. (2007). *Adobe Photoshop Forensics: Sleuths, Truths and Fauxtography*, Thompson Course Technology.

[15] Fraser, B., Murphy, C. & Bunting, F. (2005). *Real World Color Management*, 2nd Edition, Peachpit Press.

[16] Malacara, D. (2001). *Color Vision and Colorimetry: Theory and Applications*, Spie Press.

[17] Overheim, R.D. & Wagner, D.L. (1982). *Light and Color*, John Wiley & Sons.

[18] Morton, R.A. (ed) (1984). *Photography for the Scientist*, Academic Press.

[19] Fraser, B. (2007). *Real World Imaging Sharpening With Adobe Photoshop CS2*, Peachpit Press.
[20] Champod, C., Lennard, C., Margot, P. & Stoilovic, M. (2004). *Fingerprints and Other Ridge Skin Impressions*, CRC Press.
[21] Lee, H.C. & Gaensslen, R.E. (eds) (1991). *Advances in Fingerprint Technology*, CRC Press.
[22] Saxby, G. (2002). *The Science of Imaging: An Introduction*, Institute of Physics Publishing.
[23] Mitchell, E.N. (1984). *Photographic Science*, Wiley & Sons.
[24] Langford, M., Fox, A. & Smith, R.S. (2007). *Langford's Basic Photography: The Guide for Serious Photographers*, 8th Edition, Focal Press.
[25] Kodak (1990). *Kodak Photographic Filters Handbook*, Eastman Kodak Company.
[26] Kodak (1976). *Using Photography to Preserve Evidence*, Eastman Kodak Company, Pub. M-2.
[27] McKechnie, M.L., Porter, G. & Langlois, N. (2008). The detection of latent residue tattoo ink pigments in skin using invisible radiation photography, *Australian Journal of Forensic Sciences* **40**(1), 65–72.
[28] Kodak (1980). *Applied Infrared Photography*, Eastman Kodak Company, Pub. M-28.
[29] Kodak (1968). *Ultraviolet and Fluorescence Photography*, Eastman Kodak Company, Pub. M-27.
[30] Vanderberg, N. & van Oorschot, R.A.G. (2006). The use of Polilight® in the detection of seminal fluid, saliva and bloodstain and comparison with conventional chemical-based screening tests, *Journal of Forensic Science* **521**(2), 361–370.
[31] Lennard, C. & Stoilovic, M. (2004). Application of forensic light sources at the crime scene, in *The Practice of Crime Scene Investigation*, J. Horswell, ed, CRC Press, pp. 97–123.
[32] Pountney, H. (1971). *Police Photography*, Elsevier Publishing.
[33] Stoilovic, M. (1991). Detection of semen and blood stains using polilight as a light source, *Forensic Science International* **51**, 289–296.
[34] Lennard, C. (2007). Fingerprint detection: current capabilities, *Australian Journal of Forensic Sciences* **39**(2), 55–71.
[35] Peres, M.R. (ed) (2007). *The Focal Encyclopedia of Photography*, 4th Edition, Focal Press.
[36] Ray, S.F. (1999). *Scientific Photography and Applied Imaging*, Focal Press.
[37] Weiss, S.L. (2009). *Forensic Photography: The Importance of Accuracy*, Pearson Prentice Hall.
[38] Hunter, F., Biver, S. & Fuqua, P. (2007). *Light Science and Magic: An Introduction to Photographic Lighting*, 3rd Edition, Focal Press.
[39] Russ, J.C. (2001). *Forensic Uses of Digital Imaging*, CRC Press.

GLENN PORTER

Photography: Scene *see* Crime Scene Photography: US Perspective

Physical Injury *see* Aggression

Plants *see* Botany

Plethysmography *see* Sex Offenders: Treatment of

PMI *see* Human Remains and Identity

Poisons: Detection of Naturally Occurring Poisons

Introduction

A large number of plants produce compounds that may cause serious illness, injury, or even death. Some plants contain toxic compounds in all parts, whereas other plants are poisonous in some parts and edible in others. Examples include potatoes and tomatoes, which contain toxic alkaloids in the green parts of the plant, but these alkaloids are not present in the ripe potato tuber and tomato fruits.

A comprehensive list of toxic plants is not possible because of the large number of poisonous substances. Although many plants contain toxic substances, only a few species cause poisonings in humans. The actual

number of cases of plant poisoning is hard to determine. Even the numbers of fatal poisoning by plants is not always very obvious. Severe or even deadly poisonings with plants are relatively rare; the frequencies and the plants involved are regionally different.

A published report of the Poison Information Centre in Chile showed that only 0.43% of the medical consultations from January 1998 to June 2000 were with respect to intoxications with plants or mushrooms [1]. Compared to Chile, the Poison Information Centre of northern Germany reported in 2005 that 12% of medical consultations were related to the use and misuse of plants or plant products. Most of these consultations were due to accidental ingestion of the plant material or the misuse of herbal drugs of abuse [2].

Severe and fatal poisonings by plant materials are very rare.

The Poison Information Center, Mainz, Germany, has recorded 31 severe and fatal poisonings caused by plants from January 1995 to July 2007 [3]. The majority of these cases (55%) were caused by an abuse of plant material; only five cases (16%) were a suicide attempt. A total of 5 of the 31 severe poisonings were recorded as fatal.

A study analyzing the self-poisoning fatalities in rural Sri Lanka from March 2002 to March 2003 has shown that 44 of 198 people died from ingestion of the plant oleander [4]. The higher number of plant poisonings in Sri Lanka compared with other countries may also reflect the availability of drugs or poisons used in suicide attempts.

In Australia, the National Coroners Information Service (NCIS) recorded that 9 persons of 131 400 recorded deaths had died from plant or mushroom poisoning in the period between July 2000 and July 2007 [5]. In general, severe or fatal plant poisonings are not very common and are limited to a small number of poisonous plant species.

Manner of Poisoning with Plants

Unintentional Ingestion of Plants and Plant Material

Unintentional ingestion of plant material not only occurs in children, especially younger ones, but also in adults. It is obvious that the reason for accidental ingestion in different age groups is completely different. Whereas children discover their environment by trying bits of plants that appear attractive to them, adults ingest them accidentally [6].

For children, the attractive parts of plants are mostly fruits and seeds with colorful seed coats. Examples include the fruits of deadly nightshade, which look similar to cherries, and the red seed coats of the yew. In the case of yew, the seed coats are very sweet and are not poisonous, but the seeds inside are very poisonous. Children can also be attracted by fruits that are similar to edible plants. A good example is the fruit of laburnum, which looks like peas or beans.

For adults, the reason for an unintentional ingestion of plant material is usually a mix-up. Poisoning occurs as a result of ingestion of poisonous plant due to confusion. The affected persons are usually looking for alternative sources of food or self-medication with herbal drugs. A common example is the collection of meadow saffron instead of wild garlic. Also the mix-up of mushrooms is very common. Many case reports have shown that the deadliest mushroom is the death cap, which looks similar to edible champignons [7].

Accidental ingestion of plant material is also possible by taking herbal medication that has been either incorrectly prepared or in which the wrong plant material has been used. A common problem is the use of aconitum in traditional Chinese medicine, which can be toxic when prepared incorrectly.

Intended Ingestion of Plant Material

Besides unintentional ingestion of plant materials, plants have always been, and are still, used in homicides and suicide attempts. Historical records show that plant materials have caused the death of many famous personalities. An example is the death of Socrates, which was caused by the application of the plant hemlock [8]. But even today, poisonous plants are used from time to time in homicide cases. However, the suicidal ingestion of natural poisons is more common. In developed countries, drug- or poison-related suicide attempts are mostly in connection with prescriptive drugs. However, case reports describe the use of very poisonous plants like deadly nightshade, aconite, and yew in suicides. These patients are often well educated and know about the toxicity of the plants they use. As an example, one patient was growing aconite in his garden, harvested the root of the plant, and prepared an extract, which he ingested and injected. He had

studied literature about the toxicity of this plant and had even shown the literature to an emergency physician [3].

In rural parts of developing countries, suicidal ingestion of plants is far more common due to a lack of availability of prescription drugs [4].

Ingestion of Plant Material Due to an Abuse

The use of plants and plant material for psychoactive effects has been known and recorded for millenniums. The knowledge of the psychoactive effects and the toxic side effects were extensively observed by shamans and other users and has been passed down from generation to generation. In recent years, the use of herbal drugs for their psychoactive properties has become increasingly popular among illicit drug users. The reasons for this increase are diverse. Some of the herbal drugs are not scheduled, and are generally easily accessible, and are sometimes considered to be safe since they are natural.

The assumption of less toxicity is simply not true. Plants often contain pharmacologically highly active compounds. In addition, using hallucinogenic poisonous plants carries the danger of an overdose. The amount of active ingredients in the plant is strongly dependent on geographical and climate conditions. A consumption of two leaves in spring might be hallucinogenic, whereas the consumption of two leaves in autumn may be fatal [6].

Overdoses due to an abuse of plant materials are often described for plants like deadly nightshade [9], datura [10], and kath [11].

Even a nonoverdose situation might cause dangerous behavior. It is well known that the anticholinergic syndrome triggered by atropine or scopolamine causes hallucinations and dry, hot skin. The drug users therefore try to cool down by jumping into water, risking a death by drowning. Hallucinations can also sometimes cause self-inflected injuries. The NCIS (Australia) recorded two cases of death due to self-induced stab wounds under the influence of magic mushrooms [5].

Description of Selected Plants and Fungi

In the following sections toxicologically important plants and some fungi are described in detail. To avoid confusion, scientific names of the plants are used. The scientific names have the advantage that only one name exists for each plant, whereas the commonly used names may not be unique and can be used for more than one plant. Table 1 translates the most commonly used English names of the plants into their scientific names. In the following sections, selected plants and fungi, which are often involved in plant-related poisonings, are described in monographs. The monographs are listed in alphabetical order of the scientific names of the plants. Figure 1 provides representative photographs of some poisonous plants.

Table 1 List of common names of poisonous plants and their associated botanical names

Common English name	Botanical name
Aconite	*Aconitum napellus*
Angel's Trumpet	*Datura stramonium*
Autumn crocus	*Colchicum autumnale*
Belladonna	*Atropa belladonna*
Castor bean	*Ricinus communis*
Castor oil plant	*Ricinus communis*
Deadly nightshade	*Atropa belladonna*
Death cap	*Amanita phalloides*
Devine cactus	*Lophophora wiliamsii*
Dwale	*Atropa belladonna*
European yew	*Taxus baccata*
Golden chain tree	*Laburnum anagyroides*
Hemlock	*Conium maculatum*
Jimson weed	*Datura stramonium*
Kath	*Catha edulis*
Kava	*Piper methysticum*
Laburnum	*Laburnum anagyroides*
Laurier rose	*Nerium oleander*
Magic mushroom	*Psilocybe mexicana*
Meadow saffron	*Colchicum autumnale*
Mescal button	*Lophophora wiliamsii*
Miraa	*Catha edulis*
Monkshood	*Aconitum napellus*
Naked lady	*Colchicum autumnale*
Oleander	*Nerium oleander*
Peyote	*Lophophora wiliamsii*
Poison hemlock	*Conium maculatum*
Quat	*Catha edulis*
Thornapple	*Datura stramonium*
Wolf's Bane	*Aconitum napellus*
Yaqona	*Piper methysticum*
Yew	*Taxus baccata*

Aconitum Napellus

The herbaceous perennial plant *Aconitum napellus* (Figure 1), known as aconite or monkshood, grows to



Figure 1 Flower and leaves of *Aconitum napellus*

a height of 1 m, and is native to western and central Europe. The English name "monkshood" is derived from a part of the dark blue flower of *A. napellus* that has the shape of a cylindrical helmet. The toxicity of *A. napellus* is mentioned in Greek mythology, where it is described as the first poisonous plant. In fact, aconite is the most poisonous plant in central Europe due to its alkaloid-like aconitine.

All species of the gender *Aconitum* are highly toxic and have been used for many centuries as arrow poison or common poison in homicides. Roots, leaves, flowers, and seeds are all exceedingly poisonous to man and livestock. The main active alkaloid in plants from the genus *Aconitum* is aconitine [6].

An accidental poisoning due to a mix-up with other nutrient plants is quite unlikely owing to the lack of similarity. Also, children are not attracted by the plant very often. However, aconite is a source of many fatal poisonings described in case reports [12]. The most common causes of a poisoning with plant material of the genus *Aconitum* are either a suicide attempt or an incorrect preparation of the plant in traditional Chinese medicine. Aconite is used in the traditional Chinese medicine [13], and can be freely purchased from herbal shops around the world. An incorrect preparation, or a wrong dosage of the plant material, may produce a fatal dose of the toxic alkaloid.

Amanita Phalloides

Amanita phalloides is a deadly poisonous mushroom, commonly known as *death cap*. This mushroom appears after periods of rain from late summer to the end of autumn. Death cap originates from Europe, but can be found worldwide. The fungal fruiting body of *A. phalloides* has a convex yellowish or greenish cap, usually 5–15 cm across. The stem is up to 20-cm long. The flesh of the fruiting body is white. With similarities to champignons, parasol mushrooms, and paddy straw mushrooms, this fungus can be confused with other fungi. The toxicity of *A. phalloides* has been extensively studied since the largest number of deadly mushroom poisonings are caused by this species. All parts contain amatoxins and phallotoxins, and the toxin most responsible for the deadly effects is alpha-amanitin [14]. The peptide alpha-amanitin is a very stable compound and even prolonged cooking or maceration with salt does not decrease the toxicity. The toxin acts by inhibition of RNA polymerase II and therefore blocks protein synthesis.

The lethal dose of amanitin is only 0.1 mg kg^{-1}, which means that a medium-sized mushroom can cause the death of a human. Death caps have been reported to have a pleasant taste, and a mix-up cannot be recognized by the taste. The symptoms of a poisoning with *A. phalloides* are initially gastrointestinal in nature. Colicky abdominal pain, vomiting, and watery diarrhea characteristically start approximately 8 h after ingestion. These symptoms resolve in a period of two or three days, giving the false sign of a remission. After a few days, a hepatic and renal failure causes symptoms like jaundice, delirium, seizures, and coma. Death occurs usually 6–16 days after ingestion of the mushroom.

The treatment of a poisoning with death cap includes gastric lavage, activated charcoal, the correction of metabolic acidosis, and an intravenous antidote treatment with silibinin. Silibinin is extracted from blessed milk thistle (*Silibum marianum*) and is supposed to prevent the uptake of amatoxins into liver cells. In some cases, liver transplants have been necessary.

Atropa Belladonna

The perennial plant *Atropa belladonna* (Figure 2) is commonly known as *deadly nightshade*. Other common names such as dwale, death's herb, or witch berry give an impression of its toxicity and use in the middle age. The toxicity and pharmacological effects of the plant are part of the etymology of the botanical name. The genus *Atropa* is named after the goddess



Figure 2 Berry and leaves of *Atropa belladonna*

Atropos, who is known in Greek mythology to cut the life thread. The species name *belladonna* is Italian for beautiful lady and originates from its historical use of the berry juice by women to dilate their pupils [6].

All parts of *A. belladonna* contain toxic tropane alkaloids. Even though the root contains that highest alkaloid concentration, the most dangerous parts, in terms of accidental intoxication, are the berries because of their attractive look and sweet taste. Besides accidental ingestion, poisonings are reported after an abuse of *A. belladonna* due to the hallucinogenic properties of the tropane alkaloids.

The main alkaloids present in *A. belladonna* are l-hyoscyamine and l-scopolamine. Even though only l-hyoscyamine is present in the plant, l-hyoscyamine is converted to a racemic mixture of 50% l-hyoscyamine and d-hyoscyamine. This conversion is either as a result of extraction or release after ingestion. This racemic mixture is called *atropine*.

Atropine acts pharmacologically via blocking acetylcholine receptors of the muscarine subtype. The blockage of these receptors causes symptoms like tachycardia, dilated pupils, decreased gastrointestinal motility, dry hot skin, and dry mouth due to a decreased sweat and saliva production. Apart from these peripheral effects, atropine also affects the central nervous system and causes agitation, disorientation, and hallucinations [15].

Catha Edulis

The evergreen shrub *Catha edulis*, which is native to tropical East Africa and the Arabian Peninsula, is known as *khat* or *qat*. In these countries, the use of khat for therapeutic and recreational purpose has been an integral part of the local culture for centuries. In the second part of the twentieth century, the consumption changed to an uncontrolled abuse and spread throughout other continents. For example, more than 2300 kg of khat was confiscated at Frankfurt Airport (Germany) in 1998 [16]. This new aspect of khat consumption has raised increased concern in international organizations. In 1980, the World Health Organization classified khat as a drug of abuse that is able to produce dependence.

The main psychoactive alkaloids of khat are cathinone and its metabolite cathine, also known as *norpseudoephedrine*. Cathinone was found to have a pharmacological profile closely resembling that of amphetamine. Cathinone acts as a central nervous system stimulant and shows sympathomimetic effects by releasing catecholamines from presynaptic storage sites. Experiments have shown that cathine acts like cathinone, but is less effective. Both substances are controlled substances in many countries due to khat abuse.

Colchicum Autumnale

Colchicum autumnale is a perennial plant that grows from corms. The most common English names are autumn crocus, naked lady, and meadow saffron. These names represent the similarity of the flower to crocuses. However, this similarity is limited to the appearance of the flowers. In contrast to most other plants, the plant flowers in autumn, long after the leaves have died back, and therefore the English name naked lady. The fertilized fruits emerge from the ground with the new leaves appearing the following spring [6].

Colchicine is the main active alkaloid of *Colchicum*. It is present in all parts of the plant, but particularly in the corm, seeds, and flowers. Colchicine inhibits the function of microtubules and therefore acts as a cytotoxic. It also reduces the activity of leukocytes and lymphocytes. This mechanism of action is suspected to be the reason for its success in the treatment of gout. The plant has been used to treat gout for more than 2000 years. Although colchicine has many side effects, it is still recommended as a treatment for acute gout [17].



Figure 3 Habitus of *Conium maculatum* [Image by Thomas Schoepke – http://www.plant-pictures.com]

Conium Maculatum

The 1.5–2-m tall biennial plant *Conium maculatum* (Figure 3) is known as *hemlock* or *poison hemlock*. The smooth green stem of hemlock is characteristically spotted with purple on the lower half. The leaves are finely divided and the white flowers are small and clustered in umbels. The plant is similar to fennel, parsley, or wild carrot. The root of hemlock, white and fleshy, is often unbranched and shows a similarity to parsnip. The toxicity of hemlock has been known in Greece since 399 BC, as the symptoms of the lethal poisoning of Socrates were described by his pupil Plato. In the middle ages, the medicinal use of hemlock was very limited due to its known high toxicity [8]. *Conium maculatum* is native to Europe but has been introduced and grows as a weed in Asia, North America, and Australia.

All parts of the plant are poisonous. The highest concentrations of the toxic alkaloids are found in unripe seeds, but concentrations vary significantly depending on temperature, moisture, and the season. The main toxic alkaloids of hemlock are coniine and γ-coniceine. The exact mechanism of action of these toxic alkaloids is not known. The primary action is on the central nervous system with symptoms similar to nicotine poisoning. The most common symptoms of hemlock poisoning are problems in movement, dilation of pupils, slow and weak pulse, heavy salivation, and nausea. After a severe poisoning, coma and death from respiratory failure are possible [18].



Figure 4 Flower, leaves, and unripe fruit of *Datura strammonium*

Datura Stramonium

The annual plant *Datura stramonium* (Figure 4), which is often found in nutrient-rich soils, grows to a height of up to 1 m. This plant of the nightshade family is commonly known as *thornapple, Jimsonweed*, or *Angel's Trumpet*. The large flowers of *Datura* are white, erect, and tubular. The tubular shape and its psychoactivity are related to the name Angel's Trumpet. The name thornapple is related to the fruits, which are spiky large green capsules containing numerous black seeds. The plant is distributed generally throughout temperate and subtropical regions. The *Brugmansia* species are similar to *Datura* in botanical appearance and are often cultivated in pots as house plants [6].

All parts of *Datura* and *Brugmansia* are toxic and contain tropane alkaloids. As already described for *A. belladonna*, the main alkaloids are l-hyoscyamine and l-scopolamine. Owing to the presence of these alkaloids, these plants are often abused for their hallucinogenic properties. This increasing misuse has forced a prohibition by law in Florida against planting Angel's Trumpets.

The symptoms of a poisoning as well as the treatment are similar to those described for *A. belladonna*.

Laburnum Anagyroides

Laburnum anagyroides (Figure 5), also known as *laburnum*, grows as a shrub or a small tree. The plant has yellow flowers in pendulous racemes, and therefore it is known in German as "Goldregen", which is often translated as golden rain acacia.



Figure 5 Flower of *Laburnum anagyroides* [Image by Thomas Schoepke – http://www.plant-pictures.com]

Because the plant is attractive and frost tolerant, laburnum is very popular as ornaments in parks and gardens. The fruits are silky hairy pods; the unripe pods are similar to bean or pea pods. The seeds are similar to small beans [6].

All parts of laburnum are toxic and contain quinolizidine alkaloids. The main toxic alkaloid is cytisine, with the highest concentration detected in the ripe seeds and seed pods. Knowing about the toxicity, laburnum has been used in traditional medicine by American Indians, who have consumed the seeds for their emetic effects during rites and magical practices. During the Second World War, the leaves of laburnum were used as a tobacco substitute due to its similar effects. Studies have also shown that cytisine is effective as an aid to smoking cessation [19].

Cytisine binds with a high affinity to nicotinergic acetylcholine receptors. Like nicotine, cytisine acts as a blocking agent on the central nervous system via an overstimulation of these receptors. Owing to the mechanism of action, the symptoms of a poisoning are similar to a nicotine overdose. The central-stimulating effects can cause delirium and convulsions [19]. Death is possible through a respiratory paralysis or failure of the circulatory system. The treatment of a poisoning with laburnum is symptom orientated, with no specific antidote available [6].

Lophophora Williamsii

The small spineless cactus *Lophophora williamsii* is commonly known as *peyote*. This cactus grows extremely slowly and flowers sporadically. It takes in the wild up to 30 years to reach the size of a golf ball and to produce the first flowers. The small pink fruits of peyote are sweet tasting and delicate. The cactus is native to southern United States and Mexico, but is cultivated all over the world.

A recent study has shown prehistoric use of peyote by native North Americans. A radiocarbon study dated dried cacti, so-called mescal buttons, to the time interval 3780–3660 BC. Because in these dated buttons psychotropic alkaloids were present, it was concluded that they were used for their psychoactive effects [20].

The main reason for the psychoactive effects of peyote is due to the presence of the phenethylamine alkaloid mescaline. The effects of peyote and mescaline in humans are well studied. Native peyote cults used the cactus because it produces rich visual hallucinations. These effects were also used in psychiatric studies as a chemically induced model of mental illness. The mechanism of action is similar to that of lysergic acid diethylamide (LSD) or psilocin. These substances act as partial agonists at 5-hydroxytryptamine (5-HT) receptors. Although the acute toxicity of peyote or mescaline is not as high as other herbal alkaloids, fatalities have been described. Owing to the strong hallucinations, fights among drug abusers or self-harm situations are not uncommon.

Nerium Oleander

The evergreen shrub *Nerium oleander* (Figure 6), simply known as *oleander* is native to northern parts of Africa and the Mediterranean region. The scientific name is deduced from their preference to grow near water; Nero is the Greek word for water. The leaves of oleander are thick and leathery, dark green and narrow lanceolate, and up to 20-cm long. The leaves grow typically in pairs or whirls of three. The flowers of oleander are white, pink, or yellow, are up to 5 cm in diameter, and grow in clusters. Even though the flowers of *N. oleander* can be slightly yellow, the so-called yellow oleander is *Thevetia peruviana*. Both oleander and yellow oleander can easily be grown in warm subtropical regions and are extensively used in parks and roadsides worldwide [6].

Both plants are highly toxic and contain cardiac glycosides of the cardenolide type in all parts. The highest glycoside concentrations were detected in the



Figure 6 Flower and leaves of *Nerium oleander*

seeds of oleander, and the main glycoside present in these plants is oleandrin. Oleandrin acts like other cardiac (cardiac) glycosides of the cardenolide type by inhibition of the sodium potassium exchange, which causes an increased calcium level in heart cells, which increases contractions of heart muscle cells. An overdose of the cardiac glycosides causes dysrhythmia and a possible heart block. The first scientific study showing the therapeutic usefulness of cardiac glycosides was published in 1785 by Withering [21]. The therapeutic range of the cardiac glycosides is narrow. Many fatal and severe poisonings have been reported after ingestion of oleander or other plants containing cardiac glycosides, such as *Digitalis purpurea* (foxglove), *Adonis vernalis* (pheasant's eye), or *Convallaria majalis* (lily of the valley). The typical symptoms of an overdose with cardiac glycosides are dizziness and vomiting, followed by cardiac arrhythmias [6].

The treatment of an overdose is dependent on the severity of the poisoning. For treatment of severe poisonings, an antibody-based antidote has been developed to specifically remove the glycosides. In case of poisoning with plants or self-prepared extracts, this antidote is unfortunately poorly effective. The antidote removes therapeutically applied pure glycosides, but it is likely that some cardioactive glycosides are not removed [6].

Piper Methysticum

The common English name for the western pacific plant *Piper methysticum* is kava. Other names for the plant are 'awa', used in Hawaii, or 'yaqona', common in Fiji. Kava is an evergreen bush growing up to 3 min height, with heart-shaped leaves up to 20 cm in length. The plant is closely related to black pepper (*Piper nigra*) and also has a spicy taste. Kava is psychoactive, indicated by the scientific species name *methysticum*, which is Greek for intoxicating. Kava is consumed on many western pacific islands and Australia [6]. Traditionally Kava is prepared by grinding or chewing the rhizome, which is mixed with water or coconut milk. The effects after consumption of kava are talkative and euphoric behavior, anxiolytic effects, sense of well-being, clear thinking, and relaxed muscles. The plant contains a mix of kavalactones and kavapyrones. Extracts of the plant were introduced into modern medicine as a mild anxiolytic. After the report of some deaths due to its medicinal use, kava medicines were banned. Kava-containing medicine causes acute liver failure [22]. The traditional use of kava by Pacific Islanders and by some aboriginal communities is not believed to be associated with liver damage. A recent study has shown that kava feeding in rats does not cause liver damage [23]. Further investigations are necessary to demonstrate the long-term safety of kava preparations.

Psilocybe Species

The small brown mushroom genus *Psilocybe* sp is best known for their psychoactive properties and are therefore called "magic mushrooms". The fruiting bodies of the magic mushrooms are small to medium in size and typically show a brown coloration. Hallucinogenic species of *Psilocybe* can be found in temperate regions throughout the world [24]. The psychoactive species contain as active ingredients psilocin and psilocybin. These compounds are structurally related to serotonin, a neurotransmitter in the central nervous system. Psilocin and psilocybin cause effects like the synthetic drug of abuse LSD. The effects of intoxication with magic mushrooms are mainly hallucination; the acute toxicity of the compounds is low. However, the consumption of these compounds is dangerous. Severe hallucinations can cause self-inflected injuries. Deaths due to self-induced stabbings following consumption of magic mushrooms have been recorded [5]. Two patients died from the wounds after they had stabbed themselves in the chest while under the influence of magic mushrooms.



Figure 7 Leaves of *Ricinus communis*



Figure 8 Fruit and leaves of *Taxus baccata*

Ricinus Communis

Ricinus communis (Figure 7), the castor oil plant, is cultivated all over the globe for oil production. The annual bushy shrub is believed to have its origin in tropical Africa, but grows nowadays worldwide. The plant with glossy palmately divided leaves is often used for ornamental purposes in gardens. The fruit of *R. communis* is a soft, spiny capsule with three almost oval seeds, which are the so-called castor beans. The attractive seeds have a hazelnut-like taste and contain 45–55% of fatty oil. The seeds also contain up to 25% protein. The protein fraction contains the highly toxic lectin ricin, which is poisonous via ingestion, inhalation, or injection. Ricin acts via an inhibition of protein synthesis. As a lectin, it binds to glycoproteins, facilitating the entry of the toxin into the cytosol [25].

After ingestion of ricin, the symptoms are nausea and diarrhea. In severe poisonings, liver and renal dysfunction, and possibly death, occurs. After inhalation, coughing and dyspnea can occur. These symptoms can progress to respiratory distress and death. The injection of ricin causes symptoms of general weakness and myalgias. Death is possible due to hypotension and multiorgan failure.

Ricin poisoning is possible by accidental or suicidal eating of the seeds. After oral ingestion of the seeds, the toxicity depends on how well the seeds are chewed. Some factors make the castor beans dangerous, such as their attractive appearance and the stability of ricin toward proteolytic enzymes. The treatment of a poisoning with ricin is recommended to be symptomatic. No specific antidote for ricin poisoning is available [25].

Taxus Baccata

The conifer *Taxus baccata* (Figure 8) is widely known as yew, and is often considered to be the oldest plant in Europe. The age of some yew trees is estimated to be 5000 years. The plant is a small to medium-sized tree, which grows relatively slowly. The leaves of this conifer are dark green and lanceolate, with a length of up to 4 cm. The plant has very characteristic single seed cones surrounded by a bright red colored berrylike structure, called an *aril*. All parts of the plant are highly toxic, except the arils. This enables the cones (including the arils) to be eaten by birds, but the seeds remain undamaged in the bird's droppings. The plant contains pseudoalkaloids of the taxane type. The main compound responsible for the toxicity of the European yew (*T. baccata*) is taxine B [6].

Some taxane-type pseudoalkaloids are also important in the treatment of cancer. In recent years, Taxol®, which contains paclitaxel, has become of particular interest in the treatment of ovarian, breast, and non-small-cell lung cancer. Paclitaxel has been isolated from the Pacific yew, *Taxus brevifolia*. Because of the extensive use of Taxol® and the fact that more than 1000 trees are needed to obtain 1 kg of paclitaxel, pharmaceutical companies have found an alternative source of the drug. A precursor of paclitaxel is isolated from cultivated European yew plants, and paclitaxel is synthesized using this precursor. The mechanism of action of taxanes is an inhibition of cell division by stabilization of the microtubuli. Therefore, these substances have cytotoxic effects.

In the case of an overdose with yew plants, the symptoms are nausea, dizziness, abdominal pain,

shallow breathing, and tachycardia. Death can occur as a result of respiratory paralysis, with the heart in diastolic arrest. The treatment of a poisoning with yew is symptomatic, with no specific antidote available [6].

Analytical Methods

Because of the diversity of the toxic compounds in plants and mushrooms, a general laboratory screening procedure is not possible. Therefore, various specific methods for detection have been developed. Table 2 summarizes the major active compounds and the most common detection methods of the plants described in this article.

Table 2 List of botanical names, the corresponding main toxic compounds, and the common detection of these compounds in biological fluids

Scientific name	Main toxic compound(s)	Common detection in biological fluids
Aconitum napellus	Aconitine	HPLC, LC-MS
Amanita phalloides	Amanitine	Immunoassay, LC-MS
Atropa belladonna	Hyoscyamine, scopolamine	GC-MS, LC-MS
Catha edulis	Cathinone	GC-MS, LC-MS
Colchicum autumnale	Colchicine	HPLC, LC-MS
Conium maculatum	Coniine	GC-MS, LC-MS
Datura stramonium	Hyoscyamine, scopolamine	GC-MS, LC-MS
Laburnum anagyroides	Cytisine	LC-MS
Lophophora wiliamsii	Mescaline	GC-MS, LC-MS
Nerium oleander	Oleandrine	Immunoassay, LC-MS
Piper methysticum	Kavaine	HPLC
Psilocybe mexicana	Psilocin, psilocybin	GC-MS
Ricinus communis	Ricin	Immunoassay
Taxus baccata	Taxine B	LC-MS

Methods for detection of naturally occurring toxic compounds include high-performance liquid chromatography (HPLC), liquid chromatography mass spectrometry (LC-MS), gas chromatography mass spectrometry (GC-MS), and various immunoassay techniques. The use of HPLC is common for detection of compounds like aconitine, colchicines, and the ingredients of kava. Kavalactones and kavapyrones are commonly used for the detection of aconitine and colchicine. The modern and more sensitive LC-MS techniques have been applied to many naturally occurring substances, especially if low concentrations have to be detected. The use of GC-MS techniques is applicable if the substances are thermostable and provides a powerful tool for detection of alkaloids like atropine, cathinone, and mescaline. While GC-MS techniques are useful, LC-MS techniques are preferred for the detection of low concentrations in blood.

The use of immunoassay techniques is possible for some poisonous compounds of herbal origin. It has been shown that the cardiac glycosides of oleander can be detected using an immunoassay technique for cardiac glycosides. This technique is not able to differentiate between the cardiac glycosides. A differentiation is possible using HPLC techniques. Also the detection of ricin is commonly performed using immunoassay techniques, such as radioimmunoassay (RIA) and enzyme-linked immunosorbent assay (ELISA). Owing to the physicochemical properties of the protein ricin, chromatographic techniques are more challenging and cannot at the present time be easily confirmed by conventional chromatographic techniques due to its large molecular size.

Detection methods using immunoassay techniques were also developed for the proof of death cap poisoning. Amanitin can be detected in urine up to 24 h after ingestion by the use of immunoassay techniques.

A postmortem detection of the toxic substances is often challenging. Determination of these is often likely in urine, but for some compounds like colchicine bile is the most useful specimen for postmortem analysis. Some toxins like amanitin are unlikely to be detected in postmortem specimen because death occurs many days after ingestion.

Some substances like psilocin and psilocybin are chemically unstable; hence concentrations need to be interpreted carefully.

See **Toxicology: Initial Testing** for further information on general drug detection techniques.

Summary and Conclusions

Fatal plant and mushroom poisonings are relatively rare. The diversity of poisonous compounds and their widespread occurrence in plants and fungi makes a comprehensive list of dangerous flora impossible. Besides, the majority of fatal cases are limited to a small number of plants and mushrooms. These poisonings can occur either after unintentional or intended ingestion or after abuse for their hallucinogenic effects. While plants and fungi can be dangerous to humans, fatal cases are avoidable. However, the analytical detection of a poisoning from an unknown plant or fungus can be a challenge for toxicologists.

References

[1] Manriquez, O., Varas, J., Rios, J.C., Concha, F. & Paris, E. (2002). Analysis of 156 cases of plant intoxication received in the Toxicologic Information Center at Catholic University of Chile, *Veterinary and Human Toxicology* **44**(1), 31–32.

[2] Poison Information Centre Goettingen (2005). *Harmonized Annual Report 2005*.

[3] Ritter-Weilemann, I. (1995–2007). *Annual Reports*, Poison Information Centre, Mainz.

[4] Eddleston, M., Gunnell, D., Karunaratne, A., de Silva, D., Sheriff, M.H. & Buckley, N.A. (2005). Epidemiology of intentional self-poisoning in rural Sri Lanka, *The British Journal of Psychiatry* **187**, 583–584.

[5] Australian National Coroners Information System (NCIS), search conducted in July 2007. www.ncis.org.au.

[6] Frohne, D. (2005). *Poisonous Plants: A Handbook for Doctors, Pharmacists, Toxicologists, Biologists and Veterinarians/Dietrich Frohne and Hans Jrgen Pfnder*, 2nd Edition, Manson Publishing, London.

[7] Trim, G.M., Lepp, H., Hall, M.J., McKeown, R.V., McCaughan, G.W., Duggin, G.G. & Le Couteur, D.G. (1999). Poisoning by *Amanita phalloides* ("deathcap") mushrooms in the Australian Capital Territory, *The Medical Journal of Australia* **171**(5), 247–249.

[8] Daugherty, C.G. (1995). The death of Socrates and the toxicology of hemlock, *Journal of Medical Biography* **3**(3), 178–182.

[9] Jaspersen-Schib, R., Theus, L., Guirguis-Oeschger, M., Gossweiler, B. & Meier-Abt, P.J. (1996). Serious plant poisonings in Switzerland 1966–1994. Case analysis from the Swiss Toxicology Information Center, *Schweizerische Medizinische Wochenschrift* **126**(25), 1085–1098.

[10] Boumba, V.A., Mitselou, A. & Vougiouklakis, T. (2004). Fatal poisoning from ingestion of *Datura stramonium* seeds, *Veterinary and Human Toxicology* **46**(2), 81–82.

[11] Giannini, A.J. & Castellani, S. (1982). A manic-like psychosis due to khat (*Catha edulis* Forsk.), *Journal of Toxicology. Clinical Toxicology* **19**(5), 455–459.

[12] Elliott, S.P. (2002). A case of fatal poisoning with the aconite plant: quantitative analysis in biological fluid, *Science and Justice* **42**(2), 111–115.

[13] Chan, T.Y., Chan, J.C., Tomlinson, B. & Critchley, J.A. (1993). Chinese herbal medicines revisited: a Hong Kong perspective, *Lancet* **342**(8886–8887), 1532–1534.

[14] Wieland, T. (1967). The toxic peptides of Amanita phalloides, *Fortschritte der Chemie Organischer Naturstoffe* **25**, 214–250.

[15] Rang, H.P. (1987). *Pharmacology*, 2nd Edition, H.P. Rang & M.M. Dale, eds, Churchill Livingstone, Edinburgh, pp. 547–567.

[16] Toennes, S.W., Harder, S., Schramm, M., Niess, C. & Kauert, G.F. (2003). Pharmacokinetics of cathinone, cathine and norephedrine after the chewing of khat leaves, *British Journal of Clinical Pharmacology* **56**(1), 125–130.

[17] (2006). Gout: finally, diagnosis and treatment guidelines. A European task force offers the first recommendations on dealing with this painful arthritic condition, *Health News*, **12**(10), 10–11.

[18] Drummer, O.H., Roberts, A.N., Bedford, P.J., Crump, K.L. & Phelan, M.H. (1995). Three deaths from hemlock poisoning, *The Medical Journal of Australia* **162**(11), 592–593.

[19] Tutka, P. & Zatonski, W. (2006). Cytisine for the treatment of nicotine addiction: from a molecule to therapeutic efficacy, *Pharmacological Reports* **58**(6), 777–798.

[20] El-Seedi, H.R., De Smet, P.A., Beck, O., Possnert, G. & Bruhn, J.G. (2005). Prehistoric peyote use: alkaloid analysis and radiocarbon dating of archaeological specimens of Lophophora from Texas, *Journal of Ethnopharmacology* **101**(1–3), 238–242.

[21] Wade, O.L. (1986). Digoxin 1785–1985. I. Two hundred years of digitalis, *Journal of Clinical and Hospital Pharmacy* **11**(1), 3–9.

[22] (2002). Kava kava may cause irreversible liver damage, *South African Medical Journal* **92**(12), 961.

[23] DiSilvestro, R.A., Zhang, W. & DiSilvestro, D.J. (2007). Kava feeding in rats does not cause liver injury nor enhance galactosamine-induced hepatitis, *Food and Chemical Toxicology* **45**(7), 1293–1300.

[24] Musshoff, F., Madea, B. & Beike, J. (2000). Hallucinogenic mushrooms on the German market – simple

instructions for examination and identification, *Forensic Science International* **113**(1–3), 389–395.
[25] Audi, J., Belson, M., Patel, M., Schier, J. & Osterloh, J. (2005). Ricin poisoning: a comprehensive review, *Journal of the American Medical Association* **294**(18), 2342–2351.

JOCHEN BEYER

Poisons: Natural *see* Poisons: Detection of Naturally Occurring Poisons

Police Interviews and Interrogations *see* Confessions: Evidentiary Reliability of

Police Use of Force

Traditionally, the hallmarks of policing involve officers responding to criminal and noncriminal calls for police service, detecting and preventing criminal activity, investigating past crimes, and enforcing laws. To carry out these activities, the police have institutional authority to use force. Such authority, power, or legal right comes from rules of law, which recognize that the police must sometimes use coercive action against law-violating citizens to accomplish legal objectives. Because police agency records and published statistics such as the annual figures reported by the Bureau of Justice Statistics consistently show that some citizens are willing to violate the law, use of force by the police remains a central part of their occupation (*see also* **Policing and Critical Incident Teams**).

When the police use force, they have the discretionary power to choose from a range of possible responses available in a given law enforcement activity. Responses can include the police using their presence or show of authority to using deadly force to deal with law-breaking behaviors by citizens. Whenever the police invoke their legal right to use force against citizens, there are court decisions regarding legal standards for making the right force choice.

Courts in the United States, where the use of force by police is subject to constitutional restraints, generally apply three major legal decisions that clarify standards surrounding the appropriateness of police use of force: *West v. Atkins* [1], *Graham v. Connor* [2], and *Tennessee v. Garner* [3]. Under the umbrella of *West v. Atkins*, the police must act under color of law to be liable for use-of-force actions against citizens that amount to violations of police authority. The police must make objectively reasonable use-of-force choices under the *Graham* rules, which include a balancing test, subjective, objective, hindsight, and totality of circumstances tests. If the police decide to use deadly force against law-violators, then the *Garner* rules highlight three conditions in which deadly force may be reasonable: deadly force is necessary, suspects are dangerous, and the police are able to warn suspects.

Standards Governing Acting Under Color of Law

Before courts consider allegations that the police used wrongful or excessive force against citizens, they make legal determinations on whether the police were acting in their official capacity; that is, whether the police used force while "acting under color of law" whereby they exercised a power or right granted by law [1]. Generally, the use of force by the police against law-violating citizens is a law enforcement activity normally carried out under color of law.

Not all uses of force by the police, however, amount to acts under color of law. The question of official conduct is not always clear especially during off-duty hours, during acts of self-defense that happen in personal circumstances [4], or during acts of force that occur in secondary employment settings such as security venues. When making color of law determinations in these situations, courts weigh heavily the nature of the officer's behavior [5]. They have considered behaviors such as flashing of a badge [6], wearing a police uniform [7], identifying themselves as police officers [4], driving a marked police vehicle [6], or brandishing a department issued weapon [8], in concert with force as acts under color of law. Courts may also consider off-duty police acts within their jurisdiction and on which they file

official police reports as behaviors that constitute acting under color of law [9].

Although sometimes the question of official conduct is ambiguous, courts have employed a three-prong test in color of law determinations: public function test, state compulsion test, and nexus test [10, 11]. For example, under the public function test, the court would consider whether an officer's use of force was a state action normally reserved for the police and usually performed by them during their official duties, or a personal action [12]. Making investigatory stops, arrests, or performing searches are uses of force that the police usually carry out under color of law [8, 13].

Under the state compulsion test, the court would determine whether the state or government compelled the officer to use force to enforce a government interest such as arresting a citizen who commits a crime of domestic violence. The central question is whether the use of force was an action required by law or police department policy [7]. Finally, the nexus test would involve the court's consideration of whether the officer's use of force involved conduct closely linked to or associated with the state [14]. For example, the action of an off-duty officer who grabs and arrests a citizen who commits a larceny may constitute state action because there is a sufficiently close nexus between the officer's conduct and the state's regulation of arrest powers.

While courts have employed the police function, state compulsion, and nexus tests in making legal decisions about police uses of force that fall under color of law, they have not used them in a formulaic fashion. Instead, they have considered the totality of circumstances surrounding each particular use of force case when employing them. To hold the police liable for wrongful uses of force that amount to constitutional violations, courts must first find whether they were clothed with official authority.

Standards Governing Objective Reasonableness

Whenever the police use force while acting under color of law, they must ensure that their use-of-force tactics do not violate the constitutional rights of citizens. Police uses of force against free citizens mostly occur in the course of an arrest, investigatory stop, or other seizure. The Fourth Amendment and its "reasonableness" standard is the precise constitutional right that provides free citizens protection against unreasonable seizures or excessive force by the police. In *Graham v. Connor* [2], the court established five substantive tests for judging the reasonableness of police use of deadly or nondeadly force against citizens: balancing test, subjective test, objective test, hindsight test, and totality of circumstances test.

Using the balancing test, courts weigh the rights of free citizens under the Fourth Amendment against the interests of the government to take action against them. For example, an officer has a legal interest to stop or seize a motorist who travels through a stop sign without stopping. Yet, the motorist has a right to travel freely along the highway without police interference if the officer cannot substantiate a legal interest to carry out the traffic stop.

Where the police establish a legal interest to take action against citizens, the Fourth Amendment recognizes that the interest carries with it the need to use some degree of physical coercion or threat [15]. Built into the application of the balancing test is a principle of proportionality: Is the officer's act of force in the correct relationship to the citizen's violation of law? For example, a suspect disobeys an officer's verbal commands during an arrest. This form of behavior is less severe than is the suspect shooting at the officer. Both suspect behaviors require some degree of force by the officer to handle them and complete the arrest, but at obviously different levels. Therefore, courts balance the amount of force the police use against the amount of force they need to use in a particular situation [16].

When employing the objective test, courts consider whether another well-trained officer under the same set of circumstances would observe and conclude that the use of force by the police was reasonable. It requires a retrospective investigation and opinion by the well-trained officer (or expert) who has special knowledge of issues surrounding police use of force. Both carrying out the investigation and giving an opinion involve seeing through the lens of the officer on scene: what the officer's observations were; what the officer's observations meant; what the officer's experience and situational knowledge were; and what the officer's realty was. The expert weighs heavily the citizen's behavior because it has a large degree of power in influencing the officer's choice of force [17]. In the light of the particular circumstances, the expert conveys to the court whether he or she

believes that the officer's choice of force was within the range of objectively reasonable options available because there is no precise formula for determining a single or best force option.

Under *Graham*, the subjective test discounts the personal motivations of the police because they have no value in the court's application of the Fourth Amendment's reasonableness standard. For example, an officer's evil intentions do not raise a Fourth Amendment claim of excessive force when his or her use of force was objectively reasonable. Alternatively, the officer's good intentions do not make an objectively unreasonable use of force constitutional.

Using the hindsight test, courts would not consider every grab or push by the police even if they later seem unnecessary as violations of the Fourth Amendment's protection against unreasonable seizures. Courts make allowances for errors by the police who must at times make split-second decisions about using force in situations that are tense and uncertain, and that evolve rapidly. What gives rise to constitutional violations are more than police acts that amount to mere mistakes. For example, having probable cause to arrest a person but arresting the wrong person, or having a warrant to search a house but carrying it out at the wrong house would not in every case violate the Fourth Amendment's reasonableness standard. The hindsight test takes into account police mistakes and uses the at-the-moment perception of another well-trained officer on scene as the basis for a reasonableness inquiry and not the after-the-fact perception of others.

The last *Graham* test – totality of circumstances – involves evaluating the circumstances of a use of force event. Because the Fourth Amendment's reasonableness standard is not capable of an exact definition, its application requires attention to the totality of circumstances. The central question is whether the totality of circumstances justifies a particular use of force by the police. Circumstances that the police know before and when they use force are relevant in courts' determinations of reasonableness [18].

In *Graham*, the court suggests careful attention to the severity of the crime; whether the suspect poses an immediate threat to the police or others nearby, the suspect attempts to escape police custody, or the suspect fights against arrest. On the face of these circumstances, deciding to use deadly force to shoot a suspect of a felony crime who resists the police and attempts to escape arrest might seem permissible. The calculus of reasonableness in deadly force cases, however, requires some attention to the *Garner* [3] rules because *Graham v. Connor* is a nondeadly force case.

Standards Governing Deadly Force

The only US Supreme Court decision that deals directly with the use of deadly force by the police is *Tennessee v. Garner* [3]. Under the *Garner* rules, police may use weapons or other use-of-force tactics that amount to deadly force to prevent the escape and make the arrest of fleeing felons under three conditions. First, the police have probable cause to believe that a suspect poses a threat of significant physical harm to them or others: the danger condition. The immediacy of the threat and the dangerousness of it are the cornerstones of this condition. The dangerousness element suggests that suspects who threaten the police or others with weapons "or" suspects who commit crimes where they cause or threaten to cause significant physical harm are dangerous because of their violent or potential violent behavior. Under *Garner*, the "or" aspect of dangerousness suggests that the possession of a weapon or the threat to use one is not necessary to satisfy dangerousness. Courts, however, have considered suspects armed with guns, knives, or flashlights, and suspects who have used a vehicle as a weapon against an officer or have attempted to seize an officer's weapon as dangerous [19–24]. They have recognized murder, bank robbery, and armed robbery as felony crimes that demonstrate dangerous behaviors that justify the use of deadly force [25–27].

The immediacy element of a suspect's physical threat to the police or others suggests that the period during which the threat occurs and the threat actually happens is important. Unfortunately, not all courts measure this period the same. Some may apply an "imminent" yardstick whereby significant physical harm is about to happen such as a suspect pointing a handgun at an officer [28]. Others may use more than a yardstick whereby they consider the "unpredictability" of a dangerous suspect, which is not easily foreseeable and measurable [29]. Nevertheless, courts carefully examine the totality of circumstances in all cases when making determinations about immediacy.

Second, the police may only use deadly force against a dangerous suspect when necessary: the

necessary condition. Unless the suspect satisfies the danger condition, deadly force is unnecessary. What is necessary or needed, however, may invite hindsight arguments. For example, in *Plakas v. Drinski* [30], Plakas ran at an officer and tried to use a fireplace poker to kill him. The officer shot and killed Plakas. In this case, the plaintiff argued that the officer had nondeadly force options available and that the officer did not try them. The officer could have used a spray or a police dog to disarm Plakas; deadly force was unnecessary. At both the district and federal court levels, the courts ruled in this case that the officer's force was reasonable. The Fourth Amendment's reasonableness standard does not require police to use less intrusive uses of force. Federal courts have rejected necessary arguments that the police could have used nondeadly force options when deadly force ones were reasonable [31].

Third, the police must warn – where feasible – a suspect of their intention to use deadly force: the warning condition. Unless the suspect satisfies the danger condition, both a warning and deadly force are unnecessary. The use of deadly force may also be unnecessary if a dangerous suspect submits to an arrest after a warning.

Whether the police make decisions to use deadly or nondeadly force options to seize free citizens, their choices must fall within the range of objectively reasonable options available in particular situations. The Fourth Amendment's reasonableness standard is the appropriate test in claims that the police used excessive force in the course of an arrest, an investigatory stop, or other seizures. However, not every police abuse of force occurs under these conditions. An officer who uses excessive force against a prisoner would violate either the Fourteenth Amendment's due process clause or the Eighth Amendment's prohibition against cruel and unusual punishment. It is important then to identify the specific constitutional right that is the precise textual source of an officer's use-of-force conduct [2].

References

[1] West v. Atkins, 487 U.S. 42 (1988).
[2] Graham v. Connor, 490 U.S. 386 (1989).
[3] Tennessee v. Garner, 471 U.S. 1 (1985).
[4] Huffman v. County of Los Angeles, 147 F.3d 1054, 1058 (9th Cir., 1998).
[5] Stengel v. Belcher, 522 F.2d 438, 441 (6th Cir., 1975).
[6] Neuens v. City of Columbus, 303 F.3d 667 (6th Cir., 2002).
[7] Roe v. Humke, 128 F.3d 1213, 1216 (8th Cir., 1997).
[8] Abraham v. Raso, 183 F.3d 279, 287 (3d Cir., 1999).
[9] Kappeler, V.E. (2006). *Critical Issues in Police Civil Liability*, Waveland Press, Illinois.
[10] Ellison v. Garbarino, 48 F.3d 192, 195 (6th Cir., 1995).
[11] Wolotsky v. Huhn, 960 F.2d 1331 (6th Cir., 1992).
[12] Bonsignore v. City of New York, 683 F.2d (2nd Cir., 1982).
[13] Pickrel v. City of Springfield, 45 F.3d 1115 (7th Cir., 1995).
[14] Cooper v. Parrish, 203 F.3d 937, 952 (6th Cir., 2000).
[15] Terry v. Ohio, 392 U.S. 1 (1968).
[16] Flores v. City of Palacios, 381 F.3d 391 (5th Cir., 2004).
[17] Adams, K. (1999). What we know about police use of force, in *Use of Force by Police: Overview of National and Local Data*, J. Travis, J.M. Chaiken & R.J. Kaminski, eds, U.S. Department of Justice, National Institute of Justice and Bureau of Justice Statistics, Washington, DC. pp. 1–14.
[18] Palmquist v. Selvik, 111 F.3d 1332 (7th Cir., 1997).
[19] Butler v. City of Detroit, 386 N.W.2d 645 (Mich. App. 1985).
[20] Ealy v. City of Detroit, 375 N.W.2d 435 (Mich. App. 1985).
[21] Haineze v. Allison, 216 F.3d 1081 (5th Cir., 2000).
[22] Nelson v. County of Wright, 162 F.3d 986 (5th Cir., 1988).
[23] Pittman v. Nelms I.I.I., 87 F.3d 116 (4th Cir., 1996).
[24] Rhiner v. City of Clive, 373 N.W.2d 466 (Iowa, 1985).
[25] Ford v. Childress, 650 F.Supp. 110 (D.C. Ill. 1986).
[26] Ryder v. City of Topeka, 814 F.2d 1412 (10th Cir., 1987).
[27] Trejo v. Wattles, 654 F.Supp. 1143 (D. Colo. 1987).
[28] Boyd v. Baeppler, 215 F.3d 594 (6th Cir., 2000).
[29] Hegarty v. Somerset County, 53 F.3d 1367 (1st Cir., 1995).
[30] Plakas v. Drinski, 19 F.3d 1143 (7th Cir., 1994).
[31] Scott v. Henrich, 39 F.3d 912 (9th Cir., 1994).

FRANK J. GALLO

Policing and Critical Incident Teams

Police calls for service occasionally require the police to resolve high-threat or special-threat situations such as barricaded suspects, hostage situations, drug raids, or warrant services. Although infrequent, these events are significantly different from usual police work that is often less dangerous: They require some degree of

special handling. In US law enforcement circles, such special teams have become known by the acronym SWAT, which stands for special weapons and tactics. In other countries, such special teams may be designated by other terms. Since the 1965 Watts riots in Los Angeles, California, the need for SWAT teams (at times called *police paramilitary units, critical incident teams, special response teams*, and other names in the police literature) to deal with special-threat situations that regular officers are traditionally unprepared to handle has become popular among American law enforcement agencies [1].

Police departments of all sizes have working SWAT teams and some are developing them [2, 3]. They most often employ, when available, their SWAT teams to deal with rare high-threat police call outs. Officers who are SWAT members receive special training to work as a team and to use special weapons and force tactics to handle the most dangerous call outs. Because their behaviors involve special uses of force, there are some common risks related to the legal process.

The adequacy of a SWAT team and its training in particular uses of special weaponry, technology, and tactics to confront the most dangerous situations may give rise to a legal action when the team fails to train in the proper use of issued equipment and use of force tactics. When a SWAT team uses force, its choices must be "objectively reasonable." In cases of alleged use of excessive force, criminal and civil liability can attach itself in the form of a team's wrongful actions under Titles 18 and 42 of the US Code respectively. There are legal risks surrounding the intervention of a SWAT team to control and overcome hostage takers where the main concern is the protection of human life and the avoidance of hostage and bystander injuries.

Adequate Training

Police departments that have SWAT teams often call upon them to resolve the most dangerous police–citizen contacts. SWAT teams commonly carry out high-threat warrant services where the police believe that suspects are armed and dangerous. They execute high-threat narcotics search warrants where drug dealers seem likely to defend their drugs and homes using weapons. Sometimes, SWAT team interventions require team members to confront, disarm, and arrest suspects who hold citizen prisoners.

To assist SWAT teams in handling and controlling extreme police callouts, police departments employ their SWAT team members with special weaponry that are not available to regular officers. For example, some police agencies issue high-powered rifles, automatic firearms, diversionary devices, ballistic shields, and chemical munitions [4]. They authorize their SWAT teams to carry out tactical operations that involve the use of special tactics such as warrant, warrantless, or "no-knock" forced building entries. Because police departments arm their SWAT teams with special weapons and tactics to help them to enforce laws and make arrests, they need to train them on their appropriate uses. A failure to train may amount to a "conscious choice" or "deliberate indifference" to the constitutional rights of citizens to be free from unreasonable uses by SWAT teams [5].

Without training, SWAT teams are likely to make mistakes when deploying SWAT such as using diversionary devices to help them to carry out no-knock warrants. It is not reasonable to expect common officers to know the right force option without training [6]. SWAT teams must receive training in specialized tasks they are likely to perform on-the-job [5]. A failure to train them might give rise to a federal cause of action. Community stakeholders could consider police agencies to be deliberately indifferent to the needs of SWAT teams to receive specialized training.

In resolving the responsibility to train, police departments must focus on the adequacy of its training programs to meet the plausible conditions under which its SWAT teams works [5]. Courts suggest that training require officers to make judgments on the use of varying degrees of force [7] and present officers with situations that reflect real-life work conditions [8]. Knowing the prevalence of SWAT team call outs and the unique facts surrounding them can be the basis for making informed training decisions. Use of expert witnesses on the proper application of guidelines enacted by a department often provides the crucial factual information to courts when litigation ensues.

Reasonable and Excessive Force

Free citizens raise a Fourth Amendment claim when they allege that a SWAT team used excessive force against them in a law enforcement capacity (*see also* **Police Use of Force**). The Fourth Amendment prohibits unreasonable searches and seizures. It protects

free citizens from the police using excessive force against them. There is no prevailing definition of what is excessive force. The Supreme Court, however, imposed an "objective reasonableness" standard for reviewing claims of excessive force by the police [9]. In such cases and under the *Graham* rules [9], the trier (or judge) or triers (or jury) of facts must first weigh the rights of free citizens under the Fourth Amendment against the interests of the government to take action against them (or balancing test). The Fourth Amendment recognizes that the right to take action against free citizens is associated with the need to use some degree of physical coercion or threat [10]. Second, consider whether a well-trained officer under the same set of circumstances would observe and conclude that the actions by the police were reasonable (or objective test).

Third, discount the subjective motivations of the police because they have no value in a court's appraisal of excessive force (or subjective test). Evil intentions do not raise Fourth Amendment claims when force options were objectively reasonable. Good intentions do not make objectively unreasonable force options constitutional. Fourth, evaluate the circumstances of the police call out. That is, what is the severity of the crime, does the suspect pose an immediate threat to the police or others nearby, and is the person actively resisting or fleeing arrest (or totality of circumstances test). Because the "objective reasonableness" standard is not capable of providing an exact definition, its application requires attention to the totality of circumstances. Only circumstances known to the police before and at the time they use force are relevant [11]. The central question is whether the totality of circumstances justifies a particular use of force by the police.

Fifth, recognize that not every push or shove by the police even if it later seems unnecessary violates the Fourth Amendment (or hindsight test). The standard of "objective reasonableness" makes allowances for errors made by the police who must at times make split-second decisions about force in situations that are tense and uncertain, and that evolve rapidly [12].

Finally, examine carefully at whether the severity of force by the police puts suspects at risk of death "or" serious bodily harm (or deadly force test). Under the *Garner* rules [13], police may deploy weapons or use tactics that amount to deadly force in three conditions. First, if deadly force is necessary to prevent escape (or necessary condition). Second, if the police have probable cause to believe that a suspect poses a significant threat of death "or" serious bodily harm to them "or" others nearby (or dangerous condition). In this condition, the court suggests that suspects who threaten the police or others with weapons and suspects who commit a crime and cause or threaten to cause serious bodily harm are dangerous because of their violent or potential violent behavior. Third, if possible, the police must warn a suspect of their intention to use deadly force against him or her (or warning condition).

Whether SWAT teams use deadly or nondeadly force options against suspects of crimes, allegations of excessive force are possible especially when tactical operations involve the use of SWAT not employed by regular officers. For example, plaintiffs have argued that the uses of certain police tactics such as deploying diversionary devices are excessive. In numerous court cases, however, the courts have suggested that uses of diversionary devices are permissible under the *Graham* rules [9] when the totality of circumstances support their use, when the police use discretion and do not deploy them routinely as a matter of custom, and when the police receive training in their appropriate use [14–16].

A high-risk forced building entry by a SWAT team also raises concern of excessive force. In general, police need consent, exigent circumstances, or an arrest or search warrant to enter buildings. Carrying out a forced building entry will trigger a judicial review of its lawfulness and excessiveness especially when any injuries occur. The more on-scene time a SWAT team has with a police call out such as negotiating with a barricaded suspect, the less the team can rely on exigency to carry out a warrantless entry, search, or seizure. A SWAT team avoids some excessive force liability for using forced building entry tactics when it obtains prior judicial approval.

Questions about excessive force may occur when police regularly execute as a matter of policy or custom warrantless building entries in critical incident situations regardless of the totality of circumstances. For example, in *O'Brien v. City of Grand Rapids* [17], the court held the city liable for a warrantless entry into the home of an armed barricaded suspect with whom the police had contact time for 6-hours. The police here had a custom of executing warrantless building entries in critical incidents regardless of the circumstances.

Sometimes, when police fail to reevaluate prior to entry the circumstances justifying a no-knock warrant, it may be unreasonable or excessive to carry out the warrant. For example, in *United States v. Singer* [18], the court said that if during the period between obtaining a no-knock warrant and executing it the police receive reliable information that exigent or dangerous circumstances no longer exist, the police must reevaluate their plan of entering a home without first knocking and announcing.

In cases of alleged use of excessive force, criminal and civil liability can attach itself in the form of a team's wrongful actions under Title 18 of the US Code section 242 and Title 42 section 1983 respectively. Commonly referred to as *sections 242* and *1983*, police shall not deprive any citizen within their jurisdiction of any rights secured by the Constitution and laws while acting under the color of law. In an action of law (criminal action for deprivation of rights) or redress (civil action for deprivation of rights), police are liable to any persons injured under sections 242 and 1983. Civil liability can also come in the form of specific state statutes that are similar to those of the federal government. For those members of SWAT teams, call outs are at times fraught with certain legal – criminal or civil – problems, dangers, or difficulties.

Hostage and Bystander Injuries

Police have the responsibility to safeguard the well-being of the community. There is the possibility of injuries, however, when tactics to handle dangerous police call outs involve the use of physical force to capture and arrest law violators. Generally, no deprivation of rights occurs when the police use reasonable care, but accidentally shoot hostages or bystanders during police–citizen encounters. For example, in *Green v. Denison* [19], a suspect fired a bullet that shattered glass, which blinded a bystander. An officer fired a bullet that accidentally hit the suspect's girlfriend. Both injured persons sued and claimed that the police had a duty to protect them from harm. The court dismissed the claims and held that police officers are not liable for injuries or damages to the general public that arise from their acting within the scope of their authority or duty. It distinguished duty owed to the public from duty owed to particular persons targeted by the police. Protecting the police or giving them official immunity from personal liability arising from making discretionary or best judgment decisions allows them to perform their job without distractions.

In *Lee v. Williams* [20], a police deputy unintentionally shot a hostage during a shootout with armed suspects. Postincident litigation brought about a § 1983 civil deprivation of rights claim against the deputy. The court ruled that there was no Fourth Amendment seizure of the hostage because the deputy did not intend to shoot him, but did intend to shoot the armed suspects.

For officers involved in making tactical decisions about handling hostage situations, the courts have established a philosophy of "human life is the main concern." For example, in *Downs v. United States* [21], a Federal Bureau of Investigation (FBI) SWAT team carried out a forceful assault against armed suspects who hijacked an airplane, held passengers as hostages, but released some of them during negotiations. The outcome of the assault approach was the deaths of hostages. The court ruled that the safety of hostages is more important than the arrest of suspects.

There is no constitutional obligation to have specially trained SWAT teams deal with nonnormal police calls for service such as hostage situations [22]. Actions by regular or non-SWAT officers, who generally do not possess the same training, experience, and equipment that SWAT officers posses to handle the most dangerous police call outs, might not automatically result in injuries or police behaviors that would shock the conscious of courts. For SWAT and non-SWAT officers involved in making tough tactical decisions, (*see* **Aggression**) postincident outcomes might involve criminal or civil action for deprivation of rights. Fourth Amendment principles (*see* **Seizures: Behavioral**) are equally applicable (*see* **Threat Assessment: School**) to both SWAT and non-SWAT officers who might injure citizens.

References

[1] Clark, J.G., Jackson, M.S., Schaefer, P.M. & Sharpe, E.G. (2000). Training SWAT teams: implications for improving tactical units, *Journal of Criminal Justice* **28**, 407–413.

[2] Kraska, P.B. & Cubellis, L.J. (1997). Militarizing mayberry and beyond: making sense of American paramilitary policing, *Justice Quarterly* **14**, 607–629.

[3] Kraska, P.B. & Kappeler, V.E. (1997). Militarizing American police: the rise and normalization of paramilitary units, *Social Problems* **44**, 1–18.
[4] Williams, J.J. & Westall, D. (2003). SWAT and non-SWAT police officers and the use of force, *Journal of Criminal Justice* **31**, 469–474.
[5] Canton v. Harris, 489 U.S. 378 (1989).
[6] Walker v. City of New York, 974 F.2d 293 (2nd Cir., 1992).
[7] Allen v. Muskogee, 119 F.3d 837 (10th Cir., 1998).
[8] Popow v. Margate, 476 F.Supp. 1237 (D.N.J. 1979).
[9] Graham v. Connor, 490 U.S. 386 (1989).
[10] Terry v. Ohio, 392 U.S. 1 (1968).
[11] Ford v. Childers, 855 F.2d 1271, 1276 (7th Cir., 1988).
[12] Johnson v. Glick, 481 F.2d 1028, 1033 (2nd Cir., 1973).
[13] Tennessee vs. Garner, 471 U.S. 1 (1985).
[14] Commonwealth v. Garner, 423 Mass. 735, 772 N.E.2d 510 (1996).
[15] Langford v. Gates, 43 Cal. 3d 21, 729 P.2d 822 (1987).
[16] United States v. Myers, No. 94-20013-01, 1194 WL 324582 (10th Cir., 1997).
[17] O'Brien v. City of Grand Rapids, 23 F.3d 990, 999 (6th Cir., 1994).
[18] United States v. Singer, 943 F.2d 758, 763 (7th Cir., 1991).
[19] Green v. Denison, 738 S.W.2d 861 (Mo. 1987).
[20] Lee v. Williams, 138 F.Supp.2d 748 (E.D. Va. 2001).
[21] Downs v. United States, 522 F.2d 990 (6th Cir., 1975).
[22] Salas v. Carpenter, 980 F.2d 299, 309–10 (5th Cir., 1992).

Related Articles

Daubert v. Merrell Dow Pharmaceuticals

Police Use of Force

FRANK J. GALLO

Pollen *see* Palynology

Polymorphism: Genetic

Natural genetic variation, within a population, is seen by the existence of more than one type of allele for a locus. Loci with several commonly occurring alleles are useful for profiling purposes because, although some individuals may share the same at alleles at a locus, the probability of them sharing the same alleles at all loci reduces dramatically as the number of loci considered increases. For profiling purposes, scientists use loci that comprise sequences with simple tandem repeats (STR) (*see also* **Short Tandem Repeats**). These are regions of DNA that have short sequences of DNA that are repeated as multiple blocks, like carriages of a train. Any one person will have two copies (alleles) of the locus, one from mother and one from father, and each allele will have its own number of STRs. These may be the same (homozygote) or different (heterozygote).

DNA profiling tests the polymorphism, or variation, of the alleles at a set of loci for a sample. For the sake of convenience, the different alleles of any one locus are given a name that is the same as the number of STR contained within the allele's DNA sequence. Thus, a person with profile D3/13,16 is a heterozygote with one allele at D3 that has 13 repeats and a second allele that has 16 repeats.

Polymorphisms in a population have evolved as a result of mutations and the passage of time. In the case of STR, mutations are thought to be due to a low rate of slippage of DNA synthesis of the repeated sequences during the process of making sperm and egg cells. Such mutations are then passed on to a new generation and over time become a measurable proportion of a population. A rule of thumb for a polymorphism is that it occurs in at least 0.05% of a population. Exactly, how common any allele is depends on many factors, including chance, selective breeding, population size, intermingling of populations, and so on. The allele frequencies of a polymorphic locus will remain stable if the population is in Hardy–Weinberg equilibrium (*see also* **Hardy-Weinberg Equilibrium**), but are often different for different populations.

SCOTT BADER

Popper Theory of Falsifiability *see* Falsifiability Theory

Postmortem *see* Autopsy

Postal: Going *see* Homicide: Multiple (Behavior)

Postmortem Biochemical Examinations

Introduction

Functional Causes of Death

The diagnosis of *functional causes of death* is, on one hand, based on mostly sparse postmortem findings and, on the other hand, considerably on postmortem biochemical alterations, which frequently originate from illnesses with internal causes and subsequent dysregulations such as diabetes mellitus, alterations of kidney and liver function, and imbalances of water and electrolytes. It is not rare that combinations of such dysregulations with problematic overlappings are seen due to close physiological and biochemical links [1–4] (*see also* **Cardiac and Natural Causes of Sudden Death**; **Natural Causes of Sudden Death: Noncardiac**).

Postmortem Biochemical Estimations and Differences to Clinical Biochemistry

Postmortem biochemical analyses may represent the main clue to the diagnosis of functional causes of death. One of the main problems is to be able to apply clinical biochemical values on postmortem conditions. On one hand, there are big unpredictabilities regarding general postmortem changes in body fluids. On the other hand, biochemical values in postmortem specimen may well represent more or less the results of changes taking place during agony or the early postmortem period. Contrary to clinical biochemical estimations, values obtained postmortem do not necessarily allow conclusions regarding the mechanism of death. Postmortem diagnostic procedures, therefore, require a critical way of looking at them [5].

Obtaining the Appropriate Specimen

Body fluids are usually obtained during postmortem examination. In cases with a limited external examination specimen (e.g., cerebrospinal fluid (CSF), vitreous humor, blood, and urine) can also be taken by cannulation (suboccipital access, puncture of an eyeball, dissection of a femoral vein, and puncture of the urinary bladder). The cranial cavity and the eyeballs provide a relatively good protection of the enclosed body fluids against decomposition effects. After obtaining vitreous humor, the eyeballs should be refilled with water due to cosmetic reasons.

The volume of CSF to be found varies from 50 ml (baby) to 135 ml (adult). A few milliliters are sufficient for the postmortem biochemical analyses and there is usually no problem to get blood-free CSF. Approximately 1–2 ml of vitreous humor can be obtained by the puncture of both eyeballs. Aspiration of small parts of the retina is of no further relevance. Postmortem blood should be taken from the heart and a (peripheral) femoral vein and urine from the bladder immediately after dissection (a few milliliters per specimen) [6].

"Near-Table" Methods

During the postmortem examination, several screening tests with stripes and tablets can be carried out regarding glucose, bilirubin, or acetone. Furthermore, there are a number of electronic test devices on the market, which can be used for screening purposes. These "near-table" methods are useful to confirm or exclude certain differential diagnoses at the time of the autopsy (further information: http://www.roche-diagnostics.com).

Glucose Metabolism and Diabetes Mellitus

General Aspects of Diabetic Coma

Diabetic coma is a life-threatening complication of diabetes mellitus. Owing to a relative or an absolute insulin deficit, there is a typical rise of blood sugar with the possibility of acute complications or damage to blood vessels and nerves after longer duration. Depending on the age group, the incidence of diabetes mellitus varies between 2 and 5%. Causes for coma may be the onset of an unknown diabetes, lack of insulin injections, or increased requirement of

Table 1 Postmortem biochemical values in case of alterations of glucose metabolism

Dysfunction	Parameter	Compartment	Results
Coma (in general)	Sum value[a]	Cerebrospinal fluid	$\sum > 415\,\text{mg dl}^{-1}$[b]
		Vitreous humor	$\sum > 410\,\text{mg dl}^{-1}$[c]
	HbA_{1c}	Blood	>12.1%[d]
	Glucose	Urine	$>25\,\text{mg dl}^{-1}$[e]
Ketotic coma	Acetone	Blood, cerebrospinal fluid	$>21\,\text{mg dl}^{-1}$[f]
		Vitreous humor	$>5\,\text{mg l}^{-1}$[f]
		Urine	
Hypoglycemia	Sum value	Cerebrospinal fluid	$\sum < 50–80\,\text{mg dl}^{-1}$
		Vitreous humor	$\sum < 100–160\,\text{mg dl}^{-1}$

[a] According to Traub: concentrations of glucose and lactate
[b] Mean value = $500–600\,\text{mg dl}^{-1}$
[c] Mean value glucose = $300–950\,\text{mg dl}^{-1}$; mean value lactate initially = $80–160\,\text{mg dl}^{-1}$, after 20 h = $210–260\,\text{mg dl}^{-1}$
[d] Mean value = 13–15%, nondiabetics = 9.15%
[e] Most of coma cases $>50\,\text{mg dl}^{-1}$, partly $2000–4000\,\text{mg dl}^{-1}$
[f] Coma: mean value = $100–150\,\text{mg l}^{-1}$

insulin due to acute infections, poor diet, operations, gastrointestinal diseases, or myocardial infarction. Twenty-five percent of all diabetic comas are so-called manifestation comas with previous unknown diabetes. Infections are the most frequent triggers for coma onset (approximately 40% of the cases). The frequency of fatal coma among known diabetics is between 0.5 and 1.5% with a peak in the age group of 40–60 years. The overall lethality from coma varies from 5 to 25% and rises to 70% with coma of longer duration. Lethality of diabetic coma is tenfold in 70-year-old individuals compared to 30-year-old patients. Furthermore, the risk for coma in juveniles is four- to sevenfold higher than in adults [1, 2] (see Table 1).

Types of Diabetic Coma

Typically, diabetes mellitus type I is associated with ketonemic coma, whereas hyperosmolar coma normally results from type II diabetes mellitus. A lack of insulin causes a rise of blood glucose with subsequent loss of fluids and electrolytes. In addition, increased lipolysis is used to compensate the deficit of energy resulting from the inhibition of glucose metabolism leading to increased levels of ketone bodies with metabolic acidosis. The latter may be excessive ($500–1000\,\text{mg l}^{-1}$ acetone or higher), whereas hyperglycemia remains mostly moderate ($250–600\,\text{mg dl}^{-1}$). Hyperosmolar coma is more rare (approximately 10–20% of the cases) and associated with relative lack of insulin causing reduced peripheral utilization of glucose with simultaneous release of glucose from the liver. Low levels of insulin prevent ketosis due to inhibition of lipolysis. Therefore, it is typical to find excessive hyperglycemia (often above $1000\,\text{mg dl}^{-1}$) with lacking or only mild ketosis.

Diabetic coma may lead to fatal outcome via different pathophysiological pathways. There is a cardiovascular type with leading oliguria or a renal type with acute kidney failure. Moreover, there exists pseudoperitonitis type with the symptoms of an acute abdomen. Typical accompanying diseases of fatal diabetic decompensation may be myocardial infarctions, apoplexy, embolism, pneumonia, pancreatitis, pyelonephritis, and a predisposition for lactic acidosis [5].

The most important body fluids for postmortem diagnostic purposes are CSF and vitreous humor using the so-called sum value according to Traub, which provides a combined calculation to compensate postmortem alterations of blood glucose level due to glycolysis, accordingly [7].

Glucose

The hourly metabolic decrease of glucose in CSF is about $10–15\,\text{mg dl}^{-1}$ but may vary between approximately 5 and $45\,\text{mg dl}^{-1}$. The hourly rate is below $1\,\text{mg dl}^{-1}$ 100 h postmortem. Given normal metabolic conditions, therefore, zero levels are reached after 10–12 h. Longer persisting glucose levels are indicative of antemortem hyperglycemia [8].

The speed of postmortem glycolysis depends on a number of factors, e.g., the temperature and duration of body storage. Postmortem glycolysis is slower in diabetics compared to nondiabetic individuals, whereas obesity accelerates degradation of glucose. Isolated assessment of elevated glucose levels in CSF requires critical reserve (normal range: *ca.* 50–90 mg dl^{-1}), because multiple other dysregulations may be accompanied by the same symptom as carbon monoxide poisoning, acute cardiac death, brain trauma, strangulation, protracted agony, asphyxia, pneumonia, and pancreatitis. This aspect has also to be taken into account regarding other body fluids [6].

Lactic Acid (Lactate)

The product of postmortem glycolysis is lactate (normal level in CSF *ca.* 9 mg dl^{-1}). Its concentration increases postmortem with a rate of approximately 10–15 mg dl^{-1} up to the 10th hour after death. After this time, the increasing rates vary considerably. Under differential diagnostic aspects also other disorders may cause hyperlactacidemia, e.g., tumors, respiratory insufficiency, severe chronic inflammations, uremia, especially inflammations of the central nervous system or alcohol-induced with lack of thiamine, physical strain, and alimentary factors (e.g., strict fasting).

Sum Value

This combined method, according to Traub, compensates arithmetically for the postmortem production of lactate from glucose by using a "sum value". It is based on the fact that 1 mol of glucose produces, via glycolysis, two moles of lactate so that the concentrations can be added using milligrams per deciliter. If the "sum value" exceeds 362 mg dl^{-1} in CSF, the probability of fatal diabetic coma is about 89%, if other, e.g., toxicological and morphological, alterations can be excluded. In cases of diabetes mellitus, the "sum value" remains almost stable up to the 200th hour postmortem. If there are nondiabetic causes of death, the "sum value" increases up to the 30th hour postmortem, but remains nearly stable afterwards. Although the formula, according to Traub, has to be used under critical view, the "sum value" may be considered the most important criterion for the diagnosis of fatal diabetic coma. However, the author's research has revealed that it is more realistic to increase the limit "sum value" in CSF to 415 mg dl^{-1} (upper limit of the 95% confidence interval in cases of cardiac death), with cases of diabetic coma ranging on average between *ca.* 500 and 600 mg dl^{-1} [9, 10].

Vitreous Humor

The calculation method according to Traub may also be applied on vitreous humor. The glucose level herein is about 50–85% of the serum glucose. Values for postmortem glucose concentrations vary from 20 mg dl^{-1} (nondiabetics) to 90 mg dl^{-1} (known diabetics), but wide variation ranges have to be taken into account. Owing to slower glycolysis in vitreous humor compared to CSF, normal glucose values may be found as long as two days postmortem. In cases of fatal coma, glucose values between *ca.* 300 and 950 mg dl^{-1} may be found. Lactate values are already around 80–160 mg dl^{-1} in the intramortal period and between 210 and 260 mg dl^{-1} approximately 20 h postmortem. The upper limit value is 410 mg dl^{-1} and if this is exceeded, it can be taken as a strong indication of fatal diabetic coma, given the condition that other possibly competing mechanisms can be excluded. The procedure is said to be applicable until the 10th postmortem day [11–15].

Blood Glucose

Blood sugar levels alone are only of low diagnostic relevance, if at all limited to blood from the femoral veins within the 1st and 2nd hour postmortem in which the level is *ca* 40–100 mg dl^{-1}. Contrary to this, glucose levels in central blood (right ventricle) may easily reach 1000 mg dl^{-1} and over due to postmortem hepatic glycogenolysis. Normally, postmortem glycolysis (approximately 13 mg dl^{-1} h^{-1}) results in complete metabolization of the blood glucose within 6–8 h. This leads to a corresponding increase of lactate up to 180 mg dl^{-1} after 1 h and *ca.* 450–680 mg dl^{-1} after 12–24 h. Especially due to postmortem diffusion of serum and its components from surrounding tissues into blood vessels, the "sum value" cannot be used [10].

Hemoglobin A_{1c}

This glycosylated fraction of hemoglobin represents an important parameter regarding a basic diagnosis of

diabetes mellitus. Owing to the fact that kinetics of its formation is depending on time and concentrations, HbA_{1c} can be used as a long-term indicator of diabetic conditions (so-called blood sugar memory for *ca.* 120 days). Levels of 6–8% (maximum of 10%) are consistent with a normal glucose metabolism, whereas higher concentrations are indicative of inappropriate metabolic conditions (hyperglycemias in the past). Periods of increased blood sugar have to last 6–8 h minimum to cause significant rises of HbA_{1c} due to its slow reaction kinetics. Furthermore, the prefinal and postmortem drop of the pH value in blood, due to formation of lactate, are likely to result in a reduction of HbA_{1c} because of separation of its unstable component. Blood sugar also decreases rapidly after death. The stable part of hemoglobin A_{1c} makes up approximately 90% of the whole. For example, hyperglycemia around 360 mg dl^{-1} takes around 12 h to cause an increase of HbA_{1c} of 1.3% absolute. In reverse, a reduction of around 5% needs around seven days.

There has been found a positive connection between "sum value", urine glucose concentration, and HbA_{1c} level. This means that there usually is a coincidence of elevated "sum value", high urine glucose, and HbA_{1c}. Hemoglobin A_{1c} has proven to be relatively stable versus autolysis especially in hemolyzed blood and can be measured postmortem in frozen samples and also in samples stored in a normal fridge. It has been revealed that storage at temperatures between +4 and −80 °C does not cause any relevant changes to the HbA_{1c} concentrations. The result is independent from the actual total hemoglobin level because HbA_{1c} is measured as percentage of the current hemoglobin value [16].

Falsely elevated hemoglobin A_{1c} concentrations can be found due to increased HbF levels in cases of thalassemia or advanced renal failure. In principle, HbA_{1c} has proven to be a reliable parameter for the basic diagnosis of diabetes mellitus without being too liable for interferences. It is also possible to measure other glycosylated proteins such as fructosamine, but assessment is rather difficult. The mean levels of HbA_{1c} in cases of diabetes mellitus differ considerably from those in nondiabetic individuals and are around 12.1% in diabetic coma (range: ~13–15%). However, the lower portion of the range in case of diabetes mellitus may overlap with the upper portions of the range in nondiabetic cases as it has been shown for the "sum value" [17].

Ketone Bodies

The ketotic type of diabetic coma is characterized by an increased level of ketone bodies in blood and other body fluids (acetone and acetylacetate *ca.* 25–35%, β-hydroxyl-butyrate *ca.* 65–75%; normal values for acetylacetate 0.8–2.4 mg l^{-1}, for β-hydroxyl-butyrate 2.5–9.8 mg l^{-1}). Estimation of acetone may easily be carried out in connection with blood alcohol analysis using headspace chromatography. The normal concentrations for free acetone range from 2.3 to 2.5 mg l^{-1} in nondiabetic patients and may reach 23 mg l^{-1} in diabetics. The levels are almost independent from the postmortem interval [2, 5].

The level of acetone in CSF with diabetes mellitus differs considerably from those seen with nondiabetic causes of death, especially in cases of diabetic coma, with an obvious association regarding an elevated "sum value". If other causes can be ruled out, acetone levels exceeding 5 mg l^{-1} are suspicious of diabetes mellitus. Ketotic coma may be associated with levels higher than 100 mg l^{-1}, but ketonemia is rarely seen if the blood glucose concentration is only 200 mg dl^{-1} and below. According to the author's research, acetone levels in ketotic coma exceed 21 mg l^{-1} in most of the cases, with mean values in this group of 100–150 mg l^{-1}. Single cases may show levels of more than 1000 mg l^{-1}. Nondiabetic factors that might cause elevated ketone levels are, e.g., chronic hepatic and renal diseases, pancreatitis, shock, chronic alcoholism and isopropanol poisoning (levels up to 160 mg l^{-1}) as well as protracted fasting (acetone levels may exceed 5000 mg l^{-1}) [18].

Urine

As the fourth column of postmortem diabetes mellitus diagnostics, an examination of urine can reveal important clues. Urine glucose levels higher than 25 mg dl^{-1} (maximum in healthy individuals) may be indicative of diabetes. Diabetic coma is sometimes associated with urine glucose concentrations above a few 1000 mg dl^{-1}, but usually higher than 500 mg dl^{-1}. These excessively high values only show very small overlapping with other cause of death groups, although positive findings for glucose in urine alone are only of lower value. Glucosuria is a rather frequent nonspecific symptom, e.g., due to brain trauma, myocardial infarction, intoxication, apoplexia, and leukemia. Likewise, glucosuria may

be absent even in cases of manifest diabetes mellitus caused by diabetic glomerulosclerosis itself or postmortem degradation. Ketone bodies are likely to be found in urine longer than 24 h postmortem. Concentrations exceeding $0.5\,mg\,dl^{-1} = 5\,mg\,l^{-1}$ of free acetone may be indicative of ketotic dysregulation. However, a positive test for ketonuria is not a proof for ketonemia, because the kidneys have a relatively high clearance rate for ketone bodies. Furthermore, there are multiple conditions that might cause considerable ketonemia (see above). Hyperosmolar coma is typically characterized by a lack of ketonemia (approximately 30% of diabetic comas) [1, 2, 5].

Lactic Acidosis

There are some secondary effects of lactic acidosis that might gain special forensic medical relevance. For example, moving potassium to the extracellular space may cause hyperkalemia (see below). Acidosis decreases the reactivity versus catecholamines with a negative-inotrope effect on the heart. Severe acidosis may result in massive reduction of the kidney blood circulation leading to acute renal failure. Diabetic coma can also cause acidosis by production of β-hydroxyl-butyrate and acetylacetate (see above). Lactic acidosis plays an important role, particularly regarding overlapping with postmortem diagnosis of diabetes mellitus. Considerable amounts of lactic acid are being released during shock and hypoxia, due to poor perfusion caused by diabetes mellitus, following renal failure, hepatic diseases, and ethanol/methanol intake, rarely as complication of a treatment with biguanides or due to severe lack of thiamine with chronic increased alcohol intake. The conditions can be exacerbated by chronic renal failure due to reduced excretion of acids and also by an increased loss of bicarbonate resulting from diarrhea and/or vomiting.

The central causal mechanism is an increased concentration of pyruvate from protein catabolism together with a lack of oxygen, so that energy can still be provided by glycolysis. Accumulation of lactate happens more frequently in diabetics than in other patients what is due to disturbances of oxygen supply and alterations of metabolic activities. The clinical picture is characterized by gastrointestinal discomfort, muscular spasms, central nervous disturbances and deep frequent respiration. The severe type of biguanide-induced lactic acidosis shows a lethality rate of over 50%.

Patients suffering from chronic alcoholism represent a special risk group regarding fatal lactic acidosis and ketotic coma as well. There are often only very few and/or nonspecific morphological findings. On one hand, considerable ketonemia may follow acute alcoholization (free acetone from 74 to $400\,mg\,l^{-1}$), but, on the other hand, high "sum values" may also result in this condition. Their range (*ca.* $294–594\,mg\,dl^{-1}$) can also be associated with fatal diabetic coma. Given the precondition that diabetes mellitus and other competing mechanisms can be ruled out, ketotic coma or lactic acidosis has to be considered as a cause of death in such cases. The lower limiting values for the "sum value" are *ca.* $300–400\,mg\,dl^{-1}$, for acetone in blood around $90\,mg\,l^{-1}$ and 6% for HbA_{1c} [19, 20].

Hypoglycemia (Endogenous vs. Exogenous Hyperinsulinism)

Although fatal hypoglycemia appears to be a rather rare event among forensically examined death cases, they might be the source of serious diagnostic problems. Under clinical conditions, hypoglycemia is diagnosed if the blood glucose level lies below $40\,mg\,dl^{-1}$ *or* if the so-called Whipple's triad can be found. It comprises blood glucose level below $45\,mg\,dl^{-1}$, symptoms of hypoglycemia, which disappear under administration of glucose. Multiple circumstances may be responsible for hypoglycemia in individuals with an empty stomach, e.g., insulinomas and other tumors, severe hepatic diseases, uremia and glycogenoses. The initial manifestation of diabetes mellitus may also be accompanied by reactive hypoglycemia as well as alterations of gastric mobility, vegetative instability, or massive alcohol intake with simultaneous lack of food due to inhibition of gluconeogenesis. The autonomous or glucopenia-associated spectrum of symptoms includes hyperorexia, nausea, restlessness, sweating, tachycardia, endocrine neuropsychologic disorder, primitive automatisms, risk of convulsions, and focal signs with apoplectiform symptoms. The final state with somnolence, coma, and central alterations of respiration and circulation until death has forensic medical relevance.

Hypoglycemias due to exogenous causes are mostly seen with an existing diabetes mellitus. Important mechanisms are accidental or intentional overdosage of insulin or sulfonyl-urea derivates with

subsequent reactive hypoglycemia. Such a situation may arise from lack of regular alimentation due to intercurrent diseases without changing the doses of antidiabetic drugs. Other possibilities for hypoglycemias can be interferences with drugs which decrease the blood sugar level indirectly or unusual physical strains. However, types of hypoglycemia with a forensic medical impact are those caused by overdosages of antidiabetics.

The so-called factitious hypoglycemia needs special attention. It is caused by (unnecessary) administration of insulin or sulfonyl-urea derivates and can be seen in connection with psychic alterations (e.g., borderline personality disorder) or suicidal intention. It is rare to find a primary criminal background, e.g., cases of homicide. The most important diagnostic criterion of this type of hypoglycemia is that it happens independently from alimentation. Affected persons often have relations to professional health care or are relatives of known diabetics [21].

The calculation procedure regarding a "sum value" can also be used for the diagnosis of hypoglycemia. Consequently, low sum values in CSF in vitreous humor below *ca.* $50-80\,\mathrm{mg\,dl^{-1}}$ or rather $100-160\,\mathrm{mg\,dl^{-1}}$ are strongly indicative of fatal hypoglycemia. This conclusion is particularly supported by simultaneously high insulin levels suggesting that estimation of insulin levels and also of c-peptide postmortem is essential. In case of endogenous secretion, insulin and c-peptide are both found elevated. If there is exogenous hypoglycemia due to administration of insulin, the level of c-peptide will be noted as much lower than normal. Contrary to this, there are usually increases of insulin and c-peptide concentrations following an intake of sulfonyl-urea derivates, but, in diabetic individuals, often rather high insulin levels can be seen without any indication of hypoglycemia. The procedure has also proven to be reliable in cases of suspected hypoglycemia in car drivers [22, 23].

Postmortem estimations of insulin levels can be carried out by radioimmune assay (RIA) and have revealed levels very similar to those of healthy individuals in blood from a femoral vein and also from the heart. Nevertheless, postmortem concentrations of insulin in blood from the right ventricle may be increased to about 10-fold of normal values due to release of insulin after death. Putrefaction may cause problems as well. Furthermore, single estimations have a wide variation and, therefore, cannot be used as the only criterion for the diagnosis of insulin-based hypoglycemia. Sometimes it is possible and useful to have a proof of suicidal insulin injection by analyses of the tissues close to the injection site. It is a strict rule that the postmortem diagnosis of hypoglycemias must be based on a combined assessment of different criteria and can only be made "per exclusionem". According to this, especially cardiac diseases, cerebral hemorrhages, pulmonary embolism, strangulation/asphyxia, ruptures of vessels and intoxications have to be ruled out. Estimations of insulin should always be carried out in peripheral venous blood or CSF/vitreous humor because diffusion of insulin from the pancreas via the portal vein might take place postmortem. The "sum value" calculated from glucose and lactate levels is of special importance (see above) [24, 25].

Alterations of Liver Function

In case of an advanced stage of *hepatic cirrhosis* from different causes, it is not rare that there develops an alteration of liver metabolism, often resulting in potentially reversible complications, due to retention of neurotoxic substances in blood with decompensation and final *hepatic failure*. Suspicion may arise from the previous medical history, desolate housing conditions, known alcohol abuse, and sometimes the presence of jaundice. Acute deterioration of hepatic insufficiency with a danger of hepatic coma originates from an increased production of ammonia due to a high portion of proteins in the intestinal contents that may be caused by gastrointestinal hemorrhages (especially esophageal varicosis due to alcoholism), protein-rich nutrition, febrile infections with increased protein catabolism and drugs (e.g., benzodiazepines, analgesics). Clinically, the advanced stage is characterized by permanent drowsiness but patients can be woken up, later on hepatic smell, and electroencephalogram (EEG) alterations. This picture leads to coma with unmistakable "foetor hepaticus" and massive EEG changes until fatal outcome with total hepatic failure [1, 2].

The terms *acute hepatic insufficiency* or *endogenous hepatic coma* describe a failure of the liver function *without previously existing chronic liver disease.* Contrary to the chronic hepatic failure, decompensation can occur suddenly without any indications from the medical history. Important morphological findings are dermal and scleral jaundice

and, clinically, disturbances of blood coagulation and consciousness (somnolence, coma). It is especially the fulminant type with a duration of less than seven days, which may gain forensic medical relevance. Important causes are viral hepatitis (65%) and hepatotoxic substances (30%) such as medication (Paracetamol), drugs, chemicals (CCl_4), or poisons from mushrooms (Amanita phalloides). This elucidates the importance of accompanying *toxicological analyses.* Potentially fatal complications may be brain edema (80%; most frequent cause of death), gastrointestinal hemorrhages (50%), as well as hypoglycemia and renal failure with electrolyte imbalances.

The typical enzymes of liver metabolism represent important parameters, which can also be examined postmortem, as well as bilirubin. The daily bilirubin production comes to approximately 510 $\mu mol\,l^{-1}$ (30 $mg\,dl^{-1}$; normal value up to 1.1 $mg\,dl^{-1}$). Hepatic failure is typically associated with an increased level of serum bilirubin causing jaundice if it exceeds 34 $\mu mol\,l^{-1}$ (2 $mg\,dl^{-1}$). A differentiation between direct bilirubin bound to biglucuronide and nondirect bilirubin bound to albumin is only useful under clinical aspects. *Postmortem bilirubin levels* may well be compared with those obtained antemortem. Differences are only ranging in the area of 0.1 $mg\,dl^{-1}$, especially in death cases showing jaundice. During the postmortem period, there can be seen a slight but steady increase (*ca.* 0.2 $mg\,dl^{-1}$ after 2 h and 0.7 $mg\,dl^{-1}$ after 20 h). Furthermore, there is an increase of *enzymes typical for the liver* (*GPT glutamate pyruvate transaminase, GGT gamma glutamyl transferase, and AP alkaline phosphatase)* as well as of *ammonia* (>100 $mg\,dl^{-1}$; normal value below 0.05 $mg\,dl^{-1}$) primarily not only in blood but also in other body fluids (CSF, vitreous humor). However, clinical reference ranges of values can only be used as a basis for assessment. Most of the bilirubin in CSF belongs to the conjugated type, often associated to hypokalemia and hypoglycemia [6].

Disturbances of Kidney Function

Chronic renal failure represents the result of a nonreversible reduction of the function of both kidneys. Important causes are, e.g., diabetes mellitus (nephropathy, *ca.* 35%), hypertension (*ca.* 25%), chronic inflammations (*ca.* 15%) and abuse of analgetics (*ca.* 1%). The chronic reduction of the renal function can also show acute decompensation leading to unexpected sudden death, which is not an unusual development during diabetic coma. The *compensated chronic phase* showing only a functional reduction of a low degree and the phase of *compensated retention* (azotemia, creatinine levels up to 6 $mg\,dl^{-1}$) are not associated with symptoms of uremia. Preterminal renal failure with creatinine levels above 8 $mg\,dl^{-1}$ plus symptoms of uremia is called *decompensated retention. Terminal renal failure* (uremia) showing creatinine levels over 10 $mg\,dl^{-1}$ is associated with massive symptomatology of uremia. During the phase of *decompensated retention* (preterminal phase), there may be seen edematous changes, cardiac failure, gastroenteritis due to uremia and neuropathy. The *terminal phase* is characterized by acute life-threatening symptoms, such as neuropathy and encephalopathy, overhydration with pulmonary edema, bleeding tendency, coma, and death (see Table 2).

Acute renal failure or acute renal insufficiency represents a mostly *reversible reduction of the renal function* with loss of urine production and increasing retention parameters (urea, creatinine). Fifteen percent of the cases with acute renal failure show polyuria or normuria with an increase of retention values being the only symptom. Without sufficient therapy, e.g., dialysis, acute renal failure mostly has a fatal outcome. Sometimes bilateral necroses of the renal cortex can be seen. There are multiple possible causes for *acute renal failure,* such as alterations of the blood circulation, toxins, medication (antirheumatics, cytostatics, and antibiotics), chemicals (glycols), and inflammatory or vascular processes.

The most critical clinical phase is the third one with polyuria and extensive loss of water/electrolytes and simultaneous increase of urea and creatinine. Fatal complications may occur associated with other organs, e.g., shock lung, cardiac failure and arrhythmia, and cerebral edema with further central nervous complications. The most significant biochemical changes of acute and chronic renal failure are increased levels of urea and creatinine, electrolyte imbalances (often decreased with acute renal failure) and also a reduced concentration of urine [1, 2, 6].

Creatinine

Under postmortem conditions, an increased level of creatinine in CSF and vitreous humor can be indicative of renal failure (normal value 0.6–1.4 $mg\,dl^{-1}$).

Table 2 Postmortem biochemical values in cases of renal failure (insufficiency)

Dysfunction	Parameter	Clinical values	Compartment	Results
Compensated retention	Creatinine	$\leq 6\,mg\,dl^{-1}$		
Preterminal failure (f)		$>8\,mg\,dl^{-1}$		
Terminal failure		$>10\,mg\,dl^{-1}$		
RF ruled out			CSF/VH	Creatinine $<2.5\,mg\,dl^{-1}$
RF possible				$2.5–4.0\,mg\,dl^{-1}$
RF primary fatal				$>4.0\,mg\,dl^{-1}$
Normal values (urea–nitrogen)/urea[(a)]			Blood (heart)	Maximum 179 (83) $mg\,dl^{-1}$ (mean value = 102 (47) $mg\,dl^{-1}$)
			CSF	Maximum 197 (92) $mg\,dl^{-1}$ (mean value = 89 (41) $mg\,dl^{-1}$)
Uremia (urea–nitrogen)/urea[(a)] dysfunction[(b)]		(First 13 hpm)	Blood and CSF	$>200\,mg\,dl^{-1}$ (93)
		CSF/blood (heart)	CSF	Blood (heart)
RF ruled out		Urea $<100\,mg\,dl^{-1}$	Creatinine $<2.5\,mg\,dl^{-1}$	Creatinine $<3.5\,mg\,dl^{-1}$
RF possible		$100–200\,mg\,dl^{-1}$	$2.5–4.0\,mg\,dl^{-1}$	$3.0–4.5\,mg\,dl^{-1}$
RF primary fatal		$>200\,mg\,dl^{-1}$	$>4.0\,mg\,dl^{-1}$	$>4.5\,mg\,dl^{-1}$

RF, renal failure; CSF, cerebrospinal fluid; VH, vitreous humor; pm, postmortem

[(a)] Urea–nitrogen $\times$ 2148 ($mg\,dl^{-1}$) = urea ($mg\,dl^{-1}$)

[(b)] Different method of assessment (see text and references)

During the early postmortem interval, the creatinine concentration is rather stable. In healthy individuals, the mean values are $1.6\,mg\,dl^{-1}$ (8 h postmortem), $1–2\,mg\,dl^{-1}$ (12 h postmortem) and $3–4\,mg\,dl^{-1}$ (24 h postmortem). Therefore, reliable assessment is possible for pathological levels if the specimens are obtained during the early postmortem period.

Renal failure can be ruled out if the creatinine level is below $2.5\,mg\,dl^{-1}$. It is possible if its concentration ranges between 2.5 and $4.0\,mg\,dl^{-1}$ and renal failure is to be considered as the primary cause of death with levels exceeding $4.0\,mg\,dl^{-1}$, given CSF being obtained within the first hours postmortem. After death, the normal relation between creatinine levels in serum and CSF remains almost the same.

On one hand, problems may arise from a connection between renal damage and creatinine level. On the other hand, high creatinine values are seen without any or only slight alterations of the kidneys. However, there is also the possibility that advanced kidney damage coincides with levels below $4\,mg\,dl^{-1}$. It must be pointed out that disturbances of the circulation and toxicemia may cause creatinine retention but that partial renal function can still be in place during uremia [6].

Urea

In case of renal failure, there exists a close relation between the levels of urea in serum and CSF (normal range: $13.8–34.6\,mg\,dl^{-1}$). The urea level in CSF is approximately three fourths of the serum value. However, there have been reported reduced levels in CSF and also slight increases in blood from the femoral veins and also in liquor compared to antemortem values and also independent from the cause of death. If renal diseases can be excluded, such changes may be due to agonal or postmortem effects. Furthermore, there is a rising difference between the

concentrations of urea in liquor and blood with the postmortem interval increasing. Often postmortem values are slightly higher compared to intravital estimations. However, this increase is lower if the intravital concentration has been rather high. In case of manifest renal insufficiency, possibly with uremia, there are usually considerable differences to the levels found in healthy individuals.

There is an arithmetical connection between urea–nitrogen and urea, which is as follows: urea–nitrogen $\times$ 2.148 (mg dl^{-1}) = urea (mg dl^{-1}). Urea levels in CSF above 20 mg dl^{-1} (9.3 mg dl^{-1} urea–nitrogen) are indicative of renal disease, whereas the postmortem "normal values" for blood from the heart is 179 mg dl^{-1} maximum (83 mg dl^{-1} urea–nitrogen), with a mean value of 102 mg dl^{-1} (47 mg dl^{-1} urea–nitrogen). The corresponding concentrations in CSF are 197 mgYdl^{-1} (92 mg dl^{-1} urea–nitrogen) with a mean value of 89 mg dl^{-1} (41 mg dl^{-1} urea–nitrogen). Contrary to this, urea levels in CSF and blood from the heart do usually exceed 200 mg dl^{-1} (93 mg dl^{-1} urea–nitrogen) during the first 13 h postmortem in the case of uremia from all imaginable causes [26].

Diagnosis

Postmortem estimation of creatinine and urea levels in blood from the heart (left ventricle preferred) and CSF have important relevance regarding the postmortem diagnosis of renal failure. The following ranges of values can be differentiated for a practicable combined diagnostic procedure:

Urea below 100 mg dl^{-1} in CSF/blood, creatinine below 2.5 mg dl^{-1} in liquor and below 3.5 mg dl^{-1} in blood: renal failure can be excluded.

Urea 100–200 mg dl^{-1} in CSF/blood: renal failure possible if there is an additional creatinine level of 2.5–4.0 mg dl^{-1} in liquor and of 3.0–4.5 m gdl^{-1} in blood from the heart. Urea above 200 mg dl^{-1} in CSF or blood: renal failure represents the primary cause of death if creatinine levels in liquor simultaneously exceed 4.0 and 4.5 mg dl^{-1} in blood from the heart [2, 6].

Water- and Electrolyte Imbalances

The regulation of the water and electrolyte balance aims to maintain *isotonia* and *isovolumia* within the intravasal space. Sodium, chloride, and bicarbonate show the highest extracellular concentrations, whereas potassium and phosphoric esters predominate in the intracellular space. Owing to the fact that the relation between extracellular fluid volume and water exchange is much lower in *infants* than in adults, water imbalances may develop much earlier and be life-threatening.

It is not rare that electrolyte imbalances occur due to other diseases such as diabetes mellitus, chronic alcoholism, and nutritive disturbances. There are some types of dysregulations, which can lead to sudden unexpected death and may therefore be of forensic medical relevance. *Isotonic dehydration* is characterized by extracellular loss of sodium and water in isotonic relation, e.g., during the polyuric phase of acute and chronic renal failure, vomiting and diarrhea, pancreatitis and peritonitis, and due to dermal loss (following burn injuries). The main mechanism of *hypotonic dehydration* is salt depletion together with extracellular deficit of water. Delirium and convulsions are typical cerebral symptoms, which have to be considered as causes of sudden death. *Hypertonic dehydration* (with hypernatremia) leads to a deficit of free water in the extracellular and also in the intracellular space and is caused e.g., by a lack of water supply, dermal loss (sweating), and also via the lungs (e.g., hyperventilation from infections and fever), the kidneys (diabetic coma), and the gastrointestinal tract (diarrhea, vomiting). The typical morphology comprises tinting of the skin, sunken eyes, dry surface of the galea and/or dry cutting areas of organs. A biochemical pattern was proposed as diagnostic tool. The so-called dehydration pattern consists of an elevation of sodium $>$155 mmol l^{-1}, chloride $>$135 mmol l^{-1}, and urea $>$40 mg dl^{-1}. Persisting imbalances also result in corresponding alterations within the CSF (osmotic gradient) [27, 28].

Regarding the postmortem diagnosis of *water and electrolyte imbalances,* measurements of the pH is of no value. Estimations of electrolytes in CSF and vitreous humor can only be of limited meaningfulness. On one hand, the pH strongly depends on the state of the body, and, on the other hand, liquor often becomes sanguinolent when it is obtained so that there may be considerable alterations especially to electrolytes. Centrifugation may be of certain help, but cannot remove all components originating from damaged erythrocytes. This is why liquor from the lateral ventricles should be obtained, because after

12–24 h there are no differences to lumbar liquor [29].

Potassium

Disturbances of the potassium balance can gain forensic medical relevance because they have been described to occur not only isolated but also in connection with other diseases and sudden death (acute myocardial failure due to arrhythmias). Particularly, intestinal or renal loss or insulin treatment of diabetic coma are likely to result in *hypokalemia* ($<3.6\,\text{mmol}\,l^{-1}$). The main causes of *hyperkalemia* ($>5.0\,\text{mmol}\,l^{-1}$) are acute renal failure, chronic renal insufficiency, or extensive tissue damages. The main possible complications are disturbances of conduction, ventricular flutter, and fibrillation, which may lead to asystolia (acute danger to life with potassium levels $>6.5\,\text{mmol}\,l^{-1}$) [1, 2, 6].

Estimation of potassium in blood and serum specimens obtained postmortem have proven not to be reliable due to extremely fast and intense potassium release from cytolysis. In CSF, the potassium value can reach up to sevenfold of the normal level within the first 10 h postmortem, but the range of variation is rather wide. The potassium content of liquor is, to a large extent, independent from the serum level and in infants lower than in adults (normal range: *ca.* $2.1–4.6\,\text{mmol}\,l^{-1}$). Contrary to this, the increase of the potassium concentration in vitreous humor has been reported to be regular. This can provide certain conclusions regarding the time of death within the first 12 h postmortem. There seem to be no other relevant disturbances from other diseases on the potassium content of vitreous humor except hepatic failure. Furthermore, there do not exist any comprehensible associations between the concentration differences of sodium and potassium, which appear to allow further reliable conclusions [30–34].

Sodium/Chloride

There is an extracellular decrease of *sodium* parallel to an increase of potassium (see above) postmortem. As a general rule, there is a variation of the sodium level within CSF mostly corresponding to the serum concentration (*ca.* $128–157\,\text{mmol}\,l^{-1}$), except situations with severe infections of the central nervous system.

Without differentiation regarding the mechanisms of death, sodium levels in CSF and serum are usually found within the normal range, but the variation range differs considerably from intravital values (*ca.* $123–205\,\text{mmol}\,l^{-1}$). Although there is a distinct decrease of sodium in CSF and serum after death, its concentration in vitreous humor remains rather stable up to 30 h postmortem, followed by an almost linear decrease in the following 50 h. Sodium levels above $155\,\text{mmol}\,l^{-1}$ and below $130\,\text{mmol}\,l^{-1}$ in adults and larger differences outwith the normal range in children can be indicative of hypernatremia or hyponatremia antemortem. Sodium levels in *fluid obtained from the pericardial sac* show distinct correlation to the postmortem interval, namely, a decrease of approximately $0.4\,\text{mmol}\,l^{-1}$ during the first 85 h after death, but also with a wide range of variation.

The level of *chloride* in CSF is approximately 20% higher compared to serum and shows a range of *ca.* $110–129\,\text{mmol}\,l^{-1}$ in healthy individuals. The postmortem changes of chloride are comparable to those of sodium (see above), so that there happens also a typical decrease of the chloride concentration in plasma and CSF. The levels of chloride and sodium in vitreous humor appear to be almost "parallel" and remain nearly constant for over 30 h postmortem. However, any close correlations between chloride values and causes of death or time could not be identified postmortem.

Calcium

The homoeostasis of calcium has an important impact onto the neuromuscular conduction. *Hypocalcemia* (total Ca $<2.2\,\text{mmol}\,l^{-1}$, ionized Ca $<1.1\,\text{mmol}\,l^{-1}$) results in pathological reflexes or arrhythmia. Causes of *hypocalcemia* (total Ca $>2.7\,\text{mmol}\,l^{-1}$, ionized Ca $>1.3\,\text{mmol}\,l^{-1}$) are chronic osteolytic or endocrine processes in most of the cases, which may be the reason for sudden unexpected deaths via electrolyte imbalances with arrhythmias, somnolence, and coma.

Under postmortem conditions, the serum calcium concentration is constant for *ca.* 10 h with a slight increase thereafter (normal range in healthy individuals: $1.96–2.60\,\text{mmol}\,l^{-1}$). The calcium contents of CSF reflect approximately the serum level of ionized calcium. In vitreous humor, calcium levels are much more stable and there is less influence of agonal and postmortem effects [1, 2].

Diagnosis

Postmortem diagnosis of imbalances of electrolyte and water metabolism cannot be based on isolated single parameters. Assessment must always include a synopsis of different values. Furthermore, the postmortem interval has to be taken into account in each case. Postmortem biochemical analyses regarding electrolyte imbalances are believed to be most successful in cases being characterized by elevations of parameters such as states of dehydration. One main disadvantage are the wide range of variation referred to single analyze results. This requires a combined interpretation of different values with consideration of all morphological and toxicological findings as well as the possibility of combined dysregulations (e.g., kidney and glucose metabolism).

High Excitation and Hypothermia

A state of *high excitation* is characterized by a massive release of catecholamines, especially in situations with mechanical restraints and also in cases of prolonged agony. Such stress situations can be classified by estimation of adrenaline and noradrenaline levels using high performance liquid chromatography (HPLC) in serum, CSF, and vitreous humor. Analyses in different compartments is useful to achieve semiquantification of the intensity of stress and its impact on the mechanism of death.

Particularly increased noradrenaline levels in CSF and vitreous humor are indicative of a protracted stress reaction. The author's research has revealed massively increased catecholamine concentrations, partly exceeding the normal ranges many times (adrenaline values in vitreous humor and CSF 100–8000 $ng\,l^{-1}$; noradrenaline levels 4000–70 000 $ng\,l^{-1}$ (normal ranges in serum: adrenaline 20–120 $ng\,l^{-1}$ and noradrenaline 150–170 $ng\,l^{-1}$)). Especially high noradrenaline levels indicate a longer duration of stress [35–37].

Hypothermia can also cause a massive release of catecholamines in the sense of intense stress. The levels are within the ranges of high excitation with the noradrenaline concentrations being considerably higher than those of adrenaline (10- to 32-fold) comparable to cases with prolonged agony. Contrary to this, adrenaline levels often exceed those of noradrenaline in death cases with short agony. Death due to hypothermia results in mean quotients adrenaline/noradrenaline considerably <1, whereas quotients >1 are typical for short agony (e.g., myocardial infarction, head trauma) being indicative of higher adrenaline levels [38].

Additional analyses of volatile substances (ethanol, methanol, propanol-1, propanol-2, and acetone) usually show elevated acetone concentrations in all compartments being indicative of hypothermia, but basically only in cases that are ethanol-free. Acetone and propanol-2 are then altered equally. If relevant alcoholization is found, both substances can only be found in very low or physiological ranges that is indicative of an antilipolytic effect of ethanol (acetone $>35\,mg\,l^{-1}$ if blood alcohol level is $<10\,mg\,dl^{-1}$ vs. $<5\,mg\,l^{-1}$ if blood alcohol level is $>185\,mg\,dl^{-1}$) [39].

Conclusions and Final View

Regarding the postmortem diagnosis of *fatal diabetic coma,* morphological findings are only of indicative value. Therefore the diagnosis "death due to diabetic coma" always has to be a *synopsis comprising medical history, macromorphology and histology completed by postmortem biochemistry.* Specimens (CSF, vitreous humor, blood, and urine) should be obtained if there is any suspicion on disturbances of the glucose metabolism. Parameters of major relevance are "sum value" and hemoglobin A_{1c}, which are found to be elevated in most cases of fatal coma (above $415\,mg\,dl^{-1}$ and 12.1%, respectively). The level of free acetone usually exceeds $21\,mg\,l^{-1}$ and urine glucose concentration exceeds $500\,mg\,dl^{-1}$. A correct diagnosis always requires a combination of a minimum of three positive values e.g., increased "sum value", hemoglobin A_{1c} positive, and elevated acetone concentration *or* increased "sum value" and several indicative findings within macromorphology and histology (*see also* **Natural Causes of Sudden Death: Noncardiac**).

Under forensic medical aspects, the diagnosis of fatal diabetic coma can only be made *per exclusionem.* Consequently, other mechanisms of death, e.g., intoxications, have to be ruled out. However, overlapping with other causes of death appears to be rather typical and common. Owing to the fact that the whole diagnostic procedure can only be carried out *per exclusionem,* the only area of overlapping causing problems is that with "natural

causes of death" because myocardial infarctions or pulmonary embolism may both represent real complications of diabetic coma and can as well cause metabolic decompensation to preexisting diabetes mellitus. Especially in cases with acute myocardial infarctions differentiation may be problematic, but contrary to such acute changes, the situation is different with chronic alterations as e.g., narrowing coronary arteriosclerosis or myocardial scars. With such preconditions, the higher the relevance of positive biochemical findings is, the more intensive they appear to be (very high "sum value" and acetone level, etc.).

Postmortem biochemical examinations can also provide help in cases without morphological causes of death outwith the field of diabetes mellitus so that specimens of body fluids should also be obtained. Often analyses on certain parameters sensibly complement postmortem morphological diagnostics as in cases of liver disease, chronic renal failure, and electrolyte imbalances. Preliminary studies have also been carried out on the usefulness of other body compartments (e.g., synovial fluid) for a range of examinations as well as for further biochemical parameters (e.g., troponin T) [40, 41].

It has to be mentioned that urea levels in blood and CSF are likely to be elevated in case of *chronic kidney disease* and furthermore slightly following death, but this increase has been found to be considerably lower in liquor compared to blood. Postmortem diagnosis of renal insufficiency can be made with urea levels above $200\,\mathrm{mg\,dl^{-1}}$ (urea-nitrogen in excess of $93\,\mathrm{mg\,dl^{-1}}$). Creatinine concentrations seem to remain widely unaltered in all body fluids postmortem. The most reliable examinations are possible in CSF with a level below $1.6\,\mathrm{mg\,dl^{-1}}$, expected to be typical in individuals without kidney disease. However, postmortem biochemistry can only represent one pillar of the procedure to establish the cause of death as such results are unsuitable to be used as the only diagnostic criterion.

Especially, a combined spectrum of postmortem biochemical values is of most relevant meaningfulness, regarding the diagnosis of fatal metabolic disturbances. They may be strongly indicative of chronic or acute mechanisms and diseases, although, of course, no clinical diagnoses based on postmortem findings can be made. *Exclusion of any competitive mechanisms* is of special importance. A final diagnose regarding the cause of death can only be made by the inclusion of medical history, macromorphology, histology findings, postmortem biochemical results, and toxicology and *per exclusionem* only.

References

[1] Kernbach-Wighton, G. (2003). The diagnosis of functional causes of death, in *Practice of Legal Medicine*, B. Madea, ed, Springer, Berlin, Heidelberg, New York, Tokyo, pp. 239–244.

[2] Kernbach-Wighton, G. (2006). Possibilities of postmortem biochemical diagnostics. Starting points, measurement techniques, evaluation and conclusions, *Rechtsmedizin* **16**, 27–36.

[3] Madea, B. & Musshoff, F. (2007). Postmortem biochemistry, *Forensic Science International* **165**, 165–171.

[4] Coe, J.I. (1972). Use of chemical determinations on vitreous humour in forensic pathology, *Journal of Forensic Sciences* **17**, 541–546.

[5] Kernbach, G. & Brinkmann, B. (1983). Postmortem patho-biochemistry used for the diagnosis of fatal diabetic coma as cause of death, *Der Pathologe* **4**, 235–240.

[6] Kernbach-Wighton, G. (2003). Post mortem biochemical examinations, in *Handbook of Forensic Medicine*, Springer, Berlin, Heidelberg, New York, Tokyo, Vol. 1, pp. 1060–1069.

[7] DiMaio, V.J.M., Sturner, W.Q. & Coe, J.I. (1977). Sudden unexpected deaths after the acute onset of diabetes mellitus, *Journal of Forensic Sciences* **22**, 147–151.

[8] Gormsen, H. & Lund, A. (1985). The diagnostic value of postmortem blood glucose determinations in cases of diabetes mellitus, *Forensic Science International* **28**, 103–107.

[9] Traub, F. (1969). Method for the diagnosis of fatal disturbances of glucose metabolism from post mortem findings (diabetes mellitus and hypoglycemia), *Zentralblatt fuer Allgemeine Pathologie* **112**, 390–399.

[10] Kernbach, G., Püschel, K. & Brinkmann, B. (1986). Biochemical measurements of glucose metabolism in relation to cause of death and postmortem effects, *Zeitschrift für Rechtsmedizin* **96**, 199–213.

[11] De Letter, E.A. & Piette, M.H. (1998). Can routinely combined analysis of glucose and lactate in vitreous humour be useful in current forensic practice? *The American Journal of Forensic Medicine and Pathology* **19**, 335–342.

[12] Gagajewski, A., Murakami, M.M., Kloss, J., Edstrom, M., Hillyer, M., Peterson, G.F., Amatuzio, J. & Apple, F.S. (2004). Measurement of chemical analytes in vitreous humour: stability and precision studies, *Journal of Forensic Sciences* **49**, 371–374.

[13] Karlovsek, M.Z. (2004). Diagnostic values of combined glucose and lactate values in cerebrospinal fluid and vitreous humour-our experiences, *Forensic Science International* **146**(suppl 1), 19–23.

[14] Sippel, H. & Möttönen, M. (1982). Combined glucose and lactate values in vitreous humour for post-mortem diagnosis of diabetes mellitus, *Forensic Science International* **19**, 217–222.

[15] Sturner, W.Q., Sullivan, A. & Suzuki, K. (1983). Lactic acid concentrations in vitreous humour: Their use in asphyxial deaths in children, *Journal of Forensic Sciences* **28**, 222–230.

[16] John, W.G., Scott, K.W.M. & Hawkroft, D.M. (1988). Glycated haemoglobin and glycated protein and glucose concentration in necropsy blood samples, *Journal of Clinical Pathology* **41**, 415–418.

[17] Khun, H.M., Robinson, C.A., Brissie, B.M. & Konrad, R.J. (1999). Post mortem diagnosis of unsuspected diabetes mellitus established by determination of decendent's haemoglobin A1c level, *Journal of Forensic Sciences* **44**, 643–646.

[18] Coe, J.I. (1993). Postmortem chemistry update. Emphasis on forensic application, *The American Journal of Forensic Medicine and Pathology* **14**, 91–117.

[19] Brinkmann, B., Fechner, G., Karger, B. & DuChesne, A. (1998). Ketoacidosis and lactic acidosis – frequent cause of death in chronic alcoholics? *International Journal of Legal Medicine* **111**, 115–119.

[20] Osuna, E., Garcia-Villora, A. & Perez-Carceles, M.D. (1999). Vitreous humour fructosamine concentrations in the autopsy diagnosis of diabetes mellitus, *International Journal of Legal Medicine (Tokyo)* **112**, 275–279.

[21] Kernbach-Wighton, G. & Püschel, K. (1998). On the phenomenology of lethal applications of insulin, *Forensic Science International* **93**, 61–73.

[22] Kernbach-Wighton, G., Sprung, R. & Püschel, K. (2001). On the diagnosis of hypoglycaemia in car drivers – including a review of the literature, *Forensic Science International* **115**, 89–94.

[23] Kernbach-Wighton, G. & Püschel, K. (2003). The evidence of carbohydrate metabolism disturbances in traffic delinquents, *Legal Medicine* **5**, 237–239.

[24] Logemann, E., Pollak, S., Khalaf, A.N. & Petersen, K.G. (1993). Postmortem diagnosis of exogenous insulin administration, *Archiv für Kriminologie* **191**, 28–36.

[25] Winston, D.C. (2000). Suicide via insulin overdose in nondiabetics: the New Mexico experience, *The American Journal of Forensic Medicine and Pathology* **21**, 237–240.

[26] Zhu, B.L., Ishikawa, T., Michiue, T., Tanaka, S., Zhao, D., Li, D.R., Quan, L., Oritani, S. & Maeda, H. (2007). Differences in postmortem urea nitrogen, creatinine and uric acid levels between blood and pericardial fluid in acute death, *Legal Medicine (Tokyo)* **9**, 115–122.

[27] Madea, B. (1996). Post mortem diagnosis of water and electrolyte imbalances, *Rechtsmedizin* **6**, 141–146.

[28] Madea, B. & Lachenmeier, D.W. (2005). Postmortem diagnosis of hypertonic dehydration, *Forensic Science International* **155**, 1–6.

[29] Mulla, A., Massey, K.L. & Kalra, J. (2005). Vitreous humour biochemical constituents: evaluation of between-eye differences, *The American Journal of Forensic Medicine and Pathology* **26**, 146–149.

[30] Henßge, C., Knight, B., Krompecher, T., Madea, B. & Nokes, I. (1995). *The Estimation of the Time of Death in the Early Postmortem Period*, Arnold, London.

[31] James, R.A., Hoadley, P.A. & Sampson, B.G. (1997). Determination of postmortem interval by sampling vitreous humour, *The American Journal of Forensic Medicine and Pathology* **18**, 158–162.

[32] Madea, B., Herrmann, N. & Henßge, C. (1990). Precision of estimating the time since death by vitreous potassium – comparison of two different equations, *Forensic Science International* **46**, 277–284.

[33] Madea, B. & Rödig, H. (2006). Time of death dependent criteria in vitreous humour: accuracy of estimating the time since death, *Forensic Science International* **164**, 87–92.

[34] Rognum, T.O., Hauge, S., Oyasaeter, S. & Saugstad, O.D. (1991). A new biochemical method for estimation of postmortem time, *Forensic Science International* **51**, 139–146.

[35] Zhu, B.L., Ishikawa, T., Michiue, T., Li, D.R., Zhao, D., Quan, L., Oritani, S., Bessho, Y. & Maeda, H. (2007). Postmortem serum catecholamine levels in relation to cause of death, *Forensic Science International* **173**, 122–129.

[36] Kernbach-Wighton, G. & Saternus, K.S. (2004). On the post mortem diagnosis of hypothermia, in *Hypothermia. Clinical, Pathomorphological and Forensic Features. Research in Legal Medicine*, M. Oehmichen, ed, Schmidt-Römhild, Lübeck, Vol. 31, pp. 221–229.

[37] Kernbach-Wighton, G. & Saternus, K.S. (2007). Postmortem biochemical estimations in cases of fatal hypothermia (catecholamines and volatiles), *Romanian Journal of Legal Medicine* **15**, 32–38.

[38] Kernbach-Wighton, G., Sprung, R., Kijewski, H. & Saternus, K.S. (2002). Maximum excitement and sudden death, in *Restraints of Excited Persons. Sudden Death in Hospital and Custody. Research in Legal Medicine*, K.S. Saternus & G. Kernbach-Wighton, eds, Schmidt-Römhild, Lübeck, Vol. 28, pp. 55–74.

[39] Kernbach-Wighton, G. & Saternus, K.S. (2006). Post mortem biochemical estimations in cases of lethal high excitation, *Romanian Journal of Legal Medicine* **14**, 251–261.

[40] Zhu, B.L., Ishikawa, T., Michiue, T., Li, D.R., Zhao, D., Oritani, S., Kamikodai, Y., Tsuda, K., Okazaki, S. & Maeda, H. (2006). Postmortem cardiac troponin T levels in the blood and pericardial fluid. Part 1. Analysis with special regard to traumatic causes of death, *Legal Medicine (Tokyo)* **8**, 86–93.

[41] Madea, B., Kreuser, C. & Banaschak, S. (2001). Postmortem biochemical examination of synovial fluid – a preliminary study, *Forensic Science International* **118**, 29–35.

GERHARD KERNBACH-WIGHTON

Postmortem Interval: Anthropology

One of the most difficult tasks for the forensic anthropologist is establishing the postmortem interval (PMI) – i.e., time since death. The determination of PMI on skeletal remains is affected by enormous limitations, which have to do with the great variability in decomposition rates.

Postmortem phenomena (such as putrefaction and other types of decomposition) severely alter markers, which are essential for the evaluation of the PMI (along with identification of the body, interpretation of manner of disposal of the body, and the manner of death, in particular signs of trauma). To correctly and thoroughly read and interpret human remains in this sense, intrinsic factors (pertaining to the body) may not be sufficient and it is becoming increasingly evident that the environment (i.e., extrinsic factors such as plants, pollen, algae, moulds, macrofauna, and microfauna) may be crucial for answering the question of PMI [1–8]. Much of the difficulty in determining the time since death and/or permanence of a body in a specific environment stems from the lack of systematic observations and research on the decomposition modalities in different environments of the human body. Postmortem changes have been dealt with extensively in the literature but are very specific and bear too little an interdisciplinary approach [1–28]. Many physicochemical changes begin to take place in the body immediately or shortly after death and progress in a fairly orderly manner until the body disintegrates. Each change has its own time factor or rate. Unfortunately, these rates of development of postmortem changes are strongly influenced by unpredictable endogenous and environmental factors. Consequently, the longer the PMI, the wider is the range of estimate as to when the death probably occurred. In other words, the longer the PMI, the less precise is the estimate of the time of death. Shortly after death, enzymes that occur naturally in the body begin the degradation processes (autolysis), subsequent decay by bacteria and fungi is purely aerobic and, ideally, leads to the entire skeletonization of the corpse. According to an old rule of thumb, 1 week of putrefaction in air is equivalent to 2 weeks in water, which is equivalent to 8 weeks buried in the soil, given the same environmental temperature. The interval between the time of death and final skeletonization is governed by the environment in which these processes occur, while the processes of degradation on the soil surface, which is biologically highly active, usually reach completion within several weeks, the time required for the decomposition of interred bodies takes between 3 and 12 years under favorable conditions. Under unfavorable conditions, the processes might require up to a hundred or even thousands of years before completion [29–43]. Once severe decay sets in, tools for determining PMI and permanence in a specific environment move from forensic pathology to forensic anthropology and taphonomy (i.e., the study of decaying organisms over time and the effects environmental factors may have on them) and to the multidisciplinarity of other forensic sciences, more or less well known, such as forensic entomology, forensic botany, zoology, hydrology, etc. In fact, forensic entomology is the only one of these disciplines that has been thoroughly investigated, but may not be sufficient. Taphonomy and other disciplines (such as even biomolecular diagenesis) must take their stand in the medicolegal world in order to solve the above-mentioned issues. This means that a multifactorial approach dealing with intrinsic and extrinsic factors must be sought. Very few studies exist in this sense, perhaps, because the world of natural sciences, anthropology, taphonomy, and forensic medicine are still too far apart. Until now, the following steps have been taken. Tenessee's body farm has led to several observations on decomposition processes of human remains in specific environments, but these deal mainly with macroscopic observations, entomological, and chemical ones. As regards human decomposition processes, sporadic case studies exist on this issue, on the surface, in soil, and in water, which, however, only deal mainly with body intrinsic and entomological factors. Even useful chemical aspects still need to be explored. For example, the articulated processes including physical, chemical, biochemical, and microbiological changes occurring during postmortem events have to be held responsible for the onset of odour. Odour, a complex mixture of volatile compounds, is a complicated and difficult attribute to measure. The electronic nose is an instrument that comprises an array of electronic chemical sensors with a partial specificity and an appropriate recognition system capable of recognizing simple or complex odors. Substantial work has been done to



Figure 1 Human remains of a young woman found in a woody area in northern Italy. Much mummified tissue still clings on to the bone surface. Postmortem interval was known to be eight months. Remains were found in spring

apply electronic nose technology for environment and food quality monitoring and evaluation in order to identify, recognize, and discriminate different odors. In food, animal species for human nutrition, spoilage odors as key indicators for fresh or not fresh fish (in conjunction with other factors, such as oxidative changes and growth of microrganisms), and time development of these volatile components in both fish and cereals have been identified. However, no studies are available on the precise identification and possible quantification of odorous molecules from corpses, and this argument at present is the domain of dog conductors, who train dogs to recognize a possible grave localization utilizing their ability to track decomposition odors. The repeatability of cadaver decay in animal models has been treated in an experimental study on the effect of freezing and thawing and on the decomposition of organic matter. Some work has been done for intrinsic changes of the pig body (and some entomology) particularly in forests and in a marine environment, but very few reports combine the data on intrinsic and extrinsic factors. From these few reports, interesting data emerge: the great variability of decomposition and the need to combine decomposition with environmental data. Botany (roots, leaves, and pollen) and zoology (diatoms, plankton, etc.) seem to have a crucial role (Figures 1, 2). The significant contribution of annual growth in woody tissue for PMI estimation has been proven and many have stressed the importance of extrinsic factors – however, the world of combined extrinsic and intrinsic factors still remains uninvestigated.

For the case of skeletonized or partly skeletonized human remains, it is frequently difficult to say whether such remains belong to someone who



Figure 2 Botanical remains can be of crucial importance. The image shows a grain of pollen in microscopy that denotes the season of deposition on human remains, and therefore a term *antequem non* since death

died 5, 10, 20, or even more years before. Variables dependant on the environment, climate, location, and even animal intervention can seriously modify decomposition rates and preservation of human remains in an unpredictable manner. The exposition to the sun, temperature, climatic rainfall, and wind conditions as well as soil chemical characteristics influence bone surface appearance and may alter PMI estimation.

A rough distinction at times can be made between archaeological and forensic remains. Some scientists claim that it is possible to distinguish among recent bones (less than 10 years), old bones (less than 50 years), and ancient bones according to a more or less dry appearance of the bone. In the past, PMI on bones

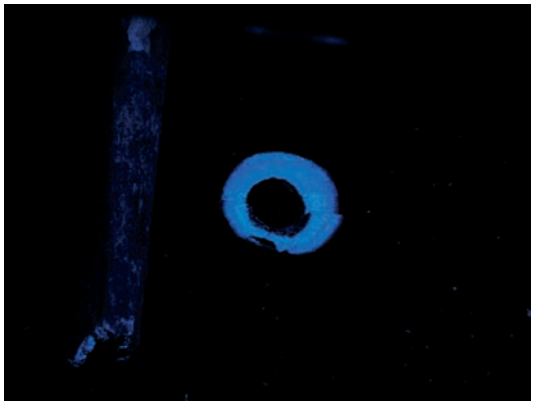


Figure 3 Cross section of a femur with a PMI of 5 years showing a greater degree of fluorescence with UV light

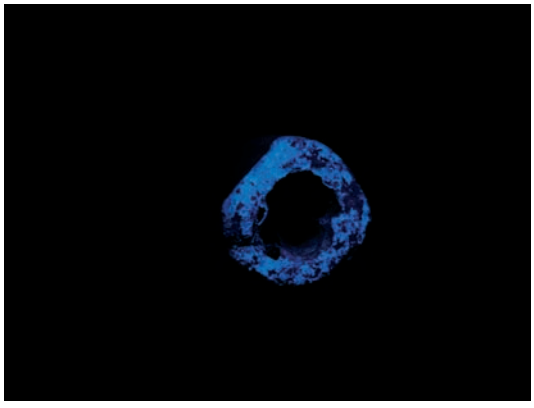


Figure 4 Cross section of a femur with a PMI of 50 years showing a lesser degree of fluorescence with UV light

has also been estimated according to the quantity of organic material still visible within the bone matrix by illumination with UV light (Figures 3 and 4) or the microscopic appearance of bone (Figures 5 and 6).

These methods, such as the macroscopica appearance, are, however, dependant on too many variables affecting degradation, which may not reflect the actual length of PMI. In other terms, a skeleton with a short PMI may look more degraded, if it was exposed to harsh environmental factors, than a skeleton with a longer PMI with less severe environmental variables. In spite of attempts to standardize fluorescent and microscopic patterns, these tests give no unequivocal classification, and therefore can provide only indicative data on PMI.

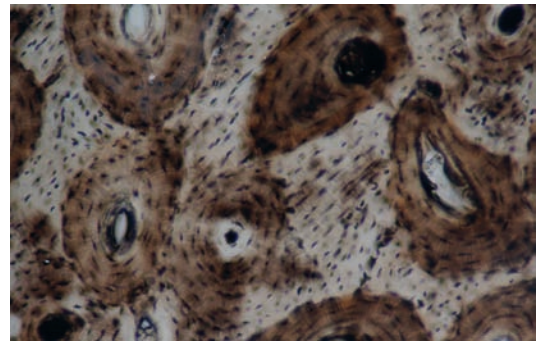


Figure 5 Microscopical aspect (100×) of fresh bone (PMI of several months). The structure of bone is clearly visible

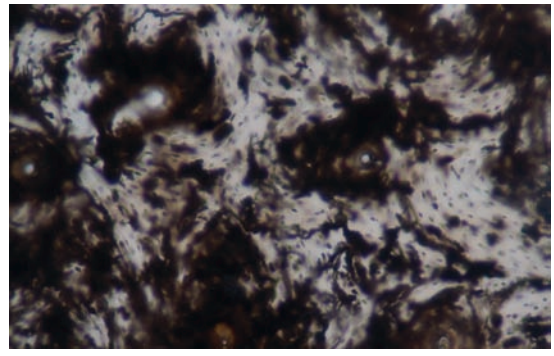


Figure 6 Microscopical aspect (100×) of old bone (PMI of 50 years) that shows severe degradation, destruction of the microscopical structure, and fungal infiltration

This is why in cases of retrieval of skeletal remains, the context becomes fundamental. The botanical analysis of leaves, roots, and seeds on and around the remains, for example, can give a *postquem non* or *antequem non* temporal limit by studying the elements that have colonized the site after the deposition. A PMI indication can also be reached throughout the analysis of clothes and personal belongings found at the site. Entomological study of insects colonizing remains may provide a more accurate indication about the age of deposition; corpses are in fact colonized by different insect populations in accordance with season, climatic conditions, environmental characteristics, and decomposition stage.

Radiochemical tests are another, more expensive, tool by which the anthropologist may try to determine PMI [35–43]. These are based on the decay rates of radioisotopes fixed in bone during life. The

use of Carbon-14 provides accurate information about the PMI in case of archeological remains, but its importance in forensic material is reduced because its before period (BP) date refers to the 1950s. Strontium 90 may also be useful; its high atmospheric concentration is due to massive explosions and nuclear weapon pollution that occurred during World War II. Another method benefits from the high levels of artificially introduced Carbon-14 in terrestrial organisms by thermonuclear devices between 1950 and 1963. It is based on the observation that different tissues of the body have a diverse time of development and a diverse turnover rate. Quantifying radiocarbon values in different tissues and placing these values in the bomb-curve, taking care of noting the age of the individual and other factors, allows an estimation of the date of death.

All such methods, however, still need calibration and testing for contamination – and remain exclusive to few laboratories and are quite expensive.

References

[1] Anderson, G.S. & Hobischak, N.R. (2004). Decomposition of carrion in the marine environment in British Columbia, *Canadian International Journal of Legal Medicine* **118**(4), 206–209.

[2] Archer, M.S. (2004). Rainfall and temperature effects on the decomposition rate of exposed neonatal remains, *Science and Justice* **44**(1), 35–41.

[3] Aturaliya, S. & Lukasewycz, A. (1999). Experimental forensic and bioanthropological aspects of soft tissue taphonomy: 1. Factors influencing postmortem tissue desiccation rate, *Journal of Forensic Sciences* **44**(5), 893–896.

[4] Bell, L.S., Skinner, M.F. & Jones, S.J. (1996). The speed of postmortem change to the human skeleton and its taphonomic significance, *Forensic Science International* **82**(2), 129–140.

[5] Boddingtom, A., Garland, A.N. & Janaway, R.C. (1987). *Death, Decay and Reconstruction*, Manchester University Press, Manchester.

[6] Clark, K., Evans, L. & Wall, R. (2006). Growth rates of the blowfly, *Lucilia sericata*, on different body tissues, *Forensic Science International* **156**(2–3), 145–149.

[7] Coe, J.I. (1973). Postmortem chemistry: practical considerations and a review of literature, *Journal of Forensic Sciences* **19**, 13–32.

[8] Courtin, G.M. & Fairgrieve, S.L. (2004). Estimation of postmortem interval (PMI) as revealed through the analysis of annual growth in woody tissue, *Journal of Forensic Sciences* **49**(4), 781–783.

[9] Evans, W.E. (1963). *The Chemistry of Death*, CC Thomas, Springfield.

[10] Galloway, A., Birkby, W., Jones, A.M., Henry, T.E. & Parks, B.O. (1989). Decay rates of human remains in an arid environment, *Journal of Forensic Sciences* **34**(3), 607–616.

[11] Catts, E.P. & Goff, M.L. (1992). Forensic entomology in criminal investigations, *Annual Review of Entomology* **37**, 253–272.

[12] Haglund, W.D. (1988). Contribution of rodents to postmortem artifacts of bone in animal scavenged human skeletons, *Journal of Forensic Sciences* **33**(4), 985–997.

[13] Haglund, W.D. (1993). Disappearance of soft tissue and the disarticulation of human remains from aqueous environments, *Journal of Forensic Sciences* **38**(4), 806–815.

[14] Haglund, W.D. (1997). Dogs and coyotes: post-mortem involvement with human remains, in *The Postmortem Fate of Human Remains*, W.D. Haglund & M.H. Sorg, eds, CRC Press, Boca Raton.

[15] Haglund, W.D. & Reay, D.T. (1993). Problems of recovering partial human remains at different times and locations: concerns for death investigators, *Journal of Forensic Sciences* **38**(1), 69–89.

[16] Haglund, W.D., Reay, D.T. & Swindler, D.R. (1989). Canid scavenging/disarticulation sequence of human remains in the Pacific Northwest, *Journal of Forensic Sciences* **34**(3), 587–606.

[17] Haglund, W.D. & Sorg, M.H. (1997). *The Postmortem Fate of Human Remains*, CRC Press, Boca Raton.

[18] Haglund, W.D. & Sorg, M.H. (2002). *Advances in Forensic Taphonomy*, CRC Press, Boca Raton.

[19] Hobishak, N.R. & Anderson, G.S. (2002). Time of submergence using acquatic invertebrate succession as markers of decompositional change, *Journal of Forensic Sciences* **47**(1), 142–151.

[20] Huntington, T.E., Higley, L.G. & Baxendale, F.P. (2007). Maggot development during morgue storage and its effect on estimating the post-mortem interval, *Journal of Forensic Sciences* **52**(2), 453–458.

[21] Introna, F., Di Vella, G. & Campobasso, C.P. (1999). Determination of postmortem interval from old skeletal remains by image analysis of luminol test results, *Journal of Forensic Sciences* **44**(3), 535–538.

[22] Kahana, T., Almog, J., Levy, J., Schmeltzer, E., Spier, Y. & Hiss, J. (1999). Marine taphonomy: adipocere formation in as eries of bodies recovered from a single shipwreck, *Journal of Forensic Sciences* **44**(5), 897–901.

[23] Lopes de Carvalho, L.M. & Linhares, A.X. (2001). Seasonality of insect succession and pig carcass decomposition in natural forest area in southeastern Brazil, *Journal of Forensic Sciences* **46**(3), 604–608.

[24] Mann, R.M., Bass, W.M. & Meadows, L. (1990). Time since death and decomposition of the human body: variables and observations in case and experimental field studies, *Journal of Forensic Sciences* **35**(1), 103–111.

[25] Micozzi, M.S. (1986). Experimental study of postmortem change under field conditions: effects of freezing, thawing and mechanical injury, *Journal of Forensic Sciences* **31**(3), 953–961.

[26] Micozzi, M.S. (1991). *Postmortem Change in Human and Animal Remains*, Charles C Thomas Publisher, Springfield.
[27] Pollard, A.M. (1997). Dating the time of death, in *Studies in Crime: An Introduction to Forensic Archaeology*, J. Hunter, C. Roberts & A. Martin, eds, BT Batsford, London.
[28] Prieto, J.L., Magana, C. & Ubelaker, D.H. (2004). Interpretation of postmortem change in cadavers in Spain, *Journal of Forensic Sciences* **49**(5), 918–923.
[29] Rodriguez, W.C. & Bass, W.M. (1983). Insect activity and its relationship to decay rates of human cadavers in East Tennessee, *Journal of Forensic Sciences* **28**, 423.
[30] Sledzik, P.S. (1998). Forensic taphonomy: postmortem decomposition and decay, in *Forensic Osteology: Advances in the Identification of Human Remains*, K.J. Reichs, ed, Charles C Thomas Publisher, Springfield.
[31] Sorg, M.H., Dearborn, J.H., Monahan, E., Ryan, H.F., Sweeney, K.G. & David, E. (1997). Forensic taphonomy in marine contexts, in *The Postmortem Fate of Human Remains*, W.D. Haglund & M.H. Sorg, eds, CRC Press.
[32] Spenneman, D.H. & Franke, B. (1995). Decomposition of buried human bodies and associated death scene materials on coral atolls in the tropical Pacific, *Journal of Forensic Sciences* **40**(3), 356–367.
[33] Swift, B., Lauder, I., Black, S. & Norris, J. (2001). An estimation of the post-mortem interval in human skeletal remains: a radionuclide and trace element approach, *Forensic Science International* **117**(1–2), 73–87.
[34] Swift, B. (1998). Dating human skeletal remains: investigating the viability of measuring the equilibrium between ^{210}Po and ^{210}Pb as a means of estimating the post-mortem interval, *Forensic Science International* **98**(1–2), 119–126.
[35] Turner, B. & Wiltshire, P. (1999). Experimental validation of forensic evidence: a study of the decomposition of buried pigs in a heavy clay soil, *Forensic Science International* **101**(2), 113–122.
[36] Ubelaker, D.H. (2001). Artificial radiocarbon as an indicator of recent origin of organic remains in forensic case, *Journal of Forensic Science* **46**(6), 1285–1287.
[37] Vass, A.A., Barshick, S.A. & Sega, G. (2002). *et al.* Decomposition chemistry of human remains: a new methodology for determining the postmortem interval, *Journal of Forensic Sciences* **47**(3), 542–553.
[38] Vass, A.A. (2001). Beyond the grave: understanding human decomposition, *Microbiology Today* **28**, 190–192.
[39] Vass, A.A., Bass, W.M., Wolt, J., Foss, J. & Ammons, J. (1992). Time since death determinations of human cadavers using soil solution, *Journal of Forensic Sciences* **37**(5), 1236–1253.
[40] Vass, A.A., Smith, R.R., Thompson, C.V., Burnett, M.N., Dennis, A., Synstelien, J.A., Dulgerian, N. & Eckenrode, B.A. (2004). Decompositional odor analysis database, *Journal of Forensic Sciences* **49**(4), 760–769.
[41] Verhoff, M.A., Wiesbrock, U.O. & Kreutz, K. (2004). Macroscopic findings for the exclusion of a forensic relevant soil embedded resting period in skeletal remains – an approach based upon literature, *Archiv fur Kriminologie* **213**(1–2), 1–14.
[42] Willey, P. & Heiman, A. (1987). Estimating time since death using plant roots and stems, *Journal of Forensic Sciences* **32**, 1264.
[43] Yoshino, M., Kimijima, T., Miyasaka, S., Sato, H. & Seta, S. (1991). Microscopical study on estimation of time since death in skeletal remains, *Forensic Science International* **49**(2), 143–158.

Related Articles

Anthropology
Botany
Entomology

CRISTINA CATTANEO AND DANIELE GIBELLI

Postmortem Toxicology: Analysis

see Postmortem Toxicology: Laboratory Analysis

Postmortem Toxicology: Artifacts

Introduction

In unnatural, sudden, violent, or unexpected deaths, the investigator often needs evidence whether a foreign compound is present in autopsy material. Moreover, quantification of a drug is necessary to state whether its amount is sufficient to cause, prevent, or be involved directly in the death [1]. The interpretation of an analytical result is often the most difficult aspect of forensic toxicology. It is different from the situation in clinical toxicology, and largely dependent on the type and quality of specimens provided as well

as on storage and analytical procedures [2, 3]. Interpretation and evaluation may become more difficult in decomposed or embalmed cases. Nevertheless, in these cases, drug degradation or formation of artifacts is more likely to be expected [4, 5].

Artifacts in forensic postmortem toxicology are substances or drug concentrations present in body fluids or tissues during analysis that do not correspond to the genuine drug or drug level present in the body at the time of death. Because of the immediate history of the treatment of the deceased, the postmortem time period as well as the collection, storage, processing, and analysis of postmortem specimens, these artifacts remain an inherent part of postmortem forensic toxicology. For the analyst, postmortem artifacts may be an intricate and demanding challenge. To avoid errors in the evaluation of an analytical result, it is essential to be aware of and to identify artifacts.

This article illustrates various individual artifacts, explains their origin, and discusses techniques to recognize them and minimize their effects.

Antemortem Factors

In death investigation, quantification of a foreign substance is often necessary to state whether the amount of a particular compound is compatible with fatal poisoning or more in accordance with therapeutic concentrations in the underlying case [6]. In living beings, basic pharmacokinetic concepts and models provide estimates of the quantitative relationship between the dose of a drug and the observed blood or tissue concentration. This field is concerned with drug liberation, absorption, distribution, metabolism, and excretion as well as the relationship of these processes to the intensity and time course of the drug's therapeutic and adverse effects. A large pharmacologic variability can be observed in living individuals. Factors likely to contribute to the individual's response to a drug and to influence the disposition of a drug are given in Table 1 [1, 2].

In acute poisoning, death may usually occur before steady state has been reached, and the arterial blood concentration can be appreciably higher than the venous blood concentration. Studies on arteriovenous differences in drug concentrations during lifetime are rare, most comprehensive data for humans are provided by Chiou [7]. For example, arterial plasma concentrations of amitriptyline were found to be up to fourfold greater than venous concentrations during the absorption/distribution phase. With diazepam and lidocaine, the initial arteriovenous differences were of approximately 2 orders of magnitude and lasted for about 60 min after dosing.

Table 1 Factors likely to contribute to pharmacokinetic variability [1, 2]

Different preparations of the drug and clandestine manufacture
Dose, route of administration, and frequency
Pharmacokinetics
Age, gender, race, and genetic disposition
Body weight, exercise, nutrition, condition, and disease state
Genetic variation in drug metabolism
Tolerance
Coadministration of other drugs or alcohol

Disease affects the way drugs are absorbed, distributed, metabolized, and excreted. The liver and the kidneys play a central role in the disposition kinetics of most drugs [8]. Patients with renal disease excrete considerably less unchanged drugs than patients with normal renal function. If a drug is only eliminated by hepatic metabolism, however, its clearance should not be markedly altered. This rather simple theory does not apply to all metabolized drugs. For example, patients with impaired renal function may experience severe and prolonged respiratory depression when treated with morphine. Although morphine's metabolism and excretion is not impaired in renal insufficiency, accumulation of morphine-6-glucuronide occurs. There is evidence that morphine-6-glucuronide is an active metabolite [9]. Sulfonurea drugs, such as glibenclamide, are likely to cause significant hypoglycemia due to accumulation of the active metabolite in renal insufficiency [10]. The effects of liver disease on the pharmacokinetics of drugs are unpredictable due to multiple effects that liver disease produces. In alcoholic cirrhosis, the half-life of diazepam is increased about fourfold over control values [11]. An impaired metabolism and changes in the apparent volume of distribution with chlordiazepoxide have also been observed in severe liver disease, but not with oxazepam and lorazepam, whose elimination involves glucuronidation only [12]. Liver dysfunction also affects hepatic blood flow, which is important in the disposition of drugs with a high hepatic extraction ratio [13]. For example, the coefficient of variation for morphine

serum concentrations during four consecutive days of stable oral morphine treatment at stable clinical symptoms ranged from 13 to 103% [14]. This day-to-day variation has been attributed to fluctuations in hepatic blood flow [15].

Not only is a high degree of variability routinely found between subjects, but a wide range of blood levels may also be seen in the same subject taking a drug on different occasions. The renal clearance of digoxin is significantly lower during a period of immobilization than during a period of normal physical activity leading to an artifactual increase in digoxin concentration [16]. A large pharmacokinetic variability is consistently observed with drugs that are subject to a substantial presystemic metabolism and/or a high hepatic clearance such as desipramine, where plasma levels differed by 30-fold in patients treated with 25 mg desipramine orally three times a day [17].

Alcohol is a common finding in medicolegal death investigations. Many drugs interact with alcohol, thereby altering the mechanism or effects of the alcohol and the drug involved [18]. Dorian *et al.* [19] have shown that the area under the curve of both amitriptyline and nortriptyline was increased in the presence of ethanol because of a reduction in amitriptyline hepatic clearance.

At present, drug–drug interactions are widely recognized, although the underlying mechanisms are still far from being fully evaluated [20]. A great deal of variability, both pharmacokinetic and pharmacodynamic, has been seen with midazolam in patients on mechanical ventilation in an intensive care unit [21]. The half-life of midazolam ranged from less than 2 to about 10 h. In addition, it was difficult to establish a relationship between consciousness and the concentration of the drug. These differences were associated with the state of the patient and the variety of coadministered drugs. Far less is known about drug–nutrient interactions resulting in therapeutic failure or drug toxicity [22]. For example, the activities of cytochrome P450 (CYP) 3A4, CYP2C9, and CYP1A2 may be increased by use of St. John's Wort, and kava may inhibit CYP1A2, CYP2D6, and CYP3A4 [23]. Carbamazepine causes a loss of biotin, which may lead to poor seizure control, and plays a role in the drug's adverse effects [24].

In drug overdose or critically ill persons, pharmacokinetics of a drug are likely to be very different from the common subjects of pharmacokinetic reports, due to diminished cardiac output, falling blood pressure, decreased tissue perfusion, impaired ventilation, tissue hypoxia, acidosis, dehydration, as well as renal and hepatic failure. With few exceptions, there is little information on this point. Major changes in the clearance and volume of distribution of morphine were found in trauma and burned patients, and sepsis may induce a decrease in the hepatic metabolism of drugs [25].

Also, formation of endogenous compounds may be considered. An elevated blood isopropanol concentration, which is usually attributed to alcohol misuse, can be observed following physical exercise. Isopropanol is also detectable under pathophysiological conditions in diabetes mellitus, and in liver and gastrointestinal diseases. In these cases, the blood acetone concentration is also increased [26]. In ketotic states, isopropanol may function as a shunt regenerating NAD^+, thus contributing to the maintenance of metabolic stability.

Not only analytical results may be biased by clandestine manufacture of drugs, which are frequently impure or adulterated, but substances present with these drugs can also be classified as diluents, adulterants, impurities of manufacture, and impurities of origin. Adulteration may occur by the addition of bulking agents or to fake the pharmacological effect. For example, lidocaine is a common adulterant in cocaine preparations [27]. A comprehensive overview on clandestine drug synthesis including pharmacological activities associated with the analogs and impurities is given by Soine [28].

The presence of tolerance will usually make interpretation of results difficult. Tolerance may result from a decreased efficacy at the receptor site or an increased metabolism due to enzyme induction. The presence of a drug analyte in blood or urine can be used to document recent exposure only. By providing information on exposure to drugs over time, hair analysis may be useful [29]. Although there are still controversies, particularly concerning drug incorporation, the influence of cosmetic treatment, and external contamination, long-term information on an individual's drug use may be accessible through hair analysis [30, 31].

Effects of treatment during resuscitation or hospitalization may produce artifacts. In persons who have undergone emergency medical treatment, the medication given by the emergency physician should be provided, for, even without restoration of the heart

action during cardiopulmonary resuscitation, high concentrations of lidocaine in the left cardiac chamber were observed. Obviously, substantial amounts of intubation-related lidocaine had been absorbed by the trachea during cardiac massage [32]. Treatment with intravenous fluids for a period of time before dying often presents interpretative problems in alcohol and drug findings, and blood and tissues may even be devoid of detectable amounts. Devices that automatically deliver medication by the parenteral route can lead to artificially high blood concentrations postmortem [33]. Transdermal patches, e.g., fentanyl containing devices left on the body, will give rise to locally high drug concentrations of the drug [34].

All these factors affect a drug concentration that will be found in the body after death. However, experience has told that there are many additional and even unique aspects of postmortem changes, which are discussed in the following passage.

Changes Occurring after Death

Postmortem Redistribution

Drug concentration is likely to change after death. Early indications that a change does exist are from Curry and Sunshine [35] reporting large differences in the concentrations of barbiturates in blood obtained from different anatomical sites. A comparison of ante- and postmortem drug levels already indicates that postmortem drug concentrations do not necessarily reflect concentrations at the time of death. A study on six cases revealed postmortem blood concentrations as high as, or higher than the antemortem circulating blood concentration [36]. For example, an 11.7-, 3.9-, and 2.6-fold increase could be noticed for dothiepin, amitriptyline, and methadone in postmortem blood. In general, drugs with wide concentration ranges observed in central (c) and peripheral (p) blood exhibiting a high c/p ratio tended to have a high postmortem/antemortem concentration **Postmortem Toxicology: Interpretation**. In contrast, no significant difference was observed in samples taken at admission and autopsy in heroin fatalities [37]. At present, the attempt to estimate antemortem concentrations from postmortem measurements is prone to considerable error [38–42].

Knowledge on the mechanisms causing artifactual increases or decreases in drug concentration during the postmortem period is still limited. These changes are gathered under the generic term of *postmortem redistribution*. The underlying mechanisms of postmortem redistribution as far as known have been reviewed by Pelissier-Alicot *et al.* [43], Yarema and Becker [44], and references cited therein. The physicochemical and pharmacokinetic properties of a drug are probably favoring factors. Organs such as the gastrointestinal tract, the lungs, the liver, and the myocardium are major sources of postmortem redistribution. Cell and tissue modification during agony, autolysis, and putrefaction are also involved.

Potential factors that have been recognized to govern postmortem redistribution are summarized in Table 2. It appears that drugs that have an apparent volume of distribution $>3\,l\,kg^{-1}$, and are sequestered in tissue and present in extracellular fluid are candidates for postmortem redistribution. Accordingly, amitriptyline exhibited significant postmortem distribution, whereas morphine and its glucuronides indicated only a trend for higher concentrations in heart blood compared with femoral or subclavian blood [37, 45, 46].

Many drugs are sequestered antemortem in organs, and postmortem redistribution may either occur by diffusion through blood vessels or from the lumen of a body cavity toward surrounding organs. The vascular pathway may depend on the blood remaining fluid after death. Investigations on postmortem

Table 2 Major factors governing postmortem drug distribution

Physicochemical and pharmacokinetic properties of the drug	Size, shape, charge, pK_a, partitioning behavior, lipophilicity, volume of distribution, binding to proteins and/or red cells, affinity toward tissues, decreasing or residual metabolic activity during the perimortem and early postmortem time period
Environmental conditions	Initial concentration, pH, orientation of solute flux, temperature, time, blood coagulation and hypostasis, blood movement due to pressure and fluidity changes, position of the corpse, lysosomal enzyme activities, and bacterial invasion

diffusion from gastric residues in a human cadaver model using amitriptyline, paracetamol, and lithium carbonate revealed high concentrations in liver and lungs, whereas diffusion into gallbladder bile, cardiac, and aortic blood was less severe [47]. Diffusion of ethanol from the stomach into blood is not a problem in alcohol analysis. For an intact stomach containing 400 ml of 10% ethanol, contamination of a femoral vein sample was minimal [48]. However, it should be considered that regurgitation of alcohol or drugs from the stomach, esophageal sphincter relaxations, and severe blunt trauma resulting in the rupture of internal organs, especially of the stomach, facilitate postmortem diffusion processes. These circumstances may favor artificially elevated blood concentrations [49, 50].

Drugs accumulated in lungs are rapidly released inducing elevated drug concentrations in thoracic and heart blood samples as well as in liver specimens. These changes were assessed in animal and cadaver models, and were, e.g., also seen in methamphetamine-associated deaths [51]. Redistribution from the lungs seems more intense than redistribution from the gastrointestinal tract due to the large surface of the alveoli, the thin membranes, and the high vascularization. The early rise in dothiepin levels in thoracic blood in a rabbit model reflects postmortem redistribution from the lungs, where the drug was heavily concentrated [52]. Fuke *et al.* [53] observed that torso blood samples showed less toluene after gastric instillation than after tracheal instillation. Also, a higher toluene concentration was present in the left lobe of the liver than in the right with gastric instillation. The pleural and peritoneal fluids are regarded as vehicles for drug exchanges between the lungs and the liver [43].

The redistribution effects from the liver are far more complex, and may occur via the hepatic vessels or directly to adjacent organs, such as the stomach or the gall bladder [39]. Postmortem decreases in concentrations of fluoxetine and norfluoxetine, in both liver and lungs, occurred along with increases in blood concentrations in a dog model [54]. In rats, administered amitriptyline in the liver lobes had high but variable drug concentrations among tissues. Lobes lying closest to the stomach had the highest drug concentrations [55]. In general, a great part of the surface of the left liver lobe being in close contact with the stomach wall will be more involved in postmortem redistribution. Also, it may be difficult to correctly assign the source of hepatic concentrations postmortem [43].

A substantial increase in heart blood could be noticed for drugs, such as calcium channel blockers or cardiac glycosides, which are highly bound to cardiac tissue. In a series of digoxin cases, the drug concentration was invariably higher in heart blood specimens than in peripheral blood samples [56, 57]. The lungs, the liver, and the stomach may serve as further drug sources. A useful compilation of drug concentrations in heart and femoral blood has been published by Dalpe-Scott *et al.* [58].

Variations in postmortem drug concentration largely depend on both the time since death and the site of blood collection. Differences in drug concentrations collected from different anatomical sites have been reported for numerous drugs including imipramine, diphenhydramine and codeine [59], methadone [60, 61], doxepin, clomipramine, barbiturates [39], amitriptyline [45], cimetidine [62], methamphetamine [51], cocaine [63], digoxin [56], zopiclone [64], methylenedioxymethamphetamine, and methylenedioxyamphetamine [65]. A review on site-dependent differences is given by Prouty and Anderson [41] and Baselt [38].

It is evident that postmortem redistribution is governed by the postmortem time interval. Compared to studies on site-dependent differences, only few data on time-dependent differences are available. In a fatal case of dihydrocodeine intoxication, site-to-site differences of the parent drug and major metabolites were very small, probably due to steady state, an apparent volume of distribution of $1.0-1.3\,l\,kg^{-1}$ for dihydrocodeine, the fluidity of blood as well as very early postmortem blood sampling [66].

Some case reports have determined that there is little evidence of time-dependent variability, which may be due to delayed sampling. Temporal changes in drug concentrations that have been studied in animal models revealed significant changes to occur already during the early postmortem period [55, 67–70]. The level of dothiepin in cardiac and pulmonary blood samples steadily increased to reach 400% of its original concentration at 8-h postmortem [71]. In a dog model, 2-h postmortem concentrations of fluoxetine and norfluoxetine were 2.2- to 6.0-fold higher than antemortem concentrations, but did not significantly differ from 12-h postmortem concentrations [54]. An overview on drugs in which redistribution is likely to occur or in which postmortem redistribution probably

does not occur is provided by Drummer [3], Leikin and Watson [4], and Baselt [38].

Major Changes of the Media and the Analyte Occurring after Death

General Remarks. As indicated in the section "Postmortem Redistribution", body fluids and tissues as well as drugs present in these specimens are subject to fundamental changes during the postmortem time period. Alterations of the media may considerably impact drug analysis. Decomposition of a corpse involves the processes of autolysis and putrefaction. Enzymes naturally present in the body induce autolytic changes; putrefaction is due to destruction by microorganisms. The onset of autolysis is rapid in cells with high concentrations of hydrolytic enzymes, such as the pancreas and the gastric mucosa, and is slower in the cells of the heart and the liver. Being the specimen of choice for detecting, quantifying, and interpreting drug concentrations, a more detailed review of postmortem changes in blood is given.

Postmortem Changes of Blood with Regard to Drug Analysis. Blood is a complex mixture that contains solubilized proteins, dissolved fats, solids, and suspended cells. Serum or plasma is traditionally used in clinical settings because blood affords advanced handling in the laboratory procedures [6]. Drug concentrations provided in literature are usually determined from these fluids. Analytical results obtained from postmortem blood are compared valuably with levels previously reported in therapeutic and toxic conditions [38, 72–74]. However, separation of red blood cells from postmortem blood is usually not possible, and its composition may remarkably differ from a blood sample obtained from a living person.

Changes may already occur during agony [75]. Hypoxia reduces the intracellular pH value, thus inducing an increased accumulation of basic drugs into the cells. Neutral or acidic drugs are less affected. Intracellular acidification and changes in ionic strength lead to a damage to the lysosomal membrane, and, subsequently, to enzymatic digestion of the cell membrane and components. Drugs that are concentrated in the cell are redistributed at this stage into the extracellular compartment. After death had occurred, there is a rapid progress in postmortem redistribution processes due to disintegration of physiological and anatomical barriers (see the section "Postmortem Redistribution"). Postmortem vascular permeation has been shown in an *in vitro* model using morphine and its glucuronides [76].

In addition to the immediate postmortem dropping of the pH value up to 5.5, a decrease in blood–water content can often be observed [75, 77]. There exist strong variations in the water content of postmortem blood ranging from 59 to 89%. Both hemoconcentration and altered partition behavior affect original drug levels.

As the permeability of all cell membranes increases, hemolysis occurs. In addition to hemolysis seen with most specimens, blood coagulates postmortem, and then becomes liquid again. The effectiveness of these two processes will determine whether postmortem blood is clotted, fluid or partially clotted, and partially fluid [78]. Blood clots distribute unevenly in the body. A few hours after death, hypostasis occurs by sedimentation of blood and serum to the lower parts of the body due to gravitation. As a result, concentration measured for any drug exhibiting unequal distribution between red cells and serum may be biased by blood hypostasis, irregular clotting, and hemolysis.

Average distribution ratios between whole blood and plasma are given for major drugs in Table 3 [61, 79–88]. Further data of blood-to-plasma concentration ratios are provided by Baselt [38] and Iten [89]. For some drugs, varying blood-to-plasma ratios have been observed between individuals. In patients, chlorpromazine erythrocyte concentrations tended to correlate with plasma concentrations, but the erythrocyte/plasma concentration ratio varied from 0.61 to 2.00 among patients [90].

Ratios may not only vary between drugs but also differ between a particular drug and corresponding metabolites. Some caution is advisable using these data. Most of them are derived from *in vitro* partition experiments using systems composed of plasma water, plasma proteins, and erythrocytes. When spiked blood is diluted with autologous plasma water, erythrocytes always discharge the compound overproportionally, compared to plasma proteins [91, 92]. Also, ratios may differ depending on whether the sample had been collected from a living person or a corpse. For example, the plasma-to-whole blood concentration ratios of cannabinoids were found to

Table 3 Blood-to-plasma concentration ratios for some drugs of forensic interest

Drug	Ratio	Reference
Amitriptyline	1.0–1.1	[79]
Nortriptyline	1.5–1.7	
Cocaine	1.00	[80]
Diazepam	0.70	[79]
Oxazepam	1.00	[81]
Ethanol	0.74–0.90	[82]
Methadone	0.75	[83]
	1.00	[61]
Morphine	1.02	[84]
Morphine glucuronides	Dependent on hematocrit	[85]
Δ^9 - Tetrahydrocannabinol	0.55, 0.66	[86, 87]
11-Hydroxy-Δ^9-tetrahydrocannabinol	0.57, 0.58	[86, 87]
11-Nor-9-carboxy-Δ^9-tetrahydrocannabinol	0.62	[88]

be very similar and their individual coefficient of variation to be very low in samples taken from living individuals. However, data obtained postmortem suggest that the distribution of cannabinoids is scattered over a wider range of values compared to those determined in living subjects. Also, cannabinoids favored postmortem "serum", the mean ratio between the blood supernatant and whole blood being 2.4, but only 1.6 in samples collected from living people [88]. Generally, the differences observed between blood or plasma are considered to be less important compared to the changes in concentration that may occur prior to sampling.

Invasion of intestinal flora into tissues and body fluids occurs rapidly after death, especially at ambient or elevated temperatures. Postmortem blood samples taken 6 h after death in patients who had died of causes other than infectious diseases were tested positive for bacteria [93], whereas in a study on heart blood samples collected 85 h after death bacteriologic cultures gave negative results [94]. Microbial enzymes hydrolyze and transform lipids, carbohydrates, and proteins. As a result, the pH value of blood slowly increases again during the postmortem interval [75, 95].

Postmortem Alterations in Drug Concentrations. Degradation as well as formation of drugs during the postmortem interval as competing processes to postmortem redistribution has been observed. Drugs concentration may change due to chemical and physical degradation, metabolic formation or breakdown (Table 4). Problems also arise from interfering substances that are endogenously produced during autolysis and putrefactive processes (see the section "Postmortem Redistribution").

Systematic investigations on the time dependence of detectability of drugs or poisons in a putrefying body do not exist. Some case reports revealed that drugs such as morphine or atropine may be identified in specimens from exhumed corpses or from stored tissues many months after death [96]. Comprehensive data on positive drug findings in putrefied bodies had been published by Arnold *et al.* [97]. For example, phenobarbital, bromazepam, sulpiride, and promethazine could be successfully identified in highly putrefied materials. The interpretative value of these measurements is limited, however.

Recovery of organophosphorous pesticides, such as parathion or malathion, was less successful compared to paraquat, whereas recovery of organochlorine compounds was >72% from putrefactive materials [99, 100]. Stevens [95] studied the stability of 56 drugs and drug-related compounds added to drug-free liver homogenates. Degradation was elevated in samples exposed to fly-borne bacteria. From the results, the following molecular structures are assumed to be prone to putrefactive decomposition: oxygen that is bonded to nitrogen as in nitro groups or *N*-oxides, sulfur, which forms part of a heterocyclic ring, and aminophenol structures. A variable decomposition rate, which was observed for dothiepin in bacteria-contaminated liver and blood specimens, was suggested to be due to differences in bacterial activity [52].

Table 4 Possible mechanisms operating on drugs postmortem, examples modified according to [98]

Mechanism	Example(s)
Chemical instability	
Hydrolysis	Heroin, cocaine, *O*-acyl-, and *N*-glucuronides
Oxidation	Sulfur-containing drugs and morphine
Metabolic instability	
Esterases (endogenous)	Hydrolysis of ester-type drugs Hydrolysis of phase-II metabolites Reduction, e.g., of nitrobenzodiazepines Oxidation, e.g., of thioridazine
Metabolic production	Ethanol, γ-Hydroxybutyrate, carbon monoxide, and cyanide

The possible role of enteric bacteria in the bioconversion of nitrobenzodiazepines was studied in detail by Robertson and Drummer [101]. It is well known, that in deaths involving flunitrazepam, considerably higher concentrations of 7-aminoflunitrazepam than that of the parent compound can be detected in blood [102]. The conversion of the respective 7-aminometabolites of nitrobenzodiazepines by individual bacteria in blood was species dependent. Significantly higher rates were found for obligate anaerobic species than for facultative anaerobic species, and there was little difference among the species for their ability to metabolize nitrazepam, flunitrazepam, or clonazepam. The conversion rate was slowed down by keeping the corpses at 4 °C. The effect of pH variation on the metabolic activity of different bacterial species was variable.

Postmortem degradation does also concern acid-labile conjugates, such as the ester glucuronides of 11-nor-9-carboxy-Δ^9-tetrahydrocannabinol, propofol, or diflunisal [103, 104] as well as the more stable ether glucuronides [105]. Conversion of morphine glucuronides to free morphine by residual glucuronidase activity or some bacterial enzymes is a most prominent example [106]. The bacteria most likely involved are originating from the gastrointestinal tract. Being the most prominent among them, *Escherichia coli* is an important source of β-glucuronidase activity. Preferential hydrolysis of morphine-3-glucuronide to free morphine by bacterial enzymatic activity was ascertained in *in vitro* experiments, and was shown to depend on storage time, temperature, and initial degree of putrefaction [107].

Bacteria, yeast, and fungi can also produce some compounds in the postmortem blood, the most prominent representative being ethanol. In contrast to drug metabolism, which may persist some time after death, the physiological metabolism of ethanol is assumed to cease at the time of death [108]. A considerable site-dependent variation of the ethanol concentration in blood samples had been observed even in cases where signs of putrefaction could not be noticed [109]. These differences were mainly attributed to death occurring during the absorptive phase where differences between arterial and venous blood exist. Ethanol absorbed from the gastrointestinal tract distributes throughout the body according to the water content of the corresponding tissue or body fluid. The postmortem change in blood alcohol content closely following the change in blood water content, correction for water content has been recommended [110]. In principle, the water content of blood decreases with the time after death. However, there are some exceptions, e.g., drowning experiments with animals indicated a dilution of alcohol in blood [111].

Postmortem changes of ethanol are well documented [112]. The majority of the cases attributed to neoformation did not have significant ethanol concentrations (<0.07%). However, there are a few case reports demonstrating that a significant amount of ethanol up to 0.22% may be produced [113, 114]. The amount of ethanol formed during the postmortem time interval depends on the antemortem conditions, the species of microorganisms present, the availability of substrates, and the storage conditions of the body prior to collection of samples for toxicological analysis **Alcohol: Analysis**. In bodies that were stored refrigerated, even in the presence of microbial species capable to produce ethanol, alcohol formation

could not be established within 4–24 h after death [113, 115]. Comprehensive reviews on the variety of organisms capable of producing ethanol and on the microbiology and biochemistry of postmortem ethanol formation are given by Corry and Huckenbeck [116, 117]. Potential substrates are carbohydrates, glucose, lactate, ribose, and amino acids, and glycolysis is thought to be the primary process for ethanol production [116, 118].

For postmortem synthesis of ethanol is difficult to be accurately established, further putrefactive products have been suggested as indicators to differentiate postmortem-formed ethanol from antemortem-ingested ethanol. Along with ethanol, other short chain alcohols, such as 1-propanol, isopropanol, 1-butanol, 2-butanol, isobutanol, and isoamylalcohol, are produced by microorganisms. Methanol is considered not to arise from microbial synthesis. In case reports, *in vitro* and animal studies, most commonly 1-propanol, could be detected [113]. Estimation of the ethanol synthesis based on the 1-propanol or other short chain alcohol levels seems questionable due to their highly variable formation rates as well as for they may derive from ingested alcoholic beverages. Huckenbeck [117] found that different Clostridium and Proteus species produced α- and γ-aminobutyric acid and δ-aminovaleric acid along with ethanol, but a quantitative relationship between postmortem ethanol and putrefactive amino acid production could not be established.

Postmortem cyanide production could give improper estimation of cyanide poisoning from blood analysis. Seto [119] suggested that cyanide is released from methemoglobin cyanide complex by heat denaturation, and diffuses out of the blood vessels via plasma during which it encounters albumin. Also, liberation by superoxide anion radical oxidation of proteins has been assumed. However, these pathways have been judged to be of minor importance compared to cyanide production during storage or as an analytical artifact.

γ-Hydroxybutyrate (GHB) has been widely abused. It is also a naturally occurring compound present in the mammalian central nervous system. Small amounts are formed from γ-aminobutyric acid with subsequent reduction to form GHB again. Besides its premortem production, GHB can be produced as a postmortem artifact [120]. When GHB concentrations are low at a postmortem time interval of more than a few hours, they may be falsely implicated as a cause of death [121].

Artifacts in Specimens Stored in Formaldehyde or Collected after Embalmment

Sometimes it is necessary to perform analysis on pathologic specimens stored in formaldehyde or on samples that have been collected from an embalmed body. The embalming procedure may have diluted the blood, and/or may have partially or completely removed drugs or poisons present at the time of death from major blood vessels [122]. Embalming fluids are typically based on formaldehyde with final concentrations of 5–20% at acidic or neutral pH values, and may usually also contain alcohols [123]. Formaldehyde is a highly reactive chemical agent, and may mask, alter, or destroy a drug during the fixation or embalming process. Previous studies have demonstrated that formalin can react with various drugs including, e.g., amphetamines [124], barbiturates [125], and benzodiazepines [126] in a time, pH value, and formalin concentration-dependent manner. Most likely, reaction pathways are through hydrolysis and/or methylation via the Eschweiler–Clarke reaction. In addition, body fluids and tissues are difficult to extract due to denaturation.

Results from liver tissue specimens collected from drug-related suicides indicated that some methylation of nortriptyline to amitriptyline had occurred during formalin fixation and storage, but consistent ratios could not be established [122]. In a different study, conversion of nortriptyline to amitriptyline was confirmed [127]. For the formation of amitriptyline was only proportional to the loss of nortriptyline at a pH value of 9.5, it was suggested that the metabolite may be susceptible to degradation in formaldehyde containing solutions at lower pH values. *N*-methylation has also been reported for amphetamine, methamphetamine, and fenfluramine. Disappearance of fenfluramine was higher at elevated formaldehyde concentrations, and an increase in pH directly correlated with the appearance of *N*-methyl fenfluramine [128].

Some drugs appeared to be reasonably stable in formalin. In a study examining chemically fixed tissue specimens of poisoned rabbits, a formalin solution containing 10% formaldehyde (pH 7.4) was found to be most suitable for diazepam- and

chlorpromazine-containing tissues for at least 28 days with respect to the recovery of the analytes [129]. Succinylcholine only slowly degraded in embalmed rat tissue samples [130]. It appears that the embalming fluid maintaining acidic conditions contributed to the drug's stability in fixed tissue, whereas succinylcholine is very rapidly hydrolyzed in blood. However, some compounds, such as alprazolam and midazolam, decompose more rapidly under acidic conditions [126].

Stability of volatile substances in formalin-fixed tissues was tested for ethanol, diethyl ether, chloroform, and toluene in a rabbit model by the intravenous route [131]. All volatile compounds were still detectable after a 14-day fixation period. Compared to nonfixed materials, in the fixed tissue samples concentrations had decreased in the following order: ethanol > diethyl ether >> chloroform > toluene.

Leaching of drugs into the fixing/storing solution is evident. In a sildenafil-related death, a comparison of the quantitative values of sildenafil in fixed tissues and those in the same tissues at autopsy revealed a mean decrease of 74%. However, the total recovery from both the fixed tissue and the particular formalin solution was 95% with regard to the original quantity in the same tissue before fixation [132]. Similar observation was reported on the detection and quantification of morphine and strychnine in fixed samples and formalin solutions [133, 134].

Analysis of fixed samples may create problems with regard to the isolation of the analyte and damage of the technical equipment. Iffland *et al.* [135] succeeded in determining carbon monoxide in clots of heart blood collected from an embalmed body following release of carbon monoxide from the sample by nitric acid. Phenobarbital was detected in fixed brain tissue as well as in the formalin solution in a poisoning case using ultraviolet (UV) spectrometry and thin layer chromatography, whereas gas chromatography (GC) was not applicable due to interferences and damage of the column [136].

Acquisition of Specimens

General Considerations

The purpose of sampling is to provide a representative part of the whole that is suitable for analysis and reliable interpretation. Sampling is the most important step in drug analysis because an analytical result will never be better than the sample from which it is derived. Specimens available in postmortem toxicology investigations can be numerous and variable, and may be selected on the basis of the case history, requests, legal aspects, and availability in a given case. So far, a harmonized protocol for sampling in suspected poisoning or drug-related death has not been established [49, 137]. Generally, the specimens routinely collected at autopsy include fluids, such as blood from peripheral sites and heart blood, urine, bile, cerebrospinal fluid, vitreous humor and gastric contents, and organs, particularly liver [49, 138]. The main collection artifact is contamination, but can be reduced by sampling before the autopsy, if appropriate. It is difficult, if not impossible, to acquire quality specimens once autopsy has been completed. An appreciation of how contaminants may be introduced is also important.

Sampling Artifacts

Incorrect Selection and Acquisition of Samples. Samples taken for analysis should always be chosen bearing in mind the disposition of the drug in the body. An incorrect or insufficient sampling will severely affect case investigation **Toxicology: Analysis**. In fatalities requiring quantitative determination, a blood specimen is preferably taken from the femoral vein prior to autopsy, for contamination may be avoided, and this site is less affected by postmortem changes (see the section "Changes Occurring after Death"). If the femoral vein is ligated prior to sampling, the sample is likely to be relatively uncontaminated by blood from the major organs [40]. Cardiac blood is regarded as unsuitable for quantitative analysis of drugs. Diffusion out of the stomach can artificially raise the cardiac blood concentration, and a sample may equally contain blood that has drained from the lungs, the vena cava inferior, the aorta, and the subclavian veins. If a heart blood specimen is sampled through the chest wall, one should be aware that the sample is contaminated with thoracic fluid and gastric contents. Postmortem specimens obtained from patients who have died several days after a drug-related episode are likely to give negative results [2]. It is essential to perform toxicology investigations on specimens obtained on or soon after admission to hospital. In fire victims, blood that is not collected from body regions excluded from severe burning can contain falsely elevated carbon monoxide levels due

to diffusion of environmental carbon monoxide and binding to hemoglobin. Relying on a result from a single specimen may be misleading. It is recommended to collect blood from at least two different sites or, if not available, along with other specimens, at least.

Urine is a valuable specimen, because it can easily be tested, and drugs and drug metabolites are usually found in high concentrations **Drug Testing: Urine**. A sample collected during autopsy may be contaminated by blood and should preferably be taken prior to autopsy by puncture of the abdominal wall. A positive identification indicates recent drug use, but does not indicate when or how much drug was ingested. There is little correlation between urine and blood concentrations. If death takes place quickly, urine findings can be negative. Analysis of a specimen may not be a reliable means to reveal ingestion of drugs, e.g., sertraline and norsertraline, which are present in very low concentrations or were not renally cleared [38]. Analysis of a urine specimen may reveal exposure to organophosphate compounds or analysis of toluene, xylene, and trichloroethylene via identification of their major metabolites in cases where blood analysis will fail to detect these compounds [38]. Local anesthetics, e.g., lidocaine, which are used on catheters, are a common finding in urine samples.

Gastric contents are a useful specimen to rapidly discover drug overdose, for oral ingestion is the major administration route of prescribed drugs. The total amount of a drug or poison remaining in the gastric contents is far more important than its concentration. A low absolute amount does not rule out the possibility of an overdose. Also, a low drug concentration in the stomach may arise from passive diffusion from the blood into stomach contents. This phenomenon is frequently observed in drugs being weakly basic in nature [33]. Since the gastric contents can be largely inhomogeneous, the entire specimen should be submitted or mixed before an aliquot is taken. The odor of gastric contents or colored material can potentially point to a specific agent, although, e.g., blue stains may result from parathion or flunitrazepam ingestion.

Bile represents a collection and storage depot for many xenobiotics and corresponding metabolites that have a biliary excretion and are subject to enterohepatic cycling. To avoid contamination of surrounding tissues, the gall bladder should be tied off before removing from the liver. Drug concentrations can be significantly higher in bile than in blood. The mean bile to blood ratios varied from about 1 for acetaminophen and amphetamine to about 2000 for desmethylclobazam. In several cases, a drug could be identified in bile, but was not detectable in blood [139]. A qualitative finding in bile may indicate previous or chronic exposure to a drug or poison.

Cerebrospinal fluid and vitreous humor are aqueous and transparent fluids, which are useful to screen for a variety of drugs [140]. Both cerebrospinal fluid and vitreous humor also contain very little proteins. Therefore, drugs highly bound to proteins or lipophilic in nature tend to be found in lower concentrations in these fluids than in blood [49, 141]. A major limitation to the use of cerebrospinal fluids is the small database of reference values [142], whereas vitreous humor has been used to analyze a larger number of drugs [143–150]. Many studies have stressed the usefulness of vitreous humor for alcohol analysis [151]. However, the wide variation of vitreous humor to blood ethanol ratios must alert when results from vitreous humor are used to estimate the concentration in femoral venous blood [152].

Tissue specimens collected for postmortem toxicology investigations include liver, kidney, lung, brain, skeletal muscle, and adipose tissue [49, 151]. Tissue samples may be useful in cases with an extended postmortem time period and whenever body fluids are not available. Extensive data had been published for liver and kidney, less for brain and lung specimens [38, 153]. Drug concentrations in liver were found to be site dependent (see the section "Postmortem Redistribution"), and sampling from deep within the lobe has been recommended [154]. Concentrations in brain may also significantly vary from one region to another [155]. The within-case variability of drug levels observed in muscle specimens supports the opinion, that drug analysis on skeletal muscle is rather qualitative than quantitative in nature. Muscle specimens had also been considered for alcohol analysis, but the muscle to blood–ethanol concentration was found to depend on the time course of ethanol absorption, distribution, and elimination [156]. A skin specimen or a cube of muscle may support evidence of the route of drug administration [49]. As skin acts as a temporary drug reservoir, the specimen should always be excised together with a

random specimen preferably taken from a similar site to act as control [157].

Hair is an ideal specimen for determining, e.g., chronic arsenic and mercury poisoning [158]. Also, numerous drugs and poisons have been detected in hair in recent years (*see* **Hair: Toxicology**) [159, 160]. The amounts deposited in hair are functions of both ingestion/exposure and of the metabolic regimen. Segmentation of the hair can assist in estimating the time of exposure. A decrease in drug concentration in the proximal sections of hair may indicate a decrease in tolerance to the drug. There are many factors that influence drug concentrations in hair: biological factors such as hair structure and pigmentation, individual factors such as drug use, customs culture or race including hair care, environmental factors, and also methodological factors [30]. External uptake of drugs from blood, vomit, or putrefactive fluids leads to artificially elevated drug levels in hair, which will not be fully removed by common wash procedures [31].

Specimen Containers and Preservation. The use of appropriate specimen containers and preservatives can be critical with regard to ultimately identify a substance in an individual specimen. Specimens must be collected in separate, clean containers, which should be filled up to minimize evaporation of volatiles and oxidative losses of drugs.

The best materials to collect and store fluids or tissue specimens are glass containers. Sampling into a glass container is a must, if solvent abuse or an anesthetic death is suspected. Essential are also aluminum foil- or Teflon-lined lids to prevent gas escaping and to minimize drug adsorption. Solid tissue samples may also be placed in nylon bags, which are tightly sealed [138]. Most types of plastic containers are suitable for the collection of tissue specimens in drug-related fatalities. Disposable hard plastic tubes or Nalgene® bottles with screw caps are also recommended for collection of body fluids for breakage of these containers upon freezing had not been observed [161]. The use of evacuated tubes is less desirable, for sample contamination from plasticizers used in their manufacture may occur. For collection tubes containing gel separators, a gross contamination by toluene, 1-butanol, ethyl benzene, and xylene has been reported [162]. Evaluating the container before routinely collecting specimens in it might reduce production of artifacts.

Obligatory recommendations for specimen preservations do not exist. Specimen preservatives are generally not required for specimens other than blood. In addition to a preserved blood sample, an unpreserved specimen available to the toxicologist is optimal, for preservation strategies depend on the target analytes and are not unique. Fluoride preservation with a final concentration of 1–5% sodium fluoride by weight is recommended for postmortem analyses of alcohol, cocaine, cyanide, and carbon monoxide. Postmortem synthesis of ethanol can be effectively inhibited, whereas hydrolysis of cocaine can only be slowed down by fluoride preservation [163]. A general problem with ester-type drugs is the presence of esterases; as degradation takes place even at 4 °C, immediate freezing of specimens is recommended [164]. Artificial production of GHB has been observed in blood samples not collected in fluoride-containing tubes [165], and also in specimens collected in citrate buffer [121]. Fluoride preservation must not be used when organophosphorous chemicals are involved. For example, rapid degradation of metrifonate (dichlorvos) was found to be favored by the presence of sodium fluoride, esterases, elevated temperatures, and alkaline condition [166]. Early acidification of the specimen and storage at −80 °C was recommended by Heinig *et al.* [167].

Acidification may also stabilize cocaine or labile conjugates such as *N*-glycosides [5, 168]. Ascorbic acid may be used as an antioxidant. For example, losses in olanzapine during storage at 4 °C may be reduced by the addition of 0.25% ascorbic acid [169]. Apomorphine, like most catechols, is prone to oxidation to quinones, unless ascorbic acid is added as an oxidant [170]. However, the presence of an antioxidant may have reverse effects. During storage reduction of the *N*-oxide metabolites of chlorpromazine, of samples containing antioxidants, resulting in an increase in the concentration of the parent drug, has been observed [171]. Anticoagulants are not recommended for postmortem blood samples because these additives may also affect drug concentration. Blood concentration of morphine in ethylene diamine tetraacetic acid (EDTA) tubes was 4.8% higher than in heparin tubes [172], which also contain phenolic preservatives such as cresol [171].

Stability During Storage

In clinical chemistry, stability is defined as the capability of sample material to retain the initial value of a measured quantity for a defined period within specific limits when stored under defined conditions. The individual maximal permissible instability is preferably linked to the criteria of analytical imprecision, and is expressed as the critical difference [173]. This procedure allows defining the maximum permissible storage time for an analyte in a particular specimen at a given condition.

In postmortem investigations, such consideration starts at the time of sampling and covers the time until analysis. The presence and extent of alterations since the time of death can only be estimated, part of them even cannot be avoided, and they cannot be undone. Therefore, knowledge on degradation mechanisms in a particular matrix and on resulting breakdown products is important.

A review on poisons, drugs, and heavy metals suspected to be unstable is given by Leikin [4] and Ellenhorn [174]. Unfortunately, the particular biological matrix, which may play a role in the stability of an analyte, has not been considered. For example, in urine, cocaine is chemically stable at a pH of less than 7.0. In blood samples, even acidic conditions do not prevent cocaine to be metabolized by residual esterase activities. Interestingly, cocaethylene seemed to be more stable in postmortem specimens than cocaine. Muscle as well as brain tissues were considered to be the specimen of choice for testing both cocaethylene and cocaine [168]. Valuable information on artifacts is provided by Baselt [38] and Drummer [175]. Comprehensive data on the stability of drugs of abuse in blood were reported by Levine and Smith [176] already in 1989. A recent update was given by Skopp and Pötsch [177]. Data on the stability of drugs in tissues are rare.

In principle, all volatile compounds such as aerosol propellants, anesthetic gases, carbon monoxide, ethanol, and organic solvents are unstable during storage. Losses of up to 25% of blood toluene have been observed in glass tubes stored unopened for 7 days at room temperature, and considerably higher losses were noted in glass tubes with rubber stoppers [178]. There is an artifactual rise in carbon monoxide level in unpreserved blood specimens since bacterial action can result in both the production of carbon monoxide and the denaturation of hemoglobin [38]. The formation of toxicologically significant concentrations of cyanide in postmortem tissue has been demonstrated, which was attributed, in part, to conversion of thiocyanate to cyanide and breakdown of proteins [179, 180]. Conversely, a significant decrease in blood cyanide concentration has been attributed to mechanisms that include evaporation, thiocyanate formation, and reaction with specimen components [181]. Temperature and cyanide concentration are apparently important factors in these changes [182].

Another important consideration when measuring drugs is the stability of corresponding metabolites. Highly labile metabolites such as sulfate conjugates and *N*- and acylglucuronides may rapidly be converted back to the unconjugated compound and consequently result in falsely elevated concentrations. Examples are the *N*-sulfate metabolite of minoxidil, *N*-glucuronide metabolites of nomifesine, and 11-nor-9-carboxy-Δ^9-tetrahydrocannabinol glucuronide [103].

Some of the degradation mechanisms seen during storage are similar to those observed during autolysis and putrefaction (see the section "Major Changes of the Media and the Analyte Occurring after Death"). Generally, degradation of a drug occurs through hydrolysis, oxidation or reduction processes, and is generally slowed down by decreasing storage temperatures and preservation of the sample (see the section "Specimen Containers and Preservation"). These processes are due to endogenous enzyme activities, e.g., esterases still operating in the sample, chemical reactions or to enzyme activities such as glucuronidase following bacterial invasion during the postmortem interval (see also Table 4) [98].

Experimental investigations on time-dependent changes give information on the reaction type involved in drug degradation [183], and may guide to a more proper estimation of the drug level at the time of sampling. Hydrolysis of ester-type drugs generally exhibited an apparent first-order reaction kinetic, whereas an oxidation can often be described by a second-order reaction kinetic. Investigations on the reaction type involved in the degradation of forensic relevant drugs have already been performed, e.g., for morphine, morphine glucuronides, cocaine, benzoylecgonine, ecgonine methyl ester, lysergic acid diethylamide (LSD), and 11-nor-9-carboxy-Δ^9-tetrahydrocannabinol glucuronide [103, 105, 163, 184].

The mechanisms involved in the breakdown of benzodiazepines are poorly understood. Hydrolysis and reduction are suggested to be involved in their degradation. Degradation of nitrobenzodiazepines occurs very rapidly, whereas other benzodiazepines do not appear to be as unstable [101, 185]. Besides bacteria, major influence factors in the degradation of nitrobenzodiazepines are an increased temperature and the absence of sodium fluoride [186]. Chlordiazepoxide tends to form desoxychlordiazepoxide during storage, and further degrades to nordiazepam, which also represents a metabolite and an artifact in the analysis of the parent drug by GC [187].

Terbutaline was shown to be stable in spiked postmortem blood at room temperature for 7 days. In contrast, a loss of 83% was observed for fenoterol in spiked postmortem blood at the same conditions. Only 7% of the initial concentration was present after 6 months at 4 °C. The instability of fenoterol is most likely a result of the presence of the phenolic group attached to the side-chain nitrogen atom, which is susceptible to oxidation [188].

There is evidence that different degradation mechanisms operate in blood depending on its source, either obtained from living individuals or collected from corpses. An experimental investigation on morphine and its glucuronides in spiked fresh blood and plasma revealed that oxidation primarily affected drug stability, whereas in postmortem samples stored under the same conditions, hydrolysis of morphine glucuronides was assumed to be the predominant reaction [105].

Sometimes, metabolites or breakdown products are far more stable than the parent drug. Complete degradation of furazolidone occurred in muscle tissue stored at 4 °C during 24 h. Even storage in liquid nitrogen did not fully stabilize furazolidone [189] indicating that analysis of a metabolite, which is reasonably stable, may be favored [190]. The poor stability of cocaine, benzoylecgonine, and ecgonine methyl ester is well documented [163, 183, 191]. Alternate analysis for ecgonine representing a rather stable breakdown product may be performed in highly putrefied specimens or specimens stored for long periods of time.

Most drugs or poisons are probably stable in biological materials for months, particularly if frozen and special arrangements may prevent loss of an analyte by degradation (see the section "Specimen Containers and Preservation").

Analytical Artifacts

Although the history of a sample appears to be most relevant to the production of artifacts, analytical artifacts may also be considered during the isolation and identification of an analyte. For most drugs and poisons, a two-stage testing is usually employed, comprising a preliminary screening test followed by confirmatory analysis, which should offer a higher degree of specificity for the analyte than the first test. A gross overview on common methods used in postmortem toxicology investigations has been given by Hearn and Walls [1] as well as by Drummer [3]. There is little to differentiate in analytical procedures used in other forms of forensic toxicology with respect to postmortem toxicology investigations.

Available immunoassays can yield positive results in urine for metabolites from most of the common benzodiazepines, except lorazepam and flunitrazepam [1]. These drugs may not be detected until hydrolyzed [192]. Interaction with putrefactive amines is commonly seen in immunoassays for amphetamine-type drugs [193]. Interactions are not limited to matrix effects or the structural and conformational similarity of compounds. Further information on cross-reactivity and potential mechanisms is given by Richardson [2]. In addition, turbid, highly colored, or opaque specimens can interfere with the detection principle affording an intensive preextraction step.

Chromatography has been the mainstay of drug analysis for many years. Often, a special pretreatment or homogenization according to the specimen's nature and/or a more sophisticated cleanup extraction of putrefied materials is required for all forms of chromatography. Extraction into an organic solvent for partial purification of a biological fluid or a tissue homogenate is still widely used. Lipophilic compounds will be readily extracted by nonpolar solvents. The more polar solvents being partially miscible with water will also remove water-soluble materials. As a result, drug conjugates are also transferred into the organic phase. If phase-II metabolites are subsequently hydrolyzed, this will lead to erroneously high levels of the parent drug [103, 194]. The phenomenon of conjugate instability is most common with labile conjugates such as *N*- or *O*-acyl-glucuronides [103, 104].

The chemical properties of a solvent can give rise to chemical reactions of the solvent itself and the insidious breakdown products or stabilizers it

may contain. For example, the effect of phosgene in chloroform has been reported as a cause of artifactual formation of carbamate derivatives during extraction of tricyclic antidepressants [195].

The problems associated with solvent extraction can be minimized by using solid-phase extraction. Because of its far greater variability, one must be well informed on the mechanisms of interaction to maintain proper control on the separation. If blood cells are not disrupted or if particles are still present in the sample, flow rates and reproducibility are altered.

The extraction efficiency of a drug or metabolite from a postmortem specimen may be variable from case to case, or even from site to site within the same corpse. It may also be markedly different than from a particular blank specimen used for calibration. The use of stable isotope internal standards may provide a higher degree in the accuracy of the analytical results. Unfortunately, few deuterated standards are commercially available for drugs, metabolites, or artifacts.

It is often necessary to evaporate extracts to reconstitute them into small volumes for transfer to chromatography. The most common routes of sample loss are adsorption onto glassware and volatilization of the drugs, e.g., of amphetamines. Adsorption losses can be prevented, e.g., by silanization of glassware or by including a polar solvent, such as amyl alcohols, as an additive to the extracting solvent or prior to evaporation. The problem of amphetamines' volatilization can be solved by converting them to their nonvolatile hydrochloride salts [6].

Gas chromatography coupled to mass spectrometry (GC/MS) is generally accepted as unequivocal identification for most drugs, and provides best confirmatory information. Also, liquid chromatography coupled to mass spectrometry(LC/MS) is an emerging technique [3]. There are several drawbacks with GC/MS analyses; some of them are included in Table 5.

It is now recognized that determination of major metabolites and degradation products along with the parent drug is essential to avoid misinterpretation of the data due to artifacts [196]. Such a stability-indicating assay is one that can accurately and selectively differentiate the intact drug from its potential decomposition products.

Table 5 Potential problems with GC/MS analyses[(a)]

Erroneous identification at a low analyte concentration
Misidentification due to interfering substances
Inadequate information due to similar (barbiturates) fragmentation behavior
Production of low mass fragment ions only (tricyclic antidepressants)

[(a)] Reproduced from Ref. 1. © Taylor & Francis Group, 1998

Conclusions

Artifacts must be accepted as an integral part of postmortem toxicology, frequently interfering with a straightforward interpretation of the analytical results.

Reporting all the details of the scene investigation and terminal events as well as of the social and medical history will aid to recognize important issues that may occur during the antemortem phase with respect to postmortem findings.

There is general agreement that drug concentrations are site dependent and tend to change with time, most significant alterations occurring already perimortem or rapidly after death. Heart blood concentrations are often higher than those of peripheral specimens. However, there does not appear to be a way to predict the relationship between drug levels derived from various specimens nor the manner in which the concentrations may change with time. To circumvent the problem of postmortem redistribution, it is recommended that blood be sampled from a peripheral vessel along with at least a second specimen taken from a different site, liver from deep within the right lobe and lung rather from the apex than the base. Tissue samples can be of value to assess the significance of a drug in the death of an individual provided that a sufficiently large database has been established.

Tables of therapeutic, toxic, or fatal ranges or correcting for blood-to-plasma ratio cannot be applied without restrictions. As the exact mechanisms of postmortem redistribution are not fully understood and most likely a combination of several factors, exact estimates of the dose are unreliable.

Postmortem degradation as well as formation of compounds have been observed, and may vary between tissues and body fluids as well as from drug to drug. For example, ethanol may be formed in postmortem blood in variable and nonpredictable amounts. In severely putrefied, embalmed, or formalin-fixed tissue, fundamental changes of both the matrix and the drug should be considered. Complete degradation or even removal of a drug may have occurred.

Sampling poses a high risk of contamination and/or production of artifacts. Correct sampling prior to or during autopsy is as essential as is the use of appropriate containers and preservatives including deep freezing. Some drugs or metabolites may undergo further decomposition during storage for several months.

Changes in materials and target analytes often require modifications of routinely applied analytical procedures; the most valuable appears to be a stability-indicating assay.

A thorough cooperation of the pathologist and toxicologist will enable to handle some of the problems addressed above. Some postmortem artifacts will never be resolved, whereas others are amenable to further elucidation from both case reports and experimental studies.

References

[1] Hearn, W.L. & Walls, H.C. (1998). Introduction to post-mortem toxicology, in *Drug Abuse Handbook*, S.B. Karch, ed, CRC Press, Boca Raton, pp. 863–873.

[2] Richardson, T. (2000). Pitfalls in forensic toxicology, *Annals of Clinical Biochemistry* **37**, 20–44.

[3] Drummer, O.H. (2004). Postmortem toxicology of drugs of abuse, *Forensic Science International* **142**, 101–113.

[4] Leikin, J.B. & Watson, W.A. (2003). Post-mortem toxicology: what the dead can and cannot tell us, *Journal of Toxicology Clinical Toxicology* **41**, 47–56.

[5] Skopp, G. (2004). Preanalytic aspects in postmortem toxicology, *Forensic Science International* **142**, 75–100.

[6] Chamberlain, J. (1995). *Drugs in Biological Fluids*, 2nd Edition, CRC Press, Boca Raton.

[7] Chiou, W.L. (1989). The phenomenon and rationale of marked dependence of drug concentration on blood sampling site. Implications in pharmacokinetics, pharmacodynamics, toxicology and therapeutics (Part I), *Clinical Pharmacokinetics* **17**, 175–199.

[8] Gibaldi, M. (1991). *Biopharmaceutics and Clinical Pharmacokinetics*, 4th Edition, Lea & Febiger, Philadelphia.

[9] Angst, M.S., Buhrer, M. & Lotsch, J. (2000). Insidious intoxication after morphine treatment in renal failure: delayed onset, *Anesthesiology* **92**, 1473–1476.

[10] Harrower, A.D. (1991). Pharmacokinetics of oral antihyperglycemic agents in patients with renal failure, *Clinical Pharmacokinetics* **31**, 111–119.

[11] Klotz, U., Avant, G.R., Hoyumpa, A., Schenker, S. & Wilkinson, G.R. (1975). The effects of age and liver disease on the disposition and elimination of diazepam in adult man, *The Journal of Clinical Investigation* **55**, 347–359.

[12] Wilkinson, G.T. (1978). The effects of liver disease and aging on the disposition of diazepam, chlordiazepoxide, oxazepam and lorazepam in man, *Acta Psychiatrica Scandinavica Supplementum* **274**, 56.

[13] Williams, R.L. & Mamelok, R.D. (1980). Hepatic disease and drug pharmacokinetics, *Clinical Pharmacokinetics* **5**, 528–547.

[14] Klepstad, P., Hilton, P., Moen, J., Kaasa, S., Borchgrevink, P.C., Zahlsen, K. & Dale, O. (2004). Day-to-day variations during clinical drug monitoring of morphine, morphine-3-glucuronide and morphine-6-glucuronide serum concentrations in cancer patients. A prospective observational study, *BMC Clinical Pharmacology* **4**, 4–7.

[15] Vermeire, A., Remon, J.P., Rosseel, M.T., Belpaire, F., Devulder, J. & Bogaert, M.G. (1998). Variability of morphine disposition during long-term subcutaneous infusion in terminally ill cancer patients, *European Journal of Clinical Pharmacology* **53**, 325–330.

[16] Wagner, J.G. (1988). Inter- and intrasubject variation of digoxin renal clearance in normal adult males, *Drug Intelligence and Clinical Pharmacy* **22**, 562–567.

[17] Hammer, W., Idestrom, C.M. & Sjöqvist, F. (1967). Chemical control of antidepressant drug therapy, in *Proceedings of the 1st International Symposium on Antidepressant Drugs*, S. Garratini & M.N.G. Dukes, eds, Exerpta Medica, pp. 301–310.

[18] Tanaka, E. (2003). Toxicological interactions involving psychiatric drugs and alcohol: an update, *Journal of Clinical Pharmacy and Therapeutics* **28**, 81–95.

[19] Dorian, P., Sellers, E.M., Reed, K.L., Warsh, J.J., Hamilton, C., Kaplan, H.L. & Fan, T. (1983). Amitriptyline and ethanol: pharmacokinetic and pharmacodynamic interaction, *European Journal of Clinical Pharmacology* **25**, 325–331.

[20] Olkkola, K.T. & Ahonen, J. (2001). Drug interactions, *Current Opinion in Anaesthesiology* **14**, 411–416.

[21] Oldenhof, H., de Jong, M., Steenhoek, A. & Janknegt, R. (1988). Clinical pharmacokinetics of midazolam in intensive care patients, a wide interpatient variability? *Clinical Pharmacology and Therapeutics* **43**, 263–269.

[22] Santos, C.A. & Boullata, J.I. (2005). An approach to evaluating drug-nutrient interactions, *Pharmacotherapy* **25**, 1789–1800.

[23] Boullata, J.I. (2005). Natural health product interactions with medication, *Nutrition in Clinical Practice* **20**, 33–51.

[24] Mock, D.M. & Dyken, M.E. (1997). Biotin catabolism is accelerated in adults receiving long-term therapy with anticonvulsants, *Neurology* **49**, 1444–1447.

[25] Berkenstadt, H., Segal, E., Mayan, H., Almog, S., Rotenberg, M., Perel, A. & Ezra, D. (1999). The pharmacokinetics of morphine and lidocaine in critically ill patients, *Intensive Care Medicine* **25**, 110–112.

[26] Kalapos, M.P. (2003). On the mammalian acetone metabolism: from chemistry to clinical implications, *Biochimica et Biophysica Acta* **1621**, 122–139.

[27] Moriya, F. & Hashimoto, Y. (2000). Determining the state of the deceased during cardiopulmonary resuscitation from tissue distribution patterns of intubation related lidocaine, *Journal of Forensic Sciences* **45**, 846–849.

[28] Soine, W.H. (1986). Clandestine drug synthesis, *Medicinal Research Reviews* **6**, 41–76.

[29] Kintz, P. (2004). Value of hair analysis in postmortem toxicology, *Forensic Science International* **142**, 127–134.

[30] Pötsch, L. & Skopp, G. (2004). Inkorporation von Fremdsubstanzen in Haare, in *Haaranalytik. Technik und Interpretation in Medizin und Recht*, B. Madea & F. Muhoff, eds, Deutscher Ärzteverlag, Köln, pp. 31–98.

[31] Thorspecken, J., Skopp, G. & Pötsch, L. (2004). In vitro contamination of hair by marijuana smoke, *Clinical Chemistry* **50**, 596–602.

[32] Moriya, F. & Hashimoto, Y. (2000). Criteria for judging whether postmortem blood drug concentrations can be used for toxicologic evaluation, *Legal Medicine* **2**, 143–151.

[33] Jones, G.R. (1998). Interpretation of post-mortem drug levels, in *Drug Abuse Handbook*, S.B. Karch, ed, CRC Press, Boca Raton, pp. 970–985.

[34] Larsen, R.H., Nielsen, F., Sorensen, J.A. & Nielsen, J.B. (2003). Dermal penetration of fentanyl: inter- and intraindividual variations, *Pharmacology and Toxicology* **93**, 244–248.

[35] Curry, A.S. & Sunshine, I. (1960). The liver to blood ratio in cases of barbiturate poisoning, *Toxicology and Applied Pharmacology* **2**, 602–606.

[36] Cook, D.S. & Braithwaite, R.A. (2000). Estimating antemortem drug concentrations from postmortem blood samples: the influence of postmortem redistribution, *Journal of Clinical Pathology* **53**, 282–285.

[37] Gerostamoulos, J. & Drummer, O.H. (2000). Postmortem redistribution of morphine and its metabolites, *Journal of Forensic Sciences* **45**, 843–845.

[38] Baselt, R.C. (2004). *Disposition of Toxic Drugs and Chemicals in Man*, 7th Edition, Biomedical Publications, Foster City.

[39] Pounder, D.J. & Jones, G.R. (1990). Post-mortem drug redistribution – a toxicological nightmare, *Forensic Science International* **45**, 253–263.

[40] Pounder, D.J. (1994). The nightmare of postmortem drug changes, in *Legal Medine 1993*, C.H. Wecht, ed, Butterworth Legal Publishers, New Hampshire, pp. 162–191.

[41] Prouty, R.W. & Anderson, W.H. (1990). The forensic implications of site and temporal influences on postmortem blood-drug concentrations, *Journal of Forensic Sciences* **35**, 243–270.

[42] Drummer, O., Forrest, A.R.W., Goldberger, B. & Karch, S.B. (2004). Forensic science in the dock. Postmortem measurements of drug concentration in blood have little meaning, *British Medical Journal* **329**, 636–637.

[43] Pelissier-Alicot, A.L., Gaulier, J.M. & Marquet, P. (2003). Mechanisms underlying post-mortem redistribution of drugs: a review, *Journal of Analytical Toxicology* **27**, 533–544.

[44] Yarema, M.C. & Becker, C.E. (2005). Key concepts in post-mortem drug redistribution, *Clinical Toxicology* **43**, 235–241.

[45] Pounder, D.J., Owen, V. & Quigley, C. (1994). Postmortem changes in blood amitriptyline concentration, *The American Journal of Forensic Medicine and Pathology* **15**, 224–230.

[46] Logan, B.K. & Smirnow, D. (1996). Postmortem distribution and redistribution of morphine in man, *Journal of Forensic Sciences* **41**, 37–46.

[47] Pounder, D.J., Fuke, C., Cox, D.E., Smith, D. & Kuroda, N. (1996). Postmortem diffusion of drugs from gastric residue, *The American Journal of Forensic Medicine and Pathology* **17**, 1–7.

[48] Pounder, D.J. & Smith, D.R.W. (1995). Postmortem diffusion of alcohol from the stomach, *The American Journal of Forensic Medicine and Pathology* **16**, 89–96.

[49] Hepler, B.R. & Isenschmid, D.S. (1998). Specimen selection, collection, preservation, and security, in *Drug Abuse Handbook*, S.B. Karch, ed, CRC Press, Boca Raton, pp. 873–889.

[50] Pounder, D.J. & Yonemitsu, K. (1991). Postmortem absorption of drugs and ethanol from aspirated vomitus – an experimental model, *Forensic Science International* **51**, 189–195.

[51] Mijazaki, T., Kojima, T., Yashiki, M., Wakamoto, H., Iwasaki, Y. & Taniguchi, T. (1993). Site dependence of methamphetamine concentrations in blood samples collected from cadavers of people who had been methamphetamine abusers, *The American Journal of Forensic Medicine and Pathology* **14**, 121–124.

[52] Pounder, D.J., Hartley, A.K. & Watmough, P.J. (1994). Postmortem redistribution and degradation of dothiepin. Human case studies and an animal model, *The American Journal of Forensic Medicine and Pathology* **15**, 231–235.

[53] Fuke, C., Berry, C.L. & Pounder, D.J. (1996). Postmortem diffusion of ingested and aspirated paint thinner, *Forensic Science International* **78**, 100–207.

[54] Pohland, R.C. & Bernhard, N.R. (1997). Postmortem serum and tissue redistribution of fluoxetine and norfluoxetine in dogs following oral administration of fluoxetine hydrochloride (Prozac), *Journal of Forensic Sciences* **42**, 812–816.

[55] Hilberg, T., Bugge, A., Beylich, K.M., Ingum, J., Bjorneboe, A. & Morland, J. (1993). An animal model of postmortem amitriptyline redistribution, *Journal of Forensic Sciences* **38**, 81–90.

[56] Vorpahl, T.E. & Coe, J.I. (1978). Correlation of antemortem and postmortem digoxin levels, *Journal of Forensic Sciences* **23**, 329–334.

[57] Aderjan, R. & Mattern, R. (1980). Zur Wertigkeit postmortaler Digoxin-Konzentrationen im Blut, *Zeitschrift fur Rechtsmedizin* **86**, 13–20.

[58] Dalpe-Scott, M., Degouffe, M., Garbutt, D. & Drost, M. (1995). A comparison of drug concentrations in postmortem cardiac and peripheral blood in 320 cases, *Canadian Society of Forensic Science Journal* **28**, 113–121.

[59] Jones, G.R. & Pounder, D.J. (1987). Site dependence of drug concentrations in postmortem blood – a case study, *Journal of Analytical Toxicology* **11**, 186–190.

[60] Levine, B., Wu, S.C., Dixon, A. & Smialek, J.E. (1995). Site dependence of postmortem blood methadone concentrations, *The American Journal of Forensic Medicine and Pathology* **16**, 97–100.

[61] Milroy, C.M. & Forrest, A.R.W. (2000). Methadone deaths: a toxicological analysis, *Journal of Clinical Pathology* **53**, 277–281.

[62] Berg, M.J., Lantz, R.K., Schentag, J.J. & Vern, B.A. (1984). Distribution of cimetidine in postmortem tissue, *Journal of Forensic Sciences* **29**, 147–154.

[63] Hearn, W.L., Keran, E.E., Wie, H. & Hime, G. (1991). Site-dependent postmortem changes in blood cocaine concentrations, *Journal of Forensic Sciences* **36**, 673–684.

[64] Pounder, D.J. & Davies, J.I. (1994). Zopiclone poisoning: tissue distribution and potential for postmortem diffusion, *Forensic Science International* **65**, 177–183.

[65] Elliott, S.P. (2005). Methylenedioxymethamphetamine and methylenedioxyamphetamine concentrations in ante-mortem and post-mortem specimens in fatalities following hospital admission, *Journal of Analytical Toxicology* **29**, 296–300.

[66] Skopp, G., Klinder, K., Pötsch, L., Zimmer, G., Lutz, R., Aderjan, R. & Mattern, R. (1998). Postmortem distribution of dihydrocodeine and metabolites in a fatal case of dihydrocodeine intoxication, *Forensic Science International* **95**, 99–107.

[67] De Letter, E.A., Clauwaert, K.M., Belpaire, F.M., Lambert, W.E., Van Boxclear, J.F. & Piette, M.H.A. (2002). Post-mortem redistribution of 3,4-methylenedioxy methamphetamine (MDMA, "ecstasy") in the rabbit. Part I: experimental approach after in vivo intravenous infusion, *International Journal of Legal Medicine* **116**, 216–224.

[68] De Letter, E.A., Belpaire, F.M., Clauwaert, K.M., Lambert, W.E., Van Boxclear, J.F. & Piette, M.H.A. (2002). Post-mortem redistribution of 3,4-methylenedioxy methamphetamine (MDMA, "ecstasy") in the rabbit. Part II: post-mortem infusion in trachea or stomach, *International Journal of Legal Medicine* **116**, 225–232.

[69] Hilberg, T., Rogde, S. & Morland, H.J. (1999). Postmortem drug redistribution – human cases related to results in experimental animals, *Journal of Forensic Sciences* **44**, 3–9.

[70] Quatrehomme, G., Bourett, F., Liao, Z. & Ollier, A. (1994). An experimental methodology for the study of postmortem changes in toxic concentrations of drugs, using secobarbital as an example, *Journal of Forensic Sciences* **39**, 1300–1304.

[71] Pounder, D.J., Hartley, A.K. & Watmough, P.J. (1994). Postmortem redistribution and degradation of dothiepin. Human case studies and an animal model, *The American Journal of Forensic Medicine and Pathology* **15**, 231–235.

[72] Moffat, A.C., Osselton, M.D. & Widdop, B. (eds) (2004). *Clarke's Analysis of Drugs and Poisons*, 3rd Edition, Pharmaceutical Press, London, Vol. 2.

[73] Schulz, M. & Schmoldt, A. (2003). Therapeutic and toxic blood concentrations of more than 800 drugs and other xenobiotics, *Pharmazie* **58**, 447–474.

[74] Winek, C.L., Wahba, W.W., Winek Jr, C.L. & Winek Balzer, T. (2001). Drug and chemical blood-level data 2001, *Forensic Science International* **122**, 107–123.

[75] Schleyer, F. (1958). *Postmortale, klinisch-chemische Diagnostik und Todeszeitbestimmung mit chemischen und physikalischen Methoden*, Thieme, Stuttgart, pp. 39–41.

[76] Skopp, G., Lutz, R., Pötsch, L., Ganßmann, B., Klinder, K., Schmidt, A., Aderjan, R. & Mattern, R. (1997). An in vitro experiment for postmortem vascular permeation. The passage of morphine and morphine glucuronides across a vascular wall, *Journal of Forensic Sciences* **42**, 486–491.

[77] Brettel, H.F. (1970). Über Beziehungen zwischen dem Abfall der Blutalkoholkonzentration und dem Wasserverlust des Blutes nach dem Tode, *Blutalkohol* **7**, 54–64.

[78] Thomsen, H. (1998). *Post Mortem Platelets in Blood, Thrombi and Haematomas*, *Research in Legal Medicine*, Schmidt-Römhild, Lübeck, Germany, Vol. 20.

[79] Maguire, K.P., Burrows, G.D., Norman, T.R. & Scoggins, B.A. (1980). Blood/plasma distribution ratios of psychotropic drugs, *Clinical Chemistry* **26**, 1624–1625.

[80] Jeffcoat, A.R., Perez-Reyes, M., Hill, J.M., Sadler, B.M. & Cook, C.E. (1989). Cocaine disposition in humans after intravenous injection, nasal insufflation (snorting), or smoking, *Drug Metabolism and Disposition* **17**, 153–159.

[81] Shull Jr, H.J., Wilkinson, G.R., Johnson, R. & Schenker, S. (1976). Normal disposition of oxazepam in acute viral hepatitis and cirrhosis, *Annals of Internal Medicine* **84**, 420–425.

[82] Payne, J.P., Hill, D.W. & Wood, D.G.L. (1968). Distribution of ethanol between plasma and erythrocytes in whole blood, *Nature* **217**, 963–964.

[83] Inturrisi, C.E., Colburn, W.A., Kaiko, R.F., Houde, R.W. & Foley, K.M. (1987). Pharmacokinetics and pharmacodynamics of methadone in patients with chronic pain, *Clinical Pharmacy and Therapeutics* **41**, 392–401.

[84] Hand, C.W., Moore, R.A. & Sear, J.W. (1988). Comparison of whole blood and plasma morphine, *Journal of Analytical Toxicology* **12**, 234–235.

[85] Skopp, G., Pötsch, L., Ganssmann, B., Aderjan, R. & Mattern, R. (1998). A preliminary study on the

distribution of morphine and its glucuronides in the subcompartments of blood, *Journal of Analytical Toxicology* **22**, 261–264.

[86] Widman, M., Agurell, S., Ehrnebo, M. & Jones, G. (1974). Binding of (+)- and (−)-delta-1-tetrahydrocannabinols and (−)-7-hydroxy-delta-1-tetrahydrocannabinol to blood cells and plasma proteins in man, *The Journal of Pharmacy and Pharmacology* **26**, 914–916.

[87] Hanson, V.W., Buonarati, M.H., Baselt, R.C., Wada, N.A., Yep, C., Biasotti, A.A., Reeve, V.C., Wong, A.S. & Orbanowsky, W.M. (1983). Comparison of ^{3}H-and ^{125}I-radioimmunoassay and gas chromatography/mass spectrometry for the determination of Δ^{9}-tetrahydrocannabinol and cannabinoids in blood and serum, *Journal of Analytical Toxicology* **7**, 96–102.

[88] Giroud, C., Menetrey, A., Augsburger, M., Buclin, T., Sanchez-Mazas, P. & Mangin, P. (2001). Delta(9)-THC, 11-OH-Delta(9)-THC and Delta(9)-THCCOOH plasma or serum to whole blood concentrations distribution ratios in blood samples taken from living and dead people, *Forensic Science International* **123**, 159–164.

[89] Iten, P.X. (1994). *Fahren unter Drogen- oder Medikamenteneinfluss*, Institut für Rechtsmedizin der Universität Zürich, Zürich.

[90] Linnoila, M. & Dorrity, F. (1978). Measurement of plasma and erythrocyte chlorpromazine and N-monodesmethylchlorpromazine levels by gas chromatography with a nitrogen sensitive detector, *Acta Pharmacologica et Toxicologica* **42**, 264–270.

[91] Flanagan, R.J., Amin, A. & Seinen, W. (2003). Effect of post-mortem changes on peripheral and central whole blood and tissue clozapine and norclozapine concentrations in the domestic pig (Sus crofa), *Forensic Science International* **12**, 9–17.

[92] Highley, M.S. & De Bruijn, E.A. (1996). Erythrocytes and the transport of drugs and endogenous compounds, *Pharmaceutical Research* **13**, 186–195.

[93] Morris, J.A., Harrison, L.M. & Partridge, S.M. (2006). Postmortem bacteriology: a re-evaluation, *Journal of Clinical Pathology* **59**, 1–9.

[94] Reinhardt, G., Zink, P. & Legler, F. (1973). Bacteriological findings in cardiac blood of a cadaver, *Beiträge zur Gerichtlichen Medizin* **31**, 311–314.

[95] Stevens, H.M. (1984). The stability of some drugs and poisons in putrefying human liver tissues, *Journal – Forensic Science Society* **24**, 577–589.

[96] Drummer, O.H. (1997). Stability of drugs postmortem: a *review*, *Proceedings of the 35th Annual Meeting of The International Association of Forensic Toxicologists*, Padua, pp. 13–17.

[97] Schmidt, G. (1969). Postmortale Veränderungen von Arzneistoffen und Giften in Organen und Körperflüssigkeiten einschließlich Neubildung von Störsubstanzen, in *Gadamers Lehrbuch der chemischen Toxikologie und Anleitung zur Ausmittelung der Gifte*, R. Preu, ed, Vandenhoeck & Ruprecht, Göttingen, pp. 189–241.

[98] Arnold, D., Naeve, W. & Arnold, W. (1984). Toxikologische Befunderhebungen an Fäulnisleichen, *Zeitschrift fur Rechtsmedizin* **93**, 151–164.

[99] Tsunenari, S., Yonemitsu, K., Uchimara, Y. & Kanda, M. (1981). The influence of putrefactive changes on the determination of paraquat in autopsy materials, *Forensic Science International* **17**, 51–56.

[100] Kiyofuji, T. (1970). Studies on the extraction of poisonous substances from cadaveric materials of synthetic organo-pesticides poisoning and effects of postmortem changes on their extraction, *Journal of the Kumamoto Medical Society* **44**, 767–817.

[101] Robertson, M.D. & Drummer, O.H. (1998). Stability of nitrobenzodiazepines in postmortem blood, *Journal of Forensic Sciences* **43**, 5–8.

[102] Drummer, O.H., Syrjanen, M.L. & Cordner, S.M. (1993). Deaths involving the benzodiazepine flunitrazepam, *The American Journal of Forensic Medicine and Pathology* **14**, 238–243.

[103] Skopp, G. & Pötsch, L. (2002). Stability of 11-nor-delta(9)-carboxy-tetrahydrocannabinol glucuronide in plasma and urine assessed by liquid chromatography-tandem mass spectrometry, *Clinical Chemistry* **48**, 301–306.

[104] Shipkova, M., Armstrong, V.W., Oellerich, M. & Wieland, E. (2003). Acyl glucuronide drug metabolites: toxicological and analytical implications, *Therapeutic Drug Monitoring* **25**, 1–16.

[105] Skopp, G., Pötsch, L., Klingmann, A. & Mattern, R. (2001). Stability of morphine, morphine-3-glucuronide, and morphine-6-glucuronide in fresh blood and plasma and postmortem blood samples, *Journal of Analytical Toxicology* **25**, 2–7.

[106] Moriya, F. & Hashimoto, Y. (1997). Distribution of free and conjugated morphine in body tissues in a fatal heroin overdose: Is conjugated morphine stable in postmortem specimens, *Journal of Forensic Sciences* **42**, 736–740.

[107] Carroll, F.T., Marracini, J.V., Lewis, S. & Wright, W. (2000). Morphine-3-d-glucuronide stability in postmortem specimens exposed to bacterial enzymatic hydrolysis, *The American Journal of Forensic Medicine and Pathology* **21**, 323–329.

[108] Felby, S. & Nielsen, E. (1994). The postmortem blood alcohol concentration and the water content, *Blutalkohol* **31**, 24–32.

[109] Iffland, R. & Palm, W. (1979). Untersuchungen zur postmortalen Alkoholverteilung in Butgefäßen und Körperflüssigkeiten, *Blutalkohol* **16**, 81–96.

[110] Brettel, H.F. (1973). Der Korrekturfaktor bei der gaschromatographischen Leichenblutalkoholbestimmung, *Blutalkohol* **10**, 120–124.

[111] Huckenbeck, W. & Bonte, W. (2003). Alkohol, in *Handbuch Gerichtliche Medizin*, B. Madea & B. Brinkmann, eds, Springer Berlin, Heidelberg, New York, pp. 469–485.

[112] Kugelberg F.C., Jones A.W. (2007). Interpreting results of ethanol analysis in postmortem specimens: a review of the literature, *Forensic Science International* **165**, 10–29.

[113] O'Neal, C.L. & Poklis, A. (1996). Postmortem production of ethanol and factors that influence interpretation, *The American Journal of Forensic Medicine and Pathology* **17**, 8–20.

[114] Zumwalt, R.E., Bost, R.O. & Sunshine, I. (1982). Evaluation of ethanol concentrations in decomposed bodies, *Journal of Forensic Science Society* **27**, 549–554.

[115] Hansen, A.C. (1994). Validity of postmortem alcohol determination, *Ugeskrift for Laeger* **156**, 55–57.

[116] Corry, J.E.L. (1978). Possible sources of ethanol ante- and post-mortem: its relation to the biochemistry and microbiology of decomposition, *The Journal of Applied Bacteriology* **44**, 1–56.

[117] Huckenbeck, W. (1999). *Experimentelle Untersuchungen Zum bakteriell induzierten Alkohol- und Aminosäurenstoffwechsel im Blut der menschlichen Leiche*, Habilitationsschrift, Universität Düsseldorf, Germany.

[118] Bogusz, M., Guminska, M. & Markiewicz, J. (1970). Studies on the formation of endogenous ethanol in blood putrefying in vitro, *Journal of Forensic Medicine* **17**, 156–168.

[119] Seto, Y. (1996). Stability and spontaneous production of blood cyanide during heating, *Journal of Forensic Sciences* **41**, 465–468.

[120] Berankova, K., Mutnanska, K. & Balikova, M. (2006). γ-Hydroxybutyric acid stability and formation in blood and urine, *Forensic Science International* **161**, 158–162.

[121] Karch, S.B. (2001). GHB. Club drug or confusing artefact? *The American Journal of Forensic Medicine and Pathology* **22**, 266–269.

[122] Winek, C.L., Zaveri, N.R. & Wahba, W.W. (1993). The study of tricyclic antidepressants in formalin-fixed human liver and formalin solutions, *Forensic Science International* **61**, 175–183.

[123] Gannett, P.M., Daft, J.R., James, D., Rybeck, B., Knopp, J.B. & Tracy, T.S. (2001). In vitro reaction of barbiturates with formaldehyde, *Journal of Analytical Toxicology* **25**, 443–449.

[124] Tirumalai, P.S., Shakleya, D.M., Gannett, P.M., Callery, P.S., Bland, T.M. & Tracy, T.S. (2005). Conversion of methamphetamine to N-methylmethamphetamine in formalin solutions, *Journal of Analytical Toxicology* **29**, 48–53.

[125] Cingolani, M., Cippitelli, M., Froldi, R., Tassoni, G. & Mirtella, D. (2005). Stability of barbiturates in fixed tissues and formalin solutions, *Journal of Analytical Toxicology* **29**, 205–208.

[126] Tracy, T.S., Rybeck, B.F., James, D.G., Knopp, J.B. & Gannett, P.M. (2001). Stability of benzodiazepines in formaldehyde solutions, *Journal of Analytical Toxicology* **25**, 166–173.

[127] Dettling, R.J., Briglia, E.J., Dal Cortivo, L.A. & Bidanset, J.H. (1990). The production of amitriptyline from nortriptyline in formaldehyde-containing solutions, *Journal of Analytical Toxicology* **14**, 325–326.

[128] Gannett, P.M., Hailu, S., Daft, J., James, D., Rybeck, B. & Tracy, T.S. (2001). In vitro reaction of formaldehyde with fenfluramine: conversion to N-methyl fenfluramine, *Journal of Analytical Toxicology* **25**, 88–92.

[129] Nishigami, J., Takayasu, T. & Ohshima, T. (1995). Toxicological analysis of the psychotropic drugs chlorpromazine and diazepam using chemically fixed tissues, *International Journal of Legal Medicine* **107**, 165–170.

[130] Forney Jr, R.B., Carroll, F.T., Nordgren, I.K., Petterson, B.M. & Holmstedt, B. (1982). Extraction, identification and quantitation of succinylcholine in embalmed tissue, *Journal of Analytical Toxicology* **6**, 115–119.

[131] Takayasu, T., Saito, K., Nishigami, J., Ohshima, T. & Nagano, T. (1994). Toxicological analysis of drugs and poisons in formalin-fixed organ tissues, *International Journal of Legal Medicine* **107**, 7–12.

[132] Pagani, S., Mirtella, D., Mencarelli, R., Rodriguez, D. & Cingolani, M. (2005). Postmortem distribution of sildenafil in histological material, *Journal of Analytical Toxicology* **29**, 254–257.

[133] Cingolani, M., Froldi, R., Mencarelli, R., Mirtella, D. & Rodriguez, D. (2001). Detection and quantitation of morphine in fixed tissues and formalin solutions, *Journal of Analytical Toxicology* **25**, 31–34.

[134] Cingolani, M., Froldi, R., Mencarelli, R. & Rodriguez, D. (1999). Analytical detection and quantitation of strychnine in chemically fixed organ tissues, *Journal of Analytical Toxicology* **23**, 219–221.

[135] Iffland, R., Madea, B. & Balling, P. (1988). Diagnose Kohlenmonoxidvergiftung nach Einbalsamierung und Exhumierung, *Archiv fur Kriminologie* **182**, 100–106.

[136] Tsoukali-Papadopoulou, H. (1987). Elucidation of a poisoning case from the analysis of formalin in which brain tissue was preserved, *Forensic Science International* **34**, 63–65.

[137] Brinkmann, B. (1999). Harmonisation of medico-legal autopsy rules, *International Journal of Legal Medicine* **113**, 1–14.

[138] Forrest, A.R.W. (1993). Obtaining samples at post mortem examination for toxicological and biochemical analyses, *Journal of Clinical Pathology* **46**, 292–296.

[139] Vanbinst, R., Koenig, J., Di Fazio, V. & Hassoun, A. (2002). Bile analysis of drugs in postmortem cases, *Forensic Science International* **128**, 35–40.

[140] Maurer, H.H. (1999). Systematic toxicological analysis procedures for acidic drugs and/or metabolites relevant to clinical and forensic toxicology and/or doping control, *Journal of Chromatography. B, Biomedical Sciences and Applications* **733**, 3–25.

[141] Jones, G. (2004). Postmortem toxicology, in *Clarke's Analysis of Drugs and Poisons*, A.C. Moffat, M.D. Osselton & B. Widdop, eds, 3rd Edition, Pharmaceutical Press, London, Chicago, pp. 94–108.

[142] Jenkins, A.J. & Lavins, E.S. (1998). 6-Acetylmorphine detection in postmortem cerebrospinal fluid, *Journal of Analytical Toxicology* **22**, 173–175.

[143] Scott, K.S. & Oliver, J.S. (1999). Vitreous humor as an alternative sample to blood for the supercritical fluid extraction of morphine and 6-acetylmorphine, *Medicine, Science, and the Law* **39**, 77–81.

[144] Ritz, S., Harding, P., Martz, W., Schütz, H.W. & Kaatsch, H.J. (1992). Measurement of digitalis-glycoside levels in ocular tissues: a way to improve postmortem diagnosis of lethal digitalis-glycoside poisoning? I: digoxin, *International Journal of Legal Medicine* **105**, 149–154.

[145] Anastasos, N., McIntyre, I.M., Lynch, M.J. & Drummer, O.H. (2002). Postmortem concentrations of citalopram, *Journal of Forensic Sciences* **47**, 882–884.

[146] Furnari, C., Ottaviano, V., Sachetti, G. & Mancini, M. (2002). A fatal case of cocaine poisoning in a body packer, *Journal of Forensic Sciences* **47**, 208–210.

[147] Decaestecker, T., De Letter, E., Clauwaert, K., Bouche, M.P., Lambert, W., Van Boxclear, J., Piette, M., Van den Eeckhout, E., Van Peteghem, C. & De Leenheer, A. (2001). Fatal 4-MTA intoxication: development of a liquid chromatographic-tandem mass spectrometric assay for multiple matrices, *Journal of Analytical Toxicology* **25**, 705–710.

[148] Gock, S.B., Wong, S.H., Stormo, K.A. & Jentzen, J.M. (1999). Self-intoxication with morphine obtained from an infusion pump, *Journal of Analytical Toxicology* **23**, 130–133.

[149] Wogoman, H., Steinberg, M. & Jenkins, A.J. (1999). Acute intoxication with guaifenesin, diphenhydramine, and chlorpheniramine, *The American Journal of Forensic Medicine and Pathology* **20**, 199–202.

[150] Scott, K.S. & Oliver, J.S. (2001). The use of vitreous humor as an alternative to whole blood for the analysis of benzodiazepines, *Journal of Forensic Sciences* **46**, 694–697.

[151] Caplan, Y.H. & Levine, B. (1990). Vitreous humor in the evaluation of postmortem blood ethanol concentrations, *Journal of Analytical Toxicology* **14**, 305–307.

[152] Jones, A.W. & Holmgren, P. (2001). Uncertainty in estimating blood ethanol concentrations by analysis of vitreous humor, *Journal of Clinical Pathology* **54**, 699–702.

[153] Musshoff, F., Padosch, S., Steinborn, S. & Madea, B. (2004). Fatal blood and tissue concentrations of more than 200 drugs, *Forensic Science International* **142**, 161–210.

[154] Pounder, D.J., Adams, E., Fuke, C. & Langford, A.M. (1996). Site to site variability of postmortem drug concentrations in liver and lung, *Journal of Forensic Sciences* **41**, 927–932.

[155] Merrick, T.C., Felo, J.A. & Jenkins, A.J. (2001). Tissue distribution of olanzapine in a postmortem case, *The American Journal of Forensic Medicine and Pathology* **22**, 270–274.

[156] Garriott, J.C. (1991). Skeletal muscle as an alternative specimen for alcohol and drug analysis, *Journal of Forensic Sciences* **36**, 60–69.

[157] Skopp, G., Potsch, L., Eser, H.P. & Moller, M.R. (1996). Preliminary practical findings on drug monitoring by a transcutaneous collection device, *Journal of Forensic Sciences* **41**, 933–937.

[158] Kijewski, H. (1993). *Die Forensische Bedeutung der Mineralstoffgehalte in menschlichen Kopfhaaren*, Schmidt-Römhild, Köln, pp. 53–95.

[159] Sachs, H. & Kintz, P. (1998). Testing for drugs in hair. Critical review of chromatographic procedures since 1992, *Journal of Chromatography. B, Biomedical Sciences and Applications* **713**, 147–161.

[160] Pragst, F., Rothe, M., Spiegel, K. & Sporkert, F. (1998). Illegal and therapeutic drug concentrations in hair segments – A timetable of drug exposure? *Forensic Science Reviews* **10**, 81–111.

[161] McCurdy, W.C. (1987). Postmortem specimen collection, *Forensic Science International* **35**, 61–65.

[162] Dyne, D., Cocker, J., Streete, P.J. & Flanagan, R.J. (1996). Toluene, 1-butanol, ethylbenzene and xylene from sarstedt monovette serum gel blood collection tubes, *Annals of Clinical Biochemistry* **33**, 355–356.

[163] Klingmann, A., Skopp, G. & Aderjan, R. (2001). Analysis of cocaine, benzoylecgonine, ecgonine methyl ester, and ecgonine by high pressure liquid chromatography-API mass spectrometry and application to a short-term degradation study of cocaine in plasma, *Journal of Analytical Toxicology* **25**, 425–430.

[164] Barrett, D.A., Dyssegaard, A.L.P. & Shaw, P.N. (1992). The effect of temperature and pH on the deacetylation of diamorphine in aqueous solution and in human plasma, *The Journal of Pharmacy and Pharmacology* **44**, 606–608.

[165] Ferrara, S.D., Frison, G., Tdeschi, L. & LeBeau, M. (2001). γ-Hydroxybutyrate (GHB) and related products, in *Drug-Facilitated Sexual Assault*, M.A. LeBeau & A. Mozayani, eds, A Forensic Handbook, Academic Press, San Diego, pp. 107–126.

[166] Moriya, F., Hashimoto, Y. & Kuo, P.L. (1999). Pitfalls when determining tissue distributions of organophosphorus chemicals: sodium fluoride accelerates chemical degradation, *Journal of Analytical Toxicology* **23**, 210–215.

[167] Heinig, R., Zimmer, D., Yeh, S. & Krol, G.J. (2000). Development, validation and application of assays to quantify metrifonate and 2,2-dichlorovinyl dimethylphosphate in human body fluids, *Journal of Chromatography. B* **741**, 257–269.

[168] Moriya, F. & Hashimoto, Y. (1996). The effect of postmortem interval on the concentrations of cocaine and cocaethylene in blood and tissues: an experiment using rats, *Journal of Forensic Sciences* **41**, 129–133.

[169] Olesen, O.V. & Linnet, K. (1998). Determination of olanzapine in serum by high-performance liquid chromatography using ultraviolet detection considering the

easy oxidability of the compound and the presence of other psychotropic drugs, *Journal of Chromatography. B* **714**, 309–315.

[170] Smith, R.V., Wilcox, R.E. & Humphrey, D.W. (1980). Stability of apomorphine in frozen plasma, *Research Communications in Chemical Pathology and Pharmacology* **27**, 183–186.

[171] Curry, S.H. & Evans, S. (1976). A note on the assay of chlorpromazine N-oxide and its sulphoxide in plasma and urine, *The Journal of Pharmacy and Pharmacology* **28**, 1467–1468.

[172] Westerling, D., Bengtsson, D.I., Thysell, C. & Hoglund, P. (1996). The influence of preanalytical factors on concentrations of morphine and metabolites in patients receiving morphine, *Pharmacology and Toxicology* **78**, 82–85.

[173] Stamm, D. (1982). A new concept for quality control of clinical laboratory investigations in the light of clinical requirements and based on reference method values, *Journal of Clinical Chemistry and Clinical Biochemistry* **20**, 817–824.

[174] Ellenhorn, M.J. (ed) (1997). Appendix H. The poisoned patients and the laboratory – "the flanagan tables", in *Ellenhorn's Medical Toxicology: Diagnosis and Treatment of Human Poisoning*, 2nd Edition, Williams & Wilkins, Baltimore, pp. 1929–1933.

[175] Drummer, O.H. (2001). *The Forensic Pharmacology of Drugs of Abuse*, Arnold, London, New York, New Delhi.

[176] Levine, B. & Smith, M.L. (1990). Stability of drugs of abuse in biological specimens, *Forensic Science Reviews* **2**, 147–157.

[177] Skopp, G. & Pötsch, L. (2002). Zur präanalytischen Phase chemisch-toxikologischer Untersuchungen. II: Stabilität forensisch relevanter Substanzen in Blut-, Plasma- oder Serumproben – eine Bestandsaufnahme, *Rechtsmedizin* **12**, 195–202.

[178] Saker, E.G., Eskew, A.E. & Panter, J.W. (1991). Stability of toluene in blood: its forensic relevance, *Journal of Analytical Toxicology* **15**, 246–249.

[179] Curry, A.S., Price, D.E. & Rutter, R.C. (1967). The production of cyanide in post mortem material, *Acta Pharmacologica et Toxicologica* **25**, 339–344.

[180] Egekeze, J.O. & Oehme, F.W. (1980). Thiocyanate to cyanide: revisited, *Clinical Toxicology* **16**, 127–128.

[181] Ballantyne, B., Bright, J.E. & Williams, P. (1974). The post mortem rate of transformation of cyanide, *Forensic Science* **3**, 71–76.

[182] Ballantyne, B. (1976). Changes in blood cyanide as a function of storage time and temperature, *Journal of Forensic Sciences* **16**, 305–310.

[183] Giorgi, S.N. & Meeker, J.E. (1995). A 5-year stability study of common illicit drugs in blood, *Journal of Analytical Toxicology* **19**, 392–398.

[184] Skopp, G., Pötsch, L., Mattern, R. & Aderjan, R. (2002). Short-term stability of lysergic acid diethylamide (LSD), N-desmethyl-LSD, and 2-oxo-3-hydroxy-LSD in urine, assessed by liquid chromatography-tandem mass spectrometry, *Clinical Chemistry* **48**, 1615–1618.

[185] Robertson, M.D. & Drummer, O.H. (1995). Postmortem drug metabolism by bacteria, *Journal of Forensic Sciences* **40**, 382–386.

[186] El Mahjoub, A. & Staub, C. (2000). Stability of benzodiazepines in whole blood samples stored at varying temperatures, *Journal of Pharmaceutical and Biomedical Analysis* **23**, 1057–1063.

[187] Entwhistle, N., Owen, P., Patterson, D.A., Jones, L.V. & Smith, J.A. (1986). The occurrence of chlordiazepoxide degradation products in sudden deaths associated with chlordiazepoxide, *Journal of Forensic Sciences* **26**, 45–54.

[188] Couper, F.J. & Drummer, O.H. (1999). Postmortem stability and interpretation of beta 2-agonist concentrations, *Journal of Forensic Sciences* **44**, 523–526.

[189] McCracken, R.J., Blanchflower, W.J., Rowan, C., McCoy, M.A. & Kennedy, D.G. (1995). Determination of furazolidone in porcine tissue using thermospray liquid chromatography-mass spectrometry and a study of the pharmacokinetics and stability of its residues, *The Analyst* **120**, 2347–2351.

[190] McCracken, R.J., McCoy, M.A. & Kennedy, D.G. (1997). The prevalence and possible causes of bound and extractable residues of the furazolidone metabolite 3-amino-2-oxazolidinone in porcine tissues, *Food Additives and Contaminants* **14**, 287–294.

[191] Isenschmid, D.S., Levine, B.S. & Caplan, Y.H. (1989). A comprehensive study of the stability of cocaine and its metabolites, *Journal of Analytical Toxicology* **13**, 250–256.

[192] Meatherall, R.C. & Fraser, A.D. (1998). CEDIA dau: a reformulation, *Journal of Analytical Toxicology* **22**, 270–273.

[193] Kupiec, T., DeCicco, L., Spiehler, V., Sneed, G. & Kemp, P. (2002). Choice of an ELISA assay for screening post-mortem blood for amphetamine and/or methamphetamine, *Journal of Analytical Toxicology* **26**, 513–518.

[194] Mauden, M., Skopp, G., Mattern, R. & Aderjan, R. (2000). GC/MS-Bestimmungen von THCCOOH im Serum: Vergleich verschiedener Aufarbeitungsmethoden und Einfluß von THCCOOH-Glucuronid, *Blutalkohol* **37**, 45–53.

[195] Wester, R., Noonan, P., Markos, C., Bible Jr, R., Aksamit, W. & Hribar, J. (1981). Identification of carbamate derivatives formed during chloroform extraction of tricyclic antidepressants in urine, *Journal of Chromatography* **209**, 463–466.

[196] Gadkariem, E.A., El-Obeid, H.A., Abounassif, M.A., Ahmed, S.M. & Ibrahim, K.E. (2003). Effects of alkali and simulated gastric and intestinal fluids on danazol stability, *Journal of Pharmaceutical and Biomedical Analysis* **26**, 743–751.

GISELA SKOPP

Postmortem Toxicology: Interpretation

Introduction

The detection of drugs and other substances in biological tissues, such as blood, represents the first stage in the application of toxicology to the forensic sciences. The presence of drug including situations where quantitative data are available requires careful interpretation.

With few exceptions, interpretation requires a thorough understanding of the circumstances of the case and an advanced knowledge of how the substances detected by the analyses interact with the body. Thus, knowledge of both the pharmacokinetics (effect of drug) and the pharmacology of the substance is required and must be carefully related to the known circumstances of the case.

This article provides an overview of the basic pharmacokinetics of drugs and how route of administration and the overall health of the person can influence the interpretation of toxicological results. Several examples are included to illustrate how toxicological data can be misinterpreted.

Other articles in this encyclopedia provide details of the expected effects of substances; see sections on drug classes: alcohol, amphetamines, cocaine, benzodiazepines, and the opioids (opiates). (*See also* **Behavioral Toxicology**).

Basic Pharmacokinetics

To understand the way substances such as drugs are absorbed and the time course of their presence in the body, it is necessary to understand some basic effects of drugs in the body, known as *pharmacokinetics* [1, 2].

The main pharmacokinetic phases can be segregated as follows: absorption, distribution, and elimination.

Absorption

Drugs that are swallowed rely on the release of the drug or other dose form in the stomach or small intestine. This can occur through disintegration of a tablet or through a controlled release of drug from the tablet matrix. Controlled release of drugs is designed to slow the absorption, to (usually) prolong the drug's actions, or sometimes to protect the stomach from potentially harmful drug (i.e., enteric-coated tablets). Except for some acidic drugs (e.g., acetylsalicylic acid), drugs are primarily absorbed in the small intestine and then mainly in the upper sections (jejumen). The delay from swallowing to first appearance of drug in the blood stream can be typically 15–30 min and will occur over many hours.

The time to the maximum blood (or serum) concentration is termed T_{max} (units: time such as hours), while the concentration at this time is termed C_{max} (units: mass per volume, i.e., milligrams per liter).

Many drugs once absorbed (or their subsequent metabolites) can be excreted into the bowel through bile and become reabsorbed further down the gastrointestinal tract. This is known as *enterohepatic recirculation* and can lead to an apparent delay (or even a second peak) in the absorption of drugs. Morphine is a common example in which morphine glucuronide metabolites secreted into bile find their way back into the bowel and are subsequently reabsorbed.

Unabsorbed drug is present in feces and can represent a significant proportion of the administered drug.

A number of pharmaceutical formulations provide a controlled release of drug from a tablet matrix. This is used to control the absorption of drug into the body and provide a longer duration of action of the drug. This is typically used for drugs with short biological actions and has the net result in reducing the need for repeated doses within a 1-day period. For example, morphine and oxycodone can be given twice daily rather than four times daily when formulated into a sustained delivery tablet or capsule.

Distribution

Once the drug is absorbed and has entered the blood stream, it is distributed to all parts of the body. The uptake of drug into tissues and organs will depend on access of drug to all parts of the tissue (e.g., blood supply) and the relative affinity of the drug-to-tissue components. There is considerable variation of drug uptake for different drugs and also between individuals. For some drugs, the affinity of drug to a tissue can be manyfold higher than the blood (e.g., THC in fat and muscle tissue).

The distribution phase is variable but usually requires some hours or days of exposure before some form of steady-state situation is reached.

Elimination

All foreign substances are eventually removed by the body; the rate depends on the drug and its ability to be metabolized (e.g., by liver) and excreted by the kidneys and other organs. Some drugs, e.g. cocaine, are rapidly metabolized to less active compounds and excreted within hours, whereas methamphetamine may only be excreted within days. The metabolism of drugs can be quite complex and often involve multiple pathways. In some cases, metabolites are also biologically active and contribute to the pharmacological response in a person (e.g., methamphetamine is metabolized to the active amphetamine) [3].

The time to halve the blood concentration (following the peak concentration) for methamphetamine can be over 1 day. This time is called *half-life*. This value applies to the terminal elimination phase once the drug has been fully absorbed and distributed to bodily tissues. There is considerable variation in half-lives among individuals, even for the same drug. Half-lives measured before the terminal elimination phase is dominant will generally underestimate the terminal elimination rate.

Some pharmacokinetic data including terminal elimination half-lives for common drugs of abuse are given in Table 1.

The term *clearance* is often used as another measure to quantify the removal of drugs from the body and represents a composite of all forms of drug removal. This includes a combination of kidney excretion, liver metabolism, and other sources of drug removal.

For volatile substances, elimination can also occur through expiration, e.g., alcohol (ethanol), solvents, although this is still a relatively minor source of elimination compared to metabolism and excretion through the kidneys.

Route of Administration and Bioavailability

The proportion of drug available to the body when compared with another route of administration is termed *bioavailability*. This term usually refers to the comparative availability of drugs that are orally taken compared with the same dose given intravenously. Drugs that have a bioavailability of 100% are completely absorbed orally and are not metabolized prior to entering the blood supply. For example, morphine has a bioavailability of 25% when given orally as tablets, meaning that only 25% of morphine is available to the body after oral administration. In total, 75% is either not absorbed or is metabolized prior to entering the blood. When morphine is given by intravenous injection, the bioavailability is 100%. Moreover, the injection has delivered the drug to the blood stream almost instantaneously by passing the absorption phase. Drugs given by intravenous injection will have an immediate intense effect on the person. In abuse situations, such as in the use of heroin, this can precipitate a cardiorespiratory collapse and sudden death [4].

Substances given by other modes of administration, e.g., nasal insufflation (snorting), inhalation (volatile substance abuse), smoking, etc., have different rates of drug absorption.

Table 1 Typical pharmacokinetic data for some common drugs

Drug	Dose range (mg)[a]	Typical blood concentrations ($mg\,l^{-1}$)[b]	Half-life (days)[c]
Alprazolam	0.5–4	0.05–0.2	0.3–1
Amphetamine	From 10	0.1–0.2	0.3–1.5
Diazepam	5–40	0.1–0.6	0.8–2
Cocaine	From 25	0.1–0.5	0.6–4
Codeine	8–60	0.1–0.3	0.1–0.2
MDMA	50–150	0.1–0.3	0.4–1
Methadone	5–120	0.1–0.3	0.6–3
Methamphetamine	From 10	0.05–0.2	0.5–1.5
Morphine	From 5	0.1–0.4	0.1–0.4

[a] Usual dose range
[b] Typical blood concentrations following common doses seen in forensic cases
[c] Pharmacokinetic half-life of terminal elimination phase
MDMA, 3,4-methylenedioxymethamphetamine

Single versus Multiple Doses

The pharmacokinetics of drugs do not generally change with repeated doses of drug; however, depending on the time between doses, there can be carryover from previous dose(s). For example, a drug with a half-life of 12 h will need about five times this

for the drug to be removed from the body. Hence, the administration of a further dose earlier than 60 h (5 × 12 h) will result in some accumulation of drug from dose to dose. In practice, multiple doses are given at least daily, and sometimes two or three times daily; hence, drugs with half-lives of more than a several hours will result in larger pharmacological responses on repeated dosing [1].

Methadone, a drug related to morphine, is used widely to treat dependency to heroin and other opioids. It has a half-life of about 24 h and when given once daily, the blood concentrations increase substantially over the first 5 days of treatment. This can cause potentially fatal toxicity if the initial doses are too high for the established tolerance to opioids [5]. To avoid this phenomenon, low starting doses are recommended, with daily monitoring for the first week of treatment to ensure optimal safety (and response) for the subject.

The interpretation of blood concentrations in a person on a drug such as methadone is further complicated by the accumulation of drug with repeated doses. Hence, the only way a toxicological result can be properly interpreted is to establish whether the drug was likely to have been taken as one (larger) dose or by repeated (smaller) doses.

Predicting Blood Concentrations

There is considerable variation in the way humans respond to drugs. A standard dose, even when corrected for body weight, will show considerable pharmacokinetic variability from one person to another and will even vary in the same person when the substance is given on separate days. This is because any one of the three processes (absorption, distribution, and elimination) will affect the plasma concentration *versus* time profiles. This difference is increased when the drug is given orally when compared with injection, since variability in absorption also occurs.

Figure 1 illustrates schematically what would be expected of typical person-to-person variability for an orally administered drug when both T_{max} and C_{max} are quite different, and the overall area under the plasma concentration *versus* time profile indicates the amount of drug available to the body.

It is important to understand that each drug has its own pharmacokinetic properties. These include the rate of absorption, the degree of distribution in bodily tissues, and the rate of metabolism and elimination. Standard texts provide details of the relevant pharmacokinetic factors to provide a guide as to the effects of the particular drug [2, 6].

A number of physiological factors can further affect blood concentrations of drugs. These include any disease that alters absorption, distribution, and elimination. The most common diseases are those of the liver and the kidneys, since these are the most important organs involved in the elimination of drugs. Liver is a major organ that metabolizes drugs to (more water soluble) metabolites that are more likely to be excreted by the kidneys [7]. Heart disease, such as congestive heart failure, can also affect drug clearance since blood flow through vital organs is reduced. Advanced age (over 70 year) will usually result in reduced organ function, leading to a reduced ability to process drugs. Older (but otherwise

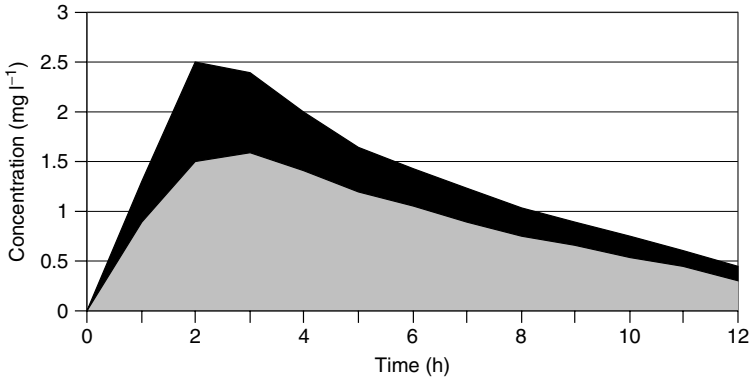


Figure 1 Stylized blood concentration *versus* time profiles showing possible diverse profiles in two different persons given the same dose of drug orally

healthy) persons will often require lower doses for their weight than their much younger counterparts.

The net consequence is that there is substantial variability in blood (and plasma/serum) concentrations for a given dose of substance even if the time of administration is known. In a great proportion of forensic cases, the time of dosage is unknown and in many instances the mode of administration (e.g., oral, intravenous, nasal insufflation, etc.) is usually assumed.

Postmortem Artifacts

In situations involving the interpretation of toxicology results in deceased persons, it is probable that concentrations of drugs (and other substances) would have changed from the time of death [3, 8].

These changes have been described in **Postmortem Toxicology: Artifacts**.

Postcollection Artifacts

Some substances are unstable chemically and can degrade postcollection, particularly, if optimal storage conditions are not maintained. This applies not only to volatile substances such as alcohol but also to nonvolatile substances that are not chemically stable under the storage conditions and, in particular, biological matrix.

These changes have been described in **Postmortem Toxicology: Artifacts**.

Examples

Table 2 summarizes examples of how a particular drug concentration in blood can be differentially interpreted based on the available information.

See **Cocaine**; **Opioids**; **Amphetamine**.

Repeated Use of Drugs

In the case of methamphetamine, where the drug can accumulate with repeated use and produce much greater concentrations than after single doses, it is not possible to infer from a concentration the likely consequence of the drug.

Similarly, with long-acting opiates such as methadone, accumulation occurs from one dose to another, leading to apparent elevated concentrations after some days of use. The interpretation is further compounded by the neuroadaptation that occurs to drugs with repeated use, leading to tolerance of potentially harmful effects [5].

The context of the death is critical to understanding the role, if any, of the drugs detected. For example, a person can die from the effects of using too much cocaine (usually from an adverse effect on the heart), but equally well a person can die from a

Table 2 The effect of circumstances on the interpretation of toxicology results

Drug	Blood concentration ($mg\,l^{-1}$)[a]	Circumstances	Likely interpretation[b]
Methamphetamine	0.5	Single oral dose, 2 h postdose	Moderate-to-high doses (<50 mg)
		Several doses over a few days, oral, 2 h postdose	Low-to-moderate doses (~50 mg)
		Single oral dose, 24 h postdose	High dose (>100 mg)
Morphine	0.5 (free)	Single oral dose of morphine, 2 h postdose	Moderate dose, potentially toxic (>50 mg)
		Single intravenous dose of heroin, 2 h postdose	Moderate dose, potentially toxic (>50 mg)
		Multiple oral doses of sustained release morphine (days)	Low-to-moderate doses, likely to be safe (<50 mg doses when given a number of times daily)
Cocaine	0.5	Single oral dose, 1 h postdose	Moderate-to-high doses (<50 mg)
		Single snorted dose, 2 h postdose	Low-to-moderate doses (~50 mg)
		Multiple snorted doses over 4 h	High dose (>100 mg)

[a] Peripheral clinical blood collected without reference to any postcollection artifacts

[b] Doses are only guides and should not be taken literally and will also depend on health and size of person

completely unrelated event such as a fatal shooting or motor vehicle crash and have the same concentration of cocaine (and its metabolites) as the person who died from the effects of cocaine. There is no relationship between the concentration of drug and an outcome for most substances when assessed in isolation.

This lack of relation between blood concentration and response is also seen with other drugs of abuse as well as most prescription drugs.

Natural Disease

As one might expect, the presence of natural disease can complicate the interpretation of drug effects. For example, a person with heart disease, such as an enlarged heart or atherosclerosis (blocked arteries), is more likely to suffer from an adverse reaction to amphetamines than an otherwise healthy person. In fact, the use of chronic amphetamine is linked to the premature development of heart disease itself [9].

Lung disease including infections, such as pneumonia, can increase the harmful effects of opioids by further weakening the ability to withstand compromised respiratory function caused by the opioids themselves (e.g., morphine and heroin).

Multiple Drugs

In most forensic cases, more than one substance is present in the person concerned [3]. When this occurs, it is likely that the combined effects need to be considered. The use of cocaine in a person using an opioid (e.g., heroin) is more dangerous than either drug alone, even though one drug is a stimulant and the other is a depressant.

Commonly, alcohol is seen in the presence of another drug. Unless the concentration of alcohol is toxicologically insignificant ($<0.02\,\mathrm{g}\ 100\,\mathrm{ml}^{-1}$), alcohol will enhance the effects of the other drug(s). Benzodiazepines that are ordinarily relatively safe drugs will be substantially more toxic in the presence of significant amounts of alcohol. Combined drug use is also more likely to lead to behavioral changes. (*See also* **Alcohol**; **Benzodiazepines**; **Behavioral Toxicology**).

Summary

In summary, in a forensic case, a large range of factors affects the interpretation of a drug concentration. It is essential to gather as much information as physically feasible as well as to interpret the information in the context of the case. There is no substitute to proper collection of both medical and forensic evidence.

References

[1] Gibaldi, M. (1991). *Biopharmaceutics and Clinical Pharmacokinetics*, Lea & Febiger, Philadelphia.

[2] Moffat A.C., Osselton W.D. & Widdop B. *Clarke's Isolation and Identification of Drugs*, (2004). The Pharmaceutical Press, London.

[3] Drummer, O.H. & Odell, M. (2001). *The Forensic Pharmacology of Drugs of Abuse*, Arnold, London.

[4] Gerostamoulos, J., Staikos, V. & Drummer, O.H. (2001). Heroin-related deaths in Victoria: a review of cases for 1997 and 1998, *Drug and Alcohol Dependence* **61**, 123–127.

[5] Caplehorn, J.R. & Drummer, O.H. (1999). Mortality associated with New South Wales methadone programs in 1994: lives lost and saved, *The Medical Journal of Australia* **170**, 104–109.

[6] Baselt, R.C. (2004). *Disposition of Toxic Drugs and Chemicals in Man*, Year Book Medical Publishers.

[7] Morgan, D.J. & McLean, A.J. (1995). Clinical pharmacokinetic and pharmacodynamic considerations in patients with liver disease: an update, *Clinical Pharmacokinetics* **29**, 370–391.

[8] Drummer, O.H. (2007). Post-mortem toxicological redistribution, in *Essentials of Autopsy Practice*, G. Rutty, ed, Springer Verlag, London.

[9] Karch, S.B., Stephens, B.G. & Ho, C.H. (1999). Methamphetamine-related deaths in San Francisco: demographic, pathologic, and toxicologic profiles, *Journal of Forensic Science* **44**, 359–368.

Olaf H. Drummer and Dimitri Gerostamoulos

Postmortem Toxicology: Laboratory Analysis

Introduction

As a major application of forensic toxicology, postmortem toxicology assists pathologists, coroners, and/or judges (in criminal matters) in determining the cause and manner of death. It assists in the assessment

of whether injury, toxicity of drug(s) or poison(s) or natural causes (e.g., disease) has any relevance to the death. On the other hand, the circumstance in which the death occurs determines the manner of death, i.e., misadventure, suicide, homicide, etc.

There are several situations that usually require postmortem toxicological analysis:

1. when the direct cause of death is suspected to be drug or poison related;
2. when the cause of death is suspicious or unknown, and investigation is needed to establish cause and/or manner of death (including drug or poison related);
3. when the cause of death is known, toxicological analysis is needed to clarify the manner of death. For instance, in a fatal traffic accident, an investigation is needed to ascertain any influence of alcohol and/or drug(s) on driver;
4. when involvement of drugs or poisons need to be ruled out to support negative pathological findings during autopsies;
5. when detected drugs may give clues as to any underlying disease process, i.e., anticonvulsant drugs, antidiabetic drugs, etc.;
6. when there is a statutory requirement for autopsies and subsequently toxicological examinations are deemed necessary by an authority such as coroners.

The use of postmortem analysis as a crucial evidence in court can be dated back to 1840 when Dr Mattheiu Orfila, a Spanish-born French physician, was asked by the court to investigate a case of Marie LaFarge for the murder of her husband using arsenic, a common poison used in those days [1, 2]. Before Orfila's investigation, toxicological analysis of arsenic was found positive in the food but was not detected in the stomach content of the victim using Marsh test [3]. However, Orfila discovered that the test had been inappropriately performed, and arsenic was later detected in the victim's body. As a consequence, Marie was found guilty of murder.

According to the history of poisons written in an Egyptian manuscript, Ebers Papyrus, approximately 1500 BC, mandrake, hemlock, opium, aconites, and certain metals from the natural sources were known for their poisonous properties and they had been used as weapons or in torture [4]. A poison can be defined as any substance that when taken in a sufficient quantity will cause intoxication or even death. This dose-dependent relationship was first explained by Paracelsus (1493–1541) with the statement "Sola dosis facit venenum" (only dose determines the poison). For instance, even the well-known deadly poison such as cyanide (CN), arsenic, or carbon monoxide may not be harmful if it is inhaled or ingested in a minute quantity. On the other hand, substances as innocuous as drinking water or minerals such as potassium or sodium, if taken in excessive quantity, could induce death. To analyze postmortem samples for a wide variety of potential intoxicants, which are unknown to the toxicologist in most of the cases, is a challenging task.

Specimens for Toxicological Analysis

The collection of specimens in an autopsy is the first and most important step in toxicological examination as the availability of proper biological specimens can maximize the chance of obtaining meaningful analytical findings to assist in determining the cause and manner of death. Since deterioration of specimens increases with postmortem time interval, biological specimens should be collected as soon as practicable after death. Additionally, in most cases, autopsy can only be performed once; the collection of specimens after autopsy is rarely possible.

Unlike in clinical toxicology, where serum/plasma and urine are usually available, the choice of specimens in postmortem toxicological investigation can be extensive and variable. The specimens selected for analysis can vary depending on the case and the information provided to the pathologist during the investigation. Additionally, the instrumentation and methodologies available to the laboratory will also determine what a laboratory can do. Importantly, the specimens obtainable will play a major role, since some specimens may not be available such as urine (if bladder is emptied or because of decomposition) [5, 6]. Generally, the pathologist has the final say on the specimens collected. Prudent facilities will collect a full set of specimens even if not immediately needed since they can be kept and stored if analysis is not immediately required.

The most common specimens used for general toxicological examinations of drugs and poisons in postmortem cases are blood, urine, and vitreous

Table 1 Suggested postmortem specimens to be collected[a]

Type of cases	Specimens collected
General cases	Blood
	Urine
	Vitreous humor
General cases (blood not available)	Cavity fluid (for screening)
	Liver
	Urinary bladder washing (if urine not available)
	Vitreous humor (if any)
Drug- or poison-related cases (suicide cases)	Plus gastric contents
Drug- or poison-related cases (suspicious cases)	Plus gastric content and liver
Gaseous or volatile substances	Plus lung and brain tissue
Heavy metal poisoning	Plus liver, kidney, and hair

[a] Other specimens can also be collected if deemed necessary

humor. When these are not available, tissue samples such as liver, brain, lung, muscle, and bones and body fluids such as bile, pleural effusions, and other specimens (e.g., hair and nails) are all useful in postmortem analysis. Table 1 summarized the specimens recommended, if available, to be collected on the basis of the types of cases.

Blood

Blood is often the specimen of choice for detecting, quantifying, and interpreting drug concentration in postmortem toxicology as most of the literature data are based on their examination in blood [7–10]. Drug concentration in blood is useful for establishing any recent ingestion of the drug in question and to determine the effect of drug on the deceased at the time of death. Blood sample is preferentially taken from peripheral sites such as the femoral (upper leg) or subclavian region rather than from the cardiac region in order to avoid contamination from abdominal fluids and contents, and to reduce artifactual rises in blood concentration due to postmortem redistribution [11]. However, diffusion of drug from urinary bladder to femoral blood can take place if a large amount of urine contains high concentration of the drug [12]. For cases involving hospital treatment before death, clinical specimens obtained soon after admission and immediately before death, whenever appropriate, should also be investigated particularly when poisoning is suspected before admission into hospital (*see* **Postmortem Toxicology: Artifacts**).

If vascular blood cannot be obtained, sampling can be made from the thoracic or the abdominal cavity. However, the composition of these samples will be markedly different from the whole blood and should therefore only be used to qualitatively determine the presence of drugs or poisons. In case of advanced putrefaction, pleural fluid should be collected for screening [6].

Urine

Urine is a convenient specimen for toxicological screening. This is because (i) a relatively high concentration of drugs and their metabolites accumulate in urine and (ii) the drug detection time window in urine is usually longer than that in blood, thereby facilitating detection of any possible exposure to potential drug(s)/poison(s). Immunoassay can be performed directly on urine specimens for the detection of certain drug classes, especially for those commonly abused drugs. However, there is no correlation of urine drug concentration with pharmacological effects because of the time difference between drug absorbed into the bloodstream and drug eliminated into the urine. In addition, in acute drug-related deaths where survival time (probably less than 15 min) is short, drug may not be excreted into the urine. Therefore, when both urine and blood are available, blood cannot be substituted by urine for the screening of drugs and poisons, and if found, the quantitation of drugs or poisons should preferably be performed on blood. Unfortunately, urine cannot be collected in cases due to perimortem voiding or decompositional changes.

Liver

Liver is the major organ for detoxification. Many toxic substances are present in the liver in higher concentrations than in the blood. It is easily collected and can be readily homogenized. Consequently, liver is used to supplement the blood concentration data, or may also be the only specimens available in cases of advanced putrefaction. Usually, the liver from deep within the right lobe is preferred to avoid possibility of gastric diffusion from stomach or from mesenteric circulation [13]. Liver should be finely diced and homogenized in water or a dilute buffer with a minimum water/buffer-to-liver ratio of 1 : 1. The usual method for extraction of biological fluid can be applied to liver homogenate, provided it is properly validated. However, the more complicated matrix present in liver may require additional cleanup to allow better instrumental detection. Unfortunately, the limited toxicological data from the literature limit the ability to interpret liver concentration data [14].

Vitreous Humor

Vitreous humor is the major fluid component of the eye, and is well protected inside the eyeball so that it is less subject to contamination and bacterial action and has little protein content. In conjunction with alcohol levels in blood and/or urine, quantification of alcohol in vitreous humor is useful to assist in distinguishing between alcohol intake before death and postmortem alcohol formation. In addition, this specimen is particularly useful for determining common antidepressants, digoxin, sodium, chloride, glucose, and metabolites related to renal functions (e.g., urea nitrogen, uric acid, and creatinine). Thus, it should be collected whenever possible.

Gastric Contents

As oral ingestion is a major route of drug administration, gastric contents are important to investigate potential poisoning and in case of overdose or acute poisoning, high concentration of drugs or poisons will be detected. In many cases of acute poisoning, undissolved capsules or tablets may be discovered through visual inspection of the contents, allowing relatively simple drug or poison identification. The total amount of a drug or poison present in the gastric contents is more important than its concentration. When supported by blood and/or tissue findings, a large quantity of the parent drug in the gastric contents as compared to the prescribed dose would indicate drug overdose. Therefore, the total volume of gastric contents should be measured. Recent intake of strong alkali or acids prior to death can easily be determined by pH measurement. In addition, an alkaline pH may also be due to the ingestion of CN and screening test for CN, such as Ferroin test [15], should be considered. The presence of a characteristic odor in gastric content can be a useful indicator for certain toxic substances. Some of the examples are listed in Table 2.

Other Specimens

Sometimes analysis of specimens other than those mentioned may be more appropriate. For instance, bile can be useful for screening of drugs and poisons as a number of drugs, such as morphine, benzodiazepines, and their glucuronide metabolites, ketamine, etc., are present in higher concentrations in bile than in blood. Bile can also be considered as an alternative screening specimen when urine is not available. Lung tissue is particularly useful in the analysis of volatile substances, such as hydrocarbons and other solvents or gases. Brain is useful for the detection of drugs such as antidepressants, narcotics, and halogenated hydrocarbons that act primarily on the central nervous system. In deaths due to chloroform poisoning, a high chloroform concentration can be found in the brain [16]. Kidney is useful in the investigation of heavy metal poisoning. The high concentration of metal deposited in kidney is often associated with structural damage that may be characterized histologically. As the growth rate of hair is approximately 0.6–1.4 cm per month, it provides a longer drug surveillance window, in a scale of weeks to months, than that of urine and blood. Hair is used

Table 2 Example of characteristic odor in relation to common toxic substances

Odor	Indication
Alcohol	Ethanol
Bitter or burnt almond	Cyanide
Garlic	Organophosphate insecticides
Antiseptic	Chloroxylenol

to evaluate prior exposure to heavy metals, such as arsenic, lead, and mercury, and is now extended to the analysis of a wide range of organic drugs and poisons to provide information on the chronic use or long-term exposure to toxic substances [17].

In case of serious putrefaction, body fluids and tissues including blood, urine, and liver may no longer be available. Under such circumstances, other alternative specimens like muscular tissue, hair, or bone should be considered for toxicology screening but the quantification of drugs and poisons in these samples will be of limited toxicological significance (*see* **Hair: Toxicology**).

Table 3 lists the recommended amount, if available, of specimens taken for analysis and their intended purposes.

All specimens collected should be stored in tightly sealed containers at low temperatures (usually below 4 °C for short-term storage during analysis and −20 °C for longer term storage). Except for blood, there is no special preservation required for specimens collected. For blood samples, one of the bottles collected should be preserved by addition of at least 2% w/v of sodium fluoride or equivalent to suppress the postmortem production of alcohol, γ-hydroxybutyrate, CN, and carbon monoxide and reduces hydrolysis of some drugs, such as cocaine to benzoylecgonine [6, 18].

Additional Information Required Prior to Analysis

Most drug- or poison-related deaths do not present characteristic features pathologically as those found in many disease-oriented deaths, such as cancers. Strategy for postmortem sampling and subsequent analysis should be based on a detailed knowledge about the case in respect of the presence of any foreign substances suspected to be related to the fatality. To facilitate toxicologists in devising the appropriate analytical methods, interdisciplinary discussions between professionals involved in the case (at least between forensic pathologist and toxicologist) prior to commencement of toxicological investigation is recommended. The following information should be provided if available:

1. Evidence(s), such as drug paraphernalia (e.g., syringe), poisons or medications, empty containers, and/or packaging inserts, suspected to be related to drugs or poisons found at the scene

Table 3 Minimum recommended amount of specimens taken for postmortem analysis

Specimen	Where to obtain (amount)	Principal uses
Blood	Peripheral (femoral or subclavian preferred) (2 × 10 ml)	Useful to confirm recent drug use and concentration, if a drug found, is useful for toxicological interpretation
	Peripheral (preserved using 2% w/v sodium fluoride or equivalent) (at least 2 ml)	Mainly for alcohol and for analysis of some drugs requiring preservation
Urine	(>10 ml)	Useful for broad class of drug screening but may not be useful on timing of drug administered
Liver	Deep within right lobe preferred (>100 g)	Useful solid tissue samples to supplement blood data especially when blood is not available but limited literature data available
Bile	(At least 5 ml or whole gall bladder)	Useful fluid for drug screening especially when urine is not available
Vitreous humor	(All available ~2–5 ml)	Useful in alcohol analysis and, if necessary, some other drugs (e.g., digoxin, glucose, urea nitrogen, uric acid, creatinine, and antipsychotic drugs)
Gastric contents	(All)	Indicative of recent drug administration
Hair	Identify distal and proximal end (>50 mg)	Drug use history provide information of drugs/poisons and metal exposure in scales of months
Lung	(50 g)	For volatile substances (such as H_2S and chloroform poisoning)
Brain	(50 g)	Brain may be useful in infant drug deaths or for volatile poison cases

should be seized as analysis of these items may be valuable for determining the subsequent type of analysis to be performed on the biological specimens. In some cases, household products, such as caustics, solvents, or pesticides, may also provide useful clues. While circumstantial evidence can provide hints, it can never substitute the analysis of body fluids or tissues because substance(s) found at the scene may not necessarily be connected to the death and often other substances are detected that are not obvious from the circumstances.

2. A comprehensive list of medications (especially recent medications prescribed) given to the deceased with relevant medical history should be provided.
3. Any physical abnormalities identified during autopsy may be indicative of intoxication or poisoning and a list of examples is given in Table 4 [6]. In this case, additional analysis targeting for the presence of possible toxic substances may be required.

Analytical Aspects in Postmortem Toxicology

Ingested substances are metabolized, being broken down or transformed into other species before they are excreted. Hence, identification of the original ingested material often involves considerable complications, for example, heroin is first metabolized into 6-acetylmorphine and then to morphine, while Δ^9-tetrahydrocannabinol (THC), an active ingredient in cannabis, is first converted to 11-hydroxy-Δ^9-tetrahydrocannabinol and then to 11-nor-Δ^9-tetrahydrocannabinol-9-carboxylic acid.

In addition, the active constituent in a regular dose ranging from grams or milligrams is diluted to a concentration usually in the range of micrograms or nanograms per milliliter of body fluids or per gram of tissue by way of dispersion throughout the body. The analytical method to be utilized must be capable, both in terms of sensitivity and specificity, of detecting the target substances at low concentrations and in complicated biological matrices (*see* **Postmortem Toxicology: Artifacts**).

Table 4 Useful findings related to toxic substances observable during autopsy

	Possible Indication
Color of skin	
Cherry red to bright red	Carbon monoxide or cyanide
Grayish to brownish	Nitrate, nitrite or aniline
Nasal/oral cavity	
Residues of powder or colored material	Intransal drug use (e.g ketamine, cocaine), ingestion of tablet or capsule residues
Oral cavity/ gastointestinal tract	
White, corrosive staining	Hydrochloric acid
Black-brown, corrosive staining	Sulphuric acid
Glass-like, reddish necrosis	Alkaline agents (e.g. sodium hydroxide)

Analytical Techniques

There are a wide variety of analytical techniques available for the analysis of toxic substances in biological specimens. The most common techniques used in modern toxicology laboratories include various immunoassays with different detection principles, color tests (e.g., Ferroin for CN, Marsh test for arsenic, etc.), instrumental chromatographic techniques using high-performance liquid chromatography (HPLC) and gas chromatography (GC) coupled with various detectors. To effectively apply chromatographic techniques, an extraction procedure is required to separate the intended drugs/poisons from biological matrices followed by reconstitution in appropriate solvents compatible with the requirements of the intended instrumentation.

Immunoassays. A number of immunoassays intended for antemortem analysis can also be used for postmortem analysis especially when urine is available. Immunoassay is based on the principle that the drug is detected by its ability to displace or block the binding of a fixed amount of labeled drug molecules present in the reagent. The label can be a fluorescent molecule (e.g., fluorescence polarization immunoassay (FPIA)), an enzyme (e.g., cloned enzyme donor immunoassay (CEDIA), enzyme multiplied immunoassay technique (EMIT), and enzyme-linked immunosorbent assay (ELISA)), a radioactive isotope (e.g., radioimmunoassay (RIA)) or other substance that can be detected by means of instrumentation. Some assays can distinguish between bound

and free labeled drug in a mixture are known as *homogenous immunoassay*, such as CEDIA, EMIT, and FPIA. For heterogenous immunoassays, such as ELISA and RIA, they require a washing step to separate the bound labeled complex from free labeled reagent prior to analysis. Thus, homogenous immunoassays are easier to be automated and less labor intensive than heterogenous ones. In general, immunoassays are fast, sensitive and, in homogenous immunoassays, direct detection can be achieved without sample purification. It is used for screening of common abused drugs, such as opiates, amphetamines, cocaine, cannabinoids, phencyclidine, and barbiturates. In addition, prescribed drugs, such as propoxyphene and tricyclic antidepressants, can also be screened with the use of specific reagents. For postmortem screening, antibodies having broad drug selectivity within a class of drugs (e.g., sympathomimetic amines) are preferred over those that are sensitive to a specific drug (e.g., methamphetamine) as it allows the screening of drugs within the same class. The sensitivity can be further increased by prior hydrolysis of glucuronide or sulfate conjugates of some drug classes, such as cannabinoids, opiates, and benzodiazepines.

Cutoff values, often applied to workplace drug testing, especially for abused drugs should be used cautiously in postmortem cases since the presence of low drug concentrations can be of forensic significance. In an acute drug-related death, for instance, the drug may not have sufficient time to be excreted into urine before death resulting in a low drug concentration in urine. False positives may occur, either from structurally related drugs or from metabolites of other drugs that are recognized by the antibodies. For instance, phenethylamine, a common putrefactive product from decomposed bodies, can cause a false-positive response to the amphetamines class test when using immunoassays. Cross-reactivity can also be due to chemicals with similar structures to the intended analytes. For example, pholcodine gives rise to a positive response to the opiates reagent of FPIA. Thus, for those samples, which give positive screening results, confirmation tests should be performed, preferably using chromatographic techniques with MS detection.

Extraction Techniques. The usual extraction techniques involve either liquid–liquid extraction (LLE) or solid-phase extraction (SPE).

LLE involves the selective partitioning of the compound of interest into one of two immiscible phases by a judicious choice of extraction solvents. Although there are some developments in SPE techniques, in recent years [19], the traditional LLE technique is still a common extraction method used in postmortem specimens. The method has the advantage of efficient extraction of drugs and poisons present in a wide concentration range and the absence of adsorption loss frequently associated with a solid surface. However, for some postmortem blood or tissues, problems associated with the formation of stable emulsion, variable extraction efficiency and endogenous interferences due to autolysis may occur. To extract acidic and basic/neutral drugs from biological fluids or tissues using LLE, separate extractions using appropriate organic solvent(s) with the addition of acidic and basic buffer, respectively, are required. Various methods using LLE for drug extraction in postmortem specimens show that there is a wide choice of solvents or mixture of solvents with similar extraction efficiency and selectivity [18, 20].

The mechanism of SPE based on the selective partition of one or more components between two phases, one of which is a solid sorbent while the second, mainly a liquid, is more complicated than LLE. Extraction is accomplished by adsorption of the analytes onto the solid sorbent followed by washing with an appropriate solvent to remove the unwanted matrix before eluting the analytes. Compared with LLE, SPE has the advantages of low solvent consumption, provision of cleaner extracts, and ease of automation and high extraction efficiency for certain specific drugs requiring a smaller sample volume. Thus, it provides an excellent alternative to the traditional LLE for the extraction of postmortem samples. Unlike LLE, untreated sample cannot be applied directly onto SPE. Pretreatment procedures, e.g., protein precipitation and centrifugation, may result in a significant loss of analytes due to adsorption or occlusion onto the precipitated constituents. Ion exchange resins have shown to be efficient in the extraction of acidic drugs [21]. On the other hand, mixed-mode SPE, being capable of extracting acidic, basic, and neutral drugs, is suitable for general unknown screening (GUS) [22] (*see* **Toxicology: Initial Testing**).

Chromatographic Techniques. GC and HPLC, coupled with various detectors, are common instrumentations to screen for a wide range of organic toxic

substances. A combination of mass spectrometry with either GC or, more recently, HPLC is the definitive technique to establish proof of structure of unknown substances. With the extensive development of commercial MS technology at an affordable cost, gas chromatography mass spectrometry (GC-MS) and high-performance liquid chromatography mass spectrometry (HPLC-MS) (or liquid chromatography mass spectrometry (LC-MS)) become increasingly popular tools employed in toxicological analyses.

GC is one of the most frequently used techniques for separating, identifying, and quantifying a parent drug and its metabolites, other coadministered drugs and endogenous compounds. This technique can be coupled with various detectors from the more universal flame ionization detector (FID) to specific detectors such as electron capture detector (ECD) and nitrogen phosphorus detector (NPD).

FID is useful for the detection of alcohol, other volatile organic compounds and many other organic drugs and poisons. NPD is sensitive to nitrogen- and phosphorus-containing compounds and is useful for the detection of drugs and poisons, such as antidepressants, antipsychotic drugs, benzodiazepines, opiates, cocaine and its metabolites, organophosphorus insecticides, etc. The use of FID and NPD either alone [23–25] or in a combination of both [26] has been shown to be useful for GUS of organic drugs and poisons. ECD is particularly sensitive to halogenated compounds (e.g., chlorinated insecticides), nitriles (e.g., CN) or nitrogen-containing compounds (e.g., benzodiazepines, nifedipine, and zopiclone). Unfortunately, the above detectors can only provide retention time data without any additional information for structural identification. Thus, more sophisticated techniques, such as GC-MS or LC-MS, are recommended for confirmation (*see* **Confirmation Testing: Toxicology**).

The high separation power of capillary GC coupled with a highly selective MS detector has been currently regarded as the "gold standard" in GUS for drugs and poisons. The availability of a well-established and standardized ionization technique (electron impact (EI) at 70 eV) has facilitated the construction of large databases of reference mass spectra for library search and many useful spectral libraries relevant to toxicological screening are now available. However, GC is not suitable for direct analysis of polar compounds although derivatization can partly solve the problem and mass spectral libraries containing a practically complete coverage of trimethylsilyl (TMS) derivatives are available commercially [27].

HPLC is capable of dealing with the analysis of a wide range of both volatile and nonvolatile compounds. The reversed phase mode column is at present the most common separation method applied in toxicological screening. Unlike GC, derivatization is not necessary for the analysis of polar and thermolabile compounds and this advantage certainly promotes HPLC as a better alternative for compounds not amenable to GC. Detection is often aided by diode-array detectors (DAD), which acquire UV–visible spectra continuously during a chromatographic run and the chromatogram is extracted and plotted at preselected wavelengths. The combined technique of high-performance liquid chromatography diode-array detector (HPLC-DAD) is another suitable technique for GUS of compounds, covering a wide range of polarity, stability, and molecular masses. Since metabolic transformation, in many cases, does not affect the ultraviolet (UV) chromophores of the molecule, one of the added values of this technique in GUS for drugs and poisons is that the low selectivity of DAD facilitates the detection of metabolites. In addition, compounds belonging to the same class with similar chemical structures often display similar absorbance patterns allowing unknown compounds not previously identified are tentatively assigned for further analysis. However, one of the drawbacks of HPLC-DAD is that the resolution of HPLC is usually inferior compared with GC, and UV spectroscopy is less sensitive than that of MS. In addition, for compounds possessing weak UV–visible or without any characteristic UV–visible absorbency, identification using HPLC-DAD is difficult in terms of both sensitivity and specificity.

LC-MS is useful for the analysis of compounds that are not amenable to GC-MS. These compounds include lysergic acid diethylamide (LSD), glucuronide conjugates, such as morphine-3 or 6-glucuronide or for some very potent or large molecules that other techniques are not sufficiently sensitive for detection (e.g., colchicines, cardiac glycosides such as digoxin and digitoxin, β-agonists such as salbutamol and terbutaline) [26].

The most common ionization mode used in LC-MS techniques is atmospheric pressure ionization (API), which mainly comprises of different versions of electrospray ionization (ESI) and atmospheric

pressure chemical ionization (APCI) interfaces. With simpler sample preparation, LC-MS is extended to certain analyses originally performed by the less-specific HPLC-DAD or even GC-MS. For these reasons, LC-MS, which combines an almost universal separation process with the most specific and sensitive type of detector, has become a promising alternative approach to GC-MS and HPLC-DAD in toxicological analysis [28–30].

However, there are several drawbacks in the use of LC-MS. First, the ion-suppression effect, especially when operating in the ESI mode, is a well-known problem: signal of the intended analyte is often suppressed because of the presence of coeluting compounds/interferents so that the intended analyte may be underestimated or even overlooked. Thus, the effect of ion suppression on the signal of the intended analyte should be cautiously evaluated before use. If ion suppression does occur, changes in experimental conditions, such as the sample cleanup method, chromatographic conditions relating to the mobile phase, the elution column, and the internal standard, should be considered. Moreover, only certain volatile buffers and mobile phase can be used for LC-MS to be compatible with the MS requirement. Thus, the separation efficiency is often inferior compared with HPLC-DAD although it may not be a problem for tandem MS because separation of analytes can be made in the MS/MS mode.

Systematic Toxicological Analysis

The usual practice in toxicological examination begins with the preliminary identification of alcohol and screening of a wide spectrum of acidic, neutral, and basic organic drugs or poisons. If a toxic substance(s) is detected, confirmatory and, if necessary, quantitative testing has to be performed. In general, a positive identification is achieved using at least two independent analyses and preferably based on different analytical principles. Using GC-MS or LC-MS, confirmation and quantification can be simplified into one single analysis. Quantification of drugs in blood, liver, and gastric content as dictated by the case, provides more meaningful interpretative information. Reference concentrations of many compounds in blood in therapeutic, toxic, and even fatal levels have been published [7–10]. Although limited, useful references are also available for some compounds in liver [14]. It should be noted that while substances found in excretory fluids such as urine or bile are useful qualitatively, quantification of drugs and poisons in these fluids usually has limited interpretative values.

A comprehensive and systematic analysis for the presence of chemical substances of toxicological significance is termed systematic toxicological analysis (STA). There are thousands of potentially harmful substances ranging from poisonous gases (e.g., carbon monoxide and hydrogen cyanide (HCN), and hydrogen sulfides), food (e.g., ethanol), deadly poisons (e.g., CN salts and arsenic), abused drugs, pesticides, toxins from natural sources, a wide variety of prescribed drugs and even household products, etc. It is impossible to design a single analytical scheme to cover all these substances with distinctly different chemical and physical properties. Screening for a wide scope of drugs and poisons is, however, possible by grouping substances of similar properties for analysis. The most effective strategy includes a series of standard general screening procedures supplemented by as many special methods as required. A combination of immunoassays with chromatographic techniques is usually employed to detect a wide range of substances. Immunoassays detect classes of drugs with similar structures while chromatographic techniques detect large groups of drugs with similar extraction properties, polarity, and detection characteristics.

General toxicological screening usually involves the following tests:

1. alcohol determination;
2. immunoassay screening;
3. GUS of organic drugs and poisons;
4. other specific tests as required such as carboxyhemoglobin, CN, etc.

Table 5 shows the suggested screening tests based on different case nature. Since it is virtually impossible to screen for all toxic substances in every case, a rational selection of case-specific analysis in addition to STA will be necessary. In general, the additional analysis to be conducted is primarily based on the information provided or specific requests made by the relevant parties such as pathologists and the police.

An example of an analytical scheme used for general toxicology screening and scope of analytes likely to be detected is shown in Figure 1.

Alcohol Determination. Alcohol or ethyl alcohol (ethanol) is the most common drug found in

Table 5 Suggested analysis for sample collected[(a)]

Type of cases	Test conducted	Specimen tested
General cases	Alcohol	Blood, urine, and vitreous humor
	Immunoassay drug screening	Urine
	General organic drugs/poisons screening	Blood (urine)
General cases (blood not available)	Alcohol	Urine and vitreous humor
	Immunoassay drug screening	Urine
	General organic drugs/poisons screening	Liver/cavity fluid
Drug or poison suspected cases	Plus general organic drugs/poisons screening	Gastric content
Fire death	Plus carboxyhemoglobin, cyanide	Blood
Heavy metal poisoning	Plus metal analysis	Blood, kidney, and liver
Gaseous and volatile organic related	Plus sulfide	Blood, lung, and brain
	Plus volatile organic	Blood, lung, and brain

[(a)] Specific test(s) not covered may be added if deemed appropriate.

postmortem toxicology cases. In particular, ethyl alcohol is one of the leading causes of death by poisoning. In contrast to alcohol determination in living subjects, which is generally more straightforward, analysis of body fluids taken from cadavers is more likely to be contaminated with volatile substances, such as methanol and formaldehyde used in embalming processes, and abnormal metabolic products, such as acetone resulting from fasting or diabetic ketoacidosis. Therefore, techniques for analyzing postmortem alcohol are recommended to allow separation of most, if not all, low-boiling compounds eluting in the same range as ethyl alcohol [31].

Headspace GC coupled with FID is usually the method of choice and it can be used for simultaneous analysis of methanol, acetaldehyde, ethanol, isopropanol, and acetone using either *n*-butanol or *n*-propanol as an internal standard. A combination of blood (preferably femoral region), vitreous humor, and urine, if available, should be used for alcohol analysis to aid in the interpretation of the state of absorption and to avoid misinterpretation of blood alcohol concentrations due to diffusion of undigested alcohol from the stomach, and postmortem alcohol production due to bacterial action. The presence of significant amount of alcohol in blood together with the absence of alcohol in urine and vitreous humor cast doubt in the ingestion of alcohol prior to death and is indicative of endogenous alcohol production (*see also* **Alcohol: Analysis**).

Immunoassay Screening. Screening for drugs by immunoassays in urine is commonly pursued for the main classes of abused drugs (Figure 1), which include amphetamines, benzodiazepines, cannabinoids (metabolite of cannabis), benzoylecgonine (metabolite of cocaine), and opiates (morphine and codeine). The scope for drug screening can be extended to include opioids (e.g., methadone), barbiturates, and tricyclic antidepressants (e.g., amitriptyline/nortriptyline).

As urine is not always available in postmortem cases, screening for drugs and poisons, especially abused drugs, in blood or plasma may be considered and may even be desirable to establish what substances may be affecting the person. Blood, plasma, or even specimens such as bile or liver homogenate can be screened by the urine-based immunoassay test kits, but pretreatment with solvent or protein precipitation are required and validation should be made before use. In addition, techniques intended for analysis of blood, such as RIA and ELISA can be used on postmortem samples, and they tend to be more sensitive than those kits designed for detecting drugs in urine but modified for testing blood. Screening of potent drugs like digoxin or structurally related cardiac glycosides, such as bufadienolides present in Chansu (a remedy from toad venom), can be effectively performed in plasma using a digoxin reagent by various immunoassay techniques [32].

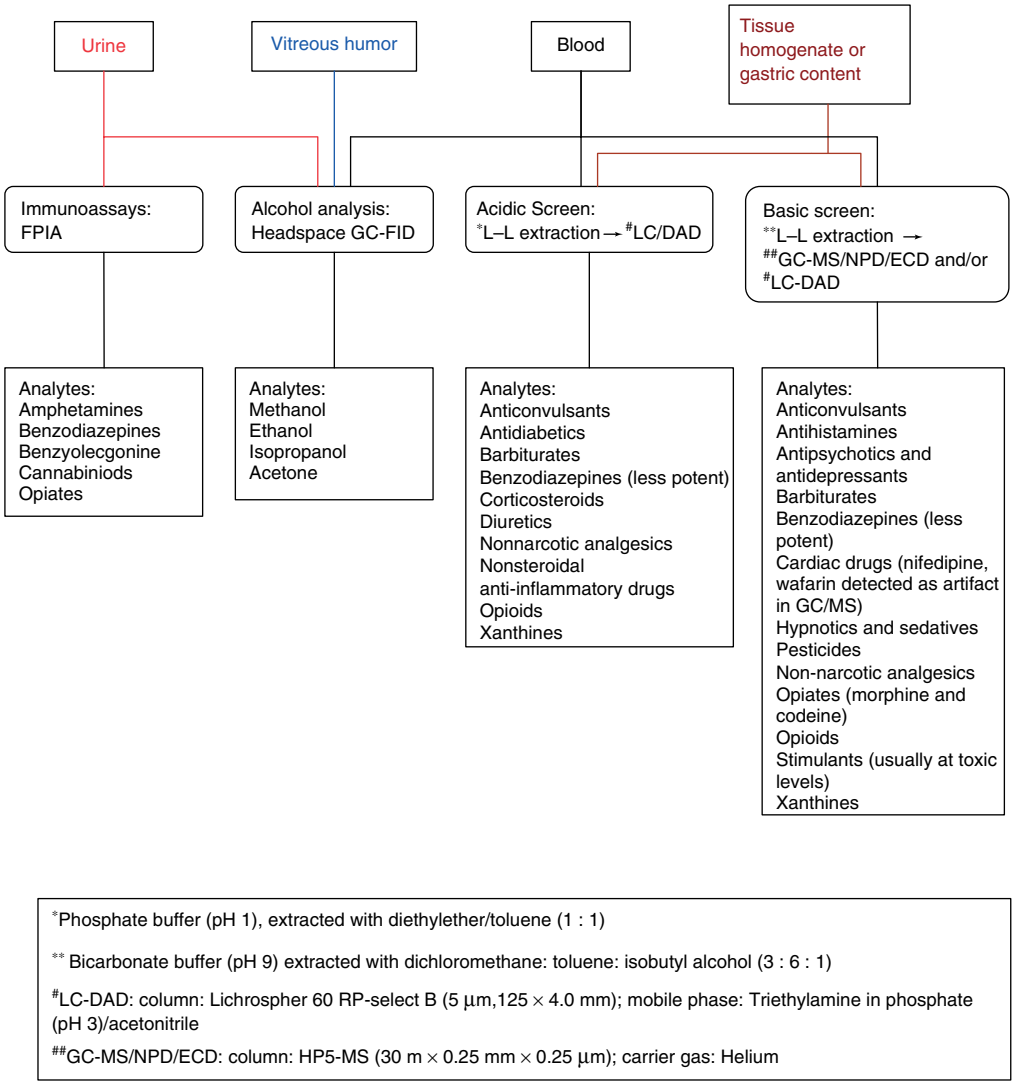


Figure 1 An example of general toxicology screening scheme

General Unknown Screening (GUS) for Organic Drugs and Poisons. As shown in Figure 1, acidic drugs can be extracted by adding an acidic buffer such as phosphate buffer (pH 1) to the samples followed by extraction with toluene: diethyl ether (1 : 1). Other methods, such as protein precipitation of blood with acetonitrile [33] or ammonium chloride salting-out of blood and other tissues with ethyl acetate followed by a washing step with hexane [34] can also be employed for the analysis of acidic and some neutral drugs. Analysis is usually made by HPLC-DAD using a reverse-phase column in conjunction with a mobile phase at acidic pH with gradient elution. Identification is accomplished by automatic library search based on a preinstalled, commercially available, or an in-house developed UV–visible spectral library. The retention times are useful to aid in identification but it should be established in-house using authentic standards according to the type of column and the mobile phase used. Nonnarcotic analgesics (e.g., paracetamol, salicylic acid, mefenamic acid, and sometimes propoxyphene), nonsteroidal anti-inflammatory

drugs (e.g., celecoxib, naproxen, and ibuprofen), diuretics (e.g., furosemide and hydrochlorothiazide), anticonvulsants (e.g., carbamazepine, phenobarbital, and phenytoin), antidiabetics (e.g., glicazide), opioids (e.g., methadone), barbiturates (e.g., secobarbital), corticosteroids (e.g., hydrocortisone), the less potent benzodiazepines (e.g., midazolam, diazepam, and sometimes estazolam), and xanthines (e.g., caffeine and theophylline) can be detected using HPLC-DAD after acidic extraction.

Drugs and their metabolites usually have similar UV–visible spectra because metabolic transformations, such as desmethylation and hydroxylation, do not usually affect the chromophores of the drugs. Yet, a drug and its metabolites usually have different retention times (e.g., midazolam and α-hydroxymidazolam, clomipramine, and desmethylclomipramine). Although it may be useful to extend the detection capability to drug metabolites even without the authentic standards, one should exercise caution to avoid misinterpretation when a peak slightly different in retention time and UV–visible spectrum from the parent drug is observed. As many drugs and poisons do not have characteristic UV–visible spectra, retention time becomes the main identification parameter. Thus, whenever standards of parent drugs and/or their metabolites are available, their retention times should be established in-house and their authenticity preferably confirmed by other techniques, such as GC-MS.

Extraction of basic drugs is accomplished by a carbonate buffer (pH 9) followed by extraction with a dichloromethane/toluene/isobutyl alcohol (3 : 6 : 1) mixture. Other combinations of alkaline buffers and extraction solvents such as borate buffer (pH 8.5) extracted with dichloromethane/isopropanol (9 : 1) [35], and using SPE with suitable cleanup procedures compatible with GC analysis [18–20] are also applicable. The use of more than one type of detectors (e.g., NPD and MS) coupled to GC would definitely give a greater coverage of possible drugs and poisons. The extracted sample could be analyzed sequentially by injection into two GCs equipped with different types of detectors or by splitting the effluent of the sample from one GC into two different detectors for simultaneous detection.

Examples of drugs and poisons found in the basic fraction include hypnotics and sedatives (e.g., zopiclone and zolpidem), benzodiazepines (e.g., diazepam, midazolam, estazolam, and bromazepam), antihistamines (e.g., chlorpheniramine, brompheniramine, and promethazine), antipsychotics and antidepressants (e.g., amitriptyline/nortriptyline, cyclobenzaprine, sertraline, clomipramine, and trihexylphenidyl), opiates (e.g., codeine, acetylmorphine, and sometimes monomorphine), opioids (e.g., methadone and meperidine), anticonvulsants (e.g., carbamazepine, phenobarbital, and phenytoin), and pesticides (e.g., melathion, dimethoate, and tetramethylene disulfotetramine). Stimulants such as amphetamines and cocaine may be detected when present at high concentrations. In addition, artifacts due to the presence of certain cardiac drugs such as warfarin and nifedipine may be encountered.

Some basic drugs that are too polar or thermally labile cannot be detected by GC and are not acid extractable for HPLC-DAD analysis. In such case, the basic extract can be analyzed using HPLC-DAD in addition to GC. Typically, the method permits the detection of cardiac drugs (e.g., warfarin, metoprolol, propranolol, dipyridamole, and amiodarone), anti-inflammatory drugs (e.g., ofloxacin), and herbal ingredients (e.g., tetrahydropalmatine).

Alternatively, derivatization prior to GC analysis can increase the volatility and hence thermal stability for those drugs that are not easily detected by GC or GC-MS [36]. Among the available derivatization procedures, trimethylsilylation is the most common method because of its versatility for derivatizing many different functional groups such as hydroxyl, carboxyl, amidic, and some amine groups under relatively mild conditions. Furthermore, the additional mass gain by silylation improves the specificity of the mass spectral information [37]. Typical examples include morphine and sympathomimetic amines. Other derivatized agents such as trifluoroacetylation for basic and neutral drugs, and extractive methylation using methyl iodide or formation of ethereal diazomethane for acidic drugs are also applicable [38, 39].

Some compounds do not exhibit characteristic mass spectra in GC-MS. In this case, unambiguous identification can be difficult if mass spectral matching is solely relied on because many compounds can give seemingly high matching scores. Typical examples include amino-containing compounds with predominant base peaks at m/z 44 ($C_2H_6N^+$) (e.g., amphetamine, methylenedioxyamphetamine MDA), 58 ($C_3H_8N^+$) (e.g., doxepin, cyclobenzaprine), 72 ($C_4H_{10}N^+$) (e.g., methadone, promethazine), or 98

($C_6H_{12}N^+$) (e.g., trihexylphenidyl, thioridazine) and usually associated with very low intensity fragments in the high mass range. If these compounds are well separated chromatographically, they can only be identified on the basis of their differences in retention time. For those substances having similar retention time in GC, however, another technique should be considered for confirmation.

The application of LC-MS to GUS is still not extensive in spite of its versatile analytical capability as well as sensitive and specific detection of a wide range of drugs and poisons. One of the problems is the soft ionization of LC-MS, which produces mass spectra that are not compatible with those generated by EI at standardized 70-eV ionization potential. Hence, the very large libraries of standardized EI spectra of chemicals, drugs, poisons, and their metabolites applicable to GC-MS are not valid for LC-MS. Thus, a totally different strategy for mass spectral identification of compounds is required for LC-MS.

Many efforts have been made to build up large mass spectral database applicable for GUS of compounds of toxicological interest on different types of LC-MS and promising progress has been demonstrated. Identification has been based on single quadrupole, ion trap, triple quadrupole, hybrid linear ion trap, and time of flight (TOF) [28–30]. The hybrid linear ion-trap LC-MS/MS showed promising development for building up a reference MS/MS library for the same type of instrument. This LC-MS/MS technique is unique in that the third quadrupole of the triple quadrupole MS can be operated in either a standard quadrupole MS for multiple reaction monitoring (MRM) experiment or as linear ion trap to produce highly sensitive enhanced product ion (EPI) scan in an information-dependent acquisition (IDA) experiment [40, 41]. The first detection step involves IDA survey scan containing a number of preselected MRM target drugs and metabolites where ions are accumulated and then filtered in the third quadrupole. In case of a signal above a preset intensity threshold is detected for an MRM transition, the EPI scan of the precursor ion is triggered to yield product ion mass spectra at various preselected collision energies. Finally, the resulting EPI mass spectra are then searched against a prebuilt mass spectral library for identification of drugs present in the sample. Alternatively, LC-MS/TOF has provided a novel approach for comprehensive drug screening. The use of LC-MS/TOF method has the advantages to provide a relatively high mass accuracy (~5 ppm) with reasonable resolution (5000–10 000 full width at half maximum). In this approach, a library is established to contain toxicological relevant compounds that consist of molecular formula and calculated monoisotopic accurate masses. Identification of drugs/metabolites was based on their accurate mass, retention time if a reference material is available, and drug metabolite patterns. Furthermore, the matching of theoretical and measured isotopic patterns of a compound introduces an additional parameter to allow unambiguous identification of compounds present in the sample. This approach allows substance identification even without any reference standards and retention time data [42] (*see* **Confirmation Testing: Toxicology**).

Other Specific Tests

When analysis of specific types of drugs or poisons that are not covered under the general toxicology screening scheme is required, additional tests will have to be performed. Figure 2 shows some of the examples on the analysis of certain specific types of drugs and poisons.

Drug-detection Techniques

Acid Back-Extraction. Although many basic drugs with amine functional groups such as those basic psychotropic drugs and antihistamines can be detected using aforementioned basic screening, an additional acid back-extraction after basic extraction can further improve GC-MS detection by producing a cleaner extract. Examples of psychiatric drugs include tricyclic antidepressants, phenothiazines antipsychotic drugs, tetracyclic antidepressants, butyrophenones, and serotonin reuptake inhibitors. In addition, better detection of other drugs like antiparkinson drugs (e.g., trihexylphenidyl) can also be achieved using this method (Figure 2).

Ketamine, amphetamines, and their analogs can be detected at moderate-to-high level in the basic extract using GC-MS/NPD (Figure 1). To increase the detection sensitivity down to therapeutic or even subtherapeutic levels, a simple acid back-extraction cleanup after basic extraction with diethyl ether (Figure 2) coupled with LC-MS/MS (ion-trap MS) analysis can be considered.

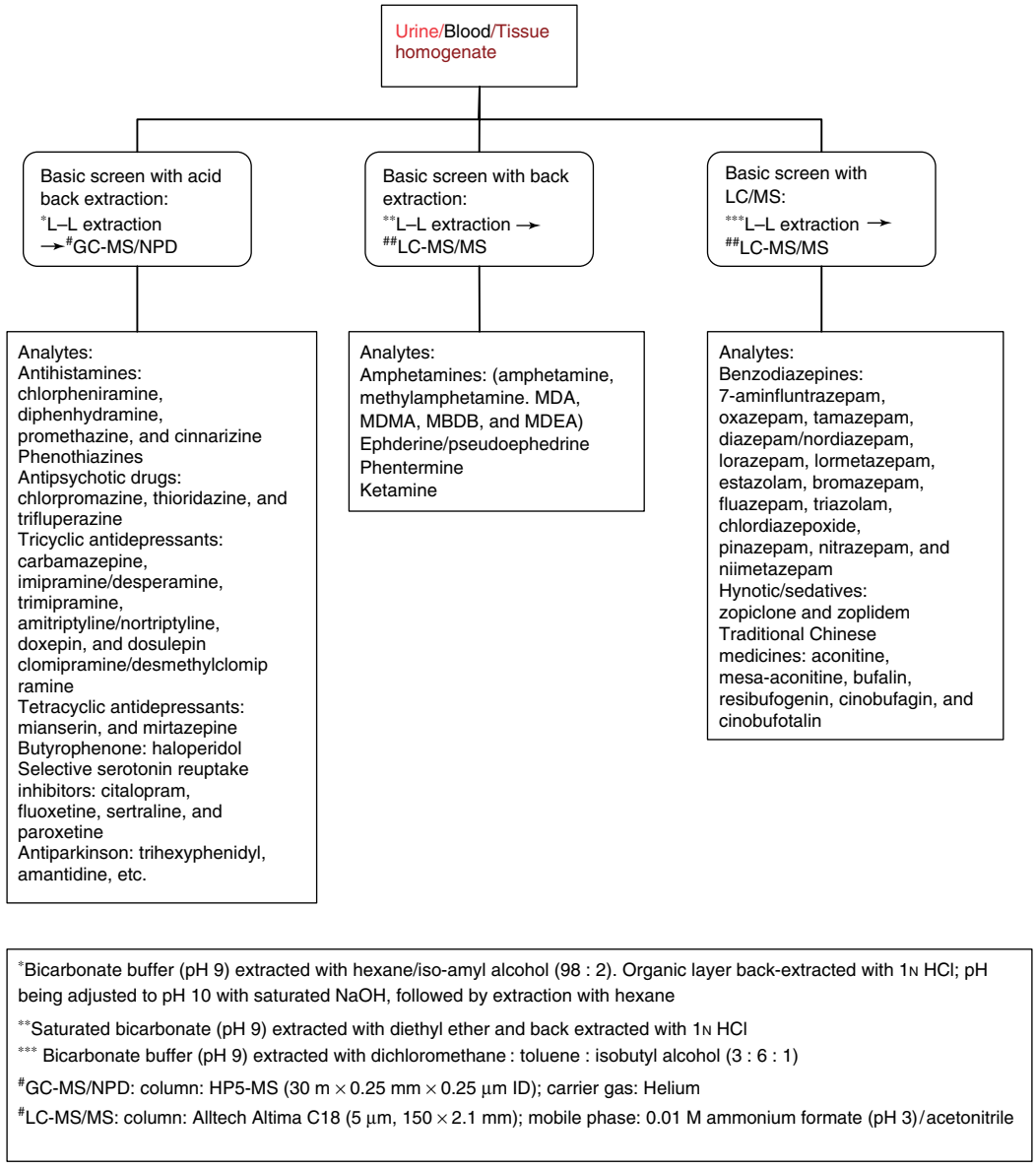


Figure 2 Examples of additional specific drug tests

Basic Extract Analyzed by LC-MS. LC-MS is complementary to the analysis of nonvolatile and thermally labile drugs/poisons that are not amenable to GC or GC-MS. In addition, LC-MS provides a superior sensitivity compared with the HPLC-DAD method, a good example being the LC-MS analysis of benzodiazepines: a structurally diverse class of pharmaceuticals. It is available as prescribed drugs with some of them (e.g., diazepam, midazolam, and nimetazepam) having been widely abused. Benzodiazepines and their metabolites represent one of the most common drug types found in postmortem specimens. While the relatively less potent benzodiazepines (e.g., midazolam and diazepam/nordiazepam) can be screened by the general screening procedures, many

potent or thermally unstable benzodiazepines (e.g., lorazepam, triazolam, and oxazepam) cannot be readily detected by standard screening techniques (Figure 1). To cater for a more comprehensive screening of benzodiazepines, the basic fraction from general toxicology screening can be reconstituted into aqueous methanol and subjected to LC-MS analysis (Figure 2). The same extract can also be used for the analysis of a wide range of targeted basic/neutral drugs and poisons including traditional Chinese medicines (e.g., aconitine and mesa-aconitine) and toxic ingredients found in Chan Su (e.g., bufalin, resibufogenin, cinobufagin, and cinobufotalin).

Analysis of Poisoning by Small Molecules. Carbon monoxide (CO), CN, and sulfide are well-known small and highly toxic molecules. Analysis of these molecules should be considered whenever poisoning due to these compounds is suspected.

Carbon Monoxide (CO). Determination of carbon monoxide poisoning will be required when a known source of CO, such as coal gas, burnt charcoal, or automobile exhaust, is located at the scene. On the other hand, it would be of forensic interest to decipher whether a fire victim had died because of CO poisoning, or had already died before the fire broke out. The saturation ratio of carboxyhemoglobin (COHb) in blood is determined by (i) simultaneous spectrophotometric (e.g., CO-Oximeter) measurement of COHb level and total hemoglobin in blood [43] or (ii) analysis of COHb through the release of CO by adding saponin and potassium ferricyanide to blood, followed by catalytic conversion of CO to methane, which is quantified by gas chromatography flame ionization detector (GC-FID) [44]. Analysis of COHb should be made using blood samples; other biological fluids, such as pleural effusion, are not recommended because a considerable amount of postmortem CO can possibly be generated in these samples by bacterial action on hemin [45].

Hydrogen Sulfide (H_2S) and Its Metabolites. Analysis is performed when H_2S poisoning is suspected, usually in industrial accidents, sewers, or ship holds where H_2S poisoning together with oxygen deficiency is suspected. Analysis is mainly based on the detection of H_2S or its metabolite, thiosulfate. Sulfate is also produced due to H_2S exposure but endogenous levels of sulfate in blood and urine are relatively high making it not an analyte of choice. GC-MS analysis can be performed for the ionized form of H_2S after derivatization using pentafluorobenzyl bromide [46]. The pH of the derivatized mixture is made acidic to suppress the production of sulfide in the blood due to decomposition of sulfur-containing compounds, such as cysteine. A similar derivatization procedure prior to GC-MS method can also be used for the analysis of thiosulfate with the exception that tetradecyldimethylbenzylammonium, a phase-transfer catalyst used in H_2S determination is substituted by ascorbic acid/sodium chloride [46, 47].

Cyanide (CN). Potassium or sodium salts of CN are used in metallurgy and electroplating industries and are relatively easy to obtain. Thus, its involvement in suicide and even homicide is not uncommon in some parts of the world. In addition, incomplete combustion of nitrogen-containing compounds such as urethane at fire scenes can produce HCN as one of the poisonous gases. Thus for fire victims, in addition to measuring the COHb saturation, CN levels in blood should also be determined. Inhalation of HCN or ingestion of CN can be fatal through inhibition of cytochrome oxidase causing cellular anoxia. CN in blood can be determined by automated headspace gas chromatography electron capture detector (GC-ECD). Blood is acidified with sulfuric acid in the presence of silver sulfate to produce HCN, which is diffused into the headspace to react with chlorine produced by chloramine-T to form cyanogen chloride, which is analyzed by GC-ECD [48].

Volatile Analysis. In addition to the analysis of alcohols, determinations of other volatile chemicals including various organic solvents are also of forensic significance. Sudden death due to volatile substance abuse (VSA) is not uncommon particularly for an inexperience user because controlling of dose is often difficult. Solvent from thinners (e.g., toluene and xylene), halogenated solvents (e.g., chloroform and dichloromethane), hydrocarbons (both aliphatic and aromatic) such as gasoline and kerosene, and fuel gas (e.g., butane) are commonly abused substances. Apart from VSA, analysis of volatile substances may also be required in certain circumstances. For instance, chloroform is commonly used for industrial purposes such as solvent and extracting reagents, and death

can occur accidentally and sometimes, in homicide cases, with the intent to incapacitate the victims. In addition, analysis for ingredients present in kerosene and gasoline may be useful in fire death cases.

Most volatile substances are stable in blood. The specimens should be stored at low temperature (i.e., less than 4 °C) in tightly sealed glass container, preferably with anticoagulant such as heparin. Analysis of tissues such as brain and lung may prove useful since high concentrations of volatile substances may be detected.

Headspace GC-FID is the method of choice for volatile chemical analysis. If structural identification is required, GC-MS should be employed. The method used for alcohol determinations can be extended for the analysis of other volatile chemicals using same columns (e.g., Elite-BAC1 or BAC-2) with some modification of experimental settings. Examples of compounds that can be analyzed are hexane, dichloromethane, chloroform, hexane, toluene, xylene, diethylether, and ethylacetate. For the analysis of low-boiling hydrocarbons, such as methane, ethane, propane, butane, and pentane, specific columns (e.g., GC-GASPRO) intended for analysis of volatile compounds can be used.

Heavy Metal Analysis. Although many metals are known to cause toxic effects, only a few are regarded as important toxic hazards: these include arsenic, lead, cadmium, thallium, and mercury. In addition, lithium is a psychiatric drug used for the treatment of manic-depressive disorder. There are many methods available for metal analysis in biological specimens, such as electrochemical, atomic absorption, and flame emission spectrophotometry, inductively coupled plasma coupled with either emission spectroscopy (ICP-AES) or inductively coupled plasma coupled with mass spectrometer (ICP-MS). ICP-MS after digestion of the biological specimens with concentrated mineral acid, such as nitric acid, is recommended. It is because ICP-MS allows a simultaneous screening of both metals and nonmetals as well as selective quantification of a single element with low detection limits. The availability of stable isotopes for most of the metal further enhances the accuracy of metal quantification by isotope dilution method.

Quality Systems and Assay Validation

A proper quality assurance program is an essential element for any analytical laboratory in order to ensure that the results generated are accurate, reliable, and traceable. It is even more important for a forensic toxicology laboratory since the results will be closely scrutinized in the courts of law. A quality manual pertaining all policies and procedures relevant to the reliability and traceability of the analytical results should be clearly written. Criteria affecting the quality of analytical results include the quality of materials used (such as reference standards, reagents, and chemicals), analytical methods adopted such as procedures and instrumentations used and their validity, sampling, and the chain of custody.

All analytical methods used must be properly validated. Suitable internal standards should be used in chromatographic assays, so that any systematic errors affecting the analyte can be compensated for by the internal standard. Compounds with chemical structures similar to the targeted analyte (e.g., deuterated analog) are preferably selected. For a batch of qualitative analyses, a negative control and a control representative of the analytes should be included. Any possible interfering factors that might adversely affect the analysis should be indicated. Quantitative analytical methods must be validated by determining the limit of detection, linearity range, precision, accuracy, and selectivity.

References

[1] Bertomeu-Sánchez, J.R. & Nieto-Galan, A. (2006). *Chemistry, Medicine And Crime: Mateu J B Orfila (1787–1853) And His Times*, Science History Publications, Sagamore Beach.

[2] Innes, B. (2000). *Bodies Of Evidence*, Reader Digest.

[3] Marsh, J. (1836). Account of a method of separating small quantities of arsenic from substances with which it may be mixed, *Edinburgh New Philosophical Journal* **21**, 229–236. The Marsh test is performed by treating the sample with sulfuric acid and arsenic-free zinc in a glass apparatus (such as test tube). The presence of arsenic is indicative by formation of silvery-black deposit on a ceramic bowl holding above the glass apparatus.

[4] Fenton, J.J. (2002). *Toxicology: A Case-Oriented Approach*, CRC Press, Boca Raton.

[5] Drummer, O.H. & Odell, M. (2001). *The Forensic Pharmacology of Drugs of Abuse*, Arnold Publishers.

[6] Skopp, G. (2004). Preanalytical aspects in post-mortem toxicology, *Forensic Science International* **142**(2–3), 75–100.

[7] Baselt, R. (2004). *Cravey R Disposition of Toxic Drugs and Chemicals in Man*, 7th Edition, Chemical Toxicology Institute, Foster City.

[8] Moffat, A.C., Osselton, M.D., Widdop, B. & Galichet, L.Y. (eds) (2004). *Clarke's Analysis of Drugs and Poisons*, 3rd Edition, Pharmaceutical Press.

[9] TIAFT (The International Association of Forensic Toxiologists), *TIAFT reference blood level list of therapeutic and toxic substances 18-10-2005*, (2005). http://www.tiaft.org/
tmembers/ttvidx.html (Electronic version).

[10] Winek, C.L., Wahba, W.W., Winek Jr, C.L. & Winek, T.W. (2001). Drug and chemical blood-level data 2001, *Forensic Science International* **122**(2–3), 107–123.

[11] Barnhart, F.E., Bonnell, H.J. & Rossum, K.M. (2001). Postmortem drug redistribution, *Forensic Science Reviews* **13**(2), 102–129. Postmortem redistribution (PMR) refers to the change in drug concentrations after death leading to differences in drug concentrations of postmortem blood samples taken from different sites. PMR arises from incomplete distribution of the drugs and/or poisons at the time of death, passive release or from vascular pathway from major organs such as gastrointestinal tract, liver, lungs, and myocardium, and later on cell autolysis and the putrefactive process.

[12] Moriya, F. & Hashimoto, Y. (2001). Post-mortem diffusion of drugs from bladder into femoral venous blood, *Forensic Science International* **123**(2–3), 248–253.

[13] Pounder, D.J., Adams, E., Fuke, C. & Langford, A.M. (1996). Site to site variability of post-mortem drug concentrations in liver and lung, *Journal of Forensic Sciences* **41**(6), 927–932.

[14] Musshoff, F., Padosch, S., Steinborn, S. & Madea, B. (2004). Fatal blood and tissue concentration in more than 200 drugs, *Forensic Science International* **142**(2–3), 161–210.

[15] Schilt, A.A. (1958). Colorimetric determination of cyanide, *Analytical Chemistry* **30**, 1409–1411. Ferroin test is performed by the addition of two drops of Ferroin reagent, 1,10-phenanthroline ferrous sulfate followed by the addition of 2 ml of chloroform. A positive result is characterized by the formation of a purple color in the chloroform (lower) phase.

[16] Gettler, A.O. & Blume, H. (1931). Chloroform in the brain, lungs, and liver, *Archives of Pathology* **11**, 554–560.

[17] Kintz, P. (2004). Value of hair analysis in postmortem toxicology, *Forensic Science International* **142**, 127–134.

[18] Drummer, O.H. (2002). Gerostamoulos j post-mortem drug analysis: analytical and toxicological aspects, *Therapeutic Drug Monitoring* **24**(2), 199–209.

[19] Franke, J.P. & de Zeeuw, R.A. (1998). Solid-phase extraction procedures in systematic toxicological analysis, *Journal of Chromatography. B, Biomedical Sciences and Applications* **713**(1), 51–59.

[20] Drummer, O.H. (1999). Chromatographic screening techniques in systematic toxicological analysis, *Journal of Chromatography B: Biomedical Sciences and Applications* **733**(1–2), 27–45.

[21] Koves, E.M. (1995). Use of high performance liquid chromatography – diode array detector in forensic toxicology, *Journal of Chromatography. A* **692**(1–2), 103–119.

[22] Chen, X.H., Franke, J.P., Wijsbeek, J. & de Zeeuw, R.A. (1993). Study of lot-to-lot reproducibilities of bond elut certify and clean screen DAU mixed mode solid phase extraction columns in the extraction of drugs in whole blood, *Journal of Chromatography* **617**, 147–151.

[23] Lo, D.S., Chao, T.C., Ng-Ong, S.E., Yao, Y.J. & Koh, T.H. (1997). Acidic and neutral drugs screen in blood with quantitation using microbore high-performance liquid chromatography-diode array detection and capillary gas-chromatography-flame ionization detection, *Forensic Science International* **90**(3), 205–214.

[24] Drummer, O.H., Horomidis, S., Kourtis, S., Syrjanen, M.L. & Tippett, P. (1994). Capillary gas chromatographic drug screen for used in forensic toxicology, *Journal of Analytical Toxicology* **18**(3), 134–138.

[25] Sims, D.N., Felgate, P.D., Felgate, H.E. & Lokan, R.J. (1991). Application of a simple extraction procedure using aqueous ammonia to the analysis of basic drugs in blood by gas chromatography, *Forensic Science International* **49**(1), 33–42.

[26] Huang, Z.P., Chen, X.H., Wijsbeek, J., Franke, J.P. & de Zeeuw, R.A. (1996). An enzymatic digestion and solid phase extraction procedure for the screening for acidic, neutral, and basic drugs in liver using gas chromatography for analysis, *Journal of Analytical Toxicology* **20**(4), 248–254.

[27] Pfleger, K., Maurer, H.H. & Weber, A. (1998). *Mass Spectral Library of Drugs, Pesticides, Poisons and their Metabolites*, 3rd revision, Hewlett Packard, Paol Alto.

[28] Marquet, P. (2002). Progress of liquid chromatography-mass spectrometry in clinical and forensic toxicology, *Therapeutic Drug Monitoring* **24**(2), 255–276.

[29] Van Bocxlaer, J.F., Clauwaert, K.M., Lambert, W.E., Deforce, D.L., Van den Eeckhout, E.G. & de-Leenheer, A.P. (2000). Liquid chromatography-mass spectrometry in forensic toxicology, *Mass Spectrometry Reviews* **19**(4), 165–214.

[30] Maurer, H.H. (2005). Multi-analyte procedures for screening for and quantification of drugs in blood, plasma or serum by liquid-chromatography-single stage or tandem mass spectrometry (LC-MS or LC-MS/MS) relevant to clinical and forensic toxicology, *Clinical Biochemistry* **38**(4), 310–318.

[31] Garriott, J.C. (1996). *Analysis of Alcohol in Post-mortem Specimens of Medicolegal Aspects of Alcohol*, 3rd Edition, J.C. Garriott, eds, Lawyers & Judges Publishing Co., Chapter 6.

[32] Datta, P. & Dasgupta, A. (2002). Effect of Chinese medicines Chan Su and Danshen on EMIT 2000 and Randox digoxin immunoassays: wide variation in digoxin-like immunoreactivity and magnitude of interference in digoxin measurement by different brands of the same product, *Therapeutic Drug Monitoring* **24**(5), 637–644.

[33] Drummer, O.H., Kotsos, A. & McIntyre, I.M. (1993). A class-dependent drug screen in forensic toxicology using diode array detector, *Journal of Analytical Toxicology* **17**(4), 225–229.
[34] Ojanperä, I., Rasanen, I. & Vuori, E. (1991). Automated quantitative screening for acidic and neutral drugs in whole blood by dual column capillary gas chromatography, *Journal of Analytical Toxicology* **15**(4), 204–208.
[35] Bogusz, M. & Erkens, M. (1995). Influence of biologic matrix on chromatographic behavior and detection of acidic, neutral and basic drugs examined by means of a standardized HPLC-DAD system, *Journal of Analytical Toxicology* **19**(1), 49–55.
[36] Hemmersbach, P. & de la Torre, R. (1996). Stimulants, narcotics and beta-blockers: 25 years of development in analytical techniques for doping control, *Journal of Chromatography. B, Biomedical Applications* **687**(1), 221–238.
[37] Segura, J., Venura, R. & Jurado, C. (1998). Derivatization procedures for gas chromatographic-mass spectrometric determination of xenobiotics in biological samples, with special attention to drugs of abuse and doping agents, *Journal of Chromatography. B, Biomedical Applications* **713**(1), 61–90.
[38] Maurer, H.H. & Arlt, J.W. (1998). Detection of 4-hydroxycoumarin anticoagulants and their metabolites in urine as part of a systematic toxicological analysis procedure for acidic drugs and poisons by gas chromatography-mass spectrometry after extractive methylation, *Journal of Chromatography. B, Biomedical Sciences and Applications* **714**(2), 181–195.
[39] Neill, G.P., Davies, N.W. & McLean, S. (1991). Automated screening procedure using gas chromatography-mass spectrometry for identification of drugs after their extraction from biological samples, *Journal of Chromatography* **565**(1–2), 207–224.
[40] Marquet, P., Saint-Marcoux, F., Gamble, T.N. & Leblanc, J.C. (2003). Comparison of a preliminary procedure for the general unknown screening of drugs and toxic compounds using a quadrupole-linear ion-trap mass spectrometer with a liquid chromatography-mass spectrometry reference technique, *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences* **789**(1), 9–18.
[41] Mueller, C.A., Weinmann, W., Dresen, S., Schreiber, A. & Gergov, M. (2005). Development of a multi-target screening analysis for 301 drugs using a QTrap liquid chromatography/tandem mass spectrometry system and automated library searching, *Rapid Communications in Mass Spectrometry* **19**(10), 1332–1338.
[42] Pelander, A., Ojanpera, I., Laks, S., Rasanen, I. & Vuori, E. (2003). Toxicological screening with formula-based metabolite identification by liquid chromatography/time-of-flight mass spectrometry, *Analytical Chemistry* **75**(21), 5710–5718.
[43] Brehmer, C. & Iten, P.X. (2003). Rapid determination of carboxyhemoglobin in blood by Oximeter., *Forensic Science International* **133**(1–2), 179–181.
[44] Lewis, R.J., Johnson, R.D. & Canfield, D.V. (2004). An accurate method for the determination of carboxyhemoglobin in post-mortem blood using GC-TCD, *Journal of Analytical Toxicology* **28**(1), 59–62.
[45] Kojima, T., Nishiyama, Y., Yashiki, M. & Une, I. (1982). Postmortem formation of carbon monoxide, *Forensic Science International* **19**(3), 243–248.
[46] Kage, S., Ito, S. & Kishida, T. (1998). A fatal case of hydrogen sulfide poisoning in a geothermal power plant, *Journal of Forensic Science* **43**(4), 908–910.
[47] Kage, S., Nagata, T. & Kudo, K. (1991). Determination of thiosulfate in body fluids by GC and GC/MS, *Journal of Analytical Toxicology* **15**(3), 148–150.
[48] Odoul, M., Fouillet, B., Nouri, B., Chambon, R. & Chambon, P. (1994). Specific determination of cyanide in blood by headspace gas chromatography, *Journal of Analytical Toxicology* **18**(4), 205–207.

WING-CHI CHENG

Postmortem Toxicology: Specimens *see* Toxicology: Analysis

Postpartum Psychosis

Introduction

Postpartum psychosis (PPP; also known as *puerperal psychosis*) was perhaps first described by Hippocrates in the fourth century BC. He described an acute onset of confusion, hallucinations, delirium, and insomnia [1, 2]. In 1865, the French physician Marce published his observations and study of perinatal disorders [3]. By the 1800s, symptoms of PPP were believed to be related to lactation, and the term *milk fever* was even used [1]. This "lactational insanity" was the basis for many Infanticide laws across the world. *The Yellow Wallpaper*, a monograph from 1899 [4], though controversial, may illustrate a woman's struggles with PPP.

The time in a woman's life when she is at greatest risk of psychosis or mental illness is in the postpartum period [5]. Fortunately PPP is rare, occurring after approximately 1–2 per thousand births. PPP often has a dramatic presentation, within the first several weeks of childbirth [6], but may begin within just days of giving birth. Early on, symptoms may include sleep disturbance and restlessness. Symptoms may evolve to include either depressed or elevated mood or both, agitation, delusions, hallucinations, and depersonalization. Women may believe they are being persecuted by the baby. Risks may include suicide, child neglect, or infanticide. PPP is considered a true psychiatric emergency, and mothers often require psychiatric hospitalization. Some specialized treatment units allow mothers to be hospitalized with their infants.

The Diagnosis of Postpartum Psychosis

Though cases may begin during pregnancy [2], more often they begin shortly after childbirth. Most cases of PPP begin within a couple weeks of delivery, with an abrupt onset [1, 2, 7]. By this time the mother has most often been discharged home with her fragile infant, rather than remaining in the hospital. Sleep deprivation is a potential trigger [8].

Risk of PPP may be related to hormonal shifts after birth (primarily the drop in estrogen) [7], stressors (such as marital problems), biology (bipolar disorder), and family history – genetic studies are underway [9], and can also be triggered by menstruation or cessation of lactation [2]. Some researchers have reported that the first pregnancy is a risk factor for PPP [2, 10]. Reports of recurrence rates of PPP after further pregnancies range from one in seven to women with bipolar disorder or schizoaffective disorder have a 50% risk of another episode of PPP [7, 11]. However, this rate is modified by prophylactic medication treatment [12]. Delivery complications may also elevate risk [10].

Common symptoms of PPP include symptoms of psychosis, with an impaired concept of reality and fluctuating delirium. Confusion, bizarre delusions (fixed, false beliefs) and behavior, hallucinations (unusual perceptual experiences that can be tactile, olfactory, or visual in addition to auditory), mood lability (ranging from depression to euphoria), and disorganized thinking may occur [1, 13]. For example, a woman might be delusional that she has special powers, that God has chosen her baby to be sacrificed, or that the baby is defective. Because of the lack of insight, suspicion, or conspiracy theories, women may not reveal their symptoms to others.

Among women with PPP, 72–88% have bipolar disorder or schizoaffective disorder, while 12% have schizophrenia [7]. In distinction to PPP, in schizophrenia delusional thinking and hallucinations often have a more gradual onset.

The psychiatric reference book DSM-IV-TR [14] does not list a specific diagnosis for PPP. Some debate exists within the field. According to the DSM-IV-TR, brief psychotic disorder or psychotic disorder not otherwise specified are diagnoses used for PPP, or a woman's symptoms may meet criteria for an affective (mood) episode [14]. Further even, the postpartum period is also defined differently depending on the group defining it; postpartum is 4 weeks according to the DSM-IV-TR, 6 weeks for ICD-10, 3 months in some epidemiological studies, and up to 1 year in some investigations [15].

Comparison with Postpartum Depression and Other Disorders

Postpartum blues or "baby blues" occur in approximately half to three-quarters of mothers [7]. Baby blues are not synonymous with postpartum depression (PPD); they are transient and not as severe. Symptoms usually occur in the first week after giving birth, and are self-limited. Symptoms may include anxiety, mood swings from sadness to irritability, and insomnia.

PPD symptoms are not the same as those of PPP. Approximately 10–15% of mothers experience PPD [3, 7]. Symptoms may occur within a few weeks to a year after giving birth. In PPD, a woman's predominant mood is sad or depressed, and she may lose enjoyment in her activities. Other symptoms may include insomnia, even an inability to sleep when the baby is sleeping. Her appetite may be abnormal. She may be fatigued and have difficulty concentrating. She may lack the interest in caring for her appearance, or even for her baby. She may experience feelings of worthlessness or hopelessness, and may have difficulty bonding with the infant. She may experience suicidal thoughts or thoughts of harming the infant. Symptoms of anxiety often also occur.

Psychiatrists must also differentiate psychotic thoughts and behavior from obsessive thoughts, which are also common in the postpartum [5, 7]. In obsessive compulsive disorder (OCD) or other anxiety disorders, women may experience intrusive thoughts that they have difficulty getting out of their head. These thoughts often represent worries, and the women, in distinction to women with psychosis, are not out of touch with reality. If these women have thoughts or fears, for example of harming someone, they usually realize that these thoughts are not of something that they would ever do and try to avoid the thoughts.

In the evaluation of a postpartum woman with psychiatric symptoms, physicians may consider that the differential diagnosis includes medical problems such as thyroiditis, Late Onset Tay Sachs disease, and vitamin B_{12} deficiency. Laboratory tests, computed tomography (CT) scan, or magnetic resonance imaging (MRI) of the brain may be indicated to rule out alternative organic diagnoses.

Treatment of Postpartum Psychosis

Because of the rapidly evolving and devastatingly severe nature of symptoms, PPP is often a psychiatric emergency. Women with a history of PPP or bipolar disorder have a 100-fold increase in psychiatric hospitalization during the postpartum [1]. In this disorder, treaters will often need to enlist help of family and other support. In some cases, child protective services may need to become involved.

Mood stabilizing medications (such as valproate or lithium) are mainstays of treatment, related to the frequently underlying bipolar disorder. Electroconvulsive therapy (ECT) is another consideration. Though not without risks of its own, ECT may provide for a rapid symptomatic improvement. Atypical antipsychotics such as olanzapine may have a role in the treatment of PPP as well [15]. (These agents are increasingly being used for not only the treatment of psychotic disorders but also of mood disorders.) Typical (older) antipsychotic agents may not lead to remission of symptoms [5]. Side effects of medications, such as sedation, which could impair her ability to respond to her infant, should be considered.

Because of the rarity of PPP, compared to other psychiatric disorders, there are fewer studies regarding evidence-based treatment. PPP is often considered to be a bipolar disorder unless proven otherwise.

Whether the mother is bottle-feeding or breast-feeding is another consideration, as different medications present different potential risks to breast-fed infants. The infant may require monitoring by a pediatrician. Other mothers may prefer to bottle-feed, to avoid any potential risk. With a supportive partner, bottle-feeding at least at night may have the added benefit of allowing the mother with PPP to sleep through the night.

Healthcare providers should educate patients about PPP. Support of families may help alleviate some of the stress associated with PPP [11]. It is important to consider prevention of episodes in future pregnancies/postpartum periods. Women with bipolar disorder are often treated prophylactically, where possible, with lithium or another mood stabilizer [7].

Risks in PPP

Mothers who experienced delusions that the baby is evil or a devil or not truly theirs were more likely to be abusive toward the infant [16]. Mothers with PPP "are at risk of injuring their children through practical incompetence or misguided delusions". [1, p. 106] Women with childbearing-related onset of psychiatric illnesses have reported homicidal ideation more frequently [13]. Risk of suicide is also present in PPP [17].

Neonaticide is specifically the murder of the neonate by the parent in the first 24 h of life [18]. Obviously, as most PPP has not yet begun at the time of birth, but rather has an onset several days to weeks later, there is a lack of relation of PPP to neonaticide in the great majority of cases. However, some cases of neonaticide do involve psychotic mothers [19].

Infanticide is often considered to include the murder of infants less than 1 year of age at the hands of the parent. (However, infanticide is a nonspecific term; in Biblical times, murder of a child at the hand of the state may have been considered infanticide.) Filicide, more precisely, is the murder of a child by the parent. If a woman is suffering from PPP, the physician should consider her risk of harming herself and/or her infant. Though there have been reported cases of infant murder while in the hospital [20], the overwhelming majority of cases occur outside the hospital. In psychiatric studies, mothers who kill their children often have

experienced psychosis, suicidality, and depression [21]. Often, alcohol or substance use, limited social support, and a maternal history of abuse are found as well.

The mother's motive for killing her infant or child may fall into any of several categories: *altruistic, acutely psychotic, fatal maltreatment, unwanted child*, or *spouse revenge* [22]. In an altruistic filicide, the mother kills her child out of love. A mother who sees her child suffering from cerebral palsy or leukemia may feel that the loving thing to do is spare the child from suffering. Though difficult to fathom, because of her depressed or psychotic outlook on life, another mother may believe that it is in the child's best interest to go to Heaven, rather than living an awful life on earth. The mother may be suicidal and not wish to leave her child in the hopeless world that she is departing. She may be psychotic, believing that if her child were to live, he or she would be tortured or raped or kidnapped, and may reason that the only humane choice is to kill him.

Alternatively, in an acutely psychotic filicide, the mother kills her child for no comprehensible reason. She may be responding to hallucinated voices commanding her to kill. She may kill in the throes of epilepsy or the confusion of delirium. She may put the baby in the oven instead of the dinner, for no rational reason but borne out of severe confusion and illness.

Fatal maltreatment filicide is the most common type of filicide overall. A child may die because of chronic abuse or chronic neglect. Certainly mothers suffering from psychosis or depression may be abusive, but so may many other mothers who do not have a mental illness. Too, mothers who are severely mentally ill and out of touch with reality may have difficulty providing for their infants' many needs. In an unwanted child filicide, a child who is not desired is killed. In a spouse revenge filicide, the rarest type, a child is killed in order to cause emotional pain in the other parent.

Newborns have a total dependence on their caregivers. If their caregiver is a woman with unidentified PPP, lacking in social support, that is highly concerning. A 4% risk of infanticide has been estimated in untreated PPP [23]. Even if a woman with PPP has support from her social network, if those who care do not understand the risks associated with PPP, there is cause for concern.

Infanticide Laws and Defenses

The British Infanticide Act was initially enacted in 1922, and was reformulated in 1938. The Act allows a woman, still recovering from giving birth, who kills her infant in the first 12 months of life to be charged with infanticide (akin to manslaughter) rather than murder, because "the balance of her mind is disturbed by reason of her not having fully recovered from the effect of giving birth to the child" [24].

Infanticide laws exist in over two dozen other nations, including Canada and Australia [25]. However, the legal criteria for infanticide vary across nations, including in New Zealand the murder of children up to age 10 [26]. In Luxembourg, there is a stricter penalty for child homicide.

However, a causal connectional between mental illness and the crime does not always occur in practice [27]. And, if a psychotic postpartum mother kills her infant and her older child, she could be charged for murder of the older child and infanticide for the killing of the infant. Also, one acutely psychotic mother who killed her 13-month-old baby might not qualify for infanticide though another mother who killed her 11-month-old baby *via* fatal maltreatment borne out of frustration might.

In the United States, there is no such infanticide legislation. The wake of the Andrea Yates child murder case led to proposals for American infanticide legislation [3, 28]. In the United States, the not-guilty-by-reason-of-insanity (NGRI) defense is used in some cases of maternal filicide. Throughout the United States and the world, there are various laws regarding NGRI (*see* **Insanity: Defense**). NGRI laws often include that a mental illness caused an actor not to know her act was wrong, and may include that she was unable to conform her conduct to the requirements of the law.

The diagnosis of PPP may be the mental illness on which the insanity defense is predicated. A woman with PPP and an alternative sense of reality may delusionally believe that she is doing what is right in killing her infant. Another woman may be unable to control her behavior because of manic psychosis. Still another may kill, believing that the hallucinations telling her to do so are from God. However, a mother who fatally abuses her infant may be unlikely to qualify for an NGRI finding because of the aforementioned requirements.

PPP can be difficult to prove in the legal sense – symptoms may have rapid resolution prior to psychiatric evaluation, and there has existed disagreement in the field as well. Psychiatric evaluation may be expedited in cases where PPP is believed to be related to the murder.

Summary

PPP may consist of symptoms including hallucinations, delusions, confusion, insomnia, mood swings, and loss of contact with reality. It must be differentiated from depression, anxiety, and medical disorders. Mothers may decompensate rather rapidly, within several weeks of delivery. Early recognition may be critical. These mothers often merit emergent treatment. Risks may include infant neglect, abuse, infanticide, or maternal suicide. Mothers with PPP who kill their infants may qualify for infanticide or insanity defenses.

References

[1] Attia, E., Downey J. & Oberman, M. (1999). Postpartum psychoses, in *Postpartum Mood Disorders*, L.J. Miller, ed, American Psychiatric Press, Washington, DC, pp. 99–117.

[2] Brockington, I. (1996). *Motherhood and Mental Health*, Oxford University Press, Oxford.

[3] Nelson, K.E. (2004). Postpartum psychosis and women who kill their children: making the punishment fit the crime, *Developments in Mental Health Law* **23**, 23–36.

[4] Gilman, C.P. (1892). *The Yellow Wallpaper*.

[5] Wisner, K.L., Gracious, B.L., Piontek, C.M., Peindl, K. & Perel, J.M. (2003). Postpartum disorders: phenomenology, treatment approaches, and relationship to infanticide, in *Infanticide: Psychosocial and Legal Perspectives On Mothers Who Kill*, M.G. Spinelli, ed, APPI, Washington, DC.

[6] Kendell, R.E., Chalmers, J.C. & Platz, C. (1987). Epidemiology of puerperal psychoses, *British Journal of Psychiatry* **150**, 662–673.

[7] Sit, D., Rothschild, A.J. & Wisner, K.L. (2006). A review of postpartum psychosis, *Journal of Women's Health* **15**(4), 352–368.

[8] Sharma, V., Smith, A. & Khan, M. (2004). The relationship between duration of labour, time of delivery, and puerperal psychosis, *Journal of Affective Disorders* **83**(2–3), 215–220.

[9] Jones, I. & Craddock, N. (2007). Searching for the puerperal trigger: molecular genetic studies of bipolar affective puerperal psychosis, *Psychopharmacology Bulletin* **40**(2), 115–128.

[10] Blackmore, E.R., Jones, I., Doshi, M., Haque, S., Holder, R., Brockington, I. & Craddock, N. (2006). Obstetric variables associated with bipolar affective puerperal psychosis, *British Journal of Psychiatry* **188**, 32–36.

[11] Robertson, E. & Lyons, A. (2003). Living with puerperal psychosis: a qualitative analysis, *Psychology and Psychotherapy* **76**(4), 411–431.

[12] Stewart, D.E., Klompenhouwer, J.L., Kendell, R.E. & van Hulst, A.M. (1991). Prophylactic lithium in puerperal psychosis, *British Journal of Psychiatry* **158**, 393–397.

[13] Wisner, K., Peindl, K. & Hanusa, B.H. (1994). Symptomatology of affective and psychotic illnesses related to childbearing, *Journal of Affective Disorders* **30**, 77–87.

[14] American Psychiatric Association (2000). *Diagnostic and Statistical Manual*, Text Revision, 4th Edition, American Psychiatric Association.

[15] Sharma, V., Smith, A. & Mazmanian, D. (2006). Olanzapine in the prevention of postpartum psychosis and mood episodes in bipolar disorder, *Bipolar Disorders* **8**(4), 400–404.

[16] Chandra, P.S., Bhargavaraman, R.P., Raghunandan, V.N. & Shaligram, D. (2006). Delusions related to infant and their association with mother-infant interactions in postpartum psychotic disorders, *Archives of Women's Mental Health* **9**(5), 285–288.

[17] Lindahl, V., Pearson, J.L. & Colpe, L. (2005). Prevalence of suicidality during pregnancy and the postpartum, *Archives of Women's Mental Health* **8**(2), 77–87.

[18] Resnick, P.J. (1970). Murder of the newborn: a psychiatric review of neonaticide, *American Journal of Psychiatry* **126**, 58–64.

[19] Putkonen, H., Weizmann-Henelius, G., Collander, J., Santtila, P. & Eronen, M. (2007). Neonaticides may be more preventable and heterogeneous than previously thought-neonaticides in Finland 1980–2000, *Archives of Women's Mental Health* **10**, 15–23.

[20] Mendlowicz, M.V., da Silva Filho, J.F., Gekker, M., de Moraes, T.M., Rapaport, M.H. & Jean-Louis, F. (2000). Mothers murdering their newborns in the hospital, *General Hospital Psychiatry* **22**(1), 53–55.

[21] Friedman, S.H., Horwitz, S.M. & Resnick, P.J. (2005). Child murder by mothers: a critical analysis of the current state of knowledge and a research agenda, *American Journal of Psychiatry* **162**, 1578–1587.

[22] Resnick, P.J. (1969). Child murder by parents: a psychiatric review of filicide, *American Journal of Psychiatry* **126**, 73–82.

[23] Altshuler, L.L., Hendrick, V. & Cohen, L.S. (1998). Course of mood and anxiety disorders during pregnancy and the postpartum period, *The Journal of Clinical Psychiatry* **59**(Suppl. 2), 29–33.

[24] Oberman, M. (1996). Mothers who kill: coming to terms with modern American infanticide, *American Criminal Law Review* **34**, 2–109.

[25] Friedman, S.H. & Resnick, P.J. (2007). Child murder by mothers: patterns and prevention, *World Psychiatry* **6**, 137–141.
[26] Dean, P.J. (2004). Child homicide and infanticide in New Zealand, *International Journal of Law and Psychiatry* **27**, 339–348.
[27] d'Orban, P.T. (1979). Women who kill their children, *British Journal of Psychiatry* **134**, 560–571.
[28] Connell, M. (2002). The postpartum psychosis defence and feminism: more or less justice for women? *Case Western Reserve Law Review* **53**, 143.

SUSAN HATTERS-FRIEDMAN

Posttraumatic Stress Disorder

Exposure to psychological trauma is thought to be a risk factor for the development of many mental health disorders including posttraumatic stress disorder (PTSD), acute stress disorder (ASD), other anxiety disorders, depressive disorders, somatic disorders, substance abuse disorders, and psychotic disorders. The specific mental health symptoms that typically become manifest following exposure to trauma appear to vary from culture to culture [1–3] and from era to era, e.g., somatic and psychotic symptoms were much more common in the traumatized World War II (WWII) combatants [4] than they were in traumatized Vietnam combatants [5]. The term *PTSD* was first introduced into the official Western classification of psychiatric disorders in 1980 with the publication of the third edition of the Diagnostic and Statistical Manual for Mental Health Disorders or DSM-III [6]. Prior to 1980, the disorder was referred to by a number of names including soldier's heart in the Civil War, shell shock and traumatic neurosis in WWI, combat fatigue and war neurosis in WWII, and gross stress reaction in the 1970s [7]. PTSD has also been referred to by a number of somewhat pejorative terms to include secondary gain neurosis, compensation neurosis, and litigation neurosis owing to its association with malingering and factitious disorder [8, 9]. Malingering involves the intentional production or substantial exaggeration of physical or psychological symptoms and/or dysfunction in order to secure external incentives, e.g., avoiding work, evading criminal responsibility, and obtaining disability compensation [10]. Factitious disorder also involves intentional feigning or exaggeration of physical or psychological symptoms or dysfunction, but the motivation for doing so is to maintain the sick role, the victim role, or the wounded soldier role rather than obtaining external incentives [10]. Malingering and factitious disorder often coexist; and for the sake of brevity, the term *malingering* henceforth will be used to denote both malingering and factitious disorder.

Six criteria currently define PTSD per the most recent edition of the DSM, the DSM-IV [10]. First, the person must have experienced, witnessed, or have been confronted with a traumatic event that involved an actual or threatened death or serious injury of someone, or involved a threat to the physical integrity of self or others. In addition, the person's emotional response to the traumatic event must have involved intense fear, helplessness, or horror. Second, the person must persistently reexperience the traumatic event in the form of nightmares, intrusive recollections during the waking state, flashbacks (illusions and hallucinations) in which the person feels as if the event were reoccurring, and/or intense psychological distress or physiological reactivity upon exposure to stimuli that resemble an aspect of the traumatic event. Third, the person must persistently avoid stimuli associated with the traumatic event and/or experience a numbing of general responsiveness. Fourth, the person must experience persistent hyperarousal symptoms, e.g., hypervigilance, exaggerated startle reflex, and irritability. Fifth, the person must manifest reexperiencing, avoidance/numbing, and hyperarousal symptoms for more than one month posttrauma. Sixth, the person must manifest clinically significant distress or impairment in social, occupational, or other important areas of functioning. The diagnostic criteria for ASD are very similar to those for PTSD – the major differences being the presence of dissociative symptoms in ASD, but not in PTSD, and the PTSD symptom triad of reexperiencing, hyperarousal, and avoidance/numbing symptoms are experienced for less than one month in ASD, rather than for more than a month as they are in PTSD [10].

A voluminous amount of research has been conducted over the course of the past 25 years on psychological trauma, ASD, and PTSD. One of the

more surprising findings is that most individuals in the United States will be exposed to one or more traumatic events in the course of their lifetime – between 68 and 90% depending on the criterion and measurement methods used [11–14]. Perhaps even more surprising is the finding that only a relatively small percentage of those individuals exposed to trauma go on to develop PTSD – between 9 and 14% depending on the stressor criteria and measurement methods used [11–16]. Individuals with certain characteristics have been found to be at greater risk than others to develop PTSD subsequent to trauma – namely, women [17] and individuals with a history of psychiatric disorders as well as individuals with a family history of psychiatric disorders [11–16]. Conversely, individuals with other characteristics appear less likely to develop PTSD subsequent to trauma – namely, men [17] and individuals with good intelligence [5, 15, 18] as well as those individuals with a substantial amount of a personality trait termed *hardiness* [19–22]. Hardy individuals are those who stay committed to life (as opposed to being alienated), who believe in their own ability to control and influence the course of events (as contrasted with a sense of being powerless), and who view negative life events as challenges to be met and overcome (as opposed to view these events as insurmountable threats). Similarly, some individuals respond to traumatic events and PTSD with enhanced mental health, altruism, empathy, spirituality, and prosocial values or what has been termed *posttraumatic growth* [23, 24].

Some types of trauma appear to be more pathogenic than the other types. The most pathogenic appears to be trauma that is repetitive and prolonged (as opposed to a single and brief), man-made (as opposed to natural catastrophes), and particularly gruesome such as is often found in combat, war atrocities, and torture situations [24–27]. Single, brief, and time-limited traumatic events also vary in terms of their pathogenic potential. For example, a situation in which an 18-month-old child is run over in his own driveway by a neighbor and thereafter dies shortly in the arms of his mother is far more likely to provoke PTSD in the neighbor and the mother than the bystanders who are horrified as they witness the tragedy but who have no personal involvement with the child, the mother, or the neighbor [24–27].

Another somewhat surprising finding is that those individuals who develop PTSD subsequent to trauma often spontaneously remit or improve over time without the benefit of formal treatment to the point that they no longer meet the criteria necessary to be diagnosed with PTSD. Estimates vary, but approximately 50% of those with PTSD appears to spontaneously remit within two years posttrauma with the majority of the remission taking place in the first year [5, 12, 16, 28–31]. Those who do not remit spontaneously and who do not remit by way of treatment have an elevated risk to develop other mental health disorders and substance use disorders [11, 12, 14]. Those who do remit are at no greater risk to develop other mental health disorders than those who were never traumatized in the first place [11, 12, 14].

Another important finding is the level of vocational and social dysfunction varies considerably in the PTSD-afflicted individuals. Some individuals function at a high level vocationally and socially, and only experience occasional episodes of acute PTSD symptoms typically provoked by trauma anniversary dates or by severe situational stressors such as the death or serious illness of a loved one. Still others with PTSD are severely and chronically incapacitated emotionally, socially, and vocationally. This usually occurs when the individual has been exposed to a particularly severe trauma or to multiple traumas, has severe psychopathology prior to the trauma, and/or when severe comorbid disorders develop subsequent to the PTSD such as addictions, personality disorders, or psychosis [11–14].

Assessment

Prior to the publication of the DSM-III in 1980 [6], PTSD was frequently misdiagnosed in the form of false negatives (PTSD was present but not diagnosed) simply because it was not recognized as a valid diagnostic entity. Since then, it appears that the frequency of PTSD diagnoses have steadily risen as has the misdiagnosis of PTSD in the form of false positives (PTSD is diagnosed as present when it is not). Described below are some of the errors that are frequently being made today.

Trauma

One of the more common diagnostic errors being made is to equate stressful life events with trauma.

For example, a highly contentious divorce, being fired from a job, being betrayed by a close and trusted family member, or observing a tragic and deadly event unfold on television are often highly distressing and can generate many PTSD-like symptoms; but such events lack the critical element of being personally life-threatening and thus fail to meet one of the essential criteria necessary to diagnose PTSD [27]. For the same reason, being fearful while on sentry duty in a noncombat zone and learning of the deaths of other soldiers killed in a helicopter crash (but not being personally involved with them nor witnessing their deaths) generally does not constitute trauma. Malingers often fabricate traumatic events or they may fabricate their involvement with actual traumatic events. In either case, their stories often have inconsistencies in their retelling, inconsistencies with factual accounts of the traumatic events or in the events leading up to them, and/or implausibilities embedded within their stories [27]. For example, a malinger claiming to have been a sniper in Vietnam could not recall the manufacturer or power of the scope that was mounted on an unlikely weapon he said he employed – a rifle commonly used in Vietnam by infantrymen but generally considered to be inaccurate at long distances.

PTSD Symptoms

Another common error is to diagnose PTSD on the basis of the presence of reexperiencing, hyperarousal, or avoidance/numbing symptoms by themselves [27]. None of these symptoms are unique to PTSD and all can be found in several other mental health disorders to include generalized anxiety disorder, phobia, panic disorder, and substance-induced mood and anxiety disorders. A causal link must be established between this triad of PTSD symptoms and a traumatic stressor before other mental health disorders can be safely ruled out and PTSD diagnosed. To do this, the triad of PTSD symptoms must have first become manifest after trauma exposure has occurred, not before. Most individuals who develop PTSD following trauma exposure will experience the initial onset of the triad of PTSD symptoms within days or weeks of trauma exposure. Occasionally individuals do delay the initial onset of PTSD symptoms for as long as a year or so posttrauma, but reports of delayed onset several years after trauma exposure should be suspect, particularly if the individual reports his or her mental health to be sound in the intervening period. The triad of PTSD symptoms typically wax and wane in concert with situationally induced stress once the symptoms first become manifest. Reports of severe PTSD symptoms that show little variability across time, stress, and situations should be suspect. Flashbacks wherein an individual with PTSD relives a traumatic experience in a delusional state and out of contact with reality do occur, but they are rare. When flashbacks do occur, they are typically provoked by high levels of situational stress coupled with substance abuse and poor sleep in individuals with a history of severe trauma and severe PTSD symptoms. Reports of frequent, prolonged, and unprovoked flashbacks should be suspect. Reexperiencing symptoms in the form of nightmares, in which the traumatic experiences are veridically recapitulated in the PTSD-afflicted individuals' dreams, are typical. PTSD nightmares usually are accompanied by substantially more body movement than is found in conventional nightmares. PTSD nightmares also normally contain little in the way of fantasy at the outset; although as time progresses, PTSD nightmares may come to incorporate contemporary problems and issues in them as well as some fantastic elements. Conventional nightmares are largely fantastic and contain little in the way of realistic trauma material. Individuals with PTSD typically avoid situations that are connected with, or similar to, the traumatic events. Malingers generally do not avoid such reminders to the same extent. For example, malingers often relish telling war stories and often watch war movies and war footage on the news – things that most individuals with combat-induced PTSD would generally try to avoid. Malingerers tend to over-report or exaggerate the severity and frequency of hyperarousal and reexperiencing symptoms, and they tend to report these symptoms in interviews in an overly dramatic or histrionic fashion. On psychological tests that are sensitive to malingering, malingers tend to endorse obvious PTSD items but not subtle PTSD items; they also tend to endorse extreme but rare PTSD symptoms, and they tend to produce overall "fake-bad" test profiles. For a thorough review of the literature regarding psychological testing and malingering, see the article on forensic matters [32] in the text by Wilson and Keane [27].

Disability

The level of psychological, social, education, and vocational dysfunction related to PTSD varies considerably from one person to the other [27, 32]. Uncomplicated PTSD that is induced by a single, time-limited traumatic event in a mentally healthy adult who has good social support and who resides in a safe and healthy environment generally produces psychological distress in the neurotic (moderate) range along with mild to moderate dysfunction in social/educational/vocational functioning. The level of psychosocial dysfunction tends to markedly increase in those PTSD-afflicted individuals who heavily rely on substances and/or high levels of avoidance to cope with their PTSD symptoms [29, 30]. More psychosocial dysfunction is also generally seen in PTSD-afflicted individuals who have low social support and reside in unsafe, impoverished environments. The most severe psychosocial dysfunction is generally found among those PTSD-afflicted individuals who have substantial pre-existing psychopathology and/or develop such pathology subsequent to the PTSD, e.g., substance abuse disorders, psychotic disorders, and personality disorders. Severe psychosocial dysfunction is also likely to be found in the PTSD-afflicted individuals whose PTSD was induced by prolonged, repetitive, and particularly gruesome trauma, e.g., atrocities, and torture. Psychosocial dysfunction should never be assumed if PTSD is present. It should always be independently assessed across vocational, educational, social, and recreational domains. Malingers tend to exaggerate their psychosocial dysfunction, and they often have marked discrepancies between their capacities for work and play. For example, a malinger claimed high levels of PTSD-related vocational dysfunction yet coached his son's soccer team and routinely hosted social events in his home and in the community for team members and their parents.

Malingering

Few, if any, studies have rigorously examined the prevalence of malingering in either clinical or forensic settings. There are, however, a number of case studies and studies that have employed samples of convenience that have estimated the prevalence to vary from 1% [33] to 50% [34] depending on the setting and the population being examined. For a review of the literature in this regard, see the article on forensic matters [32] in the text by Wilson and Keane [27]. In most clinical settings, mental health professionals rely almost exclusively on individuals' self-reports to diagnose PTSD that are taken by the way of unstructured interviews or structured interviews such as the Clinician-administered PTSD Scale [35]. Given clinicians' general lack of skepticism and their desire to give clients the benefit of the doubt, many malingerers appear to be able to feign PTSD symptoms and trauma exposure well enough in such interviews to avoid being detected. This generally does not pose a problem provided the prevalence of malingering is low as it is in most clinical settings. However, in a few clinical settings and nearly all forensic settings, malingering is likely to be prevalent. This necessitates taking additional assessment measures to rule maligning out. Such measures usually entail collaborating clients' self-reports of trauma exposure, collaborating their self-reports of PTSD symptoms, and collaborating their self-reports of psychosocial disability by way third party reports, reviews of public and private records (e.g., media accounts, police records, military records, and medical records), and the administration of structured interviews that are sensitive to dissimulation such as the Structured Interview of Reported Symptoms [36] as well as psychological tests that have scales that are sensitive to dissimulation such as the Minnesota Multiphasic Personality Inventory 2 (MMPI2) [37]. Gathering such information usually adds considerable time and expense to the assessment process.

It is generally a good idea to rate individuals in terms of the malingering indicators listed below before diagnosing them with PTSD. The more malingering red flags that are present then the greater the likelihood of malingering. The number of red flags that should trigger efforts to collaborate individuals' self-reports will vary from setting to setting depending on the costs for misdiagnosing PTSD if malingering is overlooked *versus* the costs for conducting the more thorough assessment necessary to rule out malingering. The costs will also be strongly influenced by the relative prevalence rates of PTSD versus malingered PTSD in the population of interest, which are likely to vary from time to time. For instance, malingered PTSD was probably relatively rare in veterans who approached the United States

Department of Veteran Affairs for services throughout the 1980s – about 5% according to one estimate [38]. Thus, the costs of occasionally misdiagnosing PTSD because malingering was overlooked were relatively small then. But clinical observation and some empirical evidence [39–41] now suggest that malingered PTSD is far more common in veterans seeking services from the US Department of Veteran Affairs than it once was. Consequently, the costs of misdiagnosing PTSD because malingering is overlooked are probably substantially greater than they once were. These costs not only include disability compensation being erroneously awarded to malingers but also include the costs of utilizing scarce clinical resources to treat malingers that could be better utilized if these resources were devoted to treating those in greater clinical need.

Malingering (Factitious Disorder) Red Flags

1. Criminal proceedings

Malingering should always be suspected when a diagnosis of PTSD is likely to mitigate an individual's criminal culpability.

2. Disability proceedings

Malingering should always be suspected when a diagnosis of PTSD is likely to result in disability compensation for an individual.

3. History of disability claims

Malingering should always be suspected when an individual has a history of filing disability claims.

4. Antisocial personality disorder

Malingering should always be suspected when an individual has an antisocial personality disorder or antisocial personality traits as evidenced by a long history of significant legal infractions and/or social irresponsibility.

5. Poor work history

Malingering should be suspected when an individual is functioning as a marginal member of society as manifested by a poor work history, no permanent address, a vagabond lifestyle, etc. Severe PTSD or PTSD complicated by severe comorbid disorders can result in a poor work history so it is important to examine an individual's work history prior to the trauma as well as after.

6. Substance abuse disorder

Malingering should be suspected when an individual presents with a substance abuse disorder. Substance abuse disorders sometimes do evolve as a consequence of PTSD so it is important to examine the individual's substance usage before the trauma as well as after.

7. Marked disparities in psychosocial functioning

Malingering should be suspected when an individual claims substantial vocational impairment subsequent to PTSD yet retains the capacity to successfully engage in social and/or recreational activities.

8. Atypical presentational styles

Malingering should be suspected if an individual presents with extreme emotional displays and/or bizarre symptoms when discussing their trauma. At the other extreme, malingering should also be suspected if an individual presents in a bland, matter-of-fact manner when discussing their trauma. Almost all PTSD-afflicted individuals will have some emotional distress and difficulty in discussing their trauma, but it will not be so severe as to disrupt the assessment process nor will it have a histrionic quality to it. Malingerers tend to present themselves as victims and tend to externalize, condemn, and blame others for their misfortune with accompanying anger, whereas individuals with PTSD tend to internalize and perceive themselves as being primarily responsible for their misfortune with accompanying guilt and shame. Malingering should also be suspected when an individual idealizes their life and work prior to the trauma while simultaneously attributing any and all shortcomings they currently have to the trauma and/or PTSD. Malingering should also be suspected when an individual behaves in an evasive, noncooperative, or hostile manner during the assessment. Such presentational styles are atypical of PTSD-afflicted individuals in most assessment settings.

9. Atypical trauma

Malingering should be suspected when traumatic events cannot be independently verified. Malingering should also be suspected when the traumatic events contain one or more implausible elements or when there are significant inconsistencies noted from one retelling of the trauma story to the next. Malingering should also be suspected when severe PTSD is claimed, the individual has no

pre-existing psychopathology, and the trauma is a single, brief, and not a particularly gruesome tragedy.

10. Atypical PTSD symptoms

Malingering should be suspected when an individual reports a number of atypical symptoms or symptom patterns to include: substantial time delays between traumatic events and the initial onset of the PTSD symptom triad, the complete cessation of the PTSD symptoms for substantial periods of time owing to the heavy use of substances, PTSD symptoms that fail to vary in frequency and severity in concert with situational stressors, the presence of rare PTSD symptoms such as flashbacks occurring in the absence of high stress and substance abuse and a history of severe PTSD symptoms, nightmares that contain substantial fantasy and little trauma content, little in the way of avoidance of situational stimuli associated with the traumatic events, the endorsement of obvious PTSD symptoms but not subtle PTSD symptoms on psychological tests as well as overall fake-bad test profiles.

Treatment

Treatment for individuals with PTSD takes a variety of forms depending on the severity and chronicity of the PTSD symptoms, the severity and type of any comorbid disorders that might be present, the characteristics of the community and prevailing mental health service delivery system, and the characteristics of the individuals themselves, e.g., motivation, secondary gain, age, ethnicity, and gender [42, 43]. PTSD that is of recent onset and is uncomplicated by other mental health disorders or problematic environmental circumstances (e.g., homelessness or dangerous environments) generally can be effectively treated on an outpatient basis with antidepressant medications [42, 43] and exposure-based cognitive-behavior therapy [42, 43]. Exposure-based CBT safely exposes PTSD-afflicted individuals to trauma-related stimuli in imagination and/or *in vivo* while simultaneously facilitating the assimilation of their traumatic events into healthy beliefs or cognitive schema regarding themselves and their world [42, 43]. It is also often beneficial to treat PTSD-afflicted individuals' co-occurring situational problems by way of supportive problem-solving while simultaneously treating their uncomplicated, recent onset PTSD way of antidepressants and CBT – exposure. This is particularly true in the case of soldiers recently returning from a war zone. Nearly all returning soldiers will be struggling with the task of transitioning from the military back into civilian life, with all of its attendant housing, job, education, family, and social support stressors; and some of the soldiers will be struggling with recent onset PTSD symptoms as well. In most cases, the stress of their readjustment issues will aggravate and prolong the PTSD symptoms of those who have them. By reducing the readjustment-related stress, the PTSD symptoms are likely to be lessened and the spontaneous remission process is likely to be strengthened.

Hospitalization for the treatment of PTSD is seldom necessary. With the possible exception of one study [44], most of the evidence suggests that there is little sustained therapeutic benefit from treating chronic PTSD in specialized inpatient PTSD units [45–49]; and such hospitalizations appear to have iatrogenic effects for a significant minority of those who undergo such treatment [44–49]. Treatment of PTSD complicated by severe comorbid disorders is usually dictated by the nature of the comorbid disorders rather than the PTSD *per se* [42, 43]. For example, those with PTSD accompanied by severe depression and significant suicidal risk usually require hospitalization to treat the depression and suicidal threat before beginning treatment for the PTSD. In other cases of chronic, severe, and highly complicated cases of PTSD, it is sometimes best to approach treatment as one would approach an incurable and chronic medical illness. That is, the treatment should focus on palliative care and try to help the individual manage his or her psychiatric symptoms rather than directly focus on the traumatic events and PTSD symptoms in an effort to eliminate them, as would be the case with exposure-based CBT [42, 43].

Summary

In summary, approximately 75% of all adults will be exposed to one or more traumatic events during the course of their lives. About 10% of those exposed to trauma will develop PTSD subsequent to the trauma, and about 50% of those who do develop PTSD will spontaneously remit within two

years posttrauma. Those who do not remit are at increased risk to develop additional disorders such as mood disorders and substance abuse disorders. Some individuals are at greater risk to develop PTSD than the others, particularly those with pre-existing psychopathology. Multiple trauma that is gruesome, prolonged, and man-made is more pathogenic than is less repetitive and less gruesome forms of trauma. The psychosocial dysfunction stemming from PTSD varies greatly from person to person depending on the severity of the trauma, the severity of the PTSD symptoms, and the type and severity of the comorbid disorders that are often present. Recent onset PTSD can be effectively treated by way of CBT exposure-based psychotherapy and antidepressant medications on an outpatient basis. The treatment of chronic PTSD in specialized PTSD inpatient units appears to be largely ineffective and possibly harmful for a significant minority of those who undergo it. PTSD appears to be frequently misdiagnosed in the form of false positives today. Malingering should always be suspected in forensic settings. Malingering should also be suspected in clinical settings when a significant number of malingering red flags are present. Collaborating clients' self-reports of trauma, PTSD symptoms, and psychosocial dysfunction are usually required to rule out malingering.

References

[1] Miller, K., Kulkarni, M. & Kushner, H. (2006). Beyond trauma-focused psychiatric epidemiology: bridging research and practice with war-affected populations, *The American Journal of Orthopsychiatry* **76**, 409–422.

[2] Kagee, A. & Naidoo, A. (2004). Reconceptualizing the sequela of political torture: limitations of a psychiatric paradigm, *Transcultural Psychiatry* **41**, 46–61.

[3] Jenkins, J. (1996). Culture, emotion and PTSD, in *Scurfield Ethnocultural Aspects of Post-traumatic Stress Disorder*, A. Marsella, M. Friedman, E. Gerrity & R. Scurfield, eds, American Psychological Association, Washington, DC, pp. 165–182.

[4] Kardiner, A. (1941). *The Traumatic Neuroses of War*, P.B. Hoeber, New York.

[5] Kulka, R., Schlenger, W., Fairbanks, J., Hough, R., Jordan, B., Marmar, C. & Weiss, D. (1990). *Trauma and the Vietnam War Generation*, Brunner/Mazel, New York.

[6] American Psychiatric Association (1980). *Diagnostic and Statistical Manual of Mental Disorders*, 3rd Edition, APA Press, Washington, DC.

[7] American Psychiatric Association (1968). *Diagnostic and Statistical Manual of Mental Disorders*, 2nd Edition, APA Press, Washington, DC.

[8] Resnick, P. (1997). Malingering of posttraumatic disorders, in *Clinical Assessment of Malingering and Deception*, R. Rogers, ed, Guilford, New York, pp. 130–152.

[9] Mendelson, G. (1984). Follow-up studies of personal injury litigants, *International Journal of Law and Psychiatry* **7**, 179–187.

[10] American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition, APA Press, Washington, DC.

[11] Breslau, N., Kessler, R., Chilcoat, H., Schultz, L., Davis, G. & Andreski, P. (1998). Trauma and posttraumatic stress disorder in the community: the 1966 detroit area survey of trauma, *Archives of General Psychiatry* **55**, 626–632.

[12] Breslau, N. (2002). Epidemiologic studies of trauma, posttraumatic stress disorder, and other psychiatric disorders, *Canadian Journal of Psychiatry* **47**, 923–929.

[13] Breslau, N. & Kessler, R. (2001). The stressor criterion in DSM-IV posttraumatic stress disorder: an empirical investigation, *Biological Psychiatry* **50**, 699–704.

[14] Kessler, R., Sonnega, A., Bromet, E., Hughes, M. & Nelson, C. (1995). Posttraumatic stress disorder in the National Comorbidity Survey, *Archives of General Psychiatry* **52**, 1048–1060.

[15] Brewin, C., Andrews, B. & Valetine, J. (2000). Meta-analysis of risk factors for posttraumatic stress disorder in trauma-exposed adults, *Journal of Consulting and Clinical Psychology* **68**(5), 748–766.

[16] McFarlane, A. (1989). The aetiology of post-traumatic morbidity: predisposing, precipitating, and perpetuating factors, *The British Journal of Psychiatry* **154**, 221–228.

[17] Breslau, N., Davis, G., Andreski, P., Peterson, E. & Schultz, L. (1997). Sex differences in posttraumatic stress disorder, *Archives of General Psychiatry* **54**, 1044–1048.

[18] McNally, R. & Shin, L. (1995). Association of intelligence with severity of PTSD symptoms in Vietnam combat veterans, *The American Journal of Psychiatry* **156**(6), 936–938.

[19] Kobasa, S. (1979). Stressful life events, personality and health, *Journal of Personality and Social Psychology* **37**, 1–11.

[20] Linley, P. (2003). Positive adaptation to trauma: wisdom as both process and outcome, *Journal of Traumatic Stress* **16**, 601–610.

[21] Bonanno, G. (2004). Loss, trauma, and human resilience, *The American Psychologist* **59**, 20–28.

[22] Waysman, M., Schwarzwald, J. & Solomon, Z. (2001). Hardiness: an examination of its relationship with positive and negative long term changes following trauma, *Journal of Traumatic Stress* **14**(3), 531–547.

[23] Linley, P. & Joseph, S. (2004). Positive change following trauma and adversity: a review, *Journal of Traumatic Stress* **17**, 11–21.

[24] Wilson, J. (2006). *The Posttraumatic Self*, Routledge, New York.
[25] Roth, S., Newman, E., Pelcovitz, D., van der Kolk, B. & Mandel, F. (1997). Complex PTSD in victims exposed to sexual and physical abuse, *Journal of Traumatic Stress* **10**(4), 539–555.
[26] Yehuda, R., Southwick, S. & Giller, E. (1992). Exposure to atrocities and severity of chronic posttraumatic stress disorder in Vietnam combat veterans, *American Journal of Psychiatry* **149**(3), 333–336.
[27] Wilson, J. & Keane, T. (2004). *Assessing Psychological Trauma and PTSD*. Guilford, New York.
[28] Green, B. (1994). Psychosocial research in traumatic stress: an update, *Journal of Traumatic Stress* **7**(3), 341–362.
[29] Perkonigg, A., Pfister, H., Stein, M., Hofler, M., Lieb, R., Maercker, A. & Wittchen, H. (2005). Longitudinal course of posttraumatic stress disorder and posttraumatic stress disorder symptoms in a community sample of adolescents and young adults, *American Journal of Psychiatry* **162**, 1320–1327.
[30] Karamustafalioglu, O., Zohar, J., Guveli, M., Gal, G., Bakim, B., Fostick, L., Karamustafalioglu, N. & Sasson, Y. (2006). Natural course of posttraumatic stress disorder: a 20-month prospective study of Turkish earthquake survivors, *The Journal of Clinical Psychiatry* **67**, 882–889.
[31] Shlosberg, A. & Strous, R. (2005). Long-term follow-up of PTSD in Israeli Yom Kippur war veterans, *The Journal of Nervous and Mental Disease* **193**, 693–696.
[32] Wilson, J. & Moran, T. (2004). Forensic/clinical assessment of psychological trauma and PTSD in legal settings, in *Assessing Psychological Trauma and PTSD*, J. Wilson & T. Keane, eds, Guilford, New York, pp. 603–636.
[33] Keiser, H. (1968). *The Traumatic Neurosis*, JB Lippincott Co., Philadelphia.
[34] Miller, H. & Cartlidge, N. (1972). Simulation and malingering after injuries to the brain and spinal cord, *Lancet* **1**, 580–585.
[35] Blake, D., Weathers, F., Nagy, L., Kaloupek, D., Klauminizer, G., Charney, D. & Keane, T. (1990). *Clinician–Administered PTSD Scale*, National Center for PTSD, West Haven.
[36] Rogers, R., Bagby, R. & Dickens, S. (1992). *Structured Interview of Reported Symptoms*, Psychological Assessment Resources, Odessa.
[37] Butcher, J., Dahlstrom, W., Graham, J., Tellegen, A. & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for Administration and Scoring*. University of Minnesota Press, Minneapolis.
[38] Lynn, E. & Belza, M. (1984). Factitious posttraumatic stress disorder, *Hospital and Community Psychiatry* **35**, 697–701.
[39] Frueh, B., Hamner, M., Cahill, S., Gold, P. & Hamlin, K. (2000). Apparent symptom overreporting in combat veterans evaluated for PTSD, *Clinical Psychology Review* **20**(7), 853–885.
[40] Frueh, B., Smith, D. & Barker, S. (1996). Compensation seeking status and psychometric assessment of combat veterans seeking treatment for PTSD, *Journal of Traumatic Stress* **9**(3), 427–439.
[41] Freuh, B., Gold, P. & de Arellano, M. (1997). Symptom overreporting in combat veterans evaluated for PTSD: differentiation on the basis of compensation seeking status, *Journal of Personality Assessment* **68**, 369–384.
[42] Wilson, J., Friedman, M. & Lindy, J. (2001). *Treating Psychological Trauma and PTSD*, Guilford, New York.
[43] American Psychiatric Association (2004). *Practice Guideline for the Treatment of Patients with Acute Stress Disorder and Posttraumatic Stress Disorder*, APA Press, Washington, DC.
[44] Creamer, M., Morris, P., Biddle, D. & Elliott, P. (1999). Treatment outcome in Australian veterans with combat-related PTSD, *Journal of Traumatic Stress* **12**(4), 545–558.
[45] Fontana, A. & Rosenheck, R. (1997). Effectiveness and cost of the inpatient treatment of PTSD, *American Journal of Psychiatry* **154**, 758–765.
[46] Creamer, M., Forbes, M., Biddle, D. & Elliott, P. (2002). Inpatient versus day hospital treatment for chronic, combat-related PTSD, *The Journal of Nervous and Mental Disease* **190**(3), 183–189.
[47] Fontana, A., Rosenheck, R. & Spencer, H. (1993). *The Long Journey Home: The Third Progress Report on Specialized PTSD Programs*, Northwest Program Evaluation Center, Department of Veteran Affairs Medical Center, West Haven.
[48] Johnson, D., Rosenheck, R., Fontana, A., Lubin, H., Charney, D. & Southwick, S. (1996). Outcome of intensive inpatient treatment for combat-related PTSD, *The American Journal of Psychiatry* **153**, 771–777.
[49] Hammarberg, M. & Silver, S. (1994). Outcome of treatment for posttraumatic stress disorder in a primary care unit serving Vietnam veterans, *Journal of Traumatic Stress* **7**(2), 195–216.

Related Articles

Behavioral Science Evidence
Deception: Detection of
Disaster Mental Health
Malingering: Forensic Evaluations
Rape Trauma Syndrome
Recollective Accuracy of Traumatic Memories
Syndromes: Psychological

LARRY D. SMYTH

PowerPlex

Introduction

The ability of polymerase chain reaction (PCR)-based short tandem repeat (STR) multiplexes to successfully analyze human DNA from a diverse range of circumstances is at the cornerstone of forensic DNA science. A contemporary version of these sophisticated systems has been the GenePrint PowerPlex range of multiplex STR systems (Promega Corporation, Madison, WI; Table 1).

These systems originally emerged in 1997 with PowerPlex 1.1 and progressed iteratively towards the current 16-locus PowerPlex®16 autosomal STR multiplex PCR system. The PowerPlex®16 offers extremely high discriminating power due to the combination of 15 STR loci plus the sex marker amelogenin [1, 2]. There is also the PowerPlex Y male-specific STR multiplex [3], which has specific advantages for the analysis of sex-assault cases and population genetic testing. The PowerPlex range of tests has shown its applicability to all forms of human identification casework, including kinship and disaster victim identification cases.

References

[1] Spreecher, C.J., Krenke, B., Rabbach, D., Hennes, L., Amiott, E., Nassif, N. & Mandrekar, P. (2000). The PowerPlex 16® system: development and validation, in *Proceedings of 11th International Symposium on Human Identification*, Promega Corporation, Madison, WI.

[2] Krenke, B., Tereba, A., Anderson, S.J., Buel, E., Culhane, S., Finis, C.J., Tomsey, C.S., Zachetti, J.M., Masibay, A., Rabbach, D.R., Amiott, E.A. & Sprecher, C.J. (2002). Validation of a 16-locus fluorescent multiplex system, *Journal of Forensic Sciences* **47**, 773–785.

[3] Krenke, B.E., Viculis, L., Richard, M.L., Prinz, M., Milne, S.C., Ladd, C., Gross, A.M., Gornall, T., Frappier, J.R., Eisenberg, A.J., Barna, C., Aranda, X.G., Adamowicz, M.S. & Budowle, B. (2005). Validation of male-specific, 12-locus fluorescent short tandem repeat (STR) multiplex, *Forensic Science International* **151**, 111–124.

Related Articles

Microsatellites

SIMON J. WALSH

Predecisional Bias: Jury *see* Jury Dynamics

Premenstrual Syndrome

The Medical Perspective

The idea that hormones control a woman's emotional state has been entrenched in populist thinking. For centuries, physicians tried in vein to link the cyclical hormonal pattern of estrogen and progesterone to

Table 1 GenePrint PowerPlex® STR multiplex systems produced for use in forensic DNA profiling

Name	No. of loci	Target loci included
PowerPlex (1.1 and 1.2)	8	TH01, TPOX, CSF1PO, vWA, D16S539, D13S317, D7S820, D5S818
PowerPlex 2.1	9	Penta E, D18S51, D21S11, TH01, D3S1358, FGA, TPOX, D8S1179, vWA.
PowerPlex ES	9	D3S1358, TH01, D21S11, D18S51, vWA, D8S1179, FGA, SE33 (also known as *ACTBP2*), Amelogenin
PowerPlex 16	16	D3S1358, vWA, FGA, TH01, TPOX, CSF1PO, D5S818, D13S317, D7S820, D8S1179, D21S11, D18S51, D16S539, Penta D, Penta E, Amelogenin
PowerPlex Y	12	DYS19, DYS385a, DYS385b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439

premenstrual syndrome (PMS). After all, it seemed so logical. Despite these efforts, medical science was never able to make the connection work. Attempts to treat PMS with hormonal treatments continuously failed. Hysterectomy and oopherectomy rarely afforded significant relief.

The breakthrough came with the discovery of serotonin, a ubiquitous neurotransmitter, now well established to underlie moodiness, anxiety, irritability, and hostility – symptoms commonly seen in PMS. Consequently, on 6 July 2000, fluoxetine became the first food and drug administration (FDA) recognized and approved treatment for the severe form of PMS, known as *premenstrual dysphoric disorder* (PMDD).[a]

So, the answer was not hormones – it was serotonin. Despite an understanding of the chemical etiology, PMS itself is a remarkably common experience and cannot itself be considered abnormal or pathological. Most women who experience PMS have mild or moderate symptoms and are not socially or occupationally impaired by them. However, a small subset of women experience significant impairment and disability in the premenstrual phase. Psychiatry has chosen to define the women suffering from this severe form of PMS as having PMDD. While deemed as an area of interest and research, PMDD is not considered to be a mental illness. It can be classified alternatively as a depressive disorder, not otherwise specified (coded 311); this category includes disorders with depressive features that do not meet the criteria for major depressive disorder. The sensitivity of the political correctness of this issue is obvious. The american psychiatric association (APA) has established diagnostic criteria, which are provided in the box below.

Legal Usage

In the 1980s, two British cases raised PMS as the basis for a diminished capacity defense reducing the quantum of guilt assessed against the accused.[b] First, in November, 1981, Sandie Smith was put on three years' probation after conviction of threatening to kill a police officer and for carrying a knife. She suffered from PMS and had committed almost 30 crimes, including arson, assault, and manslaughter, during the premenstrual period. Smith responded to progesterone therapy, advocated by English gynecologist Dr Katharine Dalton, to curb PMS. Dalton was a pioneer researcher who used her patients as a source of information for formulating what she and Dr Peter Green, a fellow researcher, named "premenstrual syndrome" in 1953.

Then, that same month Christine English pleaded guilty to manslaughter by reason of diminished responsibility. She drove her car into her lover after an argument that occurred while she was suffering from severe PMS [1]. She was conditionally discharged for 12 months.

One court martial case and an unreported US decision admitted PMS evidence as grounds for a defense. In *United States v. Morton*, the accused was charged with assault with a dangerous weapon, communicating a threat, and unlawfully carrying a concealed weapon. She pled not guilty by reason of insanity due to PMS. Morton had to establish by clear and convincing evidence that her PMS was so severe that she was unable to know and appreciate the consequences of her conduct [2]. The court held that she failed to establish insanity. It is unlikely that any US jurisdiction would allow PMS as an insanity defense.

In *People v. Santos* [3], the defendant testified in a preliminary hearing that she beat her child while in a "blackout" induced by PMS [4]. She was able to get a favorable plea bargain based on diminished capacity. Diminished capacity is an excuse defense that shifts the burden to the prosecution to disprove the excuse beyond a reasonable doubt once evidence of the excuse is admitted.[c] American states that recognize a diminished capacity defense might permit evidence of PMS to be admitted to show diminished capacity [5]. However, several states have abolished the diminished capacity defense.[d] Some states permit a diminished capacity defense murder prosecutions [6], following *United States v. Brawner* [7]. Other states permit a diminished capacity defense when the accused is charged with any crime of specific intent [8]. Federal practice under the 1983 Insanity Defense Reform Act [9] permits expert evidence on diminished capacity. Alaska allows expert testimony on diminished capacity that would permit PMS evidence during the guilt phase of trial.[e]

Diminished capacity must be established by expert opinion evidence that shows that the defendant both suffers from PMS and committed the crime charged while under the influence of PMS. Half

the states[f] and the Federal courts follow *Daubert v. Merrell Dow Pharmaceuticals Inc*. [10] and exclude expert opinion evidence if the expert's underlying scientific principles fail to meet a four-part test [11]:

1. Has the theory been tested by other researchers?
2. Has the theory been published?
3. What is the error rate?
4. Is the process generally accepted?

An expert in a *Daubert* jurisdiction must testify that PMS is a mental illness as diagnosed using diagnostic and statistical manuel of mental disorders, fourth edition, Text Revised (DSM-IV and DSM-IV(TR)) and explain the DSM-IV(TR) criteria for PMS (see Box). The expert must state that the diagnosis and criteria for PMS have been published and subject to peer review. She/he must assert the error rate for making the diagnosis of PMS. Finally, the expert should state that the criteria for diagnosing PMS and diminished responsibility are generally accepted (*see* **Expert Opinion: United Kingdom, Canada, and Australia**).

Some states follow the *Frye* rule (*see* ***Frye v. United States***) that requires the underlying scientific principles be "generally accepted" before an expert gives an opinion based on those principles [12]. Others have adopted a modified *Frye* rule with modifications using some or all of the *Daubert* factors [13]. In the *Frye* states and mixed states, the expert must testify that psychotherapists generally accept that PMS causes the sufferer to be unaware of the consequences of her action or to distinguish right from wrong.

Expert opinion evidence on PMS should be freely allowed during the sentencing phase of any trial. According to ¶ Section 5K2 of the Federal Sentencing Guidelines, the court may make a downward departure from the statutory maximum sentence for a convicted defendant on the ground of diminished capacity.[g]

Concerns about the use of PMS in the Courtroom

The forensic use of PMS has generated global controversy. Some of the principle arguments are as follows.

- **PMS is yet another sexist "woman as mad" explanation for female criminality**

Historically, deviant women have often been characterized as either "bad" or "mad" since their antisocial behavior conflicted with certain socially defined and desirable female personality traits and roles. Medical theory, from at least the mid 1800s, espoused that women's reproductive organs controlled their minds, bodies, and personalities. Such biological determinism was strongly endorsed and promulgated by psychiatry: in Freud's model of psychoanalysis, sexual temperament was conceived as a function of biology [14].

Accordingly, the nineteenth century doctors and lawyers agreed that menstruation and uterine malfunction could lead a woman to insanity or criminality. As an example, in the 1867 case *United States v. Harris*, attorneys called a psychiatrist and six other physicians who testified that the defendant was "morally insane" at the time of the homicide due to painful dysmenorrhea that led to mental derangement and hysteria. After a brief deliberation, the jury returned a verdict of not guilty by reason of insanity.[h] English courts also recognized some form of mental derangement related to the menstrual cycle as an excuse for criminality before *Harris* [15, 16]. One woman was acquitted of shoplifting in 1845 while two others were acquitted of murder in 1851. All three were found to have acted with temporary insanity due to "suppression of menstruation" [17]. One of these women had murdered her lover who had rejected her. A doctor testified that her wild eyes indicated problems with her uterus [18].

The specifics of the sociomedical theoretical explanations for female deviance shifted with time as the understanding of the female reproductive system evolved from the uterus to ovaries and then to hormones in the 1920s. With the UK cases of Smith and English discussed above, the focus in the early 1980s became PMS and its relationship to female criminality. In part, this undoubtedly reflected the growing interest in studying the female offender and etiology since statistics during the 1970s were reflecting an increasing number of crimes committed by women or, at the least, a higher number of arrests or prosecutions [19].

To some of its detractors, PMS is therefore the latest in a long line of anatomical deterministic theories of female criminality that have prevailed as

a substitute for looking at how entrenched gender inequality might contribute to crime. Instead, women's deviance has been linked with the anatomical parts that differentiate them from males and which allow them to fulfill their primary gender role of reproduction.

- **Its usage in court could stigmatize women as a whole and/or affect their equal participation in the public sphere**

There is concern that if PMS is raised as exculpatory grounds, people might generalize from the few and negatively stereotype all women or at the least those who experience any premenstrual symptoms [20]. Sensationalist captions such as "raging hormones", "premenstrual frenzy", and "Dr Jekyll and Ms Hyde" that appeared in British newspapers during the *English* and *Smith* cases can fuel imagery of women as periodically unstable and therefore unsuitable for some employment positions and/or responsibilities.

As with all medical disorders, a whole class of people with similar maladies could be stigmatized. As discussed earlier, the syndrome is fairly common, although only a minute percent manifest the symptoms that can substantially impact on their actions.

- **Its use relies upon acceptance of the legitimacy of PMS and PMDD and advocates/specialists**

All medical practitioners do not share in Dalton's belief in temporary psychosis as a symptom of the most severe PMS cases [21, 22]. The medical literature is confusing in its diversity of opinion concerning the possible connection of PMS to criminal behavior with no universally accepted medical consensus about the correlation of the severe variant with antisocial behavior.

In some countries like Australia where PMS is not raised except very infrequently in sentencing mitigation,[i] it continues to be referred to as *premenstrual tension* (PMT) and lacks acknowledgment as a legitimate medical condition. This is illustrated in a popular news feature story on PMS:

> Despite the popular belief that "it is all in the hormones", there is no convincing evidence that women with severe PMS have different hormonal fluctuations than other women . . . cause of PMS are still unknown . . . [23].

A number of female medical practitioners are quoted in the article articulating the view that women who think they have PMS may actually have clinical depression.

Yet to use it successfully a forensic expert is required. With insanity the defense must show that PMS is a disease of the mind and that the sufferer did not know the nature or quality of the act or that it was wrong (McNaughton Rule). Doing so with PMS can be highly problematic.

To raise automatism (a state in which the mind or the will does not accompany physical acts) by arguing that certain women with PMS who go hours without eating produce an excess amount of adrenalin that causes a hypoglycemic state of impaired consciousness requires an expert like Dalton to testify that hypoglycemia can be a symptom of PMS and that the defendant possessed that abnormality.

The defense with diminished responsibility or capacity must show that PMS prevented the accused from having the specific intent (*mens rea*) with hazy thinking, and impairment of self-control, judgment and willpower. Again, proof is problematic. Plus, there are several legal issues since the defense has to show that PMS is an abnormality of the mind (which is difficult since the symptoms of PMS are not even universally accepted), that it arose from an inherent cause, and that it substantially impaired the defendant's mental responsibility [24]. With potentially lengthy sentences, the convicted has the added stigma of mental illness [25].

- **It could be misused and or abused by defendants**

There is concern that PMS might be used as grounds for a defense by *non-bona fide* sufferers – either charlatans and/or women who experience some of the more mild PMS symptoms.

However, the diagnosis of PMDD can be substantiated by a heavy burden of proof, with medical evidence showing a clinically demonstrable physical disorder. A causal connection must be shown between the premenstrual symptom(s) and the criminal act. Additional evidence could include personal diaries, medical records, prior arrest record that could illustrate deviant activity correlation with the individual's premenstrual time of her cycle, and evidence by family and friends of marked premenstrual behavioral changes.

PMDD
APA Diagnostic Criteria [26]
In most menstrual cycles during the past year, symptoms (e.g., markedly depressed mood, marked anxiety, marked affective lability, decreased interest in activities) regularly occurred during the last week of the luteal phase (and remitted within a few days of the onset of menses). These symptoms must be severe enough to markedly interfere with work, school, or usual activities and be entirely absent for at least 1 week post menses.
The essential features are symptoms such as markedly depressed mood, marked anxiety, marked affective lability, and decreased interest in activities. These symptoms have regularly occurred during the last week of the luteal phase in most menstrual cycles during the past year. The symptoms begin to remit within a few days of the onset of menses (the follicular phase) and are always absent in the week following menses.
Five (or more) of the following symptoms must have been present most of the time during the last week of the luteal phase, with at least one of the symptoms being one of the first four:

1) feeling sad, hopeless, or self-deprecating;
2) feeling tense, anxious *or* "on edge";
3) marked lability of mood interspersed with frequent tearfulness;
4) persistent irritability, anger, and increased interpersonal conflicts;
5) decreased interest in usual activities, which may be associated with withdrawal from social relationships;
6) difficulty concentrating;
7) feeling fatigued, lethargic, or lacking in energy;
8) marked changes in appetite, which may be associated with binge eating or craving certain foods;
9) hypersomnia or insomnia;
10) a subjective feeling of being overwhelmed or out of control; and
11) physical symptoms such as breast tenderness or swelling, headaches, or sensations of "bloating" or weight gain, with tightness of fit of clothing, shoes, or rings. There may also be joint or muscle pain. The symptoms may be accompanied by suicidal thoughts.

This pattern of symptoms must have occurred most months for the previous 12 months. The symptoms disappear completely shortly after the onset of menstruation. The most typical pattern seems to be that of dysfunction during the week prior to menses that ends mid-menses. Atypically, some females also have symptoms for a few days around ovulation; a few females with short cycles might, therefore, be symptom free for only 1 *week per cycle.*
Typically, the symptoms are of comparable severity (but not duration) to those of a Major Depressive Episode and must cause an obvious and marked impairment in the ability to function socially or occupationally in the week prior to menses. Impairment in social functioning may be manifested by marital discord and problems with friends and family. It is very important not to confuse long-standing marital or job problems with the dysfunction that occurs only premenstrually. There is a great contrast between the woman's depressed feelings and difficulty in functioning during these days and her mood and capabilities the rest of the month. These symptoms may be superimposed on another disorder but are not merely an exacerbation of the symptoms of another disorder, such as Major Depressive, Panic, or Dysthymic Disorder, or a Personality Disorder. The presence of the cyclical pattern of symptoms must be confirmed by at least 2 consecutive months of prospective daily symptom ratings. Daily symptom ratings must be done by the woman and can also be done by someone with whom she lives. It is important that these diaries be kept on a daily basis rather than composed retrospectively from memory.
Delusions and hallucinations have been described in the late luteal phase of the menstrual cycle but are very rare. Although females with the combination of dysmenorrhea (painful menses) and premenstrual

dysphoric disorder are somewhat more likely to seek treatment than females with only one of these conditions, most females with either of the conditions do not have the other condition. A wide range of general medical conditions may worsen in the premenstrual or luteal phase (e.g., migraine, asthma, allergies, and seizure disorders). There are no specific laboratory tests that are diagnostic of the disturbance. However, in several small preliminary studies, certain laboratory findings (e.g., serotonin or melatonin secretion patterns, sleep EEG findings) have been noted to be abnormal in groups of females with this proposed disorder relative t o control subjects.

It is estimated that at least 75% of women report minor or isolated premenstrual changes. Limited studies suggest an occurrence of "premenstrual syndrome" (variably defined) of 20%–50%, and that 3%–5% of women experience symptoms that may meet the criteria for this proposed disorder. There has been very little systematic study on the course and stability of this condition. Premenstrual symptoms can begin at any age after menarche, with the onset most commonly occurring during the teens to late 20s. Those who seek treatment are usually in their 30s. Symptoms usually remit with menopause. Although symptoms do not necessarily occur every cycle, they are present for the majority of the cycles. Some months the symptoms may be worse than others. Women commonly report that their symptoms worsen with age until relieved by the onset of menopause.

Research criteria for premenstrual dysphoric disorder

A. In most menstrual cycles during the past year, five (or more) of the following symptoms were present for most of the time during the last week of the luteal phase, began to remit within a few days after the onset of the follicular phase, and were absent in the week postmenses, with at least one of the symptoms being either (1), (2), (3), or (4):
   (1) markedly depressed mood, feelings of hopelessness, or self-deprecating thoughts
   (2) marked anxiety, tension, feelings of being "keyed up", or "on edge"
   (3) marked affective lability (e.g., feeling suddenly sad or tearful or increasedsensitivity to rejection)
   (4) persistent and marked anger or irritability or increased interpersonal conflicts
   (5) decreased interest in usual activities (e.g., work, school, friends, hobbies)
   (6) subjective sense of difficulty in concentrating
   (7) lethargy, easy fatigability, or marked lack of energy
   (8) marked change in appetite, overeating, or specific food cravings
   (9) hypersomnia or insomnia
   (10) a subjective sense of being overwhelmed or out of control
   (11) other physical symptoms, such as breast tenderness or swelling, headaches, joint or muscle pain, a sensation of "bloating", weight gain.

   Note: In menstruating females, the luteal phase corresponds to the period between ovulation and the onset of menses, and the follicular phase begins with menses. In nonmenstruating females (e.g., those who have had a hysterectomy), the timing of luteal and follicular phases may require measurement of circulating reproductive hormones.

B. The disturbance markedly interferes with work or school or with usual social activities and relationships with others (e.g., avoidance of social activities, decreased productivity and efficiency at work or school).

C. The disturbance is not merely an exacerbation of the symptoms of another disorder, such as Major Depressive Disorder, Panic Disorder, Dysthymic Disorder, or a Personality Disorder (although it may be superimposed on any of these disorders).

Criteria A, B, and C must be confirmed by prospective daily ratings during at least two consecutive symptomatic cycles. (The diagnosis may be made provisionally prior to this confirmation.)

End Notes

[a.] The Massachusetts Institute of Technology holds the patent for this treatment.

[b.] The English Homicide Act of 1957 provides that "Where a person kills or is a party to the killing of another, he shall not be convicted of murder if he was suffering from such abnormality of mind ... as substantially impaired the mental responsibility for acts and omissions in doing or being a party to the killing." 5 and 6 Eliz. 2, ch. 2, § 2[1], 1957.

[c.] *LaFave* § 9.8(f)(4).

[d.] Arizona, California, Florida, Georgia, Maryland, Ohio, Oklahoma, Rhode Island, and South Carolina have abolished diminished capacity. See, e.g., *State v. Doss*, 568 P.2d 1054 (Ariz. 1977) (en banc) Cal. Penal Code § 25. Section 2.02 of the Model Penal Code abolished specific intent and diminished capacity.

[e.] See Alaska Stat. § 12.47.020.

[f.] Alaska, Colorado, Connecticut, Delaware, Idaho, Indiana, Iowa, Kentucky, Louisiana, Maine, Michigan, Mississippi, Nebraska, New Hampshire, New Mexico, North Carolina, Ohio, Oklahoma, Oregon, Rhode Island, Tennessee, Texas, West Virginia, and Wyoming. See, e.g., *State v. Coon*, 974 P.2d 386 (Alaska 1999); *People v. Shreck*, 22 P.3d 68 (Colo. 2001); *Springfield v. State*, 860 P.2d 435 (Wyo. 1993).

[g.] 18 U.S.C. Appendix Ch. Five Determining the Sentence, part K Departures 5K2.13. Diminished Capacity (Policy Statement).

[h.] See Clephane, J.O. *Trial of Mary Harris Indicted for the Murder of Adoniram Burroughs Before the Supreme Court of the District of Columbia*, 10–12 (opening statement of Joseph Bradley) (W.H. & O.H. Morrison, Washington, DC. 1865); Goldstein, A. (1997). *Nineteenth Century Gender Roles and the Murder Trial of Mary Harris*; Kaye, N.S. (1997). *Feigned Insanity*.

[i.] A search of Australian law databases such as LexisNexis and AUSTLII and the archives of the two major Australian newspapers *The Sydney Morning Herald* and *The Age* newspapers was conducted. No cases were reported in which PMS or PMT was used in an Australian court during that time period. Legal practitioners report its infrequent mention in sentencing.

References

[1] *British Legal Debate: Premenstrual Tension and Criminal Behavior*, New York Times, 29 Dec 1981, at http://www.nytimes.com/ (accessed 1 Jun 2007).

[2] 2001 CCA Lexis 202, (NMCM 99 00830 17 Jul 2001).

[3] No 1KO46229 (N.Y. Crim. Ct. 3 Nov 1982).

[4] (1983). Recent decisions: criminal law – premenstrual syndrome: a criminal defense, *Notre Dame Law Review* **59** 253.

[5] *United States v Pohlot*, 827 F2d 889 (3rd Cir. 1987); *cert denied*, U.S. 98 L Ed 2d 660, 108 S Ct 710 (1988) *LaFave*, § 9.1(a).

[6] North Carolina, Massachusetts, Oregon, New York, and Pennsylvania. See, e.g., *State v.* p., 488 S.E.2d 225 (N.C. 1997). *People v. Segal*, 429 N.E.2d 107, 444 N.Y.S.2d 588 (N.Y. 1981).

[7] 471 F.2d 969 (D.C. Cir. 1972).

[8] Hawaii, Kansas, Massachusetts, Minnesota, Mississippi, Missouri, and Tennessee. See, e.g., *State v Baker*, 691 P2d 1166 (Hawaii 1984). *State v. Grose*, 982 S.W.2d 349 (Tenn. Crim. App. 1997) appeal denied (May 11, 1998).

[9] *United States v. Brown*, 326 F.3d 1143 (10th Cir. 2003); 18 U.S.C.S. § 17 (2007).

[10] 509 U.S. 579, 113 S.C.t. 2786, 125 L.Ed.2d 469 (1993).

[11] *Daubert*, 509 U.S. 593–594 (1993).

[12] *Frye v. United States*, 293 F. 1013, 34 A.L.R. 145 (App. D.C. 1923) States following this rule include Arizona, California, the District of Columbia, Florida, Illinois, Kansas, Maryland, Minnesota, Missouri, Nebraska North Dakota, Pennsylvania, and Washington. See, e.g., *People v. Superior Court*, 137 Cal. App. 4th 353, 40 Cal. Rptr. 3d 365 (2d Dist. 2006). *Com. v. Crews*, 536 Pa. 508, 640 A.2d 395 (1994).

[13] *State v. Peters*, 192 Wis. 2d 674, 534 N.W.2d 867 (Ct. App. 1995) These states include Alabama, Georgia, Hawaii, Massachusetts, Nevada, Oregon, South Carolina, Utah, Virginia, and Wisconsin. See, e.g., *Turner v. State*, 746 So. 2d 355 (Ala. 1998).

[14] Chodorow, N. (1989). *Feminism and Psychoanalytic Theory*, Yale University Press, New Haven.

[15] Riley, T.L. (1986). Premenstrual syndrome as a legal defense, *Hamline Law Review* **9**, 193–194.

[16] D'Orban, P.T. (1983). Medicolegal aspects of the premenstrual syndrome, *British Journal of Hospital Medicine*, **30**, 404–406.

[17] Spitz, A. (1987). Premenstrual syndrome: a critical review of the literature, *Indiana Medicine* **80**(4), 378–382.

[18] Meehan, E. & MacRae, K. (1986). Legal implications of premenstrual syndrome: a Canadian perspective, *Canadian Medical Association Journal* **135**, 601–608.

[19] Horney, J. (1978). Menstrual cycles and criminal responsibility, *Law and Human Behaviour* **2**(1), 25–36.

[20] Easteal, P. (1991). Women and crime: premenstrual issues, *Trends and Issues in Crime and Criminal Justice* **31**, 1–8.
[21] Dalton, K. (1980). Cyclical criminal acts in premenstrual syndrome, *Lancet* 1070–1071.
[22] Dalton, K. (1986). Premenstrual syndrome, *Hamline Law Review* **9**(1), 143–154.
[23] Sweet, M. (1996). Periods of joy, *Sydney Morning Herald*, 1 April, p. 11.
[24] Mc Sherry, B. (1993). The return of the raging hormones theory, *Sydney Law Review* **15**, 309–310.
[25] Scutt, J. (1982). Premenstrual tension as an extenuating factor in female crime, *The Australian Law Journal* **56**, 99–100.
[26] American Psychiatric Association, (1994) Diagnostic and Statistical manual of Mental Disorders 4th ed., (Text Revision). Washington, DC.

Related Articles

Temporary Insanity

PATRICIA EASTEAL, NEIL S. KAYE AND TOM REED

Preventative Law ***see*** Therapeutic Jurisprudence

Primer Discharge Residue: Cartridge Discharge Residue ***see*** Firearm Discharge Residue: Analysis of

Printing Devices in Document Examination ***see*** Writing Instruments and Printing Devices

Privilege ***see*** Duty to Warn

Product Standard of Insanity ***see*** Insanity: Defense

Professional Judgment: Structured ***see*** Risk Assessment: Patient and Detainee

Professional Responsibility Codes for Forensic Scientists ***see*** Ethics: Codes of Conduct for Expert Witnesses

Profiles: Psychological and Behavioral

Profiles are most closely associated with psychological models developed by the Federal Bureau of Investigation (FBI) in their early searches for serial murderers [1] (*See* **Serial Homicide**). The term has since been applied to a great many diverse compilations of information, as well as to what are little more than stereotypes. Over the past decade or so, "profilers" have been a favorite topic of crime dramas and novels and in the process have been glamorized and misrepresented to the point where the facts have often been obscured. Of particular concern for the justice

system, however, has been the introduction of "profiles" as a component of psychological syndromes that have suggested new categories of "victims" and alleged profiles of the perpetrators responsible for their victimization (*see* **Syndromes: Psychological**; **Deception: Truth Serum**).

Profiles as an Aid in Law Enforcement

In 1987, Judge Charles Becton published a law review article that drew attention to a seemingly chameleon-like way in which drug courier profiles adapted to any particular set of observations [2]. Within months, the Court of Appeals for the Ninth Circuit incorporate Judge Becton's arguments into *United States v. Sokolow* [3]; a case which it had debated for over 2 years. By the late 1980s, however, criminal profiles had become a staple of law enforcement. Even the US Supreme Court could not stand by and watch them wiped out in a single sweep of a judicial pen.

The High Court went back to basics, and they found their foundation in *Terry v. Ohio* [4]. Officer McFadden, and his detention of two men casing Zucker's clothing store on an Ohio day in 1963, had given rise to the Terry Stop "stop and frisk" exception to the Fourth Amendment and the reasonable suspicion test, which provided justification, short of probable cause, under which police officers could initiate limited investigatory action. The Supreme Court knew that it could not allow any decision that it could make in *Sokolow* to turn profiles into unrestricted hunting licenses for overenthusiastic enforcement officials, nor was it willing to set up an entirely new bureaucracy to review the constitutionality of each and every profile. The Court also recognized that profiles changed over time, and that their effectiveness would evaporate if they became frozen and subject to public scrutiny.

This solution requires that every law enforcement intervention, regardless of whether or not it was set in motion by a profile, must be justifiable through the same articulation of facts – the "totality of the circumstances" – required for a *Terry* stop. Searches made under *Sokolow* are reviewed on a case-by-case basis and have been overturned if they fail to meet appropriate standards [5]. In this single decision, the Court recognized the validity of the probabilistic assumptions underlying the application of behavioral science techniques in justice administration, yet made it impossible to elevate criminal profiles into a license for Fourth Amendment abuse.

Criminal profiles continue to serve as a tool in law enforcement, but more importantly, profiles are a means of leveraging investigative expertise, training investigative and enforcement officials, and applying systematic methodologies to the fast-changing and highly mobile environment that characterizes contemporary criminal operations. Once the intervention has been initiated, however, the profile is of no further probative relevance. Arrest and even the issuance of a search warrant require probable cause, and prosecution of any resulting charges must be based upon substantive evidence of guilt. The original profile is inadmissible in support of guilt, and is presumed to be inherently prejudicial [6].

The temptation to push these limits was brought home to the British public in a highly publicized 1995 case. Three years earlier, 23-year-old Rachel Nickell took her 2-year-old son and dog for a walk. She selected Wimbledon Common as her south London destination because of its reputation for safety, but, less than an hour later, she was found soaked in the blood of some 49 stab wounds, her child clinging to her lifeless body crying "get up mummy". Police responded with the biggest murder investigation in London history – and one of the nation's early attempts at using the new science of psychological profiling.

England's first encounter with the behavioral sciences in criminal investigation came in the 1985 "Railway Rapist" case. Although profiling had become well established in the United States through the work of the FBI Behavioral Science Unit, both serial killers and the methods for catching them were only just then taking hold in Britain. David Canter, professor of investigative psychology at Liverpool University, had a background in the psychology of building design, human behavior during fires, and the psycholinguistics of hoax fire calls; but the rigor of his science prepared him well to become that nation's leading criminal profiler. His methodical work led to the conviction of John Duffy, and inspired the Robbie Coltrane character in the hit television detective series *Cracker* [7].

By the time of the Nickell murder, however, the publicity surrounding criminal profiling had attracted

a host of psychologists who had been bitten by the detective bug. It was one of these who claimed the case, produced a profile, and eventually took over much of the day-to-day police operations directed toward conviction of the suspect, Colin Stagg, targeted by the profile. Through an attractive blond undercover policewoman, the psychologist initiated an 8-month liaison with the 31-year-old Stagg in which she shared violent sexual fantasies, confessed to the ritual sexual murder of a baby and a young woman, and egged him on to match her stories; even telling him that she wished he were Nickell's murderer because "That's the kind of man I want."

Stagg never claimed credit for the killing, but from 700 pages of letters and transcribed telephone conversations and public meetings, the psychologist concluded that Stagg's fantasies, modeled upon information fed to him by those familiar with the details of the crime, revealed unique knowledge of the crime scene that could be known only by the murderer. Dragged before a judge in open court at the Old Bailey, defense quickly pointed out that Stagg had not even made good guesses – he did not know the location of the crime and had wrongly asserted that the victim had been raped.

Up until that point, Great Britain had never felt the need for an entrapment statute, but the judge recognized a "honey trap" when he saw one. Clearing the accused and acknowledging the understandable pressure on the police, the judge was, nevertheless, forced to conclude that the operation betrayed "not merely an excess of zeal, but a blatant attempt to incriminate a suspect by positive and deceptive conduct of the grossest kind". Stagg left the chaotic courtroom vowing to sue everyone involved, the police were the butt of press ridicule, and David Canter observed with typical English understatement that pulling in some "media recognized expert" can undermine "more effective, longer term development of a professional discipline" [8].

It is not just the newcomers who can make mistakes. The 1996 Summer Olympics began in the wake of the first case of suspected air terrorism on American soil, and as the investigation of the TWA flight 800 crash off New York moved ahead with commendable precision, a bomb blast rocked the Olympic festivities in Atlanta. Unwilling to stand by in the face of two national assaults, an inexperienced FBI agent allowed the press to get word that a psychological profile had identified the private security guard who first spotted the bomb as the likely perpetrator. In moments, Richard Jewell went from hero to the object of media scrutiny and scorn.

After a week of publicized investigation, in which the entire world got to see Jewell live on CNN sitting forlorn on his own front steps as the FBI picked through his apartment, the investigation yielded nothing more than a few pathetic souvenirs of a man's only moment of glory. Pressed for an explanation, an FBI spokesperson on the scene curtly informed the press that "We don't make apologies". FBI director Louis Freeh, called before a congressional investigating committee, tried to put a better face on the public relations disaster but confided privately that "We wish we never heard of Richard Jewell". [9, 10].

This is not to say that mention of a profile, or court testimony that overlaps in any way with a profile, is necessarily prohibited. Testimony can mention a profile in the context of background as to how and why a defendant was stopped and searched, if such testimony is confined to the preliminary stop and not the actual investigation, and the fact that the individual satisfied the profile is not used to impugn the defendant [11]. Moreover, the expertise reflected in profiles *can* be the subject of expert testimony to refute assertions by the defendant [12], to supply factual information helpful to the trier of fact in placing the case in context [13], or to establish motive, intent, absence of mistake or accident, or identity of a common scheme or plan [14].

Profiles as Evidence at Trial

By their tentative and proximal nature, profiles pose serious issues of both scientific and probabilistic validity [15]. Their use, and the term itself, are therefore most appropriately restricted to investigative tools. As evidence, profiles easily become little more than a means to either link a defendant to undesirable traits or stereotypes; or for the defense to attribute by association positive or sympathetic attributes to a defendant.

Behavioral science is relevant to justice because it can assist the trier of fact to understand a party in litigation as an individual (*see* **Risk Assessment**),

while profiles are a tactic to attribute the qualities of a grouping to an individual. There are circumstances in which this tactic can be attractive to either side of a case, and the law has established guidelines to restrict potential abuse.

Entrapment, the affirmative defense that alleged crimes were in reality induced by government persuasion or trickery, is a good example. An entrapment defense, as with defenses to most serious criminal charges, turns on the defendant's mental state: was the defendant an active participant or simply a passive bystander to conduct actually carried out by a government operative? Although early cases tended to hold that a defendant asserting the affirmative defense of entrapment could not establish their state of mind through expert testimony, the courts now generally consider such testimony an appropriate aid to the triers of fact [16, 17].

Most other attempts to enlist expert testimony regarding a defendant's personality profile to help disprove a criminal charge have been less successful. Introduction of psychiatric testimony regarding the "dependent personality disorder" of a defendant was excluded, for example, as support of her assertion that she was unaware that computer equipment that she sold was, in fact, stolen. Here, the court felt that imprimatur of such an official-sounding label was neither necessary nor helpful to the jury in making its assessment of the defendant's mental state [18].

Similarly, the psychological profile of a murder and robbery defendant was excluded as possible support that his crime could not have been deliberate and premeditated, holding that the profile appeared to be simply a narration of the defendant's social history with little or no rational bearing on issues of premeditation and intent [19]. Such personality testimony has also been excluded as a defense to armed robbery and assault [20], and manslaughter [21], where the defendants sought to establish that they were simply not the "type" to use a weapon.

In many instances, the testimony is simply a way in which to introduce character evidence. Testimony in support of good character is generally permissible [22], but the courts are leery of according it scientific stature [23]. Nevertheless, in two controversial decisions, lay character witness testimony has been upheld in a child molestation case [24], and a psychologist's opinion was upheld as appropriate testimony concerning a defendant charged with lewd and lascivious acts upon a child, noting that the testimony was based, at least in part, upon standardized testing [25].

Most frequently, however, courts have excluded expert testimony aimed simply at ruling out a defendant as the guilty party. This has been the case in proposed testimony as to "peaceableness" [26], psychiatrist testimony as to lack of characteristics "likely to result in abuse of infant victim" [27], psychiatrist testimony that defendant had made previous false confessions and may therefore be mentally ill and his confession untrustworthy [28], expert testimony as to defendant's remorse or lack of remorse [29], and that the defendant had undergone a religious conversion and therefore could be rehabilitated [30]. Expert testimony is also typically not allowed as to mitigation of an offense [31]. Often, such expert testimony is proffered in place of the defendant taking the stand on his or her own behalf, thus becoming subject to cross-examination. The jury system quite correctly assumes that a defendant is his own most revealing character witness, and that if character is to be made an issue it is best presented by the defendant himself [32].

Expert behavioral testimony on behalf of a defendant can also end up working against that defendant. For example, in *State v. Hunt* [33], the defendant claimed that a borderline personality prevented him from being able to form the necessary intent to be guilty of a shooting charge. The court ruled that this assertion opened the door for broad inquiry into his mental condition, and allowed the prosecution to counter the claim with expert testimony that the defendant was actually suffering from nothing more than "antisocial personality disorder". To explain how this conclusion was reached, the expert was further permitted to recount for the jury defendant's difficulties in interpersonal relationships, including his prior "bad acts".

If profile evidence has had little impact upon the ability of the accused to fashion a defense, it *has* provided a potentially devastating weapon in the hands of the prosecution. The case of Sgt. Russell Banks illustrates just how powerful and insidious prosecution profile testimony can become, even when the "expert" does not testify as to a personal conclusion about the guilt or innocence of the defendant. In this instance, a pinpoint profile that could only describe the defendant – a stepfather living with his wife

and her young daughter – combined with the known limitations of a child witness, and an aggressive child "therapist" able to lead that witness and allowed to testify as to her own conclusions, created a direct path to the defendant, which a jury would be hard pressed to ignore.

In its *Banks* holding [34], the Court of Military Appeals noted that its reversal of Sgt. Banks conviction for child rape and sodomy was consistent with the case law in both federal and state courts that has severely criticized attempts to introduce "profile" evidence to establish either guilt or innocence. As the Supreme Court of Kansas noted in the 1989 case of *State v. Clements*, "Evidence which only describes the characteristics of a typical offender has no relevance in determining whether the defendant committed a crime in question, and the only inference which can be drawn from such evidence, namely that the defendant who matches the profile must be guilty, is an impermissible one" [35]. This conclusion has been reached in cases as diverse as child molestation [36], child abuse [37], murder [38], rape [39], and shoplifting [40].

This is not to say that expert profile testimony may never be used by the prosecution. If the defendant places his own personality and character at issue, the prosecution can call experts to help rebut defense assertions [41]. In a Washington state case, a defendant who stuttered pointed to the fact that the person he allegedly assaulted was unable to identify his assailant as a stutterer. The prosecution was permitted to introduce to the jury expert scientific testimony as to the statistical percentage of probability that a stutterer would not exhibit that particular speech anomaly in certain situations [42].

Expert profile testimony as to a lack of profile can also be admissible when a defendant deviates significantly from the expectation that a lay jury may hold about people who commit particular types of crimes [24]. Finally, background testimony that does not specifically address guilt or innocence of a defendant but instead enables the jury to understand evidence that does go to guilt or innocence has been held to be permissible profile evidence.

Conclusions

The often misused term *profile* is best restricted to investigative tools used to narrow down likely suspects and establish reasonable suspicion necessary to justify an action in the field. Reasonable suspicion is reviewed on a case-by-case basis to assure protection of suspect rights. As evidence at trial, profiles easily become little more than a means to either link a defendant to undesirable traits or stereotypes; or for the defense to attribute by association positive or sympathetic attributes to a defendant. Although behavioral science evidence may be proffered for use at trial, including statistical compilations, such evidence should have relevance to the applicable party as an individual, and the term *profile* avoided when possible.

Expert testimony concerning a trait of an accused may only be used as evidence that the accused possesses such a trait. It must be left to the jury to determine whether and how such a trait may influence the facts of the case [43]. While the term *profile* may appear to give behavioral evidence an aura of scientific respectability, such labels themselves do nothing to enhance the stature of the substantive underlying observations. In fact, if anything, good science and credible observation are more readily accepted without them [44].

References

[1] For a critique of FBI profiles and a plea that behavioral evidence is admissible by the defense, see George, J.A. (2008). Offender profiling and expert testimony: scientifically valid or glorified results? *Vanderbilt Law Review* **61**, 221–260.

[2] Becton, C.L. (1987). The drug courier profile, *North Carolina Law Review* **65**, 417–480.

[3] United States v. Sokolow, 490 U.S. 1 (1989).

[4] Terry v. Ohio, 392 U.S. 1 (1968).

[5] An interesting look at the kind of reasoning used by the appellate courts in overturning a ***Sokolow*** search is provided in People v. Pullman, (Not Reported in Cal.Rptr.2d, 2002) WL 31230831 Cal.App. 1 Dist., (2002).

[6] People v. Hubbard, 530 N.W.2d 130 (Mich. Ct. App. 1995); United States v. Williams, 957 F.2d 1238 (5th Cir., 1992); United States v. Wilson, 930 F.2d 616 (Minn. 1991); United States v. Beltran-Rios, 878 F.2d 1208 (Cal. 1989); United States v. Hernandez-Cuartas, 717 F.2d 552 (Fla. 1983).

[7] Crace, J. (Feb 17, 1995) Inside the criminal mind, *New Statesman & Society*, 29; (1995 WL 14340484).

[8] *Guardian* (London) (1995). Sept 15 (1995 WL 9944184, 9944234, 9944240, 9944195 and 9944268).

[9] Yoder Jr., E.M. (Aug 2, 1996) Olympic park bombing, *San Diego Union Tribune*, B8.

[10] Sack, K. (Oct 29, 1996) Jewell lambastes FBI, media for 88-day ordeal as suspect, *Austin American-Statesman* A1.
[11] United States v. Hernandez-Cuartas, 717 F.2d 552 (Fla. 1983).
[12] People v. Lopez, 26 Cal. Rptr. 2d 741 (Ct. App. 1994); United States v. Robinson, 978 F.2d 1554 (N.M. 1992); United States v. Wilson, 930 F.2d 616 (Minn. 1991); United States v. Beltran-Rios, 878 F.2d 1206 (Cal. 1989).
[13] United States v. Khan, 787 F.2d 28 (N.Y. 1986).
[14] Wilson v. State, 871 P.2d 46 (Okl. 1994).
[15] An analysis of issues of probability as they relate to criminal profiles appears in Risinger, M. (2002). Three card monte, Monty hall, modus operandi and 'offender profiling': some lessons of modern cognitive science for the laws of evidence, *Cardozo Law Review* **238**, 193–284.
[16] State v. Woods, 484 N.E.2d 773 (Ohio Com. Pl. 1984);United States v. Hill, 655 F.2d 512 (Pa. 1981).
[17] Moore, C.D. (1995). The elusive foundation of the entrapment defense, *Northwestern University Law Review* **89**, 1151–1188.
[18] United States v. DiDomenico, 985 F.2d 1159 (Conn. 1993).
[19] Hartless v. State, 611 A.2d 581 (Md. 1992).
[20] People v. Watkins, 440 N.W.2d 36 (Mich. App. 1989).
[21] State v. Hensley, 655 S.W.2d 810 (Mo. App. 1983).
[22] Green, E.D. & Nesson, C.R. (eds) (1966). *Federal Rules of Evidence*, Little, Brown, Boston, pp. 51–58.
[23] Mendez, M.A. (1996). The law of evidence and the search for a stable personality, *Emory Law Journal* **45**, 221–238.
[24] People v. McAlpin, 812 P.2d 563 (Cal. 1991).
[25] People v. Stoll, 783 P.2d 698 (Cal. 1989).
[26] State v. Arnold, 421 A.2d 932 (Me. 1980).
[27] State v. Screpesi, 611 A.2d 34 (Del. Super 1991).
[28] Stano v. Dugger, 883 F.2d 900 (Fla. 1989).
[29] Clenney v. State, 344 S.E.2d 216 (Ga. 1986).
[30] People v. Moya, 350 P.2d 112 (Cal. 1960).
[31] People v. Masor, 578 N.E.2d 1176 (Ill. Ct. App. 1991).
[32] Mueller, C.B. & Kirkpatrick, L.C. (1996). *Evidence Under the Rules*, 3rd Edition, Little, Brown, Boston, pp. 677–679.
[33] State v. Hunt, 555 A.2d 369 (Vt. 1988).
[34] United States v. Banks, 36 M.J. 150 (CMA, 1992).
[35] State v. Clements, 770 P.2d 447, 448 (Kan. 1989).
[36] United States v. Gillespie, 852 F.2d 475 (Cal. 1988).
[37] Sloan v. State, 522 A.2d 1364 (Md. Ct. App. 1987).
[38] Sanders v. State, 303 S.E.2d 13 (Ga. 1983).
[39] State v. Percy, 507 A.2d 955 (Vt. 1986).
[40] State v. McCoy, 294 N.E.2d 242 (Ohio Ct. App. 1973).
[41] United States v. Gillespie, 852 F.2d 475 (Cal. 1988); State v. Hunt, 555 A.2d 369 (Vt. 1988).
[42] State v. Briggs, 776 P.2d 1347 (Wash. Ct. App. 1989).
[43] State v. Hicks, 649 P.2d 267 (Ariz. 1982).
[44] Hadden v. State, 670 So.2d 77 (Fla. Ct. App. 1996).

CARL N. EDWARDS

Profiling: Drugs *see* Drug Profiling

Profiling: Mitochondrial DNA *see* Mitochondrial DNA: Profiling

Property Crime *see* Firesetting

Pseudoseizures *see* Seizures: Behavioral

Psychodynamic Diagnostic Manual (PDM) *see* Psychopathology: Terms and Trends

Psychological Autopsy

Origins

The psychological autopsy originated in approximately 1958 as a result of the Los Angeles County Medical Examiner's Office consulting the Los Angeles Suicide Prevention Center for assistance in distinguishing drug-related accidental overdoses from suicides [1]. This collaboration laid down the basic principles of the psychological autopsy procedure. Edwin Schneidman, a director of the LA Suicide

Table 1 Commonalities of suicide

- Purpose: seek a solution
- Stimulus: unbearable psychological pain
- Stressor: frustrated needs
- Emotion: hopelessness, helplessness
- Cognition: ambivalence
- Perception: constriction

Adapted from Schneidman, 1996

Prevention Center, is credited with coining the term *psychological autopsy*. Schneidman's initial definition of the psychological autopsy was "a thorough retrospective investigation of the intention of the decedent" [2]. The process was partially based on his observations that many suicides shared certain common characteristics that could help identify a suicidal individual. Table 1 lists some of Schneidman's "commonalities of suicide".

Over the past 50 years, the procedure has become familiar to most suicidologists, suicide researchers and major city homicide investigators. However, a single standard definition has yet to be formally agreed upon. This article uses the following definition for the psychological autopsy: "A postmortem investigative procedure requiring the identification and assessment of suicide risk factors present at the time of death, with the goal of enabling a determination of the manner of death to as high a degree of certainty as possible".

Thus, the psychological autopsy can be conceptualized as synonymous with a postmortem suicide risk assessment. The strength of framing the psychological autopsy in this manner lies in the fact that performing formal suicide risk assessments on patients who are at risk for suicide is endorsed by overwhelming clinical consensus [3], and will be further clarified below in the section on current controversies. Regardless of the definition or method used, the quality of the assessment will depend heavily on the training, knowledge, experience and clinical judgment of the investigator.

In the push toward standardization, the psychiatric autopsy has evolved through a number of iterations. Initially, Schneidman developed a list of 14 areas of inquiry to guide the investigator when conducting a psychological autopsy [2]. In the late 1980s, the Centers for Disease Control established a list of 22 criteria to assist forensic investigators, called the Operational Criteria for the Determination of Suicide (OCDS) [4]. Shortly afterwards, suicide researchers developed the Empirical Criteria for the Determination of Suicide (ECDS). This instrument subsumed the OCDS, as well as other important criteria from the research literature [5].

While the Department of Defense had long employed the psychological autopsy method, in 2002 it published a sample model curriculum for conducting them, along with recommendations for training and peer review [6]. Finally, in 2006, leading suicidology experts and researchers proposed an initial standard protocol for lines of inquiry to improve reliability and validity [7]. This protocol, which has been further amended with the assistance of an expert from this research group (Berman, A. *Personal communication*, 2007), is presented toward the end of this article.

Despite progress in psychological autopsy research, some deaths (i.e., drug-related fatalities [8, 9]) continue to frustrate medical examiners (MEs). Thus, the psychological autopsy methods and protocol must continue to be vigorously pursued and tested.

Purpose

Psychological autopsies are an invaluable tool for assessing equivocal deaths. An "equivocal death" may be one in which the manner of death is questionable, or the circumstances surrounding the death are otherwise unclear [7]. Typical equivocal death scenarios are listed in Table 2.

When MEs perform autopsies, they attempt to classify the death into one of four categories or "modes": natural, accidental, suicide, or homicide (NASH) [10]. When a death cannot be immediately classified, it is often officially referred to as *undetermined*. Table 3 lists the basic elements of death classification.

Table 2 Typical equivocal death scenarios

- Drug-related deaths
- Autoerotic asphyxia
- Self-induced asphyxia (e.g., the "choking game")
- Drownings
- Vehicular deaths
- "Russian Roulette"
- "Suicide by cop"
- Staged death scenes

Table 3 Death classification

1. Cause: gunshot, poisoning, etc.
2. Mode: circumstances leading to cause
   a) Natural
   b) Accidental
   c) Suicide
   d) Homicide
3. Motive: reasons for the action
4. Lethality: probability of death as a result of method and circumstances chosen (low, medium, and high)
5. Intent: what the decedent wanted to happen at the time

The goals of the psychological autopsy include obtaining an in-depth understanding of the decedent's personality, behavior patterns, and possible motives for suicide. The investigator strives to obtain an objective analysis of the decedent's suicide risk enhancing and protective factors. In certain cases, an experienced investigator can use the method to help sort out the degree of risk, intent and causal factors at the time of death [3]. Ultimately, this should allow for a well informed assessment of whether or not the deceased was a likely candidate for suicide. Table 4 lists some of the more important goals of the psychological autopsy.

The psychological autopsy has utility in a variety of settings. Table 5 gives a list of some of its more common uses. As noted, it can be an extremely helpful tool for assisting MEs and homicide investigators. It has been shown to have a significant impact on MEs determination in equivocal cases [11]. It has been used for several decades to collect valuable research data about suicide that ultimately informs prevention efforts [12–14]. The vast majority of these studies suggest that mental disorder is present in a preponderance of suicides. The first generation of research driven by psychological autopsies found that more than 90% of completed suicides suffered from mental disorders, mostly mood disorders and substance use disorders [3, 15]. The second generation of psychological autopsy research has employed case-control designs, resulting in better estimations of the role of various risk factors for suicide [16].

Table 4 Psychological autopsy goals

- Identify behavior patterns – stress reactions, adaptability, habit, or routine changes
- Establish presence or absence of mental illness
- Identify possible precipitants
- Determine presence or absence of motives
- Determine presence or absence of suicidal intent
- Determine suicide risk factors – both mitigating and aggravating
- Perform a postmortem suicide risk assessment
- Establish whether or not the deceased was a likely candidate for suicide

Table 5 Psychological autopsy uses

- Assist medical examiners with "equivocal" deaths
- Research on suicide
- Insurance claims
- Criminal cases
- Estate issues, contested wills
- Malpractice claims
- Worker's compensation cases
- Product liability cases
- Organizational suicide prevention efforts

The psychological autopsy method has also been used in a forensic context in both criminal and civil courts. While courts have admitted testimony based on psychological autopsies in many civil cases, criminal courts have been more hesitant [17]. The issue of the psychological autopsy's legal admissibility will be further discussed below. In criminal cases, the psychological autopsy may be used to establish whether a decedent was likely to have committed suicide, or whether the matter should be viewed as a homicide. In some criminal cases, most notably *Jackson v. State* (Fla. 4th DCA 1989), the psychological autopsy has been used to help analyze whether an abusive relationship played a role in a suicide [18]. In the criminal case *U.S. v. St. Jean* (US Ct. of Appeals for Armed Forces, 1996), the psychological autopsy was used by the prosecution to assist in determining whether or not a suspected homicide victim had been a likely candidate for suicide.

In civil cases, the psychological autopsy has been used to help determine whether benefits are owed to the decedent's beneficiaries [10]. This often involves life insurance payments, where many policies hold that a suicide precludes benefits. However, some policies permit payment if it can be proven that the decedent's death was an "insane suicide". "Insane suicide" is a legal term that was defined by the US Supreme Court in the case of *Mutual Life Insurance Company v. Terry* (US 1872; 82: 580). The Court held that a suicide was "sane" when the "assured, being in the possession of his ordinary reasoning faculties,

from anger, pride, jealousy, or a desire to escape from the ills of life, intentionally takes his own life."

In contrast, an "insane" suicide was defined as "when his reasoning faculties are so far impaired that he is not able to understand the moral character, the general nature, consequences, and effect of the act he is about to commit, or when he is impelled thereto by an insane impulse, which he has not the power to resist . . ." Thus, a sane suicide implies the decedent had a rational understanding that his acts would result in his death, whereas an insane suicide implies the decedent was so emotionally disturbed that he did not have a rational appreciation of his actions [10].

Worker's compensation cases generally involve allegations that the decedent's employer was somehow legally responsible for his suicide. Similarly, product liability cases allege that the decedent's use of a particular product (e.g., medication) caused him to commit suicide. In psychiatric malpractice cases involving suicide, the plaintiff must prove that the doctor's negligence was a proximate cause of the decedent's suicide [19]. In addition to determinations regarding the standard of care, a postmortem suicide risk assessment must be conducted to determine the decedent's overall suicide risk and the foreseeability of the suicide.

Another potential use for the psychological autopsy may be its clinical utility in helping surviving family members better understand the tragedy and begin the grieving process [20]. The psychological autopsy may be used for other clinical purposes, such as informing an institution's morbidity and mortality conference after a client's suicide. For historical purposes and interest, psychological autopsies have been conducted on numerous public figures such as Ernest Hemingway [21], Vince Foster, (Berman, A. *Personal communication*, 2007), Howard Hughes and Marilyn Monroe.

Methods

The psychological autopsy method involves collecting and analyzing all relevant information on the deceased. This means that all applicable records are reviewed, including medical records, psychiatric records, police records, and autopsy findings. A visual inspection of the death scene via photographs is necessary, and occasionally a visit to the scene will be required. A thorough review of the decedent's writings in the form of diaries, journals, e-mails and internet correspondence is vital. The suggested protocol at the end of this article provides a list of other important sources of data. In addition to reviewing records, structured interviews of family members, relatives or friends are necessary. Thus, a psychological autopsy synthesizes data from multiple informants and records. When performed in a comprehensive manner, the method may take anywhere from 20 to 50 or more hours to complete. The overriding principle is that the greater the amount of relevant data analyzed, the more accurate the investigator's conclusions are likely to be.

Suicide risk factors vary among different populations [22]. The investigator should consider any special nuances of the deceased, such as age group [23], mental health diagnosis, gender [24] and other factors that may allow for a more precise consideration of risk factors associated with that group. This requires keeping up to date with the evolving psychiatric and suicidology literature, which is steadily becoming more detailed about risk factors in distinct diagnostic categories such as depression [25], bipolar disorder [26], and persons who are outside the care of mental health services [27]. Some individuals may display unique, individualized behaviors suggestive of increased or decreased suicide risk that will be known only by close social contacts or treating mental health professionals [22]. Thus, an understanding of the decedent's individualized risk factors and past stress reaction patterns becomes important.

Suicide Notes

There is a considerable literature on suicide notes. Research suggests that suicide notes are left only by a minority, approximately 10–33% of all suicides [15]. Regarding persons who do leave notes, available research has not found any significant differences when compared to suicides who do not leave notes. A few studies have suggested that whites and women are slightly more likely to leave suicide notes. At least one study has suggested there are no significant differences in the themes of notes between male and female suicides [28]. Themes of love and relationships were found to be more common than achievement themes in both men and women [29]. Another study found that suicide notes written by young people were longer and rich in emotions, whereas notes

written by the elderly were shorter, contained specific instructions, and were less emotional [30].

In a study of 42 suicide notes, the most common themes were: "apology/shame" (74%), "love for those left behind" (60%), "life too much to bear" (48%), "instructions regarding practical affairs postmortem" (36%), "hopelessness/nothing to live for" (21%), and "advice for those left behind" (21%) [31]. The common usage of computers, the Internet [32] and various electronic types of messaging have introduced another important source of data for the psychological autopsy. The investigator should not fail to inquire about these potential sources of information, as they may provide critical insight into the decedent's state of mind and intentions.

One important caveat regarding suicide notes is the possibility that they have been fraudulently prepared and left by another person attempting to disguise a homicide. While there has been some attempt to develop a method for distinguishing genuine from simulated suicide notes [33], more research in this area is required. In such cases, collaboration with a forensic handwriting analysis expert might be considered. Another important consideration for the investigator is the possibility that family members or others who find a suicide note may destroy it or remove it from the death scene for various motives.

Collateral Interviews

The importance of collateral interviews as part of the psychological autopsy method cannot be overstated. Careful interviews of the decedent's family members and other relevant social contacts distinguish a proper psychological autopsy from a mere analysis of demographic data and police reports. Most experts recommend a structured or semistructured approach to collateral interviews. At least one study has developed a semistructured interview for the psychological autopsy which has demonstrated inter-rater reliability [34]. For research purposes, there has been a trend toward using modified instruments such as the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, 4th ed.-Text Revision Disorders (SCID), as well as a life events calendar method, which helps identify and quantify the burden of events that may be associated with suicide [35].

Regardless of method used, collateral interviews often reveal critical information about the decedent that cannot be obtained elsewhere. Recent theories about suicide have stressed that psychiatric illness alone is not enough to fully explain an individual suicide. Rather, a stress-diathesis model has been proposed, in which the risk for suicidal acts is determined by the interplay of biopsychosocial factors and situational variables [36]. According to this model, a diathesis may be reflected in an individual's tendency to have maladaptive responses to stressors, such as acting impulsively. Such information is most likely to be obtained via collateral interviews.

Ethics and Sensitivity. An important ethical and practical consideration related to gathering collateral data is the manner in which collateral sources should be contacted and interviewed. Interviewing surviving family and friends is a very sensitive matter that must consider the survivor's reactions. Ideally, the investigator should have adequate clinical experience in order to handle survivors' reactions with appropriate sensitivity [37]. For "research purposes", it has been recommended that a two to six-month time interval between the suicide and interview be used [37]. There does not appear to be a significant relationship between the timing of the interview and the quality of information obtained when this time frame is used [38].

While a concern about untoward emotional reactions to the interview is an obvious concern, some have noted that survivors appeared to have benefited from the interview experience in terms of being able to express their feelings and receive a mental health referral if needed [37]. Presently, there is no clearly agreed upon method for initiating contact with survivors. Investigators performing a psychological autopsy for forensic legal purposes will likely be supplied with relevant phone numbers and/or addresses of potential interviewees. Often, attorneys will have previously informed the interviewees that an investigator will be contacting them. Investigators seeking interviews for research purposes may consider a letter followed by a phone call, or vice versa [10].

Postmortem Suicide Risk Assessment

A comprehensive postmortem suicide risk assessment is necessary because of the fact that there is no single pathognomonic risk factor for suicide [22]. Single risk factors do not have adequate statistical power on which to base conclusions. Particularly in the context

of forensic expert testimony, the postmortem suicide risk assessment approach is recommended [3]. This involves a careful identification and assessment of suicide risk factors present at the time of death.

Risk enhancing factors (both proximal and distal) should be carefully weighed along with risk reducing factors. When thoughtfully analyzed in the context of the totality of the decedent's circumstances, the investigator should be able to arrive at conclusion about the decedent's overall risk of suicide near the time of death. This ultimately informs the investigator's opinion about whether or not the decedent was a likely candidate to commit suicide at the time in question.

Testimony that focuses on whether the psychological autopsy yielded results consistent with an individual who committed suicide is more likely to be found admissible in court. In contrast, overreaching opinions that conclude the decedent did or did not commit suicide are more likely to be found inadmissible. For example, in the case of *State v. Guthrie* (2001 SD 61, 627 N.W. 2d 401), the court found that the expert's testimony became inadmissible when it "shifted from discussing typical suicide characteristics", to a "bold declaration" that the decedent did not die by suicide [39].

Limitations and Controversies

In both research and forensic legal settings, mental health professionals are legally and ethically obligated to note the limitations of their methods. In the case of the psychological autopsy, controversy over its limitations has existed for as long they have been performed [40]. Commonly cited controversies involving the psychological autopsy are listed in Table 6. The limitations largely involve the fact that there is no unanimously accepted standardized protocol for conducting a psychological autopsy. However, dedicated efforts are currently underway to resolve this issue [7]. Progress in this area may be somewhat dependent on the field of suicidology developing a standard, comprehensive nomenclature. For example, even the basic term *suicide attempt* may have different meanings to different investigators [41].

Table 6 Current controversies [42]

- Lack of standardized protocol
- Lack of standardized suicidology nomenclature
- Methodological problems
- Reliability of assessment instruments
- Lack of homogeneity among studies
- Bias among collateral informants

Controversy surrounding admissibility and meeting Daubert standards may be best resolved by adopting the postmortem suicide risk assessment approach. The reasoning for this is as follows: (i) in clinical practice, the standard of care requires the clinician to gather relevant information and assess the patient's level of suicide risk; (ii) performing such suicide risk assessments on patients considered to be at risk for suicide is endorsed by overwhelming clinical consensus (i.e., it is "generally accepted"); and (iii) the psychological autopsy is similar in that it is the assessment of suicide risk factors present in the decedent near the time of death [3].

This approach to the psychological autopsy should meet the process-driven Daubert criteria, particularly where *Kumho Tire Co. v. Carmichael* (1999 536 U.S. 137, 141) has held that Daubert standards are not limited only to scientific evidence, but may include "technical, or other specialized knowledge" [43]. In the event that a court found that the postmortem suicide risk assessment did not meet Daubert criteria, it should at least meet the *Frye v. U.S.* (DC COA, 1923) criteria of "general acceptance" among mental health professionals. This line of reasoning was adopted by a Louisiana appeals court when it found the psychological autopsy admissible under Daubert. The case, *In re Succession of Pardue* (La Ct. App. 2005 915 So. 2d 415), involved estate issues and testamentary capacity. The court held that the methodology used in the psychological autopsy was sufficiently reliable and generally accepted in psychiatry.

In federal court where expert testimony is governed by the Federal Rules, opinions on the basis of the postmortem suicide risk assessment method would appear to comply with Federal Rules 703 and 702, which allows expert opinions if they are "of a type reasonably relied upon by experts in the particular field", and if the testimony is "the product of reliable principles and methods" [44].

An obvious criticism of the psychological autopsy is the problem of not being able to interview the decedent so as to more accurately determine intent. One response to this issue is that the comprehensive nature of the psychological autopsy, with its

wide net of collateral data, may ultimately allow a reasonable approximation or inference of intent. Additionally, it has been argued that having no prior contact with the decedent removes the vagueness and subjectivity inherent in an interpersonal relationship with the decedent [18]. Therefore, one might argue that psychological autopsies may be more objective and less controversial than the analysis of living patients.

To date, only a restricted number of research studies have relied on standardized instruments, making the comparison of findings problematic [42]. Current studies are increasingly using the SCID, but this approach has yet to be conclusively validated for administration to proxies. However, at least two studies (of psychiatric inpatients admitted following a suicide attempt) have suggested that proxies are good judges of past history of suicide attempts, level of suicidal intent and data leading to a psychiatric diagnosis [45, 46].

The issue of bias in collateral informants requires a cautious approach by investigators. Informants' reports may be biased by many factors such as their own personal attitudes toward suicide and their emotional state at the time of the interview. Further, their recollection of circumstances may be impacted by the emotional trauma of the death of their loved one [42]. The demeanor and interpersonal style of the investigator may also become a factor. It is possible that informants may react to the investigator's personal characteristics, which may then influence the amount and type of information they are willing to divulge.

In summary, there is a considerable need for further research efforts aimed at improving the validity and reliability of the psychological autopsy [17]. There is concern that testimony based on a faulty or inadequate psychological autopsy may result in a "miscarriage of justice" [47]. Forensic experts who give testimony in court on the basis of psychological autopsy findings must be prepared to concede any relevant limitations.

Toward a Standard Protocol

In the interest of improving validity and reliability, forensic suicidology must continue to work toward an accepted, standardized protocol for the psychological autopsy [7]. Below is a recommended template that was developed by expert consensus [7], (Berman, A. *Personal communication*, 2007) and further enhanced according to relevant research and clinical findings [5–7, 10, 20, 48–50]. The protocol may be used to guide areas of investigation and lines of inquiry for collateral interviews. Use of a standardized protocol can serve as a framework for conducting a psychological autopsy. Ultimately, a standard protocol will enhance admissibility, and aid in the testability of the method [7].

The psychological autopsy method allows for a "polyperspective" that can help illuminate key aspects of an equivocal death [51]. However, the light that it shines may also be distorted by the perspective of the investigator. Therefore, it is critical that the individual performing the psychological autopsy possess adequate training. This includes having sufficient knowledge in the fields of suicidology, related mental health concepts, and basic death scene investigation. Data obtained from the death scene and physical autopsy are often highly determinative, and meticulous inspection of the death scene and related items may be necessary in certain forensic legal cases. For example, in cases of suspected "simulated suicidal hanging", there may be important evidence suggesting a homicide that was staged to appear as though it was a suicide [52]. Detailed investigation of body position, ligature placement and knot formation may be required to distinguish a suicide from a homicidal hanging [53]. Further, in cases of suspected autoerotic asphyxiation, researchers have noted characteristic death scene findings (see Appendix 1) [54].

Finally, the problems inherent in assessments of suicidal intention require that the investigator be cognizant of the limits of the data and corresponding conclusions. This is primarily because of clinical observations that suicidal individuals are often ambivalent, and may even have multiple intentions at the same time [55]. In certain cases, the presence and degree of suicidal intent may be difficult to determine because of ambivalence, denial, minimization or confusion. In contrast, the nuances of other cases may present a rather straightforward inference that the decedent's intent was to die. In difficult cases, the investigator may choose to simply report on the results of the postmortem suicide risk assessment, or provide reliable evidence that "Natural", "Accidental" and "Homicidal" causes of death can be excluded [56].

Psychological Autopsy Protocol

Records and documents

- Medical records
- Mental health records
- Police records and related witness statements
- Legal records
- Criminal records
- School records
- Financial records
- Military records
- Suicide note(s), communication or video
- Decedent's journals, diary
- Electronic data: E-mails, text messages, instant messages, and websites
- Forensic computer analysis report if available
- Autopsy report
- Toxicology report

Death scene

- Photos of death scene and site visit if necessary
- Presence of atypical wounds (see Appendix 1)
- Decedent's relationship to site
- Evidence of rescuability versus precautions taken against
- Evidence of planning and/or rehearsal
- Evidence of staged manner of death (see Appendix 1)

Demographics

- Socioeconomic status
- Employment status
- Financial status
- Age/gender/race/height/weight
- Marital status
- Educational status
- Religion and religiosity
- Adopted vs. biological family status
- Immigrant status – acculturation issues
- Residence relative to recent mobility

Recent symptoms and behaviors

- Appeared depressed, sad, tearful, or moody
- High risk depressive symptoms [25]
  - Insomnia
  - Appetite loss
  - Weight loss
  - Feelings of worthlessness and/or inappropriate guilt
  - Physical agitation
  - Depression comorbid with anxiety
- Expressed suicidal ideation or preoccupation with death
- Appeared to have made a change for the better
- Appeared anxious, or complained recently of anxiety or panic attacks
- Appeared agitated
- Behaved in an impulsive manner
- Displayed uncontrolled rage or aggressive behavior
- Demonstrated constricted thinking or "tunnel vision"
- Disclosed feelings of guilt or shame
- Appeared confused, disoriented or psychotic
- Expressed feelings of hopelessness, helplessness or worthlessness
- Engaged in excessive risk-taking behaviors
- Mental Status evidence of:
  - Impaired memory
  - Poor comprehension
  - Poor judgment
  - Hallucinations or delusions
  - Inflated sense of self or signs of magical thinking

Precipitants

- Significant losses (relationships, job, finances, prestige, self-concept, family member, moving, or anything of importance)
- Significant (or perceived) disruption of a primary relationship
- Legal troubles or difficulties with police
- Traumatic events
- Significant life changes (negative or positive, birth of child, promotion, etc.)
- Completed or attempted suicide by a family member or loved one
- Anniversary of important death, loss, etc.
- Exposure to suicide of another via media or personal acquaintance
- Preparations for death (e.g., gave away prized items, settled personal accounts, updated will, and said "goodbye" to loved ones)
- Expression of wish to reunite with a deceased loved one, or to be "reborn"

Psychiatric history

- Prior suicidal behaviors

  - Total number of past attempts
  - Dates
  - Precipitants
  - Method, lethality
  - What stopped event, if anything? How found
  - Attitude and behavior after found
- Prescribed psychotropic medications
- Observed adverse reactions to psychotropic medications
- Lack of compliance with psychotropic medications
- Efficacy of treatment (e.g., subtherapeutic doses, poor or incorrect agent choice, inadequate blood level, etc.)
- Psychiatric hospitalization (reasons, dates, diagnoses, and treatment)
- Outpatient treatment (psychiatrist, psychologist, therapist, and Primary Care Physician PCP)
- Psychotherapy at time of death (duration, quality of alliance, compliance, and diagnosis)
- Expressing concerns about "going crazy" or losing cognitive function

Physical health

- Recent visit to physician (reasons)
- Chronic pain
- Chronic, fatal or debilitating disease
- Recent reduction in physical/functional capabilities
- Current medications (compliance or recent changes)

Substance abuse

- History and pattern of alcohol, drug abuse
- Recent attempts to discontinue use
- Recent increase in pattern of use
- Degree of use at time of death (binge drinking, etc.)
- History of "accidental overdose" (when and type of drug)

Family history

- Suicide or attempted suicide
- Nonnatural deaths
- Level of support or observed closeness in nuclear and extended families
- Physical, sexual, or emotional abuse
- Substance abuse
- Violent behavior
- Affective or other psychiatric disorders

Firearm history (if relevant)

- Ownership
- Recent purchasing or obtaining (stated purpose?)
- Recent movement of gun (from where to where?)
- Pattern of weapon care and cleaning
- Pattern of storage and use
- Accidental discharges

Social Supports and Attachments

- Ability to create and maintain close personal relationships
- Ability to express feelings as needed in relationships
- Recent talk about feeling unsupported, uncared for, unimportant
- Relative success in personal relationships
- Relative success in work
- Attachment to hobbies, interests, religion, etc.
- Recent change in any of the above attachments/supports

Emotional reactivity

- History of violence toward others
- Impulsive behaviors
- Excessive rage or aggression

Lifestyle/character

- Typical coping patterns, pattern of reaction to stress
- Perfectionism
- Self-destructive behaviors (self-mutilation, deliberate self-harm, driving while intoxicated, etc.)
- Frequent crises
- Victimization behaviors (bullied or abused)
- Tendency to dissembling (hiding emotions or stoicism)

Access to care

- History of help-seeking behaviors
- Barriers to healthcare (no insurance or no accessible caregiver)

Other areas of inquiry

- Occupational history
- Personal interests, hobbies
- Gambling history
- Degree and type of religiosity
- Description of activities/behaviors in last days before death

Factors associated with suicide risk reduction

- Evidence of future-oriented thinking or behaviors (doctor's appointments, job interviews, etc.)
- Responsibility for a child under 18 (stronger for women vs. men)
- Absence of suicidal ideas or intent
- Hopefulness
- Willingness to accept help and/or treatment for psychiatric conditions
- Low symptom severity
- Good therapeutic alliance with mental health professional
- Stable, supportive marriage or spouse
- Religious prohibition

Collateral interviews

For each interview, note the following:

- Relationship to deceased
- Time interval between death and interview
- Reactions to the death (surprise, acceptance, and beliefs)
- Attitudes about suicide
- Potential biases (pending lawsuits, insurance claims, denial, etc.)
- Assessment instruments used

Appendix 1

Evidence of Alternative or Staged Manner of Death

- **Presence of Autoerotic Characteristics** [57]
  - Body partially supported by ground
  - Ligature with self-rescue mechanism (slip knot, etc.)
  - Bondage items and/or sexual masochistic behavior (genitals, nipples, etc.)
  - Male wearing female attire
  - Protective padding between ligature and body
  - Sexual paraphernalia (vibrator, pornography, and mirrors)
  - Evidence of previous autoerotic practices
- **Presence of Atypical Self-Inflicted Gunshot Wounds** [58]
  - More than one gunshot injury
  - Gunshot injury without contact or near contact
  - Uncommon entrance wound sites (back of neck or head, ear, and eye)
  - Uncommon bullet paths (downward and back to front)

Acknowledgment

The authors would like to acknowledge Alan L. Berman, PhD for his expertise and assistance, particularly with the Psychological Autopsy Protocol.

References

[1] Schneidman, E. (1996). *The Suicidal Mind*, Oxford University Press, New York.

[2] Schneidman, E. (1981). The psychological autopsy, *Suicide and Life-Threatening Behavior* **11**, 325–340.

[3] Simon, R. (2002). Murder, suicide, accident, or natural death? Assessment of suicide risk factors at the time of death, in *Retrospective Assessment of Mental States in Litigation*, R. Simon & D. Shuman, eds, American Psychiatric Publishing, Washington, D.C, pp. 135–153.

[4] Rosenberg, M., Davidson, L., Smith,J.C., Berman, A.L., Buzbee, H., Gantner, G., Gay, G.A., Moore-Lewis, B., Mills, D.H., Murray, D., O'Carroll, P.W. & Jobes, D. (1988). Operational criteria for the determination of suicide, *Journal of Forensic Sciences* **33**(6), 1445–1456.

[5] Jobes, D., Casey, J. & Berman, A. *et al.* (1991). Empirical criteria for the determination of suicide manner of death, *Journal of Forensic Sciences* **36**(1), 244–256.

[6] Ritchie, E. & Gelles, M. (2002). Psychological autopsies: the current department of defense effort to standardize training and quality assurance, *Journal of Forensic Sciences* **47**(6), 1370–1372.

[7] Snider, J., Hane, S. & Berman, A. (2006). Standardizing the psychological autopsy: Addressing the daubert standard, *Suicide & Life-Threatening Behavior* **36**(5), 511–518.

[8] Cone, E. *et al.* (2004). Oxycodone involvement in drug abuse deaths. II. Evidence for toxic multiple drug-drug interactions, *Journal of Analytical Toxicology* **28**(7), 616–624.

[9] Wolf, B.C., Lavezzi, W.A., Sullivan, L.M. & Flannagan, L.M. (2005). One hundred seventy two deaths involving the use of oxycodone in palm beach County, *Journal of Forensic Sciences* **50**(1), 192–195.

[10] Scott, C., Swartz, E. & Warburton, K. (2006). The Psychological autopsy: solving the mysteries of death, *The Psychiatric Clinics of North America* **29**(3), 805–822.

[11] Jobes, D., Berman, A. & Josselson, A. (1986). The impact of psychological autopsies on medical examiners' determination of manner of death, *Journal of Forensic Sciences* **31**(1), 177–189.

[12] Conwell, Y., Duberstein, P., Cox, C., Herrmann, J., Forbes, N. & Caine, E. (1996). Relationships of age and axis I diagnoses in victims of completed suicide: a psychological autopsy study, *The American Journal of Psychiatry* **153**, 1001–1008.

[13] Andrew, T.A., Cheng, Tony, H.H., Chen, Chwen-Chen Chen & Rachel Jenkins (2000). Psychosocial and psychiatric risk factors for suicide: case – control psychological autopsy study, *The British Journal of Psychiatry* **177**, 360–365.

[14] Henriksson, M., Aro, H., Marttunen, M., Heikkinen, M., Isometsa, E., Kuoppasalmi, K. & Lonquist, J. (1993). Mental disorders and comorbidity in suicide, *The American Journal of Psychiatry* **150**, 935–940.

[15] Berman, A. (2000). *Comprehensive Textbook of Suicidology*, The Guilford Press, New York, pp. 268–270, 319–320.

[16] Isometsa, E. (2001). Psychological autopsy studies – a review, *European Psychiatry : The Journal of The Association of European Psychiatrists* **16**(7), 379–385.

[17] Ogloff, J. & Otto, R. (1993). Psychological autopsy: clinical and legal perspectives, *St. Louis University Law Journal* **37**, 607–646.

[18] Jacobs, D. & Kline-Benheim, M. (1995). The Psychological autopsy: a useful tool for determining proximate causation in suicide cases, *Bulletin of the American Academy of Psychiatry and the Law* **23**(2), 165–182.

[19] Knoll, J. & Gerbasi, J. (2006). Psychiatric malpractice case analysis: striving for objectivity, *The Journal of the American Academy of Psychiatry and the Law* **34**, 215–223.

[20] Berman, A. (2005). Psychological autopsy, in *Encyclopedia of Forensic and Legal Medicine*, J. Payne-James, R. Byard, T. Corey & C. Henderson, eds, Academic Press, pp. 364–371.

[21] Martin, C. (2006). Ernest hemingway: a psychological autopsy of a suicide, *Psychiatry* **69**(4), 351–361.

[22] Simon, R. (2004). *Assessing and Managing Suicide Risk: Guidelines for Clinically Based Risk Management*, American Psychiatric Publishing, Washington, D.C.

[23] Houston, K., Hawton, K. & Shepperd, R. (2001). Suicide in young people aged 15–24: a psychological autopsy study, *Journal of Affective Disorders* **63**(1–3), 159–170.

[24] Pirkola, S., Isometsa, E., Heikkinen, M., Henriksson, M., Marttunen, M. & Lonquist, J. (1999). Female psychoactive substance-dependent suicide victims differ from male–results from a nationwide psychological autopsy study, *Comprehensive Psychiatry* **40**(2), 101–107.

[25] McGirr, A., Renaud, J., Seguin, M., Alda, M., Benkelfat, C., Lesage, A. & Turekci, G. (2007). An examination of DSM-IV depressive symptoms and risk for suicide completion in major depressive disorder: A psychological autopsy study, *Journal of Affective Disorders* **97**, 203–209.

[26] Marangell, L.B., Bauer, S.M., Dennehy, E.B., Wisniewski, S.R., Allen, M.H., Miklowitz, D.J., Oquendo, M.A., Frank, E., Perlis, R.H., Martinez, J.M., Fagiolini, A., Otto, M.W., Chessick, C.A., Zboyan, H.A., Miyahara, S., Sachs, G. & Thase, M.E. (2006). Prospective predictors of suicide and suicide attempts in 1, 556 patients with bipolar disorders followed for up to 2 years, *Bipolar Disorders* **8**, 566–575.

[27] Owens, C., Booth, N., Briscoe, M., Lawrence, C. & Lloyd, K. (2003). Suicide outside the care of mental health services: a case-controlled psychological autopsy study, *Crisis* **24**(3), 113–121.

[28] Salib, E., Cawley, S. & Healy, R. (2002). The significance of suicide notes in the elderly, *Aging and Mental Health* **6**(2), 186–190.

[29] Canetto, S. & Lester, D. (2002). Love and achievement motives in women's and men's suicide notes, *The Journal of Psychology* **136**(5), 573–576.

[30] Ho, T., Yip, P., Chiu, C. & Halliday, P. (1998). Suicide notes: what do they tell us? *Acta Psychiatrica Scandinavica* **98**(6), 467–473.

[31] Foster, T. (2003). Suicide note themes and suicide prevention, *International Journal of Psychiatry in Medicine* **33**(4), 323–331.

[32] Baume, P., Cantor, C. & Rolfe, A. (1997). Cybersuicide: the role of interactive suicide notes on the Internet, *Crisis* **18**(2), 73–79.

[33] Lester, D. & Linn, M. (1998). Joseph Richman's signs for distinguishing genuine from simulated suicide notes, *Perceptual and Motor Skills* **87**(1), 242.

[34] Werlang, B. & Botega, N. (2003). Semistructured interview for psychological autopsy: an inter-rater reliability study, *Suicide & Life-Threatening Behavior* **33**(3), 326–330.

[35] Seguin, M., Lesage, A., Turecki, G., Bouchard, M., Chawky, N., Tremblay, N., Daigle, F. & Guy, A. (2007). Life trajectories and burden of adversity: mapping the developmental profiles of suicide mortality, *Psychological Medicine* **37**(11), 1575–1583.

[36] Mann, J.J., Waternaux, C., Haas, G.L. & Malone K.M. (1999). Toward a clinical model of suicidal behavior in psychiatric patients, *The American Journal of Psychiatry* **156**, 181–189.

[37] Beskow, J., Runeson, B. & Asgard, U. (1990). Psychological autopsies: methods and ethics, *Suicide & Life-Threatening Behavior* **20**(4), 307–323.

[38] Brent, D.A., Perper, J.A., Kolko, D.J. & Zelenak, J.P. (1988). The psychological autopsy: methodological considerations for the study of adolescent suicide, *Journal of the American Academy of Child and Adolescent Psychiatry* **27**(3), 362–366.

[39] Kern, J. & Swier, S. (2004). Daubert, kuhmo, and its impact on south dakota jurisprudence: an update, *South Dakota Law Review* **49**, 217–249.

[40] Hansen, M. (2000). Suicidal Missions: psychological autopsies to uncover motivation in suspicious deaths are themselves now suspect, *ABA Journal* **86**, 28–29.

[41] Silverman, M. (2006). The language of suicidology, *Suicide and Life-Threatening Behavior* **36**(5), 519–532.

[42] Pouliot, L. & De Leo, D. (2006). Critical issues in psychological autopsy studies, *Suicide and Life-Threatening Behavior* **36**(5), 491–510.
[43] Kumho Tire Co. V. Carmichael, 536 U.S. 137, 141 (1999).
[44] Federal Rules of Evidence, Available at 2008. http://www.law.cornell.edu/rules/fre/rules.htm#Rule703.
[45] Conner, K., Conwell, Y. & Duberstein, P. (2001). The validity of proxy-based data in suicide research: a study of patients 50 years of age and older who attempted suicide. II. Life events, social support and suicidal behavior, *Acta Psychiatrica Scandinavica* **104**(6), 452–457.
[46] Conner, K., Duberstein, P. & Conwell, Y. (2001). The validity of proxy-based data in suicide research: a study of patients 50 years of age and older who attempted suicide, *I Psychiatric Diagnosis Acta Psychiatrica Scandinavica* **104**(3), 204–209.
[47] Ormerod, D. (2001). Psychological autopsies: legal applications and admissability, *The International Journal of Evidence and Proof* **5**, 1–31.
[48] Moscicki, E. (1997). Identification of suicide risk factors using epidemiologic studies, *Psychiatric Clinics of North America* **20**(3), 499–517.
[49] American Psychiatric Association (APA) (2003). Practice guideline for the assessment and treatment of patients with suicidal behaviors. *The American Journal of Psychiatry* Available at http://www.psych.org/psych_pract/treatg/pg/SuicidalBehavior_05-15-06.pdf.
[50] Shea, S. (2002). *The Practical Art of Suicide Assessment: A Guide for Mental Health Professionals and Substance Abuse Counselors*, John Wiley & Sons, New Jersey.
[51] Berman, A. (1993). Forensic suicidology and the psychological autopsy, in *Suicidology: Essay in Honour of Edwin S. Schneidman*, A. Leenaars, ed, Aronson, New York.
[52] Hazelwood, R. & Napier, M. (2005). Crime scene staging and its detection, *International Journal of Offender Therapy and Comparative Criminology* **48**(6), 744–759.
[53] Vanezis, P. & Busuttil, A. (1996). *Suspicious Death Scene Investigation*, Arnold, London, p. 153.
[54] Byrard, R., Hucker, S. & Hazelwood, R. (1990). A comparison of typical death scene features in cases of fatal male and autoerotic asphyxia with a review of the literature, *Forensic Science International* **48**(2), 113–121.
[55] Andriessen, K. (2006). On "Intention" in the definition of suicide, *Suicide and Life-Threatening Behavior* **36**(5), 533–538.
[56] O'Carrol, P., Berman, A., Maris, M., Moscicki, E., Tanney, B. & Silverman, M. (1996). Beyond the tower of babel: a nomenclature for suicidology, *Suicide and Life-Threatening Behavior* **26**, 237–252.
[57] Hazelwood, R., Dietz, P. & Burgess, A. (1983). *Autoerotic fatalities*, Lexington Books, Lexington, MA.
[58] Karger, B., Billeb, E., Koops, B. & Brinkman, B.(2002). Autopsy features relevant for discrimination between suicidal and homicidal gunshot injuries, *International Journal of Legal Medicine* **116**, 273–278.

Further Reading

Legal Citations

Campbell v. Young Motor Co, 211 Mont. 68, 684 P.2d 1101 (1984).
Daubert v. Merrell Dow (USSC., 1993).
Frye v. United States (D.C. COA., 1923).
In re Succession of Pardue (La Ct. App. 915 So. 2d 415 2005).
Jackson v. State, 553 So. 2d 719 (Fla. 4th DCA., 1989).
Kumho Tire Co. v. Carmichael 536 U.S. 137, 141 (1999).
Mutual Life Insurance Company v. Terry (U.S. 82, 580 1872).
State v. Guthrie (SD 61, 627 N.W. 2d 4012001).
United States v. St. Jean (U.S. Ct. of Appeals for Armed Forces, 1996).

Additional References

Cavanagh, J., Carson, A., Sharpe, M. & Lawrie, S. (2003). Psychological autopsy studies of suicide: a systematic review, *Psychological Medicine* **33**(3), 395–405.
Conner, K., Cox, C., Duberstein, P., Tian, L., Nisbet, P. & Conwell, Y. (2001). Violence, alcohol, and completed suicide: a case-control study, *The American Journal of Psychiatry* **158**, 1701–1705.
Conwell, Y., Duberstein, P., Cox, C., Herrmann, J., Forbes, N. & Caine, E. (1996). Relationships of age and axis I diagnoses in victims of completed suicide: a psychological autopsy study, *The American Journal of Psychiatry* **153**, 1001–1008.
Fruehwald, S., Matschnig, T., Koeni, F., Bauer, P. & Frottier, P. (2004). Suicide In custody: case-control study, *The British Journal of Psychiatry* **185**, 494–498.
He, X.Y., Felthouse, A.R., Holzer, C.E., Nathan, P. & Veasey, S. (2001). Factors in prison suicide: one year study in Texas, *Journal of Forensic Sience* **46**(4), 896–901.
Kovasznay, B., Miraglia, R., Beer, R. & Way, B. (2004). Reducing suicides in New York State correctional facilities, *Psychiatric Quarterly* **75**(1), 61–70.
Litman, R. (1989). 500 psychological autopsies, *Journal of Forensic Sciences* **34**(3), 638–646.
Malone, K.M., Oquendo, M.A., Haas, G.L., Ellis, SP., Li, S. & Mann, J.J. (2000). Protective factors against suicidal acts in major depression: reasons for living, *The American Journal of Psychiatry* **157**, 1084–1088.
Marttunen, M., Henriksson, M., Isometsa, E., Heikkinen, M., Aro, H. & Lonnqvist, J. (1998). Completed suicide among adolescents with no diagnosable psychiatric disorder, *Adolescence* **33**(131), 669–681.
Ohberg, A. & Lonquist, J. (1998). Suicides hidden among undetermined deaths, *Acta Psychiatrica Scandinavica* **98**(3), 214–218.
Owens, C., Booth, N., Briscoe, M., Lawrence, C. & Lloyd, K. (2003). Suicide outside the care of mental health services: a case-controlled psychological autopsy study, *Crisis* **24**(3), 113–121.

Shaw, J., Baker, D., Hunt, I.M., Moloney, A. & Appleby, L. (2004). Suicide by prisoners: National clinical survey, *British Journal of Psychiatry* **184**, 263–267.

Spellman, A. & Heyne, B. (1989). Suicide? Accident? Predictable? Avoidable? The psychological autopsy in jail suicides, *The Psychiatric Quarterly* **60**(2), 173–183.

Weinberger, L.E., Sreenivasan, S., Sathyavagiswaran, L. & Markowitz, E. (2001). Child and adolescent suicide in a large, urban area: Psychological, demographic, and situational factors, *Journal of Forensic Science* **46**(4), 902–907.

Related Articles

Autoerotic Deaths

Accident Reconstruction

Crime Scene Investigation

Mass Grave Investigation

Suicide (Behavior)

JAMES L. KNOLL, IV AND
ROBERT R. HAZELWOOD

Psychological First Aid *see* Disaster Mental Health

Psychological Profiles *see* Profiles: Psychological and Behavioral

Psychological Syndromes *see* Syndromes: Psychological

Psychological Testing

Introduction

Psychological testing is best understood within the larger context of psychological evaluation or assessment that, in turn, is seen as one discipline's approach to a more general and shared task of information collection. The *APA Dictionary of Psychology* [1] defines psychological assessment as "the gathering and integration of data in order to make a psychological evaluation, decision, or recommendation" (p. 751). Multiple tools of assessment are listed including interview, behavioral observations, tests, and other specialized instruments. A psychological test is a "standardized instrument (i.e., a test, inventory, or scale)" used for the purpose of measuring any of a variety of abilities, aptitudes, or attributes (p. 753).

This article includes discussion of psychological testing and differentiates it from other types of assessment and focuses on one area of personality testing (*see*, in contrast, **Neuropsychological Assessment**; **Neuropsychological Assessment: Child**; **Head Injury: Neuropsychological Assessment**). While general information is applicable to examinees of all ages, most of the tests that are discussed here were developed for use with adults.

Clinical Assessment and Formal Testing

Evaluation or assessment is used in all clinical disciplines and may involve elaborate, sophisticated technology, such as magnetic resonance imagery (MRI) to look for the presence of a brain lesion, or may depend on more informal, intuitive ways of combining personal observations to form hypotheses or conclusions, such as surmising a person acts depressed. Different clinicians within mental health "work up" a client or patient using common as well as unique tools and methods associated with the specialized training and expertise of the particular discipline. For example, almost all mental health clinicians assessing someone not only interview the person by asking common questions about mood, sleep, appetite, and daily functioning but also probe other areas more selectively.

Psychiatrists are more likely to consider medical or biological factors potentially contributing to emotional problems, whereas social workers may spend

more time exploring the person's interchanges with social systems. Family therapists may look more acutely at the history of family relationships and alliances, and psychologists are more likely to use inventories and psychological testing. Each discipline's approach is a blend of commonly shared tools, literature, and research arising from several disciplines as well as unique resources based on clinical specialization and scope of practice for that particular discipline.

Differences between Psychiatric and Psychological Testing

Clinical psychologists and psychiatrists both provide assessment and treatment services in the field of mental health, but the differing educational and clinical training routes shape distinct elements that each group uses for these services (*see* **Behavioral Science Evidence**). Psychiatrists, like other physicians, attend medical school for four years during which time they focus on studying chemistry, anatomy, and physiology as well as work in a wide variety of medical speciality placements before getting the MD (medical doctor) degree. They also may take courses in research, statistics, and learning theories. During psychiatric rotations in medical school and later in residency specialization (3 or 4 years), they learn about medical tests and clinical diagnostic techniques for mental health problems, study personality theories and human development, and acquire skills for treating mental illness.

For most psychiatrists, treatment is heavily based on biological and chemical factors that medicines affect (*see* **Psychopharmacology**; **Psychopharmacology: Child and Adolescent**). Assessment is therefore oriented toward finding biological factors or markers for which medical or somatic intervention is well matched. Psychiatric assessment relies heavily on clinical observation (something physicians are well trained to do) and the use of mental status evaluation (*see* **Mental Status: Examination**). A mental status evaluation involves assessing a person's orientation (knowledge of place, time, and self); ability to attend to environment and to communicate; mood and expression of feelings; and congruence or incongruence within the situation (appearing to hallucinate or act delirious). Depending on the nature of the hypothesized mental health problem, the psychiatrist may use brief tests or inventories to support or exclude certain diagnoses. These instruments may include behavioral inventories designed to collect information about classroom behavior of children; self-reported symptoms on a depression checklist; or a brief cognitive status exam looking for signs of dementia. More elaborate psychiatric testing is usually medically or biologically oriented as the psychiatrist reaches into his/her medical training to refine diagnosis (e.g., radiological tests looking for brain tumor; electroencephalogram (EEG) to document seizures or to find unusual electrical patterns in the brain; and blood tests to measure hormone levels or to check for problems such as anemia).

Psychologists, on the other hand, begin their training by attending a university graduate school for four or more years taking courses in personality theory and human development, measurement theory and test development, research statistics, learning theories, methods of mental health interventions, and principles of social contexts and human interactions. To a lesser degree, course work includes anatomy and physiology of the nervous system, human biochemistry, and genetics. Students in a psychology Ph.D. (doctor of philosophy) program complete both a masters level research project and a doctoral dissertation.

Clinical training involves multiple placements in clinical sites during graduate school followed by a 12-month internship in a mental health setting prior to graduation. Some states require an additional year of postgraduate supervised clinical work before licensure, and several areas of professional specialization offer residency training for one or two years. Some areas of specialization are clinical psychology (dealing with application of psychological knowledge and principles to address mental health problems), neuropsychology (study of brain–behavior relationships), experimental psychology (basic physiological or social factor research), and forensic psychology (application of psychological knowledge and principles to address legal issues). Each of these areas makes heavy use of assessment and testing.

When psychologists conduct an evaluation, they also rely on the common tools of observation and interview with the person and collateral sources (family members, teachers, institutional staff who have information about the person's behavior). As do psychiatrists, they conduct mental status evaluations and may also use brief inventories, checklists, or simple tests to support or dismiss diagnostic hunches. However, when psychologists proceed to more elaborate

testing, they rely most heavily upon the tools their discipline has developed. For the most part, these tests involve either measurement of cognitive abilities (e.g., intelligence, memory, and academic skills) or personality factors (e.g., traits and emotional states).

Cognitive Testing. Cognitive testing may be thought of as a way to measure how well the brain is functioning. Neuropsychology, a specialty with its own postgraduate training and certification, has developed a vast array of tests that measure not only basic intelligence but also learning and memory, sensory perception and sensory-muscle integration, reasoning and problem solving skills, language and communication abilities, and basic academic skills such as reading (*see* **Neuropsychological Assessment**; **Neuropsychological Assessment: Child**). Cognitive testing is able to identify and to document the level of skill or the degree of impaired functioning of a person. It is often used in conjunction with neurological or radiological testing that documents impaired structure of the brain or nervous system. Findings from cognitive testing may be used for such tasks as assisting with decisions about academic placement or need for special education resources (*see* **Mental Retardation**); rehabilitation treatment planning after head injury or neurosurgery (*see* **Head Injury: Neuropsychological Assessment**); diagnosis of conditions having subtle onset such as early stage dementia; measurement of progress or lack of response to remediation efforts after head injury or stroke; or documentation of progression of effects of diseases such as Parkinson's or Multiple Sclerosis. Neuropsychologists are often called upon in forensic cases involving head injury, toxic exposure, or questions of competency or capacity (*see* **Capacity to Stand Trial**; **Capacity Assessment**; **Guardianships of Adults**).

Personality Testing. If cognitive testing is thought of as a way to measure the brain's work, personality testing may be thought of as a way to assess the mind of the person. Whereas one deals with neurophysiology and often uses physical measures, the other deals with psyche and uses measures of emotion, attitude, and traits. Personality testing explores intrapsychic (internal aspects of self, value conflicts and ambivalence, moods, motivation) and interpersonal issues (social behavioral clusters, styles of interaction, orientation toward others).

Personality testing is often used to gather information to allow psychologists to describe what a person is like; how a person is different from others and to what degree; how the person functions or is likely to function with others; whether there is significant psychopathology; extent to which a person is open and transparent in self-presentation or guarded or even deceptive (*see* **Deception: Detection of and Brain Imaging**; **Deception: Detection of**); **Malingering: Forensic Evaluations**) and prognosis for improvement with treatment for mental health problems. Psychologists are often called upon in forensic cases to use personality testing to address questions involving risk assessment (*see* **Dangerousness: Risk of**); mental illness diagnosis and treatment recommendations; competency and capacity; tort cases where emotional distress claims are made (*see* **Posttraumatic Stress Disorder**); and criminal cases where mental illness factors are being presented (*see* for example **Insanity: Defense**; **Temporary Insanity**).

Properties of Formal Testing

Testing, which is a relatively circumscribed activity involving use of tests to obtain specific scores, is one part of an overall assessment that collects information from multiple sources using multiple methods. Information usually includes records that place a person's current performance in historical context, referral information that places a person in a situational context, behavioral observations, interviews, and consideration of the person's test taking attitudinal factors (fatigue, rapport, motivation, exaggeration, or defensiveness). The person–context information can be as important as the test result information, for instance when the person is observed to be disengaged from the task or hostile toward the examiner. As a result of such situational data, the psychologist interprets data cautiously or may even disregard the scores as being invalid measures. This person–situation information, whether obtained from direct observation or from validity scales embedded within tests, helps determine if test results are considered valid for interpretation.

Together with information from these other sources, test results are then used to assist in diagnosis, to identify areas for intervention, or to provide information about current functioning.

These results are integrated into a cohesive and comprehensive understanding of the person to communicate it to others including the person being tested. Indeed, providing feedback to the person often serves as an intervention itself and can be a very important part of treatment.

Formal testing refers to an assessment process using standardized instruments that have rules for administration and for scoring responses, normative information for comparing scores, and guidelines for interpretation of results. Holding these test factors constant means there should be minimal "noise" produced so that it may be logically assumed that most of the variability of scores between examinees is a result of something associated with differences between the individual test takers. Just as the use of an accurately calibrated thermometer allows a physician to verify or rule out the presence of fever to help refine diagnostic hunches, the use of psychological testing allows psychologist to verify or rule out abnormalities and to refine diagnostic hunches on the basis of test score patterns and known psychological conditions.

Reliability

Reliability is the term used to indicate the concept of score consistency on a test. It is a measure of how closely clustered or how far apart a person's scores would be with repeated administrations of the same test or with alternate versions of the test. Reliability is a measure of how much error (or "noise" as compared to "signal") is present in test scores [2, 3].

Validity

Validity is the term used to designate the concept of score accuracy of a test. It tells how well a test works in measuring what it purports to assess. Validity may be determined by comparing a test with a "gold standard" in the field, by seeing how well test scores discriminate between known groups on a particular factor, or by how well a test covers relevant aspects of a construct but avoids tapping other constructs. Just as a rifle must shoot true before a marksman can attain target accuracy, a test must have reliability (consistency of measured scores) before its validity (accuracy of measured scores) can be established [2, 3].

Standardization

Standardization refers to the process whereby a test and its components are made standard or routine in details of administration. Test items, the order of administration of items, ways of recording and scoring responses, and ways of instructing examinees about how to take the test must become standard or fixed. Standardization and adherence to instructions regarding the test should decrease the "noise" associated with a person's scores so that differences between examinees (or the same person tested at different times) should be related to actual differences rather than to errors involved in the process of obtaining the scores [2, 3].

Norms

Norms refer to the collection of scores that represent a sample group's responses on a test. Norms can be obtained when many people take the test under the same standard conditions. The mean (average) and standard deviation (SD) (measure of variability of scores) can be calculated and used to evaluate a particular person's obtained score in relation to the cluster and range of scores produced by the norm group. Raw scores are often transformed to standard scores (mean $= 100$, SD $= 15$), T-scores (mean $=$ 50, SD $= 10$), or scale scores (mean $= 10$, SD $= 3$) to create a shorthand way of conveying a person's score in relation to the sample group's norms. A well-developed test has a heterogeneous (mixed gender, ethnicity, and socioeconomic status) group of individuals whose scores are used to calculate the norms. As a result, the sample group norms should be more representative of the entire population of interest than if only a homogenous group was tested and scores calculated.

By having good normative data, it is possible to determine to what degree the examinee is similar to or different from the average person taking the test. By having test data from differing groups of people with a known condition or factor, it is possible to use a person's test results to refine diagnostic hypotheses about the examinee. By using information about a person's score in relation to the norms, the psychologist can then analyze and interpret test score data. Analysis of test data can occur at three levels: specific details about skills, deficits, or symptoms based on actual scores; level of performance or

information about severity of problem or level of skill based on relationship of score to group norms; and syndrome or pattern of scores, which generalizes to life context, based on research literature and clinical knowledge [3, 4].

Testing offers advantages over interview or observation alone by providing empirically quantified information that is usually more precise than interview impressions. Testing may cover a variety of constructs, and, in the case of personality testing, it covers a large array of traits in an efficient manner. Standardized administration and scoring means a common yardstick is used for measuring a person, and behavior is observed in a uniform context. Test norms allow scores to be compared to known groups, and research with such known groups provides the psychologist with a stimulating backdrop from which to generate and evaluate hypotheses related to test scores [5].

History of Psychological Testing

A brief history of psychological testing can provide a useful context for understanding its current state. A very influential English figure was Sir Francis Galton who in 1869 published "Classification of Men According to Their Natural Gifts" [6]. Galton pioneered use of questionnaires and rating scales, developed statistical methods for analysis of individual differences, and popularized forms of psychological testing. At the International Health Exhibition in London (1884), he and his assistants tested individuals willing to pay to learn about their vision and hearing sensitivity, muscle strength, reaction time, and memory. This focus of attention on individual abilities was in contrast to the emphasis on group data being collected by others. Wilhelm Wundt, a German psychologist who had established the first psychology lab in Leipzig five years before, considered individual differences to be errors and was far more interested in measures of perception that portrayed information about the species as a whole [7].

In the same tradition as Galton, James M. Cattell established a testing lab at the University of Pennsylvania (1888) and used tests to measure sensitivity to pain, perceptual discrimination, memory, and other cognitive skills as ways to study individual differences in intelligence. Cattell was the first to use the term *mental test* to describe his techniques [7]. In general, testing in clinical psychology testing has continued to focus more on individual distinctives, whereas measurement in physiological or experimental psychology has focused more on data that add to knowledge about collective or group characteristics of humans and other animals.

Psychological testing took a major leap forward in 1905 when Alfred Binet, a French psychologist, and Theodore Simon, a psychiatrist, collected 30 brief cognitive tests to use as an intelligence scale for the purpose of determining school entry for Parisian children. They later revised their collection of tests and developed the concept of mental age (the age at which an average child could pass a particular test). In the United States, Lewis Terman added some other tasks and developed the concept of intelligence quotient (IQ) that was found by dividing the attained mental age by the person's chronological age. Eventually, IQ came to be calculated in a different manner on the basis of comparison to group information and how far from the group mean a person's scores were [8].

Another major advancement in psychological testing occurred in 1917 when the US Army had to classify World War I (WWI) recruits for their suitability for military service as well as for possible officer training. In a period of two years, 1.7 million inductees were given group administered intelligence tests based on alterations of the Binet and Terman tests. For those literate in English, the Alpha version (verbal tests) was used; for illiterate or non-English speaking recruits, the Beta version (nonverbal tests) was used. Nonverbal or language free tests had been used in the United States at Ellis Island for screening immigrants for mental defects and had also been used in Chicago with unschooled juvenile delinquents.

David Wechsler, a psychology graduate student, enlisted in the Army and was trained to conduct individual intelligence tests for the 5% of recruits who failed one of the group-administered versions. Many years later as psychologist at Bellevue Psychiatric Hospital in New York City, he developed a different approach to intelligence testing that combined verbal and nonverbal tasks to result in a single score. He also collected normative data on a large group from the general population and compared each person's score to the group average (mean). This deviation derived IQ score (relationship of person's score to group average) provided information about how far a person's score was above or below the mean score for

the group. In 1939, he contracted with Psychological Corporation for production and sale of his test [8].

Psychological Corporation, formed in 1921, had already been marketing intelligence tests (termed *scholastic aptitude tests*) and other specialized aptitude tests to schools and industries after the war, but the number of psychological tests had been rather limited until the mid-1930s.

Oscar Buros' first edition of *Educational, Psychological, and Personality Tests of 1933 & 1934* had been only 44 pages, but his 1938 edition of what became *Mental Measurements Yearbook* was more than 400 pages long and covered 4000 tests. The current edition is multivolume and covers thousands of tests [9]. The entry of the US military into World War II brought with it the need for testing four million recruits for military assignments. Once again, a national crisis produced a surge of testing emphasis which propelled psychological testing from academic or clinical environments into the larger culture. A similar peacetime surge occurred in the Cold War with a strong focus on achievement and aptitude testing when the space race was initiated with the launch of the Russian satellite Sputnik. These intelligence, aptitude, and achievement tests fall into the general category of cognitive testing [10].

Objective Tests

In a different line of development, another contribution out of the WWI era was the self-report inventory that formed the basis for objective personality testing. Robert Woodworth's personal data sheet was developed too late for military use in the war but was used afterwards in civilian life to screen for seriously disturbed individuals. The inventory was basically a self-report of recognized symptoms associated with psychiatric problems. This approach culminated in the 1943 publication of the Minnesota Multiphasic Personality Inventory (MMPI). Starke Hathaway, a psychologist, and Charnley McKinley, a psychiatrist, had developed the set of test items empirically by selecting those items that successfully differentiated known patient groups from a general population sample (both genders but all Caucasian samples comprised mostly of hospital visitors, some airline workers, and civilian conservation corps workers in Minnesota). The inventory was remarkable for this use of empirical research that served as a basis for item selection and for its use of three validity scales designed to measure test taking attitude (defensiveness, exaggeration of symptoms, and subtle defensiveness) [7, 11].

From this line of development has come a wide range of objective tests that provide information about personality traits (e.g., introversion, dominance, and self-constraint); emotional states (anger, anxiety, and depression); behaviors (substance abuse, risk-taking, and antisocial acts); and interpersonal orientation (psychopathy, altruism, and cooperativeness). Often these tests are pencil and paper or computer administered self-report measures that have validity scales to assess manner of self-presentation. As the tests are usually in written form, it is important to assess the examinee's reading level to assure it is sufficient for the task. The results of the tests are thought to be explicit self-representations of the test taker. Test items are seen as stimuli, and self-report statements of endorsement or denial of particular items are seen as responses that define and describe the person. Some tests are also designed to be completed by collateral sources who know the person, and this information can be quite useful in assessment of children or when the examiner wants to know how others perceive the person [3].

Projective Tests

A different approach to personality assessment is the use of projective tests, best typified by the inkblots used by Hermann Rorschach, a Swiss psychiatrist. Others had previously used vague or ambiguous stimuli to elicit a person's responses, which then were interpreted on the basis of the assumption that the subject's internal struggles, fantasies, and needs were being projected onto this ambiguous but neutral stimulus material. Rorschach began experimenting with inkblots around 1910 and established a system for eliciting and scoring responses, which he published along with 10 inkblots in *Psychodiagnostik* (1921).

This line of assessment was further elaborated by the work of Henry Murray in the development of the Thematic Apperception Test (TAT) (picture cards used to elicit short stories from the subject), Goodenough's draw a person test (originally used as an intelligence measure), and various sentence completion or word association tests. Often projective tests are administered directly by the examiner who records responses of the examinee, but some projective tests are also self-administered

(house–tree–person drawings and sentence completion tasks). Regardless of the manner of administration, the stimuli are used to evoke behavior that is thought to contain implicit information that must be discovered and interpreted by the trained examiner [12].

Personality Testing

The study of the nature of personality has a long lineage dating back at least to early Greek philosophers and playwrights. In the mid-nineteenth and early twentieth centuries, scientists such as Galton, Wundt, and James moved psychology away from being a philosophical exercise to the study of actual human behavior, many times trading erudite speculation for humble inquiry. Much of the physiological and cognitive research work of the European psychologists was laboratory based study of normal behavior but lacked a theory of person. In contrast was the work of Freud, a psychiatrist who was a seminal figure in the history of personality theory who worked with disturbed patients and searched for a model of the psyche combining understanding of normal and abnormal personality that would fit neurological evolution. His treatment approach and psychoanalytic theory came from his clinical model of inquiry and observation, and his major contributions were related to abnormal personality.

Issues in Personality Test Theory

There remains a basic division in personality research regarding whether personality is known by studying normal or abnormal behavior. In the United States, Gordon Allport studied the individual and the unique combination of the individual's normal traits. Henry Murray (who developed the TAT) studied individual differences seen in normal drives or traits and how they were integrated in the person. Both assumed personality is defined by traits that are independent of psychopathology, but both taught that the individual cannot be understood by fragmenting these traits. Eysenck, who proposed a two factor theory of neuroticism/emotional stability and extroversion/introversion, and Cattell, who developed the 16 Personality Factor (16 PF) test, carried on this line of work that now has given rise to the five factor model of Costa and McCrae [13]. This model is the basis for the NEO-Personality Inventory (NEO-PI) [14]. In contrast, developers of inventories such as the MMPI [11] or the Personality Assessment Inventory (PAI) [15] have paid more attention to clinical groups and abnormal personality.

Those who attempt to reconcile normal and abnormal personality research must address the question articulated by Paul Meehl as to whether personality traits (enduring characteristics in contrast to temporary states) occur in a dimensional spectrum that ranges from normal to abnormal extremes or occur in categorical divisions (taxons or uniquely different groups). The thinking behind the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) approach to diagnosis assumes distinct categorization. Some trait theorists such as Millon posit a dimensional approach where either extreme of a trait may be pathological [16, 17].

Those who study personality are also divided as to whether traits actually exist independent of models of thought (constructive-realist) and are largely determined by physiology or whether traits are only psychological explanations of behavior, actually determined by social domains affecting expression of biological needs (socioanalytic). The debate comes out in discussion about nature versus nurture causing behavior but goes beyond to attempt to answer whether there are really inherent basic trait domains (such as hardwired extroversion/introversion or even psychopathy) or only differences evident in the way people respond to or cope with biological needs based on social factors [18].

Atheoretical Approaches to Personality Testing

In addition to the trait research, other American work has significantly influenced personality testing from an atheoretical model. Coming out of the medical exam/psychiatric interview methodology, Hathaway (psychologist) and McKinley (psychiatrist) in the 1930s and early 1940s developed the MMPI in an attempt to find an empirically derived test that would differentiate medical patients with psychopathology from normal individuals. Even though self-report methodology, which had been developed by Woodworth in World War I with the Personal Data Sheet, had come to be criticized as too easily manipulated, Hathaway and McKinley used the technique anyway. They collected a large pool of items that were associated with psychiatric and medical symptoms

and asked individuals to respond to each by endorsing or denying its occurrence in their own lives. The MMPI authors, aware that responses might be affected by efforts to distort self-presentation, created validity scales to be used in interpretation of scores. They disregarded theoretical constructs and attempted to use only empirical means to identify those particular items that produced scores able to distinguish between specific diagnostic groups of patients and nonpatients. The original hope and early emphasis with the work at the University of Minnesota and later at Menniger Clinic in Kansas was to develop a tool capable of providing psychiatric diagnosis using a practical, empirically derived approach [19, 20].

The work on the MMPI was part of a post-World War II interest in testing. Similar to the demands of WWI, the military during wartime had responded to the task of classifying millions of recruits. After the war, in the United States there was a large number of nonmedically trained mental health providers, mainly clinical psychologists, who were equipped and ready to apply science, which had been optimistically embraced by the nation, to the study and treatment of mental health problems. Testing was valued in clinical, educational, industrial, and correctional settings and was seen as a way to bring efficiency, empiricism, and objectivity to tasks in these fields.

During the rise and reign of behaviorism (1960s and early 1970s), there was strong challenge to the concept of "personality", something that was intangible and considered an unnecessary and useless myth in explaining behavior. There was no need for the concept of "self" as there was exclusive focus on what an organism did in response to particular stimuli. Testing did not go away, but there was a much stronger emphasis on behavioral assessment – identifying particular behaviors of interest and recording the circumstances and frequency of their occurrence. In the past few decades, there has been modification of this model that now includes cognitive as well as behavioral aspects of assessment and treatment [21]. With this cognitive-behavioral approach there is focus on specific assessment using specialized instruments (Beck Depression Inventory and Beck Anxiety Inventory) that make no effort to describe the person as an integrated personality or even to provide a comprehensive clinical picture. Some have suggested that these brief rating scales and inventories, which are limited in scope, represent a third type of personality assessment: rapid personality assessment instruments [3].

Currently, in many settings outside the teaching or research environments, comprehensive personality assessment is often not seen as important or is not viewed as essential even if important. In addition, managed care of mental health services has often taken a restrictive stance toward more comprehensive testing. A survey of psychologists assumed to be in private practice found that 25% did not do personality testing at all. Of the ones who did such testing at least sometimes, about half used one or both of the two widely recognized tests (MMPI and Rorschach Inkblots) although the most utilized personality test reported was sentence completion [21].

Current Psychological Test Usage

Other surveys of psychologists have likewise documented a decided trend toward less testing in recent years. Psychologists in 1959 said 44% of their time was spent in assessment, but that figure was down to 22% in 1982. In 1971, 5 of the top 10 most often used psychological tests were projective personality tests; 1 was an objective personality test (MMPI), and 3 others were IQ tests. The core of the most popular tests remained stable through the early 1990s, but the amount of time devoted to assessment declined markedly. By the end of the decade, a study on test usage conducted for the American Psychological Association (APA) found fewer than 20% of clinical psychologists spent as much as 5 hours a week doing testing although about three-quarters of neuropsychologists spent at least that much time testing. Of those clinical psychologists with at least 5 hours a week devoted to testing, approximately 40% of their service was devoted to IQ or achievement testing and one-third to personality testing.

Neuropsychological testing was the next highest category of time spent (about 20%) but was (as expected) the major use of time for neuropsychologists who also spent about 20% of their time doing personality assessment. The MMPI/MMPI-2 was the most widely used test in the combined two groups with an IQ test (Wechsler Adult Intelligence Scale-Revised (WAIS-R)) placing second overall. Clinical psychologists also tended to use the Rorschach Inkblots (ranked fourth), but neuropsychologists did not (ranked eighteenth). No other personality tests

were popular with neuropsychologists, but at least half of clinical psychologists in the survey used the TAT (ranked sixth in frequency of use) [22].

Psychological Testing in Forensic Context

In addition to neuropsychology, another area of psychology that relies heavily on assessment and testing is forensic psychology. A survey of state directors of mental health resulted in information for 41 state corrections departments. Of those responding, 40 used testing at intake while 25 used it also for pre-parole work-ups. Of those testing at intake, 26 (65%) used MMPI/MMPI-2, 18 (45%) used intelligence tests, and five (12.5%) used Rorschach Inkblots in their department protocols. Of the 25 using testing for preparole evaluations, 24 (96%) used MMPI/MMPI-2, eight (32%) used intelligence tests, and nine (36%) used Rorschach Inkblots [23].

Results of a survey of 152 forensic psychologists who were members of American Psychology-Law Society (AP-LS) or diplomates of the American Board of Forensic Psychology (ABFP) indicated they averaged 56% of their time doing forensic work, presumably largely assessment, with almost a third (29%) being spent in actual testing. When asked about most frequently used tests in particular types of adult assessments, the top five in order of weighted frequency of usage across referral questions were: MMPI-2, one of the Wechsler intelligence or memory, one of the Hare Psychopathy Checklist versions (*see* **Psychopathy Checklists**) Structured Interview of Reported Symptoms (SIRS), and PAI [24].

Another survey of 64 diplomates of ABFP asked which tests were considered acceptable for use in six areas of forensic assessment. Across areas, both the MMPI-2 and the Wechsler Adult Intelligence Scale, Third Edition (WAIS III) showed strong acceptance as did the PAI to a lesser but still major degree. The Millon Clinical Multiaxial Inventory, Third Edition (MCMI-III) was rated as acceptable in only one area (mental status at time of offense). The Rorschach *inkblots* was rated as unacceptable by the majority (52–60%) of the respondents in five of the six areas (equivocal for use in exploration of mental status at time of offense) while other projective tests were deemed unacceptable in all areas by a significant majority (60–95%). Two neuropsychological batteries (Halstead-Reitan and Luria-Nebraska) were reported to be acceptable in half of the areas as was another IQ test (Standford-Binet) [25].

Forensic Assessment: Psychology and Law

Historical Developments

Hugo Munsterberg, considered the founder of applied psychology, was the first director of the psychology lab at Harvard after leaving the University of Leipzig where he had been a student of Wilhelm Wundt. He was an early pioneer of the study of eyewitness testimony and a strong advocate of application of psychology to law. In his book *On the Witness Stand* (1908), he chastised attorneys and judges for not embracing the research findings of psychology and using them in the courtroom. He was strongly criticized and his ideas mocked, but in the 1920s law schools began hiring psychologists to teach courses. Psychology was involved in courts largely in relation to treatment and disposition of children, and psychiatrists were more involved with competency and sanity questions. But psychologists and other social scientists were involved in Brandeis briefs, the most famous being the work of Kenneth Clark and colleagues in *Brown v. Board of Education* (1954) [26]. Only after *Jenkins v. United States* (1962) [27] was the way open for psychologists to provide expert testimony on mental health issues.

Other more recent legal rulings that have significantly impacted the practice of forensic psychology have been *Daubert v. Merrell Dow Pharmaceuticals, Inc*. (1993) [28], which replaced *Frye* (1923) criteria and defined the standards for consideration of scientific expert testimony; *General Electric v. Joiner* (1997) [29], which established the authority for the trial judge to determine what proffered testimony met *Daubert* standards; and *Kumho Tire Co., LTD v. Carmichael* (1999) [30], which expanded *Daubert* criteria to other fields of technical and specialized knowledge. Although these rulings only apply to federal courts, states have their own case law decisions in these areas [31].

Prevalence of Forensic Psychological Assessment

No data were found that indicate the current extent of psychological assessment utilization by US courts, but a conservative estimate of Competency to Stand

Trial evaluations in 1993 was 50 000 [32]. A more recent work cited research indicating 4–7% of criminal cases involved referral for these competency evaluations, the most prevalent mental health evaluation for criminal courts [33]. Civil courts often refer custody case participants for psychological evaluations, and personal injury cases frequently involve mental health expert testimony. Although the data are not available, it is obvious that a significant number of legal cases involve psychological or other mental health assessment. This reliance on psychology appears to be growing as evident with the publication of this volume that verifies Munsterberg's foresight.

The practice of forensic psychology, which is heavily dependent on psychological research and assessment tools, is one of the speciality areas in which the American Board of Professional Psychology (ABPP) offers board certification or diplomate status. The APA Division 41 (AP-LS) is devoted to exploration of the interface of psychology and law. Several well respected, refereed journals (including *Law and Human Behavior; Behavioral Sciences and the Law; Psychology, Public Policy, and Law*) are devoted exclusively to publication of research in this area of disciplinary overlap.

Tensions between Clinical and Forensic Assessment Perspectives

Forensic assessment in some important ways differs from traditional clinical assessment as discussed by Melton *et al.* [34], Heilbrun [35], and Archer *et al.* [31]. Some of the major ways include purpose, scope, and understanding of who is being served (broad understanding of purpose and scope in mental health evaluation where examinee is also the client being served versus forensic assessment addressing specific legal or quasilegal question(s) regarding the examinee to assist the decision maker who is considered the client). In addition, forensic examinees are frequently mandated for an evaluation and often assumed to have significant reasons to be purposefully selective in self-disclosure so that a much stronger focus must be placed on examiner objectivity and assessment of examinee's response style (pattern of self-presentation). Because of the threats of conscious deception or selective self-presentation in forensic evaluations, there is more emphasis on use of multiple sources of data to create hypotheses or to verify information as well as strong reliance on external sources (collateral observations, historical records, and reports of others) apart from the formal assessment interactions with the examinee.

While there are some instruments specifically developed for forensic use, these tend to be structured interviews, rating scales, or tests designed for use with a particular legal application in mind (e.g., Competence Assessment Instrument for Standing Trial (CAI), Psychopathy Checklist-Revised (PCL-R), and Competence Assessment for Standing Trial for Defendants with Mental Retardation (CAST/MR)). Quite frequently other instruments, developed for nonforensic purposes, are used in a forensic assessment because of the vast research on the instruments, validity indicators built into some of the instruments, or ability for these tests to contribute to a broad understanding of the person to develop hypotheses related to factors bearing on the legal question(s). Obvious examples would include well researched personality tests, tests of malingering, and cognitive tests including IQ measures. When any test is considered for forensic evaluation, these factors are important to consider: sufficient research and norms with a population similar to that of the examinee, adequate test development and psychometric properties, and ability to link test results to conclusions regarding the referral question [35].

A major risk inherent with use of clinical tests is overinterpretation of results where adequate validity research (relating test scores to real world conditions or outcomes) does not demonstrate clear connections between the test data and specific legal question(s) [34]. This problem was pointed out in a marked way by Jay Ziskin and David Faust, two psychologists, in the 1970s and 1980s (*Coping with Psychiatric and Psychological Testimony; The Limits of Scientific Reasoning*). They criticized mental health providers for going beyond the data and research by offering forensic opinions based on personal impressions rather than scientific conclusions. As a more recent writer noted, the task should be to provide "the best that psychology has to offer but also to be candid about the limits of our science and our expertise" (p. 131) [36].

Tensions between Psychological and Legal Perspectives

Even though psychologists are often used to conduct assessments and to provide relevant information

to the courts about psychological factors having a bearing on legal issues, there remains basic tension between psychology (which is experimental, descriptive, and probabilistic in its orientation to pursue scientific truth) and law (which is adversarial, prescriptive, and decisive in its orientation to pursue social and individual justice) [34, 37]. Psychology operates from a perspective of determinism (i.e., behavior can be predicted by some combination of genetic, biological, social, interpersonal, and environmental factors) whereas law operates on an assumption of free will (i.e., behavior is the result of personal choice and responsibility as each person is a free moral agent).

Other differences are evident as law applies broad terminology to specific cases, and psychology seeks to define and operationalize terminology into quantifiable form. Questions arise about legal terms for which there are no operational definitions either in law or psychology (e.g., reasonable degree of certainty pertaining to professional opinions, best interest of the child, reasonable degree of rational understanding for competency, unable to appreciate the nature and quality or wrongfulness of acts for criminal responsibility) or where the meaning of words or use of concepts differ (insanity as a legal construct and psychosis as a mental health concept). Despite the sometimes ill-fit of the two systems of thought, courts continue to rely on psychologists for help; and psychologists continue to explore better ways to respond to the task by developing methods to tie psychological data to legal constructs in a rational manner that can be logically followed [38].

Brief Overview of Major Personality Tests

Minnesota Multiphasic Personality Inventory (MMPI and MMPI-2)

The original 550 items and norms were developed by Hathaway and McKinley in the 1930s, mostly with patients and visitors at the University of Minnesota Hospital. The instrument was developed to aid in psychiatric and medical screening [7]. A revision (MMPI-2) was published in 1989 with updated items (66 modified, 90 omitted, and 107 newly created) and contemporary norms (2600 subjects of which 5% were current psychiatric patients). Individuals in the new norm group were from seven states representing a geographic diversity and were roughly balanced for gender (56% female).

Although ethnic and socioeconomic diversity was sought, Hispanic and Asian American groups are underrepresented in the norms. The normative sample is also skewed with college graduates comprising a larger percentage than is present in the US population [20]. The current 567 item true/false inventory requires an eighth grade reading level according to the 1989 version of the manual or a sixth grade reading level according to the revised edition of the test manual published in 2001 [11]. In addition to the original three validity scales (L, F, K), there are now several others measuring tendency to acquiesce, marking responses randomly, or subtly presenting self in a consciously distorted manner. Although well respected and popularly used, the inventory is criticized for its scales being too highly intercorrelated and therefore not distinct because many items are used simultaneously for scoring on different scales.

The MMPI was originally constructed for diagnostic classification purposes, but it proved to be disappointing in that regard when used in replication studies. Meehl was influential in adapting interpretation to a larger context by focusing on patterns of scores (profiles) and code types (based on scores of highest scales in a person's profile). These profiles and code types were researched to provide psychiatric descriptions for the major groups. While code type or group profile descriptions came to be widely used, research revealed problems with temporal instability of the code types when individuals were retested [19].

Another way of using test results from the MMPI has grown through the development and use of content scales, which are comprised of items that seem to be consistently related to a similar construct. This approach led to content scales providing descriptive rather than diagnostic information about a person and has more recently led to innovations such as Restructured Clinical (RC) scales (basic clinical scales with a general demoralization factor removed) and Restructured Form (MMPI-2-RF) with RC scales and newly developed specific problem (content) scales in a shorter version. These innovations also removed the use of K-corrections in calculating scores and removed item overlap within the basic scales [19].

Personality Assessment Inventory (PAI)

The PAI, a self-report inventory comprised of 344 items to which a person marks one of four choices

about how well each item applies to self, produces scores on 22 scales and 31 conceptually derived subscales. Scales were devised based on content and internal consistency and no item is scored on more than one major scale. Raw scores are transformed to T-scores using norms from a group of 1000 community dwelling subjects matching 1995 US census projections. Comparison clinical norms are also available based on a group of 1246 patients. Construction of the PAI was based on the assumption that normal personality constructs are distributed for both patients and nonpatients according to a bell-shaped (normal) curve. However, abnormal symptoms and constructs differ markedly within the two populations and occur very infrequently in the general population. The PAI requires a fourth grade reading level. It has four validity scales measuring positive and negative self-presentation, random responding, and overendorsement of unusual items [15, 39].

NEO-Personality Inventory-Revised (NEO-PI-R)

A third objective personality measure is the NEO-Personality Inventory-Revised (NEO-PI-R) that measures five basic stable dimensions of personality based on the work of Tupes and Christal who analyzed data that had been collected by Cattell in his work on normal personality traits. These personality domains are neuroticism, extroversion, openness to ideas, agreeableness, and conscientiousness. These five scales each have six subscales that have been developed by logical and factor analytic strategies utilizing known research on normal personality traits. This inventory is comprised of 240 items marked with one of five possible responses (strongly disagree to strongly agree). It contains no validity scales. Norms were developed using 500 males and 500 females selected to match 1995 US census projections. Forms are available for both self-rating and ratings by collateral persons who know the examinee [13, 14].

Rorschach Inkblots

This projective test, a standard set of 10 inkblots, was published by Swiss psychiatrist Hermann Rorschach in 1921 and is based on the assumption that personality can best be assessed through analysis of responses to ambiguous stimuli. Such implicit knowledge is less confounded by conscious distortion than is explicit information obtained by self-report. The assumption is that self-report only describes behavior, but this projective task actually produces behavior that can be observed, recorded, scored, and analyzed. Over the next five decades after Rorschach's publication, several different systems for scoring and interpretation emerged in both Europe and the United States. John Exner, influenced by Meehl's writings about actuarial rather than intuitive approaches to test interpretation, set out to standardize rules for scoring responses so that empirical data could be developed and used in interpretation. His comprehensive system of coding caught on and has largely shaped the use of this instrument since its introduction in the mid-1970s. Use of Exner's system [40] leads to specific data that can be related to norms and empirical research to provide information about a person's ability to control stress; ways to process information or make sense of the world; patterns of thinking about self and the world; emotional state; and ways of perceiving events and relationships.

Despite criticisms and disfavor of projective tests in general, the Rorschach Inkblots, especially when scored and interpreted using the Exner system, has remained a respected instrument for personality assessment and has passed *Daubert* scrutiny enabling expert testimony based on its findings. It provides limited information about diagnosis with the notable exception of thought disorder but provides a wealth of information to enable an understanding a person as an individual [41].

Other Resources

Clinical and forensic uses of these and other personality tests are discussed at length in Butcher [42, 43], Goldstein [44], Meyer and Deitsch [45], Maruish [46], Strack [47], and Archer [48].

References

[1] Vanden Bos, G.R. (ed) (2007). *APA Dictionary of Psychology*, American Psychological Association, Washington, DC.

[2] American Psychological Association (1999). *Standards for Educational and Psychological Testing*, American Psychological Association, Washington, DC.

[3] Adams, R.L. & Culbertson, J.L. (2000). Personality assessment: Adults and children, in *Kaplan & Sadock's Comprehensive Textbook of Psychiatry*, 4th Edition, B.J. Sadock & V.A. Sadock, eds, Lippincott, Williams, & Wilkins, Philadelphia, pp. 702–722.

[4] King, R.A., Schwab-Stone, M.E., Peterson, B.S. & Theis, A.P. (2000). Psychiatric examination of the infant, child, and adolescent, in *Kaplan & Sadock's Comprehensive Textbook of Psychiatry*, 4th Edition, B.J. Sadock & V.A. Sadock, eds, Lippincott, Williams, & Wilkins, Philadelphia, pp. 2558–2586.

[5] Meyer, G.J., Finn, S.E., Eyde, L.D., Kay, G.G., Kubiszyn, T.W., Moreland, K.L., Eisman, E.J. & Dies, R.R. (1998). *Benefits and Costs of Psychological Assessment in Healthcare Delivery: Report of the Board of Professional Affairs Psychological Assessment Work Group, Part I*, American Psychological Association, Washington, DC.

[6] Sattler, J.M. (2001). *Assessment of Children, Cognitive Applications*, 4th Edition, Author, Jerome M. Sattler, San Diego.

[7] Anastasi, A. (1988). *Psychological Testing*, 6th Edition, Macmillan, New York.

[8] Boake, C. (2002). From the Binet-Simon to the Wechsler-Bellevue: Tracing the history of intelligence testing, *Journal of Clinical and Experimental Neuropsychology* **24**, 383–405.

[9] Geisinger, K.F., Spies, R.A., Carlson, J.F. & Blake, B.S. (eds) (2007). *The Seventeenth Mental Measurements Yearbook*, Buros Institute of Mental Measurements, Lincoln. Also accessible online at http://www.unl.edu/buros/.

[10] Haney, W. (1981). Validity, vaudeville, and values: A short history of social concerns over standardized testing, *The American Psychologist* **36**, 1021–1034.

[11] Butcher, J., Graham, J., Ben-Porath, Y., Tellegen, A., Dahlstrom, G. & Kaemer, B. (2001). *MMPI-2 (Minnesota Multiphasic Personality Inventory-2) Manual for Administration, Scoring, and Interpretation*, (Revised Edition), University Press, Minneapolis.

[12] Carr, A. (1980). Psychological testing of personality, in *Comprehensive Textbook of Psychiatry, III*, H. Kaplan, A. Freedman & B. Sadock, eds, Williams & Wilkins, Baltimore, pp. 940–966.

[13] Costa Jr, P.T. & McCrae, R.R. (2000). Approaches derived from philosophy and psychology, in *Kaplan & Sadock's Comprehensive Textbook of Psychiatry*, 4th Edition, B.J. Sadock & V.A. Sadock, eds, Lippincott, Williams, & Wilkins, Philadelphia, pp. 638–651.

[14] Costa Jr, P.T. & McCrae, R.R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor Inventory (NEO-FFI) Professional Manual*, Psychological Assessment Resources, Odessa.

[15] Morey, L.C. (1991). *The Personality Assessment Inventory Professional Manual*, Psychological Assessment Resources, Odessa.

[16] Strack, S. (2006). Introduction, in *Differentiating Normal and Abnormal Personality*, S. Strack, ed, Springer, New York, pp. xvii–xxviii.

[17] Halsam, N. & Williams, B. (2006). Taxometrics, in *Differentiating Normal and Abnormal Personality*, S. Strack, ed, Springer, New York, pp. 283–310.

[18] Harkness, A.R. & Hogan, R. (1995). The theory and measurement of traits: Two views, in *Clinical Personality Assessment*, J.N. Butcher, ed, Oxford University Press, New York, pp. 28–41.

[19] Ben-Porath, Y. (2006). Differentiating normal from abnormal personality with the MMPI-2, in *Differentiating Normal and Abnormal Personality*, S. Strack, ed, Springer, New York, pp. 337–382.

[20] Pope, K., Butcher, J. & Seelen, J. (1993). *The MMPI, MMPI-2, and MMPI-A in Court*, American Psychological Association, Washington, DC.

[21] Exner Jr, J.E. (1995). Why use personality tests? A brief historical view, in *Clinical Personality Assessment*, J.N. Butcher, ed, Oxford University Press, New York, pp. 10–18.

[22] Camara, W., Nathan, J. & Puente, A. (2000). Psychological test usage: Implications for professional psychology, *Professional Psychology: Research and Practice* **31**, 141–154.

[23] Gallagher, R.W., Somwaru, D.P. & Ben-Porath, Y. (1999). Current usage of psychological tests in state correctional settings, *Corrections Compendium* **24**, 1–3, 20.

[24] Archer, R.P., Buffington-Vollum, T.K., Stredny, R.V. & Handel, R.W. (2006). A survey of psychological test use patterns among forensic psychologists, *Journal of Personality Assessment* **87**, 84–94.

[25] Lally, S.J. (2003). What tests are acceptable for use in forensic evaluations? A survey of experts, *Professional Psychology: Research and Practice* **34**, 491–498.

[26] Brown v. Board of Education, 347 U.S. 483 (1954).

[27] Jenkins v. United States, 307 F.2d 637 (1962).

[28] Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).

[29] General Electric v. Joiner, 522 U.S. 136 (1997).

[30] Kumho Tire Co. Ltd. v. Carmichael, 526 U.S. 137 (1999).

[31] Archer, R.P., Stredny, R.V. & Zoby, M. (2006). Introduction, in *Forensic Uses of Clinical Assessment Instruments*, R.P. Archer, ed, Lawrence Erlbaum, Mahwah, pp. 1–18.

[32] Skeem, J., Golding, S., Cohn, N. & Berge, G. (1998). Logic and reliability of evaluations of competence to stand trial, *Law and Human Behavior* **22**, 529–547.

[33] Stafford, K. (2003). Assessment of competence to stand trial, in *Handbook of Psychology: Vol. 11, Forensic Psychology*, I. Weiner, (Series Editor) A. Goldstein, (Volume Editor), John Wiley, Hoboken, pp. 359–380.

[34] Melton, G.B., Petrila, J., Poythress, N.G. & Slobogin, C. (2007). *Psychological Evaluations for the Courts: A Handbook for Mental Health Professionals and Lawyers*, 3rd Edition, Guilford, New York.

[35] Heilbrun, K. (2001). *Principles of Forensic Mental Health Assessment*, Kluwer Academic, New York.

[36] Nicholson, R.A. (1999). Forensic assessment, in *Psychology and Law: The State of the Discipline*, R. Roesch, S.D. Hart & J.R.P. Ogloff, eds, Kluwer Academic, New York, pp. 121–173.

[37] Ogloff, J.R.P. & Finkelman, D. (1999). Psychology and law: an overview, in *Psychology and Law: The State of the Discipline*, R. Roesch, S.D. Hart & J.R.P. Ogloff, eds, Kluwer Academic, New York, pp. 1–20.
[38] Grisso, T. (2003). *Evaluating Competencies: Forensic Assessments and Instruments*, 2nd Edition, Kluwer Academic, New York.
[39] Morey, L.C. & Hopwood, C.J. (2006). The Personality Assessment Inventory and the measurement of normal and abnormal personality constructs, in *Differentiating Normal and Abnormal Personality*, S. Strack, ed, Springer, New York, pp. 451–471.
[40] Exner Jr, J.E. (2003). *The Rorschach: A Comprehensive System*, 4th Edition, John Wiley, Hoboken.
[41] Ganellen, R.J. (2006). Rorschach assessment of Normal and Abnormal Personality, in *Differentiating normal and abnormal personality*, S. Strack, ed, Springer, New York, pp. 473–500.
[42] Butcher, J.N. (ed) (1995). *Clinical Personality Assessment*, Oxford University Press, New York.
[43] Butcher, J.N. & Miller, K.B. (2006). Personality assessment in personal injury litigation, in *The Handbook of Forensic Psychology*, 3rd Edition, A.K. Hess & I.B. Weiner, eds, John Wiley, New York, pp. 140–166.
[44] Goldstein, A. (ed) (2007). *Forensic Psychology: Emerging Topics and Expanding Roles*, John Wiley, Hoboken.
[45] Meyer, R.G. & Deitsch, S.E. (1996). *The Clinician's Handbook: Integrated Diagnostics, Assessment, and Intervention in Adult and Adolescent Psychopathology*, 4th Edition, Allyn and Bacon, Boston.
[46] Maruish, M.E. (ed) (2004). *The Use of Psychological Testing for Treatment Planning and Outcomes Assessment*, 3rd Edition, Lawrence Erlbaum, Mahwah.
[47] Strack, S. (ed) (2006). *Differentiating Normal and Abnormal Personality*, Springer, New York.
[48] Archer, R.P. (ed) (2006). *Forensic Uses of Clinical Assessment Instruments*, Lawrence Erlbaum, Mahwah.

PAUL ANDREWS

Psychological Trauma ***see*** Posttraumatic Stress Disorder

Psychopathology ***see*** Compulsion, Firesetting

Psychopathology: Terms and Trends

Introductoin

Society cannot exist without norms, and yet the definition of what is normal proves to be a thorny task for both social science and the field of mental health. In particular, at the interface of behavioral sciences and the law, forensic assessment requires clarity regarding threshold issues in the determination of the presence or absence of pathology and psychical suffering. Whether societal norms are identical to norms of psychical functioning or whether these two norms diverge from one another in specific and definable ways, the line between mental health and mental illness always necessarily implies an appreciation of the individual's ability to function effectively within his environment.

This environment, however, is not a biological one but rather is societal. An individual's ability to function in the social environment is mediated through his subjective interpretation of its laws and imposed constraints, that is, his perception of and ability to integrate norms. A forensic evaluator's role is precisely to introduce and assess this subjective point of view within the parameters of legal proceedings; parameters are designed to apply to all cases equally while providing for individual differences.

Norms and Normality

Psychopathology refers to the description, explanation, and logical formulation of the individual's mental state and mental processes, with the aim of characterizing abnormalities or guiding interventions in the context of treatment. Psychoanalysis has traditionally played an important role in psychopathology, along with cognitive science, neurobiology, and epidemiology (the study of diseases within populations). The overarching aim of psychopathology is the task of demarcating mental health and mental illness, two concepts whose meanings are inextricably bound up with one another, in reference to a norm.

In an article in the September 1967 issue of *Archives of General Psychiatry*, Sabshin outlines four perspectives on normality: (i) normality as health

or adequate functioning; (ii) normality as utopia or optimal functioning; (iii) normality as average or statistically common behavior; and (iv) normality as process or constant redefinition of the human condition relative to advances in civilization [1]. Canguilhem offers an additional viewpoint, defining normality as the individual's capacity to adjust his relationship to the environment toward the restitution of a norm, when confronted with "error" or deviation due to an anomaly or defect [2].

As cited by Sabshin, Freud's work, *Civilization and its Discontents* [3], describes civilization's impact on human nature as an intrinsic potential source of pathology. The imposition of culture upon the biological organism necessarily induces discomfort, though not always illness. To the extent that society's norms are evolving and imperfect, absolute conformity to such norms would entirely efface the individual's initiative. Normality, thus involves the negotiation of desires within social parameters and the law.

Values, Theory, and the Problem of Stigma

In the previously cited issue of *Archives of General Psychiatry*, medical sociologist Strauss argued that deviant behavior becomes pathological through "visibility," through public attention [4]. Deviant behavior is assigned significance through societal attribution of harm or potential harm – a judgment of value. The process of distinguishing between deviant behaviors that are due to mental health problems and behaviors that are not due to mental pathology has been subject to complex debate.

Wakefield [5, 6] is a key contemporary author in this debate, who introduced the concept of *harmful dysfunction* as a criterion for disorder. Although a detailed summary of the debate is beyond the scope of this article, the concept of harmful dysfunction has gained credence and acceptance and requires elaboration. Among other topics, Wakefield has addressed the advantages of an approach to diagnosis that emphasizes the development of a common vocabulary for improved communication between clinicians and for research purposes, and which thereby allows for the clarification of causation of symptoms using multiple explanatory theories.

Wakefield developed the concept of harmful dysfunction in response to a lack of clarity regarding the definition of a "disorder." Although clinicians utilize the concept of disorder and are able to reach a consensus about what is and is not a disorder, the formulation of an operational definition of disorder remains problematic. The concept of harmful dysfunction involves two separate elements, both of which must be present in order for there to be a mental disorder. First, there must be *harm* of some kind to the individual or others. This is a value judgment, in reference to societal norms. However, harm alone would not define a mental disorder, since many forms of harmful behavior are simply social evaluative judgments. The second element *dysfunction* implies a cause that is in some way internal to the functioning of the individual – in other words, a disturbance of an evolutionary capacity from which the resulting harmful behavior flows. If there is dysfunction but no harm, then there is no disorder. Similarly, if there is harm but no dysfunction, there is no mental disorder.

Klein agrees with Wakefield and adds the "involuntary" nature of the dysfunction [7]. In a disorder, harm arises from a process that occurs without the individual choosing it. Spitzer remarks that the current diagnostic criteria may be overly inclusive of nonpathological syndromes, and that the testing of diagnostic criteria according to the harmful dysfunction analysis would result in more precise definitions of disorders. With regards to adjustment disorder (symptoms which arise in the context of life stressors) he writes, "the degree of distress ... is a poor marker for discriminating disorder from nondisorder. What would be more helpful would be the issue of whether the individual's response to the stressor helped or hindered the individual in dealing with the stressor. ... If the reaction, even if associated with marked distress, facilitated dealing appropriately with the stressor, then a judgment of nondisorder would be appropriate." [8].

Separation of harm from dysfunction as components of mental disorder allows for clarification of the nature of psychopathology, as defined by norms that are not entirely determined by judgments of social value. The definition of disorder as a combination of harmful manifestations and underlying dysfunction may reduce the tendency toward stigmatization of the mentally disordered person, since disorder is, thus, attributed neither to social norms nor to causes that are beyond the individual's control [9].

Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR)

In the United States, there is consensus by mental health forensic experts on the use of the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* (DSM) for attribution of diagnoses in legal proceedings, despite the fact that the DSM explicitly indicates that it is not designed for this purpose [10]. The first edition was issued in 1952, and involved narrative, descriptive paragraphs according to categories from the psychoanalytic orientation of psychopathology at that time. The second edition, with some revisions, was issued in 1968 and continued the previous edition's characterization of disorders through narrative paragraphs. A major conceptual shift occurred in 1980 with the release of the third edition, shaped in parallel with the research diagnostic criteria (RDC), which identified disorders according to specific criteria, which were required in order to assign the diagnosis. This was a significant step in the international standardization of diagnosis for the purposes of biomedical research, and the DSM was also applied in clinical and public mental health settings [11]. The strict set of criteria present for establishing a diagnosis were designed to focus on the phenomenology, or symptom description, rather than the underlying theory or causal explanation of the symptoms.

The same conceptual basis was retained in subsequent revisions, the DSM-III-R (R for revised), the DSM-IV and the DSM-IV-TR (TR for text revision, indicating that the accompanying text was updated without changing the criteria). However, the two editions of the DSM-IV emphasized empirical evidence, including field trials of the diagnostic criteria and classification, as well as data from published research [12]. A dramatically different system of classification, with a new conceptual basis, is likely in the forthcoming fifth edition of the DSM (see below).

The DSM-IV-TR divides mental disorders into broad categories and defines within each grouping a set of specific disorders with symptom lists and other criteria that must be met in order to assign the diagnosis. While some diagnoses are excluded or superseded by the presence of symptoms meeting criteria for another disorder, other diagnoses allow for the presence of multiple other disorders, sometimes with overlapping symptoms. Although the definitions require clinical judgment in assigning diagnosis, the method of arriving at a diagnosis specifies that if the clinical manifestations do not meet all of the necessary criteria for a given diagnosis, the diagnosis cannot be assigned.

For this reason, many clinically significant syndromes do not fit into specific DSM disorders, and the DSM provides for these through "not otherwise specified" diagnoses within each large category. The "not otherwise specified" diagnoses generally focus on a predominant symptom, such as anxiety or depressed mood, and indicate that such symptoms are present, without the individual meeting full criteria for a specific disorder. Diagnoses with the suffix "not otherwise specified" should not be taken to be less meaningful than specific diagnoses, since the individual's level of impairment may be as significant as that of a person who meets the criteria for a specific diagnosis. The reality of clinical practice suggests that disorders arise on a continuum with gradations from subtle to more severe and marked manifestations; the existence of "not otherwise specified" diagnoses reflects the reality that some symptomatic individuals with clinically significant distress or impairment do not fit neatly into the criteria which define research categories, but nevertheless closely resemble them.

The International Classification of Diseases, 10th Edition (ICD-10)

The International Classification of Diseases (ICD) includes diagnostic codes and criteria for all medical disorders, including mental health disorders [13]. The ninth version of the ICD was finalized 1 year after the task force for DSM-III was formed, without coordination of the criteria or nomenclature [14]. The preparation for a 10th edition of the ICD began well in advance of the work of the DSM-IV task force, making it impossible for the two systems to become identical [15]. During the subsequent development of the ICD-10 and the DSM-IV, international groups worked to improve the concordance between the two [15]. When both systems of classification are present in legal proceedings (e.g., where expert opinions are expressed in terms of DSM-IV-TR criteria and medical records contain ICD-10 coding), possible differences in specific criteria used to assign diagnoses should be clarified.

Beyond the Clinical Interview

Although the clinical interview and the mental status examination (*see* **Mental Status: Examination**) form the basis of diagnostic assessment, the examiner often has recourse to additional information, including documents, such as prior hospital records. Some of the criteria required for assigning a DSM diagnosis are difficult to confirm or disaffirm through the evaluee's self-report of symptoms or past behavior alone.

Parallel History

Some disorders are characterized by a lack of awareness of symptoms. Individuals with disorders such as schizophrenia or a personality disorder often manifest symptoms that they themselves are unable to perceive. Their symptoms are more readily perceived by others in their entourage. For this reason, information obtained from sources other than the evaluee (family members, employers, or the individual's treating physician) may be of value in determining the nature of the psychopathology. In addition, diagnostic criteria for some disorders include the time course of symptoms, and parallel history obtained from family members or others may be useful, for example, in confirming whether or not the symptoms of schizophrenia have been present for six months or longer, or whether elements suggestive of a personality disorder do in fact represent a life-long pattern, present from the time of childhood or adolescence.

Psychometric Validation

Although a large number of rating scales and standardized diagnostic instruments are available for the assessment of psychopathology (*see* **Psychological Testing**), the use of such techniques is not obligatory in forensic clinical practice for the determination of the presence of a mental disorder. Some examiners routinely administer psychometric tests, whereas others have recourse to them only in cases where there is ambiguity in the examination or where findings in the clinical interview suggest the presence of malingering (*see* **Malingering: Forensic Evaluations**). Caution must be exercised in relying too heavily upon the results of standardized testing, because the results are calibrated upon findings within a defined population as statistical probabilities and not as probabilities in an individual case. Even reporting of a highly validated and reliable examination such as the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) is usually expressed in terms such as "Individuals with this profile tend to . . ., " suggesting that the result indicates a trend that is relevant to the particular case only after appreciation in the context of findings upon clinical examination.

However, the diagnosis of a mental defect (*see* **Mental Retardation**) such as mental retardation or dementia (disturbance of memory and either language, motor activity, or complex task completion) (*see* **Neuropsychological Assessment**) does require testing beyond the clinical examination. By definition, these mental defects involve deficits in cognitive function that are sometimes subtle and require detailed evaluation with psychometrically validated techniques. In some cases, dementia may be so evident upon clinical examination that confirmation through specialized testing is unnecessary.

Genetics

Efforts in the field of neurobiology to determine the genetic basis of mental disease (*see* **Genomics and Behavioral Evidence**) have yielded results that are as yet unreliable, though some researchers propose classification of psychopathology on a genetic basis in the future [16–18]. Owing to variability in the expression of genes and environmental factors in disease causation, it appears unlikely that genetic typing will supplant clinical examination as a basis for diagnosis at any time in the near future. Even a disease with strong genetic underpinnings such as schizophrenia is present in both identical twins only approximately 50% of the time [19, 20]. Should genetic evidence in support of a diagnosis become admissible in courts, it will likely have a status similar to that of psychological testing, as data that supplements the clinical impression and not as definitive proof of diagnosis.

Neuroimaging

Recent developments in structural and functional magnetic resonance imaging (MRI) have led some jurisdictions to permit testimony from experts in brain imaging (*see* **Deception: Detection of and Brain Imaging**) as an aid in the assessment of psychopathology, though the future use of such

techniques in legal proceedings remains uncertain at the time of writing of this article.

Structural Approaches

The concept of *structure* refers to the underlying organization of internal psychical functioning, at a level beyond what is manifest. Structure is interpreted in different ways by various schools of thought, particularly within psychodynamic psychiatry and psychoanalysis. The notion refers, in general, to latent relationships between different elements of the psyche, viewed as a set [21]. Recent articles utilize the concept of structure to describe the underlying organization of psychopathology, including the integration of temperament, personality, and symptoms [22–24].

Although the various conceptualizations of structure are valuable in clinical practice, they pertain to enduring patterns relating to personality diagnosis and are seldom relevant to specific legal adjudications such as insanity at the time of the criminal act. Structural assessment of a forensic evaluee is challenging under evaluation conditions, since accurate formulation may require serial interviews over a significant period of time. Hypotheses regarding structural diagnosis may lead the forensic examiner to identify traits or historical features that he may then integrate into formulations that are pertinent to criminal mitigation or the assessment of civil damages. In addition, preliminary hypotheses regarding structural diagnosis may lead to further exploration of an individual's points of vulnerability in the setting of assesment of fitness for duty or of risk for future violence.

Psychodynamic Diagnostic Manual (PDM)

In 2006, a task force representing six psychoanalytic organizations published the result of their collaborative effort to develop a system of diagnostic classification to "complement the DSM and ICD efforts of the past 30 years in cataloguing symptoms by explicating the broad range of mental functioning," and with a wider appreciation of the patient as a "whole person" with a complex and rich life history [25]. Prior to the Psychodynamic Diagnostic Manual (PDM), some authors provided cogent psychoanalytic perspectives on DSM-IV diagnoses, reflecting the perceived need for enhanced formulations of the processes underlying descriptive diagnostic categories [26, 27]. Although the PDM includes syndromes that are recognizable to readers of the DSM-IV, it also describes other syndromes and classifies them in a manner that allows the clinician to find a basis to link diagnoses that are commonly co-occurring. In other words, in the PDM disorders are not considered as separate categories but as identifiable manifestations of the individual's underlying internal functioning.

The PDM is likely to be useful to treating clinicians, but is unlikely to supplant the DSM as a basis for expert opinion, a use for which it is not explicitly intended. The authors state quite clearly that it is not their intention to displace the DSM as a scientific classification necessary for research studies. Only time will determine whether forensic experts will refer to the PDM in supplementing their case formulations when arriving at an opinion or testifying in court. The PDM represents a consensus statement that supports what careful forensic mental health professionals already do – that is to say, to seek in detail the case-specific factors that are at issue in explaining past behavior and the likely future evolution of an individual within his personal history and in terms of idiosyncratic subjective experience.

Dimensional Approaches: the DSM-V and Beyond

Planning for the DSM-V began in 1999 and resulted in the publication in 2002 of *A Research Agenda for DSM-V* [28]. After a review of the existing literature, DSM-V workgroups were formed in 2007, with an aim to publish DSM-V by as early as 2011 [29]. A major conceptual revision of the DSM is anticipated, in light of numerous observations regarding the DSM-IV, which utilizes a *categorical* system of classification of disorders, in which the particular disorder is either present or absent. Clinicians have noted that disorders do not easily conform to such a system, since symptoms that partly meet criteria may result in significant impairments that are susceptible to treatment. Researchers have noted that categorical diagnosis may exclude individuals from studies, resulting in a skewed picture of underlying pathological processes [30].

In order to remedy such difficulties, the DSM-V workgroups began to consider the feasibility of a *dimensional* approach to diagnosis, which would

retain the categorical diagnoses currently in use, with modifiers to indicate the milder versions or partial syndromes, which are nonetheless related to the primary diagnosis. This approach is based on the assumption that psychopathology exists on a continuum, although there are characteristics which define a "prototype" of the disorder [31, 32]. Previously, the terms "categorical" and "dimensional" were most widely used in the diagnosis of personality disorders, where a wide range of assessment instruments were developed, alternately defining disorders using strict criteria for determining the presence or absence of a specific entity ("borderline personality disorder" or "antisocial personality disorder") *versus* defining disorder according to the predominance of traits that were deemed to be more or less characteristic of the individual's life pattern. The clarification of relevant dimensions in disorders other than personality disorders is likely to require the development of consensus through empirical study of the types of dimensions most characteristic of the "prototype" for each specific disorder [33, 34].

The DSM-V may also include a reorganization of disorders into *spectra*, of related disorders. Beyond the identification of relevant dimensions for specific disorders, the spectrum approach seeks to group together disorders that share similar underlying processes [35, 36]. For example, pathological gambling, eating disorders, and substance abuse could be conceptualized as disturbances arising from an underlying process of compulsion (*see* **Compulsion**), and these might therefore be more likely to occur together in the same individual.

Although the process of DSM revision is lengthy and careful, including field testing of the criteria prior to finalization, commentators observe that the major conceptual change will likely require further refinement after the DSM-V is published [12, 37]. After extensive discussion, the final published criteria for DSM disorders summarize only the essential findings of the committees. For this reason, the *DSM-IV Sourcebook* [38] – a summary of debates prior to arriving at final criteria – is a useful reference when clarification is needed, and provides further information for assessment of the intended meaning of DSM diagnoses. Hopefully, the DSM-V will also be accompanied by such a companion sourcebook to aid the reader.

Relevance to Legal Proceedings

At first blush, a dimensional approach to the diagnosis of mental disease would appear to be contraindicated in the context of legal proceedings. Case law has defined, for example, what diagnoses are eligible for consideration as a "severe mental disease" leading to criminal nonresponsibility, implying that adjudications require clear thresholds provided by categorical diagnosis. However, closer scrutiny reveals that a dimensional approach is compatible with the process of expert evaluation in cases involving mental disorders, to the extent that forensic opinions do not rely solely upon diagnosis, but rather present the relationship between particular symptoms and specific behavior or distress, along a continuum of severity. A term such as "severe mental disease" is, thus, a legal term of art that can accommodate evolution in the professional standards regarding mental pathology, in accordance with contemporary scientific definitions.

Conclusion

Psychopathology is a complex and evolving field of study, with a long historical tradition that merits a closer examination than is possible in this article. Systems of classification of mental disorders have been constructed in accordance with changing needs and new findings in research. Although manuals such as the DSM and the PDM are commercially available for purchase, works such as these should be viewed as guides for use by mental health professionals in light of their education, experience, and intuition based on clinical reasoning and the specifics of each case. Through research and consensus, it has been possible to define categories of psychopathology using standardized criteria that serve as a common language, across theoretical orientations. However, debate continues as to the adequacy of these categories, particularly given the wide variety of clinical manifestations and underlying mental processes involved in the experience of each person, and efforts are under way to revise the existing classifications in light of new findings.

References

[1] Sabshin, M. (1967). Psychiatric perspectives on normality, *Archives of General Psychiatry* **17**, 258–264.

[2] Canguilhem, G. (1989). A new concept in pathology: error (1963–1966), *The Normal and the Pathological*, Zone Books, New York.
[3] Freud, S. (1930, 1953). *Civilization and its Discontents* in J. Strachey, ed, *The Standard Edition of the Complete Psychological works of Sigmund Freud*, Hogarth Press and The Institute of Psycho-Analysis, London, Vol. 21.
[4] Strauss, A. (1967). A sociological view of normality, *Archives of General Psychiatry* **17**, 265–270.
[5] Wakefield, J.C. (1997). Diagnosing DSM-IV – Part I: DSM-IV and the concept of disorder, *Behaviour Research and Therapy* **35**(7), 633–649.
[6] Wakefield, J.C. (1999). The concept of disorder as a foundation for the DSM's theory-neutral nosology: response to Follette and Houts, part 2, *Behaviour Research and Therapy* **37**, 1001–1027.
[7] Klein, D.F. (1999). Harmful dysfunction, disorder, disease, illness, and evolution, *Journal of Abnormal Psychology* **108**(3), 421–429.
[8] Spitzer, R.L. (1999). Harmful dysfunction and the DSM definition of mental disorder, *Journal of Abnormal Psychology* **108**(3), 430–432.
[9] Dain, N. (1994). Reflections on antipsychiatry and stigma in the history of American Psychiatry, *Hospital and Community Psychiatry* **45**(10), 1010–1014.
[10] American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition, Text Revision, APA Press, Washington, DC, pp. 32–33, 37.
[11] First, M.B. (2002). The DSM series and experience with DSM-IV, *Psychopathology* **35**, 67–71.
[12] Regier, D.A., Narrow, W.E., First, M.B. & Marshall, T. (2002). The APA classification of mental disorders: future perspectives, *Psychopathology* **35**, 166–170.
[13] World Health Organization (1990). *International Classification of Diseases and Related Health Problems*, 10th Revision, Geneva.
[14] Widiger, T.A., Frances, A.J., Pincus, H.A., Davis, W.W. & First, M.B. (1991). Toward an empirical classification for the DSM-IV, *Journal of Abnormal and Social Psychology* **100**(3), 280–288.
[15] Kendell, R.E. (1991). Relationship between the DSM-IV and the ICD-10, *Journal of Abnormal Psychology* **100**(3), 297–301.
[16] Gottesman, I.I. & Gould, T.D. (2003). The endophenotype concept in psychiatry: etymology and strategic intentions, *The American Journal of Psychiatry* **160**, 636–645.
[17] Cannon, T.D. & Keller, M.C. (2006). Endophenotypes in the genetic analyses of mental disorders, *Annual Review of Clinical Psychology* **2**, 267–290.
[18] Bearden, C.E. & Freimer, N.B. (2006). Endophenotypes for psychiatric disorders: ready for primetime? *Trends in Genetics* **22**, 306–313.
[19] Cardno, A.G. & Gottesman, I.I. (2000). Twin studies of schizophrenia: from bow-and-arrow concordances to Star Wars Mx and functional genomics, *The American Journal of Medical Genetics* **97**, 12–17.
[20] Kendler, K.S. (2001). Twin studies of psychiatric illness, *Archives of General Psychiatry* **58**, 1005–1014.
[21] Vandermersch B. (1998). *Structure du Dictionnaire de la Psychanalyse*, R. Chemama & B. Vandermersch, eds, 2nd Edition, Larousse Paris, pp. 407–408.
[22] Clark, L.A. (2005). Temperament as a unifying basis for personality and psychopathology, *Journal of Abnormal Psychology* **114**(4), 505–521.
[23] Cuthbert B.N. (2005). Dimensional models of psychopathology: research agenda and clinical utility, *Journal of Abnormal Psychology* **114**(4), 565–569.
[24] Krueger, R.F., Watson, D. & Barlow, D.H. (2005. Introduction to the special section: toward a dimensionally based taxonomy of psychopathology, *Journal of Abnormal Psychology* **114**(4), 491–493.
[25] PDM Task Force (2006). *Psychodynamic Diagnostic Manual*, Alliance of Psychanalytic Organizations, Silver Spring.
[26] Gabbard, G.O. (2000). *Psychodynamic Psychiatry in Clinical Practice*, 3rd Edition, American Psychiatric Press, Washington, DC.
[27] McWilliams, N. (1994). *Psychoanalytic Diagnosis: Understanding Personality Structure in the Clinical Process*, The Guilford Press, New York.
[28] Kupfer, D.J., First, M.B. & Regier, D.A. (2002). *A Research Agenda for DSM-V*, American Psychiatric Publishing, Washington, DC.
[29] DSM-V website at http://dsm5.org (accessed Nov 4, 2007).
[30] Kraemer, H.C. (2007). DSM categories and dimensions in clinical and research contexts, *International Journal of Methods in Psychiatric Research* **16**(S1), S8–S15.
[31] Jablensky, A. (2005). Categories, dimensions and prototypes: critical issues for psychiatric classification, *Psychopathology* **38**, 201–205.
[32] Widiger, T.A. & Samuel, D.B. (2005). Diagnostic categories or dimensions? A question for the Diagnostic and Statistical Manual of Mental Disorders – Fifth Edition, *Journal of Abnormal Psychology* **114**(4), 494–504.
[33] First, M.B. (2005). Clinical utility: a prerequisite for the adoption of a dimensional approach in DSM, *Journal of Abnormal Psychology* **114**(4), 560–564.
[34] Jablensky A. (2005). Categories, dimensions and prototypes: critical issues for psychiatric classification, *Psychopathology* **38**, 201–205.
[35] Cuthbert, B.N. (2005). Dimensional models of psychopathology: research agenda and clinical utility, *Journal of Abnormal Psychology* **114**(4), 565–569.
[36] Krueger, R.F., Markon, K.E., Patrick, C.J. & Iacono, W.G. (2005). Externalizing psychopathology in adulthood: a dimensional-spectrum conceptualization and its implications for DSM-V, *Journal of Abnormal Psychology* **114**(4), 537–550.
[37] Brown, R.A. & Barlow, D.H. (2005). Dimensional versus categorical classification of mental disorders in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* and beyond: comment on the

special section, *Journal of Abnormal Psychology* **114**(4), 551–556.

[38] Widiger, T.A., Frances, A.J., Pincus, H.A., First, M.B., Ross, R. & Davis, W. (1994). *DSM-IV Sourcebook*, American Psychiatric Association, Washington, DC, 3 Vols.

SUZANNE YANG AND FRANÇOIS SAUVAGNAT

Psychopathy

Psychopathy or psychopathic personality disorder is referred to as *antisocial personality disorder* in the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* or *DSM-IV* [1] and as dissocial personality disorder in the tenth edition of the *International Statistical Classification of Diseases and Related Health Problems* or *ICD-10* [2]. Previously, it was referred to as *sociopathy* or *sociopathic personality disorder*.

Clinical Description

According to clinical descriptions over the past 200 years [3, 4], psychopathy is characterized by a broad range of symptoms in several major domains of personality functioning. In the domain of *behavioral organization*, they include lack of perseverance, unreliability, recklessness, restlessness, disruptiveness, and aggressiveness. The *emotionality* domain includes lack of anxiety, lack of remorse, lack of emotional depth, and lack of emotional stability. The domain of *interpersonal attachment* includes detachment, lack of commitment, and lack of empathy or concern for others. The domain *interpersonal dominance* includes antagonism, arrogance, deceitfulness, manipulativeness, insincerity, and glibness or garrulousness. The *cognitive* domain includes suspiciousness, inflexibility, intolerance, lack of planfulness, and lack of concentration. Finally, the *self* domain includes self-centeredness, self-aggrandizement, self-justification, and a sense of entitlement, uniqueness, and invulnerability.

Assessment and Diagnosis

In the *DSM-IV* and *ICD-10*, the diagnostic criteria for psychopathy focus primarily on symptoms from the behavioral organization domain, especially those related to violation of explicit social norms. In many civil psychiatric settings, these diagnostic criteria have adequate reliability (e.g., stability or consistency across evaluators and time) and validity (e.g., prognostic value with respect to poor treatment response, institutional misbehavior, or community violence) [1, 5]. In forensic settings, however, the *DSM-IV* and *ICD-10* criteria are less useful. Their heavy focus on criminality leads to extremely high prevalence rate – typically 50–75% or higher – in correctional offenders and forensic psychiatric patients [1, 5]. For this reason, many forensic mental health professionals prefer more comprehensive diagnostic criteria, such as the *Hare Psychopathy Checklist Revised* or *PCL-R* [6] and its progeny, such as the *Screening Version* or *PCL:SV* [7] and the *Youth Version* or *PCL:YV* [8]. These latter tests yield a lifetime prevalence rates of ~15 to 25% – about one-third the rate observed using the *DSM* criteria for antisocial personality disorder [5, 6] – and also have superior reliability and validity (*see* **Psychopathy Checklists**).

Course

Symptoms of psychopathy may emerge as early as age 6–10 [9], and it is common for adults with psychopathy to have been diagnosed in childhood or adolescence as suffering from one of the disruptive behavior disorders. Indeed, the *DSM-IV* diagnostic criteria for antisocial personality disorder require that the person met the criteria for a conduct disorder before age 15 [1]. Unfortunately, the majority of children or adolescents so diagnosed – 50 to 75% or more – spontaneously desist antisocial behavior and do not go on to develop psychopathy as adults [9, 10]. Consequently, it is recommended not to diagnose psychopathy before the beginning of early adulthood, at least 18 years old or possibly even 25 years old [1, 2].

The course of the disorder during adulthood is characterized by relative stability. For example, there is evidence of moderate diagnostic stability across periods of several months to several years [6], persistence of symptoms across adulthood [11, 12], and long-term risk for negative health outcomes

such as morbidity and mortality [13]. But there is also evidence that symptom severity may fluctuate substantially over time [14].

Prevalence

Epidemiological research in the United States indicates that the lifetime prevalence of psychopathy in the general population, according to *DSM-IV* or similar criteria, is ~1.5 to 3.5% [15, 16]; in correctional offenders, the rate is 50–75% [6].

When more comprehensive diagnostic criteria are used, the prevalence rate is considerably lower. For example, research using the *PCL-R* with correctional offenders and forensic psychiatric patients in the United States has reported lifetime prevalence rates of ~15 to 25% – about one-third the rate observed using the *DSM* criteria [6].

Gender, Age, and Sociocultural Factors

Lifetime prevalence rates of psychopathy vary across three major group factors: gender, age, and culture. First, with respect to gender, the male : female sex ratio in diagnosis is typically about 3 : 1 [15, 16]. This gender difference is not limited to a few clinical features, but is evident across the full range of symptomatology. Second, with respect to age, some epidemiological research in the United States using *DSM-III* and *DSM-III-R* criteria has reported a cohort effect, with higher lifetime prevalence rates in younger generations than in older generations [17]. Third, with respect to culture, anthropological and epidemiological research indicates that psychopathy is found across cultures [18], but there is evidence of cross-cultural differences in prevalence. For example, according to general population studies, the lifetime prevalence of psychopathy in Taiwan is much lower than that reported in the United States [19], and according to studies of correctional offenders and forensic psychiatric patients, the prevalence of psychopathy is higher in the United States than in Europe [20.]

These group differences may be due to *cultural facilitation* [20, 21]. In highly individualistic cultures such as the United States, norms and values that emphasize the importance of distinctiveness, status, self-confidence, honor, competition, and freedom from obligations to others may also foster the development of extreme manifestations of the same characteristics – for example, conceit, manipulativeness, irresponsibility, pathological dominance, and aggressiveness. Similarly, within a dominant culture, the expression of symptoms of psychopathy may be facilitated in certain subgroups, such as males or younger generations, that subscribe to more individualistic norms and values.

The group differences may also be due to inadequacies in diagnostic criteria. The current diagnostic criteria for psychopathy may be biased to reflect its prototypical manifestation in young males from individualistic cultures; if this is true, any differences due to gender, age, and culture may be smaller than suggested by research to date [20].

Comorbidity

Psychopathy has a high rate of comorbidity with substance use disorders [15, 22, 23]. This comorbidity may reflect a common etiological mechanism, or it may be that in some cases substance use disorders are a consequence or complication of psychopathy.

Psychopathy also has a high rate of comorbidity with other personality disorders, specifically, borderline, the Cluster B narcissistic, and histrionic personality disorders in *DSM-IV* or emotionally unstable and histrionic personality disorders in *ICD-10* [24, 25]. The high rate of comorbidity among them almost certainly reflects inadequacies in their diagnostic criteria (i.e., a failure to "carve nature at its joints"), as well as common etiological factors.

Low rates of comorbidity are observed between psychopathy and certain other personality disorders, specifically the Cluster C avoidant, dependent, and obsessive–compulsive personality disorders or anxious/avoidant, dependent, and anankastic personality disorder in *ICD-10* [24, 25]. The low rates of comorbidity among the disorders suggest they have independent or even competing etiologies.

The rates of comorbidity between psychopathy and most other disorders are inconsistent, unclear, or unremarkable [15, 24, 25].

Etiology

The etiology of psychopathy is unknown. Theoretical models of etiology can be divided into two main categories based on whether they view psychopathy

as a true disorder, that is, a *bona fide* form of mental abnormality.

Theoretical models of psychopathy as mental abnormality have focused on the potential causal influence of social and biological factors. Overall, the research literature supports the relative importance of biological over social factors. With respect to social factors, there are no child-rearing experiences, familial dysfunctions, or adverse life experiences that are found both frequently and specifically in people with psychopathy compared with people with other personality disorders. As noted previously, however, sociocultural factors certainly appear to play a role in the expression of the disorder [20, 21]. With respect to biological factors, researchers have reported elevated rates of prenatal trauma, neurotransmitter abnormalities, and structural abnormalities of the brain associated with symptoms of psychopathy [26–28], but none of these factors is clearly pathognomonic. Also, some adoption research has reported that the heritability of psychopathy is substantial [11, 29], but molecular genetic research has not identified genetic markers. A common theme underlying many etiological theories that focus on biological factors is that psychopathy is associated with impaired ability to experience emotions and integrate them in executive functions; this core emotional deficit results in a failure of attachment to others, inattention to cues of impending punishment, and insensitivity to reward or punishment.

Other theoretical models reject the notion that psychopathy is a mental abnormality at all. First, some interpersonal and behavioral genetic theories view psychopathy as an extreme variant of the same personality traits found in all people [30, 31]. According to these theories, psychopathy is not associated with any unique or specific causal influences and any differences between people with *versus* without the disorder are quantitative rather than qualitative in nature – that is, the differences are a matter of degree rather than of kind. Second, some sociobiological and evolutionary theories view psychopathy as an adaptation [32]. According to these theories, the human species has the genetic capacity to express traits associated with psychopathy. In sociobiological theories, the genetic capacity exists in only a minority of humans and its manifestation is only partially dependent on environmental circumstances; in evolutionary theories, the genetic capacity exists in all humans, but is manifested in only a minority of humans who are exposed to specific environmental circumstances. In both the theories, people with psychopathy have an evolutionary advantage in terms of an increased likelihood of producing offspring.

Treatment

There is no methodologically sound research on the treatment of psychopathy, and so there no good evidence that it can be successfully treated [33]. Notwithstanding methodological limitations, there is evidence that psychopathy is associated with increased risk for disruptive behavior during treatment, treatment dropout, and posttreatment recidivism [34].

Forensic Relevance

Psychopathy does not appear to impair cognitive abilities to the extent that it is relevant in the assessment of psycholegal competencies or capacities. With respect to civil-forensic evaluations, psychopathy generally is not considered by either research or law to be relevant to issues such as competence to consent to treatment, enter into contracts, or testify. With respect to criminal forensic evaluations, it is generally not considered relevant to issues such as competence to confess or stand trial, or to ability to form criminal intent.

In contrast, psychopathy does appear to impair volitional abilities to the extent that it is relevant to the assessment of risk for serious crime, and, in particular, for violence. A large body of research indicates that psychopathy is a major risk factor for disruptive behavior while institutionalized and for recidivism upon release to or while under supervision in the community [35, 36]. Explanations include the following [35]:

1. Psychopathy increases the perceived benefits of serious crime. For example, interpersonal symptoms may make demeaning, controlling, and hurting other people rewarding; and behavioral symptoms may make exciting, risk activities rewarding.
2. Psychopathy decreases the perceived costs of serious crime. For example, attachment symptoms, emotional deficit, and self symptoms may

result in failure to be deterred by anxiety, empathy, remorse, or self-punishment.
3. Psychopathy destabilizes general psychosocial adjustment. For example, all symptoms of psychopathy may increase problems in daily living, decrease social integration or cohesion, or otherwise increase interpersonal conflict.

Given the association between psychopathy and serious crime, as well as the lack of demonstrably effective treatment for psychopathy, it is understandable that psychopathy is potentially relevant in a wide range of psycholegal evaluations involving risk for serious crime. With respect to civil-forensic evaluations, psychopathy is relevant to issues such as parental capacity (i.e., risk for child abuse), employee discipline and dismissal (i.e., risk for workplace violence), and civil commitment as a sexually violent predator (i.e., risk for sexual violence). With respect to criminal forensic evaluations, it is relevant to issues such as pretrial release, sentencing, correctional classification, and community registration, notification, and supervision.

References

[1] American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition, American Psychiatric Press, Washington, DC.

[2] World Health Organization (1992). *ICD-10: International Statistical Classification of Diseases and Related Health Problems*, 10th Revision, World Health Organization, Geneva.

[3] Arrigo, B. & Shipley, S. (2001). The confusion over psychopathy (I): historical considerations, *International Journal of Offender Therapy and Comparative Criminology* **45**, 325–344.

[4] Berrios, G. (1996). *The History of Mental Symptoms: Descriptive Psychopathology Since the Nineteenth Century*, Cambridge University Press, Cambridge.

[5] Widiger, T.A., Cadoret, R., Hare, R.D., Robins, L.N., Rutherford, M. Zanarini, M., Alterman, A., Apple, M., Corbitt, E., Forth, A.E., Hart, S.D., Kultermann, J., Woody, G. & Frances, A. (1996). DSM-IV antisocial personality disorder field trial, *Journal of Abnormal Psychology* **105**, 3–16.

[6] Hare, R.D. (2003). *Manual for the Hare Psychopathy Checklist – Revised*, 2nd Edition, Multi-Health Systems, Toronto.

[7] Hart, S.D., Cox, D.N. & Hare, R.D. (1995). *Manual for the Hare Psychopathy Checklist: Screening Version (PCL:SV)*, Multi-Health Systems, Toronto.

[8] Forth, A.E., Kosson, D.S. & Hare, R.D. (2003). *Hare Psychopathy Checklist Revised: Youth Version (PCL:YV)*, Multi-Health Systems, Toronto.

[9] Goldstein, R.B., Grant, B.F., Ruan, W.J., Smith, S.M. & Saha, T.D. (2006). Antisocial personality disorder with childhood – vs adolescence-onset conduct disorder: results from the national epidemiologic survey on alcohol and related conditions, *The Journal of Nervous and Mental Disease* **194**, 667–675.

[10] Simonoff, E., Elander, J., Holmshaw, J., Pickles, A., Murray, R. & Rutter, M. (2004). Predictors of antisocial personality: continuities from childhood to adult life, *The British Journal of Psychiatry* **184**, 118–127.

[11] Burt, S.A., McGue, M., Carter, L.A. & Iacono, W.G. (2007). The different origins of stability and change in antisocial personality disorder symptoms, *Psychological Medicine* **37**, 27–38.

[12] Hare, R.D., McPherson, L.E. & Forth, A.E. (1988). Male psychopaths and their criminal careers, *Journal of Consulting and Clinical Psychology* **56**, 710–714.

[13] Repo-Tiihonen, E., Virkkunen, M. & Tiihonen, J. (2001). Mortality of antisocial male criminals, *Journal of Forensic Psychiatry* **12**, 677–683.

[14] Lenzenweger, M.F., Johnson, M.D. & Willett, J.B. (2004). Individual growth curve analysis illuminates stability and change in personality disorder features: the longitudinal study of personality disorders, *Archives of General Psychiatry* **61**, 1015–1024.

[15] Compton, W.M., Conway, K.P., Stinson, F.S., Colliver, J.D. & Grant, B.F. (2005). Prevalence, correlates, and comorbidity of DSM-IV antisocial personality syndromes and alcohol and specific drug use disorders in the United States: results from the national epidemiologic survey on alcohol and related conditions, *Journal of Clinical Psychiatry* **66**, 677–685.

[16] Narrow, W.E., Rae, D.S., Robins, L.N. & Regier, D.A. (2002). Revised prevalence estimates of mental disorders in the United States using a clinical significance criterion to reconcile 2 surveys' estimates, *Archives of General Psychiatry* **59**, 115–123.

[17] Robins, L.N., Tipp, J. & Przybeck, T. (1991). Antisocial personality, in L.N. Robins & D.A. Reiger, eds, *Psychiatric disorders in America: The epidemiological catchment area study*, Free Press, New York, USA, pp. 258–290.

[18] Cooke, D.J. (1996). Psychopathic personality in different cultures: What do we know? What do we need to find out? *Journal of Personality Disorders* **10**, 23–40.

[19] Compton, W.M., Helzer, J.E., Hwu, H.G., Yeh, E.K., McEvoy, L., Tipp, J.E. & Spitznagel, E.L. (1991). New methods in cross-cultural psychiatry: psychiatric illness in Taiwan and the United States, *American Journal of Psychiatry* **148**, 1697–1704.

[20] Cooke, D.J., Michie, C., Hart, S.D. & Clark, D.A. (2005). Searching for the pan-cultural core of psychopathic personality disorder: Continental Europe and North America compared, *Personality and Individual Differences* **39**, 283–295.

[21] Paris, J. (1998). Personality disorders in sociocultural perspective, *Journal of Personality Disorders* **12**, 289–301.
[22] Grant, B.F., Stinson, F.S., Dawson, D.A., Chou, P.S., Ruan, W.J. & Pickering, R.P. (2004). Co-occurrence of 12-month alcohol and drug use disorders and personality disorders in the United States: results from the national epidemiologic survey on alcohol and related conditions, *Archives of General Psychiatry* **61**, 361–368.
[23] Hemphill, J., Hart, S.D. & Hare, R.D. (1994). Psychopathy and substance use, *Journal of Personality Disorders* **8**, 32–40.
[24] Hart, S.D. & Hare, R.D. (1989). Discriminant validity of the psychopathy checklist in a forensic psychiatric population, *Psychological Assessment* **1**, 211–218.
[25] Hildebrand, M. & de Ruiter, C. (2004). PCL-R psychopathy and its relation to DSM-IV Axis I and II disorders in a sample of male forensic psychiatric patients in the Netherlands, *International Journal of Law and Psychiatry* **27**, 233–248.
[26] Coccaro, E.F. (2001). Biological and treatment correlates, in *Handbook of Personality Disorders: Theory, Research and Treatment*, W.J. Livesley, ed, Guilford, New York, pp. 124–135.
[27] Neugebauer, R., Hoek, H.W. & Susser, E. (1999). Prenatal exposure to wartime famine and development of antisocial personality disorder in early adulthood, *Journal of the American Medical Association* **282**, 455–462.
[28] Rainze, A., Lencz, T., Bihrle, S., LaCasse, L. & Colletti, P. (2000). Reduced prefrontal gray matter volume and reduced autonomic activity in antisocial personality disorder, *Archives of General Psychiatry* **57**, 119–127.
[29] Cadoret, R., Troughton, E., Bagford, J. & Woodworth, G. (1990). Genetic and environmental factors in adoptee antisocial personality, *European Archives of Psychiatry and Neurological Sciences* **239**, 231–240.
[30] Livesley, W.J. (1998). The phenotypic and genotypic structure of psychopathic traits, in *Psychopathy: Theory, Research, and Implications for Society*, D.J. Cooke, A.E. Forth & R.D. Hare, eds, Kluwer Academic Publisher, Dordrecht, pp. 69–79.
[31] Miller, J.D., Lynam, D.R., Widiger, T.A. & Leukefeld, C. (2001). Personality disorders as extreme variants of common personality dimensions: can the five-factor model adequately represent psychopathy? *Journal of Personality* **69**, 253–276.
[32] Mealey, L. (1995). The sociobiology of sociopathy: an integrated evolutionary model, *The Behavioral and Brain Sciences* **18**, 523–599.
[33] Dolan, B. & Coid, J. (1993). *Psychopathic and Antisocial Personality Disorders: Treatment and Research Issues*, Gaskell, London.
[34] Hemphill, J.F. & Hart, S.D. (2002). Motivating the unmotivated: psychopathy, treatment, and change, in *Motivating Offenders to Change*, M. Mc Murran, ed, John Wiley & Sons, Chichester, pp. 193–219.
[35] Hart, S.D. (1998). The role of psychopathy in assessing risk for violence: conceptual and methodological issues, *Legal and Criminological Psychology* **3**, 121–137.
[36] Douglas, K.S., Vincent, G.M. & Edens, J.F. (2006). Risk for criminal recidivism: the role of psychopathy, in *Handbook of Psychopathy*, C.J. Patrick, ed, Guilford, New York, pp. 533–554.

Related Articles

Dangerousness: Risk of
Psychopathy Checklists
Risk Assessment: Patient and Detainee

STEPHEN D. HART

Psychopathy Checklists

Psychopathy is a specific form of personality disorder (*see* **Psychopathy**). The *Hare Psychopathy Checklist-Revised* (*PCL-R*) [1, 2] and the *Screening Version of the Hare Psychopathy Checklist-Revised* (*PCL:SV*) [3] are standardized psychological tests of lifetime psychopathic symptoms in adults. They have proven to be particularly useful forensic mental health evaluations and are in wide use, both in the original English and in numerous foreign language translations.

History

In the mid-1970s, Robert Hare was dissatisfied with the assessment procedures for psychopathy then in use, as they focused primarily on impulsive, irresponsible, and antisocial behavior and had only low-to-moderate correlations with each other [4]. Hare began work on the development of measures that were more comprehensive and reliable. His first attempt, the *Psychopathy Checklist* (PCL), was distributed informally starting in about 1980 [5] and stimulated considerable interest among researchers and forensic mental health professionals. The manual for the revised PCL, or PCL-R, was published in 1991 [1] and updated in 2004 [2].

Several derivations of PCL-R were developed by Hare and colleagues. One of these was PCL:SV, which is shorter and easier to administer than PCL-R, as well as being appropriate for use in general community and civil psychiatric settings [3]. Development of PCL:SV began in the mid-1980s, concurrent with the revision of the PCL, and a test manual was published in 1995.

Format

PCL-R and PCL:SV are multi-item observer rating scales designed to assess the lifetime presence and severity of psychopathic symptoms. Ratings are based on personal interviews and third-party information (e.g., collateral interviews and official records). Each item reflects a different symptom or characteristic of psychopathy, defined in the test manual, and is rated on a 3-point scale (0 = absent, 1 = present to a limited extent, 2 = present and severe). A small number of items can be omitted if insufficient information is available to rate them. Items are summed (and prorated, if necessary) to yield total scores. Total scores can be interpreted dimensionally, relative to norms from various comparison groups, or cut-off scores can be used to make categorical diagnoses. Items can also be summed to yield factor scores, although these are used primarily for research purposes.

PCL-R is intended for use with adult correctional offenders and forensic psychiatric patients, male or female. It comprises 20 items, some of which reflect pathological personality traits and others that reflect specific forms of antisocial conduct. Item definitions average about 200 words or so in length. Total scores range from 0 to 40; a cut-off score of 30 and higher is used to diagnose psychopathy. Norms are available for a variety of reference groups. Scores can be calculated for two superordinate factors or four subordinate factors.

PCL : SV is intended for use with adults in general community and civil psychiatric settings, in addition to adult correctional offenders and forensic psychiatric patients, male or female. It comprises 12 items, all of which tap relatively broad pathological personality traits. Each item is either a simplified version of a PCL-R item or a combination and simplification of two PCL-R items. Item definitions are shorter than in PCL-R, averaging about 50 words. Total scores range from 0 to 24, with scores of 18 and higher used to diagnose psychopathy. Norms are available for reference groups of male or female correctional offenders, forensic psychiatric patients, civil psychiatric patients, and community residents. Scores can be calculated for two factors, isomorphic to the superordinate factors of PCL-R.

In forensic settings, PCL:SV can be used in conjunction with PCL-R as a screening test for psychopathy; most evaluators, however, tend to use PCL-R if the person being evaluated has a history of serious criminality (e.g., chronic or long-sentence offenders), and PCL:SV when the person does not (e.g., first-time, short-sentence, or less serious offenders; mentally disordered offenders; and forensic psychiatric patients).

Administration

PCL-R and PCL:SV are usually administered as part of a comprehensive psychodiagnostic evaluation. It takes about 20–30 min to score and interpret PCL-R, and about 10–15 min to score and interpret PCL:SV in these circumstances. Single-use assessment guides are available for both PCL-R and PCL:SV, which include a semistructured interview and space for recording relevant third-party information.

It is possible to administer the tests solely on the basis of third-party information if a person refuses or is otherwise unable to be interviewed, provided the quantity and quality of this information is sufficient. This procedure may, however, result in a score that is substantially lower than would have been obtained if the person had been interviewed.

It is not possible to administer PCL-R or PCL:SV without access to third-party information, except for certain research purposes.

Test User Qualifications

PCL-R and PCL : SV are controlled psychological tests. Independent use of the tests for clinical purposes is limited to people who are legally entitled to use psychological tests to assess and diagnose mental disorder. Test users also should have advanced education and training in individual assessment and psychological testing that would qualify them to practice as a mental health professional (e.g., graduate or medical degree). PCL-R manual also recommends

that test users should have at least 2 years experience working in forensic settings. Completion of training workshops and supervised practice in the use of PCL-R and PCL:SV is recommended, but not required. Similarly, the manuals recommend completion of 5–10 practice cases prior to clinical use of the tests.

Psychometric Properties

The psychometric properties of PCL-R and PCL:SV have been evaluated extensively within the framework of classical test theory [2, 3, 6]. The most relevant forms of reliability for the tests are structural, inter-rater, and test–retest reliabilities. Structural reliabilities are good to excellent: Item adequacy, as indexed by corrected item-total correlation, is typically 0.40–0.50; item homogeneity, as indexed by mean inter-item correlation, is typically 0.20–0.30; and internal consistency, as indexed by Cronbach's α, is typically 0.85–0.90. Inter-rater reliabilities also are good to excellent: For items, the intraclass correlation coefficient (ICC1) is typically between 0.60 and 0.80; for total scores, ICC1 is typically 0.80–0.90; and for categorical diagnoses, inter-rater agreement, as indexed by κ, is typically 0.50–0.75. Test–retest reliability of total scores and diagnoses has been examined infrequently, but appears to be good to excellent over periods of 1 week to 1 month and at least fair over periods of 6 months to 2 years.

More recently, the tests have been evaluated within the framework of item response theory (IRT) by Cooke, Hare, and colleagues [2, 7, 8]. The interpersonal, affective, and behavioral symptoms have good discriminating power; in contrast, the discriminating power of antisocial behavioral items is weaker. The interpersonal symptoms discriminate the latent trait at high levels (i.e., psychopathy), affective symptoms at moderate levels, and behavioral symptoms, at low levels.

IRT is also being used to examine potential metric bias in PCL-R and PCL : SV scores. Some researchers have reported evidence of a small but statistically significant metric bias across dominant cultures, with offenders and patients in the United Kingdom and other European countries scoring lower on PCL-R than those in Canada and the United States, given equivalent standing on the latent trait [9, 10]. This may be due to the effects of cultural facilitation, with symptoms of psychopathy being expressed more often in highly individualistic societies. There is no evidence, however, of metric bias between ethnic majority versus minority groups within a dominant culture [11]. Only recently have researchers started to investigate metric bias across gender and age.

Association Between PCL-R and PCL:SV

Cooke and colleagues have evaluated the derivation of PCL:SV using IRT methods [8]. They found a strong association between corresponding PCL-R and PCL:SV items or item pairs. They also found that PCL:SV items had a discriminating power as good as or better than the corresponding PCL-R items or item pairs.

Total scores on PCL-R and PCL:SV are also strongly associated. Cooke and colleagues [8] reported a high correlation ($r = 0.94$) between scores on the latent trait underlying both tests according to IRT analyses, and Guy and Douglas [6] reported similar high correlations ($r = 0.94–0.95$) between (raw) total scores on the tests in forensic samples. Guy and Douglas [6] also found that PCL:SV had good validity as a screening test for high-PCL-R scores in the same two forensic samples, with large areas under the curve (AUC = 0.98) according to receiver operating characteristic (ROC) analyses.

Factor Structure

Initial investigations of the dimensionality of PCL-R and PCL:SV used exploratory factor analysis (EFA) methods. The only consistent finding was the absence of a simple, unidimensional latent variable underlying the tests. Hare and colleagues proposed a structure comprising two orthogonal (i.e., correlated) factors: one reflecting interpersonal and affective symptoms, and the other reflecting behavioral symptoms and antisocial conduct [1].

Cooke and Michie [12] overcame the limitations of EFA by using confirmatory factor analysis (CFA) methods. After omitting PCL-R and PCL:SV items tapping antisocial conduct, they found strong evidence of a hierarchical three-factor structure, comprising a superordinate general factor (psychopathy) underpinned by three correlated subordinate factors (interpersonal, affective, and behavioral symptoms).

The hierarchical three-factor structure was replicable across diverse samples of offenders and forensic psychiatric patients from different nations. Importantly, it was even found when factor-analyzing diagnostic criteria for psychopathy other than PCL-R and PCL:SV.

Subsequent to Cooke and Michie's article, several researches have found support for the three-factor hierarchical structure. But others, including Hare and his colleagues, have argued that the entire item pools of PCL-R and PCL:SV should be included in factor analyses, which they say results in the addition of a fourth subordinate factor, reflecting antisocial conduct; this view was reflected in the updated PCL-R test manual [3]. Debate continues on whether antisocial conduct should be considered a primary symptom of psychopathy versus a secondary symptom, sequela, or consequence of the disorder [13].

Validity

As PCL-R and PCL:SV have been used in virtually hundreds of published studies, only a general review of research is presented here; and as the tests are strongly associated, they are reviewed together.

One important line of research has evaluated the *concurrent validity* of PCL-R and PCL:SV, that is, their association with other procedures for assessing psychopathy. Total scores on the tests have moderate to large correlations with clinical diagnoses made using other criteria, and moderate correlations with self-report measures of psychopathy [2, 3]. Both clinical diagnoses made using other criteria and self-report measures tend to correlate more highly with PCL-R and PCL:SV items reflecting behavioral symptoms and antisocial conduct than with items reflecting the interpersonal and affective features [2, 3].

With respect to *predictive validity*, total scores on PCL-R and PCL:SV are reliably associated with serious antisocial behavior, including violence, in both institutional and community settings. Summarizing a number of reviews in recent years [14, 15], several general conclusions may be drawn. First, the predictive validity of PCL-R or PCL:SV with respect to serious antisocial behavior is typically moderate in magnitude, $r = 0.20–0.30$. Second, the predictive validity tends to be larger in community settings than institutional settings. Third, the predictive validity tends to be larger when the outcome reflects general but serious antisocial conduct (e.g., "any violence"), rather than more specific or less serious antisocial conduct (e.g., "sexual violence" or "any misconduct"). Fourth, the predictive validity of psychopathy, as measured by PCL-R and PCL:SV, typically is higher than that of other established demographic, criminal history, and clinical risk factors (e.g., age, prior antisocial conduct, substance use); indeed, PCL-R and PCL:SV predict antisocial conduct about as well as do multifactor risk assessment procedures constructed theoretically or statistically. Fifth, the predictive validity of PCL-R and PCL:SV is not attributable solely or even primarily to the inclusion of items reflecting past antisocial conduct.

With respect to other aspects of *construct validity*, PCL-R and PCL:SV have been used to study the course, comorbidity, etiology, and treatment of psychopathy [16]. (*See also* **Psychopathy**.)

Forensic Applications

In the practice of forensic mental health, PCL-R and PCL:SV are used most often as part of comprehensive assessments of risk and treatability for sentencing, civil commitment, institutional classification, and release decision making. This application is supported by research on the prediction of serious antisocial conduct and on treatment response. Indeed, PCL-R and PCL:SV are incorporated explicitly into several procedures for assessing violence risk (*see* **Dangerousness: Risk of** and **Risk Assessment: Patient and Detainee**).

Forensic mental health professionals should be aware of some important limitations of PCL-R and PCL:SV [17, 18]. First, as observer ratings scales, the tests may be susceptible to distortion – unconscious or deliberate – by evaluators. To safeguard against this, evaluators should closely follow the administration instructions in the test manuals; training and supervised practice in the use of the tests may also be helpful. Second, although the tests have good psychometric properties in adult male offenders and forensic psychiatric patients in the United States and Canada, the possibility of bias due to culture and gender has not yet been ruled out. Third, the tests reflect the lifetime presence of psychopathic symptoms; this means they cannot be used to measure changes in the presence or severity of symptoms over time, or used to determine whether the person currently suffers

from psychopathy. Fourth, although comprehensive, PCL-R and PCL:SV are not exhaustive in content and relatively heavily saturated with items reflecting antisocial conduct. Evaluators should consider using PCL:SV instead of PCL-R in cases where the person being assessed does not have a serious history of criminality. Fifth, although psychopathy is a robust risk factor for antisocial behavior, PCL-R and PCL:SV scores cannot be used – either on their own or in combination with other factors – to estimate the specific probability or absolute likelihood that a given person will commit a criminal or violent act with any reasonable degree of scientific certainty.

References

[1] Hare, R.D. (1991). *Manual for the Hare Psychopathy Checklist – Revised*, Multi Health Systems, Toronto.

[2] Hare, R.D. (2003). *Manual for the Hare Psychopathy Checklist – Revised*, 2nd Edition, Multi Health Systems, Toronto.

[3] Hart, S.D., Cox, D.N. & Hare, R.D. (1995). *Manual for the Hare Psychopathy Checklist: Screening Version (PCL:SV)*, Multi-Health Systems, Toronto.

[4] Hare, R.D. (1996). Psychopathy: a clinical construct whose time has come, *Criminal Justice and Behavior* **23**, 25–54.

[5] Hare, R.D. (1980). A research scale for the assessment of psychopathy in criminal populations, *Personality and Individual Differences* **1**, 111–119.

[6] Guy, L.S. & Douglas, K.S. (2006). Examining the utility of the PCL:SV as a screening measure using competing factor models of psychopathy, *Psychological Assessment* **18**, 225–230.

[7] Cooke, D.J. & Michie, C. (1997). An item response theory analysis of the Hare psychopathy checklist-revised, *Psychological Assessment* **9**, 3–14.

[8] Cooke, D.J., Michie, C., Hart, S.D. & Hare, R.D. (1999). Evaluating the screening version of the Hare psychopathy checklist-revised (PCL:SV): an item response theory analysis, *Psychological Assessment* **11**, 3–13.

[9] Cooke, D.J., Michie, C., Hart, S.D. & Clark, D.A. (2005). Assessing psychopathy in the United Kingdom: Concerns about cross-cultural generalisability, *British Journal of Psychiatry* **186**, 339–345.

[10] Cooke, D.J., Michie, C., Hart, S.D. & Clark, D.A. (2005). Searching for the pan-cultural core of psychopathic personality disorder: Continental Europe and North America compared, *Personality and Individual Differences* **39**, 283–295.

[11] Cooke, D.J., Kosson, D.S. & Michie, C. (2001). Psychopathy and ethnicity: structural, item and test generalizability of the psychopathy checklist revised (PCL–R) in Caucasian and African-American participants, *Psychological Assessment* **13**, 531–542.

[12] Cooke, D.J. & Michie, C. (2001). Refining the construct of psychopathy: Towards a hierarchical model, *Psychological Assessment* **13**, 171–188.

[13] Cooke, D.J., Michie, C., Hart, S.D. & Clark, D.A. (2004). Reconstructing psychopathy: clarifying the significance of antisocial and socially deviant behavior in the diagnosis of psychopathic personality disorder, *Journal of Personality Disorders* **18**, 337–357.

[14] Hart, S.D. (1998). The role of psychopathy in assessing risk for violence: conceptual and methodological issues, *Legal and Criminological Psychology* **3**, 121–137.

[15] Douglas, K.S., Vincent, G.M. & Edens, J.F. (2006). Risk for criminal recidivism: the role of psychopathy, in *Handbook of Psychopathy*, C.J. Patrick, (ed), Guilford, New York, pp. 533–554.

[16] Patrick, C.J. (ed) (2006). *Handbook of Psychopathy*, Guilford, New York.

[17] Hare, R.D. (1998). The Hare PCL-R: some issues concerning its use and misuse, *Legal and Criminological Psychology* **3**, 99–119.

[18] Hemphill, J.F. & Hart, S.D. (2003). Forensic and clinical issues in the assessment of psychopathy, in *Comprehensive Handbook of Psychology: Vol. 11. Forensic Psychology*, I. Weiner (series editor) & A.M. Goldstein (volume editor), eds, John Wiley & Sons, New York, pp. 87–107.

Further Reading

Patrick, C.J. (ed) (2006). *Handbook of Psychopathy*, Guilford, New York.

Related Articles

Dangerousness: Risk of

Psychopathy

Risk Assessment: Patient and Detainee

STEPHEN D. HART

Psychopharmacology

Psychopharmacology is the study of the effects of drugs on psychological function. It is a branch of pharmacology, which focuses on the nervous system. Clinical psychopharmacology is the portion of psychiatric practice, which pertains to the use of medication as a tool to treat mental disorders [1, 2].

Clinical psychopharmacology involves an assessment process, development of a treatment plan, and subsequent assessment of the effects of treatment [3].

Psychopharmacologic testimony is often relevant in both civil and criminal cases. Establishing standard of care, and causation of damages in cases alleging negligent prescription, evaluating complex disability claims, and testamentary competency are some areas for psychopharmacological consultation. Cases of suspected poisoning or suicide may also warrant psychopharmacological consultation.

Since many crimes are committed under the influence of substances (*see* **Substance Abuse**), the effect of those substances on the defendant's mental state may be relevant (*see* **Insanity: Defense**; **Temporary Insanity**; **Mitigation Testimony**). Even though voluntary intoxication may not negate a mental state in many jurisdictions, the distinction between intoxication, mental illness, and "settled insanity" [4] requires psychiatric evaluation.

Defendants who are incompetent as a result of mental disorder (*see* **Capacity to Stand Trial**) may often be restored to competence through treatment. If the defendant refuses treatment, psychopharmacological evaluation is necessary to inform the court that medication is substantially likely to render the defendant competent to stand trial and substantially unlikely to have side effects that will interfere significantly with the defendant's ability to assist counsel in conducting a defense [5].

Basic principles of psychopharmacology are outlined in this article, along with an overview of psychotherapeutic agents, and some clinical and medical legal issues related to them. Substance abuse and dependence are reviewed in **Substance Abuse**; **Addictions**.

The study of pharmacology includes an understanding of pharmacokinetics, pharmacodynamics, and drug mechanisms. Pharmacokinetics describes what happens to a drug when it enters the body. How a drug is absorbed, metabolized, activated or inactivated, and excreted, influences how much of a drug is available at the site of action. Pharmacodynamics explains the effect of drugs on the sites of action [6].

Drug effects may vary as a result of the route of administration. For example, intravenous benzodiazepines can reliably cause short-term amnesia, which is useful to anesthesiologists. The same dose of drug ingested orally will usually not have the same effect.

Once a drug is absorbed, it enters the bloodstream to be carried through the body. Most psychoactive drugs are lipid (fat) soluble. They do not dissolve well in water. Thus, they are carried through the bloodstream bound to protein.

Many drugs are metabolized through interactions with enzymes in the liver (many are metabolized throughout the body). When a drug is metabolized, it is altered, often to an inactive or water soluble form. Sometimes the metabolites are themselves active drugs. If the drug is ingested by mouth, after it is absorbed in the intestines, it goes through the portal system to the liver for first pass metabolism, and then to the general circulation, where it can effect other organs.

Drugs that are inactivated to an efficient degree by the liver are not effective when administered orally. First pass inactivation can be bypassed through alternative routes of administration: intravenous, sublingual, or intranasal.

Drugs and their metabolites are excreted through urine, feces, and perspiration. The time required for the plasma concentration of a drug to fall by one half is the half-life of the drug ($t_{1/2}$). When the amount of drug administered in a given time equals the amount eliminated, the drug has reached steady state.

In order for a drug to act on the brain, it must cross the blood–brain barrier. Factors that affect a drug's ability to enter the brain include size and charge of the drug molecule, and its lipid solubility. For example, L-Dopa will cross the blood–brain barrier, while Dopamine (DA) will not.

A drug may have a nonspecific effect on the nervous system when it affects energy metabolism, or membrane stability. Specific drug effects arise from interaction with identifiable molecular mechanisms unique to target cells, which bear receptors for that drug. Sometimes a drug with specific effects at a low dose has general effects at a higher dose. Alcohol and general anesthetics are general depressants of the central nervous system, while drugs such as caffeine can be general stimulants [6].

Psychoactive drugs generally affect psychological function through interaction with the normal function of the nervous system [1, 6–8]. The brain contains billions of neurons supported by glia. Glia are considered supporting tissue for the nerve cells. Some form myelin, which facilitates axon function. Others are involved in metabolic activity, and the formation of the blood-brain barrier. Ongoing research suggests

that glia may, in fact, serve a more active role in brain function than previously believed [3].

The cell body contains the nucleus of the neuron. It is the center of metabolic activity. Dendrites extend and branch from it, and connect to the terminal processes of other neurons to form synapses. The axon is a single, often long, process, which arises out of the cell body to connect to other neurons.

The concentration of ions, like potassium, sodium, chloride, and calcium within the cell is not equal to the concentration outside the cell. As a result, the cell membrane carries a charge across it, like a microscopic battery. When a nerve fires, the membrane allows ions to cross, depolarizing it briefly. The depolarization travels down the length of the axon (an action potential). At the terminal process, packets of chemicals – neurotransmitters – are released across the membrane into the synapse, where the terminal process connects to the dendrite of another neuron.

The surface of the dendrite contains special proteins called a *postsynaptic receptor*. The axon may contain presynaptic receptors as well. Each receptor is specialized to react to a specific neurotransmitter. The effect may be to excite the neuron or to make excitation more difficult – to inhibit it (depending on the neurotransmitter and receptor). When the sum of excitatory minus inhibitory stimulation exceeds a threshold, the nerve will fire.

The system is regulated through feedback mechanisms mediated by (presynaptic) autoreceptors and secondary messengers. When stimulated, autoreceptors on the presynaptic neuron signal the cell to stop release of neurotransmitter. Secondary messengers within the postsynaptic neuron modulate metabolic processes in the cell, including protein production. This controls neural plasticity, including long-term potentiation, long-term depression, down regulation and up regulation of receptors.

Neurotransmitter molecules are removed from the synapse by the presynaptic neuron (which released them) through a reuptake process. Once removed, the neurotransmitter does not interact with the receptor, until it is released again. Some neurotransmitter is inactivated through enzyme-assisted metabolic degradation.

The synapse is the site of action of most psychoactive drugs.

Neurons are organized in circuits, with specialized function. The systems that are most interesting to the psychiatrist are the limbic system, the basal ganglia, and the thalamocortical system.

The limbic system is involved with the experience and expression of emotion. Basal ganglia are involved in both motor and cognitive function. The thalamocortical system is involved in sensation, movement, and cognition.

Drugs of abuse affect DA circuits in the nucleus acumbens, an area of the brain involved in learning and motivation. The reinforcing effect of such stimulation is difficult to overcome, as reflected in the rates of lapses and relapses in recovering addicts (*see* **Addictions**).

A non-comprehensive list of neurotransmitters follows:

- Dopamine (DA);
- serotonin (5-HT);
- acetylcholine (ACh);
- norepinephrine (NE);
- gamma aminobutyric acid (GABA);
- glutamate (Glu);
- endorphins;
- enkephalins;
- histamine; and
- aspartate.

Therapeutic Classes of Drugs

Drugs are often classified in terms that identify the target syndrome or symptoms for which the particular drug was first marketed. Thus, we have antipsychotic drugs, antidepressants, sedative hypnotics, stimulants, mood stabilizers, and cognitive enhancers. Over the course of time, however, drugs are often used effectively for purposes other than those for which they were initially marketed. Some antipsychotic drugs can be used for anxiety, depression, and mood stabilization. Some antidepressants are used for first line treatment of anxiety disorders. Other classification schemes based upon chemical structure or target chemical systems are sometimes employed by practitioners. Thus, various drugs may be described as benzodiazepine, or tricyclic, or selective serotonin reuptake inhibitor (SSRI).

Antipsychotic Medication

The practice of psychiatry was revolutionized in the 1950s with the introduction of chlorpromazine, the

first effective antipsychotic drug. Its use reduced the need for restraints, seclusion, and locked psychiatric facilities [6]. As the only effective treatment for psychosis at the time, it was accepted rapidly, despite its side effects.

The success of chlorpromazine stimulated the development of other phenothiazine drugs: thioridazine, fluphenazine, perphenazine, and trifluoperazine. In attempts to increase potency while reducing side effects, the pharmaceutical industry developed butyrophenones (haloperidol), thioxanthenes (thiothixene), dihydroindolones (molindone), and dibenzoxazepines (loxapine). By the 1980s, clozapine was recognized as an effective treatment for psychotic illness, which had not responded to other medications. Its significant toxicity (agranulocytosis) necessitated weekly blood counts, and precluded its use as a first line treatment in the United States. However, its atypical neurochemical profile, its lack of movement side effects, and its affect on negative symptoms of schizophrenia, suggested significant advantages over older line medications. Olanzapine, risperidone, quetiapine, ziprasidone, and aripiprazole were developed in attempts to provide clozapine's therapeutic advantages without potentiating agranulocytosis. These second-generation antipsychotics, also known as *atypical antipsychotics* are now used more frequently than neuroleptics for treatment of chronic psychosis, as their side effects are generally milder and more easily tolerated by patients.

Antipsychotic medication is indicated for treatment of schizophrenia, schizoaffective disorder, schizophreniform disorder, or psychotic symptomatology caused by medical disorders.

Symptoms of schizophrenia can be divided into positive and negative symptoms. Atypical antipsychotics are more effective than first-generation drugs for negative symptoms, which include flat or blunted affect, inactivity, diminished pleasure in activities, and poverty of thought. First-generation drugs primarily affect positive symptoms, including delusions, hallucinations, confusion, and anxiety [9–12].

Blockade of some DA receptors (D_2) is associated with both the antipsychotic effect and movement side effects of these drugs. (In fact, the first psychiatrists to use chlorpromazine did not expect to see antipsychotic effects without an extrapyramidal syndrome.) The blockade of some serotonin receptors (5-HT_2) is associated with relief of negative symptoms, as well as mitigation of movement side effects [1, 6, 7].

Side effects of antipsychotic medication include sedation, orthostatic hypotension, dry mouth, constipation, blurry vision, urinary hesitancy, extrapyramidal effects, weight gain, and metabolic effects.

First-generation antipsychotics were differentiated from each other on the basis of their side effect profiles, rather than their efficacy. These side effects, related to the drugs' effects on cholinergic, adrenergic, and histamine receptors [1] were understood in relation to differences in the structures of the side chains of these molecules.

Anticholinergic effects include dry mouth, constipation, blurry vision, confusion, and urinary hesitancy.

Adrenergic blockade may cause changes in blood pressure, and cardiac rhythm, as well as symptoms of dizziness.

Histamine blockade is associated with sedation, drowsiness, and weight gain.

Extrapyramidal side effects are caused by DA blockade. DA and ACh are neurotransmitters of the extrapyramidal motor system, involved in posture and coordinated movement. Excessive blockade of DA receptors in this system may lead to an imbalance, and one of four extrapyramidal syndromes: Akathisia, dystonia, Parkinson's syndrome, or tardive dyskinesia.

Akathisia is an uncontrolled sense of inner restlessness, which must be distinguished from anxiety. If it is mistaken for anxiety, and treated with increased dose of medication, it will be worsened.

Dystonia is spasm of the muscles, usually of the head and neck.

Parkinson's syndrome includes muscular rigidity, tremor, slowed motor responses, and diminished facial expression.

Tardive dyskinesia is an often irreversible effect of antipsychotic medication characterized by involuntary movements of the mouth and tongue, as well as of the trunk and extremities.

Akathisia, dystonia, and Parkinson's syndrome may be treated as they emerge through the use of anticholinergic agents, minor tranquilizers, or adjustment in dose of antipsychotic medication.

While various drugs have been used to diminish symptoms of tardive dyskinesia, there is no cure. Discontinuation of the antipsychotic medication will

lead to initial worsening of dyskinesia. Over time, symptoms often will remit.

Atypical antipsychotic drugs are less frequently associated with tardive dyskinesia, or other extrapyramidal side effects, than older first-generation antipsychotics.

Significant weight gain is a common problem with antipsychotic drugs, and appears to be associated with an increased risk of diabetes. Aripiprazole and ziprasidone are less likely than other agents to be implicated in weight gain.

Neuroleptic malignant syndrome is a rare complication of antipsychotic drug use. It is characterized by severe muscular rigidity, fever, increased blood pressure, increased heart and respiratory rate, and changing levels of consciousness. Laboratory testing show elevations in creatine phosphokinase (CPK), sometimes along with altered liver functions, and myoglobin (a muscle protein) in the blood or urine. Immediate treatment with supportive symptomatic measures and discontinuation of antipsychotic medication is necessary when this diagnosis is made. Treatment with DA agonists as well as benzodiazepines should be considered with high fevers.

Antidepressants

Iproniazid was an antituberculous agent developed after World War II. After its mood-elevating qualities were noticed, it was investigated, and then marketed as an antidepressant in the late 1950s. The therapeutic action was a result of inhibition of monoamine oxidase (MAO) enzymes, which break down serotonin and NE.

Iproniazid was withdrawn because of toxic effects on the liver. Other MAO Inhibitors were developed, but their use was limited because of the need for dietary restrictions to avoid potentially fatal hypertensive reactions. Recently a transdermal form of selegiline was developed, allowing fewer dietary restrictions.

Around the same time that iproniazid was developed, phenothiazine-like drugs were being tested for antipsychotic effects, with the hope that new drugs could be found without the side effects of chlorpromazine. One, imipramine, was relatively ineffective in calming psychotic patients. However, it did have a therapeutic effect on depressed patients, especially those characterized by regression and inactivity. It was marketed as the first tricyclic antidepressant.

Over the following years, several other tricyclic antidepressants were developed: desipramine, amitriptyline, nortriptyline, clomipramine, trimipramine, doxepin, protriptyline, amoxapine, and maprotiline (a tetracyclic). All block the reuptake of NE into nerve terminals. Imipramine, amitriptyline, doxepin, nortriptyline, and especially clomipramine, also block reuptake of serotonin (5-HT).

These drugs are all pharmacologically "dirty". In addition to their desired effects, they bind to a host of other receptors, with resultant unwanted side effects, including faintness, cardiac arrhythmias, constipation, dry mouth, sedation, and weight gain.

SSRIs or SRIs represent an important pharmacological advance in the treatment of depression, as pharmacologically cleaner drugs, their safety, and tolerability have led to their becoming the predominant class of prescribed antidepressants.

Venlafaxine and duloxetine, serotonin and norepinephrine reuptake inhibitors (SNRIs) may be more useful than SSRIs in treating chronic pain.

Bupropion effects DA and NE reuptake. Trazodone and nefazodone block a portion of the serotonin receptors. Mirtazapine blocks some alpha-adrenergic and serotonergic receptors. These drugs are less likely to cause sexual side effects than other antidepressants.

All antidepressants are indicated for treatment of major depression, including vegetative symptoms of appetite and sleep disturbance, fatigue, diminished sex drive, anhedonia (loss of the ability to experience pleasure), agitation, restlessness, or psychomotor retardation.

In addition, many are effective for generalized anxiety disorder, panic disorder, posttraumatic stress disorder, insomnia, and enuresis.

Tricyclic antidepressants and SNRIs are effective in chronic pain conditions. SSRIs and clomipramine are effective in obsessive compulsive disorder [13–15].

Bupropion can reduce craving for cigarettes, probably as a result of its dopaminergic effect on the nucleus accumbens.

The mechanism of action of antidepressants is not fully understood. While antidepressant drugs exert immediate effects on brain receptors, the therapeutic effect is delayed, usually 2–6 weeks. This suggests that the therapeutic effects arise from a cascade of adaptive processes to repeated administration of the drug. Increased availability of serotonin and/or NE

in the synapse evokes negative feedback mechanisms to restore homeostasis. This includes downregulation and desensitization of receptors, and alterations in tonic inhibition and stimulation mediated through secondary messengers in the neuron, which signal formation of new protein.

Newer antidepressants present fewer, less toxic side effects than older tricyclics and MAO Inhibitors. As a result, patients are more likely to comply with treatment long enough to obtain a therapeutic effect than those prescribed tricyclic antidepressants. Patient acceptability and relative safety in overdose have contributed to the popularity of these drugs.

Tricyclic antidepressants cause side effects related to their effects on the autonomic nervous system. These include dry mouth, dizziness, palpitations, blurry vision, constipation, tachycardia, orthostatic hypotension, and cardiac arrhythmias.

MAO Inhibitors may cause postural hypotension. Drug and diet interactions may precipitate hypertensive crisis or serotonin syndrome, which includes restlessness, muscle twitching, sweating, shivering, and tremor [16].

SRI and SNRI side effects include nausea, headache, delayed ejaculation, and impaired orgasm. Bupropion can cause anorexia, insomnia, and agitation. Trazodone can cause priapism. Mirtazapine is sedating and can cause weight gain.

Patients with bipolar disorder treated with antidepressants may be at risk of switching from depression to a hypomanic or manic state [17–19].

Mood Stabilizers

Cade first described the therapeutic effects of lithium on mania in 1949. It was first approved by the US FDA for treatment of acute mania in 1970s, and for prophylaxis of bipolar disorder in 1974.

While lithium is quite toxic, it is very effective for treatment of both manic and depressive phases of bipolar disorder. Several large-scale studies have demonstrated significant reduction in suicide risk associated with lithium therapy [20, 21]. There is an increase in suicide risk after discontinuation of lithium [22].

Adverse effects include acne, leucocytosis, hypothyroidism, hypoparathyroidism, nephrogenic diabetes insipidus, and kidney damage. Signs of toxicity include tremor, weakness, fatigue, nausea, vomiting, cardiac arrhythmias, hypotension, shock, stupor, coma, and even death.

Safe use of lithium requires monitoring of blood levels, since the therapeutic dose is close to the toxic dose. Evaluation of serum electrolytes, thyroid, parathyroid, renal, and hematological function is necessary at the commencement of treatment and periodically during the course of therapy.

Valproate was approved by the FDA for treatment of acute mania in 1994. It is less toxic, and easier to prescribe than lithium. It consequently has been prescribed more frequently than lithium, even though the evidence of its effectiveness for prophylaxis is not robust.

In 2003, the FDA approved the use of lamotrigine for long-term treatment of bipolar disorder. This anticonvulsant demonstrates little effect on acute mania. However, it shows antidepressant effect and prophylactic effects in bipolar patients. Other anticonvulsants, which are used as mood stabilizers, include carbamazepine, gabapentin, pregabalin, topiramate, tiagabine, oxcarbazepine, and zonisamide.

Antipsychotic drugs such as olanzapine, aripiprazole, and quetiapine are also termed *mood stabilizers* because of their demonstrated roles in treatment of bipolar disorder [23].

Anxiolytics

Chlordiazepoxide and diazepam were the first benzodiazepines, introduced in the early 1960s. They were soon widely prescribed by physicians for pathological anxiety, because they were effective and safer than meprobamate and barbiturates. Other benzodiazepines were subsequently developed and marketed, including clorazepate, alprazolam, oxazepam, and clonazepam.

All benzodiazepines exhibit anxiolytic, sedative hypnotic, muscle relaxant, and anticonvulsant effects. The primary differences among them are a result of pharmacokinetic properties, including rates of absorption and elimination.

Benzodiazepines bind to specific receptors in the brain – the GABA – BZ complex. GABA is a primary inhibitory neurotransmitter in the brain. When it is bound to its receptor, a chloride channel on the membrane is opened a little, making depolarization of the nerve more difficult. GABA works more effectively in the presence of a benzodiazepine.

Sedation and drowsiness are common side effects of benzodiazepines. Motor impairment may occur, as may memory impairment. Tolerance may occur, and physical dependence is common with prolonged treatment. Abuse and dependence (*see* **Substance Abuse**; **Addictions**), as defined by Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV) [24], are less common, but may still be problematic.

Buspirone is a unique nonsedating nonbenzodiazepine anxiolytic drug, which appears to exert its effect through actions on a subset of serotonergic receptors (5-HT_{1A}). It is effective for generalized anxiety disorder, but not for panic attacks. Its effect builds up slowly, over several weeks. Side effects may include dizziness, headaches, and nausea.

Addition of buspirone to an antidepressant can be beneficial for patients with inadequate or poor responses to an initial trial of antidepressant treatment [25–28].

Stimulants

Amphetamines and methylphenidate stimulate the release of DA and NE at the synapse. They are useful for the treatment of attention deficit disorder [29–32] (*see* **Psychopharmacology: Child and Adolescent**), and have been used for appetite suppression, fatigue, and treatment of poststroke depression.

Side effects include nervousness, insomnia, and agitation. Arrhythmias and seizures may occur. Hallucinations and delusions are uncommon with stimulants administered orally at usually prescribed doses. They are a frequent complication of inhaled or injected amphetamines.

Stimulant psychosis, including visual hallucinations, disorientation, agitation, pacing, and thought disorder, may persist for months after cessation of amphetamine use.

Modafinil is similar to stimulants, in that it increases wakefulness. It does not increase reinforcement, like other stimulants. Thus it is, in a class by itself, a "wakefulness promoting agent". It is indicated for narcolepsy, and the fatigue associated with shift work and sleep apnea.

Drug Interactions

Concurrent use of more than one drug, or a drug with other substances, may lead to altered pharmacological effect – a drug interaction. Pharmacokinetic interactions occur when an agent alters the absorption, distribution, or metabolism of a drug. Pharmacodynamic interaction involves changes at the receptor, or the biologically active site.

The absorption of buspirone and ziprasidone are enhanced in the presence of food. Antacids may decrease absorption of some antibiotics. Drugs that are bound strongly to protein, like fluoxetine, will displace other protein-bound drugs, possibly enhancing their action or leading to toxic effects. Excretion of lithium is diminished with certain diuretic drugs, leading to increased serum concentration. Grapefruit juice may inhibit Cytochrome P450 enzymes, which break down imipramine, leading to increased serum concentration. These are all pharmacokinetic interactions.

Many drugs are metabolized by the Cytochrome P450 family of enzymes, predominantly found in the liver, the gut, and the brain. The activity of these enzymes, in turn, is enhanced or inhibited by many drugs. Many significant pharmacokinetic drug interactions are mediated by changes in these enzymes [33]. Much variability in individual sensitivity to therapeutic agents can be explained by variations in the concentrations and activities of these enzymes.

Pharmacodynamic interactions can cause increased or decreased pharmacological effect. For example, alcohol can potentiate the sedative effect of hypnotic drugs. Lithium can potentiate the effect of antidepressant drugs. MAO Inhibitors can provoke a serotonin syndrome when administered with an SSRI.

Attention to the potential for drug interactions allows a psychopharmacologist to safely and effectively provide rational treatment.

Drug Testing

For most psychotherapeutic drugs, the relationship between serum concentration and clinical effect has not been established. Lithium, valproate, carbamazepine, and nortriptyline are important exceptions. The difference between the therapeutic dose and a toxic dose of lithium is small. Periodic measurement of serum concentration is necessary to safely prescribe it. The interactions of carbamazepine and valproate with the Cytochrome P450 enzymes necessitate blood testing to establish an appropriate therapeutic dose. Sometimes physicians will check on the

concentration of other prescribed drugs when toxicity is suspected. Sometimes less expensive qualitative testing can confirm or refute suspicions of noncompliance. Generally, however, adjustment of doses of psychotropic medication is made on clinical grounds.

Qualitative testing for the presence of drugs in blood, urine, saliva, or hair samples is often part of a treatment program for patients with substance use disorders (*see* **Drug Testing: Urine**; **Amphetamine**; **Benzodiazepines**; **Cannabis**; **Cocaine**; **Opioids**). Serum concentration is often presented to the forensic psychopharmacology consultant for interpretation (**Toxicology: Forensic Applications of**).

Substances associated with dependence tend to evoke tolerance over time. Thus, without additional data, one cannot usually determine from a single sample of body fluid, the level of intoxication experienced by a user. Was the blood level the result of one use or many over time? How was the drug ingested? How long ago? Were blood levels on their way up, or down? Or was this steady state?

Concentrations of drugs obtained during necropsy are often quite different than those obtained during life (Postmortem blood has been described as a fluid resembling blood that is obtained from the vasculature after death [34]). A drug that has been concentrated in solid organs and tissues will diffuse into the blood after death, increasing its concentration. This was demonstrated in the case of a man who committed suicide by ingesting an overdose of imipramine. Concentrations of imipramine and its active metabolite, desipramine, obtained during postmortem examination 7 h after death were compared with concentrations in blood samples obtained in the emergency room 2 h prior to death. Postmortem imipramine concentrations ranged from 1.8 to 7.9 times the concentration of the emergency room sample (depending on the site in the body from which the blood was drawn), while desipramine ratios ranged from 1.6 to 6 [35].

The Volume of distribution (V_d) is a hypothetical volume of body fluid that would be necessary if the total amount of drug in the body were distributed uniformly in the same concentration as in the plasma. In general, lipophilic drugs have a high volume of distribution. Caffeine, which dissolves equally in water and in fat, has a V_d of $1\,l\,kg^{-1}$. For imipramine, it is $11–16\,l\,kg^{-1}$. For sertraline, it is in the range of 50–80. Drugs with a high volume of distribution will show the greatest change in concentration postmortem (*see* **Toxicology: Analysis**).

References

[1] Schatzberg, A.F. & Nemeroff, C.B. (2004). *The American Psychiatric Press Textbook of Psychopharmacology*, 3rd Edition, American Psychiatric Press, Washington, DC.

[2] Nemeroff, C.B., Heim, C.M., Thase, M.E., Klein, D.N., Rush, A.J., Schatzberg, A.F., Ninan, P.T., McCullough, J.P., Weiss Jr, P.M., Dunner, D.L., Rothbaum, B.O., Kornstein, S., Keitner, G & Keller M.B (2003). Differential responses to psychotherapy versus pharmacotherapy in patients with chronic forms of major depression and childhood trauma, *Proceedings of the National Academy of Sciences of the United States of America* **100**(24), 14293–14296.

[3] Sadock, B.J. & Sadock, V.A. (2004). *Kaplan and Sadock's Comprehensive Textbook of Psychiatry*, 8th Edition, Lippincott Williams & Wilkins.

[4] Feix, J. & Wolber, G. (2007). Intoxication and settled insanity: a finding of not guilty by reason of insanity, *The Journal of the American Academy of Psychiatry and the Law* **35**(2), 172–182.

[5] Sell v. United States, 539 U.S. 166 (2003).

[6] Brunton, L., Lazo, J. (2005). *Goodman and Gilman's the Pharmacological Basis of Therapeutics*, 11th Edition, McGraw Hill.

[7] Seeman, P. (2004). Atypical antipsychotics: mechanism of action, *Focus* **2**(1), 48–58.

[8] Nemeroff, C.B. (1998). Psychopharmacology of affective disorders in the 21st century, *Biological Psychiatry* **44**(7), 517–525.

[9] Keefe, R.S.E., Silva, S.G., Perkins, D.O. & Lieberman, J.A. (1999). The effects of atypical antipsychotic drugs on neurocognitive impairment in schizophrenia: a review and meta-analysis, *Schizophrenia Bulletin* **25**(2), 201–222.

[10] Lieberman, J.A., Stroup, T.S., McEvoy, J.P., Swartz, M.S., Rosenheck, R.A., Perkins, D.O., Keefe, R.S.E., Davis, S.M., Davis, C.E., Lebowitz, B.D., Severe, J & Hsiao, J.K. The Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) Investigators, (2005). Effectiveness of antipsychotic drugs in patients with chronic schizophrenia, *New England Journal of Medicine* **353**(12), 1209–1223.

[11] Rosenheck, R., Cramer, J., Xu, W., Thomas, J., Henderson, W., Frisman, L., Fye, C. & Charney, D. The Department of Veterans Affairs Cooperative Study Group on Clozapine in Refractory Schizophrenia (1997). A comparison of clozapine and haloperidol in hospitalized patients with refractory schizophrenia, *New England Journal of Medicine* **337**(12), 809–815.

[12] Volavka, J., Czobor, P., Sheitman, B., Lindenmayer, J.-P., Citrome, L., McEvoy, J.P., Cooper, T.B., Chakos,

M. & Lieberman, J.A., (2004). Clozapine, olanzapine, risperidone, and haloperidol in the treatment of patients with chronic schizophrenia and schizoaffective disorder, *Focus* **2**(1), 59–67.

[13] Greist, J.H., Jefferson, J.W., Kobak, K.A., Katzelnick, D.J. & Serlin, R.C. (1995). Efficacy and tolerability of serotonin transport inhibitors in obsessive-compulsive disorder. A meta-analysis, *Archives of General Psychiatry* **52**(1), 53–60.

[14] Kaplan, A. & Hollander, E. (2004). A review of pharmacologic treatments for obsessive-compulsive disorder, *Focus* **2**(3), 454–461.

[15] Kobak, K.A., Greist, J.H., Jefferson, J.W., Katzelnick, D.J. & Henk, H.J. (2004). Behavioral versus pharmacological treatments of obsessive compulsive disorder: a meta-analysis, *Focus* **2**(3), 462–474.

[16] Looper, K.J. (2007). Potential medical and surgical complications of serotonergic antidepressant medications, *Psychosomatics* **48**(1), 1–9.

[17] Altshuler, L.L., Suppes, T., Black, D.O., Nolen, W.A., Leverich, G., Keck Jr, P.E., Frye, M.A., Kupka, R, McElroy, S.L., Grunze, H., Kitchen, C.M.R. & Post, R. (2007). Lower switch rate in depressed patients with bipolar II than bipolar I disorder treated adjunctively with second-generation antidepressants. *Focus* **5**(1), 107–110.

[18] Judd, L.L., Akiskal, H.S., Schettler, P.J., Endicott, J., Maser, J., Solomon, D.A., Leon, A.C., Rice, J.A. & Keller, M.B. (2002). The long-term natural history of the weekly symptomatic status of bipolar I disorder, *Archives of General Psychiatry* **59**(6), 530–537.

[19] Leverich, G.S., Altshuler, L.L., Frye, M.A., Suppes, T., McElroy, S.L., Keck Jr, P.E., Kupka, R.W., Denicoff, K.D., Nolen, W.A., Grunze, H., Martinez, M.I. & Post, R.M. (2006). Risk of switch in mood polarity to hypomania or mania in patients with bipolar depression during acute and continuation trials of venlafaxine, sertraline, and bupropion as adjuncts to mood stabilizers. *American Journal of Psychiatry* **163**(2), 232–239.

[20] Tondo, L., Jamison, K. & Baldessarini, R. (1997). Effect of lithium maintenance on suicidal behavior in major mood disorders, *Annals of the New York Academy of Sciences* **836**, 339–351.

[21] Tondo, L., Hennen, J. & Baldessarini, R. (2001). Lower suicide risk with long-term lithium treatment in major affective illness: a meta-analysis, *Acta Psychiatrica Scandinavica* **104**, 163–172.

[22] Baldessarini, R., Tondo, L. & Viguera, A. (1999). Discontinuing lithium maintenance treatment in bipolar disorders: risks and implications, *Bipolar Disorders* **1**, 17–24.

[23] Bauer, M.S. & Mitchner, L. (2004). What Is a "mood stabilizer"? an evidence-based response, *American Journal of Psychiatry* **161**(1), 3–18.

[24] American Psychiatric Association (1994). *The Diagnostic and Statistical Manual of Mental Disorders, (DSM IV)*, 4th Edition, American Psychiatric Association, Washington, DC.

[25] Thase, M.E., Friedman, E.S., Biggs, M.M., Wisniewski, S.R., Trivedi, M.H., Luther, J.F., Fava, M., Nierenberg, A.A., McGrath, P.J., Warden, D., Niederehe, G., Hollon, S.D. & Rush, A.J. (2007). Cognitive therapy versus medication in augmentation and switch strategies as second-step treatments: a STAR*D report, *The American Journal of Psychiatry* **164**(5), 739–752.

[26] Fava, M., Rush, A.J., Wisniewski, S.R., Nierenberg, A.A., Alpert, J.E., McGrath, P.J., Thase, M.E., Warden, D., Biggs, M., Luther, J.F., Niederehe, G., Ritz, L. & Trivedi, M.H. STAR*D StudyTeam, (2006). A comparison of mirtazapine and nortriptyline following two consecutive failed medication treatments for depressed outpatients: a STAR*D report, *The American Journal of Psychiatry* **163**(7), 1161–1172.

[27] Nierenberg, A.A., Fava, M., Trivedi, M.H., Wisniewski, S.R., Thase, M.E., McGrath, P.J., Alpert, J.E., Warden, D., Luther, J.F., Niederehe, G., Lebowitz, B., Shores-Wilson, K. & Rush, A.J. STAR*D Study Team, (2006). A comparison of lithium and T3 augmentation following two failed medication treatments for depression: a STAR*D report, *The American Journal of Psychiatry* **163**(9), 1519–1530.

[28] Rush, A.J., Trivedi, M.H., Wisniewski, S.R., Nierenberg, A.A., Stewart, J.W., Warden, D., Niederehe, G., Thase, M.E., Lavori, P.W., Lebowitz, B.D., McGrath, P.J., Rosenbaum, J.F., Sackeim, H.A., Kupfer, D.J., Luther, J. & Fava, M. (2006). Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report, *The American Journal of Psychiatry* **163**(11), 1905–1917.

[29] The MTA Cooperative Group (1999). Moderators and mediators of treatment response for children with attention-deficit/ hyperactivity disorder: the Multimodal Treatment Study of Children with attention-deficit/hyperactivity disorder, *Archives of General Psychiatry* **56**(12), 1088–1096.

[30] The MTA Cooperative Group (1999). A 14-month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder, *Archives of General Psychiatry* **56**(12), 1073–1086.

[31] Spencer, T., Wilens, T., Biederman, J., Faraone, S.V., Ablon, J.S. & Lapey, K. (1995). A double-blind, crossover comparison of methylphenidate and placebo in adults with childhood-onset attention-deficit hyperactivity disorder, *Archives of General Psychiatry* **52**(6), 434–443.

[32] Goldman L.S., Genel M., Bezman R.J., Slanetz P.J., for the Council on Scientific Affairs AMA (1998). Diagnosis and treatment of attention-deficit/hyperactivity disorder in children and adolescents, *The Journal of the American Medical Association* **279**(14), 1100–1107.

[33] Nemeroff, C., DeVane, C. & Pollock, B. (1996). Newer antidepressants and the cytochrome P450 system, *The American Journal of Psychiatry* **153**(3), 311–320.

[34] Klaasen, C.D. (2001). *Casarett & Doull's Toxicology: The Basic Science of Poisons*, 6th Edition, McGraw-Hill, New York.

[35] Pounder, D. & Jones, G. (1990). Postmortem drug redistribution – A toxicological nightmare, *Forensic Science International* **45**, 253–263.

SAMUEL I. MILES

Psychopharmacology: Child and Adolescent

Psychopharmacology's Role in Treatment of Child and Adolescent Mental and Behavioral Disorders

Psychopharmacology is the study of drug–behavior relationships and the use of medications to influence affective and emotional states and thoughts. In children and adolescents, psychotherapeutic medications are best used in conjunction with a holistic bio-psycho-social approach to identify youth's problems and at times in combination with specific nonmedication-based therapies [1]. Transactions between individual genetics, environmental, family and social stresses, and protective factors are currently considered integral to the development of or protection from psychiatric illness. The use of psychotherapeutic medication, when based on scientific evidence, has an important role in fostering improved adaptation in the face of an individual's lowered threshold toward disease. Psychotherapeutic medications are not intended to affect a cure but rather to help by reducing problematic behaviors or relieving mentally painful symptoms.

If a biologic predisposition to a mental disorder lowers the threshold at which a disease becomes evident, then understanding the nature of the risk exposure, while not always immediately evident in the chain of causality, is important. Examples include how the toxic effects of lead and mercury on the development of human nervous system lead to disease states, learning disabilities, and behavioral disorders. Likewise, fetal alcohol effects are discussed as public policy issues, while families, pediatricians, child and adolescent psychiatrists, teachers and the juvenile, and adult justice system attempt to mitigate the damage at a more individual level.

Our knowledge of the pernicious effects of multigenerational cycles of physical and mental maltreatment in childhood, leading to states of affective dyscontrol beginning in childhood and continuing throughout the life span, has expanded to include neuroanatomical and biochemical proof of the disordered states. Addictions are diseases compounded by genetic predispositions that affect motivational circuitry in the brain. Interrupting the cascading effects of untreated disease on the individual, families, and larger society is an important goal. Medication is a component in treatment. There are multiple textbooks on this subject and this article addresses only a few of the issues [2–4].

The Prescriber, Privacy, and the Court

Subpoenas to courts do not provide the authority to release confidential health information that is protected by privacy laws. The prescriber needs to assure that the person who controls protected information has consented to its release or that there has been a judicial determination that privilege does not apply. If uncertain and questioned in court, the prescriber should indicate that the information is privileged and follow the direction of the judge.

Tort Law and the Standard of Care

The use of the word prescriber rather than physician is intended to reflect the fact that many jurisdictions allow nonphysicians such as nurse practitioners to prescribe medications. The tort of negligence has been described as "conduct that falls below the standard regarded as normal or desirable in a given community for those who are perceived to be competent in carrying out their profession within the standards of reasonable skill and proficiency" [5].

In prescribing the medications for psychiatric illness, liability for malpractice does not ensue necessarily from a bad outcome but rather from deviation from the accepted use of the medication. In most jurisdictions, the standard of care in assessing negligence is a matter of medical judgment. In making this decision, rarely does the law provide an answer but rather testimony of experts, articles in learned journals and textbooks, guidelines from professional organizations, etc. will be used to advise the decision maker.

Lawsuits against prescribers of psychotherapeutic medicine often make claims of negligence in either failing to obtain an informed consent for or in the prescribing, administering, and/or monitoring of medications.

Informed Consent for Psychotheraputic Medications in Children and Adolescents

Treatment with psychotherapeutic medication requires informed consent. Applebaum and Gautheil state "Treatment without any consent or over a patient's objections may constitute a battery, but treatment after an inadequate consent is properly considered as a form of malpractice" [6].

A free person is considered to have a right to control what happens to his or her body. Most jurisdictions give the individual patient and not the prescriber of treatment the right to balance risk *versus* benefit of the procedure and to consent to it or refuse it.

Informed consent involves two essential parts: a document and a process. It is important to document it in the medical record but it is more than a signed piece of paper. Ongoing explanations help the patient make wise decisions about whether it is essential to begin or continue taking a particular treatment or medication.

The three elements of informed consent are as follows:

1. the mental *competency* to make a rational decision;
2. the *voluntary* nature of the decision; and
3. *sufficient information* to make a decision.

In children and adolescents, the issue of competence and substitute decision making are especially involved.

Competency

Statutory rules determine when the age of competence to seek or refuse psychiatric treatment occurs. This often occurs before the age of majority. Emancipated minors have full competence. Statutes and age of consent vary from one jurisdiction to another. Informed assent of the older child or adolescent helps the therapeutic relationship and improves compliance. Research trials are governed by more stringent guidelines in this regard. In the United States, the National Commission for Protection of Human Subjects of Biomedical and Behavioral Research has established age 7 as a reasonable minimum age for children involving in some kind of assent or dissent process and "parental permission" rather than "proxy consent" is considered the norm [7].

Sometimes clarification from a court as to who has authority for decision making for a child may be necessary. Divorced or separated parents with shared decision-making powers might disagree with one another's opinion. Substitute nonparent decision makers who are actively raising the child (grandparents, foster parents, social service agency workers, etc.) may not have the medical decision-making authority to consent.

Voluntary Decision Making

Regarding the issue of substitute decision making, one should consider the agency of the prescriber. Children and adolescents rarely seek psychiatric treatment on their own and are usually dependant upon their parents who act as decision makers. Because the parents are so much a part of the child's decision to be in treatment to some extent, the prescriber should be aware when he or she is acting in their agency as well as the child's. Other situations of potential for dual agency exist when clinicians are contracted to or are employed by schools or social agencies or treatment facilities. The prescriber should assure that school authorities or others are not coercing decision maker to medicate the child.

Information to Make the Decision

To make a submissible case based on negligence in obtaining informed consent, a plaintiff must show nondisclosure, causation, and injury.[a] To show nondisclosure, a plaintiff must include evidence of the risks involved and what disclosures were made by the prescriber.[b] Some jurisdictions have a professional standard of disclosure that the sufficiency of information disclosed should be what a "reasonable clinician" would reveal to his or her patient. Expert testimony of what risks a reasonable medical practitioner would disclose under the same or similar circumstances is required.[c,d] A plaintiff must also establish causation between the inadequate disclosure

and the injury.[e] The issue is whether a reasonable person in the plaintiff's position would have consented to the procedure had the proper disclosure been made.[f] The plaintiff has the burden of producing evidence from which a jury or other decision maker could determine whether a reasonable person would have consented to the procedure.[g]

Other jurisdictions use a "reasonable person" standard. This would involve disclosing the information sufficient for a reasonable patient or legal decision maker acting on that patient's behalf to make a rational decision. In those jurisdictions, the clinician should discuss the following matters with the patient and or authorized decision makers:

1. The nature of the condition for which medication is being proposed.
2. The likely outcome of nontreatment.
3. The proposed medication treatment and how and to what extent that will benefit the patient and condition.
4. The risks and side effects associated with the medication. Risks of medications or other treatments proposed that should be disclosed should include those that a reasonable person would be likely to consider significant. The standard in malpractice case law is that the information not disclosed would have been considered substantial in evaluating the risk. Exceptions to this exist when severe harm could result if disclosure were made (principle of therapeutic privilege). Transparency in self-disclosure is important. The clinician's biases and skill level should be disclosed as that might also effect a patient's decision about the risk involved. If a clinician has never treated a certain condition before or never used a certain treatment before, it may be later be judged important to have informed the decision maker of this. Disclosure of potential conflict of interests or even the appearance of such conflict is recommended.
5. Alternative treatments available with their attendant benefits and risks.

Documentation

In the defense against a claim of negligence, legible and timely documentation of prescriptions and orders along with details of the informed consent process, and instructions given to patient and caregivers support the assertion that proper care was given. Prescriptions that are difficult to read pose a hazard to the patient. The prescriber should assure patient and caregivers understand instructions about dosages and interval between doses and potential side effects is important. Warning as to the importance of parental oversight and responsibility to safeguard potentially dangerous medicines should be given.

Documentation of all other medications the patient is currently taking and consideration of drug–drug interactions is a part of clinical care.

Refills, frequency of follow-up intervals to monitor for effectiveness, appropriate compliance with therapy and side effects is important, particularly in children and adolescents with medicines that have a potential for lethality in overdose, irreversible side effects, or abuse potential.

It is important to have permission to communicate with the other individuals involved with the child or adolescent's care about all medications that are being prescribed and about signs, symptoms of side effects, and positive responses to the medications. When a patient or their decision maker will not allow this, then the prescriber's ability to treat the patient is compromised.

Law of Agency/Vicarious Liability

Multiple professional disciplines interact with individual children and families. Pediatric psychopharmacology is practiced within this greater framework. Given the multidisciplinary nature of the treatment of a child or adolescent with interactions with family, school, and larger society, there might be tensions or disagreement at times of critical decision making regarding issues such as dangerousness to self or others, diagnostic labeling, as well as treatment choice recommendations. The prescriber often does not have input on important treatment decisions. The concept of *respondeat superior* is affected by these facts. Most others interacting with the child or adolescent are not acting as agents for the prescriber. The prescriber, even a physician, collaborating with a therapist is not responsible for the actions of that therapist unless he or she is directing these activities in a supervisory manner. The nonsupervisory nature of the relationship should be made clear so that false assumptions are not made.

Medication Practices

Psychotherapeutic medications for the treatment of mental illness are considered one of the most useful and important forms of treatment available for mental illness in both adults and youth. Millions of prescriptions are written annually. Prescribing medications has become so routine and commonplace; it is easy for prescribers and patients to lose sight of the risks involved.

Multiple factors have led to the increased recognition of mental illness in children and adolescents as well as the increase in medication use including the following:

1. The nature *versus* nurture controversy of the last century yielded an increased emphasis and understanding of the biology of mental illness. Biological psychiatry with primary emphasis on the individual's biological vulnerabilities interplaying with psychosocial stressors led to popularization of the simplistic concept of "chemical imbalance". A corollary to this concept was the idea that psychotherapeutic medications righted the imbalance. This notion was applied to youth as well as adults.
2. Large epidemiologic studies such as the 1999 Methodology for Epidemiology of Mental Disorders in Children and Adolescents (MECA) showed significant functional impairment due to mental or addictive disorders in approximately 1 in 10 of the pediatric population, for a total of four million in the United States alone [8].
4. Successful marketing by the pharmaceutical industry of psychotherapeutic medications had significant impact on both prescribers and the general public in terms of *increased expectations* of relief from distressing symptoms. Given the awareness of the availability of medications helpful in treating adults, there has been a willingness to treat youth without awaiting regulatory approvals.

Pediatric Psychohamacology in Practice

It is a statement of fact that most of the medications, psychiatric or not, used in children and adolescents, with the exception of stimulants, were not extensively studied or approved for use in this population. Manufacturers regularly put in disclaimers to that effect in their labeling. Therefore, the use of "off label" (i.e., not regulatory agency approved) prescribing remains the norm rather than the exception.

Pharmacodynamics is the branch of pharmacology that studies how medicines work. In general, psychoactive medications alter the biochemical environment in the synaptic space through either the blockade or activation and enhancement (antagonism, partial agonism, and agonism) of nerve cell membrane receptors. This in turn alters the way the nerve cells communicate.

In children and adolescents, more than in the midadult period in which most new medications are studied, the brain is developing by growing, and pairing down dopaminergic, serotononergic, and noradrenergic cell networks in areas of the brain, which are important to functions of attention, mood, anxiety, and perceptions. In adult populations these brain systems are relatively stable and the medications have been studied more extensively prior to approval and widespread use. Younger, developing brain systems may be permanently affected by mechanisms that are not well understood. Because of this, children and adolescents have differential risks when administered psychotherapeutic medicines. Examples include different risks of side effects such as extrapyramidal motor side effects and dyskinesias from antipsychotics as well as possibilities of overactivation of mood and attentional systems due to increased sensitivities to antidepressants.

Pharmacokinetics studies and informs us as to the absorption, distribution, metabolism, and excretion of medicines. It is important to understand that various people and subpopulations do this differently. In children and adolescents, the phamacokinetic properties of medicines need to be studied thoroughly for the safest administration of medicines that benefit them.

Children have a higher gastric and intestinal motility than adults thus favoring more rapid absorption. Prepubescent children and adolescents usually have fewer fat stores compared with adults, which may lead to higher plasma concentration of medicines in part distributed among fat stores. The metabolic breakdown of medications and toxins in the body differs. Excretion of medicines or the metabolites of medicines, usually by the kidneys might also be affected.

Treatment

Because standard of care is determined on all clinical data available for an individual child or adolescent, the following information is intended as informational only and not intended to imply adherence or deviance from the standard. Side effects are fairly common and so management should include regular periodic monitoring and routine reassessment as to the need to continue with the medication.

History

Review of history is very important for diagnosis and treatment planning. It is not unusual for psychiatrically ill children to have comorbid medical conditions that can influence their psychiatric symptoms. Developmental history should be reviewed with specific attention to development of gross and fine motor skills, speech and language development, affective relatedness, and attachments. A medical history (allergies, adverse drug reactions, acute and chronic illnesses, hospitalizations, injuries, loss of consciousness and traumatic brain injuries, and treatments) should be obtained. Previous laboratory findings and brain imaging if available should be reviewed.

Psychosocial history should include the family of origin, adoption, foster care, exposure to influences of alleged maltreatment, abuse, trauma, and losses.

Past psychiatric treatments and results should be reviewed.

A recent physical examination by the child or adolescent's primary care provider should be reviewed. Many psychiatric medications can be teratogenic and most teen pregnancies are unplanned. Sexually active postpubertal girls should have a pregnancy test along with counseling regarding the use of appropriate methods of birth control and methods of preventing sexually transmittable disease.

Documentation of the above and appropriate laboratory tests ordered prior to initiation of medication therapies and during ongoing monitoring is important.

Diagnosis

Accurate diagnosis in children and adolescents requires clinical assessment often using multiple sources of information including interactions with, observation of, and interview of the child or adolescent, parents, caregivers, and teachers if possible. Guidelines for assessment and treatment are available from the American Academy of Child and Adolescent Psychiatry, American Academy of Pediatrics and other organizations [9, 10].

Substance Abuse. Drug and alcohol screening with clinical questions augmented by urine toxicology for substances of abuse when indicated should be routine among adolescents and preadolescents. Substance-induced behavioral effects can mimic psychiatric illness and cause or aggravate other mental illnesses. Substance use disorders and tobacco addiction need treatment along with other mental illnesses. Making appropriate referrals for treatment of addiction and dependence and identification of patients or caregivers who might be misusing medications, altering prescriptions or getting multiple prescriptions from multiple sources (doctor shopping) is important.

Examples of Therapeutic Medications. *ADHD and the Use of Stimulants and Nonstimulants in its Treatment.* Attention deficit hyperactivity disorder (ADHD) is characterized by early childhood onset of an enduring pattern of inattention and/or hyperactivity and impulsive behavior. In total, 4–12% of children are affected by the disorder [11].

Many children can contain their behaviors during office visits and so collecting information about the child's behavior in multiple settings is most useful. Well-validated and normed behavior rating scales are available and useful in assessing and measuring treatment effects. These would include the Conners' Rating Scales-Revised; Brown Attention-Deficit Scales; and Swanson, Nolan, and Pelham (SNAP-IV) [12].

Stimulant Medications. In 1937, Bradley reported on the positive effects of amphetamine on children institutionalized for neurobehavioral reasons [13]. Since then, opposing social forces have exerted pressures for either wider acceptance of these medications or for more restrictions on their use. The existence of these pressures has led scientists to explore the issue and, for the most part, demonstrate the efficacy of these medicines [14]. Concerns regarding associations with sudden cardiac death have brought these medications into greater controversy in the recent years.

Nonstimulant Medications. Atomoxetine is a selective norepinephrine reuptake inhibitor that is superior to placebo in the treatment of ADHD at appropriate doses [15]. Although the manufacturer recommends once a day dosing, avoiding adverse effects while maintaining adequate therapeutic doses often requires twice a day dosing. Because it shares features with the antidepressants it is associated with an increased risk of suicidality.

Guanfacine and clonidine are antihypertensive medications that have found usefulness. They are generally used in children who cannot tolerate or as adjuncts to the other medications.

Antidepressant Medications, Mood Disorders, and Suicide. Major depression and bipolar mood disorders are enduring and disabling conditions. Suicide is among the most feared of outcomes in the field of mental health. Ninety percent of suicides occur in the context of psychiatric illness and mood disorders are the most likely illnesses to be associated with the act [16]. Suicide is a major cause of preventable death among youth [17].

The prevalence of mood disorders is relatively low before the onset of puberty and equally distributed among the two sexes. After puberty, however, the prevalence increases to adult levels. Various epidemiologic studies point to a prevalence of 5–10% of teens and young adults suffer from major depression or other disabling depressive disorders such as dysthymia and depressive disorder not otherwise specified [18].

Pharmacotherapy with selective serotonin reuptake inhibitor (SSRI) antidepressants either with cognitive behavioral or interpersonal psychotherapy for children and adolescents with major depressive disorder had been the first line and standard of treatment until 2003 and 2004 [1] when regulators in the United Kingdom, the United States, and the European Common Market, became concerned about possible or even causal links between youth suicidality and antidepressants of all classes. UK regulators went on to advise against the use of almost all antidepressants in persons below 18 years. Black box labeling and letters of warning from manufacturers to prescribers were issued in the United States. These expressed warnings and concerns were based on collective data comparing placebo to active medication in multiple different trials, which showed increased rates of suicidality on the active medications and have been extensively reviewed [19].

The stance of regulators was contrary to the opinion of many prescribers who felt the medicines to be very helpful in the real world of clinically complex populations. The warnings had both intended and unintended consequences. Internationally, warnings had the effect of reducing prescriptions to youth by some 20% in 2004 [20]. Coincidental to this decrease in prescribing of antidepressants was a marked increase in actual suicide rates among youth. There exists epidemiologic data showing an inverse correlation between the SSRI prescriptions and suicide rates in multiple population groups in several countries [21, 22]. Antidepressant prescriptions seem to have a protective effect by reducing suicide rates such that strong arguments exist regarding the potential harm resulting from a reduction of SSRI antidepressant use [23]. Indeed, according to the United States Center for Disease Control, after having dropped some 28% from 1990 until 2004, suicide rates for youth in the United States surged 8% leading to 4599 deaths in 2004 following the issuance regulatory letters and black box warning [17]. Prescribers have been left in a conundrum of needing to treat their patients with these medicines despite alarming warnings from regulatory authorities. The best practice to follow is to assure that the diagnosis is correct, the patient makes a full informed consent and to monitor for the development of new onset of suicidal ideations or agitation.

The results of the above-noted reviews [19], which used meta-analytic methods, suggest that there is no benefit to the use of tricyclic antidepressants in children and adolescents with depressive disorders and that the risks, given their toxicity in overdose, generally outweigh the benefits of their use in this population.

SSRI antidepressant medications remain helpful in the treatment of mood and anxiety disorders [24].

Antipsychotics. Antipsychotics are used in treating hallucinations, delusions, hostility, aggression, and disorganized thinking associated with major mental illness such as bipolar disorder, major depression with psychotic features, and schizophrenia. These diseases often manifest in late adolescence or young adulthood and often go undiagnosed for several years. Antipsychotics are also used in the treatment of children with autism and mental retardation who

exhibit prominent irritability, aggression, and self-injurious behavior.

It is thought that blockade of D_2 receptors in the mesocortical and prefrontal systems is responsible for the antipsychotic effects. With older antipsychotics, the degree of D_2 receptor blockade is predictive of the potency of antipsychotic effect on hallucinations, delusions, agitation, hostility, and disorganization. Many side effects are also a function of the medication's D_2 binding properties to receptors in the basal ganglia. Drug-induced Parkinson symptoms such as feeling stiff, shuffling gait, tremor, drooling, and masked facial expression led to the mislabeled term *chemical restraint* for this group of medicines that helped many people with mental illness to return to more normal lives.

The newer second generation antipsychotics (SGAs) (olanzapine, risperidone, paliperidone, ziprasidone, quetiapine, aripirazole, and clozapine) were promoted as having fewer side effects on movement and positive effects on motivational circuitry in the prefrontal cortex. The SGAs seemed to balance out many of the untoward effects of D_2 blockade through effects on serotonin 2A receptors, making them agents of first choice. Clozapine, however, because of its potential for causing dangerous agranulocytosis, is reserved for treatment refractory disease.

The number of studies supporting their efficacy and safety in children and adolescents is limited. Children, adolescents, and young adults are particularly susceptible to antipsychotic drug-induced movement disorders (Extra-Pyramidal Symptoms (EPS), akithesia, and diskinetic side effects). The newer medicines have proven to have fewer with drug-induced movement disorders in adults but have problematic effects related to weight gain, glucose regulation, lipid regulation, sedation, prolactin levels, and cardiac conduction. It is hypothesized that because of varying proportions of blockage and, in the case of aripiprazole, the partial agonism of the dopamine receptors, as well as the above noted blockade of serotonin receptors, the effectiveness and sensitivities to side effects in children and adolescents varies markedly between one medicine and another. This means caution should be exercised in their use, careful review of improvement of targeted symptoms and monitoring for side effects. Of note, risperidone received approval from the Food and Drug Administration (FDA) for an indication in the treatment of children with autism and severe problem behaviors [25, 26].

Children and adolescents are generally treated with the lowest possible effective dose of these medications with periodic monitoring for side effects including baseline and follow-up weight measures and laboratory testing for glucose and lipid abnormalities and abnormal involuntary movements all the while examining the benefits and the need for ongoing use. Because of the potential for significant and sometimes irreversible side effects, documentation of the consideration of the risk benefit ratio prior to initiation and while continuing with these medications is important.

Mood Stabilizers. Mood stabilizers are used in the treatment of bipolar disorder and schizoaffective disorders. The term *mood stabilizer* gathered wide use after publication of reports in the late 1980s demonstrating the efficacy of carbamazepine and divalproex in adult patients with bipolar disorder giving more choices than lithium and adjunctive agents [27]. Oxcarbazapine and lamotragine are now also widely used. Use of these medications has migrated into pediatric use. These medicines are associated with potentially serious side effects; some side effects are more prominent in children and adolescents, including hirsuitism and polycystic ovary syndrome in adolescent girls associated with divalproex and increased incidences of severe rashes with lamotragine in children and adolescents. Oxcarbazapine is metabolized more rapidly in preadolescents. Patient, parent, and caregiver education as to side effects is important.

Routine laboratory evaluations include complete blood-cell count with differential and platelet counts, metabolic profile including liver function tests, pregnancy tests, thyroid function monitoring, as well as plasma drug levels of lithium, valproate, and carbamazepine.

End Notes

[a.] Wilkerson, 908 S.W.2d at p. 696.
[b.] Aiken, 396 S.W.2d at p. 673.
[c.] Aiken, 396 S.W.2d at p. 674–675.
[d.] Wilkerson, 908 S.W.2d at p. 696.
[e.] Aiken, 396 S.W.2d at p. 676.
[f.] Wilkerson, 908 S.W.2d at pp. 696–697

g. Wilkerson, at p. 697 (citing Aiken, 396 S.W.2d at p. 676).

References

[1] March, J., Silva, S., Petrycki, S., Curry, J., Wells, K., Fairbank, J., Burns, B., Domino, M., McNulty, S., Vitiello, B. & Severe, J. (2004). Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression: treatment for adolescents with depression study (TADS). Randomized controlled trial, *The Journal of the American Medical Association* **292**(7), 807–820.

[2] Martin, A. & Volkmar, F. (eds) (2007). *Lewis's Child and Adolescent Psychiatry*, 4th Edition, Lippincott, Williams & Wilkins, Philadelphia.

[3] Coffey, C. & Brumback, R. (eds) (2006). *Pediatric Neuropsychiatry*, Lippincott, Williams & Wilkins, Philadelphia.

[4] Conner, D. & Meltzer, B. (eds) (2006). *Pediatric Psychopharmacology Fast Facts*, W.W. Norton, New York.

[5] Fleming, J. (1992). *The Law of Torts*, 8th Edition, Law Book, Sidney, p. 102.

[6] Applebaum, P. & Gutheil, T. *Clinical Handbook of Psychiatry and the Law*, 4th Edition, Williams & Wilkins, Baltimore, p. 126.

[7] Broome, M.E. (1999). Consent (Assent) for research with pediatric patients, *Seminars in Oncological Nursing* **15**(2), 96–103.

[8] (1999). *Mental Health: A Report of the Surgeon General*, Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services, National Institute of Mental Health, Rockville.

[9] Kowatch, R. *et al.* (2005). Treatment guidelines for children and adolescents with bipolar disorder: child psychiatric workgroup on bipolar disorder, *Journal of the American Academy of Child and Adolescent Psychiatry* **44**, 213–235.

[10] American Academy of Pediatrics. Subcommittee on Attention-Deficit/Hyperactivity Disorder and Committee on Quality Improvement (2001). Clinical practice guideline: treatment of the school-aged child with attention-deficit/hyperactivity disorder, *Pediatrics* **108**(4), 1033–1044.

[11] Brown, R.T., Freeman, W.S., Perrin, J.M., Stein, M.T., Amler, R.W., Feldman, H.M., Pierce, K. & Wolraich, M.L. (2001). Prevalence and assessment of attention-deficit/hyperactivity disorder in primary care settings, *Pediatrics* **107**, E43.

[12] Swanson, J., Lerner, M., March, J. & Gresham, F.M. (1999). Assessment and intervention for attention-deficit/hyperactivity disorder in the schools: lessons from the MTA study, *Pediatric Clinics of North America* **46**, 993–1009.

[13] Bradley, C. (1937). The behavior of children receiving benzedrine, *The American Journal of Psychiatry* **94**, 577–585.

[14] Zametkin, A.J. & Rappaport, J.L. (1987). Neurobiology of attention deficit disorder with hyperactivity: where have we come in 50 years? *Journal of the American Academy of Child and Adolescent Psychiatry* **26**, 676–686.

[15] Michelson, D., Allen, A.J., Busner, J., Casat, C., Dunn, D., Kratochvil, C., Newcorn, J., Sallee, F.R., Sangal, R.B., Saylor, K., West, S., Kelsey, D., Wernicke, J., Trapp N.J., & Harder, D. (2002). Once-daily atomoxetine treatment for children and adolescents with attention deficit hyperactivity disorder: a randomized, placebo-controlled study, *The American Journal of Psychiatry* **159**(11), 1896–1901.

[16] Beautrais A.L., Jopyce P.R., Mulder R.T., Ferguson, D.M., Deavoll, B.J. & Nightingale, S.K. (1996). Prevalence and comorbidity of mental disorders in persons making serious suicide attempts: a case–control study, *The American Journal of Psychiatry* **153**, 1009–1014.

[17] Centers for Disease Control and Prevention WISQARS (Web-based Injury Statistics Query and Reporting System).

[18] Birmaher, B., Ryan, N.D. & Williamson, D.E., Brent, D.A., Kaufman, J., Dahl, R.E., Perel, J. & Nelson, B. (1996). Childhood and adolescent depression: a review of the past 10 years Part I, *Journal of the American Academy of Child and Adolescent Psychiatry* **35**(11), 1427–1439.

[19] U.S. Food and Drug Administration (2006). *Clinical Review: Relationship between Antidepressant Drugs and Suicidality in Adults*, Food and Drug Administration, Center for drug Evaluation and Research, Rockville.

[20] Gibbons, R.D., Brown, C.H., Hur, K., Marcus, S.M., Bhaumik, D.K., Erkens, J.A., Herings, R.M. & Mann, J.J. (2007). Early evidence on the effects of regulators' suicidality warnings on SSRI prescriptions and suicide in children and adolescents, *The American Journal of Psychiatry* **164**, 1356–1363.

[21] Ludwig, J. & Marcotte, D.D. (2005). Anti-depressants, suicide, and drug regulation, *Journal of Policy Analysis and Management* **24**, 249–272.

[22] Gibbons, R.D., Hur, K., Bhaumik, D.K. & Mann, J.J. (2006). The relationship between antidepressant prescription rates and rate of early adolescent suicide, *The American Journal of Psychiatry* **163**, 1898–1904.

[23] Gibbons, R.D., Hur, K., Bhaumik, D.K. & Mann, J.J. (2006). The relationship between antidepressant prescription rates and rate of early adolescent suicide, *The American Journal of Psychiatry* **163**(11), 1989–10904.

[24] Walkup, J. *et al.* (2002). Treatment of pediatric anxiety disorders: an open-label extension of the research units on pediatric psychopharmacology anxiety study, *Journal of Child and Adolescent Psychopharmacology* **12**(3), 175–188.

[25] McCracken, J.T., McGough, J., Shah, B., Cronin, P., Hong, D., Aman, M.G., Arnold, L.E., Lindsay, R., Nash,

P., Hollway, J., McDougle, C.J., Posey, D., Swiezy, N., Kohn, A., Scahill, L., Martin, A., Koenig, K., Volkmar, F., Carroll, D., Lancor, A., Tierney, E., Ghuman, J., Gonzalez, N.M., Grados, M., Vitiello, B., Ritz, L., Davies, M., Robinson, J. & McMahon D. Research Units on Pediatric Psychopharmacology Autism Network (2002). Risperidone in children with autism and serious behavioral problems, *The New England Journal of Medicine* **347**(5), 314–321.

[26] Research Units on Pediatric Psychopharmacology Autism Network (2005). Risperidone treatment of autistic disorder: longer-term benefits and blinded discontinuation after 6 months, *The American Journal of Psychiatry* **162**(7), 1361–1369.

[27] Emrich, H.M., Dose, M. & von Zerssen, D. (1985). The use of sodium valproate, carbamazepine and oxcarbazepine in patients with affective disorders, *Journal of Affective Disorders* **8**, 243–250.

ROBERT W. LOVELL

Psychosis ***see*** Delusions

Psychosis: Puerperal ***see*** Postpartum Psychosis

Psychotic Disorder: Shared ***see*** Temporary Insanity

Publishing in Forensics and Peer Review ***see*** Peer Review as Affecting Opinion Evidence

Pyromania ***see*** Firesetting



QiaAmp

The *QIAamp® DNA Micro Kit* DNA extraction method utilizes a silica-based spin column to separate DNA from other cellular components that are released after cell lysis. DNA binds, specifically, to the silica-gel membrane embedded within the microcentrifuge tube while other cellular components pass through. This process isolates a purified DNA extract as the washing steps remove inhibitory proteins and divalent cations, which are the cofactors for harmful nucleases. DNA is released from the silica column by the addition of an elution buffer. The entire procedure takes approximately 30 min. Lyophilized carrier RNA can be added to the lyzed sample at the beginning of the process to facilitate membrane binding. This is recommended for forensic samples; however, care must be taken to ensure that downstream quantitation methods are DNA specific. QiaAmp microextraction methods have the advantage of being a simple, reliable, flexible, single-tube technique that is highly amenable to automated liquid handling platforms. In many laboratories, QiaAmp is used as a postextraction purification step, particularly for degraded or severely compromised tissue.

Related Articles

DNA

Extraction

SIMON J. WALSH

Quality Systems: Toxicology

Introduction

Quality management (QM) aims to ensure that the activities necessary to design, develop, and provide a product or service are effective and "fit for purpose". QM including laboratory accreditation, i.e., inspection and independent certification of laboratories to ensure, as far as possible, the quality and reliability of the work produced, is becoming increasingly important [1]. A prerequisite for laboratory accreditation is to have a documented QM system. It was thought that the likelihood of examination in court was sufficient to ensure the accuracy and reliability of forensic toxicology results, but implementation of QM systems and accreditation of laboratories by regional or national accreditation bodies provide a far more robust method of ensuring quality. Obviously, all this has to be paid for and inspectors themselves have to be trained and accredited.

The ISO 9000 family (ISO 9000:2000 and 9001:2001) form a basis for many QM systems (Box 1) and are maintained by the International Organization for Standardization (ISO, http://www.iso.ch/iso/en/iso9000-14000/index.html, accessed 27f August 2007). A related set of standards, ISO 14000, is concerned with environmental management. ISO

Box 1. Some of the requirements in ISO 9001:2001

- A set of procedures that cover all key processes in the organization being managed
- Active monitoring of processes to ensure that they are effective
- Adequate record keeping
- Monitoring the quality of output, with implementation of appropriate corrective action if necessary
- Regular review of individual processes and the QM system itself for efficacy
- Implementation of a culture of continual improvement

works in collaboration with the International Electrotechnical Commission (IEC) and other metrological organizations.

ISO/IEC 17025:2005 is the main standard used by testing and reference laboratories to implement a QM system. It replaces ISO/IEC 17025:1999 (formally ISO Guide 25 and EN 45001). There is much in common with ISO 9000/9001, but ISO/IEC 17025 adds the concept of competence. There are two main sections: (i) management requirements that are primarily related to the operation and effectiveness of the laboratory QM system and (ii) technical requirements that address the competence of staff, methodology, and test/calibration equipment. ISO 15189:2007 defines standards for the operation of a medical laboratory, and is consistent with ISO 9000/9001.

Laboratory accreditation procedures and conformity assessment bodies (CABs) are assessed against ISO/IEC 17011:2004 (previously ISO/IEC Guide 58). CABs are defined as organizations providing assessment services such as testing, inspection, management system certification, personnel certification, product certification, and calibration.

Laboratory Accreditation

The International Laboratory Accreditation Cooperation (ILAC) website (http://www.ilac.org/, accessed 27 August 2007) gives details of accreditation bodies in many countries. Key elements are the Quality Policy, a statement of the quality aims of the organization, and the Quality Manual, which defines the organization's QM system. The layout of the Quality Manual should follow the outline of ISO 17025:2005.

Laboratory operations can be divided into preanalytical, analytical, and postanalytical phases (Box 2). Written procedures, usually known as *standard operating procedures* (*SOP*s), should describe all aspects

Box 2. Stages in analytical toxicology laboratory operation

- **Preanalytical**
  Procedures must be in place to advise on appropriate sample collection (including sample tubes) and to ensure the safe transport, receipt, and storage of biological samples once in the laboratory, and for arranging the priority for the analysis
- **Analytical**
  Validated (i.e., tried and tested) procedures must be used to perform the requested or appropriate analyses to the required degree of accuracy and reliability in an appropriate timescale
- **Postanalytical**
  A mechanism for reporting results and maintaining confidentiality by telephone, fax, or other electronic means and in writing must be in place. Proper interpretation of results, especially for less-common analytes, must be provided. Full records of the analysis must be kept for at least 5 years (10 or more years in forensic work) unless otherwise determined by statute. Residues of samples must be stored appropriately until disposed of safely in an agreed timeframe

of laboratory operation, including management and health and safety aspects. The Society of Forensic Toxicologists (SOFT) and American Academy of Forensic Sciences (AAFS) have published detailed guidelines for the operation of forensic toxicology laboratories, much of which is also applicable to clinical toxicology laboratories [2].

Implementation and documentation of internal audits, both vertical (when, for example, the documentation concerning the analysis of a particular sample is examined) and horizontal (such as examining operating or quality procedures for internal consistency), are important parts of laboratory accreditation. Clinical or operational audit, i.e., examining the results generated in the light of the purpose for which they were requested, although much more difficult to undertake, is also important in the accreditation process. The results of such audits provide valuable training material.

Documentation of "quality queries", i.e., instances when mistakes or failures in processing have occurred, even if the error or failure was detected and corrected before a result was issued, and implementation and monitoring of corrective action, are also important in the accreditation process.

Method Implementation and Validation

Whatever method is used for a given analysis it must be validated, i.e., it must be shown to be "fit for purpose" [3]. Assay validation should conform, as far as possible, with the US Food and Drug Administration (FDA) Center for Drug Evaluation and Research (CDER) guidance for bioanalytical method validation [4]. Data for within-day (repeatability), between-day, and total precision should be calculated according to the protocol proposed by the Clinical and Laboratory Standards Institute [5]. A number of terms important in understanding method validation are given in Table 1.

Quantitative methods must have good precision and accuracy. Selectivity (freedom from interference, specificity) is important when a single species is to be measured, but broad specificity may be useful when screening for the presence of a particular class of compounds. The recovery of the analyte, i.e., how much of the compound of interest is recovered from the sample matrix during an extraction, for example, is important if sensitivity is limiting, but need not be an issue if the lower limit of quantification (LLoQ), accuracy, and precision of the assay are acceptable.

Any quantitative analysis has associated errors, both random and systematic. In chromatographic and other separation methods, the "internal standard" method is often used to reduce the impact of systematic errors such as variations in injection volume, evaporation of extraction solvent, or mass spectrometry (MS) response during the analysis. Thus, a known amount of a second compound that behaves similarly to the analyte during the analysis, but elutes at a different place on the chromatogram or is otherwise detected independently of the analyte (the internal standard) is added at an appropriate stage in the analysis. Subsequently, the detector response of the analyte relative to the response of the internal standard is plotted against analyte concentration when constructing a calibration graph. Requirements for an internal standard for chromatographic assays are summarized in Box 3.

Stable isotope-labelled analogues are widely used as internal standards in MS (isotope dilution MS). Isotopic internal standards have virtually identical chemical and physical properties to the analyte and thus extraction, derivatization if needed, chromatography, and fragmentation are often virtually identical. However, the site of isotopic labeling should be chosen such that the bonds linking the isotope are not broken during fragmentation, as bonds involving heavier isotopes are more stable and the fragmentation pattern of the labeled compound could thus differ from that of the analyte. The vibrational frequencies of carbon–deuterium bonds are less than those of the corresponding carbon–hydrogen bonds, for example, so that deuterated compounds tend to be more stable than their unlabeled homologs. That labeled and unlabeled compounds may be partially resolved during the chromatographic analysis (deuterated analogs may elute slightly before the unlabeled analyte in GC-MS, for example) must be borne in mind not only with regard to choosing the correct integration parameters, but also because any ion suppression due to cochromatographed components may differ between the internal standard and the analyte(s). The internal standard may add to the degree of ion suppression in LC-MS, but generally both internal standard and analyte are affected equally by such phenomena [6].

Table 1 Terms used when reporting method validation

Term	Notes
Accuracy	The difference between the measured value and the accepted ("true") value
Calibration range	The range of concentrations between the highest and lowest calibration standards. This should encompass the range of concentrations found in the test samples
Coefficient of variation (CV)	An obsolete term for RSD
Higher limit of quantitation (HLoQ)	The highest concentration that can be quantified. Not always quoted, but important in assays with a clear upper "cutoff", for example, immunoassays and fluorescence assays
Internal standard	A second compound, not the analyte, added at an appropriate stage in the assay to correct for systematic errors in the analysis
Limit of detection (LoD)	The smallest amount of analyte that can be detected. Usually defined as some multiple (5, for example) of the baseline noise (signal-to-noise ratio = 5) or multiple of the SD of the blank signal
Linearity	A definable and reproducible relationship between a physicochemical measurement (e.g., UV absorption) and the concentration of the analyte, but not necessarily a straight line
Lower limit of quantification (LLoQ)	The lowest concentration that can be measured within defined limits. Usually a concentration for which the precision and accuracy have been set arbitrarily, e.g., RSD <20%
Precision	The scatter of measured values about a mean value. Usually quoted as RSD – within-assay (repeatability) and between-assay precision (reproducibility) is commonly given
Relative standard deviation (RSD)	The standard deviation of replicate measurements expressed as a percentage of the mean value: RSD = SD/Mean $\times 100\%$ Useful when comparing precision at different concentrations
Selectivity	The ability to distinguish between the analyte and some other compound
Signal-to-noise (S/N) ratio	Strictly, the response to the analyte divided by the amplitude of the random electronic noise of the detection system. In practice, the background signal due to interfering compounds is often greater than the electronic noise

Box 3. Requirements for an internal standard

An internal standard must:

- Be completely resolved from the known and unknown substances in the chromatogram, or detected selectively as in MS for stable isotopes
- Elute near to (preferably just after the last) peak(s) of interest
- Have a similar detector response (peak height or area) to the analyte(s)
- Have similar chemical and physical properties to the analyte(s)
- Undergo any derivatization reaction in the same way as the analyte(s)
- Be chemically and physically stable on storage in solution and during the analysis
- Be easily available with adequate purity

Deuterium is the most commonly used stable isotope label. This is because of not only cost, but also the availability of high purity reagents. The presence of each deuterium atom in the molecule increases the M_r of the molecule by 1 atomic mass unit (u). For high-sensitivity analyses, it is recommended that the mass of the internal standard is at least 3 u greater than the analyte to reduce interference from naturally occurring isotopes. When stable isotope-labeled analogs are not available or the procedure can be used to quantify a large number of analytes simultaneously, internal standards that have chemical properties close enough to the analyte(s) to yield reliable quantitative data are used.

Reference Compounds

A fundamental starting point in any assay is obtaining certified pure reference material, or at least the best approximation to such material that can be attained. When preparing primary standards, particular attention should be paid to the M_r of salts and their degree of hydration (water of crystallization). Analytical results are normally reported in terms of free acid or base and not of a salt. The supplier, batch or lot number, purity, expiry date, and any other relevant information supplied with a compound should be recorded. Compounds should be stored in the dark under conditions recommended by the supplier. Every effort should be made to obtain a certificate of analysis or other appropriate documentation, but supply of such documentation is at the discretion of the manufacturer.

Preparation and Storage of Calibration Solutions

Assay calibration is normally done by analyzing standard solutions containing each analyte over an appropriate range of concentrations prepared in analyte-free plasma, urine, or other appropriate fluid. The chosen medium should be analyzed prior to the addition of the analyte(s) to ensure the absence of interferences.

Balances for weighing reagents or calibration standards, and automatic and semiautomatic pipettes, must be kept clean and checked for accuracy on a regular basis. Weighing of balance check weights should be recorded. Pipette accuracy should be documented by dispensing purified water and recording the weight dispensed. Weighing of reference compounds should normally be performed by one analyst and witnessed by a second analyst, as should other steps, in preparing calibration standards and internal quality control (IQC) solutions such as calculating and performing dilutions.

The range of the calibration curve should cover the range of concentrations expected in the samples. The calibration curve should not normally be extrapolated beyond the lowest or highest standard solution. If the concentrations of some of the samples being analyzed are below the lowest calibration standard, then the assay may be repeated with the inclusion of lower concentration standards, if that is possible, or the result reported as less than the lowest standard. High concentration samples may be diluted to fit within the calibration range, provided the validity of doing so has been demonstrated during method development and if sufficient sample is available. Dilutions should be made with the same (e.g., blank plasma from the subject), or very similar (e.g., the matrix used for the calibration standards) matrix as the sample.

Neonatal calf serum is often used in the preparation of calibration solutions for plasma or serum assays. Newly prepared calibration solutions should be validated before use by comparison either with existing calibration solutions, or with IQC solutions prepared in human plasma or serum, and the results recorded. If, for any reason, neonatal calf serum proves unsuitable, human plasma or serum from an appropriate source should be used.

Blood-bank whole blood (transfusion blood) is sometimes used to prepare calibration standards. However, such blood and "plasma" derived from it (i) will usually be diluted with citrate solution, which has a high buffering capacity; (ii) may contain lidocaine and sometimes lidocaine metabolites; and (iii) may well contain plasticizers and other contaminants that may interfere in chromatographic and possibly other assays, and may alter the distribution of drugs between red cells and plasma by altering protein binding. Commercially available equine or bovine blood may suffer some of these same problems.

The question of "matrix matching" for preparation of calibration or IQC solutions is a vexed issue. Clearly true "matrix-matched" standards cannot be prepared for solid tissues such as liver or hair. Use of "blank" homogenized tissue samples or hair digests, for example, as the matrix for standard preparation is one way forward here. Analyte(s) may also be adsorbed onto hair prior to digestion.

Similarly, it is rarely possible to prepare standards in analyte-free sample from a patient or victim such that the standard has exactly the same composition as the sample submitted for analysis. The method of "standard additions" (adding known amounts of analyte to portions of the sample) is very difficult and inaccurate if only a limited sample volume is available, does not allow validation against existing calibrators, and does not permit the preparation of IQCs. Finally, "blank" human blood or urine can be used, as discussed above, but the volume of blood available is usually limited.

If newly prepared calibration standards and the IQCs meet the criteria for acceptance, portions may be transferred to labeled 3-ml plastic tubes and stored ($-20\,^{\circ}C$) until used. Cleaning records and lists of contents should be posted on all refrigerators and freezers including those used to store biological samples to meet Health and Safety requirements. Refrigerators and freezers should be fitted with failure alarms and temperature-monitoring devices and temperature records kept. Computerized temperature-monitoring devices (e.g., Tinytalk, http://www.tinytag.info/products/index.asp, accessed 10 October 2007) are available.

Calibration Graphs

Normally, a calibration graph of analyte response *versus* concentration in the calibration standards is constructed. In chromatographic assays, the response may be peak height or area, or peak height or area ratios to an internal standard. The relationship may be a straight line or a curve.

Blank samples, usually the drug-free matrix, should be prepared and analyzed along with the calibrators and unknown samples. The blank is an important part of quality control (QC). In a chromatographic assay, for example, there should be no interfering peaks in the region of the analyte(s) or internal standard(s). The response of the blank should not be so high as to limit the working range of the assay. The blank is not a zero calibrator, and normally must not be included in the calibration data. Replicate analyses of "blank" signal are necessary to define the limit of detection (LoD) and so it is important when using electronic data capture to ensure that the calibration curve is not being forced though zero.

An important criterion for an analytical method is the LoD, i.e., the minimum concentration (amount) of analyte that can be detected reliably and differentiated from any background signals measured in analyte-free samples. Defining the LoD is not as simple as it might at first seem to be, as there are random errors associated with both the blank sample and samples at or near the LoD. For chromatographic assays, the LoD might be quoted as some arbitrary multiple of the signal-to-noise ratio (S/N) such as 3, 5, or 10 times. This, of course, presupposes that the S/N can be measured and, with biological samples, the limiting factor is rarely instrument noise, but rather signals due (usually) to endogenous interfering species, in which case some multiple of the standard deviation (SD) of the blank signal may be used.

For quantitative analyses, it is usual to quote the lowest concentration or amount that can be measured with defined values of precision and accuracy; this is the lowest limit of quantitation (LLoQ). Again the criteria are arbitrary, and the relative standard deviations (RSDs) for precision and accuracy typically range from 10–15%, depending on the requirement of the assay. For trace analyses and pharmacokinetic experiments, it may be necessary to accept values as high as 20%, although the RSDs for concentrations above the LLoQ would normally be considerably lower. The concentration of the lowest calibrator may be set to the LLoQ, although for some applications it may not be necessary to measure such low concentrations. In any event, concentrations below the lowest standard should not be reported.

Reports do not always quote a higher limit of quantitation (HLoQ), particularly when the assay has a wide linear range. However, for some techniques, such as immunoassay and fluorescence detection in HPLC, the working range of the assay should be defined by both LLoQ and HLoQ. In some immunoassays, the presence of very high concentrations of an analyte can give a result suggesting the presence of a very low concentration ("hook effect").

Batch Analyses

Batch-assay calibration should normally be by analysis of standard solutions of each analyte (six to eight concentrations across the calibration range) prepared in the appropriate matrix (e.g., analyte-free neonatal calf or human serum). IQC procedures should be instituted. This involves the analysis of independently prepared solutions of known compositions that are not used in assay calibration; normally low, medium,

and high concentrations of each analyte are prepared in analyte-free human serum, for example. Calibration standards are normally analyzed in duplicate, once at the beginning and once at the end of the batch. IQC samples are analyzed at the beginning and end of the batch and also after every 5–10 test samples as appropriate. External quality assessment (EQA) samples are analyzed as appropriate to conform to the requirements of particular EQA schemes. Single-point calibration methods for emergency work must be validated and the results compared with those from multipoint calibration [7].

The performance of batch analyses (analysis of a number of samples in the same analytical sequence) and analysis acceptance criteria should be as set out in the method validation guidance [4]. Typical assay-acceptance criteria are (i) chromatography (reproducible peak shape and retention time, stable baseline) and (ii) calibration graph ($r = 0.98$ or greater, intercept not markedly different from zero), and mean IQC results within acceptable limits (generally within 10% of nominal value). Acceptance criteria for patient samples are (i) "clean" chromatogram, i.e., absence as far as can be ascertained of interferences; (ii) duplicate values (peak height ratio to the internal standard) within 10% except when approaching the limit of sensitivity of the assay when duplicates within 20% may be acceptable; and (iii) results within the calibration range of the assay. Sample analyses falling outside acceptance limits may be repeated if sufficient sample is available.

While immunoassays may be performed as batch analyses, they may, depending on the instrument used, be subjected to random access analysis. IQC in such circumstance is best performed at timed intervals at low, medium, and high concentration. Some immunoassays are used as screens for the presence of drugs or metabolites, and "cutoff" concentration values are used to define positives. It is particularly important to ensure that the analytical performance at this cutoff point is properly investigated (±25% of cutoff value) to minimize the risk of false positives and negatives arising from poor assay performance.

If a result above the calibration range is obtained, then, if possible, a portion of the sample should be diluted with "blank" plasma/serum and reanalyzed. If a sample is from a suspected overdose patient, then sample dilutions ($1 + 1$, $1 + 3$, $1 + 9$) should be made using "blank" serum at the time of the initial analysis (if the available sample volume permits) and analyzed at the same time as the normal sample analysis. Postmortem whole-blood samples or tissue digests from suspected overdose cases, for example, may be diluted $1 + 3$and$1 + 9$ with "blank" human serum prior to analysis using standard methods. If the information available suggests a massive overdose, further dilutions may be made as appropriate prior to the analysis.

Quality Control and Proficiency Testing

Once an analytical method has been validated and implemented, it is important to be able to show that the method continues to perform as intended. In qualitative work, known positive and negative specimens should normally be analyzed at the same time as the test sample. A negative control ("blank") helps to ensure that false positives (owing to, for example, contaminated reagents or glassware) are not obtained. Equally, inclusion of a true positive serves to check that the reagents have been prepared properly and remained stable.

In quantitative work, assay performance is monitored by the systematic analysis of IQC samples. Plotting the results for the IQC samples on a chart allows the day-to-day performance of the assay to be monitored and gives warning of any problems as they arise. When new batches of calibration and IQC samples are prepared, it is prudent to ensure comparability of the results obtained with those given by an earlier batch, or with the results obtained using external QC material.

FDA guidelines (for batch processing) require duplicate IQCs at three concentrations (high, medium, and low) to ensure that the assay is performing satisfactorily across the calibration range. The assay batch is deemed acceptable, provided four of the six controls (with at least one at each concentration) are within specification.

Quality Control Charts

QC charts are valuable in that they (i) produce evidence of satisfactory assay performance and (ii) give visual warning if assay performance begins to deteriorate. As with any QC method, it is important that reliable estimates of the parameters defining the control material are established. The mean value, μ, should be obtained from a minimum of 10 observations and the SD, σ, should be the between-assay

value (interassay value). Obviously, these parameters must be measured when the analysis is performing satisfactorily.

Analytical laboratories commonly use Westgard rules to establish that assay performance is adequate [8]. Five different control rules are used to assess the acceptability of an analytical sequence. Further details and a multirule worksheet can be obtained from http://www.westgard.com/mltirule.htm (accessed 27 August 2007).

External Quality Assurance

Participation in EQA or proficiency testing (PT) schemes is an important part of QM [9]. In such schemes, portions of (sometimes lyophilized) homogenized plasma, serum, whole blood, or urine specimens are sent to a number of participating laboratories. After reconstitution in deionized water, if appropriate, the specimens are analyzed as if they were real samples and the results are reported before the true or target concentrations are made known.

EQA schemes measure interlaboratory performance and allow individual laboratories to detect and correct systematic errors. The laboratories do not have to use the same analytical method, as is usually the case with collaborative trials that are designed to test the reproducibility of a particular method. The results of EQA schemes are usually given as the z-score:

$$z = \frac{x - x_a}{\sigma_p} \qquad (1)$$

where x is an individual result, x_a is the accepted, "true" value and σ_p is known as the *target value of SD*, which should be decided on the basis of what is required of the test, and should be circulated in advance. If the result needs to be measured with high precision, then a low value of σ_p would be used. Thus, z is a measure of a laboratory's accuracy and the organizer's judgment as to what is "fit for purpose". If the results of an EQA scheme are normally distributed with a mean of x_a and variance of 1, then z-scores <2 would be deemed acceptable, whereas those >3 would not.

Toxicology EQA Schemes

Quantitative EQA schemes are available for a wide range of therapeutic drug monitoring (TDM) analytes and some other poisons in many countries. The European Network of Forensic Science Institutes (ENFSI) website (http://www.enfsi.eu/page.php?uid=93, accessed 1 October 2008) and European Proficiency Testing Information System (http://www.eptis.bam.de/, accessed 1 October 2008) list EQA schemes for a range of analytes including drugs that are available in Europe, America, and Australia. The United Kingdom National External Quality Assessment Scheme (UKNEQAS) for therapeutic drug assays and the Dutch KKGT (Association for Quality Assessment in TDM and Clinical Toxicology) schemes, for example, have been operating for over 25 years.

There are, of course sometimes, concerns such as the possible effects of freeze-drying on analyte stability, and issues concerning the matrix used to prepare material for circulation (due to the cost of analyte-free human plasma or serum, neonatal calf serum may be used for some analytes). Nevertheless, from the datasets generated, scheme organizers can ascertain the methods that give the best performance, investigate sources of interference or bias, and, in extreme cases, report poorly performing methods to regulatory authorities. In the 30 years since the inception of these schemes, poorly performing assays have been identified and participants advised accordingly. The mean values reported have moved nearer to the intended value and the spread of results about the mean has been reduced [9].

EQA of qualitative work is also invaluable, as in drugs of abuse screening. Here, poor selectivity due to immunoassay cross-reactivity or gas chromatography (GC) injection port artifacts, as well as to manipulation of the sample by the patient, are real problems. Even with the ongoing educative role of EQA, errors (such as failure to detect morphine at $1\,\text{mg}\,\text{l}^{-1}$ in urine using LC-MS-MS) still occur. Often human error is the cause of such mistakes.

Well-run qualitative EQA schemes not only assess routine performance, but also add in hard-to-detect compounds or "spike" at concentrations around nationally agreed "cutoff" values. Good screening and confirmatory analytical performance from laboratories is required to ensure that the correct result is reported. Scoring of performance for a set of commonly encountered drugs such as morphine, 6-monoacetylmorphine for heroin, and 2-ethylidene-1,5-dimethyl-3,3-diphenylpyrrolidine (EDDP) for

methadone, for example, helps laboratories address any problems in their procedures.

Urine is the matrix used most commonly, but schemes to support analyses in oral fluid and indeed in other matrices, such as hair, are emerging (e.g., that organized by the Society of Hair Testing, http://www.soht.org/, accessed 15 October 2007). For urine testing, scheme cutoffs are quite generous, but laboratories may also be scored against agreed workplace cutoffs (see http://www.wdtforum.org.uk/, accessed 27 August 2007). A breakdown of the performance of each recognized method type should be given to help laboratories understand why false-positive or false-negative reports have occurred in a given distribution. Meetings with users and scheme organizers are helpful to both groups. Regular clinical/forensic toxicology case schemes have also been instituted.

User Support and Staff Training

Training is an important part of QM. Clinicians, especially Accident and Emergency staff, forensic physicians, pathologists, and police need guidance not only as to what toxicological assays are useful in a given set of circumstances but also on sample collection, transport, and storage, local assay availability and turn-around time, and the interpretation of results. Especially important is guidance on blood and tissue sampling postmortem since the homogeneity of such samples cannot be taken for granted [10] and the possibility of changes in composition having occurred since death must be borne in mind when reporting results [11].

Not only do laboratory staff providing analytical toxicology services need training in providing these services and in the interpretation of results, they will also need some training in dealing with requests for tests that will need to be referred to other analytical centers. Staff in regional or national analytical centers will require extensive training in the more complicated analytical methods that they will be called upon to use. This training must encompass not only the techniques themselves (TLC, HPLC, etc.), but also their application in analytical toxicology and related areas. Knowledge of the role of local hospital laboratories, coroners, police, and the local poisons center is important [12].

There are no internationally recognized training programs in analytical toxicology. However, a training scheme for a graduate clinical scientist specializing in analytical toxicology has been developed in the UK. The training program comprises 4 years full-time study, followed by a period of higher specialist training, in some cases, leading to the award of a research degree such as Doctor of Philosophy. The American Board of Clinical Chemistry hosts an examination in Toxicological Chemistry (http://www.abclinchem.org/tox_chem/Pages/default.aspx, accessed 30 September 2008).

Participation in continuing education (CE), continuing professional development (CPD), or revalidation programs is important when staff attain career grades, i.e., when initial and higher specialist training has been completed, and may be necessary for continued specialist registration in countries where such registration is mandatory. Compliance with CPD programs may require maintenance and external audit of personal records listing educational activities, such as scientific meetings attended, papers published, lectures given, etc. Details of such a scheme maintained by the UK Royal College of Pathologists are available (http://www.rcpath.org/index.asp?PageID=620, accessed 27 August 2007). Many other countries run similar schemes.

Nongraduate scientific staff will be normally trained in-house in specific aspects of laboratory operation, although training in the operation of newer specialized instruments may sometimes be provided by manufacturers. Proper recording of training is important.

Recording and Reporting Results

It is usually advisable to contact the laboratory by telephone in advance to discuss urgent or complicated cases. A request form must accompany the samples to the laboratory. Most clinical specimens, particularly blood and urine, may be sent by post if securely packaged in compliance with current regulations. Filter-paper adsorbed dried blood may be acceptable in certain circumstances and does not require refrigeration [11].

In forensic work, it is important to be able to guarantee the identity and integrity of the specimen. Thus, such samples should be protected during transport by the use of tamper-evident seals and should, ideally,

Box 4. Chain of custody documents

- Name of the individual collecting the specimen
- Name of each person or entity subsequently having custody of it, and details of how it has been stored
- Date(s) the specimen was collected or transferred
- Specimen or case number
- Name of the subject or deceased
- Brief description of the specimen
- Record of the condition of tamper-evident seals

be submitted in person to the laboratory by the Coroner's officer or other investigating personnel. *Chain of custody* is a term used to refer to the process used to maintain and document the history of the specimen (Box 4).

Fully validated assays must include data on the stability of the analyte in the appropriate matrix under specified storage conditions. In the absence of other information, biological specimens should be stored at 2–8 °C prior to analysis, if possible, and ideally any specimen remaining after the analysis should be kept at 2–8 °C for 3–4 weeks in case further analyses are required. In forensic work, any specimen remaining must be kept (preferably at −20 °C) until destruction is authorized by the investigating authority.

All results should be recorded in laboratory notebooks or on worksheets together with information such as the date, the name of the analyst, the name of the patient, and other relevant information, the number and nature of the specimens received for analysis, and the tests performed. All specimens received in laboratories are normally allocated a unique identifying number. This number is used when referring to the tests performed on the specimen.

UV spectra, chromatograms, calibration graphs, and other documents generated during an analysis should always be kept for a time after the results have been reported. Recording the results of color tests and TLC analyses is more difficult unless a digital camera is used. Electronic data storage is dependent on the availability of hardware and software to read the data.

Reporting the results of tests in which no compounds were detected in plasma/serum or in urine, the limit of sensitivity of the test (LoD) should always be known, at least to the laboratory, and the scope of generic tests (benzodiazepines and opioids) should be defined.

The results of urgent (emergency) analyses must be communicated directly to the client without delay, and should be followed by a written report as soon as possible. Ideally, confirmation from a second independent method, or failing this, an independent duplicate, should be obtained before reporting positive findings. However, this may not always be practicable, especially if only simple methods are available or if sample amount is limiting. In such cases, it is vital that the appropriate positive and negative controls have been analyzed together with the specimen.

When reporting quantitative results, it is important to clearly state the units of measurement used. In addition, any information necessary to ensure that the clinical implications of the result are fully understood must be available and should also be noted on the written report.

Although it may be easy to interpret the results of analyses in which no compounds are detected, such results are sometimes difficult to convey to others, especially in writing. This is because it is important to give information as to the poisons excluded by the tests performed with all the attendant complications of the scope, sensitivity, and selectivity of the analyses and other factors such as sampling variations. Because of the potential medicolegal and other implications of any toxicological analysis, it is important not to use laboratory jargon such as "negative" or "not confirmed", or sweeping statements such as "absent" or "not present". The phrase "not detected" should precisely convey the laboratory result, especially when accompanied by a statement of the specimen analyzed and the LLoQ.

While chain-of-custody procedures for sample submission are widely implemented, secure systems for reporting results are also needed – it is not unknown for a report to be altered by laboratory staff in order to subvert an investigation.

Summary

Implementation of QM principles and associated laboratory accreditation procedures helps ensure that reliable results are obtained and can be defended in court. With the increasing use of point-of-care testing (POCT) devices, the principles of QM also have to be applied to such procedures in order to ensure the reliability of results. Although QM principles are more difficult to apply to the clinical interpretation of analytical results, it is vital that this aspect of the laboratory's work is not neglected.

References

[1] Burnett, D. (2002). *A Practical Guide To Accreditation In Laboratory Medicine*, ACB Venture Publications, London.

[2] SOFT/AAFS (Society of Forensic Toxicologists/ American Academy of Forensic Sciences) (2002). *Forensic Toxicology Laboratory Guidelines*, http://www.soft-tox.org/docs/Guidelines%202006%20Final.pdf (accessed 27 August 2007).

[3] Peters, F.T., Drummer, O.H. & Musshoff, F. (2007). Validation of new methods, *Forensic Science International* **165**, 216–224.

[4] FDA/CDER (Food and Drug Administration/Center for Drug Evaluation and Research) Guidance for industry (2001). *Bioanalytical Method Validation*, http://www.fda.gov/cder/guidance/4252fnl.htm (accessed 27 August 2007).

[5] Clinical and Laboratory Standards Institute (2004). *Evaluation of Precision Performance of Quantitative Measurement Methods: Approved Guideline*, 2nd Edition, Document EP05-A∼ 2, http://webstore.ansi.org/ansidocstore/product.asp?sku=EP05%2DA2 (accessed 27 August 2007).

[6] Sojo, L.E., Lum, G. & Chee, P. (2003). Internal standard signal suppression by co-eluting analyte in isotope dilution LC-ESI-MS, *The Analyst* **128**, 51–54.

[7] Peters, F.T., Jung, J., Kraemer, T. & Maurer, H.H. (2005). Fast, simple, and validated gas chromatographic-mass spectrometric assay for quantification of drugs relevant to diagnosis of brain death in human blood plasma samples, *Therapeutic Drug Monitoring* **27**, 334–344.

[8] Westgard, J.O., Barry, P.L., Hunt, M.R. & Groth, T. (1981). A multi-rule Shewhart chart for quality control in clinical chemistry, *Clinical Chemistry* **27**, 493–501.

[9] Wilson, J.F. (2002). External quality assessment schemes for toxicology, *Forensic Science International* **128**, 98–103.

[10] Flanagan, R.J. & Connally, G. (2005). Interpretation of analytical toxicology results in life and at post mortem, *Toxicological Reviews* **24**, 51–62.

[11] Flanagan, R.J., Connally, G. & Evans, J.M. (2005). Analytical toxicology: guidelines for sample collection post mortem, *Toxicological Reviews* **24**, 63–71.

[12] Flanagan, R.J. (2004). Developing an analytical toxicology service: Principles and guidance, *Toxicological Reviews* **23**, 251–263.

ROBERT J. FLANAGAN AND ROBIN WHELPTON

Wiley Encyclopedia of
FORENSIC SCIENCE

Volume 5
R–Z
INDEX

Editors-in-Chief

Allan Jamieson
The Forensic Institute, Glasgow, UK



Andre Moenssens
Forensics and Law Center, Columbia City, IN, USA



Radiocarbon Dating

Introduction

Radiocarbon dating is traditionally considered an archeological tool rather than a forensic one. Radiocarbon or carbon-14 (^{14}C) is produced naturally in the atmosphere by cosmic ray interactions with nitrogen. Single carbon atoms in the atmosphere are chemically reactive and are quickly oxidized to carbon dioxide (CO_2). The CO_2 from the atmosphere is incorporated into plants and works its way up the food chain to label every living thing with ^{14}C. The natural atmospheric concentration of natural ^{14}C with respect to all carbon has remained relatively stable at about 1.2 parts per trillion over the past several thousand years with the exception of the radiocarbon bomb-pulse over the past half-century (*see* **Bomb-Pulse Dating**). With a radioactive half-life of 5730 years, the radioactive decay of ^{14}C is minimal within the time periods of interest in most medical forensic cases and applicable for samples over 300 years of age. Willard Libby was awarded the Nobel Prize in Chemistry in 1960 for the development of radiocarbon dating [1].

Basics of Radiocarbon Dating

Radiocarbon dating relates the $^{14}C/C$ ratio in a sample to an extensive historical record to determine the age of a sample. The rate of radioactive decay of ^{14}C is relatively slow and predictable; hence, the concentration of ^{14}C changes slowly and reliably over time once a biological material dies and stops incorporating new carbon into its structure. The decrease in $^{14}C/C$ concentration from the contemporary value is used to determine the age of the sample. The natural production of ^{14}C does vary slightly over time. Extensive geological and archeological records are used to account for the variations in atmospheric ^{14}C concentration over the chronological range of radiocarbon dating (300–50 000 years before present) [2, 3]. Any sample greater than 50 000 years old has too little ^{14}C to measure accurately. There is some debate over the accuracy of the ^{14}C calibration between 40 000 and 50 000 years before present, but it is hard to conceive of a forensic sample of this age. Fossil carbon, either coal or petroleum, is older than 60 000 years and free of ^{14}C.

Measurement of Radiocarbon Samples

Today, most ^{14}C dating analyses are conducted using accelerator mass spectrometry (AMS), although some labs still use decay counting. AMS is much faster and generally more precise than decay counting since it measures differences in carbon atom mass and is not constrained to wait for atomic decay. AMS can also use smaller samples than decay counting, an important issue when analyzing evidence. There are about 100 AMS facilities worldwide.

Sample preparation and measurement details vary among AMS facilities, depending on the type of sample to be analyzed and the design of the spectrometer. Routine radiocarbon analyses using AMS are performed on samples containing about 1 mg carbon.

Samples as small as 50 μg carbon can be analyzed at some labs, but measurement uncertainties are larger. Nearly all AMS facilities that perform high-precision dating follow these general procedures to minimize contamination from outside sources of carbon and reduce measurement backgrounds. Samples are dried completely and then combusted with excess oxygen to produce CO_2. The CO_2 is purified to remove water vapor, nitrogen, oxides of nitrogen, and oxides of sulfur. It is then reduced to graphite or elemental carbon on metal catalysts, often cobalt or iron powder.

Primary standards, secondary standards, and backgrounds are similarly processed to produce graphite, which is the form of carbon analyzed by most AMS systems. Graphite is the preferred form of carbon because it can be made easily at high purity, produces intense negative ion currents, and can be prepared at satellite labs and shipped to AMS facilities for analysis. A handful of gas-accepting ion sources that take direct feed of CO_2 exist, but they are not typically used for high-precision dating. It is important to have consistent sample source material (e.g., all carbon graphite) because different molecules ionize with different efficiencies. Methods for producing graphite for elevated biological tracing experiments are generally unsuitable for radiocarbon dating because of larger sample-to-sample variation and higher background [4, 5].

The precision of radiocarbon dating depends on the ability to measure the ^{14}C concentration in a sample and the shape of the calibration curve. It is relatively easy to achieve 0.5–0.8% precision when analyzing relatively young, full-sized samples. This measurement precision translates to a chronological uncertainty of ±30–60 years in most samples less than 10 000 years old. Samples more than 25 000 years old can still be measured to 1% precision, but uncertainty propagation from backgrounds and the calibration curve typically yields uncertainties greater than ±100 years. The conventions for reporting radiocarbon dates are described by Stuiver and Polach [6].

Samples Amenable to Dating

Anything that was alive in the past is amenable to radiocarbon dating. Charcoal, wood, straw, hair, cloth, and bones are often dated in archeological sites. Historical documents and works of art are also routinely dated to confirm that the paper or canvas is of the appropriate age for the object.

Bone

Bone is the preferred sample matrix for dating human remains. Bone's ability to resist decay while containing a relatively high concentration of carbon makes it a desirable material for traditional dating. The carbon in collagen does turnover slowly while a person is alive; hence, the ^{14}C content is really a lifetime average rather than a snapshot in time. Traditional bone dating uses a collagen extraction procedure to avoid potential complications with mineral exchange of carbonates in bone in the environment. Collagen is a protein and is not affected by environmental carbonate exchange like the mineral component of bone.

Specific procedures for collagen extraction vary slightly among labs. In general, the mineral component of bone is dissolved in acid to free the collagen into solution. A variety of washing, rinsing, and filtering techniques are then applied to purify the collagen. Once purified, it can be combusted like any organic sample.

If human remains are found without any other evidence, radiocarbon analysis of bone collagen or hair can determine whether the authorities have a crime scene (*see* **Bomb-Pulse Dating**) or an archeological site. If the ^{14}C content of the collagen is elevated above the level in 1950, the person died sometime after 1955. The decrease in ^{14}C concentration due to decay can be measured over a 100 years, but the small differences in natural production between 1650 and 1955 and the decay combine to make it very difficult to separate samples chronologically over this time. Collagen before 1650 can be clearly distinguished.

In the United States, a ^{14}C date can determine the ownership of remains or precipitate a court battle over ownership, such as the case of Kennewick Man [7–10]. The Native American Graves Protection and Repatriation Act (NAGPRA) requires institutions that receive U.S. federal funding the return of human remains and sacred artifacts to ethnic descendants when they are available.

Documents and Art Objects

Radiocarbon dating of documents and art objects is routinely done by museums to confirm that the item's age is consistent with its provenance. This approach only confirms that the paper, cloth, or canvas is of the appropriate age, but it does not confirm authenticity.

A radiocarbon analysis that contains recent bomb-pulse carbon identifies a forgery of an older item.

Acknowledgment

This work was performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

References

[1] Libby, W.F., Anderson, E.C. & Arnold, J.R. (1949). Age determination by radiocarbon content – world-wide assay of natural radiocarbon, *Science* **109**, 227–228.

[2] Reimer, P.J., Baillie, M.G.L., Bard, E., Bayliss, A., Beck, J.W., Bertrand, C.J.H., Blackwell, P.G., Buck, C.E., Burr, G.S., Cutler, K.B., Damon, P.E., Edwards, R.L., Fairbanks, R.G., Friedrich, M., Guilderson, T.P., Hogg, A.G., Hughen, K.A., Kromer, B., McCormac, G., Manning, S., Ramsey, C.B., Reimer, R.W., Remmele, S., Southon, J.R., Stuiver, M., Talamo, S., Taylor, F.W., van der Plicht, J. & Weyhenmeyer, C.E. (2004). IntCal04 terrestrial radiocarbon age calibration, 0-26 cal kyr BP, *Radiocarbon* **46**, 1029–1058.

[3] Reimer, P.J., Baillie, M.G.L., McCormac, G., Reimer, R.W., Bard, E., Beck, J.W., Blackwell, P.G., Buck, C.E., Burr, G.S., Edwards, R.L., Friedrich, M., Guilderson, T.P., Manning, S., Guilderson, T.P., Southon, J.R., Hogg, A.G., Stuiver, M., Hughen, K.A., van der Plicht, J., Kromer, B., van der Plicht, J., Manning, S. & Weyhenmeyer, C.E. (2006). Comment on "Radiocarbon calibration curve spanning 0 to 50,000 years B.P. based on paired Th-230/U-234/U-238 and C-14 dates on pristine corals" by R.G. Fairbanks *et al.* (*Quaternary Science Reviews* (2005) **24**, 1781–1796) and "Extending the radiocarbon calibration beyond 26,000 years before present using fossil corals" by T.-C. Chin *et al.* (*Quaternary Science Reviews* (2005) **24** 1797–1808), *Quaternary Science Reviews* **25**, 855–862.

[4] Vogel, J.S. (1992). Rapid production of graphite without contamination for biomedical AMS, *Radiocarbon* **34**, 344–350.

[5] Ognibene, T.J., Bench, G., Vogel, J.S., Peaslee, G.F. & Murov, S. (2003). A high-throughput method for the conversion of CO_2 obtained from biochemical samples to graphite in septa-sealed vials for quantification of C-14 via accelerator mass spectrometry, *Analytical Chemistry* **75**, 2192–2196.

[6] Stuiver, M. & Polach, H.A. (1977). Discussion: reporting of ^{14}C data, *Radiocarbon* **19**, 355–363.

[7] Chatters, J.C. (2000). The recovery and first analysis of an early holocene human skeleton from kennewick, Washington, *American Antiquity* **65**, 291–316.

[8] Bruning, S.B. (2006). Complex legal legacies: the Native American Graves Protection and Repatriation Act, scientific study, and Kennewick Man, *American Antiquity* **71**, 501–521.

[9] Huxley, A.K. & Finnegan, M. (2004). Human remains sold to the highest bidder! A snapshot of the buying and selling of human skeletal remains on eBay(R), an Internet auction site, *Journal of Forensic Sciences* **49**, 17–20.

[10] Musselman, J. (2005). Ninth Circuit limits NAGPRA to remains linked with presently existing tribes, *Ecology Law Quarterly* **32**, 707–713.

BRUCE A. BUCHHOLZ

Radiology

Within months of Roentgen's announcement of his new discovery in 1895, the "X-ray" had been employed to examine a variety of objects including the human body, suspicious packages, Egyptian mummies, welded metal, and fine art, and had developed a whole range of other such uses throughout the medical and scientific world. Radiology was quickly adopted as a tool of the forensic sciences and was employed in a wide range of scientific investigations during the initial period of discovery. Some applications were quickly superseded, but many have developed into essential tools of the forensic investigator. Radiological techniques have many applications in forensic science today, including the examination of fingerprints and documents and a range of applications within the field of forensic engineering [1] (*see also* **Friction Ridge Examination (Fingerprints): Interpretation of**; **Interpretation: Document Evidence**; **Documents: Authentication of**; **Forged and Counterfeit Documents**).

Medical Radiological Techniques

Recently, radiological images were acquired using the same basic techniques of radiography and fluoroscopy that Roentgen himself employed. While these basic techniques still form a major part of radiological investigation, they have been supplemented by newer techniques employing sophisticated computer technology, some of which rely on X-rays

and others that employ radioactive materials, sound waves, and magnetic fields. The majority of methods now employed permit images to be acquired and stored digitally.

The main medical radiological imaging techniques may be summarized as follows:

- **Radiography**

The term *radiography* is generally used to describe a static or "still" image produced by means of a single exposure to X-rays (an energy form of ionizing radiation). Such images are now usually recorded and stored digitally but were previously recorded photographically on X-ray film.

- **Fluoroscopy**

The term *fluoroscopy* is used to describe an image directly visualized in real-time motion produced by a continuous exposure to X-rays.

- **Nuclear medicine (radionuclide imaging)**

Radioactive materials or isotopes can be used to highlight internal organs or tissues by injection, inhalation, or ingestion. The radiant energy escaping from the body is detected by sensitive phosphors to create images of these body parts.

- **Ultrasound**

Ultrasound scanning uses the same principle as sonar whereby sound waves generated by a transponder are reflected by internal structures in the body and measured and converted into real-time or static images.

- **Computed axial tomography**

Computed axial tomography (CAT) or CT scanning employs a computerized X-ray machine that uses an array of photoreceptors to detect minute differences in attenuation of X-rays emitted by an X-ray tube as it rotates around the body or body part. A computer generates an image of the body part in the axial or cross-sectional plane and modern multislice CT scanners permit multiplanar, multidirectional sectional images to be displayed and/or reconstructed as three-dimensional images.

- **Magnetic resonance imaging**

Magnetic resonance imaging (MRI) utilizes strong magnetic fields to generate electromagnetic signals from chemical elements and compounds within the bodily structures. Measurement and computer analysis of these signals allows multiplanar, multidirectional sectional images to be displayed and/or reconstructed as three-dimensional images in the same manner as CT scans.

- **Interventional radiology**

Using imaging techniques to guide them, the interventional radiologist uses a range of devices and instruments to enter the body *via* small puncture wounds, not only to produce images and make diagnoses but also to treat disease processes or correct anatomical abnormalities [2].

Clinical forensic radiology employs the entire spectrum of imaging techniques to the diagnosis of disease and injury, and the forensic radiological examination of live subjects may be undertaken with any one or more of the above techniques. Forensic radiology of deceased subjects has, until very recently, been limited to the use of still-image radiography and occasionally fluoroscopy [1]. Increasingly, however, some of the more complex imaging techniques, such as MRI or CT scanning, are being employed for postmortem imaging either as a precursor to invasive autopsy or, in some cases, in place of an invasive autopsy examination.

Medical Imaging Specialists

Following Roentgen's discovery, radiology is rapidly developed as a distinct specialism within medicine and is now undertaken by highly specialized medical and scientific staff.

Radiologist

The radiologist is traditionally described as a specialist physician dedicated to the study of images of the internal structures of the body, but now also increasingly involved in the treatment of disease through "interventional radiology", a form of "keyhole surgery" in which the treatment is guided by the use of real-time imaging techniques.

Radiographer or Radiologic Technologist

The radiographer is traditionally described as the technician who produces the images from which diagnoses are made but are now also increasingly undertaking diagnosis of injury or disease from these images.

Forensic Radiologists and Radiographers

As with many other branches of forensic science, forensic radiology and radiography is practised by professionals of both disciplines with experience of the application of their specialist skills within forensic medicine. The College of Radiographers (United Kingdom), in association with the Association of Forensic Radiographers (AFR), produces professional guidelines for the provision of forensic radiography services [3] and has supported the development of education and training in this field, with a number of post-graduate courses now being offered by UK universities. From January 2007, radiographers in the United Kingdom have been eligible for registration with the Council for Registration of Forensic Practitioners (CRFP)

The Forensic Uses of Medical Radiology

The use of medical radiology within forensic medicine falls into three categories:

- diagnosis
- identification and
- detection and retrieval of concealed or embedded objects [2].

Diagnosis

Radiology assists forensic medical investigation through diagnosis of the cause of injury or death by radiological means. Forensic radiological examination may be undertaken upon deceased subjects as part of a forensic pathology examination or upon living subjects within the clinical setting. Applications for radiology of the deceased follow the indications for forensic pathological examination in both accidental and criminal deaths: sudden death, sudden infant death, suspicious deaths, adverse medical events, road traffic accidents, terrorist incidents, war crimes, and genocide. Indications for examination of live subjects may include suspected abuse of children, the elderly or other vulnerable individuals, assault or criminal injury, industrial injury compensation, medical negligence, torture, other suspected human rights abuses, etc. [4, 5] (*see also* **Autopsy**; **Cardiac and Natural Causes of Sudden Death**).

Radiology of the Deceased

As previously indicated, the use of radiological methods in forensic pathology investigation has, until recently, been restricted to radiography and, in some cases, fluoroscopy. Over the past 5 years, however, a number of institutions, both in the United Kingdom and overseas, have been investigating the potential of modern imaging methods, such as the MRI and in particular CT scanning. Perhaps the most well known of these projects is the Virtopsy® group in Bern, Switzerland [6–11]. It is thus likely that an increasing number of radiological examinations of the deceased will be undertaken by CT or MRI scanning. Given the increasing pressure on forensic pathologists and an increasing awareness and respect within society of different cultural and religious beliefs surrounding death, it is possible that, in certain cases, such as sudden, nonsuspicious death, or accidental death the noninvasive "virtual" autopsy will, in time, replace the traditional dissection.

Accidental Deaths. Forensic pathology plays a significant role in the investigation of accidents as a cause of death. Individual demonstration of all fractures present after a fatal accident is a very time-consuming process and the routine X-ray examination simplifies the procedure. In addition, some fractures are difficult to demonstrate by postmortem dissection and are more adequately demonstrated using radiological methods such as X-ray or CT scanning [2].

X-ray diagnosis is not confined to the study of bones and useful information regarding soft-tissue changes can also be obtained. CT scanning with three-dimensional reconstruction in such cases permits documentation of hard and soft-tissue injuries without the requirement for a full invasive autopsy, and this may prove to be one area where CT scanning proves to be cost-effective.

In some cases, the injection of a radiological contrast medium may be beneficial to the forensic pathology examination. Contrast media can either be "positive", i.e., radiopaque (and thus absorb X-radiation to a greater degree than the surrounding tissues) or "negative", i.e., radiolucent (and thus absorb x-radiation to a lesser degree than the surrounding tissues).

For example, a traumatic subarachnoid hemorrhage arising from a lesion in one of the intracranial arteries may be difficult to demonstrate by dissection.

However, an X-ray examination following injection of an iodine or barium-based (positive) contrast medium into the arteries prior to dissection may provide proof of rupture. In the same way, radiology may be employed to detect the presence of air (a radiolucent "negative" contrast medium) in the thorax, the tissues surrounding the heart, the abdominal cavity, or as an embolus in the heart, brain, or vascular tree to assist the pathologist to determine the cause of death. For this reason, radiography and, in particular, CT scanning is an essential tool in the examination of hyperbaric accidental death in divers (*see also* **Traffic Fatalities**).

Homicidal Death. The most obvious use of X-rays in cases of suspected homicide is to demonstrate bullets and other missiles within the body. Projectiles may be found some considerable distance away from the entry site and radiography can be used to locate these. The deployment of a dynamic imaging method, such as fluoroscopy, will accelerate the process of retrieving ballistic material. Even if the missile is recovered without the aid of radiography, fragments of bullets or of bomb casings or contents can easily be missed in an unaided dissection and important information may be lost to the ballistics expert (*see also* **Gunshot Wounds**).

Terrorist explosions are often caused by uniquely constructed homemade devices. Location and retrieval of fragments from the victims of the attack is often of vital forensic importance and radiography is particularly indicated following such events. For this reason, a full radiological survey is recommended in all such cases in order to locate all projectiles and fragments. Any radiographs that have been taken for the purpose of investigation can be retained as permanent records and, if necessary, used as exhibits at trial [12–15] (*see also* **Explosions: Scene Investigation**).

Radiology is also invaluable in the investigation of large-scale killings such as suspected war crimes and human rights abuses. Radiology will assist in the location of ballistic material, the evaluation of gunshot and/or explosive injuries and the investigation and documentation of any antemortem trauma. Systematic application of radiology in such incidents will be essential [13] (*see also* **Mass Grave Investigation**).

Other forms of homicide are also well illustrated by way of X-rays. Of significance is the radiography of the larynx in cases of strangulation in which fractures of the thyroid cartilage or hyoid bone can be demonstrated, particularly if the cartilages are calcified, a process that increases with age.

Routine radiological examination of suspicious infant deaths is specifically indicated. X-ray examination in cases of suspected nonaccidental injury may yield far more than simply demonstrating the presence of fractures. Radiology may demonstrate the nature of the trauma that caused the fracture, for example, by severe twisting, thus differentiating it from trauma sustained following an accidental fall. It may also be of significance in proving the infliction of injury at various times by demonstrating healing and fresh fractures and such evidence is essential to the concept of systematic abuse. It may also indicate parental inattention to fractures sustained; callus formation at the fracture site is particularly marked when the fracture has not been properly treated by immobilization [2]. Demonstration of active bone replacement can be more accurately determined by using radionuclide imaging, in which the uptake of injected radioisotope-labeled phosphate by bone which is being actively formed can identify an active healing process [16]. Positive findings are not, however, confined to trauma and this should be combined with radiography to make a certain diagnosis. It is of course important to rule out other possible alternative causes of multiple fractures and radiological evidence is essential to this process.

The most common cause of death from child abuse is trauma to the head [17]. Intracranial injury is frequently associated with abusive head trauma and includes subarachnoid and subdural hemorrhage, intracerebral and intracereballar hemorrhage, and massive edema. In addition, intracranial injury and, in particular, subdural hemorrhage is associated with violent shaking of an infant (the so-called shaken baby syndrome), even in the absence of direct trauma to the skull. Such injuries are best demonstrated on CT or MRI scanning and these techniques are recommended in such cases (*see also* **Shaken Baby Syndrome**; **Battered Child Syndrome**).

Examination of Live Subjects

Physical Abuse. Radiological investigation of abuse described above is of course not restricted to the examination of deceased subjects or indeed solely to the examination of children but is routinely

employed in the investigation of all forms of physical abuse of vulnerable adults, e.g., elder abuse, spousal abuse, etc (*see also* **Battered Spouse Syndrome**; **Elder Abuse: Policy**).

Accidental and Criminal Injury. Radiological examination is likely to have been undertaken in the initial diagnosis and management of cases of accidental and criminal injury that later become the subject of legal action. Such cases include road traffic accidents, assault, negligence, etc., and images and radiological reports will thus have medicolegal significance.

Human Identification

Radiology is extremely valuable in the analysis and identification of unidentified human remains and can be used in two main ways:

- deductive identification and
- determination of an individual's personal identity *via* comparative methods.

Deductive Identification

Radiology has the advantage of enabling the examination of remains in a variety of states of decomposition from fully fleshed to completely skeletonized. As such, it affords the opportunity to obtain a considerable amount of data without the need to clean and completely deflesh the remains and thus provides the investigative team with a rapid method of triage and classification by answering a number of fundamental questions:

- determination of human *versus* nonhuman remains;
- recognition of commingling of remains of more than one individual;
- evaluation of the biological profile (age, sex, stature, and ancestry); and
- recognition of embedded/hidden foreign objects and personal effects.

- **Determination of human *versus* nonhuman remains**

Radiographic examination of the bone structure and trabecular pattern can be useful in determining human from nonhuman remains [18, 19] (*see also* **Postmortem Interval: Anthropology**).

- **Recognition of commingling**

In cases involving large amounts of fragmented remains, such as a mass fatality incident, remains may be commingled and mixed with large amounts of debris and artifacts. Physical examination of such remains, particularly in cases of fire damage where there is a uniformity of discoloration of all samples retrieved, is both difficult and time consuming. In such cases, radiological examination can prove useful both in determining the presence of one or more individuals and also in identifying and locating small body parts, especially teeth, which may otherwise be overlooked in the absence of a thorough and time-consuming fingertip search [13, 14, 20] (*see also* **Length Measurement**).

- **Age estimation**

The age of a cadaver at death is particularly well estimated radiologically. In infancy and childhood, the presence of centers of ossification can be demonstrated with accuracy and the union of these centers can be observed up to approximately 25 years of age. In clinical forensic medicine, a radiograph of the wrist of a live subject can be used to determine the "bone age" of individuals claiming to be under the age of majority, either for reasons of seeking asylum or in cases where more lenient sentencing would apply in criminal cases. In later life, degenerative changes may give some indication of maturity, but analysis will only provide a wide age range. The main value of X-rays will thus be to distinguish between two or three bodies who are known as a group but are of unknown individual identity [18, 21, 22] (*see also* **Sex Determination of Remains**).

- **Sex determination**

Differentiation of sexes by skeletal radiology is unreliable until after puberty, as the features that distinguish male from female are not sufficiently developed until this point. In the case of the remains of an individual that has reached puberty, the sex can be established with some certainty and without the need for extensive cleaning processes if suitable bones are available for X-ray. The pelvis and skull are the most useful bones in this respect [18, 23, 24] (*see also* **Species Determination of Osseous Remains**).

- **Stature estimation**

Physical anthropologists are able to make estimations of stature by direct measurement from unfleshed human remains. The length of the femur is usually

used, as this has been shown to be reliable In the case of fleshed remains, the same measurements can be made radiographically, provided that correction for magnification is made [18, 25] (*see also* **Anthropology: Age Determination of Remains**).

- **Determination of ancestry**

Determination of ancestry is challenging, especially when the remains are badly decomposed or skeletonized. Similar methods to those used by physical anthropologists to determine ancestry from skeletal remains can be applied radiographically with fleshed remains. In particular, examination of the skull and mandible the distal end of the femur and the ratio of long-bone length can be useful in the determination of population ancestry [18, 26].

Comparative Identification

Radiology has long been used for the identification of human remains and is well documented. Radiological identification of human remains requires specific and unique findings on postmortem images to be matched with antemortem images of the individual. In some cases, identification can be made from a series of relatively common or nonspecific pathological anatomical changes that appear in identical locations in ante- and postmortem images. In other cases, a single unique feature is sufficient [9, 27–36]. Provided that radiological images taken in life are available, any similar radiograph taken after death can be used for comparison. Radiographs taken for medical purposes are often required to be retained for long periods. With the advent of digital imaging and the decreasing cost of digital storage, many institutions are retaining medical images far beyond their previously applied practice for X-ray film. Records are thus, on the whole, fairly accessible.

Of the structures traditionally used for comparison, the skull, sinuses and the frontal sinues, in particular, are the most useful. The number of variations in shape and size that are potentially available makes a radiographic outline almost unique to the individual. Radiographic examinations of the skull have declined dramatically since the advent of CT scanning, and antemortem radiographic data is thus less likely to be available. However, positive identification can be established by CT, either by comparison with other CT scans or some conventional radiographs [30, 37, 38], although this may prove far more complex and costly to achieve.

Abnormalities that may be useful are numerous; while some congenital abnormalities such as absence of bones, displacement, or malformation may be so rare as to be diagnostic of identity, others, such as fusion of the ribs, are not uncommon and should only be regarded as confirmatory of other circumstantial evidence. Pre-existing disease in the form of changes in bone density, arthritic deformities, etc., may provide evidence of identity if antemortem images are available for comparison. In the same way, unusual calcification in the lungs or lymph nodes or due to degenerative changes in many organs may show a pattern as individual as is that of a fingerprint. The presence of healing fractures in bones known to have been broken in life provides strong evidence of identification as does the presence of surgical prostheses or of supportive implants such as plates, pins, or orthopedic screws. Not only can the X-rays be compared but also the appliance can be removed from the cadaver and superimposed [27, 29, 31, 34] (*see also* **Identification of Human Remains**).

Radiology is particularly valuable for identification following mass disasters, and the routine use of radiological screening in such incidents can lead to rapid identification of individuals known to have had previous fractures, surgery, surgical prostheses, or known anatomical variants. Radiological screening will also facilitate the rapid localization of personal possessions, either concealed within clothing or embedded in charred tissues. It will also enable items of potential forensic significance (e.g., projectile fragments) to be located and retrieved and injuries sustained as a result of explosion to be documented [11, 13–15, 20, 39, 40] (*see also* **Biometric Devices**).

Dental identification of both the single unknown body and in the mass casualty situation is of great importance and radiography very greatly enhances the contribution of the odontologist. The presence of a single or double restoration in one or two commonly filled teeth would be insufficient to prove a positive identification as opposed to an exclusion of identification. The precise shape of the fillings is, however, likely to be unique and identification can be made from a comparison of post- and antemortem radiographs of a single filling. Dental radiography may also show up significant root shapes, socket outlines or abnormalities of tooth eruption, development, or decay [31, 41] (*see also* **Odontology**).

Other Methods

In some difficult cases, facial reconstruction may prove useful in creating a likeness of the individual that stimulates a response to a public appeal for information. Traditionally, reconstruction methods have relied on the skills of the forensic artist to model the features using clay. Computerized reconstruction using CT scanning methods has been proposed as an alternative [42] (*see also* **Facial Reconstruction**).

Detection and Retrieval of Concealed or Embedded objects

In addition to the main applications of diagnosis and identification, radiology is also regularly employed to detect and retrieve objects concealed in, on, or about the person as well as objects that have become embedded in the tissues as a result of a penetrating injury, which may have additional forensic interest to the investigator.

An example of such an application is the retrieval of shrapnel from the subcutaneous tissues of victims of a terrorist bomb attack. Radiology in such circumstances is employed not to diagnose but to guide the treatment of the patient and ultimately retrieval of fragments that may be of forensic significance to the investigation.

Smugglers and, in particular, narcotics traffickers often make use of "body packers" or "mules" to smuggle prohibited goods across borders in packages concealed in the carrier's rectum, vagina or alimentary canal. Radiology is regularly employed to detect packages of narcotics in the alimentary canal that have been swallowed by the carrier. A radiograph of the abdomen has traditionally been employed in such circumstances, but increasing sophistication on the part of smugglers has rendered packages more difficult to detect *via* this method and CT scanning may therefore be used in such cases [43, 44].

Whatever technique is employed, such examinations of the live subject should not be undertaken without the consent of the individual [3], subject to the usual exigencies of a life-threatening medical emergency that renders the patient unconscious and unable to consent. Such emergencies may follow the rupture of one or more of the packages, releasing a high dose of the drug within the alimentary canal.

References

[1] Brogdon, B.G. & Lichtenstein, J.E. (1998). Forensic radiology in historical perspective, in *Forensic Radiology*, B.G. Brogdon, ed, CRC Press, Boca Raton, pp. 13–34.

[2] Viner, M.D. (2008). *Mason's Forensic Medicine For Lawyers*, 5th Edition, S. Cowan & A.C. Hunt, eds, Tottel Publishing, Haywards Heath, pp. 428–444.

[3] College of Radiographers & Association of Forensic Radiographers (2008). *Guidance for Radiographers Providing Forensic Radiography Services*, Society & College of Radiographers, London.

[4] Viner, M.D. (2001). Forensic investigation: the role of radiography, *European Radiology* **11**(Suppl. 2), 95.

[5] Viner, M.D. (2001). Forensic investigation: the role of radiography in forensic medicine, *ISRRT Newsletter* **37**(2), 4–7.

[6] Thali, M.J., Yen, K., Schweitzer, W., Vock, P., Boesch, C., Ozdoba, C., Schroth, G., Ith, M., Sonnenschein, M., Doernhoefer, T., Scheurer, E., Plattner, T. & Dirnhofer, R. (2003). Virtopsy, a new imaging horizon in forensic pathology: virtual autopsy by postmortem multislice computed tomography (MSCT) and magnetic resonance imaging (MRI) – a feasibility study, *Journal of Forensic Sciences* **48**(2), 386–403.

[7] Thali, M.J., Yen, K., Schweitzer, W., Vock, P., Ozdoba, C. & Dirnhofer, R. (2003). Into the decomposed body-forensic digital autopsy using multislice-computed tomography, *Forensic Science International* **134**(2–3), 109–114.

[8] Brookes, J.A., Hall-Craggs, M.A., Sams, V.R. & Lees, W.R. (1996). Non-invasive perinatal necropsy by magnetic resonance imaging, *Lancet* **348**(9035), 1139–1141.

[9] Riepert, T., Rittner, C., Ulmcke, D., Ogbuihi, S. & Schweden, F. (1995). Identification of an unknown corpse by means of computed tomography (CT) of the lumbar spine, *Journal of Forensic Sciences* **40**(1), 126–127.

[10] Rocha Sdos, S., Ramos, D.L. & Cavalcanti, G. (2003). Mde applicability of 3D-CT facial reconstruction for forensic individual identification, *Pesquisa Odontologica Brasileira* **17**(1), 24–28.

[11] Rutty, G.N., Robinson, C.E., BouHaidar, R., Jeffrey, A.J. & Morgan, B. (2007). The role of mobile computed tomography in mass fatality incidents, *Journal of Forensic Science* **52**(6), 1343–1349.

[12] Viner, M.D., Rock, C., Hunt, N., Mackinnon, G. & Martin, A.W. (2006). Forensic radiography: response to the London suicide bombings on 7th july 2005, *Paper presented at the American Academy of Forensic Science 58th Scientific Meeting*, Seattle.

[13] Viner, M.D. (2008). The use of radiology in mass fatality incidents, in *Recovery, Analysis and Identification of Comingled Human Remains*, B. Adams & J. Byrd, eds, Humana Press, Totowa, pp. 145–179.

[14] Lichtenstein, J.E. (1998). Radiology in mass casualty situations, in *Forensic Radiology*, B.G. Brogdon, ed, CRC Press, Boca Raton, pp. 189–208.
[15] Kahana, T., Ravioli, J.A., Urroz, C.L. & Hiss, J. (1997). Radiographic identification of fragmentary human remains from a mass disaster, *The American Journal of Forensic Medicine and Pathology* **18**(1), 40–44.
[16] Sty, J.R. & Starshak, R.J. (1983). The role of bone scintigraphy in the evaluation of the suspected abused child, *Radiology* **146**, 369.
[17] Reece, R.M. (1994). *Child Abuse, Medical Diagnosis and Management*, Lea & Febiger, Philadelphia, Chapter 1.
[18] Brogdon, B.G. (1998). Radiological identification: anthropological parameters, in *Forensic Radiology*, B.G. Brogdon, ed, CRC Press, Boca Raton, pp. 63–96.
[19] Chilvarquer, I., Katz, J.O., Glassman, D.M., Prihoda, T.J. & Cottone, J.A. (1987). Comparative radiographic study of human and animal long bone patterns, *Journal of Forensic Sciences* **32**(6), 1645–1654.
[20] Mulligan, M.E., McCarthy, M.J., Wippold, F.J., Lichtenstein, J.E. & Wagner, G.N. (1988). Radiologic evaluation of mass casualty victims: lessons from the Gander, Newfoundland, accident, *Radiology* **168**(1), 229–233.
[21] Greulich, W.W. & Pyle, S.I. (1959). *Radiographic Atlas of Skeletal Development of the Hand and Wrist*, Stanford University Press, Stanford.
[22] Hansman, C.F. (1962). Appearance and fusion of ossification centers in the human skeleton, *The American Journal of Roentgenology, Radium Therapy, and Nuclear Medicine* **88**, 476–482.
[23] Kurihara, Y., Kurihara, Y., Ohashi, K., Kitagawa, A., Miyasaka, M., Okamoto, E. & Ishikawa, T. (1996). Radiologic evidence of sex differences: is the patient a woman or a man? *AJR. American Journal of Roentgenology* **167**(4), 1037–1040.
[24] Rogers, T. & Saunders, S. (1994). Accuracy of sex determination using morphological traits of the human pelvis, *Journal of Forensic Sciences* **39**(4), 1047–1056.
[25] Aitken, A.G., Flodmark, O., Newman, D.E., Kilcoyne, R.F., Shuman, W.P. & Mack, L.A. (1985). Leg length determination by CT digital radiography, *AJR. American Journal of Roentgenology* **144**(3), 613–615.
[26] Craig, E.A. (1995). Intercondylar shelf angle: a new method to determine race from the distal femur, *Journal of Forensic Sciences* **40**(5), 777–782.
[27] Brogdon, B.G. (1998). Radiological identification of individual remains, in *Forensic Radiology*, B.G. Brogdon, ed, CRC Press, Boca Raton, pp. 149–187.
[28] Binda, M., Cattaneo, C., Bogoni, A., Fattorini, P. & Grandi, M. (1999). Identification of human skeletal remains: forensic radiology vs. DNA, *Radiologia Medica* **97**(5), 409–411.
[29] Buchner, A. (1985). The identification of human remains, *International Dental Journal* **35**(4), 307–311.
[30] Culbert, W.L. & Law, F.M. (1927). Identification by comparison of roentgenograms of nasal accessory sinuses and mastoid processes, *JAMA : The Journal of the American Medical Association* **88**, 1634–1636.
[31] Fischman, S.L. (1985). The use of medical and dental radiographs in identification, *International Dental Journal* **35**(4), 301–306.
[32] Haglund, W.D. & Fligner, C.L. (1993). Confirmation of human identification using computerized tomography (CT), *Journal of Forensic Sciences* **38**(3), 708–712.
[33] Jensen, S. (1991). Identification of human remains lacking skull and teeth. A case report with some methodological considerations, *The American Journal of Forensic Medicine and Pathology* **12**(2), 93–97.
[34] Kahana, T. & Hiss, J. (1997). Identification of human remains: forensic radiology, *Journal of Clinical Forensic Medicine* **4**(1), 7–15.
[35] Murphy, W.A., Spruill, F.G. & Gantner, G.E. (1980). Radiologic identification of unknown human remains, *Journal of Forensic Sciences* **25**(4), 727–735.
[36] Sanders, I., Woesner, M.E., Ferguson, R.A. & Noguchi, T.T. (1972). A new application of forensic radiology: identification of deceased from a single clavicle, *The American Journal of Roentgenology, Radium Therapy, and Nuclear Medicine* **115**(3), 619–622.
[37] Nambiar, P., Naidu, M.D. & Subramaniam, K. (1999). Anatomical variability of the frontal sinuses and their application in forensic identification, *Clinical Anatomy* **12**(1), 16–19.
[38] Reichs, K. & Dorion, R.B.J. (1992). The use of computed tomography (CT) scans in the analysis of frontal sinus configuration, *Journal of the Canadian Society of Forensic Science* **25**(1), 1–6.
[39] Harcke, H.T., Bifano, J.A. & Koeller, K.K. (2002). Forensic radiology: response to the pentagon attack on september 11, 2001, *Radiology* **223**(1), 7–8.
[40] Society of radiographers radiographers help identify London bombing victims, *Synergy* (2005). September, 1.
[41] Jackowski, C., Aghayev, E., Sonnenschein, M., Dirnhofer, R. & Thali, M.J. (2006). Maximum intensity projection of cranial computed tomography data for dental identification, *International Journal of Legal Medicine* **120**(3), 165–167.
[42] Myers, J.C., Okoye, M.I., Kiple, D., Kimmerle, E.H. & Reinhard, K.J. (1999). Three-dimensional (3-D) imaging in post-mortem examinations: elucidation and identification of cranial and facial fractures in victims of homicide utilizing 3-D computerized imaging reconstruction techniques, *International Journal of Legal Medicine* **113**(1), 33–37.
[43] Beernan, R., Nunez Jr, D. & NET, C.V. (1986). Radiographic evaluation of the conaine smuggler, *Gastrointestinal Radiology* **11**, 351.
[44] Marc, B., Baud, F.J., Aelion, M.J., Gherardi, R., Diamant-Berger, O., Blery, M. & Bismuth, C. (1990). The cocaine body-packer syndrome: evaluation of a method of contrast study of the bowel, *Journal of Forensic Sciences* **35**, 345.

MARK D. VINER

Rampage Killer ***see*** Homicide: Multiple (Behavior)

Rape ***see*** Child Sexual Abuse Accommodation

Rape Trauma Syndrome

Rape trauma syndrome (RTS) is a cluster of common reactions experienced by victims of rape. The term was first used by Burgess [1], a psychiatric nurse, and Holmstrom, a sociologist, to describe the great uniformity of responses they observed among 92 survivors of rape. They divided RTS into two phases: an acute phase characterized by intense symptoms experienced in the days or weeks immediately following a rape and a reorganization phase of more moderate disturbances in the general functioning that may last for years. Subsequent research has conceptualized RTS more in terms of specific symptoms rather than general stages of recovery. RTS is also often used to describe a specific type of posttraumatic stress disorder (PTSD) (*see* **Posttraumatic Stress Disorder**).

Numerous symptoms have been reported in connection with the RTS. The most common symptoms experienced immediately after the rape are the fear and anxiety. Physiological manifestations of fear and anxiety include shaking or trembling, a racing heart, pain, tight muscles, rapid breathing, and numbness. Many survivors report feelings of denial, shock, and disbelief. The experience is often relived in intrusive thoughts and nightmares. Changes in sleeping and eating patterns are typical. Some appear to be confused and disoriented while others may be dazed and numb. The common symptoms are depression, decreased self-esteem, guilt, and self-blame. Hostility and aggression have been observed. Self-doubt and helplessness can replace previous feelings of competence and autonomy. Distorted perceptions such as distrust, pessimism, and paranoia have also been reported.

There are also a variety of more long-term reactions. Many victims develop phobias of things associated with the rape such as being alone or going out at night. Significant changes in the lifestyle such as quitting a job or moving to another residence is not unusual. Disturbances in the general functioning are often manifest by an inability to carry out the routine aspects of life or engage in social activities. Rape also tends to negatively impact the victim's enjoyment of intimate relationships and sex.

Since RTS has multiple connotations, there is no consensus regarding the number, severity, or combination of symptoms required for a diagnosis. The Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-R) American Psychiatric Association [2] does not define RTS as a disorder; therefore, PTSD is more commonly assessed after a rape. The majority of rape victims meet the criteria for PTSD shortly after a rape and about 50% after one year [3]. To meet the diagnostic criteria for PTSD a person must have experienced, witnessed, or have been confronted with a traumatic event that involved actual or threatened death or serious injury – or threat to the physical integrity of self or others – and the person's response must have involved intense fear, helplessness, or horror. Furthermore, the person must persistently re-experience the event, persistently avoid stimuli associated with the trauma and have a numbing of general responsiveness, and display persistent symptoms of increased arousal. Although there is overlap between these criteria and the aftereffects of rape, PTSD and RTS are not necessarily synonymous. Some are critical that PTSD was constructed primarily in response to the studies of adult, male war veterans and does not account for symptoms such as depression, guilt, anger, humiliation, and sexual dysfunction that are common among the rape survivors.

Applications within the Legal System

In addition to assessment and treatment considerations, an understanding of the RTS can be useful in a forensic context. Expert testimony can be used to educate a jury about rape myths, or incorrect assumptions about the causes and consequences of rape (e.g., women cannot be raped against their will,

women secretly wish to be raped, and most accusations of rape are fake). An expert can testify that the victim's psychological reactions are consistent with having been raped if the defendant claims the sex was consensual. RTS can also help explain seemingly inconsistent behavior for a rape victim (e.g., delay in reporting the attack and lack of memory). In civil suits, RTS testimony can be used to support a claim of damages. Finally, it can even be used as a defense for culpable behavior from a rape survivor (e.g., attempted murder of attacker) [4].

Court rulings concerning the admissibility of testimony on RTS and the impact of rape on victims have not been consistent. When it is admitted, there are often strict limitations on its use. Additionally, some appellate courts have reversed trial court decisions regarding RTS testimony. As of 2004, only 20 states allowed medical evidence of RTS when the issue is whether the sex was consensual [3]. There are three major issues that are examined by the courts regarding RTS testimony. First, is whether the testimony is probative or helpful to the jury. Second, is the scientific reliability of RTS (e.g., *Daubert* standard) (*see* ***Daubert v. Merrell Dow Pharmaceuticals***). Finally, and most critical, is the potential prejudicial impact of RTS testimony [5].

Court rulings have fallen on both sides of all the three of these issues. Testimony that describes common reactions to rape and general diagnostic criteria for RTS or PTSD is usually allowed. Testimony regarding whether the victim has RTS or PTSD, or that their behavior is consistent with a diagnosis, is sometimes allowed under certain conditions. Testimony that goes beyond a diagnosis and states that the defendant is telling the truth and was raped is almost never allowed. In general, the term *PTSD* is considered more acceptable in court than RTS because it is included in the DSM-IV-R and does not necessarily imply that a rape occurred [6].

References

[1] Burgess, A.W. & Holmstrom, L.L. (1974). Rape trauma syndrome, *American Journal of Psychiatry* **131**, 981–999.

[2] American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition, text revision (DSM IV-R), American Psychiatric Association, Washington, DC.

[3] Cling, B.J. (ed) (2004). Rape and rape trauma syndrome, *Sexualized Violence Against Women and Children*, Guilford, New York, pp. 13–40.

[4] Wrightman, L.S. & Fulero, S.M. (eds) (2005). Syndrome evidence: battered woman syndrome and rape trauma syndrome, *Forensic Psychology*, 2nd Edition, Wadsworth, Belmont, pp. 139–169.

[5] Frazier, P.A. & Borgida, E. (1992). Rape trauma syndrome: a review of case law and psychological research, *Law and Human Behavior* **16**, 293–311.

[6] Boeschen, L.E., Sales, B.D. & Koss, M.P. (1998). Rape trauma experts in the courtroom, *Psychology, Public Policy, and Law* **4**, 414–432.

Related Articles

Battered Spouse Syndrome

Expert Opinion: United States

Syndromes: Psychological

CARL CLEGG, KATRINA MCCOY AND WILLIAM J. FREMOUW

Rational Choice Perspective ***see*** Crime Victims' Decision to Report Crime

Reasonable Suspicion ***see*** Profiles: Psychological and Behavioral

Recollective Accuracy of Posttraumatic Stress Disorder ***see*** Recollective Accuracy of Traumatic Memories

Recollective Accuracy of Traumatic Memories

Canonical traumatic events are life-threatening experiences that produce overwhelming terror in victims who experience them. The memory of such an experience may generate symptoms of posttraumatic stress disorder (PTSD) in some people [1]. Traumatic stressors differ from ordinary stressors in terms of their capacity for threatening life and limb and for inciting extreme fear. Accordingly, researchers have investigated whether memory for trauma differs from memory for other experiences.

By definition, traumatic memories differ from nontraumatic memories in one sense: the former are about trauma, and the latter are not. But do they differ in other, nontrivial ways? For example, are traumatic memories especially likely to be accurate? Contrary to a popular view, memory does not operate like a videotape machine, faithfully storing the sensory impressions of our lives, enabling us to replay the tape upon recollection.

Rather, when people recall an episode from their lives, they reconstruct the memory from encoded elements distributed throughout the brain [2]. The reconstructive character of autobiographical memory ensures that whatever is recalled will inevitably be incomplete and possibly inaccurate. That is, not only will certain aspects of the experience be omitted during recall, but the person may inadvertently remember things differently than how they occurred. In the extreme case, a person may have a vivid "memory" for something that never occurred at all.

The aforementioned facts about memory pose problems for ascertaining the accuracy of memory, traumatic or otherwise. Only in unusual circumstances are veridical records of a trauma available that might serve as a gold standard against which to compare a person's memory. For example, robberies of stores and banks are sometimes videotaped by monitors, thereby permitting evaluation of eyewitness accounts of the event. In other cases, archival records sometimes enable researchers to gauge the accuracy of memories of war trauma [3]. In any event, for memories of trauma, corroborated by independent evidence that the trauma occurred, victims almost always recall their experience [4, 5]. This implies that victims retain, at the very least, memory for the essence or gist of what happened even when some details are altered or omitted [6].

Studies designed to ascertain the accuracy of recollections of trauma usually involve comparing later recall with previous accounts of the event provided by victims. Strictly speaking, these studies pertain to the consistency of traumatic memories overtime rather than to their accuracy *per se*. That is, there is seldom an infallible record of what happened to the person during the trauma, and so lacking such a "gold standard", psychologists have had to compare the testimony of victims long after the trauma with their testimony shortly after the trauma. For example, Wagenaar and Groeneweg [7] found that concentration camp survivors had remarkably good recollection of their traumatic experiences 40 years after they had been released from the concentration camp, relative to their accounts provided shortly after their release.

In an important longitudinal study, Porter and Peace [8] had subjects describe a recent traumatic memory and a recent very positive memory. Subjects rated the memories in terms of vividness, quality, and sensory components. The researchers recontacted subjects between 3.45 and 5 years later, and they found that memories of trauma are highly consistent over time relative to positive memories. Vividness, quality, and sensory features remained high for trauma memories, but declined sharply for positive memories. Severity of stress-related symptoms was unrelated to memory consistency.

Mechanisms of Traumatic Memory

What characterizes traumatic memories and do they differ from memories of neutral, or even positive, events? Several scholars claim that memories of traumatic events have special properties that distinguish them from ordinary memories. By this view, traumatic memories are *qualitatively* different (i.e., processed and stored differently) from other types of memories, thereby involving mechanisms different from those associated with general memory functioning [9].

This view asserts that many survivors of a trauma invoke mechanisms such as repression and dissociation, which result in traumatic amnesia for the stressful event itself. Moreover, it is argued that survivors of a trauma suffer from intrusions with strong

sensory qualities. There are several versions of this theoretical stance [10, 11], but the core assumption they have in common is that trauma has a *special* impact on the way in which memories of traumatic events are organized.

In contrast, other scholars hold that traumatic events do not impair memory for the central gist of the experience (for a discussion, see [12–14]). In fact, they can even enhance the quality of a memory. That is, traumatic events are remembered quite well [4]. Suffice to say, how people remember and forget trauma has been a highly controversial issue in psychiatry and psychology [4, 15].

Several studies on personally involving traumatic events provide evidence *against* the notion that such experiences are poorly remembered. For example, Kuch and Cox [16] studied 124 holocaust survivors and found that traumatic amnesia, with an estimated lifetime prevalence rate of 3%, was quite rare in this group. Likewise, Merckelbach *et al.* [17] found in a group of 29 Dutch concentration camp survivors only one survivor reporting mnemonic experiences that might be taken as evidence for traumatic amnesia. The authors noted that in this case there was a possibility that drug abuse contributed to the poor memory of the traumatic episode. Similarly, Geraerts *et al.* [18] found that in a sample of Croatian war veterans who had been confronted with extremely aversive events during the Balkan wars, traumatic amnesia was rarely reported. It is unclear, however, whether "amnesia" in these studies refers to an inability to recall an encoded aspect of the trauma or whether it signifies a failure to encode certain aspects of the event in the first place.

Porter and coworkers conducted several studies comparing traumatic memory characteristics with memory for other emotional experiences in life. For example, Porter and Birt [19] asked undergraduates to give detailed accounts of their most traumatic memory as well as a description of a positive emotional event, and to rate the number of details in the narrative among other characteristics such as vividness of the memories. They found that the traumatic memories were associated with greater vividness and detail in memory relative to positive memories. Relatedly, Peace *et al.* [20] found that in a sample of survivors of sexual abuse, memories of the abuse were not impaired or fragmented relative to memories related to a nonsexual trauma, or to memories related to a positive event. Instead, memories for sexual trauma were associated with a remarkably high level of vividness and detail.

Confusions about the Clinical Science

These studies indicate that victims tend to remember their traumatic experiences very well. Some clinical theorists, however, believe that a significant minority of victims may dissociate, repress, or otherwise be incapable of recalling their trauma, except under very special circumstances (e.g., psychotherapy; [21]). Unfortunately, these theorists often misunderstand the very studies they cite in support of the alleged phenomenon of repressed memory of trauma. In the following section, we briefly review the most common misunderstandings. For detailed critiques, see [4, 22]

Confusing Everyday Forgetfulness with Traumatic Amnesia

After having experienced a traumatic event, some people report difficulty concentrating and remembering things in everyday life [23]. This kind of memory impairment, however, refers to everyday forgetfulness that develops following the trauma; it does not refer to difficulty remembering the trauma itself.

Confusing Psychogenic Amnesia with Traumatic Amnesia

The term *psychogenic amnesia* is sometimes used as a synonym for traumatic amnesia. They are, however, distinct phenomena. Typical cases of psychogenic amnesia are characterized by an abrupt, immense retrograde memory loss, including loss of personal identity that cannot be attributed to a physical insult to the brain [24]. The syndrome is sometimes preceded by exposure to a stressful event, but these events are often ordinary stressors of uncertain etiologic significance (e.g., difficulties with one's job or marriage). The term *psychogenic* implies the absence of an obvious organic cause rather than an identified psychological one. Most cases of psychogenic amnesia remit within hours, days, or weeks, and often without any therapeutic intervention. With alleged traumatic amnesia, a person is unable to recall a specific traumatic event rather than being entirely unable to recall his or her past.

Confusing Organic Amnesia with Traumatic Amnesia

Clinical theorists occasionally confuse organic amnesia resulting from direct damage to the brain with amnesia resulting from psychic causes. For example, according to Brown *et al.* [25], "Dollinger [26] found that 2 of the 38 children studied after watching lightning strike and kill a playmate had no memory of the event" (pp. 609–610). Brown *et al.* [25], however, failed to mention that both amnesic children had themselves been hit by side flashes from the main lightning bolt, knocked unconscious, and nearly killed. Those children who were not struck by the lightning remembered the disaster quite well, and many had marked symptoms of psychological distress. Hence, the physical, not psychological, aspects of the lightning strike were responsible for the amnesia in these two youngsters.

Confusing Incomplete Encoding with Traumatic Amnesia

The *Diagnostic and Statistical Manual of Mental Disorders Fourth Edition Test Revision* (*DSM-IV-TR*) [1] lists "inability to recall an important aspect of the trauma" (p. 468) as a symptom of PTSD. Accordingly, some clinicians have adduced this symptom as relevant to traumatic amnesia [27]. Unfortunately, the meaning of this symptom is ambiguous. If someone does not remember an aspect of the trauma, one cannot tell (i) whether the person is unable to recall something that was encoded into memory, or (ii) whether the person never encoded the information in the first place. However, not all sensory input arising from an experience, traumatic or otherwise, gets encoded into memory. Under conditions of high arousal, attention often narrows to the most salient feature of an unfolding event such that central aspects get encoded at the expense of peripheral ones. For example, consider "weapon focus" [28]. Individuals robbed at gunpoint sometimes fail to encode the face of the robber, often because their attention is directed to the weapon. Being unable to recall the robber's appearance would not constitute amnesia because the person failed to attend and encode the face of the robber. Therefore, failing to recall something that was never encoded in the first place does not count as amnesia. Amnesia denotes that the material was encoded, but cannot be retrieved.

Confusing Nondisclosure with Traumatic Amnesia

Several studies have shown that a significant minority of adults with documented childhood abuse histories do not disclose these events [29, 30]. This failure to disclose abuse does not necessarily mean that one is unable to recall it. For example, Femina *et al.* [31] recontacted respondents who had described their abuse during a first interview, but had denied it during a second interview. When queried about the discrepancy in a follow-up interview, each respondent affirmed having recalled it during the second interview. The nondisclosing respondents gave several reasons why they had been unwilling to disclose the abuse during the second interview (e.g., did not want to talk about an upsetting experience; dislike of the interviewer). Therefore, one cannot equate failure to disclose abuse with an inability to remember it. Although it may indicate forgetting, practitioners should first rule out the more parsimonious explanation (i.e., a reluctance to disclose).

Confusing Childhood Amnesia with Traumatic Dissociative Amnesia

Most people can remember very little of their lives before age 4 or 5. Brain maturation and cognitive changes make it difficult for older children – let alone adults – to recall events encoded during their preschool years. Accordingly, a failure to recall an episode of childhood abuse from the preschool years may reflect childhood amnesia rather than traumatic amnesia. For example, several of the respondents in Williams's [30] classic study were very young when they had been taken to the hospital for suspected, and often confirmed, childhood sexual abuse (CSA). Their inability to recall the episode many years later was likely a function of childhood amnesia, not traumatic amnesia.

Confusing not Thinking about Something for a Long Time as Traumatic Amnesia

Some researchers have reported that nearly 60% of adult patients who report having been sexually abused as children answer affirmatively when asked whether there was ever "a time when you could not remember" the abuse [32, p. 24]. Such findings have been interpreted as evidence for "sexual abuse-related repression" (p. 26). An affirmative response to this question implies that the patient had spent a period

of time trying unsuccessfully to recall his or her molestation. But if the patient was entirely unaware of having been abused (i.e., had repressed the memory of the experience), why would he or she attempt to recall it in the first place?

On the other hand, if patients had interpreted this oddly worded question as asking, "Was there ever a time when you had not thought about your abuse?", then their affirmative replies make sense. Yet not thinking about something for a long time does not mean that the person was incapable of recalling it during the period when it never came to mind. Only an inability to remember encoded material should be considered amnesia. Had patients been interviewed during the time of their supposed amnesia, recollections of the abuse may very well have surfaced.

Recovered Memories

Aside from the concept of traumatic amnesia, there is much debate concerning whether a traumatic memory can be forgotten, only to be recalled later in life. The authenticity of these so-called recovered memories has often been a reason for discussion. This discussion has mainly focused on memories for CSA.

During the past decade, researchers have begun studying cognitive functioning in people who report having recovered memories of sexual abuse after many years of allegedly not having thought about it. Some research has shown that individuals reporting recovered CSA memories are more prone than other people to falsely remember words that were *not* presented in a list-learning paradigm [33, 34]. This suggests that people reporting recovered memories may be more prone in general to remembering events that they have not experienced, consistent with the possibility that some recovered memories may not refer to genuine abuse episodes.

In contrast, Schooler *et al.* [35] described several case studies of individuals who recalled apparently long-forgotten memories of corroborated CSA, demonstrating that at least some recovered memory experiences pertain to real events. In some of these cases, the partners of the women who reported recovered memory experiences said that the women had talked about the abuse *before* the recovered memory experience. The women were surprised to learn that they had talked about the abuse prior to when they thought they had first recalled it. Evidently, these women had forgotten their prior recollections. Schooler *et al.* proposed that these cases illustrate a "forgot-it-all-along" (FIA) phenomenon, in which remembering an event in a qualitatively new way (e.g., more emotionally) leads the individual to fail to recall prior occasions of recollecting that event.

Using a laboratory analog of the FIA phenomenon, we found that people reporting recovered CSA memories exhibit an FIA effect twice as large as that observed in people reporting continuous memories of CSA [36]. These findings, together with Schooler's observations, are consistent with the possibility that some recovered memories reflect genuine abuse episodes that may have been recalled before, but for which people had forgotten these previous recollections. That is, the FIA effect may result in the illusion that memories of certain events had not surfaced in many years when, in fact, people may simply forget their prior recollections.

Different Recovered Memory Experiences

Both these findings suggest radically different interpretations for how recovered memory experiences come about. How can these phenomena be integrated? Careful inspection of the nature of the recovered memory experience led us to identify two qualitatively different types of recovered memory reports. The differences in these types of reports suggest a way to understand the discrepant findings mentioned above. In one type, people come to believe that they are abuse survivors, commonly attributing current life difficulties to their repressed memories of abuse. Here, recovered memories arise following a prolonged and intensive effort to uncover the repressed memories, often by suggestions of a therapist. There are reasons to believe that some such recovered memories are illusory [37].

In the other type of recovered memory experience, people are suddenly reminded of events that they believe have not come to mind in many years. These recovered memory experiences occur spontaneously – outside therapy – when individuals encounter suitable reminders of the abuse episodes. Such recovered memories are more likely to be genuine than those surfacing during "recovered memory therapy" [37]. For example, it may be that in many such cases the individual underestimates prior knowledge about the abuse. That is, some spontaneously

recovered memory experiences may arise when people fail to remember prior recollections of an authentic abuse episode.

We examined whether there are different cognitive origins of these different types of recovered memory experiences [38]. We found a double dissociation between these groups with respect to performance on laboratory measures of memory designed to target the cognitive mechanisms that may underlie their experiences: People reporting CSA memories recovered during therapy scored high on a measure of susceptibility to false memories, but performed similarly to control subjects on a measure tapping the tendency to forget prior experiences of remembering (the FIA effect). Conversely, people with spontaneously recovered memories of abuse were especially prone to forget their prior remembering (showed larger FIA effects) but scored similarly to control subjects on false memory tasks.

This double dissociation indicates that there are important differences in the cognitive profiles of people who recover memories of CSA either through suggestive therapy or, more spontaneously, without extensive prompting or attempts to reconstruct their past. As a group, people who come to believe that they have recovered a memory of CSA through suggestive therapy showed a pronounced tendency to claim incorrectly that they have experienced events when they demonstrably have not experienced them.

To the extent that this pattern is indicative of a broader deficit in monitoring the source of one's memories, such a finding suggests that these reports of recovered memories should be viewed cautiously, as they may reflect the interaction of suggestive therapy with preexisting deficits in source memory. This concern is reinforced by recent findings that of 16 cases of memories recovered through suggestive therapy, 0% could be corroborated by independent evidence [37]. Nevertheless, it should be stressed that these findings at the group level cannot speak to the validity of any individual's recovered memory experience, and it is possible that some memories recovered in this way are accurate, even if they cannot be corroborated.

In contrast, people who report spontaneously recovering a memory of CSA showed no evidence of heightened susceptibility to the creation of false memories in the laboratory. This group did, however, show a striking tendency to forget prior recollections. Thus, even when prior accessibility of simple events studied in the laboratory could be objectively demonstrated, this group, as a whole, was significantly more likely to deny having remembered those events on previous occasions.

To the extent that performance on such a simple laboratory test is indicative of a broader vulnerability to forgetting in the face of shifts in context, such findings suggest that many members of this spontaneously recovered memory group may have failed to remember their prior thoughts about a genuine incidence of CSA (perhaps because their way of thinking about the event has changed). Consistent with that possibility, memories recovered spontaneously were corroborated by independent evidence at a significantly higher rate (40%) than those surfacing during recovered memory therapy (0%), suggesting that they often may reflect real memories of abuse [37]. These studies do not address *why* people who have spontaneously recovered memories of CSA show a stronger tendency to underestimate their prior remembering, relative to the other groups. One possibility is suggested, however, by recent findings establishing that this population shows, in laboratory measures of thought suppression, an enhanced ability to suppress unwanted thoughts, especially if those thoughts concern negative experiences [39]. If so, then memory for prior thoughts concerning the target CSA event may have been more effectively suppressed because they were unpleasant, impairing their long-term accessibility. If prior retrieval events are forgotten in this way, a subsequent recovery experience is more likely to be judged as novel, creating the experience of discovering a memory for the first time.

Finally, other research on recovered memories of CSA, most occurring spontaneously outside psychotherapy, indicates that the events were not experienced as overwhelmingly terrifying when they first occurred [40, 41]. In this research, the typical victim was nonviolently fondled by a known person, and the young school-age victim did not fully understand that he or she was being sexually exploited. The experience was upsetting, unpleasant, and puzzling, but not traumatic in the sense of being terrifying or life-threatening. These facts increased the likelihood that the victim would not think about the abuse during the ensuing years. However, after encountering reminders much later in life, the person recalls the experience, and is extremely distressed by the realization that it signified sexual abuse.

Conclusion

Several conclusions emerge from our brief survey of the literature. First, memories of traumatic experiences – overwhelmingly terrifying, often life-threatening events – are nearly always remembered very well. Second, because the mind does not operate like a videotape machine, autobiographical memories are not literal reproductions of the original experience. Therefore, even memories of trauma can be incomplete, either because certain details never got encoded in the first place or because peripheral ones have faded with time. But the gist of the experience is highly memorable. Third, questions about the relative accuracy of traumatic versus nontraumatic memories usually are addressed in terms of consistency over time rather than correspondence with a veridical record of what transpired.

Despite being subject to distortion like all memories, memories of trauma tend to be more detailed and more resistant to distortion and forgetting than memories of nontraumatic events. Fourth, there is little or no convincing evidence that memories of genuinely terrifying events that get encoded in memory can later become inaccessible to recall because of repression or dissociation. Fifth, memories of CSA, especially those that did not involve violence, repetition, and terror, can seemingly be forgotten for long periods of time. However, not thinking about something for a long time is not the same thing as being unable to recall it during the period of time when it did not come to mind. Recovering (recalling) a memory of sexual abuse does not imply that the memory had previously been repressed, dissociated, or otherwise blocked from awareness by inhibitory mechanisms. The person may simply not have encountered relevant retrieval cues during the time when the memory never came to mind.

Sixth, individuals who report having spontaneously recovered memories are more prone than others to forget their prior experiences of remembering. Hence, even when people have seeming forgotten memories of CSA, this does not confirm that the memories never surfaced during the long period of apparent forgetting. Seventh, people who recover memories of sexual abuse during psychotherapy designed to foster recall of presumably repressed memories are more prone to false memory effects in the laboratory than are people who recover their abuse memories outside of psychotherapy, and their memories are corroborated far less often than spontaneously recovered memories. Eighth, some people who recover memories of CSA did not understand it as abuse or experience it as traumatic at the time it occurred. Accordingly, because it did not evoke overwhelmingly terror when it occurred, it was especially likely to slip from awareness.

Acknowledgment

Elke Geraerts was supported by a grant from the Netherlands Organization for Scientific Research (NWO 446-06-002).

References

[1] American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders – Text Revision*, 4th Edition (DSM-IV), American Psychiatric Association, Washington, DC.

[2] Schacter, D.L. (1996). *Searching for Memory: The Brain, The Mind, and The Past*, Basic Books, New York.

[3] Frueh, B.C., Elhai, J.D., Grubaugh, A.L., Monnier, J., Kashdan, T.B., Sauvageot, J.A., Hamner, M.B., Burkett, B.G. & Arana, G.W. (2005). Documented combat exposure of US veterans seeking treatment for combat-related post-traumatic stress disorder, *British Journal of Psychiatry* **186**, 467–472.

[4] McNally, R.J. (2003). *Remembering Trauma*, Belknap Press/Harvard University Press, Cambridge.

[5] Pope Jr, H.G., Oliva, P.S. & Hudson, J.I. (1999). Repressed memories: the scientific status, in D.L. Faigman, D.H. Kaye, M.J. Saks & J. Sanders, eds, *Modern Scientific Evidence: The Law and Science of Expert Testimony*, vol. 1, pocket part, West Publishing, St. Paul, pp. 115–155.

[6] Nourkova, V., Bernstein, D.M. & Loftus, E.F. (2004). Altering traumatic memory, *Cognition and Emotion* **18**, 575–585.

[7] Wagenaar, W.A. & Groeneweg, J. (1990). The memory of concentration camp survivors, *Applied Cognitive Psychology* **4**, 77–87.

[8] Porter, S. & Peace, K.A. (2007). The scars of memory: a prospective, longitudinal investigation of the consistency of traumatic and positive emotional memories in adulthood, *Psychological Science* **18**, 435–441.

[9] van der Kolk, B.A. (1996). Trauma and memory, in *Traumatic Stress: The Effects of Overwhelming Experience on Mind, Body, and Society*, B.A. van der Kolk, A.C. Mc.Farlane & L. Weisæth, eds, The Guilford Press, New York.

[10] Brewin, C.R., Dalgleish, T. & Joseph, S. (1996). A dual representation theory of posttraumatic stress disorder, *Psychological Review* **103**, 670–686.

[11] Ehlers, A. & Clark, D.M. (2000). A cognitive model of posttraumatic stress disorder, *Behaviour Research and Therapy* **38**, 319–345.
[12] Kihlstrom, J.F. (2006). Trauma and memory revisited, in *Memory and Emotions: Interdisciplinary Perspectives*, B. Uttl, N. Ohta & A.L. Siegenthaler, eds, Blackwell, 259–291.
[13] Shobe, K.K. & Kihlstrom, J.F. (1997). Is traumatic memory special? *Current Directions in Psychological Science* **6**, 70–74.
[14] Nadel, L. & Jacobs, W.J. (1998). Traumatic memory is special, *Current Directions in Psychological Science* **7**, 154–156.
[15] Brewin, C.R. (2007). Autobiographical memory for trauma: update on four controversies, *Memory* **15**, 227–248.
[16] Kuch, K. & Cox, B.J. (1992). Symptoms of PTSD in 124 survivors of the Holocaust, *American Journal of Psychiatry* **149**, 337–340.
[17] Merckelbach, H., Dekkers, T., Wessel, I. & Roefs, A. (2003). Amnesia, flashbacks, nightmares, and dissociation in aging concentration camp survivors, *Behaviour Research and Therapy* **41**, 351–360.
[18] Geraerts, E., Kozaric-Kovacic, D., Merckelbach, H., Peraica, T., Jelicic, M. & Candel, I. (2007). Traumatic memories of war veterans: not so special after all, *Consciousness and Cognition* **16**, 170–177.
[19] Porter, S. & Birt, A.R. (2001). Is traumatic memory special? A comparison of traumatic memory characteristics with memory for other emotional life experiences. *Applied Cognitive Psychology* **15**, S101–S117.
[20] Peace, K.A., Porter, S. & ten Brinke, L. (2008). Are memories for sexually traumatic events special? A within subjects investigation of trauma and memory in a clinical sample, *Memory* **16**, 10–21.
[21] Brown, D., Scheflin, A.W. & Whitfield, C.L. (1999). Recovered memories: the current weight of the evidence in science and in the courts, *Journal of Psychiatry and Law* **27**, 5–156.
[22] McNally, R.J. (2005). Debunking myths about trauma and memory, *Canadian Journal of Psychiatry* **50**, 817–822.
[23] Wilkinson, C.B. (1983). Aftermath of a disaster: the collapse of the Hyatt Regency Hotel skywalks, *American Journal of Psychiatry* **140**, 1134–1139.
[24] Kihlstrom, J.F. & Schacter, D.L. (2000). Functional amnesia, in *Handbook of Neuropsychology*, 2nd Edition, F. Boller & J. Grafman, eds, Elsevier Science, Amsterdam, pp. 409–427.
[25] Brown, D., Scheflin, A.W. & Hammond, D.C. (1998). *Memory, Trauma Treatment, and The Law*, Norton, New York.
[26] Dollinger, S.J. (1985). Lightning-strike disaster among children, *British Journal of Medical Psychology* **58**, 375–383.
[27] Gleaves, D.H., Smith, S.M., Butler, L.D. & Spiegel, D. (2004). False and recovered memories in the laboratory and clinic: a review of experimental and clinical evidence, *Clinical Psychology: Science and Practice* **11**, 3–28.
[28] Loftus, E.F., Loftus, G.R. & Messo, J. (1987). Some facts about "weapon focus", *Law and Human Behavior* **11**, 55–62.
[29] Widom, C.S. & Morris, S. (1997). Accuracy of adult recollections of childhood victimization: part 2. Childhood sexual abuse, *Psychological Assessment* **9**, 34–46.
[30] Williams, L.M. (1994). Recall of childhood trauma: a prospective study of women's memories of child sexual abuse, *Journal of Consulting and Clinical Psychology* **62**, 1167–1176.
[31] Femina, D.D., Yeager, C.A. & Lewis, D.O. (1990). Child abuse: adolescent records vs. adult recall, *Child Abuse and Neglect* **14**, 227–231.
[32] Briere, J. & Conte, J. (1993). Self-reported amnesia for abuse in adults molested as children, *Journal of Traumatic Stress* **6**, 21–31.
[33] Clancy, S.A., Schacter, D.L., McNally, R.J. & Pitman, R.K. (2000). False recognition in women reporting recovered memories of sexual abuse, *Psychological Science* **11**, 26–31.
[34] Geraerts, E., Smeets, E., Jelicic, M., van Heerden, J. & Merckelbach, H. (2000). Fantasy proneness, but not self-reported trauma is related to DRM performance of women reporting recovered memories of childhood sexual abuse, *Consciousness and Cognition* **14**, 602–612.
[35] Schooler, J.W., Bendiksen, M. & Ambadar, Z. (1997). Taking the middle line: can we accommodate both fabricated and recovered memories of sexual abuse?, in *Recovered Memories and False Memories*, M.A. Conway, ed, Oxford University Press, Oxford, pp. 251–292.
[36] Geraerts, E., Arnold, M.M., Lindsay, D.S., Merckelbach, H., Jelicic, M. & Hauer, B. (2006). Forgetting of prior remembering in people reporting recovered memories of childhood sexual abuse, *Psychological Science* **17**, 1002–1008.
[37] Geraerts, E., Schooler, J.W., Merckelbach, H., Jelicic, M., Hauer, B. & Ambadar, Z. (2007). The reality of recovered memories: corroborating continuous and discontinuous memories of childhood sexual abuse, *Psychological Science* **18**, 564–568.
[38] Geraerts, E., Lindsay, D.S., Merckelbach, H., Jelicic, M., Raymaekers, L., Arnold, M.M. & Schooler, J.S. *Cognitive mechanisms underlying recovered memory experiences of childhood sexual abuse*. Manuscript submitted for publication.
[39] Geraerts, E., McNally, R.J., Jelicic, M., Merckelbach, H. & Raymaekers, L. (2008). Linking thought suppression and recovered memories of childhood sexual abuse, *Memory* **16**, 22–28.
[40] Clancy, S.A. & McNally, R.J. (2005/2006). Who needs repression? Normal memory processes can explain "forgetting" of childhood sexual abuse, *Scientific Review of Mental Health Practice* **4**, 66–73.
[41] McNally, R.J., Perlman, C.A., Ristuccia, C.S. & Clancy, S.A. (2006). Clinical characteristics of adults reporting

repressed, recovered, or continuous memories of childhood sexual abuse, *Journal of Consulting and Clinical Psychology* **74**, 237–242.

ELKE GERAERTS AND RICHARD J. MCNALLY

Reconstruction *see*
Reconstruction: Accident

Reconstruction: Accident

Physical Evidence in Accident Reconstruction

The objective of automobile accident reconstruction is to determine how a collision occurred, based on the physical evidence. Vehicle speed, direction of travel, and position on the roadway at impact are often the elements of interest.

The physical evidence may consist of tire marks on the road, damage to the vehicles, debris patterns, and vehicle rest positions. If a police report was prepared, the scene evidence may be documented or recorded with a varied amount of detail. Vehicles are sometimes moved from their initial rest positions by the driver, or at the direction of law enforcement. Tire marks fade with time, and debris fields are quickly cleaned by rescue crews. Generally, few accident scenes are photographed by law enforcement, witnesses, or independent photographers. Because of these variations, accident reconstruction is necessary to determine the most accurate sequence of events.

Classical reconstruction starts with the data postincident and works backward in time to determine the preincident sequence of events. The physical evidence is used to establish the pre- and postimpact motion of the vehicles, the impact orientation between the vehicles, and the vehicle impact speeds. The process of analyzing and utilizing the physical evidence to understand how the accident took place is like putting the pieces of a puzzle back together to compose the original image. Ultimately, the quantity and quality of the physical evidence determine what can be learned about how the accident happened.

A physical inspection of the accident scene provides measurements of the roadway geometry, slope, and surface characteristics, which can be used as the basis for a scaled diagram of the collision site. Any photographs of the scene, taken at the time of the accident, are used to verify vehicle rest positions and other evidence locations. Possible vision obstructions are identified and their location noted. Vehicle tire marks, gouges on the roadway surface, and any damage to curbs, poles, and signs are recorded. Photographs are taken at specific locations to demonstrate the visibility available to the drivers and to document the information or evidence that may degrade or vanish over time. Physical evidence from all available sources is collectively considered to establish the postincident conditions.

Measuring of the vehicular damage assists in the determination of the impact-force direction and quantification of collision energy. Tire damage and tire-to-vehicle interference contribute to postimpact speed calculations. Worn tires, separated drive shafts, and frame bowing are documented for vehicle speed, trajectory, and collision energy analysis.

Photogrammetry utilizes postincident photographs of the scene to reconstruct the location of evidence visible in the photos when actual measurements of the evidence were not documented, and/or the evidence is no longer visible at the scene. Computer-based analytical photogrammetry, using postincident photographs and precision measurement of unchanged scene features, may be used to reestablish the locations of tire marks, gouges, and other physical evidence. This technique can also be used to determine the damage measurements from vehicle photographs when the vehicle has been dismantled.

Vehicle Speed from Skidmarks, Digital Data Acquisition, and Computer Simulation

The speed of an automobile that skids to a stop, without impacting another vehicle or object, can be computed from the length of the skidmarks and the friction value between the vehicle tires and the roadway or driving surface. The mathematical

equation used to compute the speed of a vehicle at the start of skidding is calculated as shown in equation 1.

$$V = \sqrt{254.3 \cdot f \cdot d} \quad (1)$$

where V is the vehicle speed (kilometers per hour, kph), f is the tire-road coefficient of friction (dimensionless), and d is the length of skidmarks (m).

Equation 1 is mathematically derived from the principle of the conservation of energy – any energy in motion is converted to other forms to stop the motion (e.g., heat). Owing to the speed of the vehicle, its kinetic energy is dissipated in the work done by the tires skidding on the road. Note that the calculation is within limits and is not affected by the weight of the vehicle. Therefore, a 2000 kg vehicle will stop in the same distance as a 1000 kg vehicle from the same initial speed utilizing the same type of braking systems.

The coefficient of friction between the vehicle tires and the roadway is affected by several factors. Asphalt, dirt, gravel, and grass yield different friction values with the identical automobile tire. Water on an asphalt roadway may reduce the coefficient of friction for an automobile tire. However, tread depth on tires has a significant effect on tire–road friction on wet surfaces, or when driving in snow or mud.

Partially effective or defective brakes affect the result of a speed from skidmark calculation. One or more tires may be braking at a significantly reduced level or may not be braking at all. Adjustments to the equation must be made to account for this condition. The form of the equation is the same with the addition of a new variable that accounts for braking efficiency.

Table 1 outlines some representative values of the coefficient of friction for automobile tires on various surfaces.

Table 1 Automobile tire coefficient of friction values

Surface	Coefficient of friction
Dry asphalt	0.65–0.85
Wet asphalt	0.45–0.80
Dry dirt	0.55–0.75
Gravel	0.45–0.65
Mud	0.40–0.50
Packed snow	0.10–0.20
Ice	0.02–0.08

Many automobiles are equipped with crash data recorders (CDR). These recorders use sensor-based technologies to measure and track relevant data over time. Data is stored for later reading and enables real-time access to the information. The ability to monitor, record, and analyze vehicle and equipment behavior is an asset to accident reconstruction.

As an example, one American manufacturer provides the postimpact connections to a diagnostic link connector (DLC) or directly to the air bag module, which relates data to supplement the accident reconstruction. Data available varies by manufacturer. Examples of accessible information include vehicle speed, engine speed, brake status, and throttle position (all 5 s before impact), driver's seat belt status, passenger's air bag state (enabled/disabled), as well as change in velocity (ΔV) versus time graph for the deployment event.

Computer simulation programs can be used to predict the movement of colliding vehicles. These models incorporate vehicle dimensions and weight, vehicle inertial properties, tire properties, roadway surface friction, and driver steering and braking inputs. Predicted, postimpact trajectories from simulations that include tire mark information can be used to supplement and/or verify the results of computations based only on vehicle impact and rest positions. Scene and tire mark evidence can be imported into the simulation model for direct comparison to the simulation results. Simulation programs can include vehicle stiffness information that is used to predict the postcollision damage profiles on the vehicles. The predicted damage from the program may be compared with actual vehicle damage measurements or damage photographs to confirm the alignment of the vehicles at impact.

Speed Analysis of Angled, Two-Vehicle Collisions

Two-vehicle collisions, where the vehicles approach impact at a distinct angle to each other, are relatively common. This category of collision can occur at intersections, in parking lots, and on highways when a vehicle crosses into oncoming traffic. Establishing the speed of the vehicles at impact is typically the objective of the analysis.

The approach and departure paths of the vehicles are obtained from physical evidence including tire

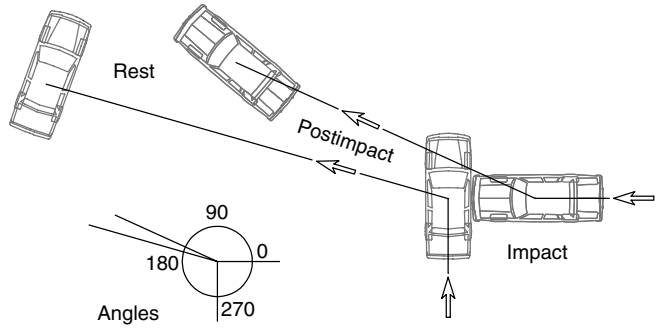


Figure 1 Principal elements of a simple, angled, two-vehicle collision

marks, points of rest of the vehicles, debris fields, and blood stains. A calculation of the postimpact vehicle speeds utilizes various forms of equation 1 or other analytical techniques. The postimpact speed analysis may become quite involved when vehicular rotation/sliding occurs before impacting a roadside object and coming to rest. One equation used to compute the postimpact speed of a vehicle that has been stopped by a roadside object after sliding is shown in equation 2.

$$V\mathrm{p} = \sqrt{V\mathrm{o}^2 + 254.3 \cdot a \cdot d} \qquad (2)$$

where V_p is the vehicular postimpact speed (kilometers per hour, kph), V_o is the vehicular impact speed with roadside object (kilometers per hour, kph), a is the deceleration rate of vehicle while sliding (g units/1 g = 9.81 m s^{-2}), and d is the distance of vehicle sliding before impact with object (m).

The impact speed at the roadside object and the vehicle deceleration rate during sliding must be determined. The impact speed at the roadside object usually involves an energy-based analysis of the vehicle damage from the object.

The postimpact deceleration rate is often computed using a technique that determines the total force on the vehicle, based on the individual forces at each tire, as the vehicle slides and rotates along the roadway. Since some tires may be locked from the initial collision, while others continue to roll, the analysis requires careful interpretation of the physical evidence and vehicle damage.

The principle of conservation of linear momentum is used to compute the impact speeds once the postimpact speeds, departure angles, and approach angles are determined. Figure 1 illustrates the simplified principal elements of a simple, angled, two-vehicle collision.

A number is assigned to each vehicle, and an angular coordinate system is established to allow the formulation of solution equations for the impact speed of each vehicle.

Using linear momentum, the equation for impact speed of vehicle 1 is calculated as shown in equation 3.

$$V_1 = \frac{m_1 \cdot V_1 p \cdot \sin(\theta_1 p - \theta_2) + m_2 \cdot V_2 p \cdot \sin(\theta_2 p - \theta_2)}{m_1 \cdot \sin(\theta_1 - \theta_2)} \qquad (3)$$

where V_1 is the vehicle 1 impact speed (kph); m_1, m_2 are the mass (kilogram) of vehicles 1 and 2, respectively; θ_1p, θ_2p are the postimpact departure angles for vehicles 1 and 2, respectively (degrees); θ_1, θ_2 are the approach angles for vehicles 1 and 2, respectively (degrees); and V_1p, V_2p are the postimpact speeds for vehicles 1 and 2, respectively (kph).

A similar equation is developed for vehicle 2. The impact speed for both vehicles may now be computed using the reconstructed speeds and angles.

Angled collisions may also change the vehicle's direction and speed. Since the preimpact and postimpact speeds and directions have been computed for the vehicles, this "speed change" quantity can be determined for each vehicle. The magnitude of speed change is one of numerous important parameters in assessing the occupant trauma potential of a collision. The speed change direction or principal direction of force is utilized to study the motion of occupants in the vehicle during the collision (termed *occupant kinematics*) as well as the use and/or effectiveness of occupant restraint systems.

Speed Analysis of Inline, Two-Vehicle Collisions

Two-vehicle collisions, with contact between the front of one vehicle and the rear of the other vehicle, with no significant angle or offset, are referred to as *rear-impact inline impacts*. In this type of collision, the impulse (force–time) vector passes through the center of mass of both vehicles. Figure 2 illustrates a rear-impact inline collision.

This kind of collision frequently occurs at low-to-moderate speeds in stop-and-go traffic and at stop signs. The struck vehicle may be stopped or moving forward when impacted from behind. Occasionally, a vehicle will back into the car behind it. Rear-impact inline collisions also occur at high speeds when a faster driver strikes the rear of a slower moving vehicle. Determining the impact speed of the striking vehicle, collision closing speed, and speed change of the struck vehicle are some generalized objectives of accident reconstruction analysis.

The principle of conservation of energy is used to analyze inline collisions where crush damage to the vehicles is significant. Specialized techniques have been developed to analyze low-speed-change, rear-end collisions with minor amounts of vehicle damage, and may involve quantifying the speed change of the struck vehicle using test data for the energy-absorbing bumper systems on the vehicles.

For collisions with significant vehicle deformation, measurements of the crush damage to each vehicle are made using a standard protocol documented in the literature published by the Society of Automotive Engineers. If the vehicles are no longer available, the crush profile may be estimated from photographs, or computed using photogrammetry and control dimensions from an exemplar vehicle. Stiffness values for each vehicle are obtained using the results of crash-safety tests or from other independent sources. The energy associated with the damage to each vehicle is then computed from the stiffness values and crush measurements.

The mathematical equation used to compute the damage energy for a vehicle with a straight crush profile, described by two damage measurements, is calculated as shown in equation 4.

$$E = w \cdot \left[G + \frac{A}{2} \cdot (C_1 + C_2) + \frac{B}{6} \cdot (C_1^2 + C_2^2 + C_1 \cdot C_2) \right] \cdot (1 + \tan^2(\theta)) \tag{4}$$

where E is the vehicle damage energy (N-m), w is the crush damage width (meters), $G = A^2/2B$, A is the stiffness coefficient for no residual damage (kg/meter), B is the damage stiffness coefficient for crush depth (kg/meter^2), C_1, C_2 are the crush depth measurements (meter), and θ is the angle between impact force and vehicular long axis (degrees).

The damage energy is computed for each vehicle separately. Using conservation of energy, an equation is developed to compute the closing speed between the two vehicles at impact. The closing speed equation is calculated as shown in equation 5.

$$Vc = \sqrt{\frac{2 \cdot Ec \cdot g \cdot (w_1 + w_2)}{(w_1 \cdot w_2)}} \tag{5}$$

where V_c is the closing speed between the vehicles (meters per second), E_c is the total damage energy for both vehicles (N-m), w_1, w_2 (N) are the weights of vehicles 1 and 2, respectively, and g is the gravitational acceleration (9.81 m s^{-2}).

The closing speed indicates the rate at which the striking vehicle is approaching the struck vehicle and is equivalent to the differential impact speed (difference between the velocities of the two vehicles). A collision where a stationary vehicle is struck by another vehicle traveling at 10 miles per hour has a closing speed of 10 miles per hour. If the struck vehicle is moving at 40 miles per hour and the striking vehicle is traveling at 50 miles per hour, in the same direction, the closing speed equals 10 miles per hour.

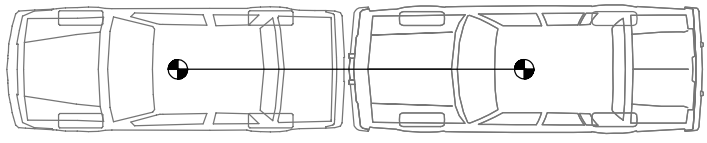


Figure 2 Illustration of a rear-impact inline vehicular collision

Theoretically, both collisions will result in the same speed changes and the same damage to the vehicles.

Once the closing speed is computed, the equations for conservation of energy and momentum are combined to determine the change in velocity (ΔV) or speed change of the struck vehicle. For a rear-end collision, where the struck vehicle is at rest initially, the speed change of the struck vehicle is calculated as shown in equation 6.

$$\Delta V_2 = \frac{1+\varepsilon}{1+w_2/w_1} \cdot V_1 \tag{6}$$

where ΔV_2 is the speed change of struck vehicle (kph), V_1 is the impact speed (closing speed) of striking vehicle (kph), ε is the coefficient of restitution (dimensionless), w_1, w_2 are the weights (N) of vehicles 1 and 2, respectively.

The coefficient of restitution is the ratio of the relative rebound velocity to the relative impact velocity and is a measure of the plasticity or elasticity of the collision. Restitution may be estimated based on the collision closing speed or computed based on test data. Several research studies have produced data for estimating the coefficient of restitution over a relatively wide range of closing speeds.

The value of the coefficient of restitution is within a range of zero (0, for a mathematically plastic collision) to one (1, for a mathematically elastic collision). Collisions tend to be more elastic at low closing speeds and more plastic at higher closing speeds. At a closing speed of 5 kilometers per hour, the coefficient of restitution may equal 0.6. At a closing speed of 25 kilometers per hour, the same two vehicles may have a coefficient of restitution of 0.2. The effect of restitution on speed change calculations is more significant at lower closing speeds.

Autopedestrian Incidents

Collisions involving automobiles and pedestrians may result in the pedestrian being projected some distance beyond the point of impact. In many autopedestrian accidents, there may not be any skidmarks from the vehicle, and therefore, a speed-from-skid mark analysis cannot be conducted. If the physical evidence can be used to determine the impact location and the postimpact trajectory of the pedestrian, a speed analysis for the striking vehicle may be conducted.

Pedestrian travel speeds toward the point of impact cannot be reliably calculated in most cases. Research data from crosswalk observations, pedestrian planning studies, and independent researchers provides a basis for estimating pedestrian speeds by age group. Usually, witness testimony is used to determine if a pedestrian is walking, jogging, or running. Pedestrian walking speeds are generally in the range of 0.31–1.83 meters per second, with an average walking speed of 1.37 meters per second. Running speed may exceed 4.27–4.88 meters per second for young adults, but drops off to about 2.44 meters per second for an adult over 60 years of age. Normally, a range of speeds is used to encompass the specific characteristics of the pedestrian.

Basic principles of physics and empirical data, developed from actual and staged collisions, are used to compute vehicular impact speed from pedestrian projection distance. The relative height of the pedestrian's center of mass must be considered in the analysis as the hood of a typical sedan will contact an adult pedestrian below their center of mass and a child pedestrian above their center of mass. Vehicle impact speed, braking level at impact, and frontal vehicular geometry, along with pedestrian center of mass height and contact location on the vehicle, are general variables influencing the postimpact trajectory of the pedestrian.

At high impact speeds, the pedestrian may pass over the top of the vehicle and not travel appreciably from the point of impact. In some autopedestrian accidents, the pedestrian is carried on the hood of the vehicle and falls to the ground after some travel distance. Not all pedestrian accidents leave physical evidence that establishes the point of impact. In these situations, the projection analysis presented here may not be applied.

One equation, derived from principles of physics, used to analyze vehicle impact speed in autopedestrian accidents is calculated as shown in equation 7.

$$Sp = \frac{V \cdot \sqrt{H}}{8} + \frac{V^2}{202} \tag{7}$$

where S_p is the pedestrian projection distance (meters), H is the pedestrian center of mass height (meters), and V is the vehicle impact speed (kilometers per hour, kph).

The above equation can be solved by simply iterating on the speed variable until a match to the projection distance is obtained, or solved directly for speed using the quadratic formula.

The graph in Figure 3 yields vehicle impact speed from pedestrian projection distance based on the results of field data from actual and staged autopedestrian impacts. This graph is for an adult pedestrian.

Since people are operating the involved vehicles, an evaluation of human performance may also be included. The disciplines of experimental psychology and human factors provide data and general guidance about the human aspects of collisions. Driver perception–response may be analyzed, using the physical evidence and testimony, to access a person's ability to avoid the collision.

Perception/Response

The analysis of automobile accidents relies upon the use of physics and engineering.

The time required for a driver to perceive and react to an unexpected hazard is associated with mental processing and physical action. The components of the process illustrate the complexity of the event:

- Detection – The perception–response time begins when the hazard enters the driver's visual field.
- Identification – Enough information about the hazard is obtained to reach a decision about what action is required.
- Decision – The driver decides what action is appropriate. The driver may brake, swerve, or change speed.
- Response – Commands to the proper muscle group carry out the chosen action.

Experimental studies show considerable variation in the way drivers respond to a given hazard. Some experimental data indicates the perception–response time depends on the evasive action chosen by the driver. Measurement of perception–response time using "real-world" driving hazards is difficult since participants cannot be involuntarily subjected to potentially dangerous situations.

A generally accepted recommendation for driver perception–response time to a reasonably clear hazard and a fairly straightforward situation is 1.5 s. Research data indicates about 85–95% of drivers will respond to the hazard in this time.

Several factors can affect the perception–response time: driver age, fatigue, alcohol, legal and illicit drugs, and driver expectancy. Some collision situations present complex and ambiguous information to the driver. Poor visibility, due to fog, rain, or darkness, complicates a driver's ability to identify a hazard. Adjustments to the perception–response time may be required when one of the components is affected by the circumstances of the collision.

The total distance required to bring an automobile to a stop includes the braking distance and the distance traveled during the perception–response interval. Braking distance increases in proportion to the square of the vehicle speed, and perception–response distance increases in direct proportion

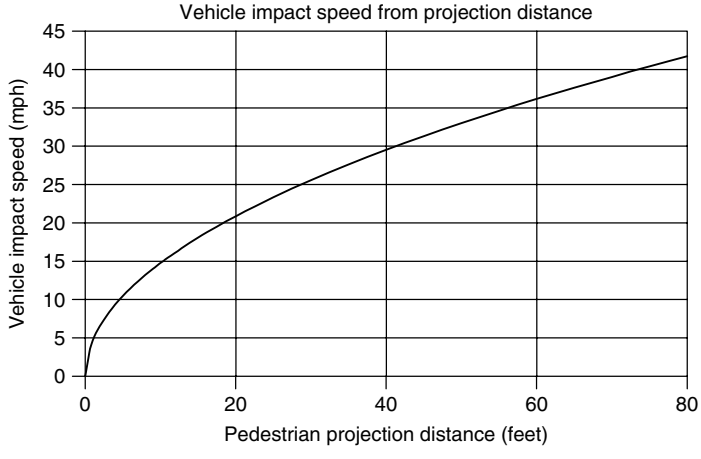


Figure 3 Vehicle impact speed from pedestrian projection distance based on the results of field data from actual and staged autopedestrian impacts (Appel [1])

to speed. Higher speeds give longer total stopping distances.

A driver's ability to avoid a collision, by braking to a stop before impact, is limited by the available stopping distance. The available distance may be determined by visibility obstructions or may be a computed value based on a speed and perception–response analysis. If the analysis shows that a driver was traveling in excess of the posted speed limit when the collision occurred, an avoidance calculation may be conducted to determine if the driver could have stopped ahead of the collision, had he or she been traveling at the speed limit.

Table 2 gives the values of the braking and perception–response distance for a range of vehicular speeds. The table also outlines total stopping distance for a dry road with a coefficient of friction of 0.76 and a driver perception–response time of 1.5 s.

Summary

Reconstruction begins with the examination of physical evidence and the accumulation of essential information. Traffic collision reports, supplemented with scene and vehicle inspections, as well as photographic analysis (from computer-based photogrammetry) may disclose crucial incident data.

After further analysis, pre- and postimpact speeds of the vehicles can be calculated. Vehicular speed prior to impact, if unknown, can be determined with skidmark length, braking efficiency, and quantification of roadway friction (coefficient of friction). The angles of impact and principal direction of force must also be considered. Angled and inline two-vehicle collisions require unique consideration, but necessitate the same insight as to the nature of the collision relevant to vehicular speeds. The magnitude of speed change is an important parameter in understanding occupant kinematics and potential trauma in a collision.

Analysis of automobile–pedestrian incidents may also indicate the nature of the collision and the vehicular speed prior to impact. Driver perception and response time, as well as the relationship of pedestrian projection distance to vehicular impact speed are often utilized in collision analysis.

Examination of available evidence in conjunction with accepted principles in order to further understand the precise nature of collisions is the strength of accident reconstruction. Depending upon the quantity and quality of such data, further analysis and calculations may determine much of the subject incident's unknown variables and lead to an understanding of passenger kinematics and potential trauma.

Table 2 Braking and perception–response distance for a range of vehicular speeds

Vehicular speed (kph)	Perception–response distance (meters)	Reaction distance (meters)	Breaking distance (meters)	Total stopping distance (meters)
5	3.8	3.8	0.1	3.9
10	7.5	7.5	0.5	8.0
15	11.3	11.3	1.1	12.4
20	15.0	15.0	2.0	17.0
25	18.8	18.8	3.1	21.8
30	22.5	22.5	4.4	26.9
35	26.3	26.3	6.0	32.3
40	30.0	30.0	8.1	38.1
45	33.8	33.8	10.5	44.2
50	37.5	37.5	13.2	50.7
55	41.3	41.3	16.3	57.6
60	45.0	45.0	19.5	64.5
65	48.8	48.8	23.1	71.9
70	52.5	52.5	27.1	79.6
75	56.3	56.3	31.6	87.9
80	60.0	60.0	36.6	96.6

Acknowledgment

This work is dedicated to the late Robert G. Liptai who was a contributing author to *Accident Reconstruction*, one of the first reference texts on accident reconstruction.

Reference

[1] Appel, H., Strurtrz, G., Gotzen, L. (1975). *Influence of Impact Speed and Vehicle Parameter on Injuries of Children and Adults in Pedestrian Accidents*, Institute of Automotive Engineering, Berlin.

Further Reading

Collins, J.C. (1979). *Accident Reconstruction*, Charles C. Thomas, Illinois.

Fricke, L.B. (1990). *Traffic Accident Reconstruction, Volume 2 of the Traffic Institute Manual*, Northwestern University Traffic Institute, Illinois.

Limpert, R. (1994). *Motor Vehicle Accident Reconstruction and Cause Analysis*, 4th Edition, The Michie Company, Virginia.

THOMAS C. CHRISTENSEN AND LAURA L. LIPTAI

Reconstruction: Facial *see* Facial Reconstruction

Reconstruction: Three Dimensional

Forensic 3D definition

Forensic 3D reconstruction is a new forensic tool to assist scientists, attorneys, judges, and juries to reconstruct a shooting case and explain material traces at scene of crime, on a suspect or victim. Forensic 3D technology uses a visual approach and presents critical evidence by use of pictures, video, and interactive 3D technology. Forensic 3D technology can be used to document crime scene with relevant traces and allows the expert to take measurements on 3D model of crime scene back in the office.

Forensic 3D technology can also be used in preliminary stages of investigation when an investigator wants to test several competing scenarios. Verbal explanations lack precision and are very often misleading when it comes to an explanation of the posture and position of a suspect, a victim, and a witness at the moment of shooting.

Using visual (multimedia) tools in the court, it is possible to present several competing scenarios of the shooting incident and estimate the likelihood of a particular scenario versus others.

Forensic 3D technology overcomes some of the limitations of the classical court reconstruction. Typical problems of the classical court reconstruction are as follows:

- Very often a scene of crime has changed so much or even has completely destroyed; hence, it is impossible to perform a real-court reconstruction.
- Very often the victim was wounded so badly that he is not able to move or demonstrate his positions or postures on crime scene during court reconstruction.
- Forensic 3D technology allows simulation of different conditions of light (day, night, artificial light sources) and weather conditions (wind, rain, snow, and fog). The insertion of sounds is also an option.
- Forensic 3D technology is effective in visualization of results of blood spatter analysis and reconstruction of position and posture of the victim and the suspect during critical moments at the crime scene.

Computer 3D reconstruction of shooting cases must always be based on the material traces found at the crime scene, the victim, and the suspect. This means a special attention must be paid to

- position of the cartridge cases;
- position of the bullet holes in the walls, doors, and similar solid and inanimate objects at the crime scene;
- gunshot residues (GSR) on a suspect's hands, clothing (face, hair, nostrils, etc.);
- GSR on a victim's hands, clothing (face, hair, nostril, etc.);
- report of a medical examiner concerning the wounding channel;

- traces of blood on a suspect's body and clothing;
- traces of blood on a victim's body and clothing;
- traces of blood on a scene of crime; and
- other traces relevant for a specific crime case.

Instead of reconstruction of the dynamics of the whole shooting incident/homicide case, it is sometimes better to reconstruct just some of the important segments of the incident that are capable of being reliable. This means it is better to freeze the suspect and the victim in their positions and postures in 3D space at the moment of the impact/hit according to the material evidence mentioned and reserve 3D reconstruction of the dynamics for the evaluation of different shooting scenarios.

Typical Input for a Forensic 3D Reconstruction

- All the data from the crime scene (measurements, sketches, maps, etc.);
- video data from surveillance cameras;
- versions of the incident provided by the suspects, victims, witnesses, etc.;
- ballistic identification of the weapon used;
- shooter identification/estimation according to the GSR from hands/clothing of the suspect;
- entrance/exit hole determination according to the GSR on clothing/skin of the victim;
- estimated position of the shooter according to known extraction of shells;
- ballistics and GSR analysis results evaluation and comparison with medical examiner report;
- result of blood spatter analysis; and
- estimation of posture and position in 3D space of the suspect and the victim at the moment the bullet hit the victim.

3D technology in the forensic field of application demands multilevel knowledge including engineering skills within the following:

1. Applications not exclusively forensic:
   (a) geodetic (by a total station, laser scanner, or photogrammetry);
   (b) cartography;
   (c) digital video editing;
   (d) 3D modeling using computer-aided design (CAD)/computer-aided modeling (CAM) software.
2. Forensic applications:
   (a) forensic ballistic examination;
   (b) GSR analysis;
   (c) forensic medical examination; and
   (d) forensic blood pattern analysis.

This is the reason why forensic 3D technology needs teamwork on the reconstruction of specific case. The team may typically include the following:

- crime mapping specialist;
- geodetic specialist;
- cartography specialist;
- specialist with skills in 3D modeling, animation, and simulation;
- specialist with skills in the digital video editing field;
- forensic ballistic expert;
- GSR analysis expert;
- forensic medical examiner;
- forensic blood spatter examiner; and
- other types of expert depending on the case.

Typical Forensic 3D Workflow

Typical forensic 3D workflow consists of the following:

- crime scene modeling;
- character modeling;
- inserting the characters in the crime scene;
- simulation of different versions of the shooting incident; and
- creation of the version of the ballistic expert.

Crime Scene Modeling

Manual crime scene modeling is a basic form of crime scene modeling. On the basis of the measurements from crime scene, we model every detail to which importance is ascribed in a specific shooting incident. The modeling is executed by the use of typical 3D modeling tools such as polygon modeling, splitting, extruding faces and edges, Boolean tools, as well as nurbs and subdivision surface tools. Manual modeling can be based on the police sketches, maps and diagrams, or photographs may be used as templates for modeling (Figures 1 and 2).

3D technology allows us to create a virtual "camera" and place it anywhere inside 3D model of crime

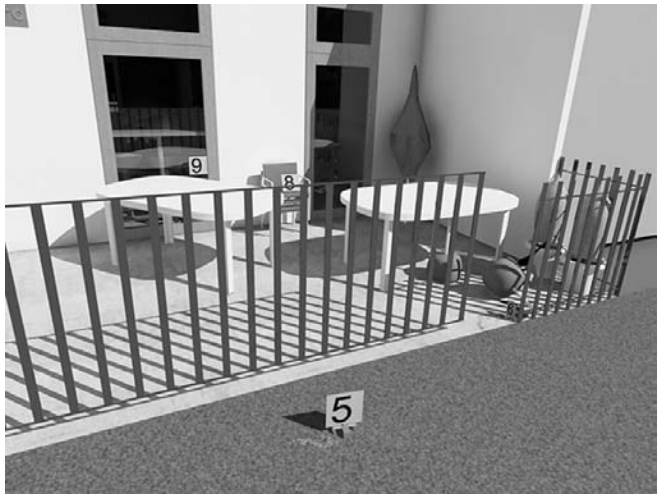


Figure 1 3D model of crime scene – exterior



Figure 2 3D model of crime scene – interior

scene. The advantage of this approach lies in the retention of freedom to choose any camera angle and position so that a shooting incident can be observed from every viewpoint and the "camera" itself can be animated. A "camera" so animated can show not only the crime scene but also a suspect and a victim "frozen" in time in some specific postures and positions in 3D space, while the "camera" is moving around and recording video sequences from different positions and angles, showing details important for the reconstruction of the dynamics of a shooting incident. This concept is known as the "*path animation of a camera*". The disadvantage of this approach is in the amount of work needed manually to model the crime scene itself. The final result is always an approximation of the reality because forensic expert decide what is important in the specific shooting case and what details may be discarded (Figure 3).

The disadvantages of this approach could be overcome by the use of a terrestrial photogrammetry programs that helps us make an exact 3D model of a crime scene based on a sequence of overlapping

photographs of the crime scene (Rollei Metric, Photo Modeler, etc.) or by the use of laser scanners (Ilris 3D, DeltaSphere, etc.).

To get an extremely precise and fast 3D model of a crime scene itself we can use a combination of computerized geodetic equipment (total station) and a laser scanner. Modern laser scanners such as Ilris 3D can scan a whole crime scene (3–1 km) in short time, with an average precision measured in centimeters or even in millimeters. The typical horizontal and vertical precision of the total station for a crime scene ranging from 200 to 300 m is about 10 mm, while laser scanners can achieve a precision measured in millimeters.

This 3D model can be imported into the 3D software and the reconstruction continued. 3D modeling of crime scene is a time-consuming job. To speed up the process it is possible to combine video taken on the scene of crime with a real video camera with computer-generated animated 3D models of the



Figure 3 High-quality 3D model of crime scene with all relevant traces

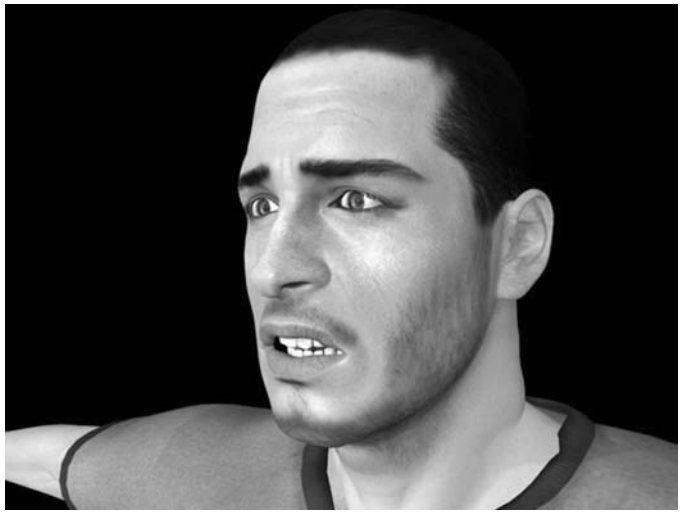


Figure 4 Detailed 3D model created using front and side photographs

suspect and the victim. Techniques used in film industry such as camera match moving and compositing can help to speed up the process. Camera match moving and compositing are special effects techniques that allow the insertion of computer-generated 3D models into video taken with real camera while maintaining correct position, scale, orientation, and motion in relation to the filmed objects in the scene of crime. Same techniques can be used to extract data from film footage taken by surveillance cameras during bank robberies, for example. Height of a perpetrator can be estimated in relation to dimensions of objects in surrounding.

Character modeling

Character modeling can be accomplished in several ways:

- Manual modeling is the most exact type of modeling and unfortunately the most complex one. Using front and side photographs and the "crossed plane method", we can make an exact 3D model of a suspect or a victim (Figure 4):
- creating 3D models of a person, skeletons, and animals using specialized software and
- creating a highly realistic 3D model of a particular victim/suspect by use of specialized laser scanners such as InSpeck.

After a 3D model of human (skin) has been made, it is necessary to create a skeletal system that will drive the parts of the skin. To have better control during the animation process, controls are created for arms, elbows, legs, knees, pelvis, and head, inserted inside the skin, which is bound to the skeletal system (Figures 5 and 6).

In most of the shooting cases where pistols or revolvers are involved, cylinders attached to a certain part of a body can approximate wounding channels. Visualization of wounding channels with a semitransparent body can be a great help in the case of multiple wounding channels (Figures 7 and 8).

Inserting Characters in a Crime Scene

3D models of a victim, suspect, and witness are imported into the 3D model of a crime scene. It is also possible to import 3D models of all the objects necessary for a particular shooting case, such as firearms, knives, cars, and trees (Figure 9).

Simulation of Different Versions of the Shooting Incident

Using character controls of imported 3D models of a suspect and a victim into a 3D model of a crime scene, their posture and position in 3D space can be controlled according to different versions of the

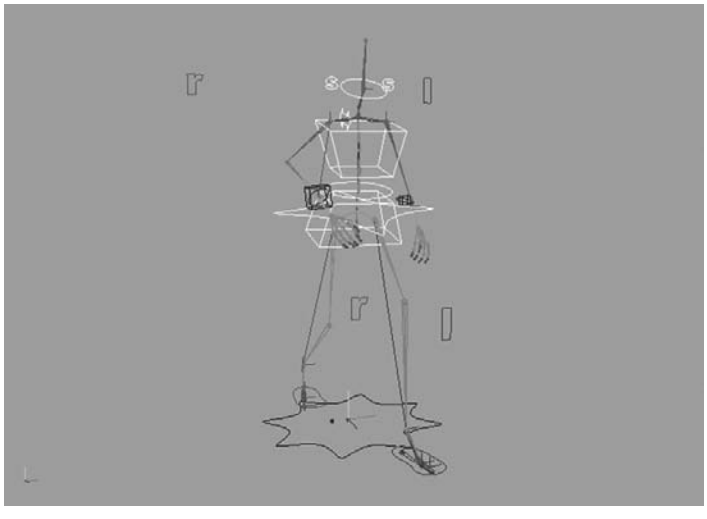


Figure 5 Detailed rigg of a character

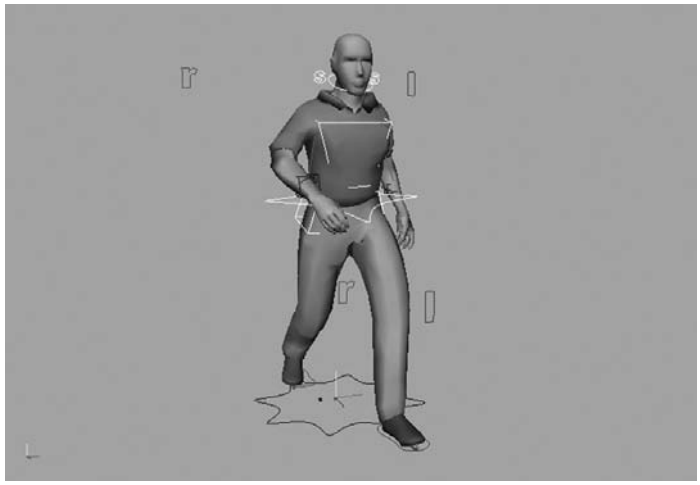


Figure 6 Rigged 3D model of a character

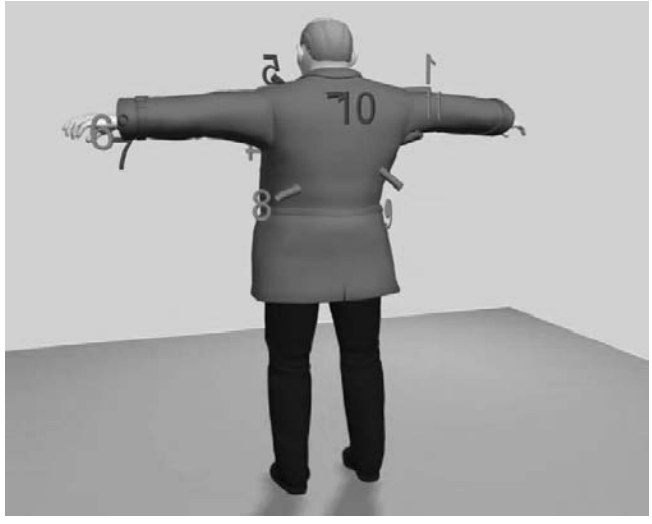


Figure 7 Visualization of wounding channels

incident and the results of forensic examinations (Figures 10 and 11).

The slow and complex work of character animation can be accelerated and greatly simplified by using motion capture systems. Those systems are capable of extracting motion from real people and exporting them in the form of motion capture files that CAD/CAM software can interpret as animation clips. Motion capture systems enable the operator simply to record the motion of a real actor and later to assign that animation clip to a particular 3D model.

Practically, in all shooting cases where several witnesses are involved, different versions of the same incident will be offered. Using the same 3D model of a crime scene suspect, victim and witnesses can be animated according to their versions of the shooting accident. With this technique it is possible to simulate different shooting scenarios. By taking into account all the relevant data from the crime scene and from

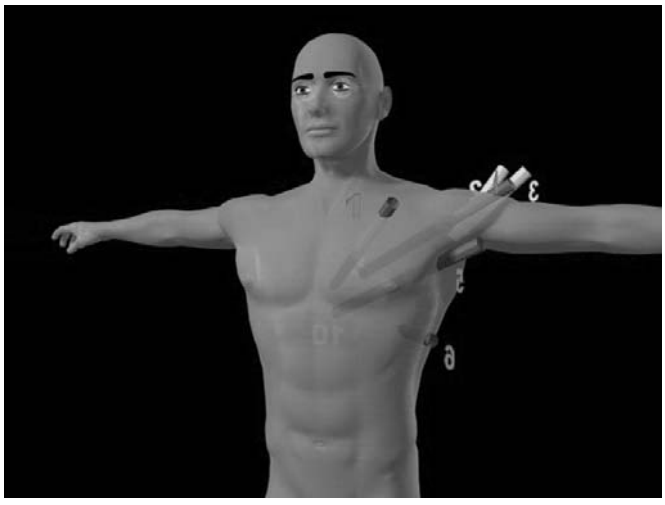


Figure 8 Visualization of wounding channels



Figure 9 Forensic 3D reconstruction of a shooting incident

the laboratory, the ballistic expert can conclude what really happened and hence formulate a version of the shooting incident. Moreover, ballistic expert can show his/her version of the incident in the form of a 3D animation or even in a form of forensic 3D interactive virtual reality (Figure 12).

Forensic 3D Interactive Virtual Reality (VRML)

Forensic 3D interactive virtual reality is a major improvement on Forensic 3D technology since Virtual Reality Modeling Language (VRML) technology permits the interactive inspection of a 3D model



Figure 10 Forensic 3D simulation of a version of shooting incident



Figure 11 Forensic 3D simulation of a version of shooting incident

of crime scene and all versions of particular shooting case. Interactive forensic 3D on the basis of the VRML technology allows virtual entrance inside the crime scene and inspection of every detail (Figure 13).

Contrary to forensic animations where it is impossible to change previously defined angle of a view of a virtual camera, interactive forensic 3D on the basis of the VRML technology allows complete freedom during "navigation" inside 3D model of scene of crime (Figure 14).

The 3D reconstruction of a shooting incident is an invaluable tool for the reconstruction of all dynamic situations where a ballistic expert is required to discriminate among competing versions of a particular shooting case (Figure 15).



Figure 12 Forensic 3D simulation of a version of a shooting incident

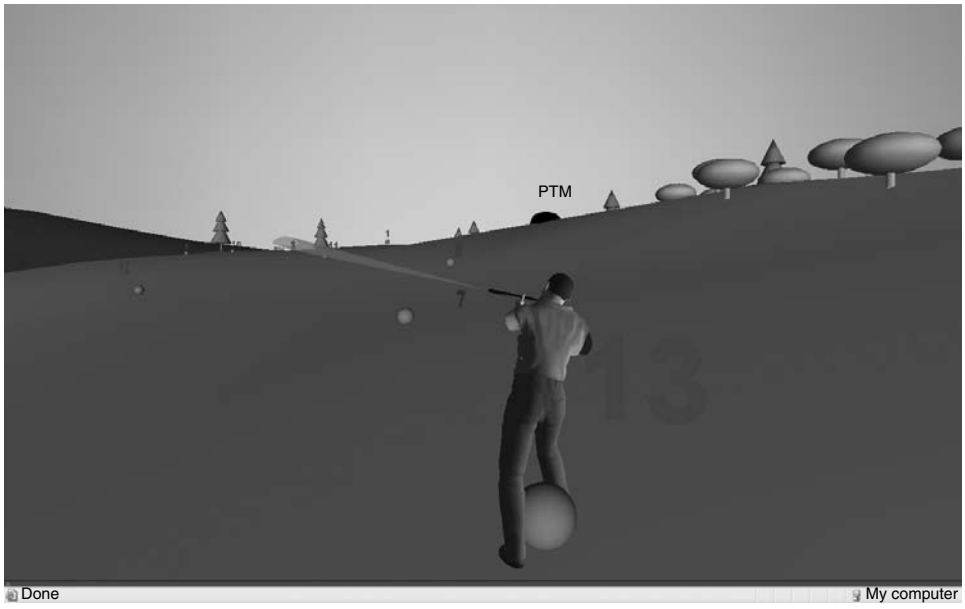


Figure 13 Forensic 3D simulation of a version of a shooting incident

The Errors and Inaccuracies that Exist in 3D Animations

The errors and insecurities that exist in 3D animations merely come from the fact that 3D models of a victim's body and wounding channels are only approximations. 3D models of people involved in a shooting case are only proxies and differ from real persons. A cylinder attached to the victim's body roughly approximates wounding channel produced by the impact of a bullet. The Virtopsy project [1] of Dr Richard Dirnhofer and Dr Michael Thali has



Figure 14 Forensic 3D simulation of a version of a shooting incident

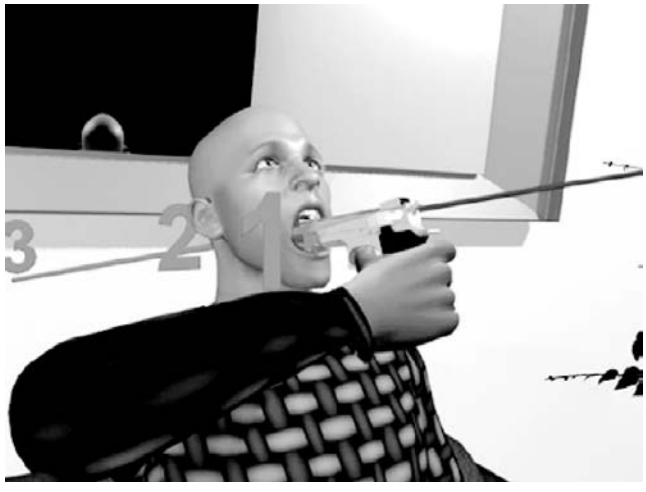


Figure 15 Forensic 3D simulation of a suicide showing version of a "thumb triggering"

a full potential to solve some errors and inaccuracies. The VIRTOPSY project is a unique project that uses imaging methods such as photogrammetry, optical surface laser scanning and multidetector computed tomography or multislice computed tomography (MDCT/MSCT), and magnetic resonance imaging (MRI) techniques [2]. Using laser scanner and photogrammetry software an exact textured 3D model of a victim is produced. Wounding channels are documented using the MDCT/MSCT and MRI techniques. Merging 3D/CAD photogrammetry with 3D radiological scanning, VIRTOPSY project reaches the highest level in 3D Surface and Body Documentation in Forensic Medicine.

The drawback of this approach is that the highly detailed 3D model of a victim and the wounding channel is not applicable for Forensic 3D reconstruction because the VIRTOPSY is applied to a static

body, whereas reconstruction is normally based on the dynamic movement of the body at the time of the incident.

Possible Future Developments of Forensic 3D

- Combination with terrestrial photogrammetry software (Roleimetric, Photo Modeler, GOM TRITOP/ATOS, etc.);
- combination with laser scanners (Ilris3D, DeltaSphere, etc.);
- combination with VIRTOPSY project;
- combination with Compositing and Camera Match Moving software (combination of 3D reconstruction with live footage: video film of a crime scene taken with a physical video camera);
- combination with motion capture systems (capturing motion files of real actors);
- combination with GSR analysis by the use of SEM/EDX;
- combination with terminal ballistics (replacing cylinder with realistic 3D model of a wounding channel);
- combination with game engines used to produce 3D games (This complex technique premises total interactivity inside crime scene with animated characters.); and
- combination with classic Photo-Robot technique (results of Photo-Robot analysis can be extended to a 3D model of the suspect).

Summary

- Forensic 3D is not a new photographic technique. It is rather a new visual analytical technique.
- Forensic 3D should be applied very carefully and be always based on the facts and material evidence.
- Forensic 3D has its limitations.
- It is reasonable to adopt forensic 3D as a novel technique and improve it.

References

[1] Thali, M.J., Braun, M., Buck, U., Aghayev, E., Jackowski, C., Vock, P., Sonnenschein, M. & Dirnhofer, R. (2005). VIRTOPSY – scientific documentation, reconstruction and animation in forensic: individual and real 3D data based geo-metric approach including optical body/object surface and radiological CT/MRI scanning. *Journal of Forensic Sciences* **50**(2), JFS2004290.

[2] Thali, M.J., Braun, M., Wirth, J., Vock, P. & Dirnhofer, R. (2003). 3D surface and body documentation in forensic medicine: 3-D/CAD photogrammetry merged with 3D radiological scanning, *Journal of Forensic Sciences* **48**(6), 486.

Related Articles

Biometric Devices
Crime Scene Investigation
Crime Scene Management
Firearm Discharge Residue: Analysis of
Facial Comparison
Image Processing and Analysis

VOJIN MASTRUKO

Reconstructive Memory ***see*** Memory: Reconstructive

Recovered Memories ***see*** Deception: Truth Serum, Recollective Accuracy of Traumatic Memories

Recovered Memories and Hypnosis ***see*** Hypnosis and Memory

Recovered Memory ***see*** Memory: Repressed

Recovered Memory Syndrome ***see*** Syndromes: Psychological

Refusing Medical Treatment ***see*** Treatment, Right to Refuse: Mental Health

Remains ***see*** Human Remains and Identity

Remains: Human, Identification of ***see*** Identification of Human Remains

Report Writing for Courts

Introduction

Expert witnesses occupy a unique position within the law and their role in legal proceedings highlights their ability to offer *opinion* evidence, rather than merely evidence of *facts*. Concomitant with the privilege of offering opinions as to the conclusions which may be drawn from the results of their work, experts bear a heavy responsibility. In many jurisdictions, expert witnesses explicitly owe a fundamental and overriding duty to the court, rather than to the party retaining them. For example, the *Civil Procedure Rules* in England and Wales state, "It is the duty of an expert to help the court on the matters within his expertise. This duty overrides any obligation to the person from whom he has received instructions or by whom he is paid" [1]. It is this duty that must be at the forefront of an expert's mind when writing a report for the court.

Given the gravity of the task and the potential for expert evidence to significantly impact upon the outcome of the legal proceedings, there arise several important steps for an expert to take when contemplating on writing a report for the court:

1. When a case file or exhibits are presented to an expert for analysis, first the expert must identify what sort of analysis or assessment needs to be done, in their expert opinion. Although working within a single discipline may provide an expert with case after case of similar evidence, it is important for the administration of justice in every individual case to consciously analyze what needs to be done *in this case*.
2. The expert must identify *why* they think that sort of analysis or assessment needs to be done in this case. This is not a rhetorical question, as lawyers, jurors, and judges who do not share the expert's specialized knowledge will not be familiar with why a particular technique or method of assessment was chosen or rejected in this case. For example, a forensic biologist may decide not to conduct low copy number (LCN) DNA testing on a set of crime scene samples. If a DNA profile has not been obtained using standard techniques, the decision not to use an LCN may seem to investigators, lawyers, jurors, or the judge to be illogical and incomprehensible. It is imperative that the expert in this scenario is conscious not only of why they chose to use standard DNA profiling methods but equally of why they did not use LCN testing (for example, because sufficient precautions were not taken to avoid contamination at the scene and of the samples) [2]. This necessarily requires specialized knowledge of the technique, method, examination, or assessment, and a conscious evaluation by the expert of the appropriateness, validity, reliability, and limitations of the options available to them in the circumstances of the individual case.
3. Which techniques, examinations, or assessments were, in fact, carried out? This may differ from what the expert anticipated in steps 1 and 2. The report must reflect what was actually done, rather than any sort of idealized version of what

the methods/techniques should have been and the results should have shown. If the results indicated that further or different examination/assessment needed to be conducted, the expert must be able to identify why this was or was not done, and what the results (or lack of results) mean. The report must reflect the work that was done and the results that were obtained, including the negative results.

4. In writing the report, the expert should be conscious of each of the previous steps. Sufficient details must be given in the report so that a lay audience is capable of comprehending the results and assessing the reliability and significance of the examination and/or assessment.

Legal Requirements of an Expert Report

Although the details vary depending on the jurisdiction, expert reports usually require consideration of some or all of the elements set out below [3]. In common law countries, these arise both from legislation and from case law; in some instances, it has been case law that precipitated changes to legislation. For instance, in England and Wales, the *Civil Procedure Rules* model the guidelines for expert witnesses set out in the *Ikarian Reefer* case [4].

1. The expert must be independent. As discussed above, the primary duty of an expert is to the court. This means that the evidence given by the expert should not vary on the basis of the side paying the expert's fee. That is not to say that the form and content of the expert's report cannot vary. Naturally, each side to a legal dispute would ask different questions of the same expert and it is entirely permissible for an expert report to focus on answering the questions asked by the side that retained them. What is important is that, if the expert was asked the *same* question by both parties, the expert's answer should be the same. Experts should consult with the lawyer who retained them to ensure that they are addressing the issues required by the retainer, and that they are not addressing matters that are not issues in the case and do not need to be included in the report.
2. The expert should not act as an advocate. If the specialized knowledge of the expert supports the case of the side which called him/her, then the expert is entitled to assist the court with this knowledge. It is not necessary, desirable, or permissible, however, for an expert to have any vested interest in seeking to support one side or the other in a legal dispute. Neither the expert's report nor oral testimony should move beyond their objective opinion and into the territory of advocacy.
3. The expert must remain within his/her area of expertise. "An expert should make it clear when a question or issue falls outside his expertise and when he is not able to reach a definite opinion, for example, because he has insufficient information"[5]. There are two important principles in this statement. First, an expert must be able to clearly identify, to themselves and to the court, their area of expertise. Judges and juries rely on expert witnesses to provide assistance in areas in which the judge/jury does not have specialized knowledge. It is important for experts to stay within their area of expertise, both when writing their report and giving evidence in court. Secondly, expert opinions and conclusions must be scientifically justifiable. The nature of the discipline or a lack of available information may make it scientifically impossible for the expert to reach a definite opinion. This is sufficient and important information for the judge or jury. It is legally, scientifically, and morally imperative that expert witnesses do not fabricate knowledge or certainty they do not have. In addition, it is necessary to identify in writing any other persons who assisted in the examinations or assessments recorded in the report. This usually includes a reference to other experts, laboratory staff, etc., who have handled/tested/participated in the testing of exhibits. This safeguards the expert who may not have handled the exhibits or samples at every step of the forensic process, but who has been asked to write a report setting out the work that has been carried out and their expert opinion of the results.
4. The expert must state the facts and assumptions that underlie his/her opinion. This is an important tool, which seeks to enable judges and juries to assess the expert's opinion. For example, under the *Civil Procedure Rules* in England and Wales, the expert's report must state the substance of all material instructions, whether written or oral, on the basis of which the report

was written [6]. In addition, if the expert is relying on professional literature as a source of facts or assumptions in their report, this should be made clear. It is not necessary to provide a thesis explaining the basis of the discipline; however, particularly if the methods used are at the forefront of science, the expert should refer the court to relevant published literature. Despite their expertise, expert witnesses are not the final word on technical, scientific, or medical questions in a legal dispute. The judge or jury must ultimately decide whether or not they accept the opinion of the expert, and to do this they must know the facts and assumptions that created the basis of the expert's opinions. When writing a report, an expert should also make clear the distinction between facts and opinions. Although the expert is allowed to give evidence of both (unlike ordinary witnesses, who are not generally allowed to offer opinions), the judge or jury needs to know which aspects of the expert's report are accepted facts and which are the opinions of the particular witness.

5. The expert must consider all material facts. In order for the expert opinion to be useful to the judge or jury, the expert must consider not only the facts that support the case theory of the side that retained them but also all of the material facts. Case law amply demonstrates instances of individuals deciding that particular pieces of information are not relevant to the case at hand, when, in fact, if each of those individual pieces of information was disclosed, a significant picture could emerge. This was the case in *R v. Ward*, where seemingly innocuous individual decisions by police officers, prosecutors, counsel, and others, combined to remove a significant body of evidence from the trial, resulting in a successful appeal by the accused. "Taken by itself, [the omitted statement] may well have been thought to have no bearing on the question of the appellant's responsibility for the offences with which she was charged, but when compared with the other [evidence that was withheld], its relevance to the crucial problem of discovering the truth or falseness of the appellant's various statements becomes apparent"[7].
6. Other requirements and considerations. Most jurisdictions specify other requirements for expert reports, which may include a list of the expert's qualifications and experience, a list of prior court testimony, a list of documents or exhibits used to prepare the report, and compensation received for the expert's opinion. Experts are advised to consult the lawyer who retained them to ensure that such requirements are met for the jurisdictions in which they have been asked to write the report. Other considerations such as the length and format of the report vary widely and should be clarified on a case-to-case basis. In all jurisdictions, it is helpful if the expert report is written using plain language, and that jargon and abbreviations are used only when necessary and are defined in the report. Draft reports should be marked as such; these may be subject to discovery in some jurisdictions (*see* **Discovery of Expert Findings**; **Discovery: Depositions**; and **Discovery: Discovery Motions**).

References

[1] Civil Procedure Rules (UK) Rule 35.3.
[2] R v Hoey [2007]. *NICC* 49 (20 *December* 2007) *per Weir J*.
[3] The legal requirements of expert reports in the United States of America are set out in Discovery in the United States: Civil Cases and Discovery in the United States: Criminal Cases.
[4] *National Justice Compania Naviera SA v Prudential Life Assurance Co Ltd (No1)* (1995). 1 Lloyd's Rep 455 per Cresswell J.
[5] Civil Procedure Rules (UK) "Experts and Assessors" Practice Direction 1.5.
[6] Civil Procedure Rules (UK) Rule 35.10.
[7] R v Ward [1993]. 1 *WLR* 619.

Related Articles

Discovery in the United States: Civil Cases

Discovery in the United States: Criminal Cases

RHONDA M. WHEATE

Repressed Memory Syndrome *see* Memory: Repressed

Repression ***see*** Memory: Repressed

Restorative Justice ***see*** Mental Health Courts

Retrograde Extrapolation: Alcohol ***see*** Reconstruction: Three Dimensional

Right to Mental Health Treatment ***see*** Treatment, Right to: Mental Health

Right to Refuse Treatment ***see*** Therapeutic Jurisprudence

Right to Refuse Treatment: Mental Health ***see*** Treatment, Right to Refuse: Mental Health

Rigor Mortis ***see*** Death: Time of

Risk Assessment

Risk assessment comprises the initial and periodic steps taken in determining the quantitative and qualitative risk related to a concrete situation from a recognized threat. Risk assessment is applied in a wide range of situations from public health to presidential protection. In the context of behavioral forensics, risk assessment can be viewed from two perspectives.

The first perspective relates to threat assessment or analysis of threat as it relates to the vulnerability of an individual, facility, or network to outside attack by an individual or group of individuals whose identity may or may not be known. (*see* **Threat Assessment: School** and **Threat Assessment: Workplace**).

The second perspective is risk assessment, which relates to the assessment of a person or persons who may present a threat. (*see* **Dangerousness: Risk of**; **Risk Assessment: Patient and Detainee**; and also **Duty to Warn**).

The principles and tools of risk and threat assessment often overlap, and have evolved steadily over the past half decade. Techniques that prove effective and appropriate in one context may not be applicable to another, however, and, therefore, procedures should be viewed in context and in relationship to specific objectives. Moreover, there are settings in which a professional assessment may both involve objective questions of external risk as well as threat to the evaluator and professional colleagues (*see,* e.g., **Violence Risk Assessment for Mental Health Professionals**).

Related Articles

Dangerousness: Risk of

CARL N. EDWARDS

Risk Assessment: Mental Health Settings ***see*** Violence Risk Assessment for Mental Health Professionals

Risk Assessment: Patient and Detainee

Risk assessment is required in a variety of legal and clinical settings. Nevertheless, there is still substantial disagreement about its scientific validity and how best to implement it into practice, which has led to extensive empirical research into its tenability. Violence risk factors can be grouped into four categories: (i) dispositional, reflecting personal traits and tendencies to act in a particular way; (ii) historical, indexing events in the past that may predispose a person to become violent; (iii) contextual, referring to elements in the current environment that are conducive to violence; and (iv) clinical, including features of mental illness, level of functioning, and substance abuse (*see also* **Duty to Warn**; **Threat Assessment: Workplace**).

Over the years the primary goal of risk assessment has shifted from simple determination of who would become violent in the future, to risk management, which allows for reduction of violence. There are two primary approaches to risk assessment: unstructured, based solely on clinical experience and professional opinion, and structured, which requires professionals to follow specific rules of identification and definition of risk factors. Structured risk assessment in turn includes two approaches: (i) actuarial, which relies on an algorithm to combine risk factors, which enter the final decision regarding future violence, and (ii) structured professional judgments (SPJs), which allow the decision maker to consider how nomothetically supported risk factors are relevant to a given individual, and what management strategies would mitigate risk. Commentary and research addressing the strengths and weaknesses of these approaches is provided.

Risk Assessment

History and Context

The task of risk assessment has come to occupy one of the more controversial positions in corrections, forensic mental health, and related fields. On the one hand, it is legally or professionally required in dozens of contexts [1, 2]. On the other hand, there remains substantial disagreement about the scientific validity of this enterprise [3, 4]. Further, even among researchers and professionals who support the enterprise in principle, there is disagreement about the best way to put it into practice (compare and contrast [5–7]).

Until the late 1960s, it was assumed that psychiatrists and other mental health professionals simply had the ability to forecast the violence of their patients [8]. A seminal and groundbreaking series of studies and commentaries in the late 1960s and early 1970s, however, seriously challenged this assumption [9–14]. In fact, according to Ennis and Litwack [10], we were merely "flipping coins in the courtroom".

Yet, various legal pressures and movements solidified a place for violence risk assessments within the law. The groundbreaking case of *Tarasoff v the Regents of the University of California* [15] established a common law duty to protect certain third parties from the violence of patients and clients. Although this duty has changed over the years, and varies according to jurisdiction, it essentially remains "good law" in many jurisdictions. Numerous legislative enactments, such as Sexually Violent Predator acts in the United States, Dangerous and Long-Term Offender law in Canada, and Dangerous Severe Personality Disorder law in the United Kingdom, all require formal risk assessments to inform the legal decision-making process (for a detailed explication of the legal and clinical evolution of the concept of "dangerousness", see **Dangerousness: Risk of**).

Needless to say, such legal entrenchment has spurred volumes of empirical research into the tenability of violence risk assessment. Early research focused on its weaknesses, claiming that mental health and other professionals should have no place in violence risk assessment [10]. More moderate professional opinion asserted that, as yet, there was not a solid demonstration of the accuracy of risk assessment, but that certain approaches (i.e., empirically based short-term predictions of violence) held more promise than others (longer term predictions of the amorphous construct of "dangerousness") [12]. Monahan's book [12] was influential in terms of charting a course for risk assessment research to investigate specific risk factors, and more empirically based predictions, as opposed to clinically based predictions. In what follows, we review what has been accomplished in the risk assessment field, starting with a review of promising violence risk factors, and then

moving to a description of the common approaches to violence risk assessment.

Risk Factors

Although the existing research corpus on violence risk factors is extensive, we categorize these into the following categories presented by Monahan *et al.* [16]. They described four areas of risk factors: (i) personal or dispositional; (ii) historical or developmental; (iii) contextual or situational; and (iv) clinical or symptom-related factors. Following is a brief description of each of the risk categories along with a summary of the current research findings on this topic.

Personal or Dispositional Risk Factors. This category of risk factors reflects personal predispositions, traits, and tendencies to behave in a particular way that elevates one's risk for violence [17]. They are thought to be generally enduring throughout the lifespan. There are a number of risk factors that fall within this domain and could be further grouped as (i) demographic variables such as age, gender, and race, (ii) personality variables such as impulsivity, anger, and psychopathy, and (iii) neurological factors such as head injury [16]. The most prominent risk markers for violence in this category, however, are impulsivity, anger, and psychopathy. On the basis of research by Raymond Novaco, anger problems and lacking the ability to control its expression have been closely linked to violence [17]. This was found to be particularly true for patients in clinical settings [17].

Further, impulsivity is often linked to aggression, and as such is considered to be a prime precursor to violence. According to Ernest Barratt, who has done extensive research on this topic, personality traits of impulsiveness and anger-hostility are associated with a particular form of aggression – namely, "impulsive aggression" [17]. Individuals possessing such personality traits are likely to experience learning difficulties and behavioral disinhibition [17]. Further, psychopathy is currently known to be a robust predictor of violence among individuals suffering from mental illness and those not so afflicted [18]. Finally, antisocial (procriminal) attitudes have been shown to be important risk factors [19].

Historical or Developmental Risk Factors. This category of risk factors essentially indexes events in the past that may affect a person and predispose him or her to act violently, and can be grouped into three subcategories: (i) social history, (ii) history of mental hospitalizations, and (iii) history of violence [17]. Such events can include modeling of violence by caregivers or authority figures, which eventually can lead people to become inclined to engage in violence. This can occur when children are the victims of abuse and/or neglect. In addition, factors that can moderate the occurrence of violence, such as parental supervision and nurturance, also fall within this domain. Further, parental loss resulting from separation, divorce, or death, as well as psychiatric hospitalization of caregivers, or familial history of drug abuse have also been linked to violence [20]. Such factors may be more detrimental for individuals under the age of 15 [16]. Furthermore, factors such as cognitive impairment are likely to interact with other factors and can essentially predispose people to act violently. Therefore, they too are included in the historical risk factor domain [20]. Also, prior violence and criminality are quite strongly related to risk for future violence, especially among individuals suffering from a mental illness [16]. Finally, factors such as group norms play a significant role in affecting the acceptability of violence, and are also incorporated in this domain. Examples of such factors include cultural and subcultural norms, or membership in particular peer groups, such as Nazis or Ku Klux Klan (KKK) groups [20].

Contextual or Situational Factors. This category includes elements in the current environment which may influence one's behavior and be conducive to violence [17]. There are a number of risk factors that fall within this category, but the most prominent ones can be grouped in the following subcategories: (i) composition of social networks and perceived social support, (ii) perceived stress, and (iii) available means for violence (e.g., weapons). Social networks comprised of high percentage of relatives and low percentage of mental health professionals appear to be positively correlated with violence [21], as do social networks comprised of antisocial peers [19]. Also, violence has been found to be linked to perceived threat and hostility from significant others, however, the predictive power of this association needs to be further examined [21]. Finally, high levels of stress, adverse neighborhood conditions and

easy access to weapons have been found to relate positively with violence incidents [16].

Clinical Risk Factors. Within this category are features of mental illness (i.e., symptoms as well as Axis I and II diagnoses), level of functioning, and substance abuse [17]. Mental illness, defined broadly, has been found to be only moderately related to violence. This relationship, however, appears to be much stronger if mental disorders are comorbid with substance abuse [16], which itself is an important risk factor, or when measured at the symptom as opposed to diagnostic level. Various studies have examined the relationship between hallucinations or delusions and violence [17]. It appears as though there is a positive relationship between hallucinations and violence, but it is nonetheless contingent upon many other factors, such as the course of the disorder or compliance with medications [17]. For instance, in a large-scale study of psychiatric patients, command hallucinations were associated with increased likelihood of violence only if the voices commanded the commission of a violent act [16]. Further, the association between delusions and mental disorders is somewhat unclear. Some studies support an association [22]. However, in others, the presence of delusions has failed to predict higher rates of violence, although nondelusional suspiciousness resulting from perceived hostility in others was positively related to violence [23]. Finally, the presence of certain personality disorders – such as antisocial (or psychopathic) or borderline–appears to be strongly associated with violence, more so than a diagnosis of a major mental illness [16].

It is important to clarify the distinction between risk domains that play role solely in risk assessment (relatively stable, or static, risk factors) versus those that can be used for the purposes of both risk assessment and management (changeable, or dynamic risk factors) [24]. Personal and historical risk factors generally fall within the domains that are mainly relevant to risk assessment. Though they may influence the intensity of risk management, they do little to direct the specific type of risk management that is needed to curtail risk [24]. Contextual and clinical factors, on the other hand, are not only relevant for risk assessment, but are also malleable, and hence important for the purposes of risk management. Since these factors can change owing to direct interventions from clinicians or as a function of circumstances surrounding the individual, they should be assessed on an ongoing basis, and form an important part of risk management [24].

The concept of dynamic risk factors – or, as Andrews and Bonta [19] have termed it, criminogenic needs – plays an important role in more contemporary models of risk assessment and management, a topic to which we now turn. We provide a brief overview of the primary models of risk assessment, how they have evolved over time, and what their main strengths and weaknesses are.

Risk Assessment Models

Risk assessment allows professionals to make informed decisions about individuals' risk for future violence, which is required in numerous legal or clinical settings. Traditionally, the primary goal of risk assessment was to determine who will become violent in the future. Recently, there has been a drastic shift within the risk assessment field in the direction of risk management where the primary objective is not only to estimate future likelihood of violence, but to assess risk factors' variability over time, so as to determine strategies that allow reduction of violence and mitigation of risk [24–27]. This reconceptualization of the primary goals of risk assessment is reflected in the evolution of risk assessment models.

There are two primary approaches to risk assessment that have been employed by professionals: unstructured and structured. Unstructured risk assessment, also known as *clinical prediction or judgment*, is based solely on clinical experience and professional opinion, and does not involve any rules for decision making [28]. Essentially, it embodies Meehl's original description of clinical prediction [29]. Structured risk assessment, on the other hand, requires professionals to follow specific rules in terms of identification and definition of risk factors, and to base their decision about the risk of future violence on those rules [28]. Structured risk assessment includes two approaches: (i) actuarial, and (ii) SPJ. In general, a structured approach includes a fixed set of risk factors and their operational definitions, and provides directions for scoring, coding, and integration of risk factors in the final decision [30]. However, while the former approach (i.e., actuarial) relies on an algorithm that is used to determine the final risk decision on the basis of a rule-based combination of risk factors,

the latter does not. Instead, SPJ requires the decision maker to determine the level of risk on the basis of the specific risk factors present, their individual relevance to the case at hand, and the anticipated degree of intervention required to mitigate risk [26]. Next, we address the strengths, weaknesses, and empirical evaluation of each of these three approaches.

Unstructured Clinical Prediction. Unstructured professional judgment probably has been the most commonly used "method" for the assessment of risk for violence (perhaps owing to a lack of validated, structured tools until well into the 1990s). Technically, it is defined by the absence of combinatory rules at the item integration phase of decision making [29]. There clearly are strengths to a purely clinical approach – it can be used in any setting, it facilitates attention to the individual case, and it tends to be highly relevant to treatment and management considerations.

Nevertheless, this approach is subject to serious limitations when it comes to forecasting the risk of any future behavior [5, 31]. First, given a lack of rules, clinicians often rely on dissimilar sources of information, which inevitably leads to lack of consistency between professionals. In addition, there is often disparity between the risk factors that are taken into consideration, where on occasion important factors are being omitted and instead irrelevant factors are attended to. Further, unstructured professional judgment is highly subjective and informal, and often does not improve upon chance [28]. Finally, the lack of standardization of risk assessment results in low accuracy owing to bias susceptibility and lack of transparency of decision making, which is particularly relevant in legal settings [30]. Currently, there appears to be an agreement among professionals that sole reliance on unstructured clinical judgment is inadequate for comprehensive and reliable risk assessment [32].

Empirical evaluation of the unstructured clinical prediction method supports these criticisms. Early research did not paint a favorable picture of such unaided clinical judgments [9, 33] and later investigations have exhibited mixed findings. Whereas the results of a number of studies provided some – if only limited – support for clinical risk assessments [34–41], many others failed to support this method [42–47]. As such, risk assessment researchers reached the consensus that the lack of procedural guidance characteristic of unstructured clinical assessments generates inadequate interrater reliability which, in turn, represents a major threat to the predictive validity of violence risk evaluations [5, 32, 48]. Consequently, structured decision-making approaches, in the form of actuarial or SPJ frameworks, have been forwarded. We review these in turn.

Actuarial Prediction. To address the shortcomings of a purely unstructured risk assessment approach, professional attention turned to the actuarial method. Technically, this involves an algorithmic combination of risk factors [29], but most (though not all – [49]) actuarial risk assessment schemes also involve empirical item selection. That is, items are selected to be risk factors on a certain risk assessment instrument because they have shown an empirical association with violence, typically in a single sample. Actuarial procedures, by relying strictly on an algorithm to combine risk factors or to make a final decision regarding future violence, tend to improve the consistency (between and within evaluators) and validity of predictions [50]. Further, since the rules for inclusion of risk factors in any given assessment remain unchanged across assessments, actuarial risk assessment approaches can be easily reviewed, or are transparent, which is a desirable feature of a risk assessment procedure.

A number of actuarial violence risk assessment tools have been developed. For instance, Quinsey and colleagues developed the "Violence Risk Appraisal Guide (VRAG)" [5, 51, 52]. Scores on this instrument were highly predictive of future violence in their construction sample (multiple $R = 0.44$). Generally, subsequent studies of the VRAG tend to show that it bears, on average, a moderate relationship to violence. However, some results evidenced generalizability concerns, particularly as it pertains to the performance of the VRAG probability bins that purport to offer precise estimates (i.e., 54%) of future violence [53]. Monahan and colleagues [6, 17, 54–57] developed a risk assessment instrument on the basis of the MacArthur study [6, 16] that, in its original sample, showed highly impressive predictive accuracy – 1% of the low-risk group was violent, compared to 76% of its high-risk group. Upon cross-validation in a new sample, however, predictive accuracy degraded substantially. Now, 9% of the low-risk group was violent, and 35–49% of the

high-risk group was violent [6]. The area under the curve dropped from 0.88 to 0.63–0.70.

These findings illustrate one of the key weaknesses of a purely actuarial approach–precision estimates based on a given sample may not generalize to new samples. In fact, the psychometric properties of actuarial risk assessment instruments cannot be guaranteed for samples other than the sample of development for which those instruments were optimized [32]. Further, actuarial approaches are rather rigid and often fail to incorporate relevant information simply because such information is not contained in the predetermined set of risk factors [58], which, again, is often based on a single sample of participants. Similarly, low base rate risk factors are often ignored by actuarial approaches. Furthermore, even though it would be theoretically possible to incorporate dynamic risk factors, most actuarial models tend to focus mainly on static risk factors, and hence are less relevant to risk management and the reduction of violence [24]. Finally, actuarial procedures that attempt to apply group-derived probability estimates at the individual level may produce highly error-prone estimates [7]. To contend with these actuarial difficulties, some researchers have developed another structured approach to risk assessment and management – SPJ.

Structured Professional Judgment. The SPJ approach to risk assessment was developed to address the limitations of actuarial and unstructured risk assessment. As in actuarial practices, the professional conducting the assessment follows theoretically, clinically, and empirically based guidelines which recommend that a fixed set of risk factors be considered [26]. There are two main differences between SPJ and actuarial approaches. First, whereas most (though not all) actuarial approaches select risk factors based on their performance in particular (usually single) samples, all SPJ measures select risk factors rationally on the basis of a review of their performance in the scientific violence literature broadly. In principle, this practice should enhance generalizability of SPJ approaches. Second, SPJ approaches do not base risk decisions on algorithms or cutoff scores, or offer specific numeric estimates of risk. This is done to avoid the pitfalls associated with such practices, including the generalizability problem mentioned earlier and the tendency of specific numeric estimates to vary considerably between samples [53].

SPJ approaches have been criticized for permitting discretion to enter the decision-making process [5]. The argument is that discretion lowers reliability and validity. However, discretion is incorporated within a structured framework, and is intended to prompt the decision maker to consider how nomothetically supported risk factors are relevant for a given individual, and what management strategies are necessary to mitigate risk.

In fact, empirical evaluation of SPJ approaches tends to support both their interrater reliability, and their predictive validity [30]. All in all, SPJ practices are more structured than traditional clinical prediction, but they also provide more flexibility than actuarial assessment practices in the final decision making by allowing professional discretion in combining risk factors to fit each individual case. Thus, the SPJ model has been forwarded to (i) maintain reliability and validity, and also to (ii) enhance the individual applicability of risk assessment and management [26]. The following section provides a brief summary of the literature on two of the most commonly used SPJ schemes.

The "historical clinical risk management-20" HCR-20 [59], one of the most researched and established SPJ guides, has demonstrated good interrater reliability, construct validity, and concurrent validity with other commonly used instruments, and generates moderate to large effect sizes with regard to violent recidivism [8, 32, 48, 60–68]. Moreover, the predictions based on the SPJs of low-, moderate-, or high-risk exhibit incremental validity beyond those solely based on the HCR-20 total numeric scores. Hence, this SPJ model of decision-making appears to yield useful structured clinical decisions mentioned earlier and beyond what can be anticipated from the actuarial scores on the instrument alone [32, 62, 69].

Another SPJ scheme, The "short-term assessment of risk and treatability" (START) [70], demonstrates promising results in terms of reliability and validity analyses [71]. In addition to the predictive accuracy of the instrument, its associated multidimensional model of violence risk assessment offers uniquely meaningful possibilities in terms of the preventative goals of such psychosocial and forensic enterprises [72]. One of the ways in which it does so is by incorporating patient strengths and weaknesses into its framework.

Conclusion

Both the actuarial and SPJ models of risk assessment hold promise for improving clinical risk assessments. The actuarial approach, however, seems best suited to group decision making, and is hampered by a potential lack of generalizability, and difficulty in meaningfully applying nomothetic data to individual cases. The reliability and validity evaluation of the SPJ approach is promising, and also attempts to enhance generalizability, individual relevance, and risk reduction. Future research should continue to evaluate the strengths and weaknesses of these two approaches, with the eventual goal of minimizing weaknesses associated with both approaches, but facilitating their strengths.

References

[1] Lyon, D.R., Hart, S.D. & Webster, C.D. (2001). Violence risk assessment, in *Law and Psychology: Canadian Perspectives*, R. Schuller & J.R.P. Ogloff, eds, University of Toronto Press, Toronto, pp. 314–350.

[2] Shah, S.A. (1978). Dangerousness and mental illness: some conceptual, prediction and policy dilemmas, in *Dangerous Behavior: A Problem in Law and Mental Health*, NIMH DHEW Publication No (ADM) 78-563), C. Frederick, ed, Superintendent of Documents, Government Printing Office, Washington, DC, pp. 153–191.

[3] Vrieze, S.I. & Grove, W.M. (2007). Predicting sex offender recidivism. I. Correcting for item overselection and accuracy overestimation in scale development. II. Sampling error-induced attenuation of predictive validity over base rate information, *Law and Human Behavior* [serial on the Internet]. Jul 3 [cited 2007 October 29]. Available from: http://www.springerlink.com/content/m8n75n180q027353/.

[4] Mossman, D. Analyzing the performance of risk assessment instruments: a response to vrieze and grove, (2007). *Law and Human Behavior*, **32**(3), 279–291.

[5] Quinsey, V.L., Harris, G.T., Rice, M.E. & Cormier, C.A. (2006). *Violent Offenders: Appraising and Managing Risk*, 2nd Edition, APA, Washington, DC.

[6] Monahan, J., Steadman, H.J., Appelbaum, P.S., Grisso, T., Mulvey, E.P., Roth, L.H., Robbins, P.C., Banks, S. & Silver, E. (2005). *The Classification of Violence Risk*, Psychological Assessment Resources, Lutz.

[7] Hart, S.D., Michie, C. & Cooke, D.J. (2007). Precision of actuarial risk assessment instruments: evaluating the 'margins of error' of group v. individual predictions of violence, *British Journal of Psychiatry* **190**(49), s60–s65.

[8] Douglas, K.S. & Webster, C.D. (1999). Predicting violence in mentally and personality disordered individuals, in *Psychology and Law: The State of the Discipline*, R. Roesch, S.D. Hart & J.R.P. Ogloff, eds, Plenum, New York, pp. 175–239.

[9] Cocozza, J.J. & Steadman, H.J. (1976). The failure of psychiatric predictions of dangerousness: clear and convincing evidence, *Rutgers Law Review* **29**, 1084–1101.

[10] Ennis, B.J. & Litwack, T.R. (1974). Psychiatry and the presumption of expertise: flipping coins in the courtroom, *California Law Review* **62**, 693–752.

[11] Kozol, H.L., Boucher, R.J. & Garofalo, R.F. (1972). The diagnosis and treatment of dangerousness, *Crime and Delinquency* **18**, 371–392.

[12] Monahan, J. (1981). *Predicting Violent Behavior: An Assessment of Clinical Techniques*, Sage, Beverly Hills.

[13] Steadman, H.J. & Cocozza, J.J. (1974). *Careers of the Criminally Insane: Excessive Social Control of Deviance*, Lexington Books, Lexington.

[14] Thornberry, T.P. & Jacoby, J.E. (1979). *The Criminally Insane: A Community Follow-Up of Mentally Ill Offenders*, University of Chicago Press, Chicago.

[15] Tarasoff v. Regents of the University of California, 551 P.2d 334 (1976).

[16] Monahan, J., Steadman, H., Silver, E., Appelbaum, P., Robbins, P., Mulvey, E., Roth, L., Grisso, T. & Banks, S. (2001). *Rethinking Risk Assessment: The MacArthur Study of Mental Disorder and Violence*, Oxford University Press, New York.

[17] Monahan, J. & Steadman, H. (eds) (1994). *Violence and Mental Disorder: Developments in Risk Assessment*, University of Chicago Press, Chicago.

[18] Douglas, K.S., Vincent, G.M. & Edens, J.F. (2006). Risk for criminal recidivism: the role of psychopathy, in *Handbook of the Psychopathy*, C.J. Patrick, ed, Guilford Press, New York, pp. 533–554.

[19] Andrews, D.A. & Bonta, J. (2003). *The Psychology of Criminal Conduct*, 3rd Edition, Anderson Publishing, Cincinnati.

[20] Klassen, D. & O'Connor, W.A. (1994). Demographic and case history variables in risk assessment, in *Violence and Mental Disorder: Developments in Risk Assessment*, J. Monahan & H.J. Steadman, eds, University of Chicago Press, Chicago, pp. 229–257.

[21] Estroff, S.E. & Zimmer, C. (1994). Social networks, social support, and violence among persons with severe, persistent mental illness, in *Violence and Mental Disorder: Developments in Risk Assessment*, J. Monahan & H.J. Steadman, eds, University of Chicago Press, Chicago, pp. 259–295.

[22] Swanson, J., Borum, R., Swartz, M. & Monahan, J. (1996). Psychotic symptoms and disorders and the risk of violent behavior in the community, *Criminal Behaviour and Mental Health* **6**, 317–338.

[23] Appelbaum, P., Robbins, P.C. & Monahan, J. (2000). Violence and delusions: data from the McArthur violence risk assessment study, *American Journal of Psychiatry* **157**, 566–573.

[24] Douglas, K.S. & Skeem, J.L. (2005). Violence risk assessment: getting specific about being dynamic, *Psychology Public Policy and Law* **11**, 347–383.

[25] Dvoskin, J.A. & Heilbrun, K. (2001). Risk assessment and release decision-making: toward resolving the great debate, *Journal of American Academy of Psychiatry and Law* **29**, 6–10.

[26] Douglas, K.S. & Kropp, P.R. (2002). A prevention-based paradigm for violence risk assessment: clinical and research applications, *Criminal Justice and Behaviour* **29**(5), 617–658.

[27] Skeem, J. & Mulvey, E. (2002). Monitoring the violence potential of mentally disordered offenders being treated in the community, in *Care of the Mentally Disordered Offender in the Community*, A. Buchanan, ed, Oxford Press, New York, pp. 111–142.

[28] Douglas, K.S. & Lavoie, J. (2006). Violence risk assessment and management: models of use and guiding principles, in *Psicologia Forense*, A.C. Fonseca, ed, Nova Almedina, Coimbra.

[29] Meehl, P.E. (1954). *Clinical Versus Statistical Prediction*, University of Minnesota Press, Minneapolis.

[30] Lavoie, J.A. & Douglas K.S. Approaches to violence risk assessment, in *Encyclopedia of Psychology and Law*, B. Cutler, ed, Sage, Thousand Oaks, in press.

[31] Grove, W.M. & Meehl, P.E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: the clinical-statistical controversy, *Psychology Public Policy and Law* **2**, 293–323.

[32] Douglas, K.S., Ogloff, J.R.P. & Hart, S.D. (2003). Evaluation of a model of violence risk assessment among forensic psychiatric patients, *Psychiatric Services* **54**, 1372–1379.

[33] Litwack, T.R. (2001). Actuarial versus clinical assessments of dangerousness, *Psychology Public Policy and Law*. **7**, 409–443.

[34] Hoptman, M.J., Yates, K.F., Patalinjug, M.B. & Marilou, B. (1999). Clinical prediction of assaultive behavior among male psychiatric patients at a maximum-security forensic facility, *Psychiatric Services* **50**(11), 1461–1466.

[35] Lidz, C.W., Mulvey, E.P. & Gardner, W. (1993). The accuracy of predictions of violence to others, *Journal of the American Medical Association* **269**, 1007–1011.

[36] McNiel, D.E. & Binder, R.L. (1987). Predictive validity of judgments of dangerousness in emergency civil commitment, *American Journal of Psychiatry* **144**(2), 197–200.

[37] McNiel, D.E. & Binder, R.L. (1995). Correlates of accuracy in the assessment of psychiatric inpatients' risk of violence, *American Journal of Psychiatry* **152**, 901–906.

[38] McNiel, D.E., Binder, R.L. & Greenfield, T.K. (1988). Predictors of violence in civilly committed acute psychiatric patients, *American Journal of Psychiatry* **145**(8), 965–970.

[39] McNiel, D.E., Lam, J.N. & Binder, R.L. (2000). Relevance of interrater agreement to violence risk assessment, *Journal of Consulting and Clinical Psychology* **68**, 1111–1115.

[40] McNiel, D.E., Sandberg, D.A. & Binder, R.L. (1998). The relationship between confidence and accuracy in clinical assessment of psychiatric patients' potential for violence, *Law and Human Behaviour* **22**(6), 655–669.

[41] Rofman, E.S., Askinazi, C. & Fant, E. (1980). The prediction of dangerous behavior in emergency civil commitment, *American Journal of Psychiatry* **137**(9), 1061–1064.

[42] Janofsky, J.S., Spears, S. & Neubauer, D.N. (1988). Psychiatrists' accuracy in predicting violent behavior on an inpatient unit, *Hospital Community and Psychiatry* **39**(10), 1090–1094.

[43] Menzies, R.J., Webster, C.D., McMain, S., Staley, S. & Scaglione, R. (1994). The dimensions of dangerousness revisited: assessing forensic predictions about violence, *Law and Human Behaviour* **18**, 1–28.

[44] Rabinowitz, J. & Garelik-Wyler, R. (1999). Accuracy and confidence in clinical assessment of psychiatric inpatients risk of violence, *International Journal of Law and Psychiatry* **22**, 99–106.

[45] Werner, P.D., Rose, T.L. & Yesavage, J.A. (1983). Reliability, accuracy, and decision-making strategy in clinical predictions of imminent dangerousness, *Journal of Consulting and Clinical Psychology* **51**(6), 815–825.

[46] Werner, P.D., Rose, T.L., Yesavage, J.A. & Seeman, K. (1984). Psychiatrists' judgments of dangerousness in patients on an acute care unit, *American Journal of Psychiatry* **141**, 263–266.

[47] Yesavage, J.A., Werner, P.D., Becker, J. & Mills, M. (1982). Short-term civil commitment and the violent patient, *American Journal of Psychiatry* **139**, 1145–1149.

[48] Douglas, K.S., Yeomans, M. & Boer, D.P. (2005). Comparative analysis of multiple measures of violence risk in a sample of criminal offenders, *Criminal Justice and Behaviour* **32**, 479–510.

[49] Andrews, D.A. & Bonta, J. (1995). *The Level of Service Inventory-Revised (LSI-R)*, Multi-Health Systems, Toronto.

[50] Grove, W.M., Zald, D.H., Lebow, B.S., Snitz, B.E. & Nelson, C. (2000). Clinical versus mechanical prediction: a meta-analysis, *Psychological Assessment* **12**, 19–30.

[51] Harris, G.T., Rice, M.E. & Quinsey, V.L. (1993). Violent recidivism of mentally disordered offenders: the development of a statistical prediction instrument, *Criminal Justice and Behaviour* **20**, 315–335.

[52] Quinsey, V.L., Harris, G.T., Rice, M.E. & Cormier, C.A. (1998). *Violent Offenders: Appraising and Managing Risk*, American Psychological Association, Washington, DC.

[53] Mills, J.F., Jones, M.N. & Kroner, D.G. (2005). An examination of the generalizability of the LSI-R and

VRAG probability bins, *Criminal Justice and Behaviour* **32**, 565–585.
[54] Monahan, J., Steadman, H.J., Appelbaum, P.S., Robbins, P.C., Mulvey, E.P., Silver, E., Roth, L.H. & Grisso, T. (2000). Developing a clinically useful actuarial tool for assessing violence risk, *The British Journal of Psychiatry* **176**, 312–319.
[55] Monahan, J. & Steadman, H.J. (2001). Violence risk assessment: a quarter century of research, in *The Evolution of Mental Health Law*, L.E. Frost & R.J. Bonnie, eds, American Psychological Association, Washington, DC, pp. 195–211.
[56] Monahan, J., Steadman, H.J., Robbins, P., Applebaum, P.S., Banks, S., Grisso, T., Heilbrun, K., Mulvey, E.P., Roth, L. & Silver, E. (2005a). An actuarial model of violence risk assessment for persons with mental disorders, *Psychiatric Services* **56**, 810–815.
[57] Monahan, J., Steadman, H.J., Applebaum, P.S., Grisso, T., Mulvey, E.P., Rothe, L.H., Robbins, P.C., Banks, S. & Silver, E. (2006). The classification of violence risk, *Behavioral Sciences and the Law* **24**, 721–730.
[58] Hart, S.D. (1998). The role of psychopathy in assessing risk for violence: conceptual and methodological issues, *Legal and Criminological Psychology* **3**, 121–137.
[59] Webster, C.D., Douglas, K.S., Eaves, D. & Hart, S.D. (1997). *HCR-20 Scheme: Assessing Risk for Violence (Version 2)*, Mental Health, Law, and Policy Institute, Simon Fraser University, Burnaby.
[60] Belfrage, H., Fransson, G. & Strand, S. (2004). Management of violent behaviour in the correctional system using qualified risk assessments, *Legal and Criminological Psychology* **9**(1), 11–22.
[61] Bloom, H., Webster, C., Hucker, S. & De Freitas, K. (2005). The Canadian contribution to violence risk assessment: history and implications for current psychiatric practice, *Canadian Journal of Psychiatry* **50**(1), 3–11.
[62] de Vogel, V. & de Ruiter, C. (2006). Structured professional judgment of violence risk clinical practice: a prospective study into the predictive validity of the Dutch HCR-20, *Psychology, Crime and Law* **12**(3), 321–336.
[63] Douglas, K.S., Ogloff, J.R.P., Nicholls, T.L. & Grant, I. (1999). Assessing risk for violence among psychiatric patients: the HCR-20 violence risk assessment scheme and the Psychopathy Checklist: Screening Version, *Journal of Consulting and Clinical Psychology* **67**, 917–930.
[64] Doyle, M. & Dolan, M. (2006). Predicting community violence from patients discharged from mental health services, *The British Journal of Psychiatry* **189**, 520–526.
[65] Gray, N.S., Hill, C., McGleish, A., Timmons, D., MacCulloch, M.J. & Snowden, R.J. (2003). Prediction of violence and self-harm in mentally disordered offenders: a prospective study of the efficacy of HCR-20, PCL-R, and psychiatric symptomatology, *Journal of Consulting and Clinical Psychology* **71**(3), 443–451.
[66] Nicholls, T.L., Ogloff, J.R.P. & Douglas, K.S. (2004). Assessing risk for violence among male and female civil psychiatric patients: the HCR-20, PCL:SV, and VSC, *Behavioral Sciences and the Law* **22**, 127–158.
[67] Nikolova, N.L., Collins, M.J., Guy, L.S., Lavoie, J.A.A., Reeves, K.A., Wilson, C.M. & Douglas, K.S. (2006). HCR-20 violence risk assessment scheme: quantitative synthesis of its application, reliability, and validity. *Poster Presented at the Annual Conference of the American Psychology-Law Society*; 2006 March, St. Petersburg, p. 62.
[68] Otto, R.K. (2000). Assessing and managing violence risk in outpatient settings, *Journal of Clinical Psychology* **56**(10), 1239–1262.
[69] de Vogel, V., de Ruiter, C., Hildebrand, M., Bos, B. & van de Ven, P. (2004). Type of discharge and risk of recidivism measured by the HCR-20: a retrospective study in a Dutch sample of treated forensic psychiatric patients, *International Journal of Forensic Mental Health* **3**, 149–165.
[70] Webster, C.D., Martin, M.-L., Brink, J., Nicholls, T. & Middleton, C. (2004). *START: The Short Term Assessment of Risk and Treatability*, St Joseph's Healthcare, Hamilton.
[71] Nicholls, T.L., Brink, J., Desmarais, S.L., Webster, C.D. & Martin, M.-L. (2006). The Short-Term Assessment of Risk and Treatability (START): a prospective validation study in a forensic psychiatric sample, *Assessment* **13**(3), 313–327.
[72] Webster, C.D., Nicholls, T.L., Martin, M.-L., Desmarais, S.L. & Brink, J. (2006). Short-Term Assessment of Risk and Treatability (START): the case for a new structured professional judgment scheme, *Behavioral Sciences and the Law* **24**, 747–766.

NATALIA L. NIKOLOVA, DIANE S. STRUB AND KEVIN S. DOUGLAS

Risk Factors: Elder Abuse ***see*** Elder Abuse: Risk

Risk Factors: Violence ***see*** Risk Assessment: Patient and Detainee

Rule 702 (USA) ***see*** Federal Rule of Evidence 702



Saliva: Drugs ***see*** Oral Fluid Toxicology

Sampling and Estimation of Quantities

Introduction

A sample is a subset of a population. The population is assumed to consist of discrete units. A sample is a subset of these discrete units. One example is that the units may be bags and the bags may contain powder, as in a drugs case. Another example is that the units may be computer files and the files may contain pornographic images.

Let the population size be N. A census would inspect all N items in the population. There are several disadvantages to a census.

- If N is large, inaccuracies inevitably occur in the inspection process because of resource constraints, hence exact results cannot be guaranteed.
- If testing is destructive, no evidence would remain for any trial.
- Inspection of all items may expose the inspectors to potential health hazards, for example, from airborne dust, physical contact, or mental stress.

Two requirements were given by Wigmore [1] for a sample, presented to show the quality or condition of the population from which it was taken, to be admissible in evidence. They are as follows:

- The population should be substantially uniform with reference to the quality in question (homogeneous).
- The sample should be of such a nature that it is fairly representative of the population.

There are three types of sampling that can be considered in a forensic context. These are simple random sampling, stratified or multistage sampling, and composite sampling. The bulk of the article considers inferences from simple random samples. The extensions to other forms of samples have to be considered carefully.

A simple random sample is one in which all members of the population are equally likely to be selected. A stratified sample is one in which there are several layers or strata in the population. For example, there may be several containers and many packets within each container. The stratified sample would take a random sample of containers and then a random sample of packets from within the random sample of containers.

Consider the following example involving a drug seizure as an illustration of a composite sample [2]. Assume that there are several packets containing tablets. A random sample of the packets is taken. From the contents of each selected packet, a certain number of core samples of tablets are removed. These are combined, placed into a mortar, and ground up

with a pestle into a fine powder. This is the composite sample. The powder is mixed by repeated tossing and stirring to look as homogeneous as possible. A number of subsamples are randomly scooped up from around the mixture and placed into weighing dishes where they are tested for the presence of an illicit drug. If a drug is present in at least one of the core samples prior to mixing, then a subsample from that mixture will, in high probability, test positive for that drug. If the subsample tests negative for that drug, then no further sampling or testing would be carried out and all packets would be declared free of that drug.

Sample sizes and the interpretation of data from samples are discussed from both the frequentist and the Bayesian perspectives.

With reasonable assumptions in the Bayesian context, a probability distribution for the true proportion of illicit units in a consignment may be derived on the basis of the scientist's prior beliefs (i.e., prior to the inspection of individual units) and the outcome of the inspection of the sample. The strength of the scientist's prior beliefs may be expressed by a probability density function. For example, if the scientist believes that the consignment is either all illicit or all licit with only a small probability of anything in-between, the probability density function that represents the scientist's prior beliefs would be U-shaped on the interval (0,1) (corresponding to 0–100%) with most weight at the 0 and the 1 ends with little weight in-between. It is possible to choose the function in such a way that the effect of the scientist's prior beliefs is very small (or very large). The choice of the model that is used to represent the uncertainty introduced by the sampling process is a subjective choice influenced by the scientist's prior beliefs. The choice of a binomial model, for example, requires assumptions about independence of the probability for each unit being illegal and the choice of a constant value for this probability.

The methods are illustrated with reference to sampling from consignments of drugs. However, they apply equally well to sampling in other forensic contexts, for example, glass [3] and pornographic images.

Simple Random Samples

The most well-known type of sample is the simple random sample. This is a sample in which every member of the sample has an equal probability of being selected at any draw from the population.

In order to draw a sample from a population, it is highly desirable to be able to associate a unique identifier with each member of the population. In certain populations, this is relatively straightforward. For example, a population of voters in an election is provided by the electoral register and the identifiers on the electoral register may be used. In other populations, it is not so straightforward. For example, in a large consignment of white tablets, it would be very difficult to associate a unique identifier to each tablet. When it is possible to construct a list, the list of unique identifiers is known as the *sampling frame*.

> In the use of random numbers, the material must first be broken up in some manner into *sampling units*. Moreover, each sampling unit must be identifiable by a serial number, actual or by some rule. For packaged articles, a rule is easy .... In the case of bulk material, ... the problem of defining usable sampling units must take place at an earlier stage of the manufacture, or in the process of moving the materials. [4]

Once the sampling frame has been constructed, the choice of a member from the population for inclusion in the sample is made at random. This can be done using electronic calculators, computers, tables of random numbers, or, in the special case of equal probability (0.5, 0.5) of selection or not, even by the simple toss of a fair coin.

It may be that it is not possible to take a simple random sample. If so, the following comments are also of relevance. The comments are made in the context of sampling for the estimation of allelic frequencies in DNA profiles but are applicable to other areas of forensic science.

> ... every case must be treated according to the circumstances within which it has occurred, and the next response is that it is always a matter of judgement .... In the last analysis, the scientist must also convince a court of the reasonableness of his or her inference within the circumstances as they are presented as evidence. ([5], pp. 44–45)

Some comments extracted from the appendix of ASTM E 105-04 are also of relevance in the discussion of randomness.

> Note that calculation of the margin of error or the risk in the use of the results of samples is possible only if the selection of the items for test is made at

> random .... For a method of sampling to be random it must satisfy statistical tests .... Randomness is obtained by positive action; a random selection is not merely a haphazard selection, nor one declared to be without bias .... One may declare that a lot of item is "thoroughly mixed" and hence that any portion, even the top layer, would give every item an equal chance of selection. In the absence of elaborate steps to mix the product, followed by careful tests for randomness, such assumptions are risky, as they often lead to wrong results.
>
> Nonrandom and judgment sampling are not of no value. If the material being inspected is known to vary but little, a 'grab' sample will be helpful in assessing the level of the characteristic concerned.
>
> Systematic sampling, in the absence of knowledge concerning the order of the material, does not permit a valid calculation of the standard error.

For small samples, the sampling units may be taken to be sampling without replacement. The hypergeometric approach is used [6] in the frequentist context. The Bayesian approach also has different methods for analyzing large and small samples.

The ratio of the sample size to the population size is known as the *sampling fraction*.

Criteria for Sample Size Determination

Frequentist Criteria for Determination of Sample Size

Frequentist procedures for choosing a sample size from a consignment are described in Tzidony and Ravreboy [7]. Distinction is drawn between an approach based on the binomial distribution and an approach based on the hypergeometric distribution. The former can be used for large consignments (greater than 50 in size) in which the sampling of units may be taken to be sampling with replacement.

Various institutions issue guidelines on sample size choice. These include the United Nations, the European Network of Forensic Science Institutes (ENFSI), and the Scientific Working Group for the Analysis of Drugs (SWGDRUG) of the US Department of Justice Drug Enforcement Administration.

United Nations guidelines differentiate between powders, licit tablets, and capsules where a high level of quality control may be assured, and illicit tablets and capsules where a high level of quality control is not assured. The guidelines are given in Table 1.

For small samples of opium, morphine, and heroin, the United Nations guidelines provide tables for sample sizes based on the hypergeometric distribution, using the criterion that an inference is made at the 95% confidence level and that 90% or more of the packages in an exhibit containing the controlled substance should be accepted as sufficient proof in such cases ([8][6]).

Guidelines on representative drug sampling are also provided by ENFSI [9], with comments.

- Sample all: advantage – 100% certainty about the composition of the population; disadvantage – excessive sample sizes for large populations.

Table 1 Sample size guidelines from the United Nations for various drugs

Consignment size N	Sample size	Drugs
Powders		
< 10	All	Methaqualone, cocaine, barbiturates
10–100	10	Cannabis, benzodiazepine, psilocybin
> 100	$\sqrt{N}$	Opium, morphine, heroin, mescaline
Tablets and capsules (licit) (du = dosage unit)		
$1 - 50$ du	$N/2$ (up to 20)	Methaqualone, barbiturates
$51 - 100$	20	Benzodiazepine, psilocybin
$101 - 1000$	30	Mescaline
> 1000	$\sqrt{N}$	
Tablets and capsules (illicit)		
≤ 10	All	Methaqualone
$11 - 27$	$3N/4$	Lysergide
≥ 28	$N/2$	
	$21 \leq N/2 \leq 50$	

- Proportion of N: advantage – simple; disadvantage – excessive sample sizes for large populations.
- A function of $\sqrt{N}$: advantage – widely accepted; disadvantage – too small for small populations, excessive sample sizes for large populations.
- $n = 20 + 10\%(N - 20)(N > 20)$: advantage – heterogeneous populations likely to be discovered before the analysis is complete; disadvantage – excessive sample sizes for large populations.

$$\text{for } N < x \qquad n = N$$
$$x \le N \le y \qquad n = z$$
$$N > y \qquad n = \sqrt{N}$$

(where x, y, and z are arbitrary numbers; $x < y$ and $x \le z < y$) advantage: United Nations Drug Control Program recommended method ($x = 10$, $y = 100$, $z = 10$); disadvantage – excessive sample sizes for large populations.

- $n = 1$: advantage – minimum amount of work; disadvantage – least amount of information on the characteristics of the seizure.

The Scientific Working Group for the Analysis of Drugs (SWGDRUG) of the US Department of Justice Drug Enforcement Administration recommendations for a sampling plan has several components.

- The issue of homogeneity.
- For inhomogeneous bulk material, several samples from different locations may be necessary to ensure that the test results are representative of the bulk material and false negative results are avoided.
- Statistical approaches are applicable when inferences are to be made about the whole population, as in
  - the probability that a given percentage of the population contains the drug of interest or is positive for a given characteristic,
  - the total net weight of the population is to be extrapolated from the weight of a sample.

Bayesian Criteria for Determination of a Sample Size

A Bayesian approach provides summaries in probabilistic terms such as

> "How big a sample should be taken for it to be said that there is a $100p\%$ probability that the proportion of units in the consignment which contains drugs is greater than $100\ \theta\%$?"

Or, for a particular case, with $p = 0.95$ and $\theta = 0.50$,

> "How big a sample should be taken for it to be said that there is a 95% probability that the proportion of units in the consignment which contains drugs is greater than 50%?"

See the SWGDRUG comments on the applicability of statistical approaches to illustrate the relevance of this approach.

Choice of Sample Size for Discrete Data

Large Consignments

A consignment of drugs containing N units is considered as a random sample from some superpopulation of units, which contain drugs. Let θ $(0 < \theta < 1)$ be the proportion of units in the superpopulation (of which the consignment is a member) that contain drugs. Let m be the number of units sampled from the consignment. The sampling fraction is m/N. Denote the numbers (out of m) that are found to contain drugs by $z(\le m)$. Then, the probability distribution for z, given m and θ, is the binomial distribution

$$Pr(z|m, \theta) = \binom{m}{z}\theta^z(1-\theta)^{m-z}; \quad z = 0, \ldots, m \tag{1}$$

The purpose of the inspection of the sample is to determine an estimate for θ.

A Frequentist Approach. The sample proportion that contains drugs $p = z/m$ provides an estimate of θ. The variance of p is given by Cochran [10] as

$$\frac{\theta(1-\theta)}{m}\left(\frac{N-m}{N-1}\right) \tag{2}$$

The term $\theta(1-\theta)/m$ is the standard deviation of a sample proportion z/m from the binomial distribution in (equation 2). The factor $(N-m)/(N-1)$ is known as the *finite population correction* (fpc). Provided the sampling fraction m/N is low, the size of the population has no direct effect on the precision of the estimate of θ. For example, a sample

of 500 from a population of 200 000 gives almost as precise an estimate of the population proportion as a sample of 500 from a population of 10 000. The estimated standard deviation of θ in the second case is 0.98 times the estimated standard deviation in the first case. Little is to be gained by increasing the sample size in proportion to the population size. Given a desired magnitude δ for the standard deviation, it is possible to choose the sample size necessary to achieve this. Set δ equal to $\sqrt{\{\theta(1-\theta)/m\}}$. Then, $m = \theta(1-\theta)/\delta^2 < 1/(4\delta^2)$, the maximum value attained when $\theta = 1/2$. A sample of size m equal to the smallest integer greater than $1/(4\delta^2)$ is the smallest sample size that satisfies the requirement that the standard deviation be less than δ.

A Bayesian Approach. This approach is based on combining a binomial likelihood for the number z of members of a sample of size m that contain drugs with a beta-distribution, with parameters α and β for the proportion θ of the superpopulation of which the consignment is a subset. The prior distribution for θ is taken to be a beta-distribution with the probability density function given by

$$f(\theta|\alpha,\beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)},$$
$$0 < \theta < 1,\ \alpha > 0,\ \beta > 0 \qquad (3)$$

denoted $B(\alpha,\beta)$, where

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \qquad (4)$$

and Γ is the gamma function, where

$$\Gamma(t+1) = t!\ \text{ for integer }\ t > 0 \qquad (5)$$

A scientist who believes that the consignment is either all illicit or all licit with a small probability of something in-between would choose values of α and β close to zero. Then, combining the beta prior (equation 3) with the binomial distribution (equation 1), the posterior distribution for θ given m, z, α, β is

$$\Pr(\theta|m,z,\alpha,\beta) = \frac{\theta^{\alpha+z-1}(1-\theta)^{\beta+m-z-1}}{B(\alpha+z,\beta+m-z)} \qquad (6)$$

a posterior beta distribution for θ, given the sample size m, the number of illicit units $z(\leq m)$, and prior parameters α and β.

Consider the criterion that the scientist wishes to be 95% certain that 50% or more of the consignment contains drugs when all units sampled contain drugs (i.e., $m = z$). Then the criterion may be written mathematically as

$$\Pr(\theta > 0.5|m,m,\alpha,\beta) = 0.95 \qquad (7)$$

where m replaces z as the second term to the right of the conditioning bar to show that $z = m$. This criterion may be written in integral form as

$$\int_{0.5}^{1} \frac{\theta^{m+\alpha-1}(1-\theta)^{\beta-1}\mathrm{d}\theta}{B(m+\alpha,\beta)} = 0.95 \qquad (8)$$

The general question of Section (3) in which p and θ are specified may be answered by finding the value of m, which solves the equation

$$\int_{\theta}^{1} \frac{\theta^{m+\alpha-1}(1-\theta)^{\beta-1}\,\mathrm{d}\theta}{B(m+\alpha,\beta)} = p \qquad (9)$$

Such integrals are easy to evaluate using standard statistical packages, values for m, α, and β being given. Given specified values for θ and p and values for α and β chosen from prior beliefs, the appropriate value of m to solve (equation 9) may be found by trial and error. See Table 2 for examples of values of p for given values of m, α, and β.

There may be concerns that it is very difficult for a scientist to formalize his prior beliefs. However, if α and β are small, large differences in the probabilities associated with the prior beliefs do not lead to large differences in the conclusions. This is not the case, however, if there is a probability of a misclassification error [11]. A frequentist approach using the hypergeometric distribution, with an adaptation to allow for false positives and false negatives is described in Faber *et al.* [12]. The methodology can be extended to allow for units that do not contain drugs.

The dependency of the sample size on the values of p and θ is illustrated in Table 3. The prior

Table 2 The probability that the proportion of drugs in a large consignment is greater than 50% for various sample sizes m and prior parameters α and β

		m			
α	β	2	3	4	5
1	1		0.94	0.97	
0.5	0.5	0.92	0.97	0.985	0.993
0.065	0.935		0.90	0.95	0.97

Permission needed from JFS

Table 3 The sample size required to be $100p\%$ certain that the proportion of units in the consignment that contains drugs is greater than θ, when all the units inspected are found to contain drugs. The prior parameters $\alpha = \beta = 1$

	p		
θ	0.90	0.95	0.99
0.5	3	4	6
0.6	4	5	9
0.7	6	8	12
0.8	10	13	20
0.9	21	28	43
0.95	44	58	89
0.99	229	298	458

Permission needed from JFS

parameters α and β are set equal to 1. Consider $p = 0.90$, 0.95, and 0.99 and consider values of $\theta = 0.5$, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99. The sample size m required to be $100p\%$ certain that θ is greater than the specified value is then given by the value of m that satisfies the equation

$$\Pr(\theta > \theta_0 | m, m, 1, 1) = 1 - \theta_0^{m+1} = p \quad (10)$$

a special case of (equation 9). The value of m is thus given by the smallest integer greater than

$$\left[\frac{\log(1-p)}{\log(\theta_0)}\right] - 1 \quad (11)$$

Variation in the prior beliefs, expressed through variation in the values of α and β, may have little influence on the conclusions, once some data have been observed. Figure 1 illustrates the prior probability that the true proportion of illegal units in a consignment is greater than a value θ, for $0 < \theta < 1$ for three choices of α and β. Figure 2 illustrates the posterior probability that the true proportion of illegal units in a consignment is greater than θ, for these choices of α and β, once four units have been examined and all are found to be illegal.

Small Consignments

Suppose now that the consignment size N is small, say $N \leq 50$. A sample of m units from the consignment is examined and $z(\leq m)$ units are found to contain drugs.

A Frequentist Approach. Consider a frequentist approach based on the hypergeometric distribution.

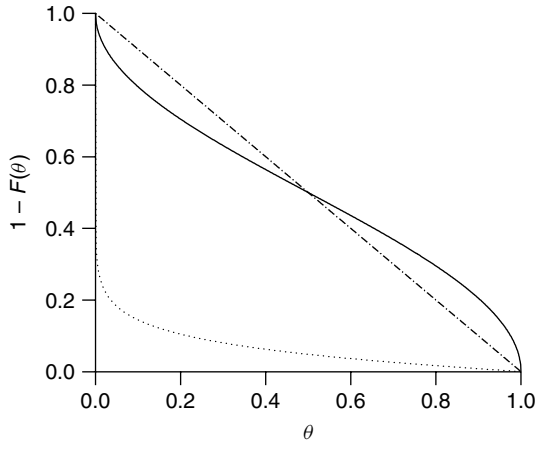


Figure 1 The prior probability $1 - F(\theta)$ that the proportion of units in a consignment is greater than θ, for various choices of α and β: $\alpha = \beta = 1(- \cdot -\cdot)$, $\alpha = \beta = 0.5(-)$, $\alpha = 0.065, \beta = 0.935(\cdots)$ [Reproduced with permission from JFS.]

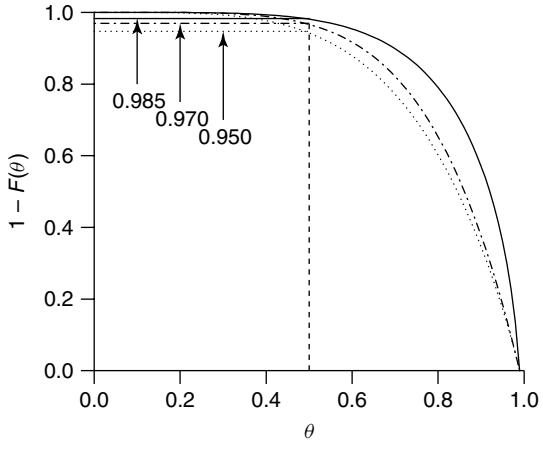


Figure 2 The posterior probability $1 - F(\theta)$ that the proportion of units in a consignment is greater than θ, for various choices of α and β: $\alpha = \beta = 1(- \cdot -\cdot), \alpha = \beta = 0.5$ $(-), \alpha = 0.065, \beta = 0.935(\cdots)$, after observation of four units all found to be illegal. The corresponding probabilities that at least 50% of the consignment contains illegal units is marked as $0.985(\alpha = \beta = 0.5), 0.970(\alpha = \beta = 1), 0.950$ $(\alpha = 0.065, \beta = 0.935)$ [Reproduced with permission from JFS.]

Examples are given in Bates and Lambert [13] and Faber *et al.* [12]. Let $R = Z + Y$ be the total number of units in the consignment that contains illicit drugs,

where Z is the number of units in the sample of size m and Y is the number of units in the remainder $(N - m)$ that contain drugs. Then, the distribution of Z is hypergeometric with

$$\Pr(Z = z) = \frac{\binom{R}{z}\binom{N-R}{m-z}}{\binom{N}{m}}, \quad z = 0, 1, \ldots, \min(R, m) \tag{12}$$

The lower limit for θ is R/N, where R is the maximum number of illicit units in the population, which satisfies the following inequality

$$\sum_{i=0}^{r} \frac{\binom{R}{m-i}\binom{N-R}{i}}{\binom{N}{m}} \leq \alpha \tag{13}$$

where r is the number of nondrug items in the sample and the confidence level is $100(1 - \alpha)\%$ [7]. When $m = z$, the inequality is

$$\frac{R!(N-m)!}{N!(R-m)!} \leq \alpha \tag{14}$$

A Bayesian Approach. A Bayesian approach for small consignments using the hypergeometric distribution is described in Coulson *et al.* [14]. A discrete prior distribution is chosen for the $(N + 1)$ possible divisions of the consignment into licit and illicit items. The likelihood function is based on the hypergeometric distribution sampling m from N and a posterior distribution obtained.

An alternative Bayesian approach for small consignments uses a so-called β-binomial distribution. This distribution provides a probability statement (as distinct from a confidence statement) about the number of units in the consignment that contains drugs.

As before, let θ $(0 < \theta < 1)$ be the proportion of units in the superpopulation that contains drugs. The probability distribution of z, given m and θ, may be taken to be binomial. For each unit, independently of the others in the consignment, the probability that it contains drugs is taken to be equal to θ. The posterior distribution of θ is another β distribution with parameters $(\alpha + z)$ and $(\beta + m - z)$.

Since the consignment size is small, a better representation of the variability of the number of units Y, which contain drugs in the uninspected section of the consignment, is obtained by considering a probability distribution for Y explicitly. There are n units in the remainder of the consignment $(m + n = N)$, which have not been inspected. Then Y (unknown and $\leq n$) is the number of units in this remainder that contain drugs. Given θ, the distribution of $(Y|n, \theta)$, similar to that of $(Z|m, \theta)$, is binomial. However, θ has a β distribution and the distribution of $(Y|n, \theta)$ and the distribution of $(\theta|m, z, \alpha, \beta)$ can be combined to give a Bayesian predictive distribution for $(Y|m, n, z, \alpha, \beta)$, a β-binomial distribution.

$$Pr(Y = y|m, n, z, \alpha, \beta) = \frac{\Gamma(m + \alpha + \beta)\binom{n}{y}\Gamma(y + z + \alpha) \times \Gamma(m + n - z - y + \beta)}{\Gamma(z + \alpha)\Gamma(m - z + \beta)\Gamma(m + n + \alpha + \beta)}, \quad (y = 0, 1, \ldots, n) \tag{15}$$

From this distribution, inferences can be made about Y, such as probability intervals or lower bounds for Y.

Quantity

The standard of proof is relevant to the estimation of quantity. In state courts in the United States in 2001, the value of Q in a drug trial was defined to be an essential element of the possession charge. As such, the value of Q had to be proved by a jury beyond reasonable doubt. However, in Federal courts, Q was considered an issue of fact and had only to be established on the preponderance of the evidence. This should be borne in mind when interpreting the estimation processes. For further information about procedures in the United States, see the Sentencing Guidelines [15]. The importance of quantity estimation was illustrated in the drug smuggling case of *U.S. v. Shonubi*. At issue was the estimation of an unknown total amount of heroin that was imported illegally in the digestive tract of the defendant during several trips from Nigeria to the United States. The only quantitative piece of information available was the amount of heroin found in the defendant's possession on his last trip. This amount was itself the result of a sampling process. Shonubi was found to be carrying internally 103 multicolored balloons, each filled with a white powdery substance. Four of the balloons were selected randomly for testing and inference. The total net weight of the heroin in all 103 balloons was estimated by multiplying the average net weight of the heroin in the four balloons by 103 [16].

The estimation of the quantity of a drug is treated in two stages. First the proportion of the units in the consignment that contain illicit drugs is modeled. Second, the total weight of the illicit material in those packets that contain anything illicit is estimated. Uncertainty in the prior belief in the proportion of packets that are "illicit" may be represented by a β distribution. It is assumed that there is no prior information for the mean and variance of the distribution of the quantity of drugs in the packages. Details of how such prior information may be considered are given in Aitken *et al.* [17].

Given the sample size, and thus an estimate of the proportion of a consignment that contains drugs and an estimate of the mean and standard deviation of the weight in the consignment, a *confidence interval* for the true quantity of drugs may be calculated [7]. A *probability interval* is appropriate in a Bayesian context. In this context, a probability distribution is associated with a parameter (Q, say) denoting the total quantity of illicit material in the consignment and probability statements of any desired kind may be made. For example, these could include the probability that Q is greater than a certain value, q say, which will be of importance in sentencing hearings; here q may be a borderline between two base offense levels.

Frequentist Approach

It is only possible to make a statement about the consignment as a whole with certainty if the whole consignment is analyzed. Once it is accepted that a sample has to be considered, then it is necessary to consider what level of proof is adequate. This is strictly a matter for the court to decide.

The method described by Tzidony and Ravreboy [7] considers the consignment as a population and the packages (or units) examined as a sample. The quantities (weights) of drugs in the units are assumed to be random variables that are normally distributed, with population mean μ and population variance σ^2, say. The mean quantity in a unit in the consignment is estimated by the mean, denoted by $\bar{x}$, of the quantities found in the sample. A confidence interval is determined for μ based on the sample size m, the sample mean $\bar{x}$, the sample standard deviation s of the quantities of drugs in the units examined, and an associated t-distribution. The general expression for the relevant confidence interval is

$$\bar{x} - t_{(m-1)}(\alpha/2)\frac{s}{\sqrt{m}}\sqrt{\frac{(N-m)}{N}} \leq \mu \leq \bar{x} + t_{(m-1)}(\alpha/2)\frac{s}{\sqrt{m}}\sqrt{\frac{(N-m)}{N}} \quad (16)$$

where $\sqrt{(N-m)/N}$ is the finite population correction factor. The term $t_{(m-1)}(\alpha/2)$ denotes the $100(1-\frac{\alpha}{2})\%$ point of the t-distribution with $(m-1)$ degrees of freedom, e.g., for the 95% confidence interval, $\alpha = 0.05$, and $t_{(m-1)}(\alpha/2)$ is the 97.5% point of $t_{(m-1)}$. The interval is the $100(1-\alpha)\%$ confidence interval for the mean quantity in a package [7].

The corresponding confidence interval for Q, the total quantity of drugs in the consignment, is obtained by multiplying all entries in the inequalities by $N\hat{\theta}$, where $\hat{\theta}$ is an estimate for θ based on the sample of size m. This gives as a $100(1-\alpha)\%$ confidence interval for Q [7].

$$N\hat{\theta}\left\{\bar{x} - t_{(m-1)}(\alpha/2)\frac{s}{\sqrt{m}}\sqrt{\frac{(N-m)}{N}}\right\} \leq Q \leq N\hat{\theta}\left\{\bar{x} + t_{(m-1)}(\alpha/2)\frac{s}{\sqrt{m}}\sqrt{\frac{(N-m)}{N}}\right\} \quad (17)$$

A corresponding $100(1-\alpha)\%$ lower bound for Q, which may be deemed more appropriate in a legal context, is given by the left-hand side of the inequality

$$N\hat{\theta}\left\{\bar{x} - t_{(m-1)}(\alpha)\frac{s}{\sqrt{m}}\sqrt{\frac{(N-m)}{N}}\right\} \leq Q \quad (18)$$

Bayesian Approach

Consider the consignment as itself a random sample from a larger superpopulation of units or packages, some or all of which contain illegal material. Then, θ $(0 < \theta < 1)$ is the proportion of units in the superpopulation, which contains drugs. The variability in θ may be modeled by a β distribution.

Let n be the number of packages in the consignment that is not examined. Then, N equals $m + n$. Let $(x_1, \ldots, x_z)$ and $(w_1, \ldots, w_y)$ be the weights of the contents of the units examined and not examined respectively, which contain drugs. It is assumed that these weights are normally distributed. The total

Table 4 Estimates of quantities q g. of drugs, in a consignment of $m + n$ units, according to various possible burdens of proof, expressed as percentages $P = 100 \times Pr(Q > q|m, z, n, \bar{x}, s)$ in 26 packages when six packages are examined (m = 6, n = 20) and z = 6, 5, or 4 are found to contain drugs. The mean ($\bar{x}$) and standard deviation (s) of the quantities found in the packages examined, which contain drugs, are 0.0425 g and 0.0073 g. The parameters for the beta prior are $\alpha = \beta = 1$. Numbers in brackets are the corresponding frequentist lower bounds using the finite population correction factor

Percentage P	Number of units examined which contain drugs			Possible burden of proof (illustrative)
	6	5	4	
97.5	0.689 (0.930)	0.501 (0.744)	0.345 (0.575)	
95	0.750 (0.968)	0.559 (0.785)	0.397 (0.613)	Beyond reasonable doubt
70	0.944 (1.067)	0.770 (0.885)	0.603 (0.704)	Clear and convincing
50	1.015 (1.105)	0.862 (0.921)	0.704 (0.737)	Balance of probabilities

Permission needed from JFS

weight, q, of the contents of the units in the consignment is then given by

$$q = z\bar{x} + y\bar{w} \tag{19}$$

Also, let $s^2 = \sum_{i=1}^{z}(x_i - \bar{x})^2/(z-1)$ be the variance of the measurements on the units that were examined and found to contain drugs. In the absence of prior information about the mean or variance of the distribution of the weights of drugs in the packages, a uniform prior distribution is used.

The distribution and corresponding probability density function of Q may be determined from the relationship $Q = z\bar{x} + y\bar{W}$. Let $f_{t,z-1}(.)$ denote the probability density function of the t-distribution with $(z-1)$ degrees of freedom. The probability density function $f(q)$ of Q is then given by [18]

$$f(q) = \sum_{y=0}^{n} f_{t,z-1}\left\{\frac{q-(z+y)\bar{x}}{sy\sqrt{\frac{1}{z}+\frac{1}{y}}}\right\} \left\{sy\sqrt{\frac{1}{z}+\frac{1}{y}}\right\}^{-1} Pr(Y=y) \tag{20}$$

An example in which a seized drug exhibit contained 26 street doses is given in Tzidony and Ravreboy [7]. A sample of six ($m = 6$) units was taken and each was analyzed and weighed. Twenty ($n = 20$) units were not examined. All six of the units examined contained drugs. The average net weight $\bar{x}$ of the powder in the six units was 0.0425 g. with a standard deviation s of 0.0073 g. A 95% confidence interval for the total quantity Q in the 26 doses is $1.105 \pm 0.175\,g$. Note that this interval incorporates the finite population correction factor from (equation 2) to allow for the relatively large sample size ($m = 6$) compared to the consignment size ($N = 26$).

The Bayesian approach described here does not require a finite population correction.

The values for Q obtained from the Bayesian argument and corresponding to appropriate percentage points of the distribution may be determined from (equation 20). Some results are given in Table 4, together with corresponding results with the method of Tzidony and Ravreboy, and in Figure 3.

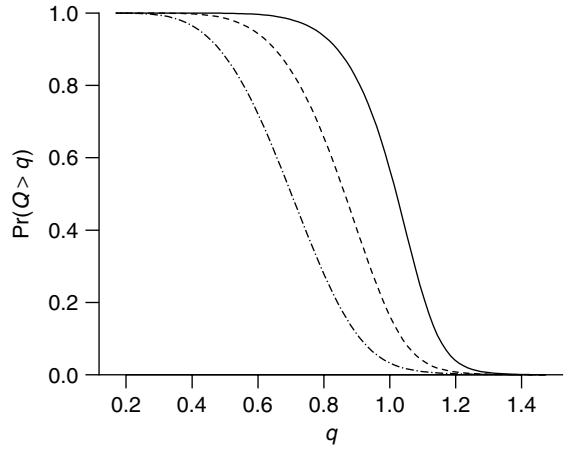


Figure 3 The probability that the total quantity Q of drugs (in grams) in a consignment of 26 packages is greater than q, using a Bayesian argument, when six packages are examined and 6(−), 5(− − −), or 4(− · −·) are found to contain drugs. The mean and standard deviation of the quantities found in the packages examined, which contain drugs, are 0.0425 g and 0.0073 g. The parameters for the β prior are $\alpha = \beta = 1$ [Reproduced with permission from JFS.]

Note: A general introduction to sampling techniques is given in Cochran [10]. Reviews of the statistical and legal aspects of the forensic study of illicit drugs are given by Izenman [2, 19]. These includes a discussion of various sampling procedures, various methods of choosing the sample size, a strategy for assessing homogeneity and the relationship between quantity and the possible standards of proof. Further comments on sampling issues are given in various chapters of Gastwirth [20], such as Aitken [21], Gastwirth *et al.* [22], and Izenman [16, 23, 24], with respect to the *Shonubi* case.

References

[1] Wigmore, J.H. (1979). *Evidence in Trials at Common Law*, Vol. 2, Little, Brown, Boston.

[2] Izenman, A.J. (2003). Sentencing illicit drug traffickers: how do the courts handle random sampling issues, *International Statistical Review* **71**, 535–556.

[3] Curran, J.M., Triggs, C.M. & Buckleton, J. (1998). Sampling in forensic comparison problems, *Science and Justice* **38**, 101–107.

[4] American Society for Testing and Materials (ASTM) (2001). *E 122-00 Standard Practice for Calculating Sample Size to Estimate with a Specified Tolerable Error, the Average for a Characteristic of a Lot or Process*.

[5] Evett, I.W. & Weir, B.S. (1998). *Interpreting DNA Evidence*, Sinauer Associates Inc, Sunderland, MA.

[6] Frank, R.S., Hinkley, S.W. & Hoffman, C.G. (1991). Representative sampling of drug seizures in multiple containers, *Journal of Forensic Sciences* **36**, 350–357.

[7] Tzidony, D. & Ravreboy, M. (1992). A statistical approach to drug sampling: a case study, *Journal of Forensic Sciences* **37**, 1541–1549.

[8] United Nations Guidelines: Methaqualone (ST/NAR/15; December, 1998); Lysergide (ST/NAR/17; January, 1989); Cocaine (ST/NAR/7; February, 1986); Benzodiazepine (ST/NAR/16; December, 1988); Cannabis (ST/NAR/8; February, 1987); Psilocybin (ST/NAR/19; December, 1989); Mescaline (ST/NAR/19; December, 1989); Opium, Morphine, Heroin (St/NAR/29/Rev.1; June, 1998).

[9] European Network of Forensic Science Institutes Drugs Working Group (2004). *Guidelines on Representative Drug Sampling*.

[10] Cochran, W.G. (1977). *Sampling Techniques*, 3rd edition., John Wiley & Sons, Chichester.

[11] Rahne, E., Joseph, L. & Gyorkos, T.W. (2000). Bayesian sample size determination for estimating binomial parameters from data subject to misclassification, *Applied Statistics* **49**, 119–128.

[12] Faber, N.M., Sjerps, M., Leijenhorst, H.A.L. & Maljaars, S.E. (1999). Determining the optimal sample size in forensic casework - with application to fibres, *Science and Justice* **39**, 113–122.

[13] Bates, J.W. & Lambert, J.A. (1991). Use of the hypergeometric distribution for sampling in forensic glass comparison, *Journal of the Forensic Science Society* **31**, 449–455.

[14] Coulson, S.A., Coxon, A. & Buckleton, J.S. (2001). How many samples from a drug seizure need to be analyzed?, *Journal of Forensic Sciences* **46**, 1456–1461.

[15] United States Sentencing Commission Guidelines Manual (2007). http://www.ussc.gov/2007guid/GL2007.pdf.

[16] Izenman, A.J. (2000c). Assessing the statistical evidence in the *Shonubi* case, in *Statistical Science in the Courtroom*, J.L. Gastwirth, eds, Springer-Verlag, New York, pp. 415–433.

[17] Aitken, C.G.G., Bring, J., Leonard, T. & Papasouliotis, O. (1997). Estimation of quantities of drugs handled and the burden of proof, *Journal of the Royal Statistical Society, Series A* **160**, 333–350.

[18] Aitken, C.G.G. & Lucy, D. (2002). Estimation of the quantity of a drug in a consignment from measurements on a sample, *Journal of Forensic Sciences* **47**, 968–975.

[19] Izenman, A.J. (2001). Statistical and legal aspects of the forensic study of illicit drugs, *Statistical Science* **16**, 35–57.

[20] Gastwirth, J.L. (ed) (2000). *Statistical Science in the Courtroom*, Springer-Verlag, New York.

[21] Aitken, C.G.G. (2000). Interpretation of evidence and sample size determination, in *Statistical Science in the Courtroom*, J.L. Gastwirth, eds, Springer-Verlag, New York, p. 1–24.

[22] Gastwirth., J.L., Freidlin, B., Miao, W. (2000). The *Shonubi* case as an example of the legal system's failure to appreciate statistical evidence, in *Statistical Science in the Courtroom*, J.L. Gastwirth, ed, Springer-Verlag, New York, p. 405–413.

[23] Izenman, A.J. (2000a). Statistical issues in the application of the Federal sentencing guidelines in drug, pornography and fraud cases, in *Statistical Science in the Courtroom*, J.L. Gastwirth, ed, Springer-Verlag, New York, p. 25–50.

[24] Izenman, A.J. (2000b). Introduction to two views on the *Shonubi* case, in *Statistical Science in the Courtroom*, J.L. Gastwirth, ed, Springer-Verlag, New York, p. 393–403.

Further Reading

American Society for Testing and Materials (ASTM) (2004). *E 105-04 Standard Practice for Probability Sampling of Materials*.

Colón, M., Rodriguez, G. & Diaz, R.O. (1993). Representative sampling of 'street' drug exhibits, *Journal of Forensic Sciences* **38**, 641–648.

United States Department of Justice Drug Enforcement Administration (2006). *The Scientific Working Group for the Analysis of Drugs (SWGDRUG)*. Recommendations.

U.S. v. Shonubi: Shonubi V: 962 F.Supp.370 (E.D.N.Y. 1997); *Shonubi IV*: 103 F.3d 1085 (2d Cir. 1997); *Shonubi III*: 895 F.Supp. 460 (E.D.N.Y. 1995); *Shonubi II*: 998 F.2d 84 (2d Cir. 1993); *Shonubi I*: 802 F.Supp. 859 (E.D.N.Y. 1992).

COLIN G. G. AITKEN

Sampling Trace Evidence

Introduction

Sampling can be defined as the removal of a part of a substance to assess the materials present in the whole. In a forensic context, sampling has a broader definition as it occurs throughout the forensic process from selection of items at the crime scene to subsampling in the laboratory.

Trace evidence by its nature is small in size such that examination commonly requires the use of a microscope [1, 2]. Sampling of trace evidence encompasses the selection, recovery, and removal of material and can include sampling of a single item, similar items, or different items. The reasons for sampling include reducing the number of analytical determinations necessary, time constraints, presentation for instrument analysis, and cost.

The first stage in sampling is defining the sampling strategy. The sampling strategy chosen is dependent on the question being asked, the purpose of sampling, the use of results. Therefore, a sampling strategy is necessary, which must satisfy all aspects above including forensic evidential requirements. The second stage is defining the target population for what is being sampled. A sampling plan is established for reasons of effectiveness and efficiency and to ensure the sampling strategy is implemented. The sampling plan forms the third stage.

Finally, the sampling methods used ensure the sampling plan is implemented [3].

This section reviews sampling of trace evidence, in particular, paint (*see also* **Paint**), glass (*see also* **Glass**), fibers (*see also* **Paint**), hairs (*see also* **Hair: Animal**), soil, and particulates (*see also* **Particles: Form**). Sampling of DNA (*see* **DNA**), explosives (*see* **Explosions: Scene Investigation**; **Explosion Debris: Laboratory Analysis of**), drugs, and firearms is beyond the scope of this section. The following aspects of trace evidence sampling are covered:

1. principles of sampling trace evidence;
2. population determination;
3. sampling plans used for trace evidence;
4. sampling questions; and
5. methods of trace evidence sample recovery.

Principles of Sampling Trace Evidence

Sampling Trace Evidence at the Crime Scene

The Locard exchange principle is quoted as "every contact leaves a trace" [4]. Usually, things are not so simple. When sampling for trace evidence, determining the type(s) of trace evidence one is looking for is crucial in the decision making process. Sampling is based on the case circumstances, the evidence type being looked for, and the items available for analysis. The investigator may be directed by the available case information, although there are occasions when information is scarce, thereby increasing the need for care in the sampling procedure so as not to compromise one evidence type while sampling for another. Sampling processes employed determine what is ultimately examined and the conclusions that can be drawn from the testing results.

When sampling for trace evidence, it is necessary to take samples from the area where the transferred trace evidence may have been deposited and take a representative sample from the original source of the trace evidence. Where a house is burgled and the suspect gains entry through a broken window, crime scene investigators seize the articles of clothing worn at the time of the alleged offense by the suspect and take an adequate representative control sample of glass from the window.

When a crime scene is examined, it is rarely practical nor is it necessary to examine all of the items available. The initial selection of items to be examined is determined by what is known to have happened (from witness, victim, and suspect statements) and where one is likely to find the evidence type. For example, if one wishes to see if there is evidence that two people were in contact, then selecting the clothing that they were wearing at the time of the alleged contact and looking for evidence of fiber transfer are required.

Sampling and Subsampling in the Laboratory

In the laboratory, the sampling plans and recovery methods chosen depend on the trace evidence type, the case circumstances, and the question being asked. For example, sampling done to determine the presence of paint, glass, particulates, etc., or to determine both the presence and quantity of the substance(s).

Some reasons for sampling in the laboratory are as follows:

- Taking samples to carry out different tests on the population of trace evidence.
- Taking samples from the population of trace evidence to carry out the same test repeatedly to determine reproducibility of results.
- Taking a sample from the population to carry out destructive testing (sample is destroyed and no further testing can be carried out on the sample selected).
- Sampling for practical considerations in relation to presentation for instrument analysis.

Once the decisions have been taken regarding the trace evidence type to sample then one has to decide how that will be done, i.e., what sampling plan will one follow and what recovery method will be used?

In essence, sampling at the crime scene and in the laboratory involves risk assessment and risk management. When sampling for trace evidence the operator is always working in a context. No matter how good the sampling process is, it introduces a level of uncertainty. Risk management of sampling involves assessing the level of uncertainty introduced by sampling and adopting sampling plan(s) and methods of recovery such that one has confidence in the sampling decisions taken and the interpretation of the results.

Population Determination

One has to decide what to sample before you decide which sampling plan to use. This is the need to define the relevant population. There are two possibilities: (a) a homogenous population where all items look the same e.g., 100 pieces of green glass that look visually similar or (b) a heterogeneous population, for example, 100 pieces of green glass and 100 pieces of clear glass. When one has a heterogeneous population, one would sample from both green and clear glass populations if the case involved a broken green and clear stained glass window, but would sample only the clear glass population if the case involved a broken clear glass window.

Population determination is based on the case circumstances and the trace material sources available [5].

Sampling Plans Used for Trace Evidence

In trace evidence analysis, the most common sampling plans used, divide into numerically based and nonnumerically based.

Numerically Based Plans

Numerical plans can be statistically based or nonstatistically based.

Statistically Based Sampling Plans. Statistically based sampling is used when one is defining a population based on the analysis of the sample taken from that population, for example, defining a population of glass fragments as being from a putative source based on the refractive index of the sample of selected glass fragments. There are two different approaches to statistical sampling, the Frequentist approach and the Bayesian (*see also* **Bayesian Networks**) approach [6]. In general, the Frequentist approach is that a fixed but unknown proportion of the population is positive and the proportion of positives in the sample can be used to estimate the proportion of positives in the population. The proportion of positives varies over different samples. This method gives a confidence level that a defined minimum in the population is positive. The most commonly used Frequentist method in trace evidence is the hypergeometric method. The following publications give a good overview of the subject [7–12].

The Bayesian approach assumes that the proportion of positives in the sample is fixed. There is no consideration for repetition of sample testing [11, 13, 14]. The Bayesian approach allows prior information about the population to be taken into account whereas the Frequentist approach does not. Prior information, for example, could be that all the suspect fibers look visually the same and were found on the victim's jumper. The disadvantage of the Bayesian method is that the values attached to the prior knowledge are subjective. However, using the Bayesian approach is more intuitive and flexible as the results obtained

from the evidence can be expressed as probability statements and prior information is taken into account [11, 13].

Nonstatistically Based Sampling Plans. Nonstatistical sampling plans have been documented in relation to drugs analysis [6]. A number of these are also used for trace evidence [12]. Examples of nonstatistical sampling plans used in relation to trace evidence include testing all the samples, testing one, and testing 20.

Nonstatistical sampling is simple to apply; however, only limited inferences can be made about the overall properties of the population [6].

Nonnumerically Based Sampling Plans

While numerically based sampling plans are well documented, the same cannot be said for nonnumerically based plans. Analysis of trace evidence involves many nonnumerically based sampling plans. Frequently used nonnumerical sampling plans are probability sampling, judgment sampling, and bulk sampling.

Probability (Random) Sampling. Probability sampling occurs when every unit in the population has a known probability of being included in the sample and is often referred to as random sampling [15]. For example, if 100 pieces of glass were removed from a suspect's jacket, randomly picking 10 of them for analysis would constitute random sampling. Depending on the trace evidence type, this may not be practical. In the case of fiber examination, if one had numerous blue fibers on an adhesive lift, one would sample the longest fibers, as long fibers allow destructive testing to be carried out while still allowing retention of some of the original fiber.

Judgment Sampling. Judgment sampling involves selecting items based on traits or aspects they have in common with other items in the case. If a pale blue car knocks down and kills a person, one would select pale blue paint found on the victim's clothes for analysis. The selection is made on the basis of the similarities the sample has with other item(s) in the case. The presence of a dark blue paint on the victim's clothes may introduce other questions, such as, could a different car have been involved?

Bulk Sampling. This type of sampling happens when a sample is taken from a large amount of material that does not consist of separate units and is often a feature of subsampling in the laboratory. Bulk sampling occurs when one takes a sample of soil from a field. Considerations include as to where the sample is taken from, how much sample is taken, and whether the sample is representative of the lot.

Trace evidence analysis frequently involves the use of many sampling plans used sequentially or at the same time. When a burglar enters a house through the window and the crime scene examiner comes to examine the scene, he uses judgment sampling to target the areas where the intruder may have left trace evidence. When a suspect is caught, judgment sampling is used to select the items of clothing to be taken from the suspect based on what the suspect was seen to have worn at the time. Judgment sampling is used in the laboratory to target paint flakes on the suspects jumper of the same color as that of the paint on the window, and bulk sampling is used to select fibers from the suspect's jumper to compare with the fibers found on the window.

Problems Associated with Sample Plan Selection. While the hypergeometric sampling is used commonly when sampling glass, there is disparity when it comes to deciding what an appropriate sample size is for the other trace evidence types. In the area of fibers, for example, Faber *et al.* [12] determined that there was a disparity in sample size analyzed between different laboratories. They observed that an objective criterion is needed to determine optimal fiber sample size in casework and suggested the use of the hypergeometric method [12]. Various international forensic organizations have been tackling this problem, e.g., European Network of Forensic Institutes (ENFSI) and European Fibres Group. If an internationally accepted norm were established for different aspects of trace evidence it would greatly enhance sampling norms across different countries. The conclusions drawn from trace evidence are largely influenced by the sampling methods and plans used. It is important that forensic scientists are aware of the methods and plans they are using, and the veracity of the conclusions they reach based on that evidence [16].

Sampling Questions

To identify the factors that influence sampling plans and methods, asking questions is helpful. The following are some questions that are relevant when making sampling decisions:

- What background information is available?
- What question(s) need to be answered?
- What will the testing results be used for?
- What type of trace evidence is of interest?
- Are there any legal requirements?
- Is identification of material all that is required?
- What accuracy and precision is required?
- What is to be sampled and how?
- Is it necessary to sequence the sampling that is to take place?
- Should statistical sampling be employed?
- Is the sampling plan chosen adequate to answer the question asked and does it fulfill the purpose?
- Is the plan chosen cost effective and efficient?
- Is there literature to back up the sampling plan chosen?
- Will it be possible to defend the conclusions drawn from the sampling plan employed?

Finding the answers to some or all of these questions should guide one to appropriate sampling plan(s) and sampling method(s).

Methods of Trace Evidence Sample Recovery

While the sampling strategy and sampling plans used influence the testing results so also do the methods of trace evidence sample recovery. Commonly used methods of recovery of trace evidence are removal of the whole item, 'hand' picking, taping, brushing, vacuuming, and scraping [4, 17].

Removal of the Whole Item

It may not always be possible or appropriate to remove trace evidence at the crime scene. In this situation the whole item is brought to the forensic science laboratory for trace evidence removal.

'Hand' picking

'Hand' picking of trace evidence is the method of choice when the quantity of evidence, size, and appearance allow. Items are usually collected by hand (gloved) or by using a sharp pointed tweezers using strong light and magnification. This method has the advantage of establishing the position of the material on the item.

Taping

The use of adhesive tape is a reliable method of removing very small particles of trace evidence [18, 19]. A short length of adhesive tape is passed repeatedly over the surface of the item ensuring the entire surface under examination is sampled. This is then applied to a sheet of rigid plastic and placed in a clean plastic bag appropriately labeled. Taping is used routinely to remove fibers, hairs, and particulates. This method protects trace evidence from contamination and allows the trace evidence of relevance to be separated from nonrelevant material. The tape-lift is examined under a microscope and the trace evidence removed by circling the evidence on the back of the tape and lifting the tape to remove it, or by circling the evidence and cutting and lifting the section of tape to remove it. The latter method prevents contamination of the lift from elements in the atmosphere.

Brushing

Brushing is the method of choice for the removal of paint, glass, and soil. A clean paintbrush or toothbrush is used to brush down the item in question and the resulting debris is collected for examination. It allows retrieval of trace evidence from areas that are difficult to access. Using this method one does not encounter the difficulty of retrieving small fragments of paint, glass, or soil intact from an adhesive lift.

Vacuuming

Vacuuming is an efficient way to remove trace evidence. However, its efficiency is a disadvantage in that recently deposited material is mixed with that deposited long ago and the volume of material retrieved lengthens the process of trace evidence retrieval. Vacuuming is performed in a systematic manner and filters changed frequently so that the debris retrieved can be associated with a specific area e.g. sleeves of a jacket. Vacuuming is not recommended for routine use because of the difficulty in interpretation, the rigorous requirements to avoid contamination, and the length of time taken to analyze the contents retrieved.

Scraping

Scraping has limited use as a method of sample removal of trace evidence. However, it can be used

to collect a control sample of a source of paint. This involves using a clean blade to remove a representative paint sample from the surface (e.g. a painted door) that is collected on clean paper, in a paper envelope, or a tweezer is used to mount the sample on a glass slide.

Selection of a suitable method of trace evidence removal is vital to minimize loss of potential evidence, avoid contamination, and to facilitate trace evidence analysis [20].

Conclusion

Trace evidence analysis is a valuable tool in crime scene investigation. When sampling for trace evidence, it is important to realize that the methods of sampling and the sampling plans used ultimately affect the evidence obtained and the interpretation of that evidence. It is vital that sampling is carried out correctly and is understood to eliminate loss of potential evidence and to avoid overestimation or underestimation of the significance of the evidence found. Currently sampling of trace evidence is an area that has received little in-depth analysis leading to a lack of uniformity in approach.

References

[1] Petraco, N. (1985). The occurrence of trace evidence in one examiner's casework, *Journal of Forensic Sciences* **30**(2), 485–493.

[2] Petraco, N. (1986). Trace evidence – the invisible witness, *Journal of Forensic Sciences* **31**(1), 321–328.

[3] Crosby, N.T. & Patel, I. (1995). *General Principals of Good Sampling Practice*, The Royal Society of Chemistry.

[4] Horswell, J. (2004). Collection techniques: present status, in *The Practice of Crime Scene Investigation*, J. Horswell, ed, CRC Press LLC.

[5] Robertson, J. (1999). Protocols for fibre examination and initial preparation, in *Forensic Examination of Fibres*, 2nd Edition, J. Robertson & M. Grieve, eds, Taylor & Francis, London.

[6] European Network of Forensic Science Institutes Drugs Working Group (2003). *Guidelines on Representative Drug Sampling*, European Network of Forensic Science Institutes Drugs Working Group.

[7] Tzidony, D. & Ravreby, M. (1992). A statistical approach to drug sampling: a case study, *Journal of Forensic Science* **37**, 1541–1549.

[8] Curran, J.M., Triggs, C.M. & Buckleton, J. (1998). Sampling in forensic comparison problems, *Science and Justice* **38**(2), 101–107.

[9] Sandercock, P.L.M. (2000). Sample size considerations for control glass in casework, *Canadian journal of Forensic Science* **33**(4), 173–185.

[10] Bates, J.W. & Lambert, J.A. (1991). Use of the hypergeometric distribution for sampling in forensic glass comparison, *Journal of Forensic Science Society* **31**(4), 449–455.

[11] Aiken, C.G.G. (1999). Sampling – how big a sample? *Journal of Forensic Sciences* **44**(4), 750–760.

[12] Faber, N.M., Sjerps, M., Leijenhorsst, H.A.L. & Maljaars, S.E. (1999). Determining the optimal sample size in forensic casework – with application to fibres, *Science and Justice* **39**(2), 113–122.

[13] Causin, V., Schiavone, S., Marigo, A. & Carrisi, P. (2004). Bayesian framework for the evaluation of fibre evidence in a double murder – a case report, *Forensic Science International* **141**, 159–170.

[14] Champod, C. & Taroni, F. (1997). Bayesian framework for the evaluation of fibre transfer evidence, *Science and Justice* **37**, 75–83.

[15] Houck, M.M. (2005). Forensic fibre examination and analysis, *Forensic Science Review* **17**(1), 17–29.

[16] Ramsey, M.H. & Ellison, S.L.R. (ed) (2007). *Measurement Uncertainty Arising from Sampling, A Guide to Methods and Approaches*, Eurachem /CITAC Guide.

[17] Saferstein, R. (ed) (1988). Collection and examination of microtraces: present status, in *Forensic Science Handbook*, Prentice-Hall.

[18] Biermann, T., European Fibres Group (1998). The advantages and disadvantages of 1:1 taping, *Proceedings of the Sixth European Fibres Group Meeting*. 10–12 June 1998, Dundee, 44–47.

[19] Pounds, C.A. (1975). The recovery of fibres from the surface of clothing for forensic examinations, *Journal of Forensic Science Society* **15**, 127–132.

[20] SWGMAT SWGoMA (1999). Trace evidence recovery guidelines, *Forensic Science Communications* **1**(3), section 5, http://www.fbi.gov.

MARY GIBLIN

Scalds *see* Injury: Burns, Scalds, and Chemical

Scanning Electron Microscopy *see* Microscopy: Scanning Electron Microscopy

Scene: Bomb, Management ***see*** Bomb Scene Management

Scene: Crime, Documentation ***see*** Crime Scene Documentation

Scene: Crime, Investigation ***see*** Crime Scene Investigation

Scene: Crime, Management ***see*** Crime Scene Management

Scene: Explosion ***see*** Explosions: Scene Investigation

Scene: Explosion, Investigation ***see*** Explosions: Scene Investigation

Scene: Fire, Investigation ***see*** Fire: Scene Investigation

Scene: Investigation ***see*** Firearms: Scene Investigation

Scene: Shooting ***see*** Firearms: Scene Investigation

School Threat Assessment ***see*** Threat Assessment: School

School Violence ***see*** Threat Assessment: School

Scientific Evidence: Behavioral ***see*** Behavioral Science Evidence

Scientific Method Compared to Legal Method

The acquisition and interpretation of information (data) used by scientists compares, superficially, to the legal method used in determining controversies in a legal forum. Both methods attempt to use a rational approach; both also use an inductive method of solving problems. The differences, therefore, do

not lie in the methodologies chosen to obtain a result, but are rather heavily influenced by the disparate purposes sought to be served [1].

Science seeks to analyze data or information obtained during prior experiments or inquiries in an essentially neutral fashion to obtain the best explanation as to why certain results were manifested, in order that the course of future events may be reliably predicted. Rarely does the scientific method focus on an individual, particular incident. The case to which it is being applied is important only insofar as it permits the scientist to formulate a general hypothesis on how to deal with similar events that are yet to occur.

The purpose of the legal method in resolving disputes or controversies is to determine what ought to be, rather than what is. After having considered certain information, the legal system is interested in knowing how findings and conclusions that arise out of the information impact on a specific case that is currently pending before the court. Thus, lawyers are less interested in what the impact of the evidence will be on other cases, but instead are concerned with applying the results to the normative values upon which behavior in our society is conditioned. While it engages in this balancing of values, the legal system may also be creating new norms and values, which will affect future behavior [1].

When differing or opposing values are represented by the parties in a legal controversy, the adversary system, which is the hallmark of the Anglo–Saxon judicial process, casts each participant into an adversary posture. Unlike the neutrality of interpretations with which the scientist seeks to clothe his testimony, the trial lawyer will be interested primarily in advocating the values that favor the party by whom he was retained [2–5].

Failing to recognize the different aims and purpose of science *vis-a-vis* those at stake in a legal contest has pitted scientists and lawyers against each other in uncomfortable battles. Lawyers often hold experts in disdain because they believe that an expert's opinion is for sale and that the expert will come to almost any conclusion that is desired by a litigating party. By contrast, scientists may feel estranged from the judicial process because of the intensive cross-examination to which they may be subjected, a process wherein the experts feel that their dignity is impugned and their honesty questioned. The acerbity of cross-examination has driven some experts to refusing to testify in court cases.

An increased understanding of both sides and the different aims that are the goal of the scientific method, as opposed to those of the legal system, will enhance the effectiveness of experts in offering testimony, and of attorneys seeking to elicit the experts' opinion testimony [6].

References

[1] Moenssens, A.A., Henderson, C.E. & Portwood S.G. (2007). *Scientific Evidence in Civil and Criminal Cases*, 5th Edition, Foundation Press, Chapter 1 at § 1.03(2).

[2] On the differences between law and science generally, see also, Channels, N.L. (1985). *Social Science Methods in the Legal Process*.

[3] Loevinger, L. (1985). Science, technology, and law in modern society, *Jurimetrics* **26**, 1.

[4] Patterson, M.R. (1999). Conflicts of interest in scientific expert testimony, *William and Mary Law Review* **40**, 1313.

[5] Meyer, C.B. (1997). Science and the law: the quest for the neutral expert witness. a view from the trenches, *The Journal of Natural Resources and Environmental Law*, **12**, 35.

[6] For a further illustration of how a scientific process may at times proceed in six different and distinct stages, see Moenssens, A.A., Henderson, C.E. & Portwood S.G. (2007). *Scientific Evidence in Civil and Criminal Cases*, 5th Edition, Foundation Press, Chapter 1 at § 1.03(2), pp. 14–15.

ANDRE MOENSSENS

Scientific Principles as Evidence in Court *see* Judicial Notice of Scientific Principles and Facts

Scientific Texts as Evidence in Court *see* Learned Treatises as Evidence

Search and Seizures *see* Police Use of Force

Seizures: Behavioral

A seizure is a set of behaviors that occurs in association with excessive, aberrant neuronal activity within the brain. Seizures are generally characterized by altered consciousness, motor dysfunction, or sensory dysfunction. Epilepsy is a chronic condition in which the patient experiences recurring seizures.

Definition, Prevalence, and Cumulative Incidence

At any given time, approximately 1% of the population manifests epilepsy. There is a total lifetime incidence rate of approximately 4% by age 80. This makes it one of the most common neurological disorders.

Classification of Seizures

A classification of seizures is important to facilitate clinical diagnosis and communication, to assess prognosis, and to evaluate whether specific drugs are the most appropriate ones for therapy. Dating back to the 1981 classification proposed by the International League Against Epilepsy (ILAE) [1], seizures are classified first into partial or generalized seizures. Seizures are generally classified as to whether they are partial (seizures that begin in a focal portion of the brain), or whether they are generalized (occurring bilaterally and diffusely without a local onset). Seizures are termed partial if there is either clinical or electroencephalogram (EEG) evidence indicating onset in a focal area of one hemisphere, whereas generalized seizures appear to begin simultaneously in both hemispheres. Partial seizures are divided into three categories: (i) simple partial (not affecting consciousness), (ii) complex partial (involving alteration of consciousness), and (iii) partial seizures evolving into secondarily generalized tonic–clonic (GTC) convulsions, which could begin as either simple partial or complex partial seizures. Simple partial seizures are marked by no loss of consciousness. They are manifested by behaviors specific to the locale of the seizure origin. Motor seizures involve either an involuntary jerking or stiffening movement or rarely paralysis of a given limb or region of the body. The focal movements can remain localized or involve a progression, for example, beginning in a finger and then spreading to the entire arm and face. The latter is called a *Jacksonian seizure*. The simple partial motor seizure may also involve versive head turning, postural changes, forced speech or inability to speak, or vocalization. Simple partial somatosensory or special sensory seizures can manifest themselves as tingling, numbness, or as specific visual, auditory (such as buzzing sound), olfactory (such as peculiar smell), or gustatory sensations. Autonomic simple partial seizures are marked by autonomic symptoms, such as nausea, sweating, or goose bumps. Simple partial psychic seizures are manifested by dysphasia (such as trouble with verbal expression), deja-vu (feeling of familiarity), or affective symptomatology (fear, elation). Complex partial seizures are marked by a loss or alteration in consciousness. These can begin as simple partial seizures, and then progress to alterations of consciousness, or can begin with loss or changes of consciousness at the onset. If the complex partial seizure is heralded by a subjective experience, we call this an aura, but the experience actually represents subjective simple partial seizure activity. Clouding of consciousness and automatisms reflect bilateral seizure spread. Automatisms are repetitive, purposeless movements of the extremities (particularly hands) and face, such as lip smacking, chewing, scratching, or rubbing. Less frequently, they can also involve complex behaviors such as reaching for an object, walking, running, or disrobing. Partial seizures can also secondarily generalize into GTC activity. A "postictal" state of confusion, drowsiness, and tiredness often follows the seizure itself. This is least pronounced with simple partial seizures and most pronounced with secondarily generalized seizures.

Generalized-onset seizures, seizures that begin bilaterally, can take different forms as well. Generalized seizures are divided into six types: (ii) absence seizures, (ii) myoclonic seizures, (iii) clonic seizures, (iv) tonic seizures, (v) tonic–clonic seizures, and (vi) atonic seizures. Generalized absence seizures

are characterized by brief (several seconds) loss of awareness or responsiveness and arrest of activity. Absence seizures (also called *petit mal*) involve a sudden loss of consciousness that typically only lasts a few seconds. There may or may not be associated minor motor activity, such as automatisms, blinking, slight twitching, decreased tone, or increased tone. GTC seizures are characterized by loss of consciousness at onset, a sudden tonic muscular contraction, and sometimes a movement of air through a closed glottis, producing a "cry", as well as cyanosis. After the tonic phase, the seizure then evolves into generalized clonic or rhythmic jerking activity from a 4- to 8-s tremor, sometimes accompanied by grunting. The clonic activity is initially fast, and the frequency of the jerks decreases before the seizure stops. GTC seizures are often accompanied by tongue biting and urinary incontinence. Generalized myoclonic seizures are brief, sudden contractions that may be generalized or confined to a group of muscles, or even a single muscle. They can be single and isolated, or occur in a cluster. Consciousness is usually preserved. Some forms of myoclonus are not epileptic and have no associated EEG discharge. Myoclonic seizures usually have a concomitant discharge on the EEG. Generalized clonic seizures are characterized by repetitive clonic jerks, without an initial tonic component. Generalized tonic seizures manifest with a muscular contraction that may vary in duration, severity, and the parts of the body involved. The muscular contractions are more sustained than those of myoclonic seizures. Generalized atonic seizures are characterized by a sudden decrease in muscle tone, which is variable in severity and extent, so that there may simply be a slight head drop, or an abrupt fall to the ground.

Status epilepticus is a state of continuous seizure activity or repetitive seizures without recovery in between seizures. When the status epilepticus seizure activity is generalized convulsive, this represents a life-threatening medical emergency due to the potential for damage to the brain. Brain damage is estimated to begin at approximately 20 min after the onset of status, so it is critical to intervene early.

Classification of Epilepsies and Epileptic Syndromes

In addition to classifying individual seizures, the epilepsy condition as a whole can be classified. In 1989, the ILAE has recommended a classification that remains in wide use [2]. It is recognized, however, that advances will necessitate a revision. Most patients with epilepsy will have only partial-onset or only generalized-onset seizures. Hence, the classification has two major categories: localization-related epilepsies (partial epilepsies is an acceptable synonym) or generalized epilepsies. Each category is further subdivided into idiopathic or symptomatic epilepsies. The idiopathic epilepsies are typically characterized by the absence of acquired brain insult and are widely considered to be of genetic origin. The symptomatic epilepsies have known insults that have caused the epilepsy. In addition, some patients are suspected to have had a brain insult, but one that cannot be definitively identified. These patients are referred to as *probably symptomatic* (cryptogenic was used as a synonym but has fallen out of favor). Most partial epilepsies are symptomatic or probably symptomatic, while most generalized epilepsies are idiopathic or probably idiopathic. However, it is recognized that many patients have both genetic predisposition to seizures as well as acquired brain insults.

Some patients have additional characteristic features that help to group them into recognized epileptic syndromes. The characteristics of syndromes may include specific age at onset, a particular combination of seizure types, a predictable response to treatment, and a variety of other clinical features.

Epidemiology

Not all persons who experience seizures have epilepsy. In fact, most do not. Epilepsy is not diagnosed if the seizure is acutely provoked by an insult such as trauma or stroke, or if the seizure accompanies exposure to certain drugs or a derangement in metabolism. Even single unprovoked seizure recurs in less than half of individuals. Of those persons who do have epilepsy, approximately half of them have no known precipitating event or injury.

Many factors can provoke seizures. Some reversible alterations that can trigger seizures include metabolic changes, such as low blood sodium, calcium, or magnesium, kidney failure, or liver failure. Many toxins and drugs can trigger seizures. This includes psychoactive prescription drugs such as tricyclic antidepressants, bupropion (Wellbutrin), antipsychotic drugs such as clozapine (Clozaril),

and illicit agents such as cocaine or phencyclidine (PCP). In addition, withdrawal from alcohol, benzodiazepines, or barbiturates can also trigger seizures. In general, seizures do not recur after removing the offending agent or after the withdrawal period has passed.

Many factors may contribute to the development of epilepsy. Any disease or condition that injures the cerebral cortex can cause epilepsy. This includes brain infections and inflammations such as encephalitis, head trauma, brain tumors, brain malformations, and degenerative conditions such as Alzheimer's disease. Chronic alcohol use increases the risk of epilepsy, and this is not due to alcohol withdrawal seizures.

Head trauma is a common etiology for epilepsy. The most epileptogenic head trauma is penetrating head injury, with an epilepsy incidence as high as 53% following such injuries. Moderate-to-severe closed head injuries are associated with a significantly increased risk of epilepsy, from 3 to 25 times greater than in the absence of head trauma [3]. Among closed head injury patients, brain contusions and intracranial hemorrhages appear to be strong risk factors for the development of seizures. The presence of skull fractures or prolonged loss of consciousness or posttraumatic amnesia (greater than one day in length) are also risk factors, though not as strongly predictive. Mild head trauma, defined by loss of consciousness or posttraumatic amnesia for less than 30 min, is associated with a slightly greater risk of epilepsy (1.5-fold increase in risk) that is not statistically significant [4]. Other potential causes of seizures or seizurelike episodes should be carefully ruled out in the mild head injury patient presenting with apparent seizures, including withdrawal from alcohol or other substances, hypoglycemia or other metabolic conditions, and psychogenic seizures. The latter are particularly common following mild traumatic brain injury [5]. Even after moderate-to-severe head injury, 33% of patients experiencing seizurelike phenomena were actually experiencing psychogenic seizures [6].

A common legal question is how to prove causality between an injury and epilepsy. Before addressing causality, it is important to first prove the presence of epilepsy. The injury has to be sufficiently severe to cause epilepsy. For example, mild head injury may not stand in court as a cause of epilepsy. Additional requirements are absence of any manifestation of epilepsy before the injury in question and absence of other known causes of epilepsy. Of course, it is also possible that a preexisting epilepsy can be worsened by an injury [7].

Diagnostic Methods

The clinical interview and medical history taken are critical to the assessment of seizures and epilepsy. The history is obtained from both the patient and observers who have witnessed typical attacks. In the description of clinical seizures, evidence for focal seizure onset is sought, for example, the presence of an aura, initial focal sensory symptoms, or motor signs. In addition, the type of aura and the first signs in partial seizures help to localize the seizure onset within the brain. The past history, including prenatal, birth, and early development, and family history may provide insight into the etiology. All of the above may provide the clinician with the opportunity to classify the seizure types and the epileptic syndrome, and identify etiologic factors. The neurological examination is important to identify the focal signs of neurological dysfunction, which would favor partial epilepsy.

If the description of the events is strongly suggestive of epileptic seizures, the EEG is an important test to help confirm the presence of a seizure tendency. The EEG is a measure of electrical brain activity, usually recorded from scalp electrodes. Patients with epilepsy frequently have EEG abnormalities in between seizures, called *interictal* abnormalities. Some of these abnormalities, such as spikes or sharp waves, are fairly specific for epilepsy and are referred to as *interictal epileptiform* discharges. In partial epilepsies, the interictal epileptiform discharges tend to be focal or regional. These discharges generally have a topographic correlation with the epileptogenic zone, the zone from which seizures are generated. Recordings obtained during seizures will generally show a rhythmic discharge. In generalized epilepsies, epileptiform discharges tend to be generalized and synchronous in the two hemispheres. Approximately 50% of persons with epilepsy will show abnormal electrical activity on their first EEG recording and more than 90% by the fourth recording. There is no apparent benefit to additional EEGs after the fourth recording.

The EEG yield can be enhanced with prolonged recording and simultaneous video monitoring, allowing for a direct correlation to be made between

the behavioral manifestations of the seizure and the changes observed on the EEG. This procedure is usually needed when there is some question about the nature of the seizure, such as whether the spells experienced by the patient might represent psychogenic seizures or pseudoseizures. The psychogenic seizure patient will evidence no EEG changes concomitant to the seizure being videotaped. EEG/video recordings also allow for the localization of seizure onset and evolution. Thus this method is particularly useful for the epilepsy patients considered for surgery.

The recording of a seizure on EEG/video can provide a definitive proof of the existence of epilepsy. The recording of interictal epileptiform discharges indicates a seizure tendency and suggests the presence of epilepsy, but cannot prove the presence of epilepsy beyond doubt. It is much harder, almost impossible, to prove the absence of epilepsy. If typical attacks are recorded and are determined to be nonepileptic, and the EEG is always normal in between seizures, then the presence of epilepsy becomes very unlikely.

In the patient with seizures, imaging techniques are useful to identify a cause of epilepsy. A structural imaging study of the brain, preferably with magnetic resonance imaging (MRI), should be sought to identify structural abnormalities that may be epileptogenic. In the epilepsy surgery candidate, positron emission tomography (PET), which measures glucose uptake by the brain, is useful to identify functional deficits corresponding to the seizure focus. The epileptogenic focus generally has decreased glucose uptake in between seizures and increased glucose uptake during a seizure.

Differential Diagnosis

Several medical conditions can imitate seizures, thus requiring careful differential diagnosis. These conditions can be broadly divided into psychiatric and physiological imitators.

Several psychiatric disorders can imitate seizures. Panic disorder is marked by spells of autonomic hyperactivity, including shortness of breath, heart palpitations, sensations of heat or cold, gastric distress, and motor tension. Panic attacks are usually characterized by intact awareness throughout the spell, including fear of having a heart attack or losing one's mind, and an absence of postictal state or incontinence. Episodic dyscontrol syndrome, intermittent explosive disorder, or "rage attacks" also must be differentiated from seizures. The patient with intermittent explosive disorder, though perhaps reporting amnesia for the event, will not evidence altered consciousness or a postictal state. Typically, precipitants for the rage events will be readily identifiable.

One of the most common imitators of seizures and epilepsy is the condition of psychogenic nonepileptic seizures (PNES), often called *pseudoseizures* or *pseudoepileptic seizures*. PNES are episodes of alteration in movements or responsiveness, which resemble epileptic seizures in some ways, but are purely emotional in nature and lack a concomitant cerebral electrical discharge. They occur most commonly through unconscious "conversion reactions", but less frequently through voluntary malingering. PNES may account for an estimated 20–30% of patients referred for intractable seizures. PNES are more common in young people, with a relative female predominance. The clinical manifestations of PNES are extremely variable. There may or may not be a reported aura. The onset is often gradual, and always occurs when the patient is awake, even if the patient appears to be asleep. PNES often include motor manifestations, but physical collapse or altered responsiveness may be the only observable manifestations. Many clinical features have been used to diagnose psychogenic attacks and to distinguish them from epileptic seizures. Features that support a diagnosis of PNES include out-of-phase upper and lower extremity movements, side-to-side head movement, and forward pelvic thrusting. Other suggestive clinical features include a gradual onset, "preictal" behavioral changes, "pseudosleep" before seizure onset, discontinuous seizure activity, prolonged duration ("pseudostatus epilepticus" is common), gradual cessation, absence of postictal state, high seizure frequency, excessive variability in ictal manifestations, nonphysiologic progression, eye closure during unresponsiveness, eye fluttering, resistance to eye opening, vocalizations consisting of gagging, retching, gasping, screaming, crying or moaning, retained consciousness and recollection of events with bilateral jerking activity, emotional displays such as crying during events, the presence of an emotional trigger, and the occurrence of events only in the presence of others. If the examiner can suggest the patient into initiating or stopping seizure activity, a psychogenic seizure is suspected. Documented incontinence, tongue biting, and self-injury during attacks

suggest epilepsy, but these features are also reported by patients with PNES. No feature alone is definitive, but a combination of the features noted above can improve the ability to distinguish PNES from epileptic seizures. An elevated serum prolactin level 15–30 min after a seizure also suggests epilepsy, but a normal level does not exclude epilepsy. Prolonged EEG/video monitoring is usually necessary for the definitive diagnosis of PNES. A history of childhood sexual or physical abuse is commonly obtained (approximately 30%). Depression, anxiety, personality disorders, and somatoform disorders are also common and are important to identify and treat. About 10–20% of patients with PNES also have epilepsy. The investigation for coexistent epilepsy is important for treatment, but the exclusion of coexistent epilepsy can be very difficult.

An important condition in which seizurelike episodes are consciously generated is Munchausen syndrome or factitious disorder. Patients with this condition have a strong intrapsychic need to fill the patient role. Munchausen's syndrome by proxy refers to individuals who have a need for another (typically their child) to fill a physically ill patient role. Affected individuals are often very knowledgeable through study of medical works and accounts of seizure symptoms. Other patients may malinger seizurelike events in order to obtain personal injury or worker's compensation benefits, or to obtain antiepileptic medication. It is possible that malingering patients choose an episodic disorder to simulate in order to avoid direct scrutiny of their symptoms in the examination room.

Among physiological disorders imitating epilepsy, syncope is a common symptom related to a transient global reduction in cerebral blood flow. A cardiac origin, such as an arrhythmia, is more of a concern in older individuals, whereas in adolescents and young adults, syncope is most often neurally mediated. Neurally mediated syncope usually features a prodrome of nausea, lightheadedness, and dimming of vision. An observer may witness pallor. This prodrome is usually longer than the aura preceding an epileptic seizure. Up to 90% of syncope is associated with motor activity, predominantly brief multifocal myoclonus that lasts for a few seconds. This is to be distinguished from epileptic tonic–clonic activity, which is longer in duration and synchronous on the two sides. Recovery is usually much faster with syncope than with a seizure, without a "postictal state". Syncope is usually not associated with tongue biting or incontinence. In neurally mediated syncope, the patient usually has enough warning to sit down or at least break his or her fall, whereas patients with epileptic seizures may collapse suddenly, without any warning. If there is doubt, an EEG can be helpful.

Other medical disorders that can be confused with seizures include migraine headaches, which can have a prodromal period marked by visual or somatosensory symptoms, vertigo, or confusion. The characteristic headache and nausea that follow the prodrome usually make the diagnosis clear, but they are not always present. Migraine symptoms tend to have a more gradual onset and a longer duration than those of seizures. Transient ischemic attacks (TIAs) are usually characterized by negative symptoms and signs, such as numbness, visual loss, or weakness, whereas seizures most often produce positive phenomena, such as paresthesias, hallucinations, or involuntary motor activity. Other conditions that can be confused with seizures are transient global amnesia, a transient amnestic state that may mimic a temporal lobe seizure, and "drop attacks of the elderly", a syndrome of sudden falling without self-described loss of consciousness or postictal state. Hypoglycemia can cause lightheadedness, a sense of hunger, and an alteration of consciousness, which can lead to syncope or in some instances seizure activity, if the patient does not ingest glucose. Movement disorders such as hemiballismus, dystonia, and myoclonus can occasionally be confused with seizures. A few sleep disorders such as parasomnias and REM behavior disorder occasionally masquerade as seizures. Parasomnias such as sleep walking, sleep talking, or night terror arise from deep, slow wave sleep, whereas seizures more often arise in light sleep. REM behavior disorder arises from an abnormal REM sleep in which inhibition of motor activity is impaired. Narcolepsy includes sleep attacks, cataplexy, or loss of tone with emotion or startle, visual hallucinations upon falling asleep or waking up, and sleep paralysis, an inability to move upon waking up.

Adverse Consequences of Seizures and Epilepsy

There is an increased risk for epileptic patients to suffer from injuries, primarily due to falls and other accidents caused by the seizures themselves. Burns due to seizures are very common and account for

0.8–3.7% of burn center admissions [8, 9]. There is an increased risk of fractures, concussions, and spinal cord injuries as well. There is likewise an increased risk of mortality among seizure patients, including those caused by drowning and other accidents. Sudden unexplained death in epilepsy (SUDEP) is one of the leading causes of epilepsy mortality. Death occurs most often in association with a seizure, and disproportionately during sleep. There is no apparent structural cause, and the most likely cause of death is either respiratory depression or cardiac arrhythmia, or a combination of both. Seizure control is the most important measure for reducing mortality due to SUDEP.

Occupational impairment and disability are conditions complicating the lives of many persons with epilepsy. A large proportion of individuals with epilepsy lose their jobs upon acquiring the illness. For many others, the threat of seizures occurring while at work is a significant burden. The Social Security Administration identifies epilepsy as a "listed impairment" for which disability benefits are awarded, though many individuals even with intractable seizures often have difficulty obtaining their benefits through the Social Security disability determination system.

Driving restrictions due to epilepsy account for a considerable portion of the impairment in daily activities encountered by epilepsy patients, and is a factor that figures heavily into epilepsy patients' ratings of their quality of life. In the United States, in 28 states patients must be seizure-free for a specific period, ranging from 3 to 12 months, before they are allowed to drive. In other states, medical advisory boards and/or the patient's physician play a role in determining fitness to drive. Many states do not shield the physician from legal liability for their decision. Factors that reduce the likelihood of accidents include the presence of a 6- to 12-month seizure-free interval, and, in the individual patient, the presence of reliable auras, monitoring and adjustment of antiepileptic drugs (AEDs), and the absence of previous accidents. Patients with an established pattern (more than six months) of seizures occurring exclusively in sleep or seizures without change in awareness or responsiveness do not need to be restricted from driving [10].

Discrimination against persons with epilepsy is a serious obstacle to employment and the maintenance of normal social relationships. The stigma perceived by epilepsy patients, and internalization of shame about their condition, are factors in their perceived quality of life. The appearance of stories about seizures in the print media have been shown to often exaggerate the risk of dying during a seizure, to overestimate the curative powers of anticonvulsant medications, and to use demonic imagery in stories about seizures. The use of the term *epileptic*, discouraged by epilepsy associations, is often used in the print media. The American with Disabilities Act provides guidelines and regulations to employers and employees in the workplace regarding seizures and the management of and responsibilities of employees with epilepsy.

Treatment

There is general agreement that treatment has to be initiated after two seizures, because the risk of recurrence is very high. However, it is not clear if treatment is necessary after a single seizure. A number of factors predict a higher risk of seizure recurrence and encourage initiation of therapy after a single seizure. These factors include the presence of a known cause of epilepsy and the presence of epileptiform discharges on the EEG. Patients with a low risk of seizure recurrence have to be involved in the treatment decision and have to balance the risk of seizures and the risks and inconvenience of medication therapy.

The initial treatment of epilepsy is with one AED, chosen based on efficacy for the specific seizure type, and taking into account the specifics of each patient. Approximately 50% of patients will become seizure-free with the first AED. For those who are resistant to the first AED, a second AED is substituted or added. Other treatment methods for treatment-resistant patients include the ketogenic diet, particularly in some forms of pediatric epilepsy. The patient is put on a restricted high fat and low carbohydrate diet, which causes ketones to accumulate in the body, with resultant anticonvulsant effects. The vagal nerve stimulator is the only device that is currently approved for epilepsy therapy. An electrode is implanted in the neck and wrapped around the left vagus nerve and a pulse generator and battery are implanted over the chest and connected to the electrode. The optimal stimulation parameters are identified through gradual titration. Approximately

two-thirds of patients with epilepsy are resistant to treatment. If the epilepsy is focal, surgical resection of the epileptogenic zone is considered and can be highly effective.

AEDs can have a problematic impact on patient behavior and cognition. Many of these agents have side effects of sedation, impaired attention, irritability, and impaired memory. Adverse effects that are of potential legal relevance include suppression of intelligence and learning ability in children taking AEDs, as well as the potential for impaired intelligence and overall neuropsychological functioning in children born to mothers taking AEDs during pregnancy. Evidence-based practice parameters promulgated by the American Academy of Neurology conclude that AEDs in general are implicated in teratogenic effects during pregnancy. The risk may be greatest for mothers taking valproic acid and carbamazepine, or for those taking multiple medications. The risk of major malformations (structural abnormalities requiring surgery to prevent death or dysfunction) is approximately doubled in children exposed to AEDs *in utero*, from 2–3% to 4–6%.

For patients who become seizure-free on seizure medications, treatment is continued for at least two years. The decision to stop the seizure medications has to take into consideration factors that predict a higher recurrence rate. These factors include an abnormal EEG and the presence of brain damage. The patient also has to understand the potential consequences of seizure recurrence. There is a small risk that recurrent seizures may not be as easy to control as the initial seizures.

Seizures, Aggression, and the Law

In criminal trials, the contention that violent behavior has occurred as a result of a seizure disorder is sometimes offered as a criminal defense. In such cases, the question is whether the defendant possessed a culpable mental state at the time of the alleged events, or whether the act was an automatic, nonvolitional one caused by a seizure or its after effects. Technically, such a defense is not an insanity defense (*see* **Insanity: Defense**), but is a defense based upon alleged automatism. The law views acts carried out in a seizure as automatic, with no volitional component, hence there is no criminal culpability (*see* **Automatism as a Defense to Crime**).

Ictal aggression, the commission of a violent act during a seizure itself, is thought to be very rare. Presumably, such an action would be part of an often-repeated, stereotyped automatism typically manifested by the person with the seizure disorder. Such aggression is rare, and would be accompanied by a confused mental state in which the person attacks another in a haphazard, involuntary fashion without a triggering event. The attack is usually limited to pushing, clutching, or shoving. According to Treiman, all documented cases of possible ictal aggression reviewed by him showed that the behavior was either (i) nonaggressive violent automatisms that were stereotyped and repetitive from seizure to seizure, (ii) reactive automatisms manifested by directed aggression after the onset of a clearly identifiable complex partial seizure, or (iii) resistive violence at the end of a complex partial or GTC seizure occurring while the individual was being restrained while still in a postictal, confused state [11].

In the case of aggressive automatisms, the patient's behavior before the episode would be normal, and no specific triggers for the aggression will be present. The onset is quite sudden, with normal behavior one moment and the next displaying inappropriate behavior. Observers will typically report that the patient suddenly stopped, began to stare, and began to show evidence of confusion and automatisms. The aggression will be poorly defined and inappropriate, with attacks on those who happen to be physically nearby, rather than on those with a psychological connection to the patient. Following the conclusion of the episode, consciousness will return. The patient will likely not recall the incident.

Treiman has described guidelines for determining whether there is a relationship between a violent event and epilepsy prior to proffering an expert legal opinion [11]. These include that the diagnosis of epilepsy should be established by a neurologist with a special competence in epilepsy, the presence of epileptic automatisms should be documented by the history and EEG/video monitoring, the presence of aggression during epileptic automatisms should be verified in a video-recorded seizure, the aggressive act should be characteristic of the patient's habitual seizures as shown in the history and should be of short duration and not in response to any external stimulus except restraint, and that the neurologist should judge whether the act followed the known sequence of behavioral changes in complex partial

seizures or whether it was too complex a behavior to have been carried out as part of an epileptic automatism.

The presence of interictal violence is also sometimes in question in cases in which an individual with a seizure disorder commits a violent act, but not during a seizure itself. There does not seem to be an association between epilepsy and violence, other than perhaps beyond the known association between mild brain dysfunction and impulsivity. There is no evidence that seizure disorder itself predisposes individuals to violent acts. Thus, there is no reason to question criminal responsibility in cases of alleged interictal violence, though Fenwick has attributed some cases of mild aggressiveness to prodromal irritability occurring prior to seizure activity [12]. Such irritability is not likely to contribute to an insanity defense, though it may lead to an argument of decreased culpability for a given act of aggression (*see* **Mitigation Testimony**).

References

[1] Commission on Classification and Terminology of the International League Against Epilepsy (1981). Proposal for revised clinical and electroencephalographic classification of epileptic seizures, *Epilepsia* **22**(4), 489–501.

[2] Commission on Classification and Terminology of the International League Against Epilepsy (1989). Proposal for revised classification of epilepsies and epileptic syndromes, *Epilepsia* **30**(4), 389–399.

[3] Annegers, J.F. & Coan, S.P. (2000). The risks of epilepsy after traumatic brain injury, *Seizure* **9**(7), 453–457.

[4] Annegers, J.F., Hauser, W.A., Coan, S.P. & Rocca, W.A. (1998). A population-based study of seizures after traumatic brain injuries, *The New England Journal of Medicine* **338**(1), 20–24.

[5] Barry, E., Krumholz, A., Bergey, G.K., Chatha, H., Alemayehu, S. & Grattan, L. (1998). Nonepileptic posttraumatic seizures, *Epilepsia* **39**(4), 427–431.

[6] Hudak, A.M., Trivedi, K., Harper, C., Booker, K., Caesar, R.R. & Agostini, M.A., Van Ness, P.C., Diaz-Arrastia, R. (2004). Evaluation of seizure-like episodes in survivors of moderate and severe traumatic brain injury, *The Journal of Head Trauma Rehabilitation* **19**(4), 200–295.

[7] Tai, P.C. & Gross, D.W. (2004). Exacerbation of pre-existing epilepsy by mild head injury: a five patient series, *The Canadian Journal of Neurological Sciences* **31**(3), 394–397.

[8] Rimmer, R.B., Bay, R.C., Foster, K.N., Jones, M.A., Wadsworth, M. & Lessard, C., Mathieson, K., Caruso, D.M. (2007). Thermal injury in patients with seizure disorders: an opportunity for prevention, *Journal of Burn Care and Research* **28**(2), 318–323.

[9] Wirrell, E.C. (2006). Epilepsy-related injuries, *Epilepsia* **47**(Suppl 1), 79–86.

[10] Krauss, G.L., Ampaw, L. & Krumholz, A. (2001). Individual state driving restrictions for people with epilepsy in the US, *Neurology* **57**(10), 1780–1785.

[11] Treiman, D.M. (1999). Violence and the epilepsy defense, *Neurologic Clinics* **17**(2), 245–255.

[12] Fenwick, P. (1989). The nature and management of aggression in epilepsy, *The Journal of Neuropsychiatry and Clinical Neurosciences* **1**(4), 418–425.

JAMES S. WALKER AND BASSEL ABOU-KHALIL

Self-Incrimination: Capacity to Waive ***see*** Capacity to Waive Miranda Rights

SEM ***see*** Microscopy: Scanning Electron Microscopy

Semen: Differential Extraction ***see*** Differential Extraction

Semen: Test for ***see*** Acid Phosphatase

Sentencing: Demographic Factors in

A jury trial is a complex process that includes many factors for jurors to consider when trying to render a judgment of guilt or determine sentencing length for a guilty defendant.[a] Evidence shows that jurors try to make good decisions by considering the case at hand, the strength of the evidence, and other appropriate factors; but there is also ample research to show that jurors' decisions are also affected by extralegal factors. In particular, demographic factors of the defendant such as race, age, and gender have also been shown to affect decisions. These factors can make expert testimony appropriate to the extent that knowledge of their impact may assist the trier of fact (*see* **Mitigation Testimony**).

Race of the Defendant

Although racial inequalities are not as prevalent as they were a century ago, subtle inequalities continue to exist in the American courtroom. Most of the research, focused on the racial disparity between Caucasians and African-Americans, reports prejudiced decisions from white jurors when the defendant is of African-American decent [1–4]. White jurors are also more likely to convict and impose harsher sentences on a member of another race than on a member of their own race. However, Caucasians are not the only ones displaying racial bias; juries predominantly made up of African-Americans are less likely to convict a black defendant compared to a white defendant who commits the same crime [2, 5, 6] and individual jurors of African-American decent are also more likely to convict a Caucasian defendant than they are to convict a fellow African-American [3, 7]. Although these studies suggest that racial bias is still an important part of the American legal system, other studies have failed to detect a racial bias in verdict decisions [e.g., [8]] and sentencing judgments [e.g., [9]]. Two relevant theories can explain these disparate findings and help explain why some studies have found an effect of racial bias while others have not.

The phenomenon of ingroup/outgroup bias is a robust finding in the psychological literature [3] and is one explanation for why jurors of one race are more lenient with members of their own race and harsher when defendants represent a different racial group. According to this explanation, individuals display a strong preference for members of their own group (e.g., race) and have negative attitudes towards outgroup members [e.g., [10, 11]]. The ingroup/outgroup bias has been found in a variety of studies and across many different contexts [e.g., [12, 13]]. For example, Pettigrew [13] found that positive behavior of ingroup members is attributed to an inherent disposition, whereas positive outgroup member behaviors are attributed to situational factors. The opposite would be true for negative behaviors. This theoretical perspective would explain why juror bias could be exhibited when the defendant is of another race than the juror. However, the ingroup/outgroup bias phenomenon cannot explain why some studies have failed to detect racial bias in juror sentencing and verdict decisions.

Gaertner and Dovidio [14] postulated a second explanation for the research findings. They suggest that a new form of racism, namely *modern racism* or *aversive racism*, can explain the failure to detect juror bias in several studies [9, 15]. According to Gaertner and Dovidio, the fact that there has been a decline in the display of overt prejudice and explicit racist beliefs does not mean that racism does not exist any more, it simply means that individuals censor their public display of racism, because society does not condone most forms of racism today and Caucasians have moved toward an egalitarian value system [4].

This implicit racism is more likely to manifest itself in public policies (e.g., affirmative action) and is usually not blatantly displayed. Gaertner and Dovidio [14] suggest that Caucasians make a conscious effort to appear egalitarian as long as they are aware that a racial issue is relevant. Thus, it is only when Caucasians interact with members of another race, and no explicit racial issue is presented, they will show prejudice.

When a motivation to appear egalitarian is "not" activated individuals will display racial bias. Taken to the jury level, verdict and sentencing decisions should be most impacted by race when racial prejudices are not made salient to jurors [3, 4, 14, 16, 17]. If race is made salient in a case, white jurors are less likely to act prejudicial when making sentencing recommendations [3] than when it is not particularly mentioned or weighted as important [3].

Finally, it should be mentioned that under certain circumstances jurors might be "more" likely to impose harsher sentences on members of their own race. When ingroup members perceive a member of their group negatively and as threatening to the positive image of the ingroup, they are more likely to evaluate this member in a negative way [18]. This phenomenon is known as the *black sheep effect* (BSE) and was first introduced by Marques and colleagues [19–22]. Kerr *et al.* [23] applied the BSE to a mock juror paradigm and found support for the notion that negatively perceived ingroup members would be punished more harshly and be given longer sentences than outgroup members. This however, was only the case when the evidence against a negatively perceived ingroup member was strong. When the evidence was weak, white jurors displayed the typical ingroup bias in their judgments.

Age of the Defendant

A defendant's age can be construed as either a mitigating or extralegal factor. On the one hand, the court system makes allowances for youthful offenders. Defendants under the age of 18 are often separated from adults, seen in courts designed specifically for youthful offenders, and court outcomes often focus on rehabilitation instead of punishment. In this way, age serves as a mitigator.

The age of the defendant is also an extralegal factor because once jurisdictional decisions are made age should not affect jurors' judgments in a particular case. However, numerous studies have shown that age does influence jurors' judgments. As with race, jurors should not be influenced by the age of a defendant. The adult court system imposes a just-desserts rationale for sentencing, meaning harsher offenses should receive harsher punishments [24, 25]. Age should not be a factor in such decision making. However, numerous studies show that this not the case [e.g., [26, 27]]. Young defendants (those under the age of 13) are often treated more leniently than older defendants and some studies have found that older adolescents who are tried as adults may be treated more harshly than adults who commit the same crime [27].

Several researchers have suggested that jurors' decisions may follow a utilitarian (rather than a just-desserts) approach [24, 25, 28], meaning that factors such as perceived deterrence, incapacitation, and rehabilitation will influence jurors sentencing recommendations. For example, the utilitarian approach would predict that younger offenders would be treated with more leniency than adult offenders because jurors may believe that a young offender has a better chance of being rehabilitated and being harmed more by prison than a middle-aged adult offender [29]. Also, an elderly offender could be perceived as less dangerous and blameworthy than a middle-aged adult offender and therefore, sentenced more leniently [25]. Both archival [30, 31] and experimental [28, 29] studies have supported the utilitarian perspective during sentencing.

Elderly Offenders

Wilbanks [31] examined secondary records for elderly (60 years old and up) and nonelderly (25–59 years old) offenders throughout the criminal justice process. Wilbanks found that nonelderly offenders were more likely to be incarcerated and sentenced to longer prison time than elderly offenders, whereas the elderly were slightly more likely to be convicted, but not necessarily sentenced to serve time in prison. This archival data illustrates that elderly offenders may be treated more leniently than middle-aged adults during sentencing, supporting the utilitarian perspective.

Evidence toward this leniency during sentencing of the elderly has also been found in empirical studies. Bergeron and McKelvie [29] presented mock jurors with either a murder or theft vignette in which the defendant was a 20-, 40-, or 60-year-old man. In the murder case, the 20- and 60-year-old defendants were treated with more leniency than the 40-year-old defendant. The 40-year-old defendant was sentenced to serve more prison time and was not recommended for parole as quickly as the 20- or 60-year-old defendants. These data are congruent with the utilitarian perspective and suggest that age is an important factor during sentencing that jurors consider, even if only unconsciously when rendering their decisions.

Juvenile Offenders

The other end of the age spectrum that is of importance during sentencing is that of adolescent, or juvenile offenders. As with elderly offenders, the age of juvenile offenders influences juror's judgments

of guilt. Several studies [27, 32, 33] have found a positive relationship between age and perceptions of guilt, meaning, the younger the juvenile defendant is, the less culpable he or she is perceived. Furthermore, younger defendants are also less likely to be convicted [33]. These findings are congruent with research on adolescent development [e.g., [34–36]] that postulates that adolescence is a period of poor decision making skills because of significant cognitive and social developmental changes, resulting in a decrease of blameworthiness of the adolescent for his or her behavior.

Though most young offenders are seen in juvenile courts, many states in the mid 1980s and early 1990s passed laws that would ease the prosecution of juveniles as adults [37, 38] allowing harsher punishments compared to the rehabilitative juvenile system [39, 40]. For example, some states passed laws that allowed the juvenile system to be bypassed entirely for children who reached certain ages and who committed specific crimes.

However, to date only three field studies exist that compared the differential outcomes for adults and juveniles tried as adults [41–43]. Eigen [41] found that juveniles who were charged with nonfelony killings in Philadelphia were more likely to be convicted than adults and more likely to receive a verdict of first- or second-degree murder (68%) than adults (38%). Strom *et al.* [43] obtained similar results when they compiled crime statistics in 75 counties in the United States. The authors found that juveniles tried as adults who were charged with violent crimes were more likely to be convicted compared to adult defendants. The most recent field study by Rainville and Smith [42] however found no difference between juveniles and adults who were prosecuted in an adult criminal court.

The above-mentioned field studies illustrate that juveniles who are charged with violent offenses and tried as adults often receive the same punishment as adults and sometimes even harsher punishments. Experimental research that has examined how juveniles are sentenced in an adult criminal court compared to adults has obtained similar findings [e.g., [26, 27]].

Tang and Nunez [27] examined undergraduate mock jurors who were either prosecution-biased (PB) or defense-biased (DB). These mock jurors read a murder trial involving a 19-year-old adult offender, a 16-year-old juvenile offender who was tried as an adult, or a 13-year juvenile offender who was tried as an adult. The authors found that PB jurors were more likely to convict the 16-year-old juvenile offender and had higher confidence in their verdicts. These results illustrate that juvenile offenders who are tried as adults and charged with a serious offense can be at a disadvantage compared to adult offenders. Other studies of younger defendants have found that the harsh treatment of juveniles may not extend to juveniles younger than 13. In fact several studies have found that children younger than 13 are treated more leniently by mock jurors [26, 44]. This suggests that a certain cutoff age exists that would predict when juvenile offenders will or will not be treated more harshly than adult defendants.

Gender of the Defendant

Gender of the defendant has received less research attention; however, several consistent findings have emerged. Early work on gender differences in sentencing decisions found that women were at an advantage compared to men [45, 46] during sentencing. Since then not much has changed with regard to the differential treatment of women and men during sentencing. The finding that women receive lighter sentences than men has become a robust one in the literature [e.g., [47]]. Extensive literature reviews [48, 49] emphasize the strength of the association between gender and sentencing and some scholars [48–50] even argue that the gender of the defendant is the strongest demographic predictor during sentencing, even more so than race or age.

The gender disparity becomes most profound at the point in the judicial process when a defendant is either sentenced to jail time or receives a nonincarcerative sanction such as probation [47]. Women are typically 12–23% less likely to spend time in jail compared to men [e.g., [25, 49, 51]]. However, if a defendant is sentenced to jail, the research is mixed and the disparity between men and women is not as profound. Some studies have found an effect of gender that suggests that women are sentenced with more leniency [25, 51–56] while others found no differences between men and women during sentencing [49, 56–59]. Some other studies have even found women to be at a disadvantage during sentencing [e.g., [44, 60]], but these studies involved cases with juvenile defendants.

Only a few studies have examined how various crime types affect this gender bias during sentencing. Rodriguez *et al.* [e.g., [47]] found in a large sample of convicted offenders in Texas that for minor offenses, such as property or drug offenses, women were sentenced less harshly than men. For violent offenses, women are just as likely as men to be sentenced to jail, however, if sentenced to jail, women received substantially shorter sentences than men.

There are two theoretical explanations that aid in explaining the differential treatment for men and women during sentencing, namely the chivalry thesis and the focal concerns theory. The chivalry thesis postulates that women are generally stereotyped as childish and not as blameworthy as men and therefore should not be punished the same way as men are [51, 57]. However, much of the current research fails to completely support this thesis. A more recent theory, the focal concerns theory, which is similar to the utilitarian approach mentioned above, stipulates that judges or other criminal justice investigators do not have enough time to adequately evaluate all the information for a given case and that their judgments often incorporate some form of human error. They therefore rely on shortcuts that will help them make the best decision at hand. Three focal points that will typically aid them in their decision making are blameworthiness, dangerousness of the defendant, and practical constraints [47]. Men may be perceived as more likely to survive prison than women [25] or may be perceived as more dangerous than women [56]. Focal concerns theory is at the core of explaining why women receive milder sentences and if sentenced to jail they may receive shorter sentences than men.

Conclusion

On the basis of the literature reviewed here, it appears that extralegal factors still play an important role in the way defendants are treated by the criminal justice system and especially during sentencing decisions [25]. However, a jury trial is far too complex to reduce jurors' decisions down to a few variables and it still remains difficult to determine with absolute certainty what factors will ultimately influence jurors' decisions. Another level of complexity that this article did not address is that of jury deliberations [e.g., [61, 62]] and how decisions after deliberations often vary greatly from that of individual jurors. Furthermore, several demographic characteristics have not been discussed in this article, such as socioeconomic status or relationship status of the defendants, which can also impact jurors' judgments in a case. Nevertheless, it can be said that demographic factors continue to influence jurors' decisions and should not be ignored by judges or lawyers during trial proceedings.

End Notes

[a.] In most states jurors do not make sentencing decisions in felony cases. Only Arkansas, Kentucky, Missouri, Oklahoma, Texas, and Virginia allow jurors to sentence defendants in felony cases.

References

[1] Fairchild, H.H. & Cowan, G. (1997). The O.J. trial: challenges to science and society, *Journal of Social Issues* **53**, 583–591.

[2] Parloff, R. (1997b). Speaking of junk science, *American Lawyer* **19**, 5–8.

[3] Sommers, S.R. & Ellsworth, P.C. (2000). Race in the courtroom: perceptions of guilt and dispositional attributions, *Personality and Social Psychology Bulletin* **26**, 1367–1379.

[4] Sommers, S.R. & Ellsworth, P.C. (2001). White juror bias: An investigation of prejudice against black defendants in the american courtroom, *Psychology Public Policy and Law* **7**, 201–229.

[5] Parloff, R. (1997). Race and juries: if it ain't broke, *American Lawyer* **19**, 5–7, 72–74.

[6] Reynolds, G.A. (ed) (1996). *Race and the Criminal Justice System: How Race Affects Jury Trials*, Center for Equal Opportunity, Washington, D.C.

[7] Mitchell, T.L., Haw, R.M., Pfeiffer, J.E. & Meissner, C.A. (2005). Racial bias in mock juror decision-making: a meta-analytic review of defendant treatment, *Law and Human Behavior* **29**, 621–637.

[8] McGowen, R. & King, G.D. (1982). Effects of authoritarian, anti-authoritarian, and egalitarian legal attitudes on mock juror and jury decisions, *Psychological Reports* **51**, 1067–1074.

[9] Mazzella, R. & Feingold, A. (1994). The effects of physical attractiveness, race, socioeconomic status, and gender of defendants and victims on judgments of mock jurors: A meta-analysis, *Journal of Applied Social Psychology* **24**, 1315–1344.

[10] Brewer, M.B. & Brown, R.J. (1998). Intergroup relations, in *The Handbook of Social Psychology*, 4th Edition, D.T. Gilbert, ed, McGraw-Hill, Boston, MA,Vol. **2** pp. 554–594.

[11] Tajfel, H. (1991). *Social Identity and Intergroup Relations*, Cambridge University Press, Cambridge, UK.
[12] Hewstone, M. (1990). The "ultimate attribution error"? A review of the literature on Intergroup and causal attribution, *European Journal of Social Psychology* **20**, 311–335.
[13] Pettigrew, T.F. (1979). The ultimate attribution error: Extending Allport's cognitive analysis of prejudice, *Personality and Social Psychology Bulletin* **5**, 461–476.
[14] Gaertner, S.L., Dovidio, J.F. (1986). *The Aversive Form of Racism*, Academic Press, San Diego, CA, pp. 61–89.
[15] McGuire, M.V. & Bermant, G. (1977). Individual and group decisions in response to a mock trial: a methodological note, *Journal of Applied Social Psychology* **7**, 220–226.
[16] Fein, S., Morgan, S.J., Norton, M.I. & Sommers, S.R. (1997). Hype and suspicion: the effects of pretrial publicity, race, and suspicion on jurors' verdicts, *Journal of Social Issues* **53**, 487–502.
[17] Skolnick, P. & Shaw, J.I. (1997). The O.J. Simpson criminal trial verdict: Racism or status shield? *Journal of Social Issues* **53**, 503–516.
[18] Taylor, S.T. & Hosch, M.H. (2004). An examination of jury verdicts for evidence of a similarity-leniency effect, an out-group punitiveness effect or a Black sheep effect, *Law and Human Behavior* **28**, 587–598.
[19] Marques, J.M. (1990). The Black sheep effect: outgroup homogeneity in social comparison settings, in *Social Identity Theory: Constructive and critical advances*, D. Abrams & M. Hogg, eds, Harvester Wheatsheat, London, pp. 131–151.
[20] Marques, J.M., Robalo, E.M. & Rocha, S.A. (1992). Ingroup bias and the "Black sheep" effect. Assessing the impact of social identification and perceived variability on group judgments, *European Journal of Social Psychology* **22**, 331–352.
[21] Marques, J.M. & Yzerbyt, V.Y. (1988). The Black sheep effect: judgmental extremity towards in-group members in inter- and intragroup situations, *European Journal of Social Psychology* **18**, 287–292.
[22] Marques, J.M., Yzerbyt, V.Y. & Leyens, J.P. (1988). The Black sheep effect: extremity of judgments towards in-group members as a function of group identification, *European Journal of Social Psychology* **18**, 1–16.
[23] Kerr, N.L., Hymes, R.W., Anderson, A.B. & Weathers, J.E. (1995). Defendant-juror similarity and mock juror judgments, *Law and Human Behavior* **19**, 545–568.
[24] Gebotys, R.J. & Roberts, J.V. (1987). Public views on sentencing: the role of the offender, *Canadian Journal of Behavioural Science* **19**, 479–488.
[25] Steffensmeier, D., Ulmer, J. & Kramer, J. (1998). The interaction of race, gender, and age in criminal sentencing: The punishment cost of being young, black, and male, *Criminology* **36**, 763–798.
[26] Nunez, N., Dahl, M.J., Tang, C.M. & Jensen, B.L. (2007). Trial venue decisions in juvenile cases: mitigating and extralegal factors matter, *Legal and Criminological Psychology* **12**, 21–39.
[27] Tang, C.M. & Nunez, N. (2003). Effects of defendant age and juror bias on judgment of culpability: what happens when a juvenile is tried as an adult? *American Journal of Criminal Justice* **28**, 37–52.
[28] Kapardis, A. & Farrington, D.P. (1981). An experimental study of sentencing by magistrates, *Law and Human Behavior* **5**, 107–121.
[29] Bergeron, C.E. & McKelvie, S.J. (2004). Effects of defendant age on severity of punishment for different crimes, *The Journal of Social Psychology* **144**, 75–90.
[30] Champion, D.J. (1987). Elderly felons and sentencing severity: intergenerational variations in leniency and sentencing trends, *Criminal Justice Review* **12**, 7–14.
[31] Wilbanks, W. (1988). Are elderly felons treated more leniently by the criminal justice system? *International Journal of Aging and Human Development* **26**, 275–288.
[32] Crosby, C.A., Britner, P.A., Jodl, K.M. & Portwood, S.G. (1996). The juvenile death penalty and the Eight Amendment: an empirical investigation of societal consensus and proportionality, *Law and Human Behavior* **19**, 245–261.
[33] Finkel, N.J., Hughes, K.C., Smith, S.F. & Hurabiell, M.L. (1994). Killing kids: the juvenile death penalty and community sentiment, *Behavioral Sciences and the Law* **12**, 5–20.
[34] Grisso, T. (1996). Society's retributive response to juvenile violence: a developmental perspective, *Law and Human Behavior* **20**, 229–247.
[35] Grisso, T., Steinberg, L., Woolard, J., Cauffman, E., Scott, E., Graham, S., Lexcen, F., Repucci, N.D. & Schwartz, R. (2003). Juveniles' competency to stand trial: a comparison of adolescents' and adults' capacities as trial defendants, *Law and Human Behavior* **27**, 333–363.
[36] Heide, K.M. (1997). Associate editor's editorial: killing words, *International Journal of Offender Therapy and Comparative Criminology* **41**, 3–8.
[37] Donegan, C. (1996). Preventing juvenile crime, *The CQ Researcher* **6**, 219–236.
[38] Kresnak, J. (1999). *Juvenile Court Turns 100, but is the Party Over? Youth Today*, The American Youth Work Center, Washington, D.C.
[39] Fagan, J.A. (1995). Separating the Men From the Boys: The Comparative Advantages of Juvenile Versus Criminal Court Sanctions on Recidivism Among Adolescent Felony Offenders, in *A Sourcebook: Serious, Violent, and Chronic Juvenile Offenders*, J.C. Howell, B. Krisberg, J.D. Hawkins & J.J. Wilson, eds, Sage Publications, Thousand Oaks, CA.
[40] Torbert, P., Gable, R., Hurst IV H., Montgomery, I., Syzmanski, L. & Thomas, D. (1996). *State Responses to Serious and Violent Juvenile Crime*, Bureau of Justice Statistics Special Report, U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention, Washington, D.C.
[41] Eigen, J.P. (1981). Punishing youth homicide offenders in Philadelphia, *The Journal of Criminal Law and Criminology* **72**, 1072–1093.

[42] Rainville, G.A. & Smith, S.K. (2003). *Survey of 40 Counties, 1998: Juvenile Felony Defendants in Criminal Courts*, U.S. Department of Justice, Washington, D.C.
[43] Strom, K.J., Smith, S.K. & Snyder, H.N. (1998). *State Court Processing Statistics, 1990-94: Juvenile Felony Defendants in Criminal Courts*, U.S. Department of Justice, Washington, D.C.
[44] Nuñez, N., Dahl, M.J. & Hess, C. (2005). Juror perceptions of juveniles who commit murder: are female defendants at a disadvantage? Paper Presented at the Meeting of the American Psychology-Law Society meeting, La Jolla, CA.
[45] Pope, C.E. (1975). *Sentencing California Felony Offenders*, U.S. Government Printing Office, Washington, D.C.
[46] Nagel, S.S. & Weitzman, L.J. (1971). Woman as litigants, *Hastings Law Journal* **23**, 171–198.
[47] Rodriguez, S.F., Curry, T.R. & Lee, G. (2006). Gender differences in criminal sentencing: do effects vary across violent, property, and drug offenses? *Social Science Quarterly* **87**, 318–339.
[48] Daly, K. & Bordt, R.L. (1995). Sex effects and sentencing: an analysis of the statistical literature, *Justice Quarterly* **12**, 141–175.
[49] Steffensmeier, D., Kramer, J.H. & Streifel, C. (1993). Gender and imprisonment decisions, *Criminology* **31**, 441–446.
[50] Spohn, C. & Holleran, D. (2000). The imprisonment penalty paid by young, unemployed black and Hispanic male offenders, *Criminology* **38**, 281–306.
[51] Farnworth, M. & Teske Jr R.H.C. (1995). Gender differences in filing court processing: testing three hypotheses of disparity, *Women and Criminal Justice* **6**, 23–44.
[52] Bushway, S.D. & Piehl, A.M. (2001). Judging judicial discretion: legal factors and racial discrimination in sentencing, *Law and Society Review* **35**, 733–764.
[53] Curran, D.A. (1983). Judicial discretion and defendant's sex, *Criminology* **21**, 41–58.
[54] Engen, R.L. & Gainey, R.R. (2000). Modeling the effects of legally relevant and extralegal factors under sentencing guidelines: The rules have changed, *Criminology* **38**, 1207–1230.
[55] Mustard, D.B. (2001). Racial, ethnic, and gender disparities in sentencing: evidence from the U.S. Federal Courts, *Journal of Law and Economics* **44**, 285–314.
[56] Albonetti, C.A. (1991). An integration of theories to explain judicial discretion, *Social Problems* **38**, 247–266.
[57] Crew, K.B. (1991). Sex differences in criminal sentencing: chivalry or patriarchy? *Justice Quarterly* **8**, 59–83.
[58] Nobling, T., Spohn, C. & DeLone, M. (1998). A tale of two counties: unemployment and sentence severity, *Justice Quarterly* **15**, 159–185.
[59] Wooldredge, J.D. (1998). Analytical rigor in studies of disparities in criminal case processing, *Journal of Quantitative Criminology* **14**, 155–179.
[60] Haegerich, T.M. & Bottoms, B.L. (2000). Empathy and jurors' decisions in patricide trials involving child sexual assault allegations, *Law and Human Behavior* **24**, 421–437.
[61] London, K. & Nunez, N. (2000). The effect of jury deliberations on jurors' propensity to disregard inadmissible evidence, *Journal of Applied Psychology* **85**, 932–939.
[62] McCoy, M.L., Nunez, N. & Dammeyer, M.M. (1999). The effect of jury deliberations on jurors' reasoning skills, *Law and Human Behavior* **23**, 557–575.

Further Reading

Ulmer, J.T. (2000). The rules have changed – So proceed with caution: a comment on Engen and Gainey's method for modeling sentencing outcomes under guidelines, *Criminology* **38**, 1231–1244.

Andre Kehn and Narina L. Nunez

Serial Homicide

Introduction

The term *serial murder* (and serial killer) was not even a part of the forensic lexicon until the 1970s, when it was popularized by one of the authors (RR), then an investigator with the Behavioral Science Unit of the US Federal Bureau of Investigation (FBI) [1]. The precise definition of serial murder has been the subject of some debate, which has somewhat hampered research efforts. However, most proposed definitions share the following elements common: (i) there have been at least two victims (some definitions require three victims), (ii) victims are killed in a noncontinuous manner (i.e., there is an emotional "cooling off" period between murders), and (iii) the murders usually involve a sexual component [2–4].

Serial murder may be a universally terrifying concept, but it is an extraordinarily rare event. In a study of the frequency of serial sexual homicide, it was found to account for only 0.5% of all homicides over a 10-year period in Virginia [5]. In contrast to the sensationalized perception that serial murder is a growing "epidemic", there does not appear to be solid evidence that this is the case. An analysis of homicide victims from 1960 to 1998 indicates that the percentages of female homicide victims have actually decreased [6]. Because the victims of serial murderers

are overwhelmingly female, it has been argued that this data supports the notion that serial murder is not increasing in frequency. However, this data does not address the rate of serial murder in which victims are men or children.

Other research has challenged prior beliefs about serial murder. For example, it had previously been widely accepted that a disproportionate number of serial murderers are Caucasian. Yet researchers have demonstrated that the number of African-American serial murderers closely corresponds to the overall proportion of African-American males in the United States [7].

When a serial murderer begins to commit homicides, the local community is often dramatically impacted. Media coverage is unrelenting, and residents generally live in fear until the perpetrator is apprehended. Residents living in a community exposed to serial murder may even experience post-traumatic stress disorder symptoms for varying periods of time [8]. Yet, despite the high level of interest, there is no universally accepted theory that adequately explains the etiology of serial murder [9]. This may be partly reflective of the fact that serial murder is an event with an extremely low base rate, and therefore difficult to study *via* rigorous scientific methods [5].

Once the concept of serial murder gained widespread recognition, one expert noted that the "tendency of the press, public, and public officials to regard such individuals as mad solely on the basis of their crimes reflects the widespread need to attribute such behavior to alien forces" [10]. This tendency may be understood as the desire to disavow the notion that another human being could commit such atrocities. Could serial murder "simply be part of the spectrum of human possibility, a brutal dark side of man, not representing demons or disease"[11]? Or, do we have reliable evidence of biopsychosocial deficits in certain cases? This article provides a broad overview of the literature and research in an attempt to address these questions.

History

While the term *serial murder* may be relatively new [12], its occurrence is not. It is quite possible that serial murderers have always been among us. Perhaps the first documented serial killer was a first century Roman woman named Locusta. She was a "professional poisoner" who lived in the time of Nero (54 AD), and was ultimately executed [13]. In the fifteenth century, Gilles de Rais, a wealthy French aristocrat, raped and killed some 100 young boys because it brought him "pleasure" [14]. In sixteenth century France, it is likely that myths such as "werewolves" were used to explain the deeds of serial murderers that were too horrifying to attribute to human beings [14]. In the United States, there have been documented cases of serial murder as far back as the 1800s.

In 1886, psychiatry professor Richard von Krafft Ebing wrote the classic *Psychopathia Sexualis*, in which he described the characteristics of individuals who appeared to obtain sexual gratification from acts of sadistic domination. The next major contribution to the understanding of serial murderers was in 1970, when forensic psychiatrist Robert Brittain produced detailed clinical descriptions of sadistic murderers he had encountered over his career [15]. Beginning in the early 1970s, media coverage of notorious cases such as Ted Bundy and the Hillside Stranglers produced a sense of urgency to study and explain the phenomenon.

In the 1980s, experts working in the FBI's Behavioral Science Unit began the pioneering study that ultimately established the foundation for scholarly research of serial murder. To emphasize the sexual nature of the crimes, and to distinguish these offenders from others who murder serially for other reasons (e.g., contract killers), the term *sexual homicide* was adopted [4]. In a sexual homicide, the perpetrator engages in some form of sexual activity before, during, or after the murder. For each individual serial sexual homicide offender, the performance and meaning of the sexual element may vary.

Typology

FBI researchers gathered data from exhaustive interviews of 36 convicted serial murderers. From this data, they were able to extract and analyze core personality and behavioral characteristics [16]. The traits and behaviors identified allowed distinctions to be made between different types of serial murderers. For ease of conceptualization and communication, offenders were categorized into either "organized" or "disorganized" types. These terms

Table 1 Serial murder organized – disorganized typology

Organized offender traits	Disorganized offender traits
Good verbal skills, socially adept	Poor verbal and social skills
May live with spouse	Loner or lives with parents
Reasonably intelligent	Low intelligence
Usually employed	Under or unemployed
Planning of crime	Little-to-no planning of crime
Ruse or con to gain control of victim	Blitz or surprise attack of victim
Targeted victim	Victim of opportunity
Crime scene: suggests control, order	Crime scene: disarray
Crime scene and death scene not the same	Crime scene and death scene often the same
Movement of body	Body left at death scene
Attempts to conceal evidence	Little-to-no attempts to conceal evidence

were initially meant to help law enforcement interpret crime scenes, and are best understood as generalized distinctions occurring on a continuum. The organized–disorganized typology provided illustrative descriptors of personality and behavior, and had the advantage of being easily grasped by both law enforcement and mental health professionals. The major organized–disorganized typology traits are listed in Table 1. The term *mixed* sexual homicide was used to describe the offender, whose crime scene reflected aspects of both the organized and disorganized types, thus lying on the continuum between them. Finally, the term *sadistic* murderer describes the offender who is primarily a sexual sadist, and derives the greatest satisfaction from the victim's response to torture [4].

Since the original organized–disorganized typology was advanced, a number of other typologies have been proposed. A clinically based typology has been advanced that categorizes perpetrators as either "compulsive" or "catathymic" [17]. The *compulsive* perpetrators are similar to the FBI's organized killers. They leave organized crime scenes, and can be diagnosed with sexual sadism, as well as antisocial/narcissistic personality disorders. The *catathymic* perpetrators leave disorganized crime scenes, and may be diagnosed with a mood disorder and varying personality traits. While the compulsive type displays emotional detachment and autonomic hyporeactivity, the catathymic type are less psychopathic. In contrast, the catathymic types are autonomically hyperreactive, and may have histories of abuse. Like the organized–disorganized typology, these types are generalized concepts, and any individual case is likely to fall on a continuum between the two.

A number of researchers have pointed out that nearly all serial murderers demonstrate a core of organized features, and that it is the degree of disorganization that most differentiates them [18]. Along a similar line of reasoning, a statistical analysis of 85 sexual murderers yielded a "five-cluster" model [19]. It was noted that most offenders shared core characteristics such as preparatory behavior, precautionary behavior, and some degree of sophistication. This core of behaviors was referred to as the *undifferentiated* pattern, and the other four patterns (predator, fury, rape, and perversion) were differentiated from it. The *predator* pattern is similar to the organized or sadistic offender as described by the FBI. This offender was characterized as older, more mobile, Caucasian and likely to be living with a partner. They tend to be well groomed, collectors of crime literature and sexual materials. This group may also operate with an accomplice. The *fury* pattern represents an unfocused, explosive obliteration of the victim, and is most similar to the FBI's disorganized type. The murder is characterized by excessive violence and overall disorganization. The chances of this offender having or not having a mental disorder are roughly equal.

The *rape* pattern consists of offenders whose primary goal is sexual intercourse. They use that amount of force that is necessary to carry out the rape. Application of force is typically minimal, and there is no indication of sexual dysfunction in the offender. Thus, this pattern may be conceptualized as a sexual assault that ultimately resulted in murder. The crime scene

will not reflect any great degree of perversion, but a relatively straightforward violent sexual attack. These offenders may have some prior peripheral acquaintance with the victim, such as living in the same apartment complex. The assault usually occurs in a single indoor location, such as the victim's home. The offense is usually unplanned, and characteristic of a younger offender who attacks a victim of opportunity. Thus, the scene may reflect elements of the FBI's disorganized type. The *perversion* pattern offender engages the victim in conversation by offering reassurances. Offenders tend to be older with bisexual or homosexual orientations, while their victims tend to be younger and male. The offense usually involves some type of pedophilic assault that culminates in murder.

In general, serial murderers do appear to share some core characteristics. However, research has begun to focus on the differences that distinguish various subgroups of serial murderers. The following descriptions represent a sampling of some serial murder subgroups that have been studied, and should not be considered exhaustive.

Female Serial Murderers

Relatively little has been written about female serial murderers as compared to their male counterparts. In a review of published literature on female serial murder, the most common motive identified was material gain [20]. Sexual or sadistic motives are believed to be extremely rare in female serial murderers. Psychopathic traits and histories of childhood abuse have been consistently reported in these women [20]. In a study of 105 female serial killers, the preferred method of killing was poisoning [21]. An analysis of 86 female serial killers from the United States found that the victims tended to be spouses, children, or the elderly [22]. Sometimes referred to as *black widow* killers, these women tend to be geographically stable and live in the same area where their offenses occurred. Their victims are not strangers, and the methods they use are covert or "low profile" [21]. On rare occasions, women may be involved with a male serial killer as a part of a serial killing "team" [23].

Perhaps one of the more high profile female serial killers in the United States was Aileen Wuornos, who was convicted of killing seven men in separate incidents. Wuornos had claimed that all of the men had raped her (or attempted to) while she was working as a prostitute. Thus, she did not fit the typical profile of a female serial killer. Her case received remarkable media and Hollywood attention. In a detailed case study analysis, it was theorized that Wuornos was biologically predisposed to psychopathy, and her abusive childhood resulted in serious attachment deficiencies [24]. Finally, her aggressive narcissism and antisocial lifestyle predisposed her to situations in which she was able to commit acts of predatory murder. Wuornos was executed by lethal injection in 2002 in Florida.

Juvenile Serial Murderers

Serial murder by children and adolescents is exceedingly rare, and little scientific information is available. In a study of six cases, juvenile serial murderers were found to exhibit signs of sexual sadism, predatory violence and the use of "hands on" methods of killing [25]. The average age of the juveniles when they committed their first murder was 14. There is somewhat more data on juvenile sexual homicide, and it is hypothesized that youths who go on to commit serial murder would have originated from this group [24]. In a study of 16 juvenile sexual homicides, 10 factors were commonly found in the perpetrators [26]. These factors are listed in Table 2. Three of these factors (impaired capacity to feel guilt, neuropsychiatric vulnerabilities, and serious school problems) were found in 100% of the offenders.

Owing to increasing law enforcement sophistication, it is possible that juvenile sexual murderers will be apprehended after their first offense, making juvenile serial murder even rarer [25]. Nevertheless, professionals working with juveniles should be sensitive

Table 2 "Big 10" factors in 16 juvenile sexual murderers [26]

1.	Impaired capacity to feel guilt
2.	Neuropsychiatric vulnerabilities
3.	Serious school problems
4.	Child abuse
5.	Family dysfunction
6.	History of interpersonal violence
7.	Prior arrests
8.	Sadistic fantasy
9.	Psychopathic personality traits
10.	Personality disorder diagnosis (especially schizoid or schizotypal)

to the fact that violent fantasies can develop early in the life of a sexual murderer [27].

Physician Serial Murderers

The study of medical serial killers has gone overlooked until relatively recently. This may be due to an unwillingness to perceive sworn "healers" as potential murderers. However, research has revealed that medical killers may actually be the most prolific of all serial killers. Doctors who serially murder their patients are considered to belong to a larger group of "career-assisted killers." The term *clinicide* has been used to describe "the unnatural death of multiple patients in the course of treatment by a doctor" [28]. Such murders may be difficult to detect, since they often occur in settings where death is expected to happen. Doctors accused of clinicide will be likely to put forth the defense that they were relieving suffering or providing euthanasia. Clinicidal doctors may have extreme narcissistic personalities, and may obtain pleasure by "determining" when a person will die.

One of the most deadly doctor serial killers may also hold the dubious distinction of being one of the most prolific serial murderers to date. Dr Harold Shipman, a UK physician, was convicted of killing 15 patients with lethal injections of narcotics. In a posttrial investigation, it was concluded that Shipman was responsible for 218 known victims [29]. Other estimates have suggested that the number is closer to 450 [30]. Most of Shipman's victims were not terminally ill, nor did they have an immediate life-threatening illness. Shipman refused to speak to anyone, and no complete psychological assessment was ever performed on him [29]. He committed suicide in prison in 2004.

Other healthcare professionals have been implicated in serial murder. In a study of 90 healthcare killers, 86% were nurses and 12% were doctors [31]. Injection was the most common method used, followed by suffocation, poisoning, and tampering with equipment. Fifty-four of the 90 cases were ultimately convicted. A total of 2113 deaths could be attributed to these 54 convicted healthcare killers.

Other Subgroups

As research has progressed, specific subgroups have been singled out for more detailed analyses. For example, sexual murderers of children have been singled out for separate study and comparison. It has been found that sexual murderers of children are more often victims of childhood sexual abuse, and have more deviant sexual fantasies than sexual murders of women [32]. In addition, sexual murderers of children are more likely than sexual murders of women to: use pornography prior to the offense, have contact with the victim, premeditate the offense, use strangulation, hide and dismember the body. Another body of research indicates that sexual murderers of elderly women present a particularly distinct profile [33, 34].

It has been suggested that sexual murderers of men may differ in certain respects from those who murder women. An initial typology has been proposed, although it was based on a study of only 10 cases [35]. The *avenger type* is usually involved in prostitution, and has a history of childhood abuse. A triggering event during a sexual exchange may elicit memories of abuse, which leads to severe expressive violence and murder. The *sexual predator type* premeditates his offense, and is motivated by deviant sexual fantasy. Victims of the sexual predator type are often adolescents or young men. The offense involves abduction, confinement, and sadistic acts. The *nonsexual predator type* sets out to rob a vulnerable, often older homosexual man. The victim may be seduced, and substance use by both victim and offender occurs prior to the offense. Usually, the victim is not sexually assaulted, and the murder may have resulted from an escalation of violence during the robbery.

Motivations

Most research on the motivations of serial murderers has focused on sexual fantasies and/or a need for control and domination. Indeed, one of the most reliable psychological findings in the mental lives of serial murderers is the presence of violent sexualized fantasy. Convicted serial murderers have consistently described a high frequency of violent fantasies that are both persistent and arousing [15, 36, 37]. In a comparison of 25 serial sexual murderers with 17 single sexual murderers, it was found that the serial murder group had a higher prevalence of paraphilias and violent sexual fantasies [38]. It was theorized that serial murderers' deviant fantasies may

be reinforced *via* repeated pairing with compulsive masturbation. This has led to the speculation that deviant sexual fantasy may serve as a rehearsal of sorts for eventual murder [39]. In fact, many cases of serial murder have involved strong components of sexual fantasy, use of pornography, or other media with sexually sadistic themes [40]. Internet and digital technology may provide a new and abundant source of deviant fantasy material for the sexual murderer [41].

A study of 28 sexual murderers suggested that they may be driven by certain "implicit theories" or cognitive schemas they had about life [42]. It was found that the offenders had commonly held beliefs such as "the male sex drive is uncontrollable", and the world is generally hostile and dangerous. Other guiding cognitions included anger and resentment toward women, and a sense of entitlement. Related findings came from a study comparing 19 sexual murderers to 16 nonmurdering rapists [43]. The sexual murderers, compared to rapists, reported higher levels of grievance toward females in childhood, higher levels of loneliness in adolescence and were more likely to portray themselves as victims, as adults.

Because a number of well-known serial murderers have served in the military, a social learning theory has been hypothesized, whereby the military may provide the associations and reinforcements to enable efficient killing [44]. Military training is seen as helping the offender objectify victims, and reinforcing the tendency to use lethal aggression. Hostility and aggression, in general, have been proposed as possible motives. This theory suggests that either anger, or anger fused with sexual arousal, drives serial murder [45]. This fusion theory was proffered to explain the murders committed by Jeffrey Dahmer [46]. Here, it is important to make the distinction between rapists who murder to silence their victims, *versus* the so-called lust-murderer who derives sexual satisfaction from the act of rape–murder [27].

Finally, a convincing argument has been put forth that anger is not *causative* in sexual homicide. This hypothesis stresses that the physiology of anger and sexual arousal are usually mutually incompatible. True anger, as opposed to aggression, has an inhibitory effect on erectile and sexual functioning. This reasoning holds that sexual motivation is the primary motive in sexual murders, whereas anger and/or control over victims are secondary [3].

Offense Behavior

The study of serial murderers' offense behavior provides important data for homicide investigators. Serial murderers often leave confusing and violent crime scenes, which must be carefully analyzed for behavioral evidence. A study that compared single and serial homicide offenses found that serial offenders targeted more women, and killed more strangers [47]. While single homicide offenders killed out of anger, serial offenders appeared to be sexually motivated. The serial offenders were significantly more likely to use strangulation, move the body from the death scene and dispose of the body in a remote location. Some commonly observed serial sexual murder offense behaviors are listed in Table 3. Serial murderers may clean the victim's genital area to remove evidence, and may take personal items from the victim as souvenirs [48]. Prostitutes may represent a common victim pool for serial killers [49].

In the analysis of serial murder crime scenes, the concepts of *staging* and *posing* may play important roles. Staging is a purposeful alteration of the crime scene by an offender to throw off investigators [52]. The offender may be concerned about being a likely suspect, and wish to direct the investigation away from him or her. Three types of staging have been described: (i) staging to appear like a suicide or accident, (ii) staging to appear like a sexual homicide, and (iii) arson to destroy evidence [53]. In contrast, posing is the positioning of the victim's body by the offender. Posing, a very rare offense behavior, often involves leaving the victim in a sexually degrading position. This may either be to shock police, or simply for the offenders own gratification [52]. Serial murderers who leave posed victims may also leave evidence of binding, stabbing, or bludgeoning.

Table 3 Serial sexual murder offense behaviors [16, 50, 51]

Binding, torture
Stabbing, biting
Attempted or completed sexual intercourse (oral, anal, vaginal)
Victim left nude or seminude
Sexual positioning of victim's body
Insertion of foreign objects into victim's body cavities
Semen on or near victim's body
Victim's personal items taken

It has been found that the geographic locations of serial murderers' offenses are critically important. The geographic sites of 155 serial killer disposal sites were analyzed, and a strong relationship was found between the location of the offender's home and the disposal site [54]. The vast majority (89%) of offenders lived within a circle defined by the disposal sites that were furthest from each other. The finding that serial murderers appear to operate in a certain spatial location around their home has been called the *circle hypothesis* [54].

Psychiatric Findings

Most data on psychiatric diagnoses of serial murderers comes from individual case studies and retrospective analyses. The majority of these studies have suggested a common constellation of diagnoses: *psychopathy, antisocial personality, sexual sadism*, and other *paraphilias* (especially voyeurism, fetishism, transvestism, and sometimes necrophilia). More recent and well-designed comparison studies have yielded similar findings. For example, a study comparing sexual murderers to other general sex offenders found that the sexual murderers had greater levels of psychopathy, sadism, fetishism, and transvestism [55]. In addition, the sexual murderers began their criminal careers earlier, and had stronger histories of fire setting and cruelty to animals.

The sexual sadism seen in serial murderers must be distinguished from sexual sadism between consenting adults that would not be considered criminal. The variant of sexual sadism seen in serial murderers is at the extreme end of the spectrum, as it ultimately involves killing for sexual excitement. Such individuals may engage in torturing victims to the point of death to obtain the "pleasure in complete domination" over them [56]. To better capture this distinction, the diagnosis *"sexual sadism, homicidal type"* has been proposed for serial sexual murderers [3]. It is noteworthy that a higher level of gratuitous and sadistic violence was found in the homicides committed by psychopathic sexual murderers when compared to nonpsychopathic sexual murderers [57].

There is considerable overlap between the construct of psychopathy, and another diagnostic construct described in serial murderers called *malignant narcissism. Malignant narcissism* has been defined as an extreme variant of narcissistic personality disorder, antisocial personality, sadism, and a tendency to paranoid reactions [58]. This clinical construct has been proposed by a number of theorists studying serial murder [59]. It is suggested that malignant narcissism is best viewed as an aspect of personality structure, as opposed to a formal mental disease. This theoretical concept has been used to explain the chronic sadistic arrogance and lack of empathy observed in most serial murderers.

Paraphilias

Paraphilias, particularly voyeurism and fetishism, have been described in many serial murderers. Over 70% of sexual murderers had these paraphilias in the early FBI studies [16]. Some individuals with voyeurism and fetishism may engage in burglaries that actually serve the purpose of gratifying these two paraphilias [60]. Offenders may steal sexually related items during a burglary, such as female undergarments. Such offenses may, in fact, be part of a progression to an eventual sexual homicide. In a descriptive review, it was found that serial murder was associated with multiple paraphilias [61]. It was speculated that multiple paraphilias served to enhance and reinforce the overall sexual experience of the offender.

Obsessive–Compulsive Traits

Focusing on the seemingly compulsive nature of the offenses, researchers have speculated about the significance of the obsessive qualities of the serial murderer, particularly the organized type. These individuals demonstrate a tendency toward orderliness, obsessive, fantasy, and ritualistic behavior (posing the body, biting, inserting objects, etc.) during their murders that suggest compulsive qualities. A study finding high rates of compulsive masturbation in sexually sadistic serial murderers may support the hypothesis of underlying obsessive–compulsive traits [62]. Other support comes from Rorschach studies of sexual murderers who evidenced high levels of obsessional thinking [63, 64]. Experts believe that these obsessive and compulsive traits, combined with higher than average intelligence, permit the organized offender to avoid apprehension.

Other Disorders

There appears to be a very low rate of psychosis among serial murderers, and approximately half

Table 4 Psychiatric disorders found in serial murderers

Psychopathy
"Malignant" narcissism
Antisocial personality disorder
Borderline personality disorder
Schizoid and Schizotypal personality disorders
Substance use disorders
Paraphilias (sexual sadism, voyeurism, fetishism, transvestism, and necrophilia)
Asperger's disorder

of perpetrators report substance use prior to their offenses [16]. At the present time, there is no conclusive evidence that specific organic factors play a causal role in the creation of a serial murderer. However, studies have found right temporal lobe abnormalities [65], as well as subtle neurological abnormalities in sexual sadists [66].

Neurodevelopmental deficits have been hypothesized as a possible contributing factor to serial murder (Table 4), most notably in the case of Jeffrey Dahmer [67, 68]. Investigators have described an association between autism spectrum disorders and a subgroup of serial murders, and propose that Dahmer may have suffered from Asperger's disorder. Along these lines, it is interesting to note that after exhaustive interviews with Dahmer, one author (RR) was impressed by the "peculiar" nature of Dahmer's presentation [69].

A possible neuropsychiatry genetic defect has been found in a few cases of sexual homicide. The sex chromosome abnormality XYY had previously been associated with criminal behavior [70]. More recently, researchers found three men with the XYY chromosome abnormality in a sample of sexual homicide offenders [71]. The rate was found to be higher than what would be expected among prisoners or the general population. All three men were diagnosed with sexual sadism and psychopathy.

Developmental Theories

A number of different psychosocial theories have been put forth regarding the developmental etiology of serial murder. Investigators have speculated that the behavior may result from a deadly convergence of (i) early childhood attachment disruptions, (ii) psychopathy, and (iii) early traumatogenic abuse [72]. However, there is conflicting evidence on the presence of child abuse in the development of serial murderers. When the FBI studied 36 serial murderers, many of them had a history of either abuse or neglect [16]. Forty-three percent reported a history of childhood sexual abuse, and 74% reported a history of psychological abuse, which typically involved humiliation. A study of 48 homicidal sex offenders found high rates of childhood abuse and removal from the family home [73]. In contrast, others have found that a majority of sexually sadistic murderers had no evidence of childhood abuse [56, 74]. One possibility accounting for these differences may be due to heterogeneity in the populations studied.

When sexual murderers with a history of sexual abuse were compared to murderers without such a history, significant differences were found [75]. Sexual murderers with a history of early sexual abuse were more likely to begin fantasizing about rape earlier. In addition, they developed more severe sexual deviancy. Besides sexual abuse, the family histories of many sexual murderers reveal unstable environments, which may predispose them to disordered early life attachments. Approximately 70% of sexual murderers' families had histories of alcohol abuse, and about 50% had family members with criminal histories [76]. It has been hypothesized that parental neglect, from either absence or preoccupation with their own life problems, may have further impaired these men's ability to form healthy attachments.

Animal cruelty has been a common finding in the childhood and adolescent developmental stages of many serial murderers. The link between animal cruelty during childhood and subsequent physical violence during adulthood has been demonstrated in a number of studies [77, 78], resulting in the addition of animal cruelty to the diagnostic and statistical manual of mental disorders, 3rd edition, revised (DSM III-R) as a symptom under the diagnosis of conduct disorder in 1987. In keeping with the developmental theme of conduct disorder symptoms, researchers have also commented on a possible link between childhood fire setting and adult serial murder [79]. The association between serial murder and enuresis has not been as strongly born out in the research to date.

Obviously, all children who are diagnosed with conduct disorder and/or engage in animal cruelty do not go on to become serial murderers. Nevertheless, it is thought that in the cases of those who do, an early "practicing" of violent and sadistic behavior on a living creature may play a role in desensitizing the individual to violence against humans. This notion

has been termed the *graduation hypothesis* to denote the progression or "graduation" from animal cruelty to sadistic acts against humans [80]. Thus, some individuals may progress past mere desensitization, toward intense pleasure from acts of cruelty and ultimately murder.

Psychodynamically oriented investigators have theorized that a sexually provocative mother may contribute to the formation of a serial murderer [17, 81]. It is important to note that this premise is far from another "blaming" of the mother theory. Rather, investigators point to documented instances of strikingly inappropriate sexual behavior on the part of the mother that would easily qualify as sexual abuse. Evaluations of some convicted serial murderers suggest that a displacement of aggression from their mothers onto to their female victims was present during their offenses.

Upon review of the developmental theories and individual case studies, the following traits are frequently observed in serial murderers: deviant sexual interests, a strong need for control of a victim, a very active deviant fantasy life, and psychopathic traits allowing for the objectification of victims. When these traits are synthesized into a gestalt, the clinical picture that emerges is one of an individual who spends excessive time in a reverie of deviant fantasy, has a tendency toward isolation, a need for totally submissive partners, and a preference for autoerotic pleasure [82]. As can be imagined, such an individual will be unlikely to have healthy relationships, and subsequently must depend on fantasy for gratification. However, at some point, mere fantasy becomes an insufficient source of pleasure for the potential serial murderer.

It is theorized that what follows is a gradually progressive series of "try outs," where he attempts to turn his fantasies into reality. For example, an offender may begin by simply following a potential victim. This may next progress to voyeurism or breaking into victim's homes [60]. During a burglary, the offender may steal fetishistic items for sexual pleasure. When this fails to provide sufficient satisfaction, the offender may progress to rape and ultimately murder. The behavior is positively reinforced over time through paired association with masturbation, making the deviant fantasies extremely refractory to extinction[38]. Each time the offender murders a victim, there is further stimulation of fantasy and an overall reinforcement of the cycle.

Assessment and Prevention

Evaluations of serial murderers may occur for a number of different purposes. Interviews may be conducted by law enforcement upon apprehension, by forensic mental health in the event that a mental disease defense is raised by the defendant, and after conviction for research purposes. Forensic assessments of suspected serial murderers are best done by those with experience evaluating psychopathic and serial sexual offenders. Dishonesty and underreporting of deviant fantasies and offenses are commonplace, and a meticulous review of collateral data prior to the evaluation is necessary. Individuals who have already confessed to murders may still be unwilling to discuss the sexual nature of their offenses for a variety of reasons, the most common being the fact that sex offenders are severely harassed by other inmates in prison. In the authors' experience, many such individuals will want to focus on the aggressive aspects of their crimes, and downplay the sexually deviant aspects.

Interviewing Strategies

Interviewing strategies and approaches will need to be tailored to both the purpose of the interview and the individual nuances of the case. Regardless of the purpose of the interview, it is essential to gather and review all relevant collateral data prior to interviewing the offender. This will not only lend important perspective but will also better allow the interviewer to realize when the offender is giving false or misleading information [27]. The use of a cointerviewer has been suggested both for safety purposes, as well as to provide added objectivity and oversight [83].

Very often, the quality of data obtained during an interview will only be as good as the rapport that the interviewer is able to establish. In studies of interviews of sex offenders, a "humanistic" style of interview has been found to yield the best results [84]. This style requires the interviewer to assume a sympathetic, nonjudgmental stance, as opposed to a "dominant" interrogation style of interview. In a study of interview styles of murderers and sex offenders, it was found that interviews marked by dominance were associated with a higher proportion of denials [85]. In contrast, a humanistic style was associated with greater admissions. Because many serial

murderers will be discerning, intelligent individuals, they will likely perceive even subtle critical or judgmental attitudes in the interviewer. Thus, verbal or nonverbal communication relating bias or disrespect should be avoided. Rather, a humanistic interview style that relies on an understanding and respectful approach is recommended.

While open-ended types of questions are recommended, the interviewer should consider using some type of interview protocol as a general guide for consistency and reliability. The interviewer should not assume that the offender's sexually related crimes are limited to those for which he has been charged or convicted. The interviewer should inquire about concurrent assaultive behavior or deviant fantasies that may have occurred during other, seemingly unrelated criminal acts such as burglary or breaking and entering. Finally, it has been observed that sexual offenders, for a variety of reasons, may be slow to reveal incriminating details [86]. Thus, repeated interviews may be a helpful consideration. Table 5 lists generally recommended strategies for interviewing serial murderers.

When serial murderers are interviewed for research purposes, it has been suggested that interviews should be conducted postconviction, as well as after the possibility of all appeals by the offender are exhausted [47]. In addition, researchers should consider offering some form of written assurance of confidentiality. These precautions have the aim of reducing the offender's motivation for giving deceptive or otherwise self-serving answers.

Prevention

Most experts believe that the prognosis for individuals who have committed serial murder is extremely poor [87, 88]. At the present time, a preventive approach has received the widest endorsement. It has been suggested that more attention be paid to prevention in child and adolescent sex offenders, given that future adult serial murderers may come from this population. Children and adolescents who demonstrate sexually sadistic fantasies, or other early warning signs should be followed closely by mental health professionals who are in a position to direct efforts toward extinguishing the reinforcing cycle, and conducting periodic risk assessments [37].

The field of juvenile sex offender treatment has been gaining increasing attention; however, it remains in its very early stages. Current reviews of treatment efficacy support the use of cognitive behavioral methods, residential treatment programs, supervision, family therapy, and psychosocial interventions [89, 90]. Pharmacotherapy remains understudied, but various psychotropic agents have been utilized, including testosterone lowering agents [91]. In a study of 21 adolescent sex offenders, the use of Naltrexone, a medication used to treat some obsessional and impulsive disorders, was associated with decreased masturbation and sexual fantasies [91]. Much more research in the area of juvenile sex offender treatment is needed to be able to establish a greater degree of confidence in treatment efficacy [92].

Possible Early Warning Signs

In an effort to help guide forensic risk assessments, a list of "10 ominous signs" has been suggested [93]. The list consists of traits, characteristics, and behaviors frequently found in the backgrounds of perpetrators of sexual homicides. It is suggested that when these signs are seen in combination, the juvenile may be predisposed to committing sexual homicides when older. The 10 ominous signs are listed in Table 6.

Table 5 Recommended interviewing strategies

Gather all relevant supplemental data
Thoroughly review data prior to interview
Consider a cointerviewer
Establish good rapport
Attempt to "understand" subject's perspective
Use "humanistic" interview style
Avoid "dominance" or interrogation interview style
Avoid a judgmental or critical approach
Consider repeated interviews

Table 6 Ominous signs [93] (when seen in combination)

Childhood abuse
Inappropriate maternal sexual conduct
Pathological lying and manipulation
Sadistic fantasy with a compulsion to act
Animal cruelty
Need to control and dominate others
Repetitive firesetting
Voyeurism, fetishism and sexual burglary
Unprovoked attacks on females and generalized misogynous emotions
Ritualistic behavior

Conclusions

Serial murderers have been described as representing an exceedingly small, yet "freakish side show in the circus of American punishment" [94]. Despite this very reasonable assertion, serial murderers captivate a disproportionately large amount of society's attention, study, and fears. In the case of serial murderers, we apparently have a substantial desire to know the face of evil. Perhaps this is driven by our fear that these "moral monsters", unbranded by physical stigmata, will be able to commit atrocities undetected. Indeed, a journalist covering the Jeffrey Dahmer trial was unable to fathom how ordinary Dahmer appeared, remarking that "there was nothing to him" [95]. Yet another reason may be that it is the tension between the killer's outer normalcy and inner deviance that elicits our tireless fascination [96]. Whatever the case, serial murderers' offenses represent the extreme pole on the continuum of human cruelty and selfishness.

In this article, some of the biological and psychological theories on serial murder have been discussed. However, it is important to recognize how limited our present understanding remains in terms of the etiology and development of serial murder, so that erroneous conclusions are not drawn. While researchers have identified traits and abnormalities common to serial murderers, there are many who possess these traits and do not go on to become serial murderers. What is it that leads some to act on their deviant fantasies while others do not? Until future research can help clarify this question, we must satisfy ourselves with the notion that "the leap from fantasy to action has much to do with character and the vicissitudes of life . . ." [97]

References

[1] Hickey, E. (1997). *Serial Murderers and Their Victims*, 2nd Edition, Wadsworth Publishing Company, Belmont.

[2] Meloy, J. & Felthous, A. (2004). Introduction to this issue: serial and mass murder, *Behavioural Sciences and the Law* **22**, 289–290.

[3] Myers, W., Husted, D., Safarik, M. & O'Toole, M. (2006). The motivation behind serial sexual homicide: is it sex, power, and control, or anger? *Journal of Forensic Sciences* **51**(4), 900–907.

[4] Douglas, J., Burgess, A.W., Burgess, A.G. & Ressler, R. (1992). *Crime Classification Manual*, Lexington Books, New York.

[5] McNamara, J. & Morton, R. (2004). Frequency of serial sexual homicide victimization in Virginia for a ten-year period, *Journal of Forensic Sciences* **49**(3), 529–533.

[6] Schlesinger, L. (2001). Is serial homicide really increasing? *Journal of American Academy of Psychiatry and the Law* **29**, 294–297.

[7] Geberth, V. & Turco, R. (1997). Antisocial personality disorder sexual sadism, malignant narcissism, and serial murder, *Journal of Forensic Sciences* **42**, 49–60.

[8] Herkov, M. & Beirnat, M. (1997). Assessment of PTSD symptoms in a community exposed to serial murder, *Journal of Clinical Psychology* **53**(8), 809–815.

[9] Holmes, S., Tewksbury, R. & Holmes, R. (2002). Fractured identity syndrome: a new theory of serial murder, in *Current Perspectives on Sex Crimes*, R. Holmes & S. Holmes, eds, Sage Publications, Thousand Oaks.

[10] Dietz, P. (1986). Mass, serial and sensational homicides, *Bullettin of the NewYork Academy of Medicine* **62**(5), 477–491.

[11] Drukteinis, A. (1992). Serial murder: the heart of darkness, *Psychiatric Annals* **22**, 532–538.

[12] Stone, M. (2001). Serial sexual homicide: biological, psychological, and sociological aspects, *Journal of Personality Disorders* **15**(1), 1–18.

[13] Newton, M. (2000). *The Encyclopedia of Serial Killers*, Facts on File, New York.

[14] Everitt, D. (1993). *Human Monsters: an Illustrated Encyclopedia of the World's Most Vicious Murderers*, Contemporary books, Chicago.

[15] Brittain, R. (1970). The sadistic murderer, *Medicine Science and the Law* **10**(4), 198–207.

[16] Ressler, R., Burgess, A. & Douglas, J. (1988). *Sexual Homicide: Patterns and Motives*, D.C. Heath, Lexington.

[17] Meloy, J. (2000). The nature and dynamics of sexual homicide: an integrative review, *Aggression and Violent Behavior* **5**(1), 1–22.

[18] Canter, D., Alison, L., Alison, E. & Wentink, N. (2004). The organized/disorganized typology of serial murder: myth or model? *Psychology, Public Policy and Law* **10**(3), 293–320.

[19] Kocsis, R., Cooksey, R. & Irwin, H. (2002). Psychological profiling of sexual murders: an empirical model, *International Journal of Offender Therapy and Comparative Criminology* **46**(5), 532–554.

[20] Frei, A., Vollm, B., Graf, M. & Volker, D. (2006). Female serial killing: review and case report, *Criminal Behaviour and Mental Health* **16**, 167–176.

[21] Wilson, W. & Hilton, T. (1998). Modus operandi of female serial killers, *Psychological Reports* **82**, 495–498.

[22] Kelleher, M. & Kelleher, C. (1998). *Murder Most Rare: The Female Serial Killer*, Praeger, Westport.

[23] Holmes, R. & Holmes, S. (1998). *Serial Murder*, 2nd Edition, Wadsworth, Belmond.

[24] Arrigo, B. & Griffin, A. (2004). Serial murder and the case of aileen wuornos: attachment theory, psychopathy, and predatory aggression, *Behavioral Sciences and the Law* **22**, 375–393.

[25] Myers, W. (2004). Serial murder by children and adolescents, *Behavioral Sciences and the Law* **22**, 357–374.
[26] Myers, W. (2002). *Juvenile Sexual Homicide*, Academic, London.
[27] Ressler, R., Burgess, A. & Douglas, J. (1983). Rape and rape-murder: one offender and twelve victims, *American Journal of Psychiatry* **140**, 36–40.
[28] Kaplan, R. (2007). The clinicide phenomenon: an exploration of medical murder, *Australasian Psychiatry* **15**(4), 299–304.
[29] Esmail, A. (2005). Physician as serial killer – the shipman case, *New England Journal of Medicine* **352**(18), 1843–1844.
[30] (2002). The shipman inquiry, *First Report, Volume One Death Disguised*, COI Communications, Manchester.
[31] Yorker, B., Kizer, K., Lampe, P., Forrest, A., Lannan, J. & Russell, D. (2006). Serial murder by healthcare professionals, *Journal of Forensic Sciences* **51**(6), 1362–1371.
[32] Beauregard, E., Stone, M., Proulx, J. & Michaud, P. (2007)Sexual murderers of children: developmental, precrime, crime, and postcrime factors,. *International Journal of Offender Therapy and Comparative Criminology* **52**(3), 253–269 .
[33] Safarik, M., Jarvis, J. & Nussbaum, K. (2002). Sexual homicide of elderly females: linking offender characteristics to victim and crime scene attributes, *Journal of Interpersonal Violence* **17**(5), 500–525.
[34] Safarik, M. & Wurtz, D. (2006). Investigation, science, and research partner to solve elderly woman's shocking homicide, *Journal of Forensic Nursing* **127**(3), 127–129.
[35] Beauregard, E. & Proulx, J. (2007). A classification of sexual homicide against men, *International Journal of Offender Therapy and Comparative Criminology* **51**(4), 420–432.
[36] Warren, J., Hazelwood, R. & Dietz, P. (1996). The sexually sadistic serial killer, *Journal of Forensic Sciences* **41**(6), 970–974.
[37] Johnson, B. & Becker, J. (1997). Natural born killers?: the development of the sexually sadistic serial killer, *Journal of American Academy of Psychiatry and the Law* **25**(3), 335–348.
[38] Prentky, R., Burgess, A., Rokous, F., Lee, A., Hartman, C., Ressler, R. & Douglas, J. (1989). The presumptive role of fantasy in serial sexual homicide, *American Journal of Psychiatry* **146**(7), 887–891.
[39] Howitt, D. (2004). What is the role of fantasy in sex offending? *Criminal Behaviour and Mental Health* **14**, 182–188.
[40] Dietz, P., Harry, B. & Hazelwood, R. (1986). Detective magazines: pornography for the sexual sadist? *Journal of Forensic Sciences* **31**(1), 197–211.
[41] Bauer, M., Tatschner, T. & Patzelt, D. (2007). Digital imaging of the dissection and sexual abuse of a corpse – an exceptional case of necrophilia, *Legal Medicine* **9**, 143–146.
[42] Beech, A., Fisher, D. & Ward, T. (2005). Sexual murderers' implicit theories, *Journal of Interpersonal Violence* **20**(11), 1366–1389.
[43] Milsom, J., Beech, A. & Webster, S. (2003). Emotional loneliness in sexual murderers: a qualitative analysis, *Sexual Abuse: A Journal of Research and Treatment* **15**(4), 285–296.
[44] Castle, T. & Hensley, C. (2002). Serial killers with military experience: applying learning theory to serial murder, *International Journal of Offender Therapy and Comparative Criminology* **46**(4), 453–465.
[45] Langevin, R. (1991). The sex killer, in *Rape and Sexual Assault*, A. Burgess, ed, Garland, New York, Vol. 3.
[46] Jentzen, J., Palermo, G., Johnson, L., Ho, K., Stormo, K. & Teggatz, J. (1994). Destructive hostility: the jeffrey dahmer case, *American Journal of Forensic Medicine and Pathology* **15**(4), 283–294.
[47] Kraemer, G., Lord, W. & Heilbrun, K. (2004). Comparing single and serial homicide offenses, *Behavioural Sciences and Law* **22**, 325–343.
[48] Beasley, J. (2004). Serial murder in America: case studies of seven offenders, *Behavioural Sciences and Law* **22**, 395–414.
[49] Brewer, D., Dudek, J., Potterat, J., Muth, S., Roberts, J. & Woodhouse, D. (2006). Extent, trends, and perpetrators of prostitution-related homicide in the United States, *Journal of Forensic Sciences* **51**(5), 1101–1108.
[50] Hill, A., Habermann, N., Berner, W. & Briken, P. (2006). Sexual sadism and sadistic personality disorder in sexual homicide, *Journal of Personality Disorders* **20**(6), 671–684.
[51] Geberth, V. (2003). *Sex-Related Homicide and Death Investigation: Practical and Clinical Perspectives*, CRC Press, New York.
[52] Keppel, R. & Weis, J. (2004). The rarity of unusual dispositions of victim bodies: staging and posing, *Journal of Forensic Sciences* **49**(6), 1308–1312.
[53] Geberth, V. (1996). *Practical Homicide Investigation: Tactics, Procedures and Forensic Techniques*, CRC Press, Boca Raton.
[54] Lundrigan, S. & Canter, D. (2001). Spatial patterns of serial murder: an analysis of disposal site location choice, *Behavioural Sciences and Law* **19**, 595–610.
[55] Langevin, R. (2003). A study of the psychosexual characteristics of sex killers: can we identify them before it is too late? *International Journal of Offender Therapy and Comparative Criminology* **47**(4), 366–382.
[56] Dietz, P., Hazelwood, R. & Warren, J. (1990). The sexually sadistic criminal and his offenses, *Bulletin of the American Academy of Psychiatry and Law* **18**, 163–178.
[57] Porter, S., Woodworth, M., Earle, J., Drugge, J. & Boer, D. (2003). Characteristics of sexual homicides committed by psychopathic and nonpsychopathic offenders, *Law and Human Behaviour* **27**(5), 459–470.
[58] Kernberg, O. (1992). *Severe Personality Disorders*, Yale University Press, New Haven.

[59] Turco, R. (2001). Child serial murder – psychodynamics: closely watched shadows, *Journal of the American Academy of Psychoanalysis* **29**(2), 331–338.
[60] Schlesinger, L. & Revitch, E. (1999). Sexual burglaries and sexual homicide: clinical, forensic, and investigative considerations, *Journal of American Academy of Psychiatry and the Law* **27**(2), 227–238.
[61] White, J. (2007). Evidence of primary, secondary, and collateral paraphilias left at serial murder and sex offender crime scenes, *Journal of Forensic Science* **52**(5), 1194–1201.
[62] Briken, P., Habermann, N., Kafka, M., Wolfgang, B. & Hill, A. (2006). The paraphilia-related disorders: an investigation of the relevance of the concept in sexual murderers, *Journal of Forensic Sciences* **51**(3), 683–688.
[63] Meloy, J., Gacono, C. & Kenney, L. (1994). A rorschach investigation of sexual homicide, *Journal of Personality Assessment* **62**, 58–67.
[64] Gacono, C., Meloy, J. & Bridges, M. (2000). A rorschach comparison of psychopaths, sexual homicide perpetrators, and nonviolent pedophiles: where angels fear to tread, *Journal of Clinical Psychology* **56**, 757–777.
[65] Hucker, S., Langevin, R., Dickey, R., Handy, L., Chambers, J., Wright, S., Bain, J. & Wortzman, G. (1988). Cerebral damage and dysfunction in sexually aggressive men, *Annals of Sex Research* **1**, 33–47.
[66] Gratzer, T. & Bradford, J. (1995). Offender and offense characteristics of sexual sadists: a comparative study, *Journal of Forensic Sciences* **40**(3), 450–455.
[67] Silva, J., Ferrari, M. & Leong, G. (2002). The case of Jeffrey Dahmer: sexual homicide from a neuropsychiatric developmental perspective, *Journal of Forensic Sciences* **47**(6), 1347–1359.
[68] Silva, J., Leong, G. & Ferrari, M. (2004). A neuropsychiatric developmental model of serial homicidal behavior, *Behavioural Sciences and Law* **22**(6), 787–799.
[69] Ressler, R. (2004). How to interview a cannibal, in *Profilers: Leading Investigators Take You Inside the Criminal Mind*, J. Campbell & D. Denevi, eds, Prometheus Books, Amherst, New York.
[70] Schiavi, R., Theilgaard, A., Owen, D. & White, D. (1988). Sex chromosome anomalies, hormones, and sexuality, *Archives of General Psychiatry* **45**, 19–24.
[71] Briken, P., Habermann, N., Wolfgang, B. & Hill, A. (2006). XYY Chromosome abnormality in sexual homicide perpetrators, *American Journal of Medical Genetics Part B (Neuropsychiatric Genetics)* **141B**, 198–200.
[72] Myers, W., Gooch, E. & Meloy, J. (2005). The role of psychopathy and sexuality in a female serial killer, *Journal of Forensic Sciences* **50**(3), 652–657.
[73] Firestone, P., Bradford, J., Greenberg, D. & Larose, M. (1998). Homicidal sex offenders: psychological, phallometric, and diagnostic features, *Journal of American Academy of Psychiatry and Law* **26**(4), 537–552.
[74] Gratzer, T. & Bradford, J. (1995). Offender and offense characteristics of sexual sadists: a comparative study, *Journal of Forensic Sciences* **40**(3), 450–455.
[75] Ressler, R. (2004). Murderers who rape and mutilate, in *Profilers: Leading Investigators Take You Inside the Criminal Mind*, J. Campbell & D. Denevi, eds, Prometheus Books, Amherst, New York.
[76] Ressler, R. (2004). The men who murdered, in *Profilers: Leading Investigators Take You Inside the Criminal Mind*, J. Campbell & D. Denevi, eds, Prometheus Books, Amherst, New York.
[77] Kellert, S. & Felthouse, A. (1985). Childhood cruelty towards animals among criminals and noncriminals, *Human Relations* **38**, 113–129.
[78] Tingle, D., Barnard, G.W., Robbins, L., Newman, G. & Hutchinson, D. (1986). Childhood and adolescent characteristics of pedophiles and rapists, *International Journal of Law and Psychiatry* **9**, 103–116.
[79] Singer, S. & Hensley, C. (2004). Applying social learning theory to childhood and adolescent firesetting: can it lead to serial murder? *International Journal of Offender Therapy and Comparative Criminology* **48**(4), 461–476.
[80] Wright, J. & Hensley, C. (2003). From animal cruelty to serial murder: applying the graduation hypothesis, *International Journal of Offender Therapy and Comparative Criminology* **47**(1), 71–88.
[81] Fox, J. & Levin, J. (1994). *Overkill: Mass Murder and Serial Killing Exposed*, Dell, New York.
[82] Grubin, D. (1994). Editorial: sexual sadism, *Criminal Behavior and Mental Health* **4**, 3–9.
[83] Cooper, G. & King, M. (2006). Interviewing the incarcerated offender convicted of sexually assaulting the elderly, *Journal of Forensic Nursing* **2**(3), 130–133, 146.
[84] Gudjonsson, G. (2006). Sex offenders and confessions: how to overcome their resistance during questioning, *Journal of Clinical Forensic Medicine* **13**, 203–207.
[85] Holmberg, U. & Christianson, S. (2002). Murderers' and sexual offenders' experiences of police interviews and their inclination to admit or deny crimes, *Behavioural Sciences and Law* **20**(1–2), 31–45.
[86] Salter, A. (1988). *Treating Child Sex Offenders and Victims: A Practical Guide*, Sage, London.
[87] Revitch, E. & Schlesinger, L. (1989). *Sex Murder and Sex Aggression*, Charles C. Thomas, Springfield.
[88] Douglas, J. & Olshaker, M. (1995). *Mind Hunter*, Pocket Books, New York.
[89] Garrido Genoves, V., Anyela Morales, L. & Sanchez-Meca, J. (2006). What works for serious juvenile offenders? a systematic review, *Psicothema* **18**(3), 611–619.
[90] Gerardin, P. & Thibaut, F. (2004). Epidemiology and treatment of juvenile sexual offending, *Paediatric Drugs* **6**(2), 79–91.
[91] Ryback, R. (2004). Naltrexone in the treatment of adolescent sexual offenders, *Journal of Clinical Psychiatry* **65**, 982–986.
[92] Reitzel, L. & Carbonell, J. (2006). The effectiveness of sexual offender treatment for juveniles as measured by recidivism: a meta-analysis, *Sex Abuse* **18**(4), 401–421.
[93] Schlesinger, L. (2001). The potential sex murderer: ominous signs, risk assessment, *Journal of Threat Assessment* **1**(1), 47–62.

[94] Kennedy, J. (2006). Facing evil, *Michigan Law Review* **104**, 1287–1304.
[95] Schwartz, A. (1993). *The Man Who Could Not Kill Enough*, Birch Lane, New York.
[96] Schmid, D. (2005). *Natural Born Celebrities: Serial Killers in American Culture*, University of Chicago Press, Chicago.
[97] Dietz, P. (1996). Quoted, in *Bad Men Do What Good Men Dream*, R. Simon, ed, American Psychiatric Press, Washington, DC, p. 312.

Related Articles

Aggression
Assault: Sexually Motivated
Profiles: Psychological and Behavioral
Psychopathy

JAMES L. KNOLL, IV, ROBERT K. RESSLER, ROBERT R. HAZELWOOD AND ANN W. BURGESS

Serial Killer *see* Homicide: Multiple (Behavior)

Serial Murder *see* Serial Homicide

Serial Number Restoration: Firearm

Introduction

Many manufactured objects bear a serial number or other identifying numbers, letters or codes, which may be removed, altered, or obliterated. These markings are produced in various ways during the manufacturing process, such as etching, casting, engraving, or stamping. A range of restoration techniques are available, which makes it possible to restore the original alphanumeric marking in sufficient detail to be able to read it.

Most firearms are marked by their manufacturers with a serial number or a combination of numbers and letters in at least one location. On some firearms, the complete serial number may be repeated on all the major components, i.e., the frame, cylinder, and barrel of a revolver. In other instances, the last few digits of the main alphanumeric mark is repeated upon even relatively small components. These can appear in obvious exposed exterior areas or, in some cases, in locations that can only be revealed by stripping the firearm. It is common practice for criminals to obliterate the identifying marks on illegally owned firearms to make the tracing of them more difficult. Forensic firearms examiners are frequently tasked with restoring these obliterated marks or finding a secondary location where the marking is repeated.

Principles

In many cases, identifying numbers, letters, or marks, are produced using procedures that involve compression such as stamping. This process induces changes in the structure underlying the impressions. If the marking is then removed by filing or grinding so that it is no longer visible, the area of the altered structure remains. By applying etching solutions, with or without heat, the area of the altered structure can be made visible, as that the area will react in a different manner to the base metal.

Close examination of an object is required to verify that a marking should be present in a particular location and that it has been obliterated. The examination should indicate the method by which the numbers, letters, or mark were originally recorded and the method of obliteration. These assessments will dictate the available restoration options. The following are some of the common methods used to place identification marks on firearms.

Etching and Electrical Discharge Machining (EDM)

Etching and electrical discharge machining (EDM) refer to the production of identifying numbers, letters,

or marks by eroding the base material. The process involves discharges of electrical spark to erode conductive materials such as carbon or stainless steel. Little or no deformation of the underlying material occurs with this process and, as a result, completely obliterated serial numbers are almost impossible to restore.

Casting

As with etching and electrical discharge machining, casting involves little or no deformation of the underlying material. Therefore, when the markings are thoroughly obliterated, they are almost impossible to restore, especially if there has been neither 'cold working' nor the addition of any foreign material to the parent metal.

Engraving

There are two general methods of engraving, namely, rotating burrs and vibrating impact tools. Rotating burrs gouge the surface but do not deform the underlying material, whereas vibrating engraving tools have various shaped tips, generally rounded or chisel pointed, and they deform the underlying material while impacting it.

Stamping

Stamping is one of the most common methods of adding serial numbers, letters, or other markings to firearms. It is also known as *punching* and usually involves forcing a shaped tool into the base material. All variations of this method involve compression (cold working) of the base material, which provides the potential for restoration of the original identifying marks.

Obliteration

Different techniques are used in the attempt to obliterate numbers, letters, or marks. The most common methods of obliteration are described below.

Filing/Grinding

Mill Bastard or Bastard (1/2 round) metal files and grinding wheels are the tools most commonly used for this purpose. The metal surface is generally removed until all visible traces of the serial number, letters, or marks are removed. Restoration using acid etching or magnetic particle processes have a high success rate.

Center Punching or Over-Stamping

A line of holes is center punched into the serial number, letters, or marks using a hammer and punch. If deep enough, these alter the underlying structure of the metal and destructive restoration techniques are unlikely to be successful.

Other number or letter stamps are sometimes used to over-stamp existing ones in an attempt to alter the original digits. Nondestructive visual examination methods can be useful in identifying the remaining key areas of the original marking.

Milling or Drilling

Milling or drilling has a similar effect on the underlying structure of metal as filing or grinding, but this often removes more material. Visual examination can be useful on drilled areas to identify any remaining key areas of the original digits, but if the milling or drilling is deep enough, destructive techniques are unlikely to be successful.

Welding

Welding essentially remelts part of the base material, adding new material from the filler rod or wire. This produces gross changes in the underlying material, and acid etching techniques are generally unsuccessful in identifying the original markings.

Nondestructive Restoration Techniques

Visual Inspection

In many instances, when a drill or a punch has been used to obliterate the original serial number, it is possible to identify individual digits from the remaining fragments that are still visible, provided these fragments contain a key portion of the digit. A low-powered stereo microscope is generally required to aid recognition of the fragments; contrast techniques can also be useful. The most common are

rubbing chalk over the erased area or using a filtered light source (e.g., a Polilight lamp) to create ultraviolet or other light frequencies that may enhance the marking.

Wood

Some firearms have identification marks in their wooden stocks or grips. If the mark is burnt in or made by cutting with a chisel or a knife, it cannot be restored once obliterated. However, if the mark has been punched in, for example, the serial number or military marks on a Lee Enfield caliber. 303 British rifle, then the same considerations apply as with metals. The simplest and most effective method of restoration is to run a jet of steam onto the erased surface. The steam softens the wood and causes the bent fibers to spring back and the broken fibers to swell. As a result, the wood will project the marking above the surface and it can be read.

Destructive Restoration Techniques

On occasions an examiner can polish the defaced area flat with wet and dry abrasive paper to find the original markings are still visible under the correct lighting conditions, without the need for further treatment. If further treatment is required, then it is important to determine the metal type before commencing treatment.

Magnetic Particle

The concept of magnetic particle restoration was developed in 1957 in the United States of America. It relies on the principle that any change in density or of molecular structure causes a change of the magnetic flow lines from the north to the south poles. Therefore the magnetic particles tend to congregate around the obliterated number, letter, or marking, making it visible.

This process requires the surface area to be smooth and polished to achieve the best results. Once prepared, the object should be placed on an AC or DC powered magnet. Industrial magnetic crack testing devices have proved particularly successful using AC or direct DC power sources. Once the object has been charged, Ardrox white background lacquer can be applied to provide contrast and then Ardrox black magnetic ink can be sprayed over the treated area. The restoration usually appears almost immediately.

Acid Etching (Stainless Steel, Mild Steel, and Iron)

Before using this technique, it is important to identify what type of metal is to be worked on. This will determine the type of chemical solution to be used.

When acid etching stainless steel, mild steel or iron, the surface area must be cleaned with acetone to ensure it is free from oil or grease. The area to be treated is then polished with a fine grade of wet and dry abrasive paper (400 Grade) to make restoration easier to see. Badly scratched or scored areas should be lightly polished with a fine abrasive pad on the dermal tool. This simply reduces the amount of labor required to achieve a smooth surface.

There is debate over whether the use of heat is a necessary step in the restoration process. The principle surrounding the use of heat is that it will anneal the metal if a temperature of approximately 300 °C is reached. An indicator of this temperature is when the metal assumes a straw color. It is essential that the metal is not heated too strongly. The reason for annealing is to enable the compressed and distorted metal to adjust itself to the conditions. If the metal is heated to red heat, the temperature is sufficiently high to soften the metal and on cooling, the metal becomes homogeneous and can no longer be differentiated. A Bunsen burner or blowlamp is all that is needed to achieve the desired result.

Once the metal has been allowed to cool, Fry's reagent, which consists of a mixture of hydrochloric acid, cupric chloride, ethanol, and water, can be used to acid etch the metal.

Modified Fry's reagent can also be used; this mixture consists of the same ingredients except ethanol, but in a different proportion.

At the completion of the etching process, it is recommended that a lacquer is used to seal the restored numbers, letters, or marks. Alternatively, light oil should be applied to the treated area to protect it from corrosion.

Acid Etching (Aluminum and Aluminum Alloys)

The most common and effective solution is a 10% sodium hydroxide and water. Preparation of the

surface metal is the same as that described for Fry's reagent. Fry's solution can still be used on aluminum or aluminum alloys, but it reacts very quickly. Therefore it should only be used when a result cannot be achieved with 10% sodium hydroxide.

Hume-Rothery solution is an alternative to modified Fry's reagent; it contains cupric chloride, hydrochloric acid, and water in different proportions. This reagent reacts very energetically with aluminum alloys and deposits a layer of loose brown copper, which must be wiped off.

The procedure for using Hume-Rothery solution is basically the same as for mild steel, stainless steel, and iron, except that the aluminum surface should not be cleaned with abrasives. Instead, any paint or grease can be removed with acetone before commencing the acid etching process.

Vinella's solution is also an effective reagent on aluminum or aluminum alloys. It contains glycerine, hydrofluoric acid and nitric acid. However, hydrofluoric acid is extremely corrosive and a dangerous substance to work with. Therefore this reagent is not recommended for Occupational Health Safety and Wellbeing (OHS&W) reasons.

Heat

The use of heat has already been mentioned in combination with acid etching techniques on stainless steel, mild steel, and iron. It can also be used by itself on these metals after the surface area has been smoothed and polished. The process is commonly used on engine blocks to restore engine numbers, but greater care needs to be exercised on gun metals because they are far more fragile.

Further Reading

ATF National Tracing Center (1999). *ATF Guide to Illegal Firearms Trafficking Investigations, Section 2, Serial Number Removal Definitions and Codes*, ATF National Tracing Center Publication, pp. 11–19.

Barabash, T. & Fahey, R.T. (1977). Non-destructive methods of restoring defaced serial numbers, *AFTE Journal* **9**(1), 23.

Brown, E.W. (2001). Serial number restoration on ruger p series aluminum alloy frames, *AFTE Journal* **33**(1), Winter, 57.

Chisum, W.J. (1966). A catalytic process for restoration of serial numbers in aluminum, *Journal of the Forensic Science Society* **6**, 89.

Collins, J.M. (1999). Modern marking and serial numbering methods, *AFTE Journal* **31**(3), Summer, 309.

Cook, C.W. (1989). Obliterated serial numbers, *AFTE Journal* **21**(2), 166.

Dragan, P. (1996). Abrasive wheels for serial number restoration preparation, *AFTE Journal* **28**(1), 21.

Du Preez, D.T. (1997). *Manual to Ballistic and Related Identification Methods*, UNISA, Pretoria, pp. 204–210, 216–224.

Finor, J.M. & Rone, C. (1998). Serial number restoration of obliterated welded on type characters, *AFTE Journal* **30**(4), Fall, 649.

Heard, B.J., (1997). *Handbook of Firearms and Ballistics: Examining and Interpreting Forensic Evidence*, John Wiley & Sons, West Sussex, pp. 213–221.

Heflin, T.M. (1984). Overstamp, *AFTE Journal* **16**(3), 12.

Kennington, R.H. (1997). One swab serial number restoration, *AFTE Journal* **29**(3), Summer, 288.

Klees, G.S. (2002). The restoration of obliterated laser-etched firearm identifiers by conventional and alternative decryption methods, *AFTE Journal* **34**(3), Summer, 264.

Knowles, M. (1985). Instant recovery of obliterated serial numbers, *AFTE Journal* **17**(3), 63.

Massiah, E.E. (1976). Techniques and formula, *AFTE Journal* **8**(2), 26.

Miller, K.E. (1972). Current assist for die stamp impression restoration, *AFTE Journal* **4**(3), 38.

Nickolls, L.C. (1956). *The Scientific Investigation of Crime, Erased Identification Marks*, Butterworth & Co, London, 150–164.

O'Reilly, W.E. (1970). Magnetic restoration of serial number, *AFTE Newsletter* **2**(3), 26.

Polk, D.E. & Giessen, B.C. (1989). Metallurgical aspects of serial number recovery, *AFTE Journal* **21**(2), 174.

Polk, D.E. & Giessen, B.C. (1975). Metallurgical aspects, *AFTE Journal* **17**(2), 38.

Sherlock, W.E. & Keating, D.M. (1995). Obliterated serial number tracking program, *AFTE Journal* **27**(4), 264.

Shoshani, E. & Klain, A. (2001). Altering a serial number, *AFTE Journal* **33**(2), Spring, 133.

Thornton, J.I. & Cashman, P.J. (1976). The mechanism of the restoration of obliterated serial numbers by acid etching, *Journal of the Forensic Science Society* **16**(1), 69–71.

Treptow, R.S. (1978). *Handbook of Methods for the Restoration of Obliterated Serial Numbers*, NASA.

Wagoner, A. (1999). Griffin's reagent for serial number restoration in stainless steel, *AFTE Journal* **31**(4), Fall, 497.

Related Articles

Firearms: Overview

Toolmarks

NICHOLAS R. MAIDEN

Sex Determination of Remains

Once one has verified that the remains are human and recent, the anthropologist must proceed to reconstruct the biological profile. This entails the collection of all possible information from the skeletal remains, which may lead to identification. This goes from sex, age, stature, and ancestry, to all dental (hence the strong link with forensic odontology) and osseous features (deformations, pathologies, nonmetric traits), and, finally, to facial reconstruction. In an ideal system, the biological profile should be compared with descriptions of missing persons in order to find the best possible matches. Then one should proceed to perform a comparison between antemortem and postmortem data by odontological, anthropological, or genetic analyses (sometimes fingerprints may also exist on the degraded remains, which are not completely skeletonized, but this is more the domain of the fingerprint expert or the forensic pathologist). Concerning the biological profile (sexing, aging, etc.), odontological aspects also will be taken into account in this Forensic Anthropology section (*see* **Anthropology: Age Determination of Remains**; **Odontology**), leaving the specific identification aspects mainly to the Forensic Odontology chapter.

Sexing is the first step in the completion of the biological profile. Determination of sex of human remains is frequently immediate even in severely decomposed cadavers, thanks to the fact that genital organs are among the last soft tissues to disappear. In skeletonized remains, however, such tissues are not present and thus one must rely on the dental and skeletal structure. In the past decade, DNA testing has started to prevail. However, although it is extremely sensitive as a technique, one must deal with the problem of extraction, degradation, contamination, and costs. Conservation of DNA depends on various environmental variables, still not completely known, even if sporadic DNA has been used for sexing even ancient remains. Anthropological criteria are quicker and cheaper for sexing; however, they too have their limits. On one hand, they are dependant on the integrity of the skeleton. Extremely fragmented remains may be impossible to sex (if crucial parts have been lost). On the other hand, they do not always give accurate and absolute results. Another limit to the skeletal analysis is subadults. Skeletons of children and adolescents cannot be sexed due to the incompleteness of morphological characteristics. Some authors have advanced the hypothesis concerning different dental dimensions and particularly the conformation of the auricular surface of the ilium, which, in infant ilia, is elevated when the sex is female and flat when it is male. However, extreme caution should be applied with such methods whose authors indeed warrant an error rate of 50–60%. Other hints on sex of a subadult can be given by the difference in maturation between the dentition and the skeleton. Girls usually have similar maturation stages for the skeleton and the teeth; boys seem to have a slower dental development. Thus a discrepancy between dental and skeletal development in a subadult may indicate a male sex. However, even this method may be full of pitfalls.

Determination of sex on adult complete skeletons, on the other hand, is quite reliable and has two main approaches: a morphological and a metric one. Both approaches are complementary and reflect sexual dimorphism in the shape and dimension in almost all bones, but particularly and more significantly, the cranium and the skull. Metric parameters reflect, on the other hand, different sizes between males and females, in particular, as concerns the diameter of articular extremities. There is, however, an intermediate zone of superimposition and only individuals whose dimensions fall below and above it can be safely classified as male or female [1–43].

The most reliable anatomical site for sexing is the hip. Female bones reflect the tendency to adapt to pregnancy and childbirth. The female pelvis is relatively lower and larger compared to the male one. Generally speaking, in front of an entire pelvis (left and right innominate and sacrum) this appearance can be quite diagnostic, as shown in Figures 1 and 2.

It is, however, less intuitive that there are other parts of the pelvis that may be, per se, indicative of sex. The most important bones remain the pubic ones, particularly that area constituting Phenice's triad, a series of three characters, which are very reliable sex indicators. These consist of (i) the ventral arc: this is an elevated bony crest, which extends across the ventral surface of the pubic body laterally from the center of the body until it reaches the ischiopubic ramus (Figures 3 and 4).(ii) the subpubic concavity, which is a small medial concavity at the medial extremity of the subpubic angle, present in females;



Figure 1 Female pelvis. The letter A is situated in the subpubic angle, which is very wide in the female



Figure 2 Male pelvis. The letter B is under a very narrow subpubic angle, typical of males. Also notice the much more narrow nature of the anterior pubic region with respect to the pelvis in Figure 1

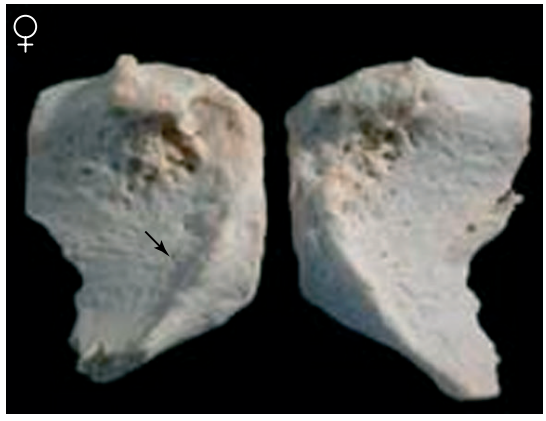


Figure 3 Pubis seen from the front. The arrow points to the ventral arc in the female pubis

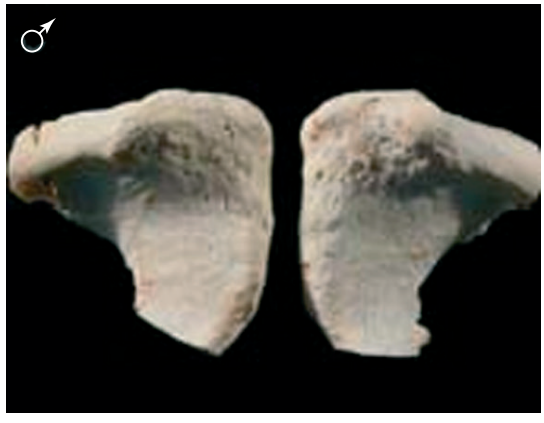


Figure 4 Isolated typical male pubic bone with no ventral arc

(iii) the shape of the ischiopubic ramus (or presence of the ischiopubic crest). The female ischiopubic ramus has a crest along the ramus and is narrower. The value of these parameters has been confirmed by several authors and this triad remains a fundamental sex indicator.

Other important sex markers of the pelvis are the shape of the sciatic notch: it is narrow in males and forms an obtuse angle in females (Figures 5 and 6).

The preauricular sulcus is an evident concavity (notch) beneath the auricular surface, which articulates the ilium with the sacrum. This is more commonly present in females than in males. Most authors will agree that the entire pelvis will reach a degree of accuracy in sexing between 95 and 98%. However, as all biological markers, a margin of overlapping of the two sexes must be taken into account.

Finally, the sacrum is also usually taken into account. A flat sacrum is generally female, while a concave one is male.

The cranium is the next best sex indicator, although reliability slides down to 80–85%.

The male cranium is, in general, larger and rougher with more marked muscular insertions.

Males tend to develop prominent crests near the muscle insertions, which move the head and mandible. These include the temporal lines, where temporal muscles insert, and the mastoid processes, which project inferiorly behind the auditory meatus.

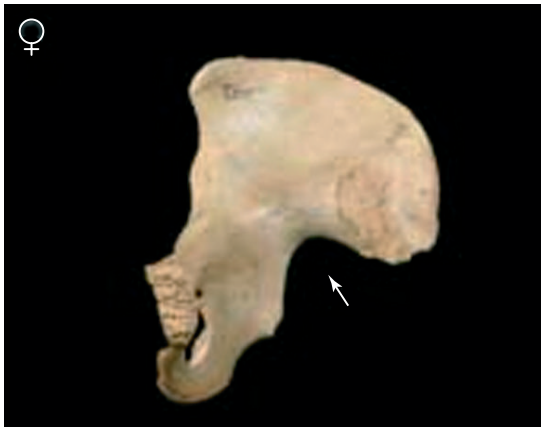


Figure 5 Female right innominate bone (medial view): the arrow indicates a large sciatic notch

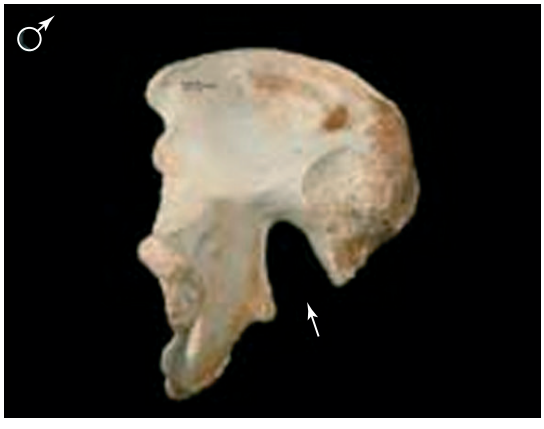


Figure 6 Male right innominate bone (medial view): the sciatic notch is much narrower



Figure 7 Female skull, where one can appreciate (1) flat glabella; (2) thin orbital margin; (3) a zygomatic arch that stops in length at the external auditory meatus; (4) small and medially inclined mastoid process; (5) (posterior) small nuchal crest; (6) no gonial eversion and delicate mandibular angle; and (7) round mandibular symphysis

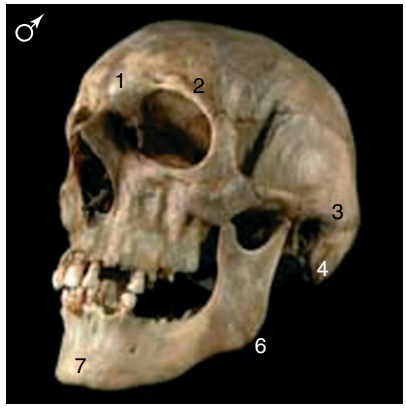


Figure 8 Male skull, where one can appreciate (1) pronounced glabella; (2) thick and rounded orbital margin; (3) a zygomatic arch that proceeds beyond the external auditory meatus; (4) large and vertical mastoid process; (5) (posterior) large nuchal crest; (6) gonial eversion and rough ridges on the mandibular angle; and (7) squared mandibular symphysis

However, such insertions may be marked even in women who have occupational stress in these areas. Ethnological studies show the development of such characters in women who prepare hides with their teeth or who carry weights on their heads.

In general, main sexual differences in the cranium can be seen in the anatomical areas shown in Figures 7 and 8.

As regards the anthropometric approach, numerous publications exist on metrical differences between sexes of almost every single bone. One can observe the diameters of the heads of humeri and femurs, and/or measure the ratios of pelvic structures, such as the ischiopubic index. An alternative is the discriminant function analysis. However, the single most commonly used parameters are the head of the femur, of the humerus and of the radius. Generally, if the diameter of a humeral head is greater than 47 mm, the person is a male and less than 43 mm it is a female; a female radial head is 21 mm or less and in males it is usually over 23 mm. The vertical diameter

of the femoral head is usually over 48 mm in males and less than 43 mm in females. There is then a vast literature on the metric analysis of various anatomical districts such as heads, diaphyses, and epicondyles of long bones and lengths and widths of tali, calcanei, sacrum, metatarsals, and metacarpals [25–43]. One advantage of metrical analysis is that it will entail a smaller error within nonexpert observers. Discriminant function analysis is also a valid, although more complex, support to sexing. It allows the combination of numerous measures in a single mathematical function in order to discriminate between two groups: male and female, based obviously on data taken from two known sample groups. With the combination of different measurements, discriminant functions give more weight to variables, which are more efficient in sexing and in distinguishing between two groups, namely, male and female. Discriminant functions have already been calculated and published for cranial, mandibular, pelvic, and other regions of the skeleton.

All these methods, however, both metric and morphological, are subject to variations within populations. It is therefore crucial, when sexing, to use morphological or metric parameters or discriminant functions tared on populations similar to the one that is being studied.

A final remark on sex should be made concerning signs of pregnancy. Some authors have indicated that notches on the dorsal surface of the pubis indicate that the woman underwent pregnancy. More marked ones even indicate several pregnancies. These theories, however, should not be taken too seriously as they have, in part, been revealed wrong by the presence of such notches in the bones of women known to have been nulliparous.

References

[1] Asala, S.A., Bidmos, M.A. & Dayal, M.R. (2004). Discriminant function sexing of fragmentary femur of South African blacks, *Forensic Science International* **145**(1), 25–29.

[2] Brown, R.P., Ubelaker, D.H. & Schanfield, M.S. (2007). Evaluation of Purkait's triangle method for determining sexual dimorphism, *Journal of Forensic Sciences* **52**(3), 553–556.

[3] Bidmos, M.A. & Dayal, M.R. (2004). Further evidence to show population specificity of discriminant function equations for sex determination using the talus of South African Blacks, *Journal of Forensic Sciences* **49**(6), 1165–1170.

[4] Case, D.T. & Ross, A.H. (2007). Sex determination from hand and foot bone lengths, *Journal of Forensic Sciences* **52**(2), 264–270.

[5] Dayal, M.R. & Bidmos, M.A. (2005). Discriminating sex in South African blacks using patella dimensions, *Journal of Forensic Sciences* **50**(6), 1294–1297.

[6] Dar, G. & Hershkovitz, I. (2006). Sacroiliac joint bridging: simple and reliable criteria for sexing the skeleton, *Journal of Forensic Sciences* **51**(3), 480–483.

[7] Patriquin, M.L., Steyn, M. & Loth, S.R. (2005). Metric analysis of sex differences in South African black and white pelves, *Forensic Science International* **147**(2–3), 119–127.

[8] Ubelaker, D.H. & Volk, C.G. (2002). A test of the Phenice method for the estimation of sex, *Journal of Forensic Sciences* **47**(1), 19–24.

[9] Asala, S.A. (2001). Sex determination from the head of the femur of South African whites and blacks, *Forensic Science International* **117**(1–2), 15–22.

[10] Franklin, D., Oxnard, C.E., O'Higgins, P. & Dadour, I. (2007). Sexual dimorphism in the subadult mandible: quantification using geometric morphometrics, *Journal of Forensic Sciences* **52**(1), 6–10.

[11] Gentry Steele, D. (1979). The estimation of sex on the basis of the talus and calcaneus, *American Journal of Physical Anthropology* **45**(3 pt 2), 581–588.

[12] Giles, E. & Elliot, O. (1963). Sex determination by discriminant function analysis of crania, *American Journal of Physical Anthropology* **21**, 53–68.

[13] Gualdi-Russo, E. (2006). Sex determination from the talus and calcaneus measurements, *Forensic Science International* **171**(2–3), 151–156.

[14] Holcomb, S.M.C. & Konigsberg, L.W. (1995). Statistical study of sexual dimorphism in the human fetal sciatic notch, *American Journal of Physical Anthropology* **97**(2), 113–125.

[15] Hu, K.S., Koh, K.S., Han, S.H., Shin, K.J. & Kim, H.J. (2006). Sex determination using nonmetric characteristics of the mandible in Koreans, *Journal of Forensic Sciences* **51**(6), 1376–1382.

[16] Kalmey, J.K. & Rathbun, T.A. (1996). Sex determination by discriminant function analysis of the petrous portion of the temporal bone, *Journal of Forensic Sciences* **41**(5), 865–867.

[17] Introna, F., Di Vella, G. & Campobasso, C.P. (1998). Sex determination by discriminant analysis of patella measurements, *Forensic Science International* **95**(1), 39–45.

[18] Iscan, M.Y., Loth, S.R., King, C.A., Shihai, D. & Yoshino, M. (1998). Sexual dimorphism in the humerus: a comparative analysis of Chinese, Japanese and Thais, *Forensic Science International* **98**(1–2), 17–29.

[19] Kemkes, A. & Göbel, T. (2006). Metric assessment of the "mastoid triangle" for sex determination: a validation study, *Journal of Forensic Sciences* **51**(5), 985–989.

[20] Kim, D.I., Lee, U.Y., Park, D.K., Kim, Y.S., Han, K.H., Kim, K.H. & Han, S.H. (2006). Morphometrics of the hyoid bone for human sex determination from

digital photographs, *Journal of Forensic Sciences* **51**(5), 979–984.
[21] King, C.A., Iscan, M.Y. & Loth, S.R. (1998). Metric and comparative analysis of sexual dimorphism in the Thai femur, *Journal of Forensic Sciences* **43**(5), 954–958.
[22] Listi, G.A. & Bassett, H.E. (2006). Test of an alternative method for determining sex from the os coxae: applications for modern Americans, *Journal of Forensic Sciences* **51**(2), 248–252.
[23] McCormick, W.F., Stewart, J.H. & Greene, H. (1991). Sexing of human clavicles using length and circumference measurements, *The American Journal of Forensic Medicine and Pathology* **12**(2), 175–181.
[24] Mac Laughlin, S.M. & Bruce, M.F. (1990). The accuracy of sex identification in European skeletal remains using the Phenice characters, *Journal of Forensic Sciences* **35**(6), 1384–1392.
[25] Mahfouz, M., Badawi, A., Merkl, B., Fatah, E.E., Pritchard, E., Kesler, K., Moore, M., Jantz, R. & Jantz, L. (2007). Patella sex determination by 3D statistical shape models and nonlinear classifiers, *Forensic Science International* **173**(2–3), 161–170.
[26] Mall, G., Graw, M., Gehring, K. & Hubig, M. (2000). Determination of sex from femora, *Forensic Science International* **113**(1–3), 315–321.
[27] Mall, G., Hubig, M., Kuznik, J., Penning, R. & Graw, M. (2001). Sex determination and estimation of stature from long bones of the arm, *Forensic Science International* **117**(1–2), 23–30.
[28] Marino, E.A. (1995). Sex estimation using the first cervical vertebra, *American Journal of Physical Anthropology* **97**(2), 127–133.
[29] Phenice, T.W. (1969). A newly developed visual method of sexing the os pubis, *American Journal of Physical Anthropology* **30**(2), 297–201.
[30] Purkait, R. (2001). Measurements of ulna – a new method for determination of sex, *Journal of Forensic Sciences* **46**(4), 924–927.
[31] Robling, A.G. & Ubelaker, D.H. (1997). Sex estimation from the metatarsal, *Journal of Forensic Sciences* **42**(6), 1062–1069.
[32] Rogers, T.L. (1999). A visual method of determining the sex of skeletal remains using the distal humerus, *Journal of Forensic Sciences* **44**(1), 57–60.
[33] Rogers, N.L., Flournoy, L.E. & McCormick, W.F. (2000). The rhomboid fossa of the clavicle as a sex and age estimator, *Journal of Forensic Sciences* **45**(1), 61–67.
[34] Rogers, T.L. (2005). Determining the sex of human remains through cranial morphology, *Journal of Forensic Sciences* **50**(3), 493–500.
[35] Steyn, M. & Iscan, M.Y. (1999). Osteometric variation in the humerus: sexual dimorphism in South Africans, *Forensic Science International* **106**(2), 77–85.
[36] Steyn, M. & Iscan, M.Y. (1998). Sexual dimorphism in the crania and mandibles of South Africans whites, *Forensic Science International* **98**(1–2), 9–16.
[37] Sutherland, L.D. & Suchey, J.M. (1991). Uses of the ventral arc in pubic sex determination, *Journal of Forensic Sciences* **36**(2), 501–511.
[38] Tanaka, H., Lestrel, P.E., Uetake, T., Kato, S. & Ohtsuki, F. (2000). Sex differences in proximal humeral outline shape: elliptical Fourier functions, *Journal of Forensic Sciences* **45**(2), 292–302.
[39] Ukelaker, D.H. & Volk, C.G. (2002). A test of the Phenice method for the estimation of sex, *Journal of Forensic Sciences* **47**(1), 19–24.
[40] Wahl, J. & Graw, M. (2001). Metric sex differentiation of the pars petrosa ossis temporalis, *International Journal of Legal Medicine* **114**(4–5), 215–223.
[41] Wescott, D.J. (2000). Sex variation in the second cervical vertebra, *Journal of Forensic Sciences* **45**(2), 462–466.
[42] Williams, B.A. & Rogers, T. (2006). Evaluating the accuracy and precision of cranial morphological traits for sex determination, *Journal of Forensic Sciences* **51**(4), 729–735.
[43] Wiredu, E.K., Kumoji, R., Seshadri, R. & Biritwum, R.B. (1999). Osteometric analysis of sexual dimorphism in the sternal end of the rib in a West African population, *Journal of Forensic Sciences* **44**(5), 921–925.

Related Articles

Anthropology

DNA: Degraded Samples

CRISTINA CATTANEO AND DAVIDE PORTA

Sex of Deceased *see* Species Determination of Osseous Remains

Sex Offenders: Treatment of

The term *paraphilia*, as defined by the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV), is used to refer to recurrent sexual fantasies, urges, or behaviors involving nonhuman objects, suffering or humiliation, children or

other nonconsenting persons. Paraphilic interests are fairly common in men and include a wide variety of behaviors typically seen in clinical populations seeking evaluation and treatment for paraphilias [1]. Those who most frequently seek psychiatric treatment for sex offenses are those accused of or involved in child molestation, voyeurism, exhibitionism, fetishism, frottage, and public masturbation [2]. The term *sex offender* is used to refer to an individual who has been legally convicted of a sex offense.

The terminology "sex offender" does not provide any information regarding the offender's reason for committing the sexual offense nor does it indicate the presence or absence of a paraphilia. The psychiatric treatment of sex offenders is distinct from other clinical populations. Recent changes in sex offender legislation have mandated treatment for offenders deemed "sexually dangerous" or "sexually violent" in psychiatric hospitals as civil committees. Such legislation has fueled long-standing debates on the diagnosis of paraphilias, the nature of mental illness, and the treatability of sex offenders [3].

Background

The efficacy of sex offender treatment is unknown. Although studies comparing treated and untreated sex offenders have been done, measurement of outcome is flawed, with recidivism rates underestimating the actual recurrence of the pathological behavior [4]. Methodological challenges such as sample selection, study design, and assessment of outcome represent significant problems in the treatment efficacy studies. The majority of treatment efficacy studies are conducted on incarcerated or civilly committed offenders. Those individuals who are incarcerated or civilly committed represent a unique population with treatment outcomes, which may not generalize to the nonincarcerated or community-residing sex offender. Many studies do not differentiate between types of offenders. For example, a study that evaluates child molesters may not differentiate between intrafamilial and extrafamilial offenders. Research suggests that incest offenders recidivate at approximately half the rate of extrafamilial child molesters [5].

A well-designed study of treatment efficacy in sex offenders would compare treated offenders with a matched, randomly assigned control group. However, the ethical and public safety implications of such a study impede its design. Sex offenders are extremely heterogeneous and when this is not recognized it is very difficult to make comparisons among treatment efficacy studies [2].

The determination of treatment success in sex offenders is arbitrarily defined. Investigators have not agreed upon a standardized measurement of improvement. Although all treatment studies report sex offender recidivism, recidivism is measured in various ways. Some studies measure recidivism based on an offender's self-report, whereas other studies define recidivism based on an offender's arrest records. Studies measuring sexual arousal by penile plethysmography define improvement as a reduction in deviant sexual arousal. In a meta-analysis by Hanson and Bussiere [6], erections to children (as measured by penile plethysmography) was the factor most highly correlated to recidivism in over 28 000 sex offenders. However, the use of penile plethysmography has not been generally accepted by the scientific community due to a lack of reliability between the different types and models of plethysmographs in addition to debate over the validity and appropriate use of penile plethysmography [7].

Biological Treatment

The scientific basis of biological treatment in sex offenders is the reduction of sexual behaviors by decreasing testosterone levels. Testosterone is a steroid hormone produced by the testes and responsible for the development of secondary sex characteristics in men. Testosterone helps maintain sex drive, the production of sperm cells, male hair patterns, muscle mass, and bone mass. Testosterone strongly influences both male and female sexual drive and the resultant sexual behavior [8]. Animal studies show that male copulatory behavior is almost entirely dependent on circulating levels of plasma testosterone [9].

The production of testosterone is mediated by follicle-stimulating hormone (FSH) and luteinizing hormone (LH), both of which are produced by the pituitary gland. LH signals the testes to produce testosterone. The hypothalamus senses whether the production of testosterone is too much or too little. If the testes are producing too little testosterone, the hypothalamus secrets a hormone called

gonadotropin-releasing hormone (GnRH) that signals the pituitary to increase production of LH, which results in an increase in testosterone production. If the testes are producing too much testosterone, the hypothalamus secretion of GnRH is inhibited. This process whereby the hypothalamus modulates the production of testosterone is called *negative inhibition*.

Surgical Treatment

Surgical treatment for sex offenders consists of two types: neurosurgery and castration. The neurosurgical procedure involves stereotaxic removal of parts of the hypothalamus to disrupt production of male hormones and decrease sexual arousal and impulsive behaviors [4]. This procedure had significant adverse effects and was considered largely ineffective. Surgical castration is the removal of the testes. The effect of surgical castration is to globally reduce available androgen by the removal of the testes [2]. Sturup [10] followed 107 castrated sex offenders and compared them with 58 sex offenders who were not castrated over a period of 18 years. The castrated individuals recidivated at a 4.3% rate. The 58 uncastrated individuals recidivated at a 43% rate. Postcastration follow-up studies provide the most comprehensive outcome data on the effect of reducing plasma testosterone and the resultant suppression of the sexual drive and deviant sexual behavior [11]. These studies have reported recidivism rates of less than 5% with follow-up periods of up to 20 years in a large sample size [11].

The introduction of antiandrogen and hormonal medications has rendered surgical castration nearly obsolete. The effects of surgical castration can be achieved through chemical castration, i.e., the use of medications to decrease testosterone production, without the invasiveness and irreversibility of surgery. Some states in the United States, such as Texas and California, mandate chemical or surgical treatment of dangerous sexual offenders. Controversy exists as to whether surgical castration should be offered as a treatment as chemical castration achieves the same results and spares the procedure.

Pharmacologic Treatment

Antiandrogens. Androgens are male sex hormones that promote the development and maintenance of male sex characteristics. Testosterone is the most common androgen in man. Antiandrogen treatment refers to treatment with drugs used to block production or interfere with the action of male sex hormones. Cyproterone acetate (CPA) and medroxyprogesterone (MPA) are the two most commonly used antiandrogen medications. CPA is not available in the United States but is widely available in Canada and Europe. Both CPA and MPA are synthetic progesterones that reduce the serum level of testosterone. Reduction of testosterone has been shown to reduce libido, erections, ejaculations, and spermatogenesis [12]. A meta-analysis of antiandrogen studies completed by Grossman *et al.* [4] suggests that recidivism rates in antiandrogen-treated offenders are less than that in untreated offenders. The studies report a spectrum of differences between patients treated with antiandrogens and those not treated, with recidivism rates as low as 1% for treated patients and as high as 68% for untreated patients [4]. In 1992, Copper *et al.* [13] performed the first direct comparison of CPA and MPA. The results suggested that MPA and CPA performed equally in decreasing sexual thoughts and fantasies, frequency of masturbation, and erection.

The risks associated with the use of antiandrogen agents include side effects such as weight gain, hyperglycemia, hot and cold flashes, liver dysfunction, hypertension, muscle cramps, phlebitis, gastrointestinal complaints, and feminization [4]. In addition there is little known about the long-term sequelae of antiandrogen treatment. The majority of antiandrogen studies follow patients for eight years or less. The patient who agrees to antiandrogen treatment should be educated about the absence of long-term data.

Hormonal Agents. Leuprolide and Triptorelin are hormonal agents referred to as *long-acting GnRH agonists*. These agents inhibit the secretion of LH with a resulting decrease in plasma testosterone levels and libido [14]. They produce a chemical castration in that the hypothalamic-pituitary axis is exhausted and there is a potent inhibition of gonadotropin [15]. Rosler and Witztum treated 30 men with severe long-standing paraphilias with triptorelin [16]. Treatment was associated with suppression of serum testosterone. A meta-analysis of pharmacotherapy of paraphilias with long-acting agonists of luteinizing hormone-releasing hormone was

conducted by Briken *et al.* [17]. In total, the studies reported on a sample of 118 treated patients. Patients previously treated with other agents like CPA, MPA, or selective-serotonin-reuptake inhibitors (SSRIs) reported better effects when taking luteinizing hormone-releasing hormone (LHRH) agonists. None of the treated patients had a relapse. The results indicated that all men showed a decrease in deviant sexual fantasies, desires, and abnormal sexual behavior. In an observational study, Krueger and Kaplan [18] treated 12 patients with various diagnosed paraphilias and comorbid psychiatric disorders with leuprolide. All patients reported a significant reduction or even cessation of deviant sexual arousal and/or interests.

The side effects associated with GnRH agonists include decreased bone mineral density or osteopenia, weight gain, hyperglycemia, diabetes, hypertension, and insomnia. The most commonly reported side effects are erectile/ejaculatory problems and gynecomastia [18]. In general, patients reported fewer side effects on GnRH agonists when compared to antiandrogens. The preliminary studies of GnRH agonist suggest that leuprolide and triptorelin may be more effective, better tolerated alternatives to antiandrogen treatment.

Selective-Serotonin-Reuptake Inhibitors (SSRIs). Although the pharmacologic mechanism is poorly understood, SSRIs have been shown to be effective in reducing paraphilic symptoms. Kafka [19] treated 21 subjects with paraphilic disorders or paraphilic-related disorders with SSRI monotherapy. A total of seventeen subjects showed a decrease in symptoms while being treated with an SSRI without significant side effects. Greenberg *et al.* [20] demonstrated that sertraline, fluvoxamine, and fluoxetine were equally effective in reducing paraphilic symptoms. Although data exists to suggest that SSRIs are effective in the treatment of paraphilias, presently there is insufficient data to conclude that SSRIs are equally efficacious as antiandrogens or hormonal agents.

Psychological and Behavioral Treatment

Psychological approaches to the treatment of sex offenders focus on direct treatment approaches to modify offenders' cognitions and attitudes. Cognitive approaches described the concept of cognitive distortions (minimizations, excuses, justifications) as one factor that may maintain offender behavior [21]. Relapse prevention and the sexual assault cycle are theoretical models purporting to describe the cognitions and behaviors of sex offenders before, during, and after their abusive acts [22]. An offender prevents relapse by the identification and arrest of the thoughts and actions, which lead to sexual offending. The majority of sex offender treatment programs offer a combination of cognitive-behavioral and relapse prevention therapies.

The rationale for the implementation of behavioral techniques to control sexual arousal is based on the clinical observation that abusive sexual fantasy is linked with abusive sexual behavior. The factor that most consistently distinguishes male sex offenders from other males is sexual arousal disorder profiles as measured by phallomteric assessment [23]. The modification of an individual's behavior is achieved by altering individual reactions to stimuli through positive and negative reinforcements. These reinforcements condition an individual by reinforcing positive behaviors and extinguishing negative behaviors. More specifically, the behavioral techniques used to control sexual arousal employ olfactory aversion conditioning, covert sensitization, and masturbatory satiation to extinguish deviant sexual arousal and reinforce appropriate sexual arousal.

Recidivism rates from cognitive and behavioral treatment programs range from 3% to 31% depending on the study [4]. In a meta-analysis by Hall [24], it was found that cognitive-behavioral treatment and antiandrogen treatment were comparable in their treatment effects and significantly more effective than behavioral treatment alone.

Models of Treatment

The treatment of individuals with paraphilias and sex offenders has primarily focused on two areas: cognitive-behavior group treatment and pharmacologic treatment. Treatment programs are based on recent research that reveals that victim empathy, remorse, responsibility training, and relapse prevention are integral components of a treatment program. A survey conducted by the Safer Society in 1994 asked sex offender treatment providers to identify the treatment modalities delivered in their program. The results revealed

that greater than 90% of programs utilized the following treatment modalities: victim empathy, anger management, and cognitive distortions; 75% of programs provided social skills/assertiveness training, approximately 42% prescribed SSRIs, and 19% prescribed antiandrogens [25]. Langevin *et al.* [26] queried the sex offenders in treatment about acceptable types of treatment. The preferred therapies were individual psychotherapy, social skills training, and group therapy, whereas aversion conditioning, castration, and sex drive–reducing drug therapy were the least acceptable forms of treatment.

Ethical and Legal Considerations

The legal and ethical issues that arise in the course of sex offender treatment are unique to this population. As such, the traditional codes of ethics employed in medical treatment are not applicable to the treatment of sex offenders. A significant percentage of sex offenders receiving treatment have been mandated by the courts as part of incarceration or to be released into the community. The concept of mandated or involuntary treatment raises the issue of whether informed consent is possible in sex offender treatment. A condition of informed consent for medical treatment is that the consent must be voluntary. Obviously, court-ordered sex offender treatment is not voluntary. Individuals who reject treatment are subject to punishment imposed by the courts. Individuals who have been ordered by the court to receive treatment do not have a choice regarding the type of treatment or the treatment provider. Sex offenders are required to complete particular programs, irrespective of any other treatment that they might be receiving, in order to gain community release or avoid imprisonment [27].

In the medical model, patients have a right to refuse various types of treatment. The right to refuse treatment is based on an individual's constitutional right to privacy. The court views sex offenders as incompetent patients, and as such the court becomes the decision maker regarding treatment.

Confidentiality and Privilege

Confidentiality refers to the physician's obligation to keep information learned in a professional relationship private from other parties. Privilege refers to the patient's right to prevent a physician from providing testimony about personal medical information. Both confidentiality and privilege are routinely breached in sex offender treatment. When individuals enter treatment, they are required to give permission for their cases to be discussed with both clinical and nonclinical personnel, correctional officers, members of their family, past and potential victims, and those associated with them and fellow offenders [27].

In response to the deviations from traditional ethical codes inherent in sex offender treatment, ethical and practice guidelines were developed by the Association for the Treatment of Sexual Abusers (ATSA). ATSA's code of ethics endorses standards of professional conduct that promote competent practice, and as such, they represent a public commitment to clients and society toward the goal of preventing sexual violence. The ATSA guidelines state that the ethical care of sex offenders is achieved by encouraging individuals to take responsibility for their behavior, that is, admission of guilt. ATSA maintains that the identification and collaborative management of risk and safety factors are indeed in the best interests of both sex offender patients and potential victims owing to the grave consequences incurred by sexual offender recidivism [28].

Castration of Sex Offenders

In 1996, California became the first state to authorize the use of either chemical or surgical castration for certain sex offenders who were being released from prison into the community. To date, additional nine states have authorized surgical or chemical castration for sex offenders, including Georgia, Montana, Oregon, Wisconsin, Florida, Iowa, Louisiana, and Texas. The American Psychiatric Association stated "these laws, which predicate release from prison on chemical castration by surgery or antiandrogenic agents, are objectionable because they are not based on adequate diagnostic and treatment considerations. They also improperly link medical treatment with punishment and social control" [2]. In some states informed consent for the procedure is not required, while in other states informed consent requires only that the offender be informed regarding the side effects [29].

Conclusion and Future Directions

Further Research

Although studies have demonstrated the effectiveness of some modalities of sex offender treatment, the scientific literature does not identify a definitive effective treatment. The literature suggests that both cognitive-behavioral therapies and antiandrogen medication decrease sex offender recidivism. However, the current sex offender legislation in the United States is not based on the literature. Rather, the legislation focuses on the preventive detention of sex offenders without concurrent evidence-based sex offender treatment. The future of research in the treatment of sex offenders lies in the development of evidence-based knowledge that will inform both clinical decisions and public policy. This effort may be afforded by interdisciplinary collaboration between the scientific community, correctional professionals, and policy makers.

Specialized Training

Many psychiatrists are unfamiliar with the fundamentals of the assessment and treatment of individuals with paraphilias. The advent of laws for the civil commitment of sex offenders has resulted in the need for psychiatrists trained in paraphilias. As a result of sex offender legislature, the treatment and release decisions regarding sex offenders have become a focus of attention for psychiatry [30]. Forensic psychiatrists or psychologists must render an opinion as to whether the sex offender has a diagnosed mental disorder and, as such, represents a risk to public safety if released from custody into the community [31]. In order to provide competent care to individuals with paraphilias, psychiatrists should be educated about the evidence-based approach to the treatment and evaluation of this population.

Education of the Public

The general public perceives sex offenders as a homogenous group of offenders who are not treatable. However, recent meta-analysis has demonstrated that sex offenders are a highly heterogeneous population with different rates of reoffending and responsiveness to treatment. The successful treatment of sex offenders is dependent, in part, on the support of the public and the courts. In order to implement an evidence-based treatment, the offender must be afforded the opportunity for treatment. The opportunity for treatment is created by an informed society that legislates treatment that is informed by scientific evidence. The education of the public is integral in the successful treatment of sex offenders.

References

[1] Crepault, C. & Coulture, M. (1980). Men's erotic fantasies, *Archives of Sexual Behavior* **9**, 565–580.

[2] The American Psychiatric Association Task Force on Sexually Dangerous Offenders (1999). *Dangerous Sex Offenders: A Task Force Report of the American Psychiatric Association*, American Psychiatric Association, Washington, D.C.

[3] Zonana, H. (1997). The civil commitment of sex offenders, *Science* **278**, 1248–1249.

[4] Grossman Linda, S., Martis, B. & Fichtner, C.G. (1999). Are sex offenders treatable? A research overview, *Psychiatric Services* **50**(3), 349–361.

[5] Marshall, W.L. & Barbaree, H.E. (1988). The long-term evaluation of a behavioral treatment program for child molesters, *Behavior Research and Therapy* **26**, 499–511.

[6] Hanson, R.K. & Bussiere, M.T. (1996). *Predictors of Sexual Offender Recidivism: A Meta-analysis*, User Report No 1966-04, Department of the Solicitor General of Canada, Ottawa.

[7] Ferrall, W.R. & Card, R.D. (1988). Advancements in physiological evaluation of assessment and treatment in the sexual aggressor, in *Sexual Aggression: Current Perspectives*, R. Prentky & V.L. Quinsey, eds, New York Academy of Sciences, New York.

[8] Davidson, J.M., Smith, E.R. & Damassa, D.A. (1977). Comparative analysis of the roles of androgen in the feedback mechanisms and sexual behavior, in *Androgens and Antiandrogens*, L. Martini & M. Motta, eds, Raven, New York, pp. 137–149.

[9] Bancroft, J. (1989). *Human Sexuality and Its Problems*, Churchill Livingstone, Edinburgh.

[10] Sturup, G.K. (1953). Sexual offenders and their treatment in Denmark and other Scandinavian countries, *International Review of Criminal Policy* **4**, 1–19.

[11] Ortmann, J. (1980). The treatment of sexual offenders, castration and antihormone therapy, *International Journal of Law and Psychiatry* **3**, 443–451.

[12] Donovan, B.T. (1984). *Hormones and Human Behavior*, Cambridge University Press, London.

[13] Cooper, A.J., Sandhu, S., Losztyn, S. & Cernovsky, Z. (1992). A double-blind placebo controlled trial of medroxyprogesterone acetate and cyproterone acetate with seven pedophiles, *Canadian Journal of Psychiatry* **37**, 687–693.

[14] Bradford, J.M. (1983). Research on sex offenders: recent trends, *Psychiatric Clinics of North America* **6**, 715–731.

[15] Bradford, J.M.W. (1983). Research in sex offenders, *The Psychiatric Clinics of North America* **6**(4), 715–733.
[16] Rosler, A. & Witztum, E. (1998). Treatment of men with paraphilia with a long-acting analogue of gonadotropin-releasing hormone, *New England Journal of Medicine* **338**, 416–465.
[17] Briken, P., Hill, A. & Berner, W. (2003). Pharmacotherapy of paraphilias with long-acting agonists of luteinizing hormone-releasing hormone: a systematic review, *The Journal of Clinical Psychiatry* **64**(8), 890–897.
[18] Krueger, R.B. & Kaplan, M.S. (2001). Depot-leuprolide acetate for treatment of paraphilias: a report of twelve cases, *Archives of Sexual Behavior* **30**, 409–422.
[19] Kafka, M.P. (1994). Sertraline pharmacotherapy for paraphilias and paraphilia-related disorders an open trial, *Annals of Clinical Psychiatry: Official Journal of the American Academy of Clinical Psychiatrists* **6**, 189–195.
[20] Greenberg, D.M., Bradford, J.M.W., Curry, S. & O'Rourke, A. (1996). A comparison of treatment of paraphilias with three serotonin reuptake inhibitors: a retrospective study, *The Bulletin of the American Academy of Psychiatry and the Law* **24**, 525–532.
[21] Murphy, W.D. (1990). Assessment and modification of cognitive distortions in sex offenders, in *Handbook of Sexual Assault: Issues, Theories, and Treatment of the Offender*, W.L. Marshall, D.R. Laws & H.E. Barbaree, eds, Plenum Press, New York, pp. 331–342.
[22] Carich, M.S., Gray, A., Rombouts, S., Stone, S., Pithers, M. & William, D. (2003). Relapse prevention and the sexual assault cycle, in *Handbook for Sexual Abuser Assessment and Treatment*, Safer Society Press, Vermont, pp. 77–104.
[23] Murphy, W.D. & Barbaree, H.E. (1994). *Assessments of Sex Offenders by Measures of Erectile Response: Psychometric Properties and Decision Making*, Safer Society Press, Brandon.
[24] Hall, G.C.N. (1995). Sexual offender recidivism revisited: a meta-analysis of recent treatment studies, *Journal of Consulting and Clinical Psychology* **63**, 802–809.
[25] Freeman-Longo, R.E., Bird, S., Stevenson, W.F. & Fiske, J. (1994). *Nationwide Survey of Treatment Programs and Models*, Safer Press, Brandon, pp. 19–20.
[26] Langevin, R., Wright, P. & Handy, L. (1988). *Sexual Abuse: A Journal of Research and Treatment* **1**(3).
[27] Glaser, B. (2003). Therapeutic jurisprudence: an ethical paradigm for therapists in sex offender treatment programs, *Western Criminology Review* **4**(2), 143–154.
[28] Levenson, J.S. & D'Amora, D.A. (2003). An ethical paradigm for sex offender treatment: response to glaser, *Western Criminology Review* **4**(2), 143–154.
[29] Scott, C.L. & Holmberg, T. (2003). Castration of sex offenders: prisoners' rights versus public safety, *The Journal of the American Academy of Psychiatry and the Law* **31**, 502–509.
[30] Lieb, R., Quinsey, V. & Berliner, L. (1998). Sexual predators and social policy, *Crime and Justice* **23**, 43–114.
[31] Screenivasan, S., Weinberger, L.E. & Garrick, T. (2003). Expert testimony in sexually violent predator commitments: conceptualizing legal standards of "mental disorder" and "likely to reoffend", *The Journal of the American Academy of Psychiatry and the Law* **31**(4), 471–485.

RENEE SORRENTINO

Sexual Abuse ***see*** Child Sexual Abuse Accommodation

Sexual Abuse: Child ***see*** Child Sexual Abuse

Sexual Abuse Accommodation Syndrome ***see*** Child Sexual Abuse Accommodation

Sexual Assault ***see*** Assault: Sexually Motivated

Sexual Assault: Child ***see*** Child Sexual Abuse

Sexual Assault: Drugs ***see*** Drug-Facilitated Sexual Assault

Sexual Fatalities ***see*** Autoerotic Deaths

Sexual Homicide ***see*** Homicide: Multiple (Behavior)

Sexual Misadventure ***see*** Autoerotic Deaths

Sexually Transmitted Infections ***see*** Northwest Juvenile Project

Sexually Violent Predator ***see*** Dangerousness: Risk of

Shaken Baby ***see*** Battered Child Syndrome

Shaken Baby Syndrome

Introduction

The diagnosis "shaken baby syndrome" (SBS) has been widely accepted for over 30 years, but recent evidence from biomechanical and clinical observational studies questions the validity of the syndrome.

Definition

The diagnosis of SBS is based on the clinical triad of encephalopathy, retinal hemorrhage (RH), and subdural hemorrhage (SDH) in infants, usually under six months of age, who may die unexpectedly or survive with greater or lesser degrees of neurological damage [1]. The term non-accidental head injury (NAHI) has been preferred as it has no implications for mechanism of injury. Other features often associated include a sole carer at the time of collapse and a clinical history that is incompatible with the severity of the injuries. The diagnosis of inflicted injury becomes less problematic if there is objective evidence of violence, such as bruises, fractures, or burns, but objective evidence of trauma has not always been necessary in making the diagnosis.

Central to the assessment of these cases is whether the triad of findings can be regarded as diagnostic of abuse with any degree of certainty. This review examines the evidence base for each element of the triad and the current biomechanical evidence regarding mechanisms of infant head injury and its pathological investigation.

History

SDH has been associated with child abuse since the mid-19th century [2]. Kempe described SDH with multiple skeletal injuries and bruises as the battered child syndrome and Caffey described long bone fractures and SDH [3–5], but it is Guthkelch [6] who developed the hypothesis that the whiplash–like movements during shaking cause the characteristic bilateral thin film SDH of the syndrome. He based his hypothesis, that shaking causes tearing of the cerebral bridging veins leading to SDH, on the biomechanical studies of Ommaya [7] who was researching adult head injury in road traffic accidents. Following Guthkelch's paper, the "shaken baby syndrome" has become widely accepted as a form of child abuse [1].

The Triad of Injuries

The three elements of the triad are encephalopathy, RH, and SDH.

Retinal Hemorrhages (RHs)

RHs have been regarded as an important indicator of inflicted injury, but many other causes of retinal bleeding are recognized in infants, for example after normal birth, raised intracranial pressure, blood dyscrasias, hemoglobinopathies, extracorporeal membrane oxygenation, cataract surgery, and accidental

trauma [8]. Postmortem indirect ophthalmoscopy has shown RHs to be more common after natural disease and accidental injury than after inflicted injury [9]. These authors also noted that infants suspected to have been abused were more likely to have ophthalmological examination in life than infants with accidental injuries or natural diseases. This bias readily distorts the true incidence of RH in non-accidental injury. Indeed Vinchon [10] noted in his study of infant head injury that "*In the construct of our study we could not obviate the circularity bias, and the evaluation of the incidence of RH in child abuse remains a self-fulfilling prophecy*". These authors did, however, suggest that the extent and nature of retinal bleeds may be more important as indicators of inflicted head injury than their existence per se [10].

The main hypotheses for genesis of RH are that it is the result of venous obstruction, which in turn may result from compression of the optic nerve by raised intracranial or intravascular pressure, even transiently, or that the tissues of the retina are torn during the act of shaking. This latter hypothesis does not withstand biomechanical scrutiny [11].

Encephalopathy

This term may be widely interpreted to include a range of clinical manifestations from feeding difficulties, vomiting, and sleepiness to seizures and fulminating cerebral edema.

The specific neuropathological features of traumatic brain injury are contusions and traumatic axonal injury. Hypoxic-ischemic injury and brain swelling are frequently seen but are not specific for trauma. Contusions are very uncommon in infant brain trauma in the absence of skull fractures. Identification of axonal injury now depends on the immunocytochemical demonstration of beta amyloid precursor protein (BAPP). This is a very sensitive marker of interruption of normal axonal flow but may be upregulated after hypoxic–ischemic injury and metabolic disruption as well as trauma (Figure 1). Distinction of traumatic axonal expression of BAPP from other causes is fraught with difficulty, and depends in part on its distribution [12], [13], [14]. Neuropathological studies have shown that in babies who die following NAHI, the underlying brain pathology is widespread hypoxic-ischemic injury and not diffuse traumatic axonal injury as previously believed [12, 13]. In this series axonal injury was seen in a limited distribution in the lower brainstem and in only a minority of cases. Radiological studies have confirmed these pathological observations [15].

This observation is important as traumatic axonal injury will lead to immediate loss of function causing clinical symptoms from the time of trauma. In contrast, hypoxic-ischemic injury and ensuing brain swelling take variable periods of time to develop and a baby so damaged may not show immediate symptoms. Even fatal brain trauma may present with a lucid interval between injury and clinical collapse [16, 17]. Lucid intervals are more frequently seen in infants less than two years of age [18], reflecting the very different responses of the infant brain to injury due to the specific intracranial pathophysiology before the skull bones fuse [19].

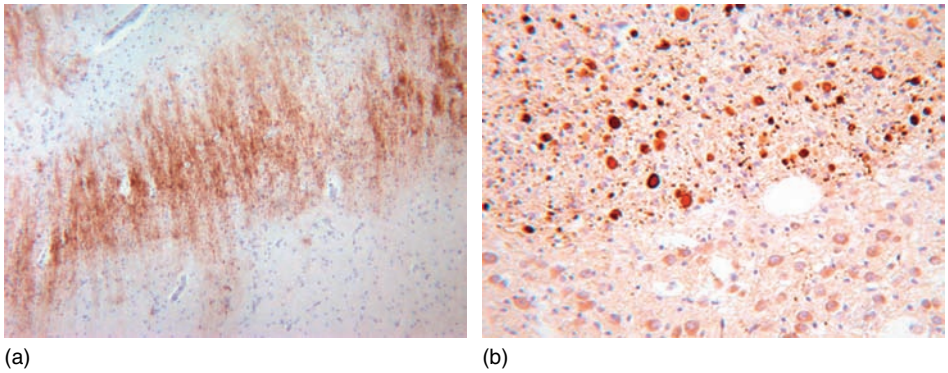


Figure 1 (a) Acute axonal injury. Bands of BAPP expression in an infarcted area of brain in acute hypoxic-ischemic injury. (b) Axonal swellings expressing BAPP restricted to the pontine cortico-spinal tracts, considered to indicate traumatic damage

Damage to the cervical nerve roots has been documented as part of the pathology of shaking injury [14]. It has not been established that this is the result of shaking, as cervical cord displacement resulting from brain swelling may also cause traction on nerve roots in the region. Autopsy studies in man and primates have shown that the spinal cord is displaced during extension and flexion of the neck [20, 21] and it remains a possibility that hyperextension and flexion could cause traction damage to nerve roots throughout the length of the spinal cord, but this has not been documented in living infants.

Subdural Hemorrhage (SDH)

SDH is perhaps the most important and consistent component of the triad. In the acutely sick infant, it is frequently the first clinical sign, identified on brain scan, to raise the question of abuse. There are no specific imaging patterns that can distinguish inflicted from accidental intracranial injury [22, 23].

Autopsy and imaging studies show that infant SDH is usually a thin bilateral film and not a thick, unilateral space occupying clot as seen in traumatic SDH in older children and adults [12, 13, 24]. This raises the question of whether the two forms have the same etiology and anatomical source.

Causes of Subdural Hemorrhage. The commonest cause of SDH in infants is said to be trauma [25] although a recent study has shown a significant incidence (26%) of birth-related SDH [26]. Other causes in infants include benign enlargement of the extracerebral spaces (BEECS), clotting disorders, hemorrhagic disease of the newborn, rare metabolic diseases, vascular malformations, and neurosurgical procedures [25, 27].

Traumatic SDH

Proposed traumatic causes of infant SDH are inflicted injury such as shaking and/or impact and accidental injuries such as falls. Impact includes blunt impact of an object on the head and that resulting from a fall or striking the moving head on a rigid surface. The biomechanical aspects of these injuries are discussed below. The vast majority of cases described as SBS have evidence of impact [28]. While the pathologist may be able to determine features indicative of impact, it is not, of course, possible to distinguish accidental from non-accidental injuries by pathology.

Low-Level Falls

Low-level falls have the potential, albeit only rarely, to cause SDH in infants and young children. Absolute height is not as important a criterion for injury as the exact nature of the fall for a particular infant, in a particular circumstance [29]. The effects of twisting, rotation, or crushing of the structures of the neck are crucial in terms of outcome. Biomechanical studies show that falls even from low levels of 3–4 ft can generate far greater forces in the head than shaking [11]. There are a number of case series demonstrating that infants and children may suffer intracranial damage including retinal and intracranial hemorrhage after falls from levels as low as 3 ft [10, 17, 30–33]. While most babies may suffer little from an apparently trivial fall, this is clearly not always the case.

Birth-Related SDH

Three studies, using magnetic resonance imaging (MRI), have shown a surprisingly high incidence of SDH after birth in asymptomatic infants. Whitby identified SDH in the first two days of life in 9% [32], while SDH was seen in up to 46% of otherwise normal neonates using higher resolution MRI scanning [26, 34]. With regard to method of delivery, ventouse or instrumental deliveries have been associated with a higher incidence of intracranial injury [35, 36]. Towner [37] found an increased incidence of intracranial hemorrhage after instrumental delivery with ventouse or forceps and emergency caesarean section, but the incidence was lower after caesarean section before labor had begun. However, it should be noted that all of Looney's cases followed normal vaginal delivery [26].

While neonates with SDH may be asymptomatic [26, 35] they may also have signs in the neonatal period including unexplained apnoea, dusky episodes, hypotonia, seizures, and lethargy [38].

Sources of SDH. Traditional belief is that in SBS the SDH results from tearing of the superficial bridging veins as they cross from the brain to the dural sinuses [6] (Figure 2). This has never been proved. Indeed it is very difficult to find documented evidence of torn bridging veins at surgery or at autopsy. Cushing, who operated on neonates with SDH and subsequently performed the autopsies wrote *"In two of the cases I have examined I have satisfied myself*

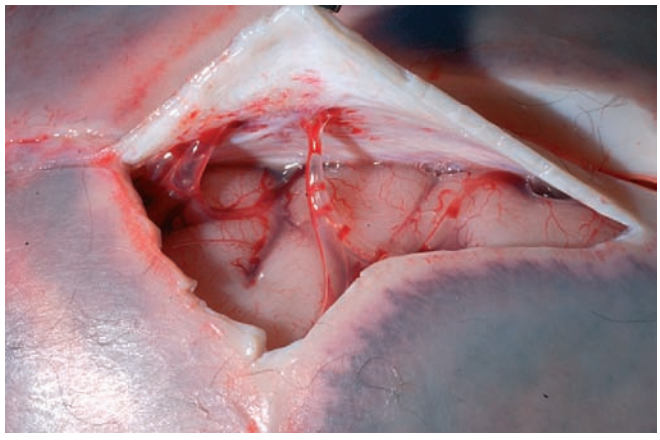


Figure 2 Infant bridging veins may be visualized by opening the skull very carefully, but they are readily torn in normal autopsy procedures. (Picture courtesy of Dr P. Lantz)

that such ruptures were present. A positive statement, however, cannot be given even for these cases, since the dissection and exposure, difficult enough under any circumstances, owing to the delicacy of the vessels is the more so when they are obscured by extravasated blood" [39]. More recently Maxeiner [40] addressed the problem by injecting radio-opaque dye into the veins at autopsy to assess their integrity after removing the top of the head in one piece, hard-boiled egg style. This approach is not widely used as it destroys much of the brain and injection pressures need to be carefully monitored if the veins are not to be ruptured artifactually.

Volpe [41] said that SDH was by no means always traumatic and suggested that in neonates without tentorial tears the bleeding may arise from the tributary veins of the dural sinuses. Autopsy studies from the older literature show bridging vein rupture is uncommon, Craig described 62 neonatal SDH, of which only 3 had torn bridging veins, all of those with overriding sutures [42]. Larroche described 700 autopsies 18% with SDH. [43] She noted an association with hypoxic-ischemic injury (Figure 3). She did not identify torn veins.

If SDH does not arise from torn bridging veins, what other sources may there be? Two obvious

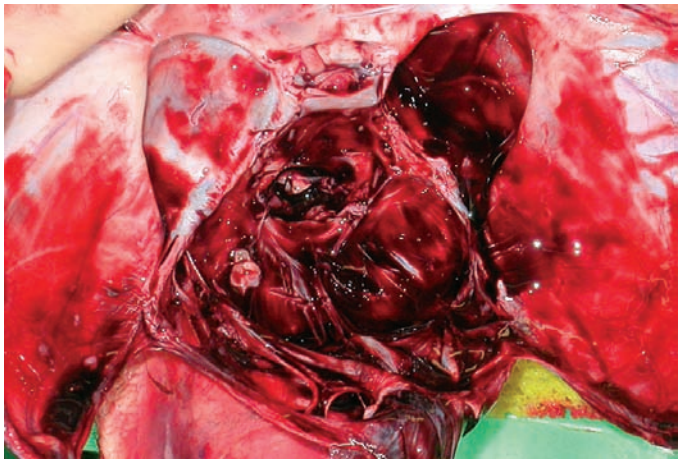


Figure 3 Fresh subdural blood seen after birth asphyxia. (Picture courtesy of Dr I. Scheimberg)

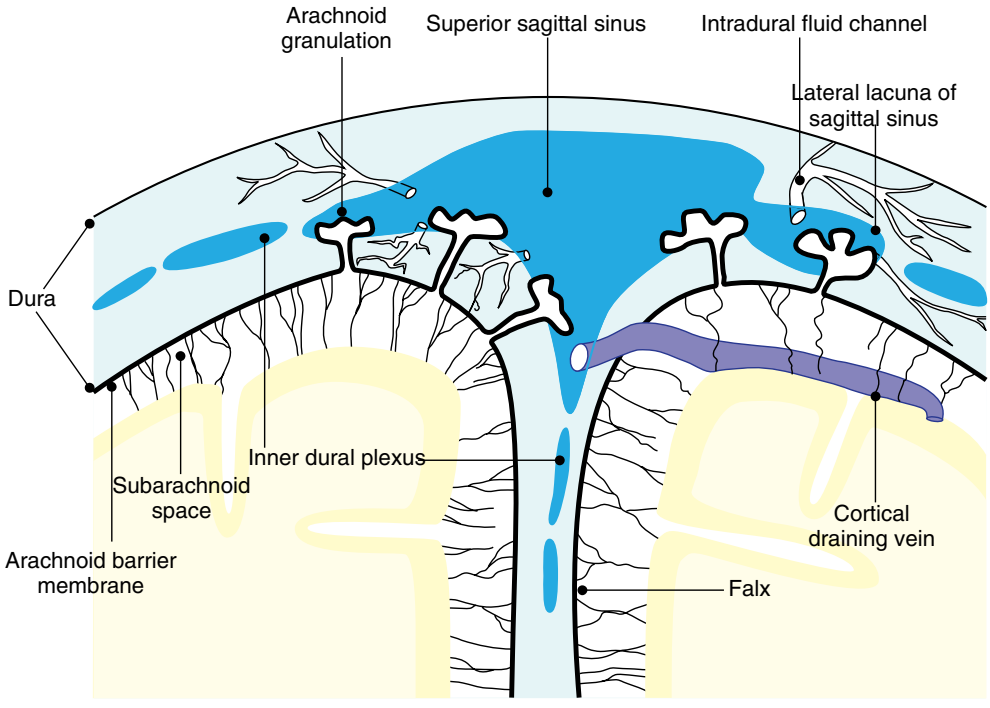


Figure 4 Diagram representing a coronal slice through the brain and dura indicating the intradural sinuses and their relationship to cortical surface veins, arachnoid granulations, and intradural fluid channels

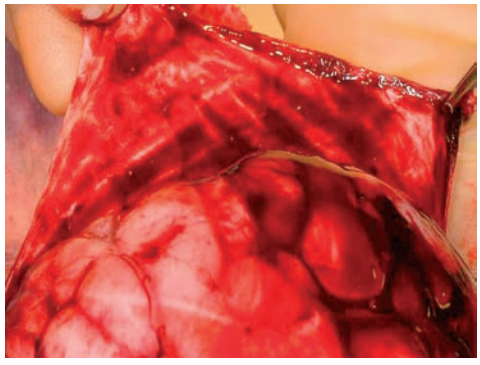
alternative sites of origin exist, the dura itself and the old subdural membranes (Figure 4).

Dural Hemorrhage

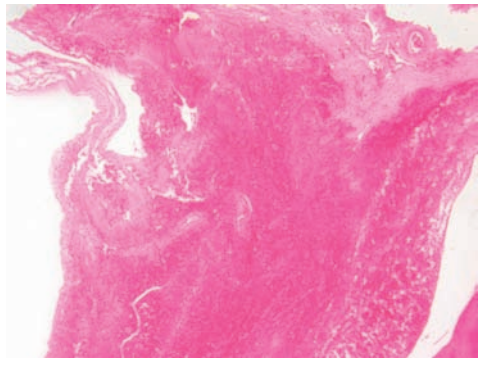
The dura is composed of two leaflets, the periosteal and the meningeal dura, separated by a thin vascular channel, which widens to form the large dural sinuses [44]. There are particularly extensive venous sinuses in the posterior falx, [45] a frequent site of high signal on brain scans in asphyxiated infants. Bleeding into the falx is well recognized in asphyxiated infants [46]. It has long been acknowledged that optic nerve sheath hemorrhage arises from the dura [47] and more recently the dura was proposed as the source of intracranial SDH in infants [48] (Figure 5). Careful microscopic examination of the dura confirms that intradural bleeding is common in asphyxiated infants, particularly in the dural folds of the falx and tentorium close to the large venous sinuses [49]. In some cases intradural bleeding leaks out on to the subdural surface leading to macroscopically evident subdural haematoma [50].

Healing Subdural Membranes

Healing of SDH is by formation of a thin, vascular membrane consisting of fibroblasts, macrophages, which often contain altered blood products, and wide thin-walled capillaries with a potential to rebleed [51] (Figure 6). It is uncommon in infants to see a double layered membrane around a localized mass of resolving clot, as seen in the elderly, probably because the infant SDH usually forms as a thin film rather than as a mass lesion. Contrast injection is required to identify the membranes radiologically [52]. In some cases, acute SDH leads to accumulation of fluid in the subdural space. The reasons for this are unknown. Fluid collections may result from immaturity of the arachnoid granulations and impaired cerebrospinal fluid (CSF) absorption [22], and be influenced by the method of treatment of the acute hematoma. Surgical evacuation or tapping may prevent later reaccumulation of fluid [53, 54]. The period of time for redevelopment of subdural fluid collections may be long, between 15 and 111 days [55]. It is likely that an important



(a)

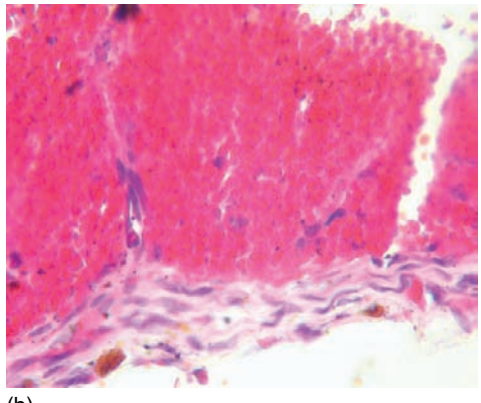


(b)

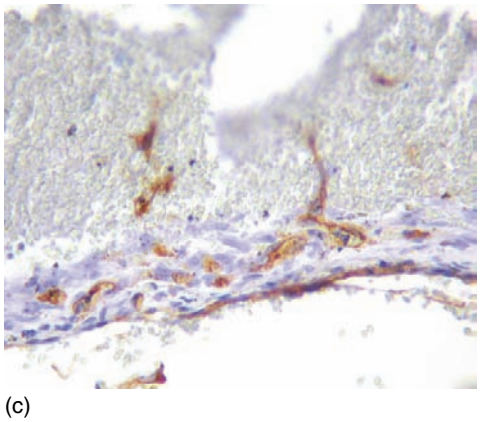
Figure 5 (a) The dura is thickened and congested and there is patchy subarachnoid and subdural blood. Autopsy 44 h after collapse following choking episode. (Courtesy of Dr I. Sheimberg.) (b) H & E stained section of falx showing it to be destroyed by massive acute bleeding



(a)

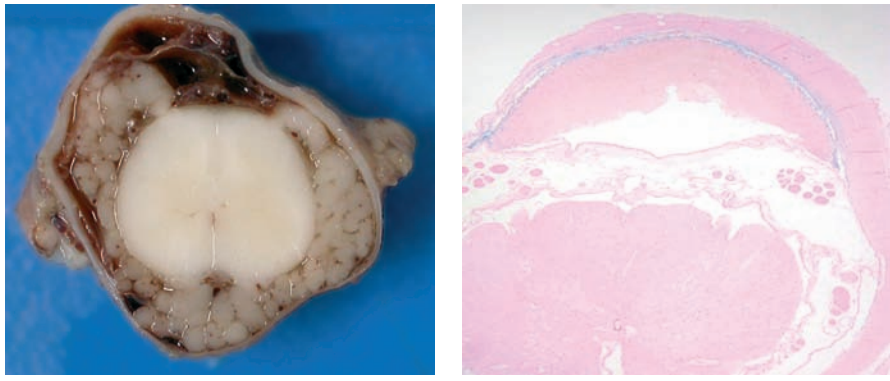


(b)



(c)

Figure 6 (a) Dural surface showing a very thin yellow-brown membrane, which has partly lifted during removal of the brain. Head injury four weeks prior to death. (b) H & E stained section of acute bleed overlying a chronic membrane, which consists of some six layers of fibroblasts between which are macrophages and new capillaries (three days after collapse with acute SDH) (c) Same section stained with CD34 to show endothelial cells. Note capillaries growing into the fresh clot



(a) (b)

Figure 7 (a) A collection of fresh subdural blood at the dorsal aspect of the sacral spinal cord. Baby died within hours of inflicted abdominal injury with acute and chronic subdural hemorrhage. (b) Microscope section showing an elliptical collection of fresh blood dorsal to the spinal cord. The blood is within a chronic subdural membrane indicated by the iron pigment, stained here by Perl's stain. Baby died three weeks after traumatic subdural hemorrhage

contribution to chronic subdural fluid accumulation is repeated rebleeding and oozing from a chronic subdural membrane [56, 57].

There is little information regarding the potential for birth-related SDH to evolve into chronic fluid collections. Whitby followed nine cases with a repeat scan at one month; none had developed a chronic collection [35]. Rooks followed 18 cases for up to 3 months, one developed a further subdural bleed [34]. However these studies could not identify membranes as contrast was not used. Chronic membranes have been seen at autopsy in up to 31% of infants dying unexpectedly without previous clinical evidence of chronic SDH [58]. In view of the potential for acute accidental SDH to evolve into a chronic collection several months later [55], it would appear likely that the same pattern would follow birth-related SDH. At this time, we simply have insufficient information.

Distribution. In the first few days after bleeding, subdural blood sediments under the influence of gravity and undergoes secondary redistribution to the most dependent part, the posterior falx and tentorium [59]. Radiological studies show that subdural blood tracks down around the spinal cord [60] and, if the spine of babies with intracranial SDH is examined at autopsy, blood is regularly seen in the subdural space and around sacral nerve roots in the most dependent parts of the dural sac (Figure 7).

Differential Diagnosis of SBS

The most common causes of the triad are impact, birth-related SDH, BEECS, coagulopathies, apnoea, asphyxia and choking, acute life-threatening events (ALTEs), osteogenesis imperfecta, osteopenia of prematurity, and metabolic diseases [14, 28, 61, 62, 63].

Choking/Asphyxia

In a considerable number of cases, vomiting and/or reflux are described at the time of collapse, and in some there is a history of feeding difficulties, gastroesophageal reflux, and choking or apnoeic episodes [14, 62]. SBS is commonly diagnosed in the first three months of life, the age of peak incidence of sudden infant death syndrome. Inhalation of feed or vomit may play a part in sudden infant death [64] and awake apnoea is associated with gastroesophageal reflux [65]. The physiological response to aspiration may be dramatic; foreign material on the larynx causes laryngospasm, which is associated with startle, cessation of respiration, hypoxaemia, bradycardia, and a doubling of blood flow to the brain [66]. These circumstances, with or even without vigorous resuscitation, may cause reperfusion injury and a pre-existing healing subdural membrane may bleed. The dura itself may become hemorrhagic and ooze blood into the subdural space (Figure 8). As long ago as 1905, Cushing suggested that coughing, choking, and



Figure 8 (a) Cortical vein thrombosis. Infant died 10 days after collapse following two choking episodes. Several surface veins are thrombosed (arrows). (b) Section of thrombosed vein shows a network of new capillaries growing into the periphery of the thrombus (CD31)

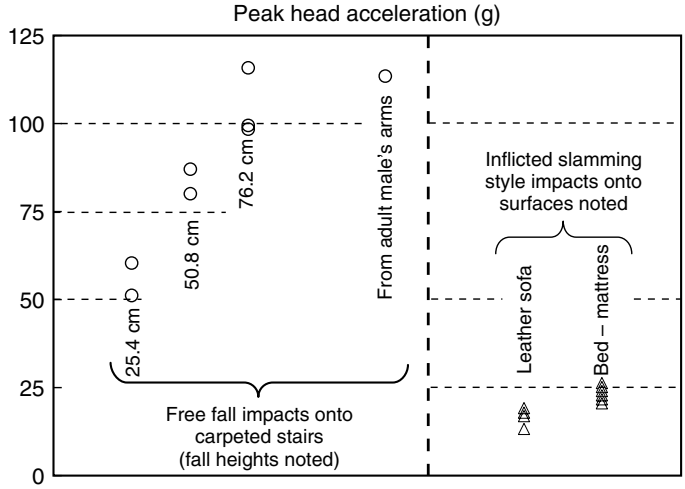


Figure 9 Comparative forces generated by dropping or shaking and slamming a dummy representing a six-month-old infant (C Van Ee, personal communication 2007)

venous congestion may explain some forms of infant SDH [39], a hypothesis recently revived by Geddes, [48, 67].

Biomechanics

Biomechanics is the application of principles of physics to biological systems and has been the mainstay of research into motor vehicle safety for six decades. It was just such research into noncontact head injury from rear-end shunts that stimulated Guthkelch to formulate his hypothesis for SBS in 1971 [6]. Ommaya [7] had caused concussion, SDH, and white matter shearing injury (diffuse axonal injury) in primates by whiplash. Guthkelch suggested that the rotational forces of shaking would cause tearing of bridging veins and bilateral subdural bleeding, although Ommaya himself warned that "*It is improbable that the high speed and severity of the single whiplash produced in our animal model could be achieved by a single manual shake or even a short series of manual shaking of an infant in one episode*".

More recent studies using "crash test dummies" indicate that impact generates far more force than

shaking (Figure 9) and that impact is required to produce SDH [68]. Cory and Jones [69] generated forces that exceeded the injury threshold for concussion, but not for SDH or axonal injury. Their adult shaker volunteers fatigued after 10 seconds. While they concluded that "*It cannot be categorically stated, from a biomechanical perspective, that pure shaking cannot cause fatal head injuries in an infant*", they noted that in their experiments there were chin and occipital contacts at the extremes of the shaking motion that could have caused impact. These authors expressed their concerns regarding the difficulties in extrapolating to human infants the findings in both dummy and animal models. Biomechanical studies have shown that falls and impact to the head produce significant rotational forces when the impacting forces are not aligned through the center of gravity of the head, due to hinging of the head on the neck. Shaking is not necessary to cause rotational acceleration.

Neck injuries may be underreported in babies dying after severe abuse [70]. In Ommaya's study, 11 of 19 primates had neck injuries; these were adult animals with mature neck structure and musculature. It is likely that the forces required to cause intracranial injury will also damage the weak infant neck [71]. In road traffic accidents, infants who suffer single severe hyperextension forces have cervical fractures, dislocations, spinal cord injury, and torn nerve roots, not SDH [72–74].

Investigation of Shaken Baby Syndrome

SBS or NAHI is most likely to occur in an infant dying suddenly under the age of six months. Autopsy should be performed with careful consideration of this diagnosis and appropriate steps taken to support or exclude it. The records of pregnancy and delivery must be carefully studied to look for any evidence of complications that could mimic NAHI. These include pregnancy disorders such as oligohydramnios, fetal hypokinesia, and prematurity, which lead to osteopenia and predispose to fractures. The birth history and method of delivery are important as SDH may arise at this time while being entirely asymptomatic in the neonatal period. Head circumference charts are important; head circumference measurements taken at birth and in the subsequent weeks may reflect abnormal head growth, which can indicate an accumulating subdural fluid collection and a propensity to rebleed.

The clinical history may give clues to other problems in the early weeks of life. Vomiting, feeding problems, and apnoeic episodes and ALTEs may indicate difficulties with coordination of breathing, sucking and swallowing, and vulnerability to choking. Any event that threatens life may also potentially end it.

The history of the baby's terminal collapse must also be carefully examined. Parents may describe events that reveal a cause for collapse. In any other field of medicine, the clinical history is regarded as the cornerstone of diagnosis and it should not be disregarded without serious critical evaluation.

The autopsy can reveal evidence of trauma such as deep bruises and fractures not seen in clinical examination. The examination of the intracranial contents is paramount. The scalp and skull require careful examination for evidence of bruising and fractures. Suture separation due to raised intracranial pressure and wormian bones can be mistaken for fractures. When the cranium is opened, the presence of any intracranial bleeding must be noted. Unclotted blood may escape from the subdural space as the skull is opened and be mistaken for bleeding from the dural sinuses. It is important to note the volume and nature of blood and the presence of xanthochromia, indicating older bleeding. As the cranium is opened, the bridging veins should be visualized and their integrity assessed. If there is a question of bridging vein rupture, histological examination may assist in establishing this. The dural sinuses and draining veins should be examined for evidence of thrombosis.

The dura must be carefully examined for evidence of older bleeding. A chronic subdural membrane may be thin and patchy and represented only by patches of light brown discoloration. Multiple samples should be taken from the dura, including the falx and tentorium, for histological examination to look for evidence of intradural bleeding and rupture onto the subdural surface. This may be the source of significant subdural blood.

The brain must be fixed for detailed histological examination.

In all of these cases, the time between collapse and death may play a significant part in the final pathology. A baby who has collapsed and becomes apnoeic with subsequent cardiopulmonary rescuscitation (CPR) and ventilation will be shocked and suffer multiorgan failure with altered clotting, loss

of integrity of vessels and membranes, oozing of blood into intracranial compartments, including the subarachnoid and subdural spaces, and development of the "respirator brain".

Review of the brain imaging in life is essential in assessing, as far as possible, just how much hemorrhage occurred at the time of collapse and how much may be the result of subsequent secondary changes. It is recognized that SDH may continue to bleed after initial onset [75] especially if a baby is very sick. Finding a large clot at autopsy may suggest traumatic rupture of a large vessel, but comparison with early brain scans may indicate that the bleed was only minor at the outset, indicating a slower oozing process with different implications for causation. It is becoming increasingly obvious that not all SDH arises from traumatic rupture of blood vessels.

Acknowledgment

I would like to thank Dr Irene Scheimberg and Dr Pat Lantz for providing pictures and Dr Chris Van Ee for valuable discussion and for preparing Figure 8.

References

[1] American Academy of Pediatrics Committee on child abuse and neglect: shaken baby syndrome: inflicted cerebral trauma. (1993). *Pediatrics* **92**(6), 872–875.

[2] Tardieu, A. (1860). *Etude medico-legale sur les sevices et mauvais traitements exerces sur des enfants*.

[3] Kempe, C.H., Silverman, F.N., Steele, B.F., Droegemueller, W. & Silver, H.K. (1962). The battered-child syndrome. *JAMA* **181**, 17–24.

[4] Caffey, J. (1972). On the theory and practice of shaking infants. Its potential residual effects of permanent brain damage and mental retardation. *American Journal of Diseases of Children* **124**(2), 161–169.

[5] Caffey, J. (1974). The whiplash shaken infant syndrome: manual shaking by the extremities with whiplash-induced intracranial and intraocular bleedings, linked with residual permanent brain damage and mental retardation. *Pediatrics* **54**(4), 396–403.

[6] Guthkelch, A.N. (1971). Infantile subdural haematoma and its relationship to whiplash injuries. *British Medical Journal* **2**(759), 430–431.

[7] Ommaya, A.K., Faas, F. & Yarnell, P. (1968). Whiplash injury and brain damage: an experimental study. *JAMA* **204**(4), 285–289.

[8] Tongue, A.C. (2007). The ophthalmologist's role in diagnosing child abuse. *Ophthalmology* **98**(7), 1009–1010.

[9] Lantz, P.E. & Stanton, C.A. (2006). Postmortem detection and evaluation of retinal hemorrhages. *American Academy of Forensic Science* **271**.

[10] Vinchon, M., Defoort-Dhellemmes, S., Desurmont, M. & Dhellemmes, P. (2005). Accidental and nonaccidental head injuries in infants: a prospective study. *Journal of Neurosurgery* **102**(4 Suppl), 380–384.

[11] Ommaya, A.K., Goldsmith, W. & Thibault, L. (2002). Biomechanics and neuropathology of adult and paediatric head injury. *British Journal of Neurosurgery* **16**(3), 220–242.

[12] Geddes, J.F., Hackshaw, A.K., Vowles, G.H., Nickols, C.D. & Whitwell, H.L. (2001). Neuropathology of inflicted head injury in children. I. Patterns of brain damage. *Brain* **124**(Pt 7), 1290–1298.

[13] Geddes, J.F., Vowles, G.H., Hackshaw, A.K., Nickols, C.D., Scott, I.S. & Whitwell, H.L. (2001). Neuropathology of inflicted head injury in children. II. Microscopic brain injury in infants. *Brain* **124**(Pt 7), 1299–1306.

[14] Shannon, P., Smith, C.R., Deck, J., Ang, L.C., Ho, M. & Becker, L. (1998). Axonal injury and the neuropathology of shaken baby syndrome. *Acta Neuropathologica* **95**(6), 625–631.

[15] Stoodley, N. (2002). Non-accidental head injury in children: gathering the evidence. *Lancet* **360**(9329), 271–272.

[16] Denton, S. & Mileusnic, D. (2003). Delayed sudden death in an infant following an accidental fall: a case report with review of the literature. *American Journal of Forensic Medicine and Pathology* **24**(4), 371–376.

[17] Plunkett, J. (2001). Fatal pediatric head injuries caused by short-distance falls. *American Journal of Forensic Medicine and Pathology* **22**(1), 1–12.

[18] Arbogast, K.B., Margulies, S.S. & Christian, C.W. (2005). Initial neurologic presentation in young children sustaining inflicted and unintentional fatal head injuries. *Pediatrics* **116**(1), 180–184.

[19] Talbert, D.G. (2006). The 'Sutured Skull' and intracranial bleeding in infants. *Medical Hypotheses* **66**(4), 691–694.

[20] Reid, J.D. (1960). Effects of flexion-extension movements of the head and spine upon the spinal cord and nerve roots 1. *Journal of Neurology, Neurosurgery, and Psychiatry* **23**, 214–221.

[21] Smith, C.G. (1956). Changes in length and position of the segments of the spinal cord with changes in posture in the monkey 1. *Radiology* **66**(2), 259–266.

[22] Vinchon, M., Noizet, O., Defoort-Dhellemmes, S., Soto-Ares, G. & Dhellemmes, P. (2002). Infantile subdural hematomas due to traffic accidents. *Pediatric Neurosurgery* **37**(5), 245–253.

[23] Ewing-Cobbs, L., Prasad, M., Kramer, L., Louis, P.T., Baumgartner, J., Fletcher, J.M. & Alpert, B. (2000). Acute neuroradiologic findings in young children with inflicted or noninflicted traumatic brain injury. *Child's Nervous System* **16**(1), 25–33.

[24] Duhaime, A.C., Christian, C.W., Rorke, L.B. & Zimmerman, R.A. (1998). Nonaccidental head injury in

infants–the "shaken-baby syndrome". *New England Journal of Medicine* **338**(25), 1822–1829.

[25] Kemp, A.M. (2002). Investigating subdural haemorrhage in infants. *Archives of Disease in Childhood* **86**(2), 98–102.

[26] Looney, C.B., Smith, J.K., Merck, L.H., Wolfe, H.M., Chescheir, N.C., Hamer, R.M. & Gilmore, J.H. (2007). Intracranial hemorrhage in asymptomatic neonates: prevalence on MR images and relationship to obstetric and neonatal risk factors. *Radiology* **242**(2), 535–541.

[27] Hymel, K.P., Jenny, C. & Block, R.W. (2002). Intracranial hemorrhage and rebleeding in suspected victims of abusive head trauma: addressing the forensic controversies. *Child Maltreatment* **7**(4), 329–348.

[28] Leestma, J.E. (2005). Case analysis of brain-injured admittedly shaken infants: 54 cases, 1969–2001. *American Journal of Forensic Medicine and Pathology* **26**(3), 199–212.

[29] Monson, K., Sparrey, C., Cheng, L., Van Ec, C. & Manley, G. (2007). Head Exposure Levels in Pediatric Falls (P48), National Neuro Trauma Society Meeting, Kansas City, USA.

[30] Kim, K., Wang, M., Griffith, P., Summers, S. & Levy, M. (2000). Analysis of pediatric head injury from falls. *Neurosurgical Focus* **8**(1), e3.

[31] Greenes, D.S. & Schutzman, S.A. (1998). Occult intracranial injury in infants. *Annals of Emergency Medicine* **32**(6), 680–686.

[32] Schloff, S., Mullaney, P.B., Armstrong, D.C., Simantirakis, E., Humphreys, R.P., Myseros, J.S., Buncic, J.R. & Levin, A.V. (2002). Retinal findings in children with intracranial hemorrhage. *Ophthalmology* **109**(8), 1472–1476.

[33] Aly-Hamdy, N., Childs, A.M., Ferrie, C.D. & Livingston, J.H., BPNA (2001). Subdural haemorrhage with bilateral retinal haemorrhage following accidental household trauma. *Developmental Medicine and Child Neurology* **43**(Suppl. 90), 24.

[34] Rooks, V.J., Eaton, J.P., Ruess, L., Petermann, G.W., Keck-Wherley, J. & Pedersen, R.C. (2008). Prevalence and evolution of intracranial hemorrhage in asymptomatic term infants. *American Journal of Neuroradiology* **29**(6), 1082–1089.

[35] Whitby, E.H., Griffiths, P.D., Rutter, S., Smith, M.F., Sprigg, A., Ohadike, P., Davies, N.P., Rigby, A.S. & Paley, M.N. (2004). Frequency and natural history of subdural haemorrhages in babies and relation to obstetric factors. *Lancet* **363**(9412), 846–851.

[36] Pollina, J., Dias, M.S., Li, V., Kachurek, D. & Arbesman, M. (2001). Cranial birth injuries in term newborn infants. *Pediatric Neurosurgery* **35**(3), 113–119.

[37] Towner, D., Castro, M.A., Eby-Wilkens, E. & Gilbert, W.M. (1999). Effect of mode of delivery in nulliparous women on neonatal intracranial injury. *New England Journal of Medicine* **341**(23), 1709–1714.

[38] Chamnanvanakij, S., Rollins, N. & Perlman, J.M. (2002). Subdural hematoma in term infants. *Pediatric Neurology* **26**(4), 301–304.

[39] Cushing, H. & Goodrich, J.T. (2000). Reprint of "Concerning Surgical Intervention for the Intracranial Hemorrhages of the New-born" by Harvey Cushing, M.D. 1905. *Child's Nervous System* **16**(8), 484–492.

[40] Maxeiner, H. (2001). Demonstration and interpretation of bridging vein ruptures in cases of infantile subdural bleedings. *Journal of Forensic Sciences* **46**(1), 85–93.

[41] Volpe, J.J. (1995). Intracerebellar haemorrhage, in *Neurology of the Newborn*, WB Saunders, p. 384–389.

[42] Craig, W. (1938). Intra cranial haemorrhage in the newborn - a study of diagnosis and differential diagnosis based on pathological and clinical findings in 126 cases. *Archives of Disease in Childhood* **13**, 89–124.

[43] Larroche, J.-C. (1977). Lesions of haemorrhagic type, mainly venous, in *Developmental Pathology of the Neonate*, Excerpta Medica, Amsterdam, p. 355–398.

[44] Haines, D.E., Harkey, H.L. & al Mefty, O. (1993). The "subdural" space: a new look at an outdated concept. *Neurosurgery* **32**(1), 111–120.

[45] Tubbs, R.S., Louis, R.G., Acakpo-Satchivi, L. & Salter, E.G. (2007). Anatomy of the falcine venous plexus. *Journal of Neurosurgery* **107**, 155–157.

[46] Friede, R.L. (1989). Hemorrhages in Asphyxiated Premature Infants, in *Developmental Neuropathology*, R. Friede, ed, Springer-Verlag, Gottingen, p. 44–58.

[47] Walsh, F.B. & Hedges, T.R. Jr. (1951). Optic nerve sheath hemorrhage. *American Journal of Ophthalmology* **34**(4), 509–527.

[48] Geddes, J.F., Tasker, R.C., Hackshaw, A.K., Nickols, C.D., Adams, G.G., Whitwell, H.L. & Scheimberg, I. (2003). Dural haemorrhage in non-traumatic infant deaths: does it explain the bleeding in 'shaken baby syndrome'? *Neuropathology and Applied Neurobiology* **29**(1), 14–22.

[49] Smith, C., Bell, J.E., Keeling, J.W. & Risden, R.A. (2003). Dural haemorrhage in nontraumatic infant deaths: does it explain the bleeding in 'shaken baby syndrome'? Geddes JE *et al.* A response. *Neuropathology and Applied Neurobiologyeuropathol Appl Neurobiol* **29**(4), 411–412.

[50] Cohen, M.C. & Scheimberg, I. (2008). Evidence of occurrence of intradural and subdural hemorrhage in the perinatal and neonatal period in the context of hypoxic ischemic encephalopathy. An observational study from two referral institutions in the United Kingdom, *Pediatric Developmental Pathology* Nov 13, 1.

[51] Friede, R.L. (1989). Subdural haematomas, hygromas and effusions, in *Developmental Neuropathology*, R. Friede, ed, Springer-Verlag, Gottingen, p. 198–208.

[52] Bruce, D.A. & Zimmerman, R.A. (1989). Shaken impact syndrome. *Pediatric Annals* **18**(8), 482–489, 492.

[53] Aoki, N. & Masuzawa, H. (1984). Infantile acute subdural hematoma. Clinical analysis of 26 cases. *Journal of Neurosurgery* **61**(2), 273–280.

[54] Loh, J.K., Lin, C.L., Kwan, A.L. & Howng, S.L. (2002). Acute subdural hematoma in infancy. *Surgical Neurology* **58**(3–4), 218–224.

[55] Hwang, S.K. & Kim, S.L. (2000). Infantile head injury, with special reference to the development of chronic subdural hematoma. *Child's Nervous System* **16**(9), 590–594.
[56] Ito, H., Yamamoto, S., Komai, T. & Mizukoshi, H. (1976). Role of local hyperfibrinolysis in the etiology of chronic subdural hematoma. *Journal of Neurosurgery* **45**(1), 26–31.
[57] Markwalder, T.M. (1981). Chronic subdural hematomas: a review. *Journal of Neurosurgery* **54**(5), 637–645.
[58] Rogers, C.B., Itabashi, H.H., Tomiyasu, U. & Heuser, E.T. (1998). Subdural neomembranes and sudden infant death syndrome. *Journal of Forensic Sciences* **43**(2), 375–376.
[59] Vinchon, M., Noule, N., Tchofo, P.J., Soto-Ares, G., Fourier, C. & Dhellemmes, P. (2004). Imaging of head injuries in infants: temporal correlates and forensic implications for the diagnosis of child abuse. *Journal of Neurosurgery* **101**(1 Suppl), 44–52.
[60] Yamaguchi, S., Hida, K., Akino, M., Yano, S. & Iwasaki, Y. (2003). Spinal subdural hematoma: a sequela of a ruptured intracranial aneurysm? *Surgical Neurology* **59**(5), 408–412.
[61] McNeely, P.D., Atkinson, J.D., Saigal, G., O'Gorman, A.M. & Farmer, J.P. (2006). Subdural hematomas in infants with benign enlargement of the subarachnoid spaces are not pathognomonic for child abuse. *AJNR. American Journal of Neuroradiology* **27**(8), 1725–1728.
[62] Hylton, C. & Goldberg, M.F. (2004). Images in clinical medicine. Circumpapillary retinal ridge in the shaken-baby syndrome. *New England Journal of Medicine* **351**(2), 170.
[63] Paterson, C.R. (2003). Clinical Review Radiological features of the brittle bone diseases. *Journal of Diagnostic Radiography and Imaging* **5**(1), 39–45.
[64] Thach, B.T. (2000). Sudden infant death syndrome: can gastroesophageal reflux cause sudden infant death? *American Journal of Medicine* **108**(Suppl 4a), 144S–148S.
[65] Spitzer, A.R., Boyle, J.T., Tuchman, D.N. & Fox, W.W. (1984). Awake apnea associated with gastroesophageal reflux: a specific clinical syndrome. *Journal of Pediatrics* **104**(2), 200–205.
[66] Miller, M.J. & Kiatchoosakun, P. (2004). Relationship between respiratory control and feeding in the developing infant. *Seminars in Neonatology* **9**(3), 221–227.
[67] Geddes, J.F. & Talbert, D.G. (2006). Paroxysmal coughing, subdural and retinal bleeding: a computer modelling approach. *Neuropathology and Applied Neurobiology* **32**(6), 625–634.
[68] Prange, M.T., Coats, B., Duhaime, A.C. & Margulies, S.S. (2003). Anthropomorphic simulations of falls, shakes, and inflicted impacts in infants. *Journal of Neurosurgery* **99**(1), 143–150.
[69] Cory, C.Z. & Jones, B.M. (2003). Can shaking alone cause fatal brain injury? A biomechanical assessment of the Duhaime shaken baby syndrome model. *Medicine Science and the Law* **43**(4), 317–333.
[70] Duhaime, A.C. & McNamara, D. (2007). Child abuse injuries: new data challenge beliefs. *Pediatric News* **41**(3), 32.
[71] Prange, M. & Myers, B. (2003). Pathobiology and biomechanics of inflicted childhood neurotrauma response, in *Inflicted Childhood Neurotrauma*, R. Reece & C. Nicholson, eds, AAP Monograph, p. 237–243.
[72] Johnston, R.A. (2004). Paediatric spinal injuries. *Injury* **35**(2), 105–106.
[73] Prange, M.T., Newberry, W., Moore, T., Peterson, D., Smyth, B. & Corrigan, C. (2007). Inertial neck injuries in children involved in frontal collisions, *Presented at the Society of Automotive Engineers World Congress and Exhibition*, April 2007, Detroit, MI. Session: Biomechanics (Part 1 of 3).
[74] Winter, S.C., Quaghebeur, G. & Richards, P.G. (2003). Unusual cervical spine injury in a 1 year old. *Injury* **34**(4), 316–319.
[75] Case, M.E., Graham, M.A., Handy, T.C., Jentzen, J.M. & Monteleone, J.A. (2001). Position paper on fatal abusive head injuries in infants and young children. *American Journal of Forensic Medicine and Pathology* **22**(2), 112–122.

Related Articles

Autopsy

Battered Child Syndrome

WANEY SQUIER

Shoe Impressions: Comparison of

see Footwear and Foot Impressions: Foot Impressions and Linking Foot to Shoe

Shooting Distance: Determination

see Shooting Distance: Estimation of

Shooting Distance: Estimation of

Introduction

The range from which a weapon has been fired is an important component in the reconstruction of firearm-related cases (*see* also **Firearms: Scene Investigation**). The firing-distance estimation is based on the examination of the appearance of the bullet entrance hole and the examination of the firearm discharge residue (FDR) patterns around the hole using various techniques. In casework the patterns obtained in a case are compared to those obtained in test firings. Although many authors in the field use the term shooting-distance determination, many forensic examiners around the world, including the Israel Police, prefer to use the term estimation instead of determination because of the intrinsic inaccuracy of the examination. The reason for this is a high variability of the FDR patterns from shot to shot when using the same weapon and ammunition. To increase the accuracy of the examination, the forensic examiner should use, for test firings, the weapon and ammunition used in the case whenever possible.

In most of the shooting cases in which there is a need for a firing-distance estimation, the victim or the victim's clothing have to be examined. In many cases, bullets hit surfaces of various parts of the human body directly without passage through any intermediate medium. In some instances other exhibits which happened to be targets of shooting have to be examined. Such exhibits may be cars, walls, doors, windows, furniture, etc. Many of these cannot be processed in the laboratory.

FDR is projected from the muzzle of a firearm in a roughly conical pattern: the larger particles travel higher distances than the smaller ones before they are stopped by the air resistance [1]. Four ranges for shooting distance as described in the literature [2–5] are based on the appearance of the bullet holes on the human body: contact range, near contact range, intermediate range and distant range. In contact wounds, the muzzle of the firearm is held against the surface of the body at the time of discharge. The appearance of tearing, scorching, soot, or the imprint of muzzle, characterizes contact wounds. Virtually, no FDR is seen around the bullet hole. In near contact wounds, the muzzle of the weapon is a few centimeters away from the body. In this range, a wide zone of powder soot overlaying seared blackened skin surrounds the wound. An intermediate range gunshot wound is one in which the muzzle of the firearm is held a few tens of centimeters away from the body, producing "powder tattooing" of the skin. In distant range no damage effects or FDR particle patterns are observed around the gunshot wound. With most handguns and ammunitions, visually detectable FDR is not found in the case of shots fired at ranges greater than 30–45 cm [1].

In this text a review of the methods for visual/microscopic, color tests and instrumental analysis of the entrance bullet holes and FDR patterns around them for shooting-distance estimation is provided. Comprehensive reviews on the subject have been published over the years [2–5]. The aspect of shooting-distance estimation for targets shot by pellet loads from shotguns may be found elsewhere [1, 2] and is not discussed here.

Visual/ Microscopic and Color Tests

Clothing Targets

FDR patterns around the entrance bullet holes consist of propellant residues and metallic residues from the bullet, e.g., lead and copper, as well as gunshot residues (GSR) (primer residues). These residues may be detected visually/microscopically if the target cloth is of a bright enough color. However in most of the cases there is need for color chemical tests or instrumental analysis in order to assess the FDR patterns around the entrance bullet holes.

Walker [6] proposed to use Griess reaction to visualize free nitrite ions (on the shot target), originating from the combustion of ester nitrates (e.g., nitrocellulose (NC) and nitroglycerine (NG) in the gunpowder. In this test, the Griess reagent consisting of sulfanilic acid and α-naphthyl amine in acetic acid aqueous solution is used. The detection of nitrite ions is based on the formation of diazonium ion from sulfanilic acid and nitrite. The diazonium ion couples with α-naphthyl amine to form an orange azo dye.

In a series of three papers [7–9], Dillon reports on the modified Griess test (MGT) as a color test for nitrites and recommends a protocol for FDR examinations in muzzle to target distance estimations. In the modified test, Dillon proposes to use

α-naphthol instead of α-naphthyl amine (the Walker test) or *N*-(1-naphthyl)-ethylenediamine dihydrochloride. According to him both the replaced reagents are carcinogenic. However, in the literature on chemical safety data [10], *N*-(1-naphthyl)-ethylenediamine dihydrochloride is not reported as being carcinogenic. In fact, in Israel this reagent is used with sulfanilamide routinely for "the local" MGT [11]. The term MGT does not refer to one specific defined test. In fact, it appears that every author who introduces any modification to the original Griess test calls it an MGT. The proposed protocol [7] includes visual, microscopic and chemical (lead and nitrites) tests. It recommends to conduct first the MGT and then the sodium rhodizonate test (SRT) for lead. The reason for this sequence is because the rhodizonate is applied directly on the target. Dillon contends that particulate lead is a random nonreproducible phenomenon, whereas the presence of vaporous lead is quite significant in that it is found principally at closer ranges.

Glattstein *et al.* reported on an improved method for shooting-distance estimation on clothing [11]. The novel part of the method includes transfer of total nitrite (nitrite ions and unburned smokeless powder residues) from the target to an adhesive lifter. After the transfer, lead and copper deposits around the bullet entrance hole are partially extracted consecutively to the Benchkote (Whatman) filter papers moistened with dilute acetic acid and ammonia solutions respectively. Their patterns are visualized by rhodizonate for lead (red color) and rubeanic acid for copper (dark green color). The MGT is carried out after alkaline hydrolysis of the smokeless powder residues on the adhesive lifter. The purpose of lifting gunpowder residues from the shot cloth target is to eliminate interferences caused by conducting MGT directly on the target, with or without the hydrolysis step. It was found that almost a complete transfer of gunpowder residues to the adhesive lifter was obtained and the vaporous lead and copper are not transferred to the adhesive lifter. The widely used MGT detects only free nitrite ions formed from the combustion of smokeless powder. The unburned smokeless powder particles cannot be detected by this method. Alkaline hydrolysis prior to the MGT has been proposed to increase the sensitivity of the test for gunpowder residues. The purpose of the alkaline hydrolysis is to cause disproportionation of the unburned NC and NG to carbonyl compounds and free nitrite ion, thus increasing the available amount of nitrite ions for MGT.

Before starting the estimation of the shooting distance it is desirable to determine that the hole is a bullet entrance hole. This can be done by applying methods for chemical visualization of lead (rhodizonate) and copper (rubeanic acid) at the perimeter of the hole [12]. From the accumulated experience in the Israel Police it has been observed that, the color tests did not give positive results on all bullet holes in clothing, although it was known that a lead bullet or a full metal jacket (FMJ, brass) bullet were used.

Persistence of FDR on Clothing Targets

Several studies dealt with possible effects of various factors on clothing items after shooting with regards to the shooting-distance estimation [7, 13–16]. Most of these found that mechanical handling of clothing or soaking them in blood, in still or running water, considerably decreases the amount of FDR around the bullet entrance holes. Emonet *et al.* [16] reported that the medical manipulations of clothing lead to an increase of the loss of visible and nitrated FDR of about 30–40%. Even *et al.* [13], on the other hand did not find a significant effect of soaking in still water on the obtained FDR patterns. Sometimes casework requests are received to estimate shooting distance on clothing items that underwent machine washing. Vinokurov *et al.* [17] conducted a study to assess the effect of machine washing or brushing of clothing items on FDR patterns around bullet entrance holes. Results show that those treatments considerably decrease the amount and density of FDR (machine washing more than brushing).

Exhibits that Cannot be Processed in the Laboratory

Glattstein *et al.* [18] examined the feasibility of the method developed for clothing [11] as described above, for additional materials such as galvanized steel, glass, plywood, and high pressure laminated plastic sheets of melamine and phenolic materials (Formica). It was found that for the above tested target materials and shooting distances the amounts and densities of the FDR detected visually (without any treatment) were considerably smaller than those obtained after chemical treatments. Total nitrite

patterns visualized on the lifters applied on the various targets were similar to those obtained on the lifters from the cotton cloth at relatively short shooting distances, i.e., up to about 25 cm. As shooting distances become greater, the number and density of nitrite spots on the lifters from all the tested materials targets decreases considerably in comparison to the lifters from the cotton cloth for the same distance; the plywood target showed the most similar results to the cotton cloth.

If there is no possibility to conduct test firing at a material similar to the evidence, then test firings may be carried out on cotton cloth. In such a case the visualized pattern of the total nitrite will be sufficient to state that the shooting distance on the evidence was equal or below the shooting distance at which similar visualized patterns of the total nitrite are obtained on the cotton cloth.

The Human Body as a Target

In many shooting cases, bullets hit surfaces of various parts of the human body (mostly the head) directly, without passage through any intermediate medium. For the purpose of assessing the shooting distance, most of the forensic literature describes only visual/microscopic methods for the examination of appearance of the wound and FDR patterns around it [2, 4, 19]. Nonetheless, sodium rhodizonate and rubeanic acid reagents were proposed for the visualization of lead and copper patterns around the gunshot wounds [20, 21]. As in cases of clothing and other objects, implementing color tests (in addition to visual/microscopic visualization) might increase the accuracy of shooting-distance estimation. An additional unique problem pertaining only to the human body is that there is no possibility to conduct test firing on the same material as in the case of other exhibits. The proposed solution for test firings was to use various simulant materials. Recommendations for those materials were based on the studies comparing those materials to the skin of some animals, like rabbits or pigs.

Glattstein *et al.* [22] examined the feasibility of applying an adhesive lifter to the entrance bullet wound in human body surfaces to visualize the total nitrite patterns, as was reported for clothing and other exhibits above. Stahling and Karlsson [23] reported a similar method for lifting and visualizing gunpowder residues from skin.

Instrumental Methods

FDR patterns around the bullet hole may be visualized by infrared (IR) photography [24] and IR imaging devices [25] due to the considerably higher absorbance of IR radiation by the soot than by other materials like fabric. X-ray radiography may also be used for that purpose because of the much higher absorbance of X-radiation by metallic deposits of the FDR than by clothing. X-ray fluorescence (XRF), atomic absorption spectroscopy (AAS) and neutron activation analysis (NAA) have also been used by some laboratories to estimate the range of shooting [3].

Recently, with the advent of the Micro-XRF technology and its increasing use in forensic science for the elemental analysis of trace evidence, some laboratories examined its feasibility for shooting-distance estimation [26].

References

[1] Rowe, W.F. (2005). Firearms identification, in *Forensic Science Handbook*, 2nd Edition, R. Saferstein, ed., Pearson Prentice Hall, Vol. 2.

[2] Sellier, K. (1991). *Shot Range Determination*, *Forensic Science Progress 6*, Springer-Verlag, Berlin, Heidelberg.

[3] Lichtenberg, W. (1990). Method for the determination of shooting distance, *Forensic Science Review* **2**, 38–62.

[4] Zeichner, A. & Glattstein, B. (2002). Recent developments in methods of estimating shooting distance, *The Scientific World Journal* **2**, 573–585.

[5] Di Maio, V.G.M. (1999). *Gunshot Wounds: Practical Aspects of Firearms, Ballistics and Forensic Techniques*, CRC Press.

[6] Walker, J.T. (1940). Bullet holes and chemical residues in shooting cases, *Journal of Criminal Law and Criminology* **31**, 497.

[7] Dillon, J.H. (1990). The modified Griess test: a chemically specific chromophoric test for nitrite compounds in gunshot residues, *AFTE Journal* **22**, 243–250.

[8] Dillon, J.H. (1990). The sodium rhodizonate test: a chemically specific chromophoric test for lead in Gunshot Residues, *AFTE Journal* **22**, 251–256.

[9] Dillon, J.H. (1990). A protocol for gunshot examination in muzzle to target distance determination, *AFTE Journal* **22**, 257–274.

[10] (1985). *The Sigma-Aldrich Chemical Company, Inc., Material Safety Data Sheet for N-(1-Naphthyl) ethylenediamine dihydrochloride*, November 2001–January 2002.

[11] Glattstein, B., Vinokurov, A., Levin, N. & Zeichner, A. (2000). Improved method for shooting distance estimation. Part I. Bullet holes in clothing items, *Journal of Forensic Sciences* **45**, 801–806.

[12] Steinberg, M., Leyst, Y. & Tassa, M. (1984). A new field kit for bullet hole identification, *Journal of Forensic Sciences* **29**, 169–176.
[13] Even, H., Bergman, P., Springer, E. & Klein, A. (1988). The effects of water-soaking on firing distance estimations, *Journal of Forensic Sciences* **32**, 319–327.
[14] Haag, L.C. (1991). A method for improving the Griess and sodium rhodizonate tests for GSR patterns on bloody garments, *AFTE Journal* **23**, 808–815.
[15] Bonfanti, M. & Gallusser, A. (1995). Problems encountered in the detection of gunshot residues, *AFTE Journal* **27**, 105–122.
[16] Emonet, F., Bonfanti, M. & Gallusser, A. (1999). Etude des phenomes physique affectant les residues de tir et engenders lors de la manipulation des habits par le personnel medical, *Canadian Society of Forensic Science Journal* **32**, 1–13.
[17] Vinokurov, A., Zeichner, A., Glattstein, B., Levin, N., Koffman, A. & Rozengaten, A. (2001). Machine washing or brushing of clothing and the influence on shooting distance estimation, *Journal of Forensic Sciences* **46**, 928–933.
[18] Glattstein, B., Zeichner, A., Vinokurov, A. & Shoshani, E. (2000). Improved method for shooting distance estimation. Part II. Bullet holes in objects that cannot be processed in the laboratory, *Journal of Forensic Sciences* **45**, 1000–1008.
[19] Heard, B.J. (1997). *Handbook of Firearms and Ballistics: Examining and Interpreting Forensic Evidence*, John Wiley & Sons, Chicester, West Sussex, England.
[20] Stone, I.C. & Petty, C.S. (1991). Interpretation of unusual wounds caused by firearm, *Journal of Forensic Sciences* **36**, 736–740.
[21] Molchanov, V.I., Popov, V.L. & Kalmykov, K.N. (1990). *Gunshot Wounds and their Forensic Medicine Examination (in Russian)*, Meditzina, Leningrad.
[22] Glattstein, B., Zeichner, A., Vinokurov, A., Levin, N., Kugel, C. & Hiss, Y. (2000). Improved method for shooting distance estimation. Part III. Bullet holes in cadavers, *Journal of Forensic Sciences* **45**, 1243–1249.
[23] Stahling, S. & Karlsson, T. (2000). A method for collection of gunshot residues from skin and other surfaces, *Journal of Forensic Sciences* **45**, 1299–1302.
[24] Eastman Kodak Company (1972). *Applied Infrared Photography*, Kodak Publication No. M-28, Eastman Kodak Co., Rochester, NY.
[25] Abrink, A., Andersson, C. & Maehly, A.C. (1984). A video system for the visualization of gunpowder patterns, *Journal of Forensic Sciences* **29**, 1223–1224.
[26] Flynn, J., Stoilovic, M., Lennard, C., Prior, I. & Kobus, H. (1998). Evaluation of X-ray microfluorescence spectrometry for the elemental analysis of firearm discharge residues, *Forensic Science International* **97**, 21–36.

ARIE ZEICHNER

Short Tandem Repeats

Introduction

Short tandem repeats (STRs) or microsatellites are discrete sequences of DNA that are repeated end on end. STR repeats in the human genome are analogous to carriages on a train. They can be repeated up to 100 times in tandem in the genome and have fragment lengths of between 100 and 400 bp [1]. Figure 1 shows the basic structure of a simple repeat STR.

The repeated DNA sequence of an STR consists of 2–5 nucleotides as di-, tri-, tetra-, and pentanucleotide repeats. More than 23 000 tetranucleotide STR loci alone have been characterized since the completion of the human genome project [2]. Tetranucleotide repeats are the most common STRs used by forensic laboratories throughout the world. Trinucleotide and pentanucleotide loci are less common. Dinucleotide loci are the most common in the human genome; however, as they are more prone to artifacts that affect interpretation, they are not used for forensic purposes [3].

STRs used for human identity testing are located in the noncoding regions of the DNA, either within genes (introns) or between genes. They have been adopted universally by forensic biology laboratories as they are short and highly polymorphic, distributed throughout the genome, and are amenable to polymerase chain reaction (PCR) and automation.

STRs are analyzed using the PCR. Primers of known sequence bind to either side of the target STR sequence (see Figure 1). Using PCR, STR profiles may be generated from as little as 0.2 ng of input DNA.

Examples of Common STRs Used for Forensic Applications

The selection criteria for STRs in forensic use are based on their robustness, size, variability between individuals (polymorphic), and ability to be combined in a single test for ease of analysis (known as *multiplexing*).

Table 1 contains a list of STRs commonly used by the forensic community for DNA analysis. This list is not intended to be exhaustive. Table 1 also includes core loci for the combined DNA index

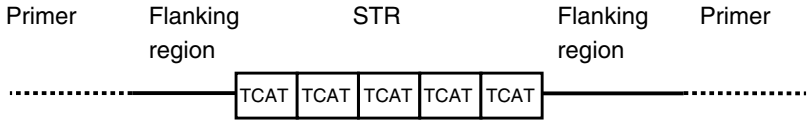


Figure 1 Diagram of an STR sequence TCAT containing five repeats

Table 1 A list of commonly used forensic STR loci and those included in the Interpol core set of loci and CODIS

			Applied Biosystems profiling kits				Promega Corp	
STR loci	CODIS	Interpol	Identifiler™	Profiler Plus™	SGMPlus™	Sefiler™	PowerPlex®16	PowerPlex® ES
CSF1PO	E		E				E	
D2S1338			E		E	E		
D3S1358	E	E	E	E	E	E	E	E
D5S818	E		E	E			E	
D7S820	E		E	E			E	
D8S1179	E	E	E	E	E	E	E	E
D13S317	E		E	E			E	
D16S539	E		E		E	E	E	
D18S51	E	E	E	E	E	E	E	E
D19S433			E		E	E		
D21S511	E	E	E	E	E	E	E	E
FGA	E	E	E	E	E	E	E	E
TH01	E	E	E		E	E	E	E
TPOX	E		E				E	
vWA	E	E	E	E	E	E	E	E
Penta D							E	
Penta E							E	
ACTBP2 (SE33)						E		E
Amelogenin			E	E	E	E	E	E

system (CODIS) and Interpol databases (discussed later).

Although not an STR, amelogenin is discussed in this work as it is routinely added to commercial STR assays for gender determination. Amelogenin is a gene that occurs on the X and Y chromosomes of the human genome [4]. The PCR primer set targets a 6 bp deletion that occurs on the X chromosome. This enables the X and Y chromosomes to be distinguished after electrophoresis. Female individuals are designated XX and males XY (see Figure 2). Amelogenin can also be useful for determining the ratio of male-to-female DNA in a mixed DNA profile.

There are several genetic variations at amelogenin that may cause difficulties in interpretation. In one, the Y chromosome fails to amplify, making the individual genotypically female [5]. Other amelogenin abnormalities that may impede interpretation include trisomic states (XXY, XYY, and XXX) and rare genetic disorders that may lead to differences of genotype and phenotype [6].

Nomenclature

Designation of Loci

An internationally accepted nomenclature for the designation of STRs is vital for interlaboratory comparison and databasing [7, 8].

STR sites analyzed on the DNA are designated on the basis of their position within the genome. Markers that are part of a gene have its name within the designation. For example, STR marker TH01 is located within intron 1 of the *t*yrosine *h*ydroxylase gene. STR markers that are not part of a gene are designated by the following pattern:

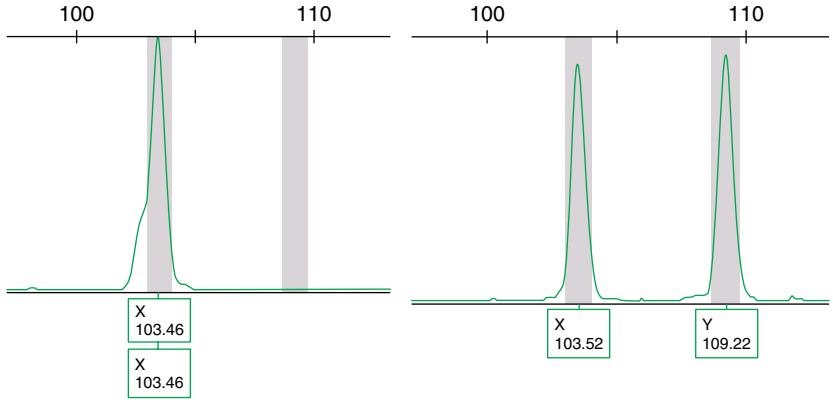


Figure 2 Amelogenin loci of a female (XX) and male (XY) individual

DxSy where D is DNA, x is the chromosome number, S indicates that the sequence only occurs once in the genome (single), and y is a counter for when the site was identified. For example, D21S11 describes the 11th site discovered on chromosome 21.

STR loci can also be preceded by the letters HUM, which indicate their human origin.

The repeat motif for each STR is defined according to the International Society of Forensic Genetics (ISFG) recommendations for nomenclature as being the first 5′ nucleotides on the GenBank[a] forward strand [9].

Designation of Alleles

Simply, STR alleles are designated by the number of repeats they contain. Repeats are designated on the basis of comparison with allelic ladders that incorporate all common alleles. STRs can be categorized by the length of their repeat unit (di-, tri-, tetra-, and pentanucleotide) and also by the type of repeat pattern they conform to. Different classes of STRs have been identified, depending on their repeat structure.

Simple repeats contain core sequences identical in sequence and length (for example, TH01). Compound repeats contain two or more adjacent simple repeats (for example, VWFA31). Complex repeats may contain several repeat blocks of variable repeat length with variable intervening sequences (for example, D21S11) [10].

Common examples of simple, compound, and complex STRs and their repeat structures are given below. The numbers outside the brackets surrounding the repeat unit sequence give an indication of the number of times the sequence may repeat.

Simple repeat, for example, TH01: $(AATG)_{5-11}$

Compound repeat, for example, VWFA31: $(ATCT)_2(GTCT)_{3-4}(ATCT)_{9-13}$

Complex repeat, for example, D21S11: $(TCTA)_{4-6}(TCTG)_{5-6}(TCTA)_3TA(TCTA)_6TCA(TCTA)_2TCCATA(TCAT)_{8-16}$

Complex hypervariable repeats are another, more complex, type of STR. A well-known example is ACTBP2 (also referred to as *SE33*). The common tetranucleotide repeat motif is $(AAAG)_n$; however, variants mono, di-, tri-, and tetranucleotides are also scattered through the locus. This makes ACTBP2 very polymorphic and for this reason it is a required STR for the German national DNA databank. However, the designation of alleles within this locus is problematic if based on solely counting repeats. For this reason, the size of the STR is measured and it is recommended that alleles be labeled "type-" to reflect this [7].

Microvariants

Where alleles at an STR locus do not contain complete repeat units, these are termed microvariants. The microvariants are designated by the number of complete units and the number of additional base pairs of the partial repeat. The most common microvariant is allele 9.3 at TH01. TH01 9.3 is actually 10 repeat units with the loss of a single adenine in the seventh unit [11]. The repeat motif for TH01 9.3 is $(AATG)_6ATG(AATG)_3$.

STR Analysis in Practise

Visualization of STRs

STRs are amplified using the polymerase chain reaction, which exponentially amplifies the target STR by a series of heating and cooling reactions. The size of the STR products is measured after PCR amplification by electrophoresis. Early STR analysis involved separating the PCR product on polyacrylamide gels followed by visualization using silver staining. Now, PCR products are labeled with fluorescent markers, which then undergo gel electrophoresis or, more commonly, capillary electrophoresis. The PCR product is loaded on the gel or capillary and subjected to an electrical charge. The negatively charged DNA travels through the medium dependent on its size. Smaller fragments of DNA travel faster through the gel or capillary. The actual size of the STR fragments is measured against two standard markers: an internal size standard that is injected with the sample and an allelic ladder that comprises common alleles with the STR loci.

The DNA fragments are labeled with fluorescent probes. As the DNA travels through the gel or capillary, the probes are excited by a laser and the fluorescence is measured by a camera. The use of different color dyes enables multiple, overlapping loci to be analyzed concurrently, which is important for multiplexed STR assays. STR profiles result in a string of numbers (the allele counts or number of repeat units for STR analyzed) associated with the loci analyzed. For this reason, the data can be easily stored and compared within databases.

Multiplexes and Commercial STR Assays

The primers targeting several STRs may be combined into one PCR reaction and amplified simultaneously. This is called *multiplex PCR* and is a rapid and convenient way to profile many STRs, thus generating a DNA profile. STRs present in multiplexed assays are chosen on the basis of their discriminating power, low levels of stutter, and compatibility with other loci. The choice of STR in commercial multiplexed assays can also be heavily influenced by demand from customers who must adhere to core loci, for example, CODIS and Interpol (discussed later).

Several different STR assays or multiplexes are commercially available for use in forensic DNA analysis and paternity or identity testing. These are used widely by the forensic community as they have been extensively validated and scrutinized in the courts of numerous countries, making them universally accepted. The use of commercial STR assays also promotes uniformity and enables database conformity.

Not all laboratories, however, use commercial STR assays. Some have developed in-house assays that may or may not have loci in common with commercial assays [12]. Indeed, even the commercially available kits do not share all loci in common with other laboratories. Two major suppliers of forensic DNA profiling kits and technology are Applied Biosystems, Foster City, California and Promega Corporation, Madison, Wisconsin. Both companies have developed several STR assays for forensic STR analysis (Table 1).

The development and optimization of STR multiplexes can be very demanding and both labor intensive and time consuming [13]. For this reason, many laboratories have opted to use commercial assays where validation papers and population databases are freely available. The quality assurance requirements for developmental validation are rightly very stringent.

Both Promega and Applied Biosystems have developed 15 loci STR multiplexes with amelogenin. Both kits incorporate all 13 CODIS loci and Interpol's standard set of loci. Initially, two STR assays were required to test all 13 CODIS loci: Promega's PowerPlex® 1.1 and PowerPlex® 2.1 or Applied Biosystem's Profiler Plus™ and COfiler™. See Figure 3 for an example of a male DNA profile generated using the 15 STR multiplex Identifiler™. Identifiler™ uses four different dyes to label the loci. The internal size standard is labeled with a fifth color.

Core Loci and Databasing (Interpol and CODIS)

STRs were first described as being potentially useful markers for forensic DNA analysis in the early 1990s [1]. The first national DNA database using STR profile data was started in the United Kingdom in 1995. This was followed by the establishment of the New Zealand National DNA Database in 1996 [14]. Several other European countries followed suit in the late 1990s. Many countries were analyzing different loci. To aid criminal investigation across the relatively open European borders, an Interpol working group proposed a European standard set of

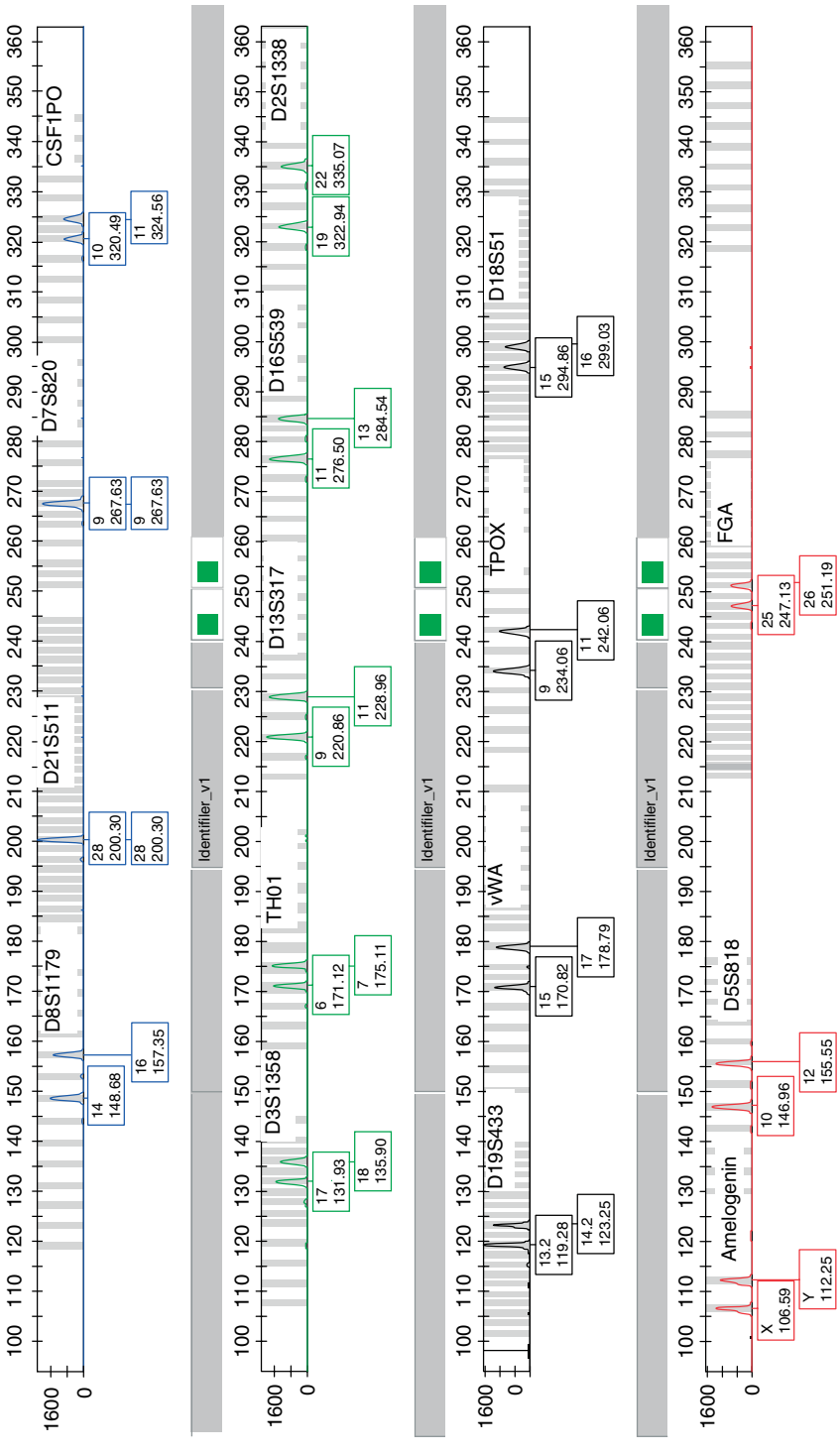


Figure 3 An Identifiler™ profile of a male individual

four loci in 1998. In 1999, this was expanded to seven loci (see Table 1) [15].

In 1997, the United States agreed to 13 core STR loci for their national DNA database, known as *CODIS* (see section "Core Loci and Databasing Interpol and CODIS").

Common sets of STR loci are required for collaboration between different testing laboratories within the same country or between different countries. Comparison of DNA profiles within national or international databases can be used to link individuals to crimes and to generate crime-to-crime links indicating recidivist offending patterns. The more core loci compared, the less likely adventitious links will be generated.

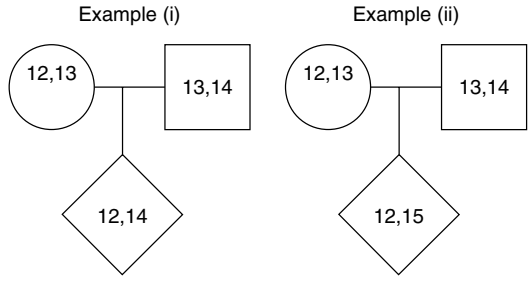


Figure 4 Two pedigrees of mother, father, and child at one loci. In example (i) the child must have inherited the "12" allele from the mother and the "14" from the father. In example (ii) a mutation has occurred most likely due to an insertion one repeat in length and the child has inherited a "15" allele

Linkage to Genetic Diseases

The number of repeats at STR loci is highly variable between individuals as they undergo no selective pressure and are thought to be noncoding. There have been reports, however, of several core STR loci being linked to genetic diseases, including certain TH01 alleles indicating susceptibility to schizophrenia [16]. To date, all claims have been disputed by further studies [17], and no STRs have been removed from the core sets of forensic loci examined.

Mutations and Rare Alleles

Mutation

STR loci are prone to mutation; the reason is their variability amongst individuals and therefore their usefulness in forensic DNA and identity testing. The high mutation rate of STRs is not problematic for routine forensic work. Generally, DNA profiles generated from crime samples are compared directly with reference profiles generated using the same DNA STR assays.

More problematic is parentage testing (paternity or maternity investigations). Mutation rates of STRs can be measured by analyzing many paternity trios (father, mother, and child) and comparing STR alleles. A child must inherit one allele from their father and one from their mother at each locus. This is demonstrated in Figure 4 with an example of two parent/child pedigrees at one STR locus. The first shows an expected pedigree and the second, a pedigree with mutation in the male germ line.

Brinkmann *et al.* investigated 10 844 parent/child sets at nine STRs used for paternity testing and reported a mutation rate of between 0 and 7×10^{-3} for the loci tested. The number of mutations in the male germ line is 5–6 times higher than the maternal germ line. Generally, the more polymorphic (variable) and complex STRs exhibit the most mutation [18]. All mutations observed in this study were due to either whole repeat loses or repeat gains (22 involving single unit repeats and 1 example of a double repeat). The stepwise mutation model has been theorized to explain how STR alleles mutate by an increase or decrease on one repeat unit [19]. The likelihood of mutation must be taken into account when investigating paternity cases as false exclusions may be possible on account of differences of only one or two alleles.

Mutation rates and their mechanisms are discussed in more detail in [20] and [21].

Other types of genetic mutations that give rise to microvariants are insertions and deletions of single nucleotides. The resulting allele often differs from full repeat units present in the allelic ladder by only a few nucleotides. These alleles are therefore sized as being "off-ladder" and many laboratories refer to them as being "rare" alleles. One example at the FGA locus is given in Figure 5.

Primer Binding Site Mutations

Discordant genotype profiles may result from the differences in primer sequences used between STR

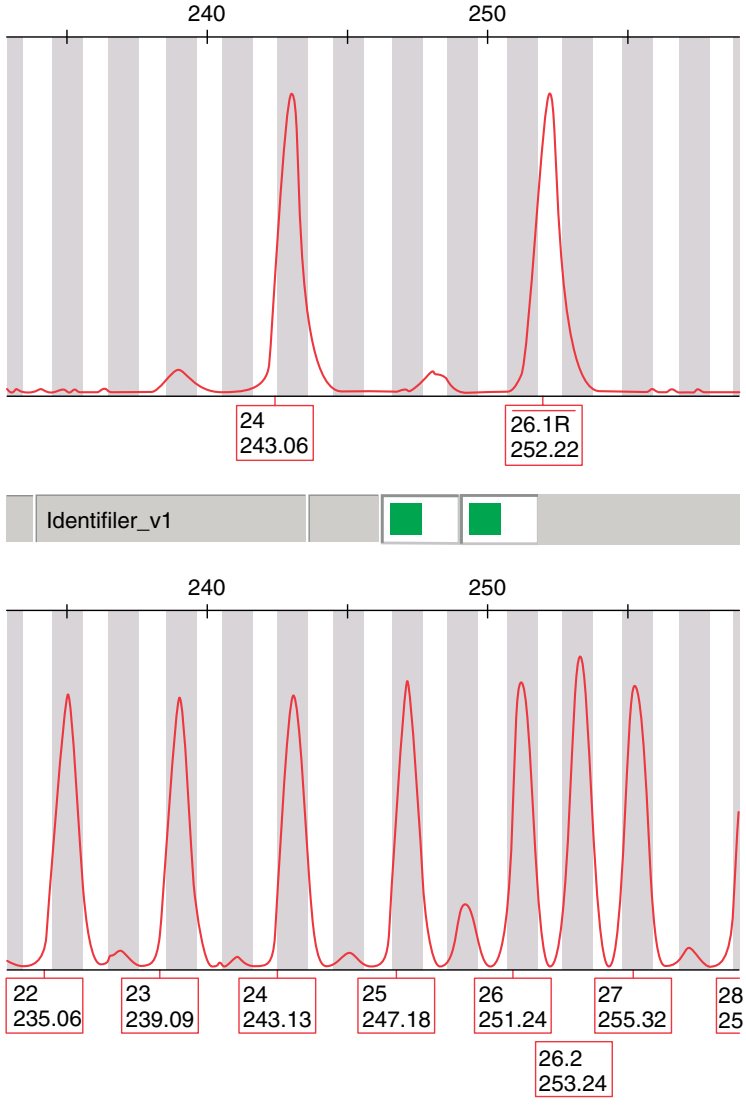


Figure 5 Rare allele at the FGA locus seen after amplification using AmpF/STR Identifiler™

assays produced by different manufacturers and different STR assays produced by the same manufacturer. The PCR amplification of STR alleles with mutations at primer binding sites may result in little or no extension during amplification. Total dropout of one allele at a locus leads to false homozygosity and the missing allele is called a *null allele*. A well-characterized primer binding site mutation is at the D8S1179 locus in the Chamorro (indigenous people of the Marianas Islands) and Filipino populations [22], caused by a single point mutation (G to A transition) in the reverse primer binding sequence.

The high frequency of homozygotes at D8S1179 in the Chamorro and Filipino populations was detected after performing a concordance study between Promega's PowerPlex®16 assay and Applied Biosystem's Profiler Plus™. In their Identifiler™ STR assay, Applied Biosystems includes an additional D8S1179 reverse primer specific for the variant

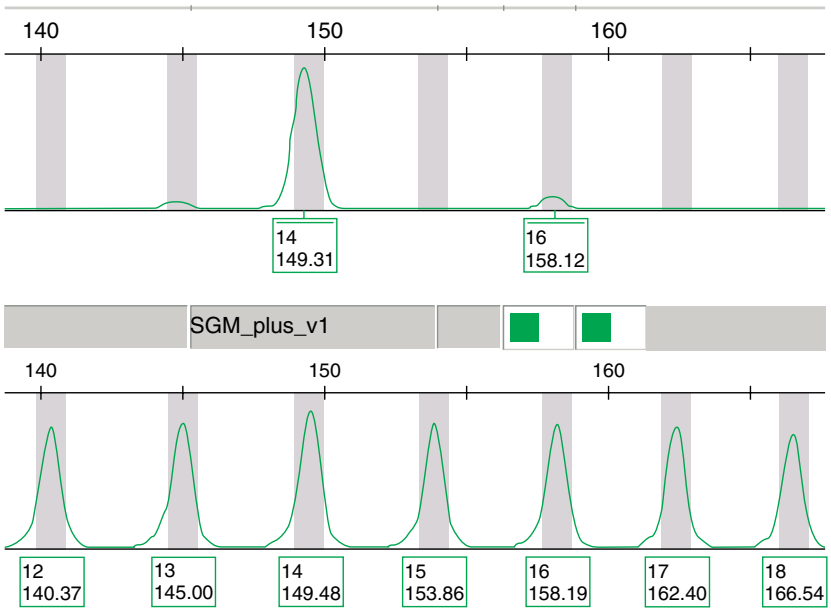


Figure 6 An example of a primer binding site mutation at D8S1179 using AmpF*l*STR SGMPlus™. This individual was subsequently typed as a D8S1179 14,16 using AmpF*l*STR Identifiler™

that results in recovery of the null alleles not previously amplified [23]. In Figure 6, a primer binding site mutation in the reverse primer of D8S1179 has resulted in only partial amplification of the second (16) allele after amplification with SGMPlus™. The amplification of this sample using Identifiler™ resulted in two alleles of equal height.

Another reported example of discordant genotypes between different commercial STR assays was at FGA where a heterozygote profile was observed using the PowerPlex™ 16 primers and a single homozygote allele was observed using the Profiler Plus™ primers [24].

When comparing profiles generated using different assays, for example between different national DNA databases or old crime profiles and reference samples subsequently typed using more recent STR assays, these discordant profiles may become problematic. Database searching can be flexible regarding the stringency with which STR profiles are compared. By incorporating one or more allowed allele mismatches between different profiles, mutations causing discordant profiles due to differing primers within assays can be investigated.

Triallelic Patterns

Triallelic patterns may be observed at individual STR loci. These can arise from trisomy (triplicated chromosomes), somatic mutation, or localized chromosomal rearrangement. There are only three known autosomal trisomies that are thought to be nonfatal; chromosome 21 resulting in Down's syndrome, 13 giving rise to Patau syndrome, and 18 giving rise to Edwards syndrome. All three result in individuals with birth defects, mental retardation, and reduced life expectancy.

The other types of triallelic patterns have been classified into two types: type 1 resulting from somatic mutations in one cell line giving rise to cells with two different genotypes and type 2 resulting from localized chromosomal rearrangement and appearing as three bands of equal height [25].

A high frequency of somatic mutations has been discovered in buccal cells of patients with oral cancer. In one study of 100 patients, 25 allelic nonconcordant profiles were detected following DNA typing of nine STR loci when compared with blood samples taken from the same individuals [26]. As for primer binding site mutations, somatic mutations and

localized chromosomal rearrangements do not complicate forensic investigations when directly comparing crime samples to reference samples. They may, however, make the interpretation of mixed DNA profiles more difficult by confusing the analyst when determining the number of individuals whose DNA may be present in a sample.

Nonstandard STRs

X-Chromosomal STRs

The application of X-chromosomal STRs (ChrX STRs) to forensic DNA analysis is not widely undertaken. Females possess two X chromosomes that recombine similar to autosomes (non sex chromosomes). Males inherit one X chromosome wholly from their mother and females share their paternal X chromosome with their father.

ChrX STR analysis is not as powerful as autosomal STR analysis for forensic identity testing; however, it may be useful for cases where traces of female DNA are present in stains of high male DNA content, for example, vaginal epithelial cells on a penis [27]. ChrX STR analysis is more useful in cases of complex parentage testing where DNA from close family members or the putative parents is not available. One example would be for investigating body remains from mass disaster cases. ChrX STRs are also useful for paternity testing of female offspring (see Case Study 1).

Case Study 1 ChrX STR typing was employed in a case of questioned paternity of a female where reference samples were unavailable for the putative father and paternal grandparents. Samples were available for two paternal "uncles" (putative father's brothers) and the child's mother. Using the ChrX STR types of the "uncles", the grandmaternal genotype was reconstructed. Using this, the putative father was excluded as being the true father as he did not carry some of the necessary paternal ChrX STR alleles [28].

Many ChrX STR loci have been identified and have been shown to be suited for forensic use due to their high variability. A commercial ChrX STR assay with eight ChrX STRs is available (Mentype® ArgusX-8 from Biotype AG) as is a website containing current research, nomenclature, and mutation rates [29].

Y STRs

Y STRs are located on the Y chromosome and are therefore only suited to profiling male DNA (*see* **Y-Chromosome Short Tandem Repeats**). Y STRs are useful in the presence of large amounts of female DNA that would have otherwise masked the male DNA, for example, male epithelial cells on vaginal swabs after digital penetration. A Y STR haplotype is also useful for determining familial relationships. As with ChrX STR analysis, Y STRs are useful in determining deficiency paternity cases, although with male offspring. Y STR profiles are haplotypes as they are noncombining and may be shared by many individuals.

Commercial Y STR assays are available from both Promega Corporation and Applied Biosystems, which test up to 12 and 17 Y STR loci, respectively [30].

miniSTRs

The Identifiler™ and PowerPlex®16 STR assays both have loci with total lengths of over 350 bp. While STRs are relatively small, which contributes to their success in the analysis of forensic samples, most of their lengths can be regions of DNA flanking the repeat units (see Figure 1). This flanking DNA increases the STR length, allowing more loci to fit into commercial STR assays. By removing the flanking sequences and moving the primers to immediately flank the repeat units, miniSTRs are created (*see* **Mini-STRs**). miniSTRs are advantageous as they allow more degraded and lower amounts of DNA to be successfully amplified by the PCR and analyzed. Lengths of miniSTRs can be up to 300 bp smaller than their counterpart PCR products in commercial standard STR assays. Generally, miniSTRs also maintain database compatibility as they target the same STR loci. However, because the primers are redesigned, some discordance is to be expected (see earlier discussion). Another disadvantage is due to the restriction in size of the miniSTR amplification product as this means that less miniSTRs can be squeezed into commercial assays [31].

The European (ENSFI and EDNAP) groups have already recognized the importance and usefulness of miniSTRs. They have recommended that European laboratories maintain their seven core loci and reengineer the primers to convert them to miniSTRs. They also recommend the adoption of three new European core loci, which are miniSTRs that have not been previously seen in major standard commercial STR assays: D10S1248, D2S441, and D22S1045 [32, 33].

Nonhuman STRs for Forensic Investigations

Plant STRs

The most common application of forensic DNA profiling to plant material is for *Cannabis sativa*. Cannabis is widely used as a recreational drug. The cultivation and possession of cannabis is illegal in many countries. However, in some countries its cultivation for fiber and seed oil may be legal. The DNA testing of *Cannabis* can be useful for forensic investigations in species identification, identifying drug versus fiber strains, linking hydroponic seizures, and, more recently, determining the source of origin (whether nationally or internationally) [34].

STR analysis of other plant material, however, has also been reported (see Case Study 2). This case highlights the potential of other plant material in forensic cases.

Case Study 2 STR analysis was undertaken on two Quercus geminata (sand live oak) leaves located in a murder suspect's vehicle. A DNA profile was generated at four STR loci for both of the leaves. This profile was compared with "reference" DNA profiles from several trees located near the burial site of the deceased. The STR profiles from the leaves located in the vehicle were different from samples collected from the trees [35].

Animal STRs

The forensic analysis of animal STRs has become more common in recent years. STR analysis of biological material from animals domestic or otherwise can link an animal to an individual where the animal is the perpetrator (for example, in dog attacks; see Case Study 3) or an individual to an animal where the link places the individual at the scene of a crime (see Case Study 4).

Case Study 3 In Budapest, in 2001, DNA was analyzed from possible saliva stains extracted from the clothing from two deceased infants thought to be killed by the family dog. DNA profiling analysis using a commercially available canine STR multiplex assay (StockMark Kit Canine I Ver.3 from Applied Biosystems) indicated that the DNA profile generated from the possible saliva stains was consistent with the DNA profile from a reference sample collected from the dog [36].

Case Study 4 In a now-famous case, an estranged husband was implicated in the murder of his wife on Prince Edward Island, Canada. Ten dinucleotide STRs were used to test snowy white cat hair found on a jacket located near the scene of the crime. The STR profile obtained corresponded to that of reference hair recovered from Snowball, the cat belonging to the estranged husband's parents [37]. Since this case, Menotti-Raymond et al. have published details for a 11 tetranucleotide STR multiplex intended for the genetic individualization of domestic cats [38].

Animal STRs are chosen and analyzed in the same fashion as human STRs. Many polymorphic STR loci have been identified for a wide variety of animals including dogs, cats, pigs, and badgers. Animal STR analysis can also be important for investigations into illegal trade, poaching, or endangered species where species identification is important.

End Notes

[a.] http://www.ncbi.nlm.nih.gov/Genbank/. GenBank is an annotated collection of all publicly available nucleotide sequences.

References

[1] Edwards, A., Civitello, A., Hammond, H.A. & Caskey, C.T. (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats, *American Journal of Human Genetics* **49**, 746–756.

[2] Collins, J.R., Stephens, R.M., Gold, B., Long, B., Dean, M. & Burt, S.K. (2003). An exhaustive DNA micro-satellite map of the human genome using high performance computing, *Genomics* **82**, 10–19.

[3] Hauge, X.Y. & Litt, M. (1993). A study of the origin of 'shadow bands' seen when typing dinucleotide repeat polymorphisms by the PCR, *Human Molecular Genetics* **2**(4), 411–415.

[4] Sullivan, K.M., Mannucci, A., Kimpton, C.P. & Gill, P. (1993). A rapid and quantitative DNA sex test: Fluorescence-based PCR analysis of X-Y homologous gene amelogenin, *BioTechniques* **15**(4), 637–641.

[5] Santos, F.R., Pandya, A. & Tyler-Smith, C. (1998). Reliability of DNA-based sex tests, *Nature Genetics* **18**(2), 103.

[6] von Wurmb-Schwark, N., Bosinski, H. & Ritz-Timme, S. (2007). What do the X and Y chromosomes tell us about sex and gender in forensic case analysis? *Journal of Forensic and Legal Medicine* **14**(1), 27–30.

[7] Gill, P., Brinkmann, B., d'Aloja, E., Andersen, J., Bar, W., Carracedo, A., Dupuy, B., Eriksen, B., Jangblad, M., Johnsson, V., Kloosterman, A.D., Lincoln, P., Morling, N., Rand, S., Sabatier, M., Scheithauer, R., Schneider, P. & Vide, M.C. (1997). Considerations from the European DNA profiling group (EDNAP) concerning STR nomenclature, *Forensic Science International* **87**(3), 185–192.

[8] DNA Commission of the International Society of Forensic Haemogenetics (1994). DNA recommendations – 1994 report concerning further recommendations of the DNA Commission of the ISFH regarding PCR-based polymorphisms in STR (short tandem repeat) systems, *Forensic Science International* **69**(2), 103–104.

[9] Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Rapp, B.A. & Wheeler, D.L. (2002). GenBank, *Nucleic Acids Research* **30**(1), 17–20.

[10] Urquhart, A., Kimpton, C.P., Downes, T.J. & Gill, P. (1994). Variation in short tandem repeat sequences – a survey of twelve microsatellite loci for use as forensic identification markers, *International Journal of Legal Medicine* **107**, 13–20.

[11] Puers, C., Hammond, H.A., Jin, L., Caskey, C.T. & Schumm, J.W. (1993). Identification of repeat sequence heterogeneity at the polymorphic short tandem repeat locus HUMTH01[AATG]n and reassignment of alleles in population analysis by using a locus-specific allelic ladder, *American Journal of Human Genetics* **53**(4), 953–958.

[12] Grubwieser, P., Zimmermann, B., Niederstatter, H., Pavlic, M. & Parson, W. (2006). Validation study and population data of 15 "new" STR loci: a highly discriminating set for paternity and kinship analysis, *International Congress Series* **1288**, 447–449.

[13] Butler, J.M. (2005). Constructing STR multiplex assays, in *From Methods in Molecular Biology Vol 297: Forensic DNA Typing Protocols*, A. Carracedo, ed, Humana Press Inc, Totowa, NJ.

[14] Harbison, S.A., Hamilton, J.F. & Walsh, S.J. (2001). The New Zealand DNA databank: its development and significance as a crime solving tool, *Science and Justice – Journal of the Forensic Science Society* **41**(1), 33–37.

[15] Martin, P.D., Schmitter, H. & Schneider, P.M. (2001). A brief history of the formation of DNA databases in forensic science within Europe, *Forensic Science International* **119**(2), 225–231.

[16] Thibaut, F., Ribeyre, J.M., Dourmap, N., Meloni, R., Laurent, C., Campion, D., Menard, J.F., Dollfus, S., Mallet, J. & Petit, M. (1997). Association of DNA polymorphism in the first intron of the tyrosine hydroxylase gene with disturbances of the catecholaminergic system in schizophrenia, *Schizophrenia Research* **23**(3), 259–264.

[17] Burgert, E., Crocq, M.A., Bausch, E., Macher, J.P. & Morris-Rosendahl, D.J. (1998). No association between the tyrosine hydroxylase microsatellite marker HUMTH01 and schizophrenia or bipolar I disorder, *Psychiatric Genetics* **8**(2), 45–48.

[18] Brinkmann, B., Klintschar, M., Neuhuber, F., Huhne, J. & Rolf, B. (1998). Mutation rate in human microsatellites: influences of the structure and length of the tandem repeat, *American Journal of Human Genetics* **62**, 1408–1415.

[19] Xu, H. & Fu, Y.X. (2004). Estimating effective population size or mutation rate with microsatellites, *Genetics* **166**(1), 555–563.

[20] Buckleton, J., Triggs, C.M. & Walsh, S.J. (2005). *Forensic DNA Evidence Interpretation*, CRC Press, Boco Raton.

[21] Butler, J.M. (2005). *Forensic DNA Typing*, Elsevier, Burlington.

[22] Leibelt, C., Budowle, B., Collins, P., Daoudi, Y., Moretti, T., Nunn, G., Reeder, D. & Roby, R. (2003). Identification of a D8S1179 primer binding site mutation and the validation of a primer designed to recover null alleles, *Forensic Science International* **133**(3), 220–227.

[23] Collins, P.J., Hennessy, L.K., Leibelt, C.S., Roby, R.K., Reeder, D.J. & Foxall, P.A. (2004). Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: the AmpFℓSTR® identifiler® PCR amplification kit, *Journal of Forensic Sciences* **49**(6), 1265–1277.

[24] Budowle, B. & Sprecher, C.J. (2001). Concordance study on population database samples using the PowerPlex™ 16 kit and AmpFℓSTR® Profiler Plus™ kit and AmpFℓSTR® COfiler™ kit, *Journal of Forensic Sciences* **46**(3), 637–641.

[25] Clayton, T.M., Guest, J.L., Urquhart, A.J. & Gill, P.D. (2004). A genetic basis for anomalous band patterns

encountered during DNA STR profiling, *Journal of Forensic Sciences* **49**(6), 1207–1214.
[26] Yang, C.H., Hsieh, L.L., Tsai, C.W., Chiou, F.S., Chou, S.L., Hsu, B.D. & Pai, C.Y. (2003). Evaluation of the DNA stability of forensic markers used in betel-quid chewers' oral swab samples and oral cancerous specimens: implications for forensic application, *Journal of Forensic Sciences* **48**(1), 88–92.
[27] Szibor, R. (2007). X-chromosomal markers: past, present and future, *Forensic Science International: Genetics* **1**, 93–99.
[28] Szibor, R., Krawczak, M., Hering, S., Edelmann, J., Kuhlisch, E. & Krause, D. (2003). Use of X-linked markers for forensic purposes, *International Journal of Legal Medicine* **117**(2), 67–74.
[29] Szibor, R., Hering, S. & Edelmann, J. (2006). A new web site compiling forensic chromosome X research is now online, *International Journal of Legal Medicine* **120**(4), 252–254.
[30] Johns, L.M., Burton, R.E. & Thomson, J.A. (2006). Study to compare three commercial Y-STR testing kits, *International Congress Series* **1288**, 192–194.
[31] Hill, C.R., Kline, M.C., Mulero, J.J., Lagace', R.E., Chang, C.-W., Hennessy, L.K. & Butler, J.M. (2007). Concordance study between the AmpFℓSTR® MiniFiler™ PCR amplification kit and conventional STR typing kits, *Journal of Forensic Sciences* **52**(4), 870–873.
[32] Gill, P., Fereday, L., Morling, N. & Schneider, P.M. (2006). The evolution of DNA databases – recommendations for new European STR loci, *Forensic Science International* **156**(2–3), 242–244.
[33] Gill, P., Fereday, L., Morling, N. & Scheeider, P.M. (2006). Letter to the editor. New multiplexes for Europe – Amendments and clarification of strategic development, *Forensic Science International* **163**, 155–157.
[34] Gilmore, S., Peakall, R. & Robertson, J. (2003). Short tandem repeat (STR) DNA markers are hypervariable and informative in *Cannabis sativa*: implications for forensic investigations, *Forensic Science International* **131**, 65–74.
[35] Craft, K.J., Owens, J.D. & Ashley, M.V. (2007). Application of plant DNA markers in forensic botany: genetic comparison of *Quercus* evidence leaves to crime scene tress using microsatellites, *Forensic Science International* **165**, 64–70.
[36] Padar, Z., Egyed, B., Kontadakis, K., Furedi, S., Woller, J., Zoldag, L. & Fekete, S. (2003). Importance of canine identification in the Hungarian forensic practise, *International Congress Series* **1239**, 897–900.
[37] Menotti-Raymond, M.A., David, V.A. & O'Brien, S.J. (1997). Pet cat hair implicates murder suspect, *Nature* **386**, 774.
[38] Menotti-Raymond, M.A., David, V.A., Wachter, L.L., Butler, J.M. & O'Brien, S.J. (2005). An STR forensic typing system for genetic individualization of domestic cat (*Felis catus*) samples, *Journal of Forensic Sciences* **50**(5), 1061–1070.

Further Reading

Butler, J.M. (2006). Genetics and genomics of core short tandem repeat loci used in human identity testing, *Journal of Forensic Sciences*, **51**(2), 253–265.
Ruitberg, C.M. & Reeder, D.J., Butler, J.M. (2001). STRBase: a short tandem repeat DNA database for the human identity testing community, *Nucleic Acids Research* **29**(1), 320–322.

JO-ANNE BRIGHT

Short Tandem Repeats: Interpretation

STR Profiles

Since around the mid-1990s, short tandem repeat (STR) profiles have been the most widely used form of DNA profile evidence, in both criminal and civil cases. They are used both to help establish the identity of the source(s) of a biological sample, and to evaluate claims of relatedness, for example, of a child with a putative father or a missing person with a parent or sibling. For further details of the biology and technology underling STR profiles, see [1]. Here, we give only brief details essential for appreciation of interpretation issues. Broadly speaking, these issues are similar for STR profiles as for other DNA profiles; for an introduction, (*see* **Evidence Interpretation: a Logical Approach**). There are some aspects that are specific for STR profiles, for example, the impact of population genetics issues, and analyses of paternity and other relationships, depend on details of the STR mutation process. Also stutter artifacts that arise in STR profiling have an important role in the analyses of unbalanced mixture profiles (for details, *see* **Mixture Interpretation: DNA**).

At STR loci, DNA sequence motifs (typically in forensic applications of length four base pairs) are repeated in tandem. The number of repeats tends to vary because of the relatively high mutation rate, around one mutation per STR locus every

500 generations, compared with about one mutation every 50 million generations at a typical genomic site. This variation can be captured cost-effectively by measuring the length of a chromosome fragment that includes the STR locus. Actually, this length is measured indirectly, inferred from the speed of the fragment under an electric field. The lengths of any flanking regions can be subtracted from the inferred length to deduce the number of repeat units. Thus, an individual's genotype at an STR locus can be reported as an unordered pair of integers, giving the number of repeats of the STR motif on each chromosome. For example, the pair (7,9) indicates seven copies of the repeat motif on one chromosome and nine on the other. Partial repeats sometimes occur: an allele designated as 9.3 has nine full repeats and one imperfect repeat having only three base pairs.

Single-Source Identification Using Likelihood Ratios

Suppose that we have a DNA sample from a crime scene, contributed by a single individual C who will be assumed here to be the offender. We also have a matching STR profile from a defendant s. Let E_d denote the two matching STR profiles. All the probabilities reported below are conditional on relevant background information, but this will be suppressed in the notation. The probability that $C = s$, given the evidence, can be written as

$$\mathrm{P}(C = s|E_d) = \frac{1}{1 + \sum_{i \in \mathcal{P}} w_i R_i} \qquad (1)$$

where $\mathcal{P}$ is the population of alternative possible culprits, and R_i denotes the likelihood ratio (LR) comparing the hypothesis $C = s$, with the alternative that C is some other individual, i:

$$R_i = \frac{\mathrm{P}(E_d|C = i)}{\mathrm{P}(E_d|C = s)} \qquad (2)$$

The *other evidence* (or *prior*) ratio w_i is the ratio of the probabilities that $C = i$ and that $C = s$, both evaluated in the light of the evidence other than E_d; that is,

$$w_i = \frac{\mathrm{P}(C = i)}{\mathrm{P}(C = s)} \qquad (3)$$

Most forensic authors define the LR with numerator and denominator interchanged. Our definition simplifies many formulas, and has the advantage that R_i can often be interpreted as a *match probability*. For us, small LRs indicate strong evidence against s.

Equation (1) is not directly useful to a forensic DNA expert in court, because the w_i lie beyond the domain of his/her expertise. Thus the expert is normally limited to advising on values for the R_i for various alternative possible suspects. However, familiarity with equation (1) is crucial for a clear understanding of evidential weight. For example, it is possible to reformulate the hypotheses so that the R_i and w_i change in value yet $\mathrm{P}(C = s|E_d)$ remains unchanged. This arises in the debate about the effect of database searches ([2], Sec 6.1). Thus, the LR is not an absolute measure of weight of evidence, but depends on the formulation of the hypotheses.

In the simplest scenario in which $R_i = r$ and $w_i = 1$ for all $i \in \mathcal{P}$, equation (1) simplifies to

$$\mathrm{P}(C = s|E_d) = \frac{1}{1 + Nr} \qquad (4)$$

where N is the size of $\mathcal{P}$. This unrealistic scenario illustrates the crucial point that a value for R_i does not measure the overall strength of the STR evidence against s. Here, N and r play equally important roles. More generally, the overall strength involves the sum of the $w_i R_i$ over all other individuals i.

The population $\mathcal{P}$ is assumed to include all the possible sources of the crime stain other than s. Note, in particular, that there is no need to invoke any "random man", which concept can cause confusion in the interpretation of DNA profile evidence ([2], Sec 8.2). Although $\mathcal{P}$ should include all realistic alternative suspects, there is some flexibility as to how many extra individuals are included. Often, it might be appropriate to include in $\mathcal{P}$ all men aged, say, between 16 and 65 living within, say, 1 h driving time of the crime scene. However, $\mathcal{P}$ could include everyone on earth except s, if desired. The value of w_i will typically be extremely small for i that reside far from the crime scene, and so it makes very little difference to equation (1) whether they are included in $\mathcal{P}$.

Since $\mathcal{P}$ can be large, it may seem impractical to compute a separate R_i for *every* member of $\mathcal{P}$. However, individuals with the same degree of relatedness with s will have the same value of R_i and hence can be grouped together to simplify equation (1). For example, the population $\mathcal{P}$ of alternative culprits may be partitioned into various categories

of direct relatives, such as siblings and cousins, while individuals apparently unrelated to s might be classified into three groups:

- same population, same subpopulation
- same population, different subpopulation
- different population.

To avoid overstating the evidence against s, if the value of R_i varies within a group then the largest value should be applied for all members of the group. Because of migrations and intermarriages, the population and subpopulation are difficult to define precisely. Nevertheless, such concepts often correspond to natural groupings, and they are useful in formulating population genetics models that can allow for the principle effects of population genetic structure on DNA profile match probabilities.

The Population Genetics Parameter θ, or F_{ST}

Suppose that the population is "UK Caucasians", and the subpopulations relevant to a particular crime might include regional subpopulations, or migrant or religious groups such as people of Polish or Jewish ancestry. STR allele frequency estimates are available from a database of individuals classified as UK Caucasians. The accuracy of the estimates will be affected both by the size of the database and the fact that it is not scientific random sample but a "convenience sample" of available individuals. Nevertheless, we can be reassured by the observation that different databases of Caucasian individuals give similar allele proportions at the STR loci in widespread forensic use. For subpopulations, however, direct frequency estimates are often not available: in practice this is always the case because the relevant subpopulation is not strictly defined. The similarity of allele frequencies in broadly defined Caucasian populations does not provide reassurance that narrowly defined subpopulations will also have similar allele proportions: their smaller sizes and distinct histories can make them more variable than large, well-mixed populations.

To deal with the problem of unknown subpopulation allele proportions, we make use of the coefficient F_{ST} [3], also called θ, which can be interpreted in terms of the variation of subpopulation allele proportions about a given reference value, which here would be taken to be the allele proportion, say p, in the UK Caucasian population. Although we cannot estimate the STR allele proportion directly relevant to the case, using θ we can, nevertheless, still calculate the required match probabilities in terms of the degree of variation of subpopulation allele proportions about p.

In some simple models, θ can also be interpreted as the probability that two alleles are descended identically from a common ancestor. Hence θ is also called a *coancestry coefficient* or *kinship coefficient*. Then θ can be thought of informally as the amount of shared ancestry within a population, and so is large for small, isolated populations with little immigration.

The Sampling Formula

Suppose that n alleles have been sampled in the subpopulation, of which m are green. Then the probability that the next allele sampled is also green can be expressed as

$$\frac{m\theta + (1-\theta)p_G}{1 + (n-1)\theta} \tag{5}$$

(see [2], Chap 5 for justification.) When $m = n = 0$, we obtain probability p_G that the first allele drawn is green. The probability that the first two alleles drawn are both green is

$$p_G(\theta + (1-\theta)p_G) = p_G^2 + \theta p_G(1-p_G) \tag{6}$$

Increasing θ raises the probability of observing two green alleles, because the second allele may be identical with the first through descent from a recent common ancestor. Conversely, increasing θ decreases the probability of a green allele followed by a blue, which is

$$(1-\theta)p_G p_B \tag{7}$$

Successive use of expression (5) can be used to build up joint probabilities for any ordered sequence of alleles.

Likelihood Ratios and Match Probabilities

Introducing the notation $x \equiv D$ to denote "x has profile D", E_d can be written succinctly as $C \equiv D$, $s \equiv D$, and the LR, equation (2), becomes

$$R_i = \frac{P(C \equiv D, s \equiv D \mid C = i)}{P(C \equiv D, s \equiv D \mid C = s)} \tag{8}$$

Here, we initially ignore the possibility of error, so that if $C = s$ then crime scene and defendant profiles

are certain to match. Assuming also that the fact that an individual committed the crime does not of itself alter the probability that they have a particular profile, equation (8) can be simplified further to

$$R_i = \frac{\mathrm{P}(i \equiv D, s \equiv D)}{\mathrm{P}(s \equiv D)} = \mathrm{P}(i \equiv D \mid s \equiv D) \quad (9)$$

Thus, under these conditions, R_i reduces to the conditional probability, called the *match probability*, that i has the profile *given* that s has it: population genetic effects arise, and can be dealt with, *via* this conditioning. Other evidence such as eyewitness reports and alibis are typically irrelevant to DNA profile match probabilities, but the background information to a case can include

- information about the relatedness of s with some other individuals;
- allele proportion information from population databases of DNA profiles;
- other relevant population genetics data and theory.

and these can be important for match probabilities.

An essential feature of the match probability is that it takes account of *both* the observed profiles that form the match. Some authors, for example [4], misleadingly refer to the population proportion of the profile as a *match probability* which is inappropriate since the concept of "match" involves two profiles, not one. Equation (9) indicates that the question relevant to forensic identification is not

> "what is the probability of observing a particular profile?"

but

> "given that I have observed an individual with this profile, what is the probability that another (unprofiled) individual will also have it?".

The parameter θ appears in our answer to this question, to take into account possible shared ancestry between the two individuals. Ignoring a defendant's coancestry with other possible sources of the crime stain is systematically unfair to him. Human population genetics is complicated, and inevitably, θ is an imperfect measure, but by choosing a suitable value defendants will not be systematically disfavored, while match probabilities remain small enough to form the basis of satisfactory prosecutions in most cases.

Single-Locus Match Probabilities

Consider the case that both i and s are homozygous for allele A. If we assume that they are unrelated, both come from the same subpopulation, and neither is inbred, then equation (9) corresponds to the conditional probability that two further alleles are both A, given a sample of two alleles that are both A. This is obtained as the product of two instances of expression (5), with $m = n = 2$ and $m = n = 3$:

$$R_i = \frac{(2\theta + (1-\theta)p_\mathrm{A})(3\theta + (1-\theta)p_\mathrm{A})}{(1+\theta)(1+2\theta)} \quad (10)$$

In the heterozygous case, under the same assumptions, we need the probability that two further alleles are A and B, given that two observed alleles are A and B. Taking two times expression (5) with $m = 1$, $n = 2$ times expression (5) with $m = 1$, $n = 3$, gives

$$R_i = 2\frac{(\theta + (1-\theta)p_\mathrm{A})(\theta + (1-\theta)p_\mathrm{B})}{(1+\theta)(1+2\theta)} \quad (11)$$

See Figure 1 for a graph showing how equations (10) and (11) vary with θ for some particular values of the p_j for $j \in \{\mathrm{A,B}\}$. Note that increasing θ does not always increase R_i. Table 1 gives numerical values of single-locus R_i for hypothetical genotypes and p_j values at four STR loci, for four values of θ that span the range used in practice.

These match probability formulas are conditional on the values of θ and the p_j. Thus, strictly, they only apply if these parameters are known exactly, which is never the case in practice. In a fully Bayesian approach, equations (10) and (11) should be integrated out with respect to probability distributions for the unknown values of θ and the p_j, based on the available background information, for example from forensic DNA profile databases and population genetics theory and data. A simpler and more interpretable approach is preferred, in which this background information is used to obtain estimates for θ and the p_j, which are then "plugged in" to the conditional formulas. Care is required to choose the most appropriate "plug-in" values for the parameters, as these need not satisfy the usual criteria for statistical estimators (see below).

Use of equations (10) and (11) implies an assumption of Hardy–Weinberg Equilibrium (HWE) within subpopulations, but not in the broader population from which the forensic database is drawn. Thus tests of deviations from HWE applied to

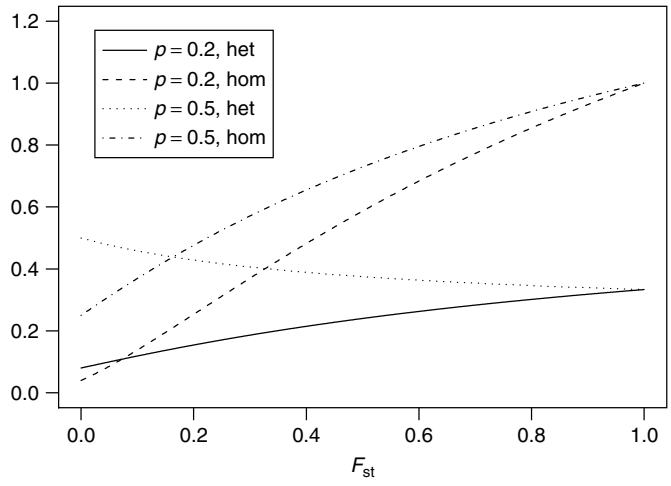


Figure 1 Single-locus match probabilities calculated using equations (10) and (11) for $\theta(= F_{ST})$ ranging from 0 to 1. In the heterozygote case, the proportions of the two alleles are both equal to p

Table 1 Single-locus match probabilities assuming i unrelated to s, for four STR loci and various values of θ, for the alleles specified in column 2 and assuming the population proportions given in column 3

STR locus	Geno-type	Population proportions	Match probability ($\times 10^3$)			
			$\theta = 0$	$\theta = 1\%$	$\theta = 2\%$	$\theta = 5\%$
D18	14, 16	0·16, 0·14	45	49	52	64
D21	28, 31	0·23, 0·07	32	37	41	54
THO1	9·3, 9·3	0·30, 0·30	90	101	112	145
D8	10, 13	0·09, 0·33	59	65	70	85

forensic databases are not directly relevant to match probability calculations. Deviations from HWE do directly affect the probability of observing a particular genotype, but not the conditional match probability given by expression (9).

Match probability formulas that take into account inbreeding within subpopulations, are more complicated than equations (10) and (11), but, nevertheless, relatively easy to apply [5]. The match probabilities are slightly increased for homozygotes, and decreased for heterozygotes; the overall effect on profile match probabilities is usually very small, but may be worth taking into account when highly inbred populations are relevant to a case.

Multiple Loci: The "Product Rule"

There has been much debate in the forensic science literature about the validity of the "product rule" for combining DNA match probabilities across loci. Combining probabilities *via* multiplication implies an assumption of statistical independence, and so the debate is equivalent to a debate about the *independence* of STR genotypes at different loci. This question can be rephrased:

> For two distinct individuals i and s, does the event that their genotypes match at one or more STR loci affect the probability that they will match at the next locus?

If i and s are directly related through one or more known common ancestors (e.g., grandparents), then matches at distinct loci are not independent, but an appropriate adjustment for the relationship can restore approximate independence. All humans are related at some level, and if i and s are apparently unrelated, this only means that the relationship between them

Table 2 Match probabilities for the four-locus STR profile of Table 1 under some possible relationships of s and i, and for various values of θ

	Match probability			
Relationship	$\theta = 0$	$\theta = 1\%$	$\theta = 2\%$	$\theta = 5\%$
Identical twin	1	1	1	1
Sibling ($\times 10^{-3}$)	17	19	20	23
Parent/child ($\times 10^{-4}$)	14	17	19	29
Half-sib ($\times 10^{-5}$)	23	29	35	61
Cousin ($\times 10^{-5}$)	6	8	10	20
Unrelated ($\times 10^{-6}$)	8	12	17	43

is unknown and presumed to be distant. However, it is not necessarily distant enough to be negligible for the purposes of calculating match probabilities.

Thus, use of the product rule based on an assumption of no relatedness at any level is unfair to defendants. However, the match probabilities given by equations (10) and (11) are conditional on a level of coancestry between apparently unrelated individuals measured by θ. Although match probabilities at many loci cannot readily be checked, the available population genetics theory and data support the view that, given a suitable value of θ, match probabilities obtained as products of the probabilities in equations (10) and (11) will not be systematically unfavorable to defendants.

Whole-profile match probabilities using the θ-adjusted product rule applied to the four STR loci of Table 1, and using various values of θ, are given in the final row of Table 2. Assuming $\theta = 5\%$ increases the four-locus match probability more than fivefold relative to the $\theta = 0$ case, and this extrapolates to more than a 50-fold increase for 10 loci, and about 200-fold for a 13-locus profile match.

Relatives of s

So far we have been considering alternative possible culprits i that are unrelated to s. Here we consider the possibility that i and s are directly related through one or more specified common ancestors. We continue to assume that the DNA profile of i is not available to the court: it would in principle be desirable to exclude close relatives of s from suspicion by examining their DNA profiles, but this is rarely possible in practice. Considering first a single locus, let Z denote the number of alleles at the locus that i and s share identical by descent (ibd) from a known, recent, common ancestor (e.g., parent or grandparent), and let

$$\kappa_j = \mathrm{P}(Z = j) \tag{12}$$

for $j = 0, 1, 2$. The values of κ_j under some regular (i.e., no inbreeding) relationships are shown in the first three columns of Table 3.

Table 3 Distribution of ibd status and coefficient of relatedness for some possible relationships of s and i. The value of κ_j is the probability that i and s share j alleles at a locus identical by descent (ibd) from a recent common ancestor, and $\overline{\kappa}$ is half the expected number of alleles shared ibd. The values for aunt, uncle, niece, nephew, grandparent, and grandchild are the same as for half-sib

Relationship	κ_0	κ_1	κ_2	$\overline{\kappa}$
Identical twin	0	0	1	1
Sibling	1/4	1/2	1/4	1/2
Parent/child	0	1	0	1/2
Half-sib	1/2	1/2	0	1/4
First cousin	3/4	1/4	0	1/8
Double first cousin	9/16	6/16	1/16	1/4
Unrelated	1	0	0	0

If $Z = 0$, then it is the same as if s and i were unrelated and we write M_2 for the appropriate match probability, either from equation (10) or (11). If $Z = 2$, a match is certain. For $Z = 1$, consider first the case that $s \equiv \mathrm{AA}$. Since one allele of i is ibd with an observed allele of s, we require the probability of observing a further A allele, given that two alleles have been observed to be both A. Using expression (5) with $m = n = 2$ we obtain

$$M_1 = \mathrm{P}(\mathrm{A} \mid \mathrm{AA}) = \frac{2\theta + (1-\theta)p_{\mathrm{A}}}{1+\theta} \tag{13}$$

If $s \equiv \mathrm{AB}$, the allele shared ibd by i and s is equally likely to be A or B. The match probability is then equivalent to the probability of observing the non-ibd allele, given that A and B have been observed:

$$\begin{aligned} M_1 &= \frac{\mathrm{P}(\mathrm{A} \mid \mathrm{AB}) + \mathrm{P}(\mathrm{B} \mid \mathrm{AB})}{2} \\ &= \frac{\theta + (1-\theta)(p_{\mathrm{A}} + p_{\mathrm{B}})/2}{1+\theta} \end{aligned} \tag{14}$$

The overall single-locus match probability for relatives is then

$$\kappa_2 + \kappa_1 M_1 + \kappa_0 M_2 \tag{15}$$

Four-locus match probabilities based on (14) and the product rule are shown in Table 2 for the four STR loci of Table 1. Notice that the value of θ still affects the match probability even when i and s are assumed to be closely related. This is because, as well as matches arising via alleles shared ibd from the recent known ancestor(s) of i and s, alleles can also be shared ibd from more distant common ancestors. However, the relative importance of θ declines as the known relationship between i and s becomes closer [6].

Values for the p_j

The population database to be used to estimate the p_j should be that most appropriate for i, the alternative possible culprit under consideration. The error arising from the database population not being exactly appropriate can be addressed using θ. Here we consider only the issue of accounting for the effects of sampling uncertainty in observed allele proportions. One way to address this issue is to estimate p_j at a heterozygous locus by

$$\widehat{p_j} = \frac{x_j + 2}{n + 4} \tag{16}$$

where x_j is the frequency of allele j in the database, and n is the total number of alleles [6, 7]. In the homozygous case, an analogous estimate is

$$\widehat{p_j} = \frac{x_j + 4}{n + 4} \tag{17}$$

These estimates can be thought of as adding both the crime-scene and defendant profiles to the database; such estimation methods are sometimes called *pseudocount* methods. They can be justified as approximations to the posterior mean given the sample allele counts and a uniform prior distribution for the allele proportions.

The value of θ

Published estimates of θ at STR loci, for subpopulations within the major human "racial" groups are often small, and typically less than 1%. See, for example, [8, 9] for estimates in some European Caucasian populations. Many other authors have estimated θ at forensic STR loci in various populations and using different methodologies; see [10] for a brief review. Typically, they report very small values.

There are several arguments for using larger values in forensic practice than suggested by these estimates. Recall that our goal is not to estimate a "best" value for θ, but to choose "plug-in" values that give a match probability similar to that which would be obtained from using the full Bayesian approach. Because of the skewness of appropriate distributions for θ, the "plug-in" value that mimics the effect of integration over θ may be much larger than, say, a maximum-likelihood estimate of θ. Moreover, published estimates of θ usually relate to the variation of allele proportions around the observed mean of the subpopulations studied, whereas in forensic applications, the reference value is the forensic database value. The θ estimate can vary substantially according to the choice of reference value, and is often much larger than when the reference value is estimated from the data. Broadly speaking, the less appropriate is the database for the genetic background of a possible culprit, the greater is the appropriate value of θ.

For these reasons (see [2], Sec 6.3, for a fuller discussion), it is suggested that a relatively large value, such as 2%, be used when both defendant and alternative possible culprit are drawn from a relatively well-mixed, large population, and perhaps 3% could be used if both are drawn from one of the large minority groups. In some small minority groups, $\theta = 5\%$ may be appropriate. These suggested values are based on informal judgements, and should not be regarded as prescriptive.

If i is not from the same racial group as s, then they have little coancestry and so a small value of θ can be justified. However, since the problem of representativeness of any database remains, as well as possibilities for some coancestry even across apparent racial groups, the use of a nonzero value of θ in every case, perhaps setting 1% as the minimum, is advocated.

Laboratory, Handling errors, and Evidence Tampering

There are at least two ways in which an observed STR profile match could have arisen even though $C \neq s$:

1. defendant and culprit happen to have matching DNA profiles and no typing error occurred ("chance match");

2. defendant and culprit have distinct DNA profiles, and the observation of matching profiles is due to an error in one or both recorded profiles ("false match").

Both (1) and (2) are typically unlikely, but (2) may not be negligible in comparison to (1), yet (2) is typically not explicitly considered in conjunction with (1). An argument that is often advanced for this is that error probabilities are difficult to assess. Even if error rates from external, blind trials are available, there will usually be specific details of the case at hand that differ from the circumstances under which the trials were conducted, and which make it more or less likely that an error has occurred.

Some critics of DNA profile evidence have argued that the match probabilities reported in court are irrelevant and potentially misleading because they relate to (1) only, whereas (2) is likely to be more important in practice. This argument has some force, and proponents of DNA profiling have been prone to mistakes and exaggeration in their attempts to discount it [11]. The chance of a false inclusion error due to genotyping anomalies is so remote as to be negligible, even relative to the match probability, because it is not the probability of any error that is relevant, but an error that causes a false match. Contamination is a distinct possibility in some settings, but because evidence and crime samples are routinely typed by different staff in different laboratories, sometimes with a substantial time gap, in many cases this possibility can also reasonably be ruled out.

This leaves the possibility of false inclusion due to handling or labeling error, or evidence tampering. A conspiracy theory, such as that police and/or judicial authorities colluded to manufacture evidence falsely linking the defendant with the crime scene, may be regarded by a reasonable juror as substantially more plausible than a 1 in a billion chance match, even when there are no particular reasons to suggest a conspiracy. For this juror, the probability of such a conspiracy based on all the other evidence, is more important than the chance match probability.

There seems no role for a forensic scientist to predict whether a juror might pursue such a line of reasoning, and hence the only reasonable option is to supply the juror with a match probability based on (1), but also to try convey an understanding of the possibilities for a false match and how these affect evidential weight.

Partial Profiles

So far we have assumed good quality crime-scene DNA samples and no profiling errors, so that a perfect match of crime-scene and defendant profiles is expected if s is the source of the crime-scene DNA. However, if the crime-scene sample is very small and/or degraded, the crime-scene profile may be subject to stochastic phenomena such as drop-out, drop-in, imbalanced peak heights and exaggerated stutter. Under these circumstances, a prosecution may proceed against s even though his profile does not exactly match the crime-scene profile. When all the crime-scene profile alleles are evident in the profile of s, but the converse does not hold, the crime-scene profile is sometimes referred to as *partial* and the prosecution case requires that drop-out has occurred. Strictly, we can only be sure that alleles have dropped out if there is a locus with no observed alleles: no conclusion about drop-out should refer to the profile of s.

The analysis of profiles in the presence of drop-out and related phenomena is a major current challenge for the forensic DNA community, and completely satisfactory solutions are not yet available. One unsatisfactory approach [12] that is often used in practice involves using the standard LR when two alleles are observed in the crime-scene profile, $2p$ when one allele is observed, and 1 when no alleles are observed. For a more advanced attempt at formulating appropriate LRs, see [13], but the software described there is not widely available.

Note that a "partial" profile that arises because some loci are not profiled does not raise any problem of interpretation: the fact that additional loci might have been profiled but were not, is irrelevant to the assessment of the loci that were profiled.

Database Searches

Many countries maintain national databases of the STR profiles of named individuals for criminal intelligence purposes. The question thus arises as to the appropriate method for assessing the DNA profile evidence when the defendant was identified following a search through a database. The number of individuals involved in such a search, and even the fact that there was a search, may not be reported to the court (*see also* **DNA Databases and Evidentiary Issues**).

Many commentators, including [4], have wrongly claimed that the fact that a DNA profile match is

more likely when it results from a search means that the evidence is weakened by the search. An analysis that focuses on the relevant question–is s the source of the crime-scene DNA profile?–shows that DNA evidence is usually slightly stronger in the database search setting than when no search has occurred. To see this, imagine an enormous database that records the STR profiles of everyone on earth. If the defendant's profile were the only one in this database to match the crime-scene profile, then the evidence against him would be overwhelming.

Although the DNA evidence may be (slightly) stronger in the context of a database search, the overall case against the defendant may be weaker because there is little or no non-DNA evidence against the defendant, see [14, 15].

Match Probability for Y Chromosome STR Profiles

Because the Y chromosome is paternally inherited, and for the most part does not undergo recombination, a Y-STR profile can be treated as a single allele, or "haplotype", characterized by the repeat counts at STR sites along the chromosome. The match probability involves the probability that a particular man has a certain haplotype given that another man has it. This is conceptually simpler than for autosomal loci because there is no need for multiplication of frequencies either within or across loci.

Using an approach analogous to that used to derive (10) and (11), the match probability for a Y haplotype with population frequency p is

$$R_i = \theta + (1 - \theta)p \tag{18}$$

Since p is typically unknown, a "pseudocount" estimator, analogous to equations (15) and (16) for autosomal loci, can be used:

$$\widehat{p} = \frac{x + 2}{N + 2} \tag{19}$$

where x denotes the database count of the observed Y haplotype.

Population structure is particularly important for Y haplotypes: values of θ are typically high, reflecting geographical clustering of males with common paternal ancestry [16]. There are only a few studies on the appropriate fine geographical scale. A specific Y haplotype that is frequent in many parts of Asia, and largely unobserved elsewhere, has been attributed to the consequences of historic Mongol conquests, possibly even to Genghis Khan himself [17]. Similar historical events on a smaller scale may have led to an unrecognized, local elevation of the concentration of a specific Y haplotype that is otherwise rare. For example, substantial microgeographical variation in forensic Y-chromosome haplotypes has been reported in the Cantabria region of Spain [18].

Although selection is thought likely to influence the distribution of Y haplotypes, because no assumption of either HWE or linkage equilibrium is made, it seems reasonable to assume that selection will not adversely affect the validity of equation (17).

Other Approaches to Assessing STR Evidence

Uniqueness

The current generation of STR profiling technology uses 10 or more loci, so that match probabilities for individuals unrelated to the defendant are typically extremely small, often less than 1 in a billion even with a generous allowance for θ. It thus seems reasonable to consider the possibility that the STR profile is unique, and if this could be established, then the need for calculating and reporting LRs could be avoided.

Although this idea is attractive, it encounters a number of difficulties, such as the choice of threshold for declaring a profile to be unique. Perhaps more importantly, the non-DNA evidence in a case also has a bearing on the question of uniqueness, but lies outside the domain of a DNA expert.

Under some simplifying assumptions on the numbers of individuals with different degrees of relatedness to the defendant, and assuming no evidence other than the STR profiles, 10 STR loci usually suffice to establish a probability greater than 99.9% that the STR of s is not shared by any individual in $\mathcal{P}$ [19].

Random Man Not Excluded (RMNE)

Another approach to the evaluation of STR profiles that avoids computing LRs is to report the inclusion probability of a crime-scene profile, or the probability that a "random man" would not be excluded as the contributor. In the case of a single-source crime-scene profile, the random man not excluded (RMNE) probability is the probability that a randomly chosen alternative possible culprit would match the crime scene

DNA profile. In this setting, the exclusion probability is numerically equivalent to the LR for i unrelated to s (but typically assumes $\theta = 0$). However, the two approaches are conceptually very different.

The rationale underlying the RMNE probability cannot adequately cope with relatives of the defendant among the alternative culprits, and can be highly misleading in the presence of drop-out (partial profiles). The idea of a random alternative suspect can lead jurors to ignore the role of the *number* of possible culprits in evidential assessments, and clear thinking about laboratory and other errors, and the effect of searches, can also be undermined. All of these aspects are readily dealt with in the LR framework.

One specific difficulty that "random man" can cause in this setting concerns the argument over which population the man is supposed to have been randomly drawn from [20]. Since "random man" is a fictional character, these arguments can never be resolved. The issue is important, since too broad a definition of the population leads to overstatement of the evidence, because a large population must contain many people sharing little ancestry with s. If we try to avoid this overstatement by specifying the narrowest possible population, we are led to the population consisting of s only, in which the match probability is one.

Numerical differences arise between RMNE and LR values for paternity, since the former makes no distinction between a homozygous and a heterozygous nonexcluded alleged father, whereas the LR correctly recognizes that the DNA evidence at this locus is about twice as strong against the homozygote. Even more important differences arise in the setting of crime-scene DNA profiles with two or more unknown contributors (*see* **Mixture Interpretation: DNA**). Here, the RMNE probability is usually taken to be the probability that a random unknown individual would have both alleles at each locus included among the alleles of the mixed crime-scene profile. This probability does not take account of the profile of the defendant, other than noting that it falls within a (usually large) class of profiles. In the case of two codefendants alleged to be contributors to the mixed crime-scene profile, the evidence can weigh more heavily against one defendant than the other, whereas the RMNE probability will be the same for both defendants.

The advantages of the RMNE probability include its being relatively easy to calculate and explain, and that it does not require any assumption about the number of contributors to a mixture. Because it ignores relevant information, the RMNE probability is usually larger than the LR, often considerably so. This statistical inefficiency is sometimes seen as a virtue, in that use of the RMNE probability is regarded as "conservative". However, conservativeness is not guaranteed, and by using an appropriate θ-value we can make the LR conservative while still relatively efficient.

Paternity

The principles of combining LRs to evaluate a probability of paternity are similar to those underpinning equation (1). However, in practice DNA evidence is often assessed differently when paternity rather than identification is at issue. Many forensic scientists, lawyers, and academic commentators seem reluctant to consider in the paternity setting the very low prior probabilities that are now often accepted for identification. Indeed, there is a shamefully high prevalence of an unjustified assumption that both s, and an unrelated "random man" i, have a prior probability 1/2 of being the father [21]. This "even prior" assumption is convenient, since in the absence of relatives of s it implies that the likelihood ratio R_i is also the posterior probability that s is not the father. It may often be based on the assumption that equal prior weight should be assigned to the, typically conflicting, claims of s and the mother m. However this ignores, for example, additional background evidence, and multiple alternative fathers.

Likelihood Ratios for Paternity

Typically, the STR evidence consists of the profiles of the mother m, alleged father s, and child c, in which case

$$R_i = \frac{\text{P(profiles of } c, s, \text{ and } m \mid \text{father is } i)}{\text{P(profiles of } c, s, \text{ and } m \mid \text{father is } s)} \quad (20)$$

Here, the fact that m is the mother of c is regarded as background information in both probabilities, but is not explicit in the notation. It is usually reasonable to assume that the probabilities that s and m have particular profiles is unaffected by whether s is the father of c, and it follows that we can rewrite equation

(20) as a ratio of conditional probabilities for the child's profile:

$$R_i = \frac{\text{P(profile of } c \mid \text{profiles of } s \text{ and } m, \text{ father is } i)}{\text{P(profile of } c \mid \text{profiles of } s \text{ and } m, \text{ father is } s)} \quad (21)$$

in which the profile probabilities for s and m have cancelled.

The denominator is easily evaluated using Mendelian transmission probabilities, and is, for example, equal to one if $m \equiv$ AB, $s \equiv$ CC, and $c \equiv$ AC.

For the numerator, we require the conditional probability of c's paternal allele, given the profiles of s and m and the hypothesis that i is the father. First, suppose that the coancestry of i and s is specified by θ, but m has no coancestry with either (for example, she is from a different racial group). Then the profile of m is uninformative about the profile of i, and the paternal allele of c may be regarded as a third allele drawn from the subpopulation of i and s, given that two observed alleles (from s) are both C. Invoking the sampling formula (5) with $m = n = 2$ we obtain

$$R_i = \text{P(C} \mid \text{CC)} = \frac{2\theta + (1-\theta)p_C}{1+\theta} \quad (22)$$

Reasoning similarly in the heterozygous case $s \equiv$ CX, the denominator is 1/2 and

$$R_i = 2\text{P(C} \mid \text{CX)} = 2\frac{\theta + (1-\theta)p_C}{1+\theta} \quad (23)$$

If i, s, and m are all assumed to have the same level of coancestry, then the sampling formula (5) can be used again but now with $n = 4$ because of the four alleles observed in s and m (*see* **Missing Persons and Paternity: DNA** and [2], Chap 7). for examples and further details.

Mutation

If an STR profile consists of around ten loci, there is very roughly a 2% probability that there will be a mutation in transmission from father to child, in which case ignoring the possibility of mutations could lead to a false exclusion. If the profiles are consistent with s being the father of c at many loci, but there is an apparent exclusion at just one or perhaps two loci (Table 4), it may still be the case that the STR evidence overall supports the claim that s is the father. A systematic treatment of this question when the mother's profile is available, but ignoring coancestry is given by Dawid *et al.* [22]. The latter omission is rectified by Ayres [23], who treats the case that any two of m, s, and i have the same level of coancestry, θ, correcting formulas given in [24] that neglected the possibility of maternal transmission.

Table 4 Three examples of scenarios in which a mutation is required to sustain the the hypothesis that m and s are the parents of c

Scenario		(i)	(ii)	(iii)
Child c	$\equiv$	AB	AB	AB
Mother m	$\equiv$	AC	AC	DD
Alleged father s	$\equiv$	CD	AD	AC

Perhaps the most common theoretical model for STR mutation is the stepwise mutation model (SMM) in which a mutant allele has either $k-1$ or $k+1$ repeat units, each with probability 1/2, where k is the current repeat number. The SMM has no stationary distribution, so that two similar populations isolated from each other do not converge under the SMM to the same allele frequency distribution. Thus, substantial between-population diversity is expected under the SMM for populations that exchange few migrants. By contrast, there is typically little between-population variation at human STR loci, suggesting high migration rates and/or the invalidity of the SMM.

In fact, the strict SMM is known to be false, for example, because the mutation rate increases with allele length and occasional two-step mutations occur, although it may provide an adequate approximation for some purposes. The SMM can easily be modified, for example, by hypothesizing a bias toward contraction mutations in long alleles, to obtain STR mutation models that do have a stationary distribution [25]. Such models can provide a better fit to observed data than the SMM, and are consistent with the observed between-population homogeneity of allele proportions at many STR loci.

Conclusion

The theory based on likelihood ratios outlined above does not provide a formulaic answer to the problem

of how to convey complex STR profile evidence to a nonscientific court. It does, however, provide a solid theoretical framework for the forensic scientist to think clearly about weight-of-evidence issues, and hence to draw well-informed conclusions about what a rational judge or juror needs to be informed of in order for them to carry out their tasks of evaluating the evidence. For further details of STR interpretation beyond those given here, *see and also* **Statistical Evidence in Court** [2, 10, 26, 27, 28]

References

[1] Butler, J.M. (2005). *Forensic DNA Typing: Biology, Technology and Genetics of DNA Markers, 2nd Ed*, Elsevier.

[2] Balding, D.J. (2005). *Weight of Evidence for Forensic DNA Profiles*, John Wiley & Sons.

[3] Wright, S. (1951). The genetical structure of populations, *Annals of Eugenics* **15**, 323–354.

[4] National Research Council (1996). *The Evaluation of Forensic DNA Evidence*, National Academy Press.

[5] Ayres, K.L. & Overall, A.D.J. (1999). Allowing for within-subpopulation inbreeding in forensic match probabilities, *Forensic Science International* **103**, 207–216.

[6] Balding, D.J. & Nichols, R.A. (1994). DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection, and single bands, *Forensic Science International* **64**, 125–140.

[7] Balding, D.J. (1995). Estimating products in forensic identification using DNA profiles, *Journal of American Statistical Association* **90**, 839–844.

[8] Balding, D.J., Greenhalgh, M. & Nichols, R.A. (1996). Population genetics of STR loci in Caucasians, *International Journal of Legal Medicine* **108**, 300–305.

[9] Balding, D.J. & Nichols, R.A. (1997). Significant genetic correlations among Caucasians at forensic DNA loci, *Heredity* **78**, 583–589.

[10] Buckleton J.S., Triggs C.M. & Walsh, S.J. (eds) (2005). *Forensic DNA Evidence Interpretation*, CRC Press.

[11] Koehler, J.J. (1996). On conveying the probative value of DNA evidence: frequencies, LRs, and error rates, *University of Colorado Law Review* **67**(4), 859–886.

[12] Buckleton, J. & Triggs, C.M. (2006). Is the 2p rule always conservative? *Forensic Science* **159**, 206–209.

[13] Gill, P., Kirkham, A. & Curran, J. (2007). LoComatioN: a software tool for the analysis of low copy number DNA profiles, *Forensic Science International* **166**, 128–138.

[14] Balding, D.J. & Donnelly, P. (1996). Evaluating DNA profile evidence when the suspect is identified through a database search, *Journal of Forensic Science* **41**, 603–607.

[15] Balding, D.J. (2002). The DNA database controversy, *Biometrics* **58**, 241–244.

[16] Jobling, M.A., Pandya, A. & Tyler-Smith, C. (1996). The Y chromosome in forensic analysis and paternity testing, *International Journal of Legal Medicine* **110**(3), 118–124.

[17] Zerjal, T., Xue, Y.L., Bertorelle, G., Wells, R.S., Bao, W.D., Zhu, S.L., Qamar, R., Ayub, Q., Mohyuddin, A., Fu, S., Li, P., Yuldasheva, N., Ruzibakiev, R., Xu, J., Shu, Q., Du, R., Yang, H., Hurles, M.E., Robinson, E., Gerelsaikhan, T., Dashnyam, B., Mehdi, S.Q. & Tyler-Smith, C. (2003). The genetic legacy of the Mongols, *American Journal of Human Genetics* **72**(3), 717–721.

[18] Zarrabeitia, M.T., Riancho, J.A., Lareu, M.V., Leyva-Cobiaan, F. & Carracedo, A. (2003). Significance of micro-geographical population structure in forensic cases: a Bayesian exploration, *International Journal of Legal Medicine* **117**(5), 302–305.

[19] Balding, D.J. (1999). When can a DNA profile be regarded as unique? *Science and Justice* **39**, 257–260.

[20] Roeder, K. (1994). DNA fingerprinting: a review of the controversy, *Statistical Science* **9**, 222–278.

[21] Koehler, J.J. (1993). DNA matches and statistics: important questions, surprising answers, *Judicature* **76**(5), 222–229.

[22] Dawid, A.P., Mortera, J. & Pascali, V.L. (1996). Non-fatherhood or mutation? A probabilistic approach to parental exclusion in paternity testing, *Forensic Science International* **124**(1), 55–61.

[23] Ayres, K.L. (2002). Paternal exclusion in the presence of substructure, *Forensic Science International* **129**, 142–144.

[24] Ayres, K.L. (2000). Relatedness testing in subdivided populations, *Forensic Science International* **114**, 107–115.

[25] Whittaker, J.C., Harbord, R.M., Boxall, N., Mackay, I., Dawson, G. & Sibly, R.M. (2003). Likelihood-based estimation of microsatellite mutation rates, *Genetics* **164**(2), 781–787.

[26] Robertson, B. & Vignaux, G.A. (1995). *Interpreting Evidence – Evaluating Forensic Science in the Courtroom*, John Wiley & Sons.

[27] Evett, I.W. & Weir, B.S. (1998). *Interpreting DNA evidence*, Sinauer Associates.

[28] Rudin, N. & Inman, K. (2002). *An Introduction to Forensic DNA Analysis*, 2nd Edition, CRC Press.

DAVID BALDING

Signature Comparison *see* Handwriting and Signatures, Comparison of

Signatures ***see*** Handwriting and Signatures, Interpretation of Comparison Results

Simulations: Computer, as Evidence ***see*** Computer Animation and Simulation Evidence

Skeletal Remains: Identification of ***see*** Human Remains and Identity; Species Determination of Osseous Remains

Skeletal remains: Identity ***see*** Trauma Analysis of Skeletal Remains

Skeleton: Trauma ***see*** Trauma Analysis of Skeletal Remains

Sleepwalking ***see*** Automatism as a Defense to Crime

Social Influence: Affect and Decision Making ***see*** Crime Victims' Decision to Report Crime

Social Isolation as Risk Factor in Elder Abuse ***see*** Elder Abuse: Risk

Sociopathic Personality Disorder ***see*** Psychopathy

Sociopathy ***see*** Psychopathy

Soil: Forensic Analysis

Introduction

Forensic soil science is the science or study of soil that involves the application of soil science, especially studies that involve soil morphology, soil mapping (assisted by existing soil maps and spatially held soil data), mineralogy, chemistry, geophysics, biology, and molecular biology to answer legal questions, problems, or hypotheses. *Soil science* is the term commonly used to study soil as a natural body in the landscape and as a resource to be managed for agricultural production, environmental waste disposal, and construction.

Soils mean different things to different people. Soil scientists (pedologists) view soils as being made up of different size mineral particles (sand, silt, and clay) and organic matter. Soils have complex

biological, chemical, physical, mineralogical, and hydrological properties that are always changing with time. Hence, soil is dynamic, teeming with organisms, and is an integral part of the environment. Agronomists, farmers, and gardeners, on the other hand, see soil as a medium for growing crops, pastures, and plants – primarily in the top 50 cm of the Earth's surface. Engineers regard soil as material to build on and excavate, and are usually concerned primarily with moisture conditions and the capacity for soil to become compacted and support structures. However, some people regard soil as "dirt" or "mud" because it makes them "dirty" when they make contact with it.

Pedology (from the Greek *pedon* = soil), is the soil science discipline concerned primarily with understanding the variety of soils and their distribution, and is most directly concerned with the key questions concerning sampling, descriptions, processes of soil formation including the quality, extent, distribution, spatial variability, and interpretation of soils from microscopic to megascopic scales [1]. This description and interpretation of soils can be used in addressing the questions 'What is the soil like?' and 'Where does it come from?' (i.e., provenance determination) in studies relating to characterizing and locating the sources of soils to make forensic comparisons.

Forensic soil scientists are more specifically concerned with soils that have been disturbed or moved (usually by human activity), sometimes comparing them to natural soils, or matching them with soil databases, to help locate the scene of crimes. Forensic soil scientists usually obtain soil samples from crime scenes and suspected control sites from which soil may have been transported by shoes, a vehicle, or a shovel. Soil properties are diverse and it is this diversity, which may enable forensic soil scientists to use soils with a degree of certainty as evidence in criminal and environmental investigations [2, 3].

Forensic soil science is a relatively new activity that is strongly "method orientated" because it is mostly a technique-driven activity in the multidisciplinary soil areas of pedology, soil survey, soil mineralogy, soil chemistry, soil molecular biology, soil geophysics, and forensic science. There are few reviews of the specific application of one soil science discipline to criminalistics at the time of writing (i.e., apart from a recent series of publications by Fitzpatrick [2], Fitzpatrick *et al.* [3], Dawson *et al.* [4], and Ritz *et al.* [5]). Conversely, there are several wider-ranging and soil-related reviews, which provide comprehensive reviews of (i) "forensic geology" [6–12] and, (ii) "geoforensics", which focuses more on the geoscience techniques such as forensic geophysics, forensic remote sensing and geological trace evidence [13–15], and (iii) archaeology [16].

Soil materials are routinely encountered as evidence by police, crime scene investigators, and forensic staff. However, most forensic and physical evidence laboratories either do not accept or are unable to adequately characterize soil materials. The main reason for this is that the morphological, mineralogical, and spectroscopic analytical knowledge required to examine and interpret such soil evidence needs a large amount of training and expertise that is not yet available in most forensic science facilities. In recent years, soil science technology has advanced dramatically and become very specialized and, for this reason, scientists and police investigation units are not using soil information as much as they did previously (e.g., application mainly of petrographic microscope data). Currently, soil analyses are generally only performed in investigations of serious crime and usually where human DNA analyses or analyses of other more commonly used types of trace evidence were not possible. Consequently, there is an opportunity for the application of soil analysis in the forensic examination of soil from a wider spectrum of routine forensic investigations. This review outlines traditional and new soil methods, as well as systematic approaches for the forensic examination of soils.

Soil as a Powerful Contact Trace

Theory of Transfer of Soil Materials from One Surface to Another as a Result of Contact

The transfer of trace evidence is governed by what has become known as the *Locard Exchange Principle* [17], which states that when two surfaces come into physical contact there is a mutual exchange of trace evidence between them. For example, the exchange can take the form of soil material from a location transferring to the shoes of a person who walked through that location. These types of transfers are referred to as *primary transfers* (e.g., evidence is transferred from the soil surface to the shoe and later recovered from the shoe, such as in the treads of

the sole or within the shoe). Once a trace material has transferred, any subsequent movements of that material, in this case from shoes (e.g., from the shoe to the carpet in a vehicles foot well), are referred to as *secondary transfers*. These secondary transfer materials can also be significant in evaluating the nature and source(s) of contact. Hence, the surface of soils can provide information linking persons to crime scenes. Higher order transfers (tertiary transfers) of trace evidence can also occur, which can present interpretative problems for forensic soil scientists because the ultimate source of trace evidence may be extremely difficult to identify.

Aardahl [18] lists the properties of the ideal trace evidence: (i) it is highly individualistic; (ii) it has a high probability of transfer and retention; (iii) it is nearly invisible; (iv) it can quickly be collected, separated, and concentrated; (v) the merest traces are easily characterized; and (vi) it is able to have computerized database capacity. In this context, glitter (i.e., entirely manmade tiny pieces of Al foil or plastic with vapor-deposited Al layer) has been considered to be the ideal contact trace [19]. Soil materials may be considered as approaching the ideal "contact trace", and the following brief discussion considers how closely they fulfill the criteria of Aardahl.

Soil is Highly Individualistic

Diversity of Natural Soils. It is important to understand and know the different kinds of soils and how they form because this helps in making accurate forensic comparisons. To determine the wide variety of soils that occur in the world, it is necessary to understand soil classification systems used to illustrate this. Soil classifications help to organize knowledge about soils, especially in conducting soil surveys. The two international soil classification systems, which are widely used, are the World Reference Base (WRB) [20] and Soil Taxonomy [21]. Many countries also have national and specialized technical classifications [22]. Soil surveys enable the depiction of soils across a landscape and soil maps are made to show the patterns of soils that exist and provide information on the properties of soils. Soil maps are produced at different scales to depict soils over (i) large areas such as the world, countries, and regions (1 : 100 000 or smaller scale), and (ii) detailed areas such as farms (1 : 10 000 or larger scale). A wide diversity of natural soils exists and each has its own characteristics (e.g., morphology, mineralogy, and organic matter composition). For example, according to the United States Department of Agriculture (USDA), which collects soil data at many different scales, there are over 50 000 different varieties of soil in the United States alone. Parent material, climate, organisms, and the amount of time it takes for these properties to interact vary worldwide.

Anthropogenic Soils. Urban soils is a class of *anthropogenic* or *anthropic soils*, a term used in several soil classification systems [20] to indicate soils that are essentially under strong human influence in urban and suburban areas. They are characterized by a strong spatial heterogeneity, which results from the various inputs of exogenous materials (e.g., compost, minerals, technological compounds, and inert, organic, or toxic wastes) and the mixing of the original (natural) soil material (e.g., parks, gardens, landscaping, and cemeteries). Mine or quarry soils are another class of anthropic soils, which are also strongly influenced soils, but found away from cities. Anthropic soils are characterized by a great ecological heterogeneity, and show special distinctness of soil properties.

These specific soils also contain a large array of historical information, which has been proved very useful in understanding and quantifying soil differences in forensic soil comparisons.

The major question posed is how can soils be used to make accurate forensic comparisons when we know that both natural and anthropic soils are highly complex and that there are thousands upon thousands of different soil types in existence? The following key issues are especially important in forensic soil examination because the diversity of soil strongly depends on topography and climate, together with anthropogenic contaminants:

- Forensic soil examination can be complex because of the strong diversity and heterogeneity of soil samples. However, such diversity, heterogeneity, and complexity enables forensic examiners to distinguish between soil samples, which may appear similar to the untrained observer.
- A major problem in forensic soil examination is the limitation in the discrimination power of the standard and nonstandard procedures and methods.

No standard forensic soil examination method exists. The main reasons for this are that examination of soil is concerned with detection of both (i) naturally occurring soils (e.g., minerals, organic matter, soil animals, and included rock fragments) and (ii) anthropogenic soils that contain manufactured materials such as ions and fragments from different environments whose presence may impart soil with characteristics that will make it unique to a particular location [e.g., material from quarries, asphalt, brick fragments, cinders, objects containing lead from glass (*see* **Examination of Fibers and Textiles**), hydrocarbons, paint chips (*see* **Paint: Interpretation**), and synthetic fertilizers with nitrate, phosphate, and sulfate]. In spite of the increasing impact of human activities on soil and the likelihood that all of Earth's ecosystems have been influenced to some extent by humans, many soils still retain their basic morphology imparted by natural soil-forming processes. These anthropogenic properties make the naturally occurring soils even more individualistic.

Soil is Easy to Characterize: Large and Trace Amounts

Historical Analysis Methods for Forensic Soil Samples: 1856–1904. Soil materials are generally easy to characterize, especially by way of the following published historical examples, which demonstrate how large amounts of soil materials have been characterized using quick morphological and light optical methods to solve crime cases. On a Prussian railroad, in April 1856, a barrel which contained silver coins was found on arrival at its destination to have been emptied and refilled with sand. Prof Ehrenburg of Berlin acquired samples of sand from stations along railway lines and used a light microscope to match the sand to the station from which the sand must have come [22]. This is arguably the very first documented case where a forensic comparison of soils was used to help police solve a crime [2]. Then, in 1887, Sir Arthur Conan Doyle published several fictional cases involving Sherlock Holmes such as "A Study in Scarlet" in Beeton's Christmas Annual of London where Holmes can "Tell at a glance different soils from each other . . . has shown me splashes upon his trousers, and told me by their colour and consistence in what part of London he had received them". In 1891, in "The Five Orange Pips", Holmes observed "chalk-rich soil" on boots. This clearly indicates that Conan Doyle was well aware of the key soil morphological properties (color and consistence) and soil mineralogy (chalk) in forensic soil comparisons. Further, as documented by Murray and Tedrow [9, 10], "October 1904, a forensic scientist in Frankfurt, Germany named George Popp was asked to examine the evidence in a murder case where a seamstress had been strangled in a bean field with her own scarf. George Popp successfully examined soil and dust from clothes for identification to solve this real criminalistic case".

Standard/Traditional Analysis Methods for Forensic Soil Samples. The methods of soil analysis used in forensic science are predicated on the size of the sample and the use to which the analytical results will be put. The aim of forensic soil analysis is to associate a soil sample taken from an item (e.g., shoes, clothing, shovel, or vehicle) by police with a specific location. To achieve this aim, the methods of analyses chosen must be able to discriminate between soil samples from different locations. Importantly, the methods used for comparing the samples must be practical (use of standard methods), inexpensive, accurate, and applicable to small and large samples.

The methodology for describing soils has been developed and refined by soil scientists for more than a century [23]. Soils from crime scenes and control sites can be investigated at least in part with traditional soil survey descriptive approaches/techniques; however, these methods must be properly adapted and new methodology must still be developed [3]. Soil morphological descriptors such as color [24], consistency, structure, texture, segregations/coarse fragments (charcoal, ironstone, or carbonates), and abundance of roots/pores are the most useful properties to aid the identification of soil materials (e.g., [23, 25]) and to assess practical soil conditions (e.g., [26]). These soil morphological descriptions follow strict conventions whereby a standard array of data is described in a sequence, and each term is defined according to both the USDA Field book for describing and sampling soils, Version 2.0 [25] and National standard systems (e.g., Australian Soil and Land Survey Field Handbook by McDonald *et al.* [27].

Soil has a High Probability of Transfer and Retention

In general, soil usually has a strong capacity to transfer and stick, especially the fine fractions in soils

(clay and silt size fractions) and organic matter. The larger quartz particles (e.g., >2-mm size fractions) have poor retention on clothes and shoes and carpets. Fine soil material (e.g., their <50–100-μm fractions) may often only occur in small quantities, as illustrated in a hit-and-run case illustrated by Fitzpatrick *et al.* [3], where a remarkably small amount of fine soil was transferred from a gravelly and stony soil on a river bank (control site) to running shoes (forensic evidence items).

Soil Can Quickly Be Collected, Separated, and Concentrated

Although soil forensic investigations are primarily performed in the laboratory, it should be emphasized that soil analyses typically begins at the crime scene and at the control sites. Hence, this section briefly summarizes the general procedures that will ensure that the collected samples are appropriate for the specific objectives of the forensic soil investigation. Soil samples must be carefully collected and handled at the crime scene or control sites using established approaches and then compared by a soil scientist with forensic science experience to ensure that the soil samples can be useful during an investigation. The size and type of samples to be taken are strongly dependent on the nature of the environment being investigated, especially the type of soil and nature of activity that may have taken place at the scene. For example, if suspect footwear is heavily coated with mud on the uppers and the ground is wet and soft, then the control sample should be collected to a depth of around 0–10 cm [3]. Samples of subaqueous soils or sediments from the bottom of river channels, streams, ponds, lakes, or dams can be obtained by pressing a plastic tube or container into the soft submerged soil/sediment and removing it with a scooping action. In deeper water, subaqueous soils/sediment samples can be taken using specialized sampling devices such as the Russian D-auger. In contrast, if the soil is very hard and dry and only the shoe tread was in contact with the soil, then carefully collect the 0–0.5 cm – or thinner. It is critical to wear clean latex gloves but do not use "talc powder" in the gloves because the layer silicate mineral "talc" will contaminate the soil sample. Always use clean tools (e.g., shovel, trowel, artist's palette knife, which are made of stainless steel). Plastic spades and trowels generally lack the strength required to dig soils, especially for most Australian soil conditions. Artist's palette knives are useful for sampling very thin layer surfaces of samples of mud or dust. Preferably place samples in "rigid plastic containers" – rather than polythene bags or paper bags because the package must keep soil lumps intact. Do not use paper envelopes because they easily tear and leak. If soil is adhering to items of clothing or shoes – first air dry the whole garment and then package whole intact sample and garment. If the soil is wet/moist or adhering in a wet/moist condition to objects (e.g., tires, vehicles, garments, or shovels) first air dry then package. However, as in the case of obvious sequential/chronologically deposited layers of soil being present, first remove the "surface layer" and then air dry. Store dry samples at room temperature and ensure containers are sealed and take appropriate caution when storing and transporting. If biological material is attached, package using clean cardboard box/paper bags because samples are prone to rapid deterioration.

Several standard methods are available for quick separation and concentration of soil materials or particles such as, for example, sieving, magnetic extraction, and heavy mineral separation (e.g., Figure 1).

Soil is Almost Invisible

Although a suspect may be unaware that soil material – especially the fine fractions (e.g., <50 μm) – has been transferred directly to the person (e.g., shoes or clothing) or surroundings (e.g., carpet in a suspect's car), soil materials are easily located and collected when inspecting crime scenes or examining items of physical evidence [2–12]. Traces of soil particles can easily and quickly be located directly using hand lenses or light microscopes. For example, Fitzpatrick *et al.* [3] successfully completed a forensic comparison of small amounts of fine yellow–brown soil adhering to a suspect's shoe with a stony/gravelly black control soil submerged in a river where a hit-and-run offender ran through. Although the black-colored control soil comprised 95% alluvial stone and coarse gravel with only 5% clay and silt, a sufficient amount of fine yellow-brown material (<50 μm) was recovered by sieving. This fine soil material closely resembled the fine soil material that was tightly trapped in grooves and treads in the rubber sole of the suspect's shoe [3]. The control sample under typical viewing conditions by the naked

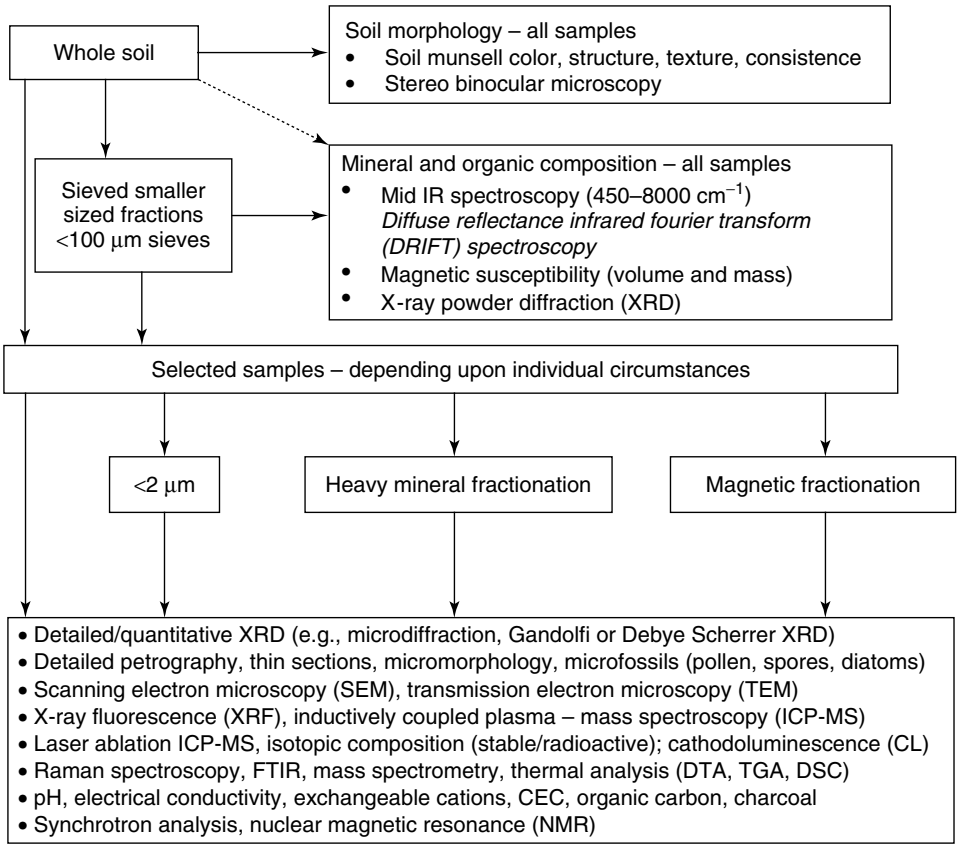


Figure 1 A systematic approach to discriminate soils for forensic soil examinations where FTIR is Fourier transform infrared spectroscopy, DTA is differential thermal analysis, TGA is thermogravimetric analysis, DSC is differential scanning calorimetry, and CEC is cation exchange capacity [3, 28]

eye did not readily observe the yellow-brown color of the fine 5% clay and silt (<50-μm fraction) fractions hidden in the extremely stony/gravelly soil until the sample was sieved and the fine fraction concentrated. This is, for example, often unlike the more obvious bright transfer colors of blood, lipstick smears, and paint chips (*see* **Paint: Interpretation**). Hence, if suspects cannot see fine soil materials adhering to their belongings, especially when they impregnate vehicle carpeting, shoes, or clothing, they will often make little effort to employ a comprehensive cleanup of soil materials.

Computerized Soil Database Capacity

Soil profiles and their horizons usually change across landscapes, and also change with depth in a soil at one location. In fact, soil samples taken at the surface may have entirely different characteristics and appearances from soil dug deeper in the soil profile. One common reason why soil horizons are different at depth is because there is mixing of organic material, in the upper horizons, and weathering and leaching, in the lower horizons.

Erosion, deposition, and other forms of disturbance might also affect the appearance of a soil profile at a particular location. For example, soils on alluvial flats with regular flooding often have clear sedimentary layers. Various soil-forming processes create and destroy layers and it is the balance between these competing processes that will determine how distinct layers are in a given soil. Some of the more common natural processes include the actions of soil fauna (e.g.,

worms, termites), and the depletion and accumulation of constituents including clay, organic matter, and calcium carbonate. In contrast, the main anthropogenic soil-forming processes that destroy layers are excavation (e.g., ploughing and grave digging) and fertilizer applications.

The mapping of the surface and subsurface of both natural and anthropogenic soils provides crucial information as to the origin of a site's specific location, function, land degradation, and management. In Australia, (e.g., [29]) and also in many developed countries in the world, soil data has been encoded into computer-compatible form. Hence, in Australia, for example, a soil map can be produced by downloading information directly from the Internet. The Australian Soil Resources Information System (ASRIS) database has compiled the best publicly available soil information available across Australian agencies into a national database of soil profile data, digital soil, and land resources maps, and climate, terrain, and lithology datasets. Most datasets are thematic grids that cover the intensively used land-use zones in Australia [29]. Hence, the first step when sampling across a region or wider area is to consult these existing/available soil maps of the region of interest in conjunction with or with help from experienced soil scientists. The areas of broadly similar soil type can then be identified as high priority areas for further sampling and comparative analyses – using morphological and analytical information. However, in the absence of obvious features, systematic sampling of the area should be conducted (cross-pattern to fully characterize the soil patterns in the area).

Common and Standardized Techniques Used by Forensic Soil Scientists

Evaluation of Degree of Similarity between Questioned Samples and Control Soil Samples

It is important to first define the word "compare" because no two physical objects can ever, in a theoretical sense, be the same [10]. Similarly, a sample of soil or any other earth material cannot be said, in the absolute sense, to have come from the same single place. However, according to Murray and Tedrow [10], it is possible to establish "with a high degree of probability that a sample was or was not derived from a given place". For example, a portion of the soil (or other earth material) could have been removed to another location during human activity. Pye [11] summarizes different schemes commonly used in the United Kingdom to convey weight of evidence relating to forms of comparisons such as trace or DNA evidence. For example, he has developed "verbal categories" ranging from 0 (no scientific evidence) to 10 (conclusive) – with no statistical significance of the ranks implied. He also states that there is a long history of the use of numerical scales in the context of evidential and legal matters.

Approaches and Methods for Making Comparisons between Soil Samples

Forensic soil scientists must first determine if uncommon and unusual particles, or unusual combinations of particles, occur in the soil samples and must then compare them with similar soil in a known location [2]. To do this properly, the soil must be systematically described and characterized using standard soil testing methods to deduce whether a soil sample can be used as evidence (Figure 1).

Methods for characterizing soils for a forensic comparison involve subdividing methods into three steps: (i) descriptive (morphological), analytical (Figure 1), and spatial information (e.g., mapping).

Soil characterization requires a multidisciplinary approach, which combines descriptive, analytical, and spatial information (e.g., mapping) steps in the following three stages:

Stage 1 – Rapid characterization of composite soil particles in whole soil or bulk samples for screening of samples (Figure 1).
Stage 2 – Detailed characterization and quantification of composite and individual soil particles following sample selection, size fractionation, and detailed mineralogical and organic matter analyses using advanced soil analytical methods (Figure 1).
Stage 3 – Integration and extrapolation of soil information from one scale to next, to build a coherent model of soil information from microscopic observations to the landscape scale (e.g., using existing soil maps or field mapping information). In forensic soil science, a provenance examination or determination, also known as geographic sourcing, has developed to identify the

origin of a sample by placing constraints on the environment from which the sample originated.

Stage1: Initial Characterization of Composite Soil Particles in Whole Samples for Screening

In the initial screening or comparison examination of whole soil samples, soil morphology, low magnification light microscopy, X-ray diffraction (XRD), diffuse reflectance infrared Fourier transform (DRIFT) spectroscopy, and magnetic susceptibility (volume and mass) are used to compare samples *via* bulk morphology, mineralogy, and organic matter characterization.

Soil Morphology – Soil Profiling. Soil morphology is defined as the branch of soil science and pedology that deals with the description, using standard terminology, of *in situ* spatial organization and physical properties of soils regardless of potential land use. Soil morphological interpretation provides a visual, quick, and nondestructive approach to screen and discriminate among many types of forensic soil samples. Morphological soil descriptors are arguably the most common and probably the simplest – and it is for this reason that all samples are characterized first using the four key morphological descriptors of color, consistency, texture, and structure (Figure 1). In many respects, the soil resembles a sandwich with these easily observed characteristics and thickness, which conveys the concept of different soil layers with different properties. In soil samples from crime scenes and control sites in question where soil may have been transported, by vehicle, foot, or shovel, a complete visual description of the soil is essential because it serves as a basis for soil identification, classification [20, 21], correlation, mapping, and interpretation [2, 3].

A checklist of six key macromorphological descriptors has been compiled from standard techniques used in soil science (e.g., [23]) for assessing the soil properties for forensic examinations. Observations of depth changes in various properties are recommended: *viz* consistence, color, texture, structure, segregations/coarse fragments (carbonates and ironstone), and abundance of roots in the different layers or horizons.

The use of petrography is a major and often precise method of studying and screening soils for discrimination in forensics (Figure 1). For example, nearly 50 common minerals (e.g., gypsum), as well as several less-common minerals can easily be seen by the naked eye, but using a hand lens or low power stereo-binocular microscope (**Microscopy: Low Power**) enables the forensic soil scientist to better detect mineral properties (e.g., particle shape and surface texture) and provide more accurate mineral identification. The petrographic microscope (**Microscopy: High Power**) is also commonly available for studying microfossils (pollen grains, grass spores, opal phytoliths, diatoms [11, 16] (*see* **Diatoms**; **Microscopy: Low Power**); and thin sections of soil samples (resin impregnated), minerals, and rocks. Thin sections of soil materials are mounted on a glass slide and viewed with the petrographic microscope under different incident light conditions through its special attachments (e.g., [30]). Where possible, such micromorphological investigations are used to supplement and verify features in macromorphological descriptions. Macromorphological and petrographic descriptors are useful in assessing soil conditions because of the following:

- They involve rapid field and laboratory assessments. Other methods, such as more detailed mineralogy (see below) and geochemistry, are complex and more costly to carry out.
- They can be used to evaluate causes for variations in soil condition induced by weathering (that may range from recent to thousands, to millions, or even billions of years), anthropogenic activities, land management, hydrology, and weather conditions.

Mineral and Organic Matter Identification and Composition. Once a familiarity with the morphology of the materials has been achieved using visual and light microscopic methods, most of the mineralogical and organic matter components in a particular whole or bulk soil sample can be determined using the following three selected methods:

X-ray powder diffraction (XRD) methods. XRD methods are arguably the most significant for both qualitative and quantitative analyses of solid materials in forensic soil science [2, 3, 31]. Extremely small sample quantities (e.g., few to a few tens of milligrams) as well as large quantities can be successfully analyzed using XRD. The critical advantage

of XRD methods in forensic soil science is based on the unique character of the diffraction patterns of crystalline and even poorly crystalline soil minerals. Elements and their oxides, polymorphic forms, and mixed crystals can be distinguished by nondestructive examinations. Part of the comparison involves identification of as many of the crystalline components as possible, either by reference to the International Centre for Diffraction Data (ICDD) Powder Diffraction File [31], or to a local collection of standard reference diffraction patterns, coupled with expert interpretation [2, 3, 31, 32].

Diffuse Reflectance Infrared Fourier Transform (DRIFT) method. The main advantages of DRIFT spectroscopy are that the analysis is nondestructive and can be rapidly applied, and that the mid-infrared portion of the electromagnetic spectrum is sensitive to organic materials, clay minerals, and quartz, due to absorption of infrared light at vibrational frequencies of the molecular functional groups [32–36]. As such, this technique is a powerful qualitative tool, which can then be used semiquantitatively to predict analytes of interest when combined with partial least-squares (PLS) (MIR-PLS) or other chemometric techniques (Stage 2).

A new rapid mid-infrared (MIR) spectroscopic method, coupled with chemometric approaches, specifically PLS (MIR-PLS) modeling has been developed by Janik *et al.* [33, 34] as applied to soils to predict soil physicochemical properties and has been routinely applied to rapidly screen and compare crime scene samples (Figure 1). Principal component analysis (PCA), which models the spectral signatures alone, is also a powerful discriminatory tool, providing an objective method of comparing the mid-infrared spectra of the soil samples being examined when enough samples are available for the technique to be viable.

Mass and Volume Magnetic Susceptibility Methods. Added to the above two rapid methods and techniques are the use of magnetic susceptibility methods, which should also always be used before moving to the more costly detailed methods (Stage 2), which require sample separation (Figure 1). Mineral magnetic techniques are a relatively recent development (post 1971) and have now become a very powerful and widely used research tool to characterize natural materials in landscapes (e.g., [37]).

Stage 2: Detailed Characterization of Composite and Individual Soil Particles

X-ray Diffraction (XRD) Methods. In many soil forensic case investigations, the amount of soil available for analyses (e.g., on clothing or soles shoes) may preclude routine bulk analyses. In such situations, it is best to use an XRD fitted with a system for analysis of extremely small samples (e.g., thin coatings or single particles of the order of 2–10 mg) loaded into thin glass capillaries or deposited onto Si low background holders for XRD analysis [38]. For analysis in a Gandolfi or Debye-Scherrer powder camera, extremely small specimens (e.g., single mineral particles and paint flakes) can be mounted on the end of glass fibers. Consequently, according to Kugler [31], X-ray methods are often the only ones that will permit further differentiation of materials under laboratory conditions. According to Murray [8], "Quantitative XRD could possibly revolutionise forensic soil examination". For example, XRD patterns can also be likened to finger print comparisons between soil samples and how closely they relate to each other [3]. However, what is the significance of the close similarity in XRD patterns to the degree of similarity in terms of mineralogical composition? If the two soil samples, for example, contain only one crystalline component such as quartz (i.e., silicon dioxide), which is very common in soils, the significance of the similarity and its evidential value in terms of comparison criteria will be low. If, however, the two soils contain four or five crystalline mineral components, some of them unusual, then the degree of similarity will be considered to be high [3].

Methods such as XRD, X-ray fluorescence (XRF), and DRIFT spectroscopy, whose results partially overlap, are used. These overlapping results confirm each other and give a secure result to the examination.

Scanning Electron Microscopes (SEM) and Transmission Electron Microscopes (TEM). Scanning electron microscopes (SEM) (*see* **Microscopy: Scanning Electron Microscopy**) and transmission electron microscopes (TEM) are frequently used to examine the morphology and chemical composition (*via* energy dispersive spectroscopy) of particles magnified to over 100 000 times their original size making them very useful for discrimination (e.g., [39–41]).

Soil minerals, fossils, and pollen spores that occur in soils [42] can be described and analyzed in detail by SEM and TEM, and are therefore very useful indicators when studying soil samples (e.g., [11, 16, 39–41]).

Elemental Analysis. The following range of more prevalent instrumental techniques are frequently used to determine the inorganic constituents in soil samples: XRF, atomic absorption spectroscopy (AAS), inductively coupled plasma (ICP) spectrometry, inductively coupled plasma-optical emission spectrometry (ICP-OES) (sometimes called *ICP-AES*), inductively coupled plasma-mass spectrometry (ICP-MS), and neutron activation analysis (NAA) [4–12]. Several geochemical techniques, using isotope ratios and geochemical signatures have been utilized in forensic work (e.g., [43]).

Biological Methods. Fossil pollen grains (*see* **Palynology**) and grass spores are preserved in many soils that are not strongly acidic (<pH 4) or alkaline (>pH 6). These reproductive particles are produced in large amounts by trees, shrubs, and grasses [7–11, 16]. Opal phytoliths (silica-rich) and calcium phytoliths are mineral deposits that form in and between plant cells. Marumo and Yanai [42] used opal phytoliths to differentiate soils with similar mineralogy. As stated earlier, FTIR can be used to characterize soil organic constituents (fats, waxes, proteins, cellulose, hemicellulose, and lignin) in soils [4, 33–36]. Other emerging soil forensic methods such as (i) plant wax markers analysis are summarized in Dawson *et al.* [4], (ii) plant fragment DNA analysis [4], and (iii) microbial fingerprinting using a variety of molecular biological techniques to analyze the diversity in soil microbial communities for forensic soil comparison [4, 44]. Several soil forensic studies have been reported [44, 45] to show that a soil bacterial community DNA profile could be obtained from small samples of soil recovered from potential crime scenes (e.g., shoes or clothing) with the profiles being representative of the site of collection.

Combined Methods. All these techniques and others listed in Figure 1 (e.g., heavy and magnetic mineral separations, routine soil chemical analysis, laser ablation, Raman spectroscopy, thermal analysis, NMR, and synchrotron analysis) in combination achieve reliable, definite, and accurate results, and provide additional information about the mineralogical, chemical, and physical properties of the suspected soil material.

Stage 3: Landform and Soil Mapping

Integration and extrapolation of soil information are necessary because soils from the crime scene and control site may constitute a highly variable continuum. Hence, integration of published soil maps or field mapping information together with other spatially held information, such as (i) terrain analysis Digital Elevation Model (DEM), (ii) regolith, geological, or vegetation maps, and (iii) remote sensing or geophysics data [13–15] are designed to study relationships between soils, landforms, and/or the stratigraphy of parent materials. This information will also ensure (i) better-informed sampling (i.e., pedometric testing of how "similar" soil on a suspect's shoe is to a scene of crime [46]), and (ii) construction of a coherent model of soil information from microscopic observations to the landscape scale (e.g., physical, chemical, or biological mechanistic process models). Decision making in forensic soil science is sometimes guided by mechanistic process models describing processes with physical, chemical, or biological mechanisms [2–15]. Some models use multiple data layers as spatial input. The data required for the analysis are frequently found in and extracted from soil and environmental databases such as geographic information systems (GISs).

Soil, regolith, and geological maps are commonly being used by forensic soil scientists in developing models to predict where sites of particular soil materials are located. An example of this relationship is from the staff in the Centre for Australian Forensic Soil Science (CAFSS) using soil maps and conducting field soil survey investigations to solve a double murder case [47, 48]. Morphological, chemical, physical, and mineralogical properties were used to identify similarities between soil found on a shovel taken from the suspect's vehicle and soil subsequently located in a quarry. Samples were indistinguishable or strongly matched in terms of all comparison criteria used, thus revealing the location of two buried bodies.

Forensic soil examination can be complex because of the diversity and heterogeneity of soil samples. However, such diversity and complexity enables

forensic examiners to distinguish between soils, which may appear to be similar. There is a general lack of expertise in this relatively new area among soil scientists. For research and practical application in this area to grow appreciably, it will need to be considered and taught as an integral part of both soil science and forensic science courses [2, 3]. Finally, an attempt should be made to develop and refine methodologies and approaches to develop a practical "soil forensic manual with soil kit for sampling, describing, and interpreting soils" [3].

References

[1] Wilding, L.P. (1994). Factors of soil formation: contributions to pedology, in *Factors of Soil Formation: A Fiftieth Anniversary Retrospective*, SSSA Special Publication 33, Soil Science Society of America, Madison, pp. 15–30.

[2] Fitzpatrick, R.W. (2008). Nature, distribution and origin of soil materials in the forensic comparison of soils, in *Soil Analysis in Forensic Taphonomy: Chemical and Biological Effects of Buried Human Remains*, M. Tibbett & D.O. Carter, eds, CRC Press, Boca Raton, pp. 1–28.

[3] Fitzpatrick, R.W., Raven, M.D. & Forrester, S.T. (2009). A systematic approach to soil forensics: criminal case studies involving transference from crime scene to forensic evidence, in *Criminal and Environmental Soil Forensics – Soil Forensics International, Edinburgh Conference Centre*, 30 October–1 November 2008, K. Ritz, L. Dawson & D. Miller, eds, Springer Science+Business Media B.V., pp. 105–127.

[4] Dawson, L.A., Campbell, C.D., Hillier, S. & Brewer, M.J. (2008). Methods of characterizing and fingerprinting soil for forensic application, in *Soil Analysis in Forensic Taphonomy: Chemical and Biological Effects of Buried Human Remains*, M. Tibbett & D.O. Carter, eds, CRC Press, Boca Raton, pp. 271–315.

[5] Ritz, K., Dawson, L. & Miller, D. (eds) (2008). *Criminal and Environmental Soil Forensics. Soil Forensics International, Edinburgh Conference Centre*, 30 October -1 November 2008, Springer.

[6] Donnelly, L.J. (2003). The applications of forensic geology to help the police solve crimes. European geologist, *Journal of the European Federation of Geologists* **16**, 8–12.

[7] Murray, R.C. (1982). Forensic examination of soil, in *Forensic Science Handbook*, R. Saferstein, ed, Prentice Hall, Englewood Cliffs, p. 1.

[8] Murray, R.C. (2004). *Evidence from the Earth: Forensic Geology and Criminal Investigation*, Mountain Press Publishing, Missoula, p. 227.

[9] Murray, R.C. & Tedrow, J.C.F. (1975). *Forensic Geology: Earth Sciences and Criminal Investigation (Republished 1986)*, Rutgers University Press, New York, p. 240.

[10] Murray, R.C. & Tedrow, J.C.F. (1991). *Forensic Geology*, Prentice Hall, Englewood Cliffs, p. 240.

[11] Pye, K. (2007). *Geological and Soil Evidence: Forensic Applications*, CRC Press, Boca Raton, p. 335.

[12] Pye, K. & Croft, D. (eds) (2004). *Forensic Geoscience: Principles, Techniques and Applications*, Special Publication 232, Geological Society of London, p. 318.

[13] Harrison, M. & Donnelly, L.J. (2008). Locating concealed homicide victims; developing the role of geoforensics, in *Criminal and Environmental Soil Forensics. Soil Forensics International, Edinburgh Conference Centre*, 30 October–1 November 2008, K. Ritz, L. Dawson & D. Miller, eds, Springer-Verlag.

[14] Ruffell, A. & McKinley, J. (2005). Forensic geoscience: applications of geology, geomorphology and geophysics to criminal investigations, *Earth-Science Reviews* **69**, 235–247.

[15] Ruffell, A. & McKinley, J. (2008). *Geoforensics*, John Wiley & Sons, Chichester.

[16] Garrison, E.R. (2003). *Techniques in Archaeological Geology*, Springer-Verlag, Berlin.

[17] Chisum, W. & Turvey, B. (2000). Evidence dynamics: Locard's exchange principle and crime reconstruction, *Journal of Behavioral Profiling* **1**(1), 1–15.

[18] Aardahl, K. (2003). *Evidential Value of Glitter Particle Trace Evidence*, Master's Thesis, National University, San Diego.

[19] Blackledge, R.D. & Jones, E.L. (2007). All that glitters is gold! in *Forensic Analysis on the Cutting Edge: New Methods for Trace Evidence Analysis*, R.D. Blackledge, ed, John Wiley & Sons, pp. 1–32.

[20] IUSS Working Group WRB (2006). *World Reference Base for Soil Resources 2006*, World Soil Resources Reports No. 103, FAO, Rome.

[21] Soil Survey Staff (1999). *Soil Taxonomy - a Basic System of Soil Classification for Making and Interpreting Soil Surveys*, *USA Agriculture Handbook*, 2nd Edition, United States Department of Agriculture, Natural Resources Conservation Service, Vol. 436, p. 869.

[22] Science and Art. (1856). Curious use of the microscope, *Scientific American* **11**, 240.

[23] Schoeneberger, P.J., Wysocki, D.A., Benham, E.C. & Broderson, W.D. (eds) (2002). *Field Book for Describing and Sampling Soils*, Natural Resource Conservation Service, National Soil Survey Center, Lincoln, Version 2.0.

[24] Munsell Soil Color Charts (2000). X-Rite, Incorporated and GretagMacbeth AG/LLC USA, Grand Rapids.

[25] Fitzpatrick, R.W., Powell, B., McKenzie, N.J., Maschmedt, D.J., Schoknecht, N. & Jacquier, D.W. (2003). Demands on soil classification in Australia, in *Soil Classification: A Global Desk Reference*, H. Eswaran, T. Rice, R. Ahrens & B.A. Stewart, eds, CRC Press, Boca Raton, pp. 77–100.

[26] Fitzpatrick, R.W., McKenzie, N.J. & Maschmedt, D. (1999). Soil morphological indicators and their importance to soil fertility, in *Soil Analysis: An Interpretation*

Manual, K. Peverell, L.A. Sparrow & D.J. Reuter, eds, CSIRO Publishing, Melbourne, pp. 55–69.

[27] McDonald, R.C., Isbell, R.F., Speight, J.G., Walker, J. & Hopkins, M.S. (1990). *Australian Soil and Land Survey. Field Handbook*, 2nd Edition, Inkata, Melbourne.

[28] Fitzpatrick, R.W., Raven, M. & McLaughlin, M.J. (2006). Forensic soil science: an overview with reference to case investigations and challenges, in *Proceedings of the First International Workshop on Criminal and Environmental Forensics*, R.W. Fitzpatrick, ed, Centre for Australian Forensic Soil Science, Perth, p. 9. http://www.clw.csiro.au/cafss/.

[29] Johnston, R.M., Barry, S.J., Bleys, E., Bui, E.N., Moran, C.J., Simon, D.A.P., Carlile, P., McKenzie, N.J., Henderson, B.L., Chapman, G., Imhoff, M., Maschmedt, D., Howe, D., Grose, C. & Schoknecht, N. (2003). ASRIS: the database, *Australian Journal of Soil Research* **41**(6), 1021–1036.

[30] Stoops, G. (2003). *Guidelines for Analysis and Description of Soil and Regolith Thin Sections*, Soil Science Society of America, Madison.

[31] Kugler, W. (2003). X-ray diffraction analysis in the forensic science: the last resort in many criminal cases. JCPDS - international centre for diffraction data, *Advances in X-ray Analysis* **46**, 1–16. http://www.icdd.com/resources/axa/vol46/v46_01.pdf.

[32] Rendle, D.F. (2004). Database use in forensic analysis, *Crystallography Reviews* **10**(1), 23–28.

[33] Janik, L. & Skjemstad, J.O. (1995). Characterization and analysis of soils using mid-infrared partial least squares. II. Correlations with some laboratory data, *Australian Journal of Soil Research* **33**, 637–650.

[34] Janik, L.J., Merry, R.H. & Skjemstad, J.O. (1998). Can mid infrared diffuse reflectance analysis replace soil extractions? *Australian Journal of Experimental Agriculture* **38**, 637–650.

[35] Van der Marel, H.W. & Beutelspacher, H. (eds) (1976). Clay and related minerals, in *Atlas of Infrared Spectroscopy of Clay Minerals and their Admixtures*, Elsevier Scientific, Amsterdam.

[36] Nguyen, T.T., Janik, L.J. & Raupach, M. (1999). Diffuse reflectance infrared Fourier transform (DRIFT) spectroscopy in soil studies, *Australian Journal of Soil Research* **129**, 49–67.

[37] Thompson, R. & Oldfield, F. (1986). *Environmental Magnetism*, Allen and Unwin, London, Chapter 2.

[38] Fitzpatrick, R.W., Raven, M.D. & Forrester, S.T. (2007). *Investigation to Determine if Shoes Seized by South Australia Police Contain Soil Materials that Compare with a Control Soil Sample from the Bank of the Torrens River, Adelaide*, CSIRO Land and Water Client Report CAFSS_027. December 2006. Restricted Report, p. 36.

[39] Smale, D. (1973). The examination of paint flakes, glass and soils for forensic purposes, with special reference to the electron probe microanalysis, *Journal of the Forensic Science Society* **13**, 5–15.

[40] Pirrie, D., Butcher, A.R., Power, M.R., Gottlieb, P. & Miller, G.L. (2004). Rapid quantitative mineral and phase analysis using automated scanning electron microscopy (QemSCAN); potential applications in forensic geoscience, in *Forensic Geoscience: Principles, Techniques and Applications*, K. Pye & D.J. Croft, eds, Geological Society, London, Special Publications 232, pp. 103–122.

[41] McVicar, M.J. & Graves, W.J. (1997). The forensic comparison of soils by automated scanning electron microscopy, *Journal of the Canadian Society of Forensic Science* **30**, 241–261.

[42] Marumo, Y. & Yanai, H. (1986). Morphological analysis of opal phytoliths for soil discrimination in forensic science investigation, *Journal of Forensic Sciences* **31**, 1039–1049.

[43] Trueman, C., Chenery, C., Eberth, D.A. & Spiro, B. (2003). Diagenetic effects on the oxygen isotope composition of bones of dinosaurs and other vertebrates recovered form terrestrial and marine sediments, *Journal of the Geological Society of London* **160**, 895–901.

[44] Petrisor, I.G., Parkinson, R.A., Horswell, J., Waters, J.M., Burgoyne, L.A., Catcheside, D.E.A., Dejonhe, W., Leys, N., Vanbroekhoven, K., Pattnaik, P. & Graves, D. (2006). Microbial forensics, in *Environmental Forensics: A contaminant Specific Guide*, R.D. Morrison & B.L. Murphy, eds, NL Elsevier, Amsterdam, pp. 227–251.

[45] Morgan, R. & Bull, P.A. (2007). The nature, philosophy and practice of forensic sediment analysis, *Progress in Physical Geography* **31**, 1–16.

[46] Junger, E.P. (1996). Assessing the unique characteristics of close-proximity soil samples: just how useful is soil evidence, *Journal of Forensic Sciences* **41**, 27–34.

[47] Zala, K. (2007). Dirty science: soil forensics digs into new techniques, *Science* **318**, 386–387.

[48] Fitzpatrick, R.W., Raven, M.D., Heath, M. & Rinder, G. (2007). How soil evidence helped solve a double murder case: a display, *2nd International Conference on Criminal and Environmental Soil Forensics*, Edinburgh, 30 October–1 November 2007. Book of Abstracts. Poster and Podcast uploaded on The Centre for Australian Forensic Soil Science (CAFSS) websites: http://www.clw.csiro.au/cafss/; http://www.csiro.au/files/files/p9ng.pdf.

ROBERT W. FITZPATRICK

Sole and Shoewear Databases ***see*** Footwear and Foot Impressions: Databases

Source Monitoring ***see*** Deception: Truth Serum; Memory: Reconstructive

Sources of DNA Evidence ***see*** DNA: Sources of

Speaker Recognition

Introduction

The global success of mobile telephony provides an opportunity for extensive use of audio recordings to search and establish links between individuals and criminal activities. However, establishing these links through speech applying forensic speaker recognition often remains the last resort for forensic investigation and as evidence in court, in the absence of any other clues. The reason lies in the following difficulties: first, the behavioral component of voice is dominant over the physical component of this biometric modality, preventing the display of highly discriminatory features yet; second, the quality of audio recordings captured in forensic conditions is most of the time far from ideal; and third, the absence of any known underlying model prevents a symbolic description of the speaker-dependent information and limits the study to recognition approaches. To tackle these pitfalls, forensic speaker recognition has been established as a multidisciplinary area of study, predominantly combining phonetics, linguistics, speech signal processing, and forensic statistics.

Speech as a Trace

Speech is a behavior developed by human beings for communication. Speech production is a complex process depending on physiological traits, like the length and shape of the vocal tract and the dynamic configuration of the organs involved in articulation. It also depends on environmental and sociolinguistic factors, like the level of education, the dialectal particularities and the linguistic context. Speech can convey up to six types of information simultaneously (referential, poetic, emotive, conative, phatic, and metalingual). It is an inherently transient phenomenon that can only be imperfectly captured as a signal. A forensic trace involving speech is a storage medium containing an analog or digitally recorded speech signal (*see* **Computers**).

Where the recording consists of a wiretapped telephone call, the trace is generally recorded in uncontrolled conditions, including undesired variations in signal quality due to background noise, transmission channels, and recording devices. Anonymous calls are generally short, from seconds to minutes. Where it is a monologue, it may be a prerecorded message, possibly modified by a filtering or editing procedure. From a lexical point of view, the themes are targeted, e.g., abuse, extortion, obscenity, and/or threats. Results of wiretapping procedures can reach hundreds of hours of recording. Their lexical content is varied but some utterances may refer to internal codes for groups or organizations.

Owing to the absence of highly discriminative features in speech and the uncontrolled recording conditions, individualization of speakers remains a very difficult field of inference of identity in forensic science at the source level, but when solved, inferences at the activity and even at the offense level are most of the time straight.

Forensic Speech Analysis

The speaker-dependent features are spread out over the information conveyed in the different levels of speech communication, but they originate in the two main stages of voice production: the language generation and the speech production. For speaker-recognition purposes, these features can be classified as high- and low-level features. High-level features refer more to the linguistic information of speech i.e., as a behavior and its understanding by human beings while low level features refer more to the speaker's vocal physiology and signal processing by automatic approaches.

The main requirements for the methods applied by human beings and computers for forensic speech

analysis are the independence of the text, the ability to handle minimal length recordings and a superior robustness regarding noise, channels, and variations in recording conditions. Efficient analysis strategies combine both high- and low-level features; many of the high-level features need more speech material than the low-level features for a reliable extraction, but they are generally more robust regarding channel variation. Therefore, the methodology for forensic analysis of human speech should ideally combine expert-based and automatic approaches.

Expert-Based Approaches

Expert knowledge has developed around three main approaches: auditory-perceptual, visual, and phonetic–acoustic.

The auditory-perceptual approach was developed during the first part of the twentieth century. It consists of a detailed auditory analysis of the parameters of the voice such as the pitch, the timbre and the voice quality, the parameters of the speech such as the articulation, the diction, the prosody (speaking rate, pauses, and intonation), any speech defects, and the parameters of the language spoken. It includes linguistic observations of lexical, phonological, morphological, syntactic and idiomatic features, the study of the extent and the variation of the dialect, the accent and the idiolect, and also takes into account paralinguistic features such as breathing patterns. The auditory-perceptual approach does not easily lend itself to validation; it remains a first step in the analysis, but is clearly insufficient if used alone.

The visual approach was developed in the 1960s in the United States under the name of 'voiceprint' technique. It comprises a visual comparison and interpretation of broadband speech spectrograms made by 'experts', most of whom had no scientific education in a speech science or related area. From its conception, the theoretical and logical grounds of the technique were contested. First, the term *voiceprint* is a misleading analogy to the fingerprint and its associated qualities of uniqueness, inalterability, and permanence; none of them apply to the voice and to the speech spectrogram. Second, confusion exists between the degree of reproducibility of the production of spectrograms (analysis) and the high degree of uncertainty associated with the individualization process (interpretation), which can be described as closer to art than science. Since the publication of the National Academy of Science (NAS) report in 1979, some US jurisdictions reject the admissibility of the technique although some still allow it. The Daubert decision (1993), advocating a reliability standard for the admissibility of scientific evidence, clearly militates against the use of this kind of nonvalidated approach, but the US jurisdictions have not adopted a uniform practice in this matter yet [1, 2] (*see* **Expert Opinion: United States**).

The acoustic–phonetic approach was developed during the 1980s. The methodology consists of a set of instrumental measurements of particular acoustic speech parameters present in the segmental and suprasegmental level of the speech. The segmental analysis focuses on the duration and the spectral distribution of energy of the segments, the vowel formant frequencies, their trajectory, and, more recently, their dynamics. The suprasegmental analysis concentrates on the measurements of the long-term features, like the long-term average spectrum of the speech signal, the long-term distribution of formants and different parameters of the fundamental frequency (F_0), as the long-term F_0, the mean F_0, and the deviation of F_0. Once measured, the parameters are interpreted using statistical information on their distributions [3, 4].

Automatic Approaches

Attempts to use computers for forensic speech analysis started in the 1970s, first using semiautomatic methods and from the 1990s using text-independent automatic methods. These automatic methods rely on speech processing, to extract speaker-dependent feature vectors from the speech signal, on pattern recognition to compute comparison scores between feature vectors, and on Bayesian statistics to estimate the evidential value of these comparison scores [5, 6].

State-of-the-art automatic speaker-recognition systems still rely on low-level speaker-dependent features, extracted from the short-time spectral level. The spectrum of the speech signal, analyzed in short-term windows, is directly related to the shape of the vocal tract, which presents speaker-dependent specificities. The spectral envelope is described efficiently using linear predictive coding (LPC) in an all-pole model with 10–16 coefficients. However, as these coefficients are correlated, the cepstrum transform has

been proposed in order to obtain pseudo-orthogonal linear prediction cepstrum coefficients (LPCC). They may also be obtained from a perceptually based mel-filter spectral analysis (mel frequency cepstral coefficients (MFCC)). Other low-level features, which capture the dynamic speech patterns, like delta and even delta–delta features have also been proposed, but LPCC and MFCC are the most widely used low-level features for automatic speaker recognition. Current systems also intend to take advantage from high-level speaker-dependent features, mainly contained in the phonotactics, the prosody, and the idiolect. In phonotactics, speaker-dependent information is embedded in the particular use and realizations of the phones and syllables that presents a highly language-dependent variability. Prosody is the combination of instantaneous energy, intonation, speech rate, and unit duration, which all exhibit speaker specificity. The idiolect contains the information related to the speaker-specific use of a language [7].

Numerous pattern-recognition methods have been developed to model and compare the speaker-dependent spectral features, i.e., vector quantization (VQ), ergodic hidden Markov models (E-HMM), artificial neural networks (ANN) and lately support-vector machines (SVM), but most of the current systems are based on Gaussian mixture models (GMM). GMM is a generative method where a mixture of multidimensional Gaussians model the probability distribution of the speaker-dependent features. Various kinds of score normalization have also been developed using prior information like the handset type or the gender information during the normalization parameter computation to improve the performance of the methods. Forensic systems mainly use a particular class of normalization techniques that rely more on the estimation of the alternative hypothesis used in the Bayesian framework than on a normalization scheme, the world-model and cohort-based normalizations [8].

Until the last decade of the twentieth century, the forensic interpretation of the results given by automatic approaches remained difficult, as solutions concentrated on decision theory and frameworks applied in commercial applications of automatic speaker recognition: speaker verification ($1:1$) or speaker identification ($1:N$ or $1:N+1$).

Nowadays, the evaluation of forensic evidence in speaker individualization cases is based on the latest developments of forensic statistics. The automatic systems not only deliver likelihood ratios but also the accuracy and the calibration of the probabilities inferred by these likelihood ratio values are tested [7, 9].

Forensic Individualization through Speech

Forensic individualization through speech is a question of inference of identity of source. In essence, the answer to the question that is provided by science remains inductive and therefore relative. Nevertheless, the use of a framework based on logic and forensic statistics to interpret the results of forensic speech analysis, allows for the most scientifically correct answer to be reached (*see* **Identification and Individualization**).

With this framework, a practitioner may report logical, robust, and balanced statistics to a court of justice, presenting the evidence according to the prosecution and the defense hypothesis as a likelihood ratio. Moreover, the examiner may combine likelihood ratios evaluated for the auditory-perceptual and acoustic–phonetic approaches based on subjective estimates of the probability of the occurrence of specific features with likelihood ratios estimates of automatic approaches, based on more objective probabilities.

The adoption of this interpretation framework by the practitioners is an ongoing process, but the forensic speaker-recognition community has pioneered from a long time in this matter actively searching for solutions to resolve the logical flaws of an interpretation given in terms of posterior probabilities of common origin for the trace and the source [6, 7, 10].

Validation and Practice

Most of the forensic laboratories still opt for one of the two approaches, even if, in many respects, the expert-based and automatic approaches appear to be complementary. For example, much larger amounts of speech data can be handled automatically than manually, which affects both validation and practice. Validation of automatic methods may be envisaged on a far larger scale than expert-based methods, bringing far more accurate figures for the former than for the second method [11]. Automatic methods are also less dependent on the language spoken, making their validation possible for different languages and

their versatility superior for practical use. In practice, automatic approaches can be used in about one-third of cases, particularly for cases with a large amount of speech material.

Expert-based methods are claimed to be more flexible for cases with qualitative and quantitative limitations and more robust for speech samples with strongly mismatched behavioral and technical conditions, or those containing linguistic or dialectal particularities. Finally, since experts are phoneticians or linguists, the expert-based methods are more easily explained in a court of justice, where they often perceive automatic methods as 'black boxes' [12].

Research and Development

Improvement in forensic speaker recognition requires to study the dependences of the results obtained with the expert-based and automatic approaches, as they appear to be complementary. This, in turn, can only be achieved through the coordination of research and development.

Methods may be developed and implemented to monitor, more carefully, the inevitable subjective components in the expert-based approaches. The development of double-blind approaches for selecting and grouping speech samples in the preanalysis phase of cases may reduce the effect of confirmation bias. It may also increase the calibration of the practitioner regarding the voice, the speech, and the language particularities of the speakers involved in the case. Studies in order to estimate the statistical probabilities of the features used in the expert-based approaches could help the practitioners to improve and calibrate the subjective probabilities they assign to these features based on their training and experience.

The relevance of the features selected may be studied using Bayesian networks and their correlations analyzed with multivariate likelihood ratio approaches. Collaborative exercises and proficiency testing are recognized tools to monitor the expert-based practice, but they encounter two major difficulties in the field of forensic speaker recognition: they are very time consuming and the language dependencies constitute a barrier to their organization at an international level.

The robustness of the automatic approaches to mismatch conditions may also be improved with further development of noise, channel, and recording distortion compensation strategies. Phonetics and linguistics may further contribute to develop more robust and automatic feature extraction and processing of higher level features from speech samples of forensic quality. Knowledge of the language dependency of both approaches, expert based and automatic should also progress, as cases may involve a broad variety of languages.

References

[1] Gruber, J.S. & Poza, F. (1995). Voicegram identification evidence, *American Jurisprudence Trials* **54**, 1.

[2] Moenssens, A.A. (2007). Spectrographic voice recognition, *Scientific Evidence in Civil and Criminal Cases*, Foundation Press.

[3] Nolan, F. (1997). Speaker recognition and forensic phonetics, in *A Handbook of Phonetic Science*, W. Hardcastle, & J. Laver, eds, Blackwell Handbooks in Linguistics, Oxford, pp. 744–767.

[4] Rose, P. (2002). *Forensic Speaker Identification*, Taylor & Francis, London.

[5] Champod, C. & Meuwly, D. (2000). The inference of identity in forensic speaker recognition, *Speech Communication* **31**(2–3), 193–203.

[6] Meuwly, D. (2001). *Reconnaissance de Locuteurs en Sciences Forensiques: l'apport d'une Approche Automatique*, Université de Lausanne, Lausanne.

[7] Ramos-Castro, D. (2007). *Forensic Evaluation of the Evidence Using Automatic Speaker Recognition Systems*, Universidad Autonoma, Madrid.

[8] Bimbot, F., Bonastre, J.-F., Fredouille, C., Gravier, G., Magrin-Chagnolleau, I., Meignier, S., Merlin, T., Ortega-G., Pretrovska-Delacrétaz, D. & Reynolds, D.A. (2004). A tutorial on text independent speaker verification, *EURASIP Journal on Applied Signal Processing* **4**(2004), 430–451.

[9] Brümmer, N. & du Preez, J. (2006). Application-independent evaluation of speaker detection, *Computer Speech and Language* **20**(2–3), 230–275.

[10] Lewis, S.R. (1984). Philosophy of speaker identification, *Police Applications of Speech and Tape Recording Analysis – Proceeding of the Institute of Acoustics* **6**(1), 69–77.

[11] Van Leeuwen, D., Martin, A., Przybocki, M. & Bouten, J. (2006). NIST and NFI-TNO evaluations of automatic speaker recognition, *Computer Speech and Language* **20**, 128–158.

[12] Broeders, A. (2004). Forensic speech and audio analysis forensic linguistics – A review: 2001 to 2004 – in *14th Interpol Forensic Science Symposium; 2004 Lyon, France*, N. Nic Daéid, eds, Forensic Science Unit, University of Strathclyde, Glasgow, UK.

DIDIER MEUWLY

Specialist Applications: Detection of Naturally Occurring Poisons

see Poisons: Detection of Naturally Occurring Poisons

Specialist Applications: Illicit Drug Analysis

see Drug Analysis; Drug Profiling

Species Determination of Osseous Remains

The first question that may come up when dealing with much degraded skeletal remains is whether they are human or not. In fact, at times it is even necessary to verify whether the "remains" are actually composed of osseous or dental material, and not other types of organic or inorganic substance. Small pieces of burnt plastic or other mineral material may resemble bone very closely, and further microscopic or chemical investigations are needed to verify the proper origin. This happens quite frequently with burnt remains, as can be seen from the examples given in Figures 1 and 2.

Once it is determined that the fragment is truly dental or skeletal material, the next step is to verify the species of origin and, usually, whether it is human or not. Investigating authorities need to know whether the fragments belong to a human, for the obvious criminal implications. The facility with which one can determine the origin of such fragments depends on how complete the remains are [1–10]. If the remains, for example, bones or teeth, are voluminous, frequently macroscopic morphological characteristics allow for species identification. Dimensions, shape, and structure are the fundamental parameters. Entire bones, especially if one has experience and an idea of the comparative anatomy of different mammals, are relatively simple to distinguish from human bones, unless some gross malformative pathology is involved. More difficult to interpret are the bones of hands and feet, especially phalanges, quite similar among different animals (for example, man and bear). The articular surfaces of animals are frequently more profoundly beveled and very different



Figure 1 Small fragments (the longest of 8 mm) of presumed bone found in a small metal container (usually used for holy hosts of the catholic ritual) of a young man who had been arrested for attempted murder and was known to be implicated in satanic rites. Investigating authorities wanted to know whether these burnt fragments could be bone, and, if so, if they belonged to the human species

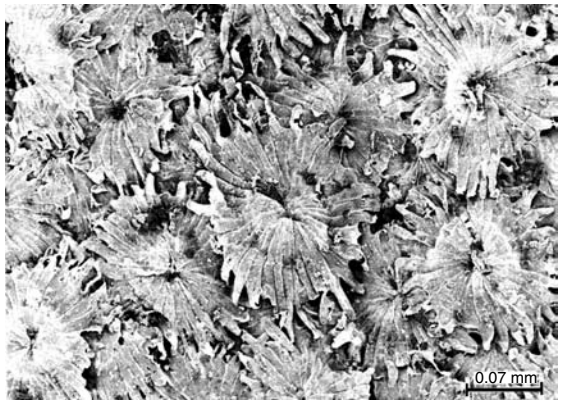


Figure 2 Scanning electron microscopy image of one of the fragments. The material was not of skeletal but of botanical origin; it was an olive gem. The pieces were part of the holy ashes stolen from a church (these are, in fact, produced from burnt olive branches) for satanic ritual purposes, but they did not involve human remains

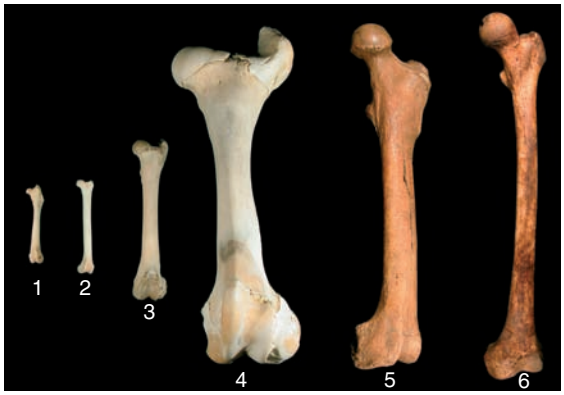


Figure 3 Left femurs of different species. (1) Rabbit (*Oryctolagus cuniculus*); (2) cat (*Felis catus*); (3) goat (*Capra hircus*); (4) cow (*Bos taurus*); (5) bear (*Ursus spelaeus*); (6) man (*Homo sapiens*). One can observe some similarities between man and bear

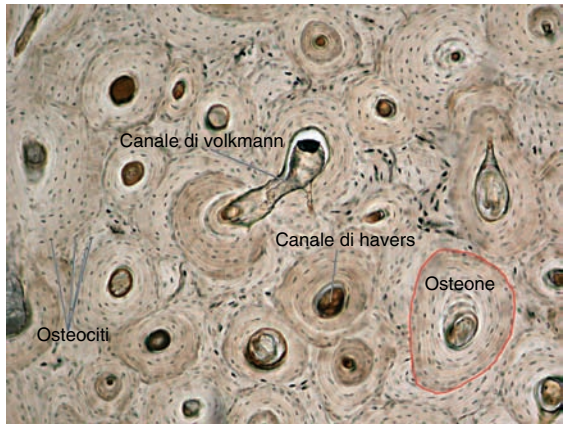


Figure 4 Microscopic structure of bone. Section of undecalcified bone (100×) showing secondary osteons, Haversian canals, and osteocytes

from human ones. Figure 3 shows the comparative morphology between different femurs of commonly found animals.

It is far less easy to determine the species of origin when the material is much more fragmented – for example, with fragments of bone diaphyses or fragments of cranial vault. In some cases, the surface of the bone may have a more coarse texture as opposed to humans or have a different kind of cross section. Cross sections of bird bones, which may apparently seem similar to infant bones, show a very thin cortex with internal lamellae of bone crossing the hollow shaft – infant bones will not have this design. Furthermore, the thickness of the bone cortex is greater in large mammals with respect to the entire diaphyseal diameter. Bones of the cranial vault may also be problematic. Here, the design of vascular sulci may be a good giveaway for human origin, but other areas of the cranium may be more problematic. Fragments of the nasal bone, for example, of small mammals may have a very similar texture to human ones. Even dental fragments can be tricky. The entire tooth is almost never a problem to an expert osteologist; however, fragments of root or crown can be difficult. In particular, fragments of crowns of goat, sheep, and bovine incisors may be similar to those of humans, as pig premolars may also be problematic. It is, therefore, sometimes impossible to perform a diagnosis based merely on macroscopic morphological features. In these cases, one may resort to two other types of methodologies: histological and biomolecular. The choice of one or the other must be performed considering the state of preservation of the fragment. DNA techniques are extremely sensitive; however, they are frequently impaired by severe degradation and calcination. In such cases, microscopic analyses may be a better bet.

Study by transmission light microscopy of thin (100 μm) undecalcified sections (decalcified sections may be used but are more complicated to prepare) prepared simply by grinding the section, which has been fixed onto a slide with resin, down to a thickness of 100–150 μm, allows one to observe microscopic characteristics of the bone, which may help in the diagnosis of species. Bone microscopic structure is made of primary and secondary osteons (Figure 4), fragments of osteons, and lamellar bone, which are in constant remodeling. The fundamental issue for species determination is the design and distribution of the osteons, shape of the single osteons, and their dimensions.

Osteons of the human species are usually distributed irregularly within the bony matrix. A classical means of exclusion is the presence of plexiform bone, i.e., with osteons set in parallel rows. Basically, human bone is nonstratified, vessels run in a longitudinal manner. Furthermore, osteon circumference can be quite important. Some animals have irregularly shaped osteons, with folds along the external circumference. The circumference of human osteons is always neat, regular, and circular (Figures 5–7).

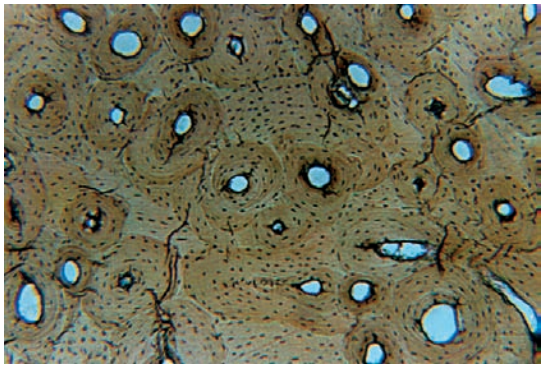


Figure 5 Thin undecalcified section of man (*Homo sapiens*) bone, which shows the typical human osteon design with a random distribution

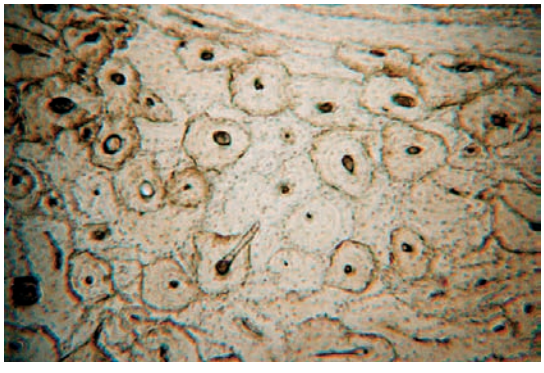


Figure 6 *Felis catus* (cat), which shows smaller Haversian canals with irregular margins of each osteon

However, other mammals have regular osteons also. In this case, it is necessary to perform a metric analysis of the secondary osteon.

Dimensions of the osteon and Haversian canals may be telling. These structures are obviously smaller in small animals and reach maximum dimensions with man. For example, the area of the cross section of a human osteon is usually around $0.039\,\mathrm{mm}^2$; the dog, on the other hand, is $0.018\,\mathrm{mm}^2$. Mammals larger than man do not have significantly larger osteons.

An algorithm including Haversian duct area and lowest and highest diameters, which can indicate human origin of bones, has been devised.

$$D = -3.99 - 0.07(C) + 0.04(D_{\mathrm{m}}) + 0.07(D_{\mathrm{M}}) \quad (1)$$

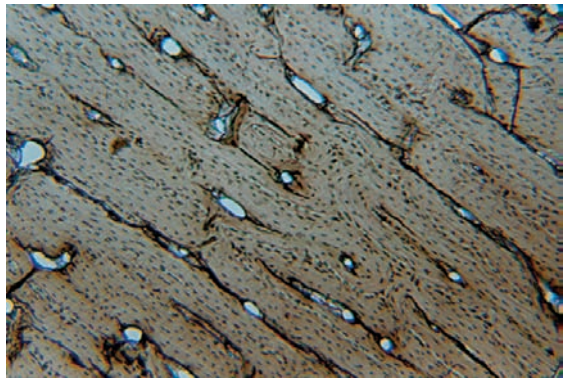


Figure 7 The presence of plexiform bone in horse (*Equus caballus*). Notice the parallel distribution of lamellar bone and osteons

where C is the Haversian canal area; D_{m} is the Haversian canal lowest diameter; D_{M} is the Haversian canal highest diameter. Where D is a positive number a human origin is established with an 80% accuracy; if negative it should be non human. Nevertheless, this method has been tested so far only on long bones; we still ignore variability of osteon characteristics in other bone districts [11].

Similar remarks may be said for the histological structure of teeth. In these cases, however, scanning electron microscopy (SEM) microscopy is the key method, since differences among certain groups of animals can be noticed, especially in enamel prism disposition. Amelogenesis, the process of production of enamel, leaves significant signs in the structure of teeth. Crystallites are organized in 5-μm fibers called *prisms*. Crystallites within and around these prisms change in their orientation. There are three different kinds of pattern they form.

- Pattern 1 has discrete prisms with complete margins and has an oval-to-circular pattern.
- Pattern 2 shows prisms with a horse-shoe section.
- Pattern 3 is similar to pattern 2 but has more of a keyhole appearance.

The latter is typical of humans, but not unique. Therefore, identification of patterns 1 and 2 may reveal a nonhuman origin, but pattern 3 is shared by humans, other primates, and even some lagomorphs (e.g., rabbits and hares). With respect to bone histology, however, dental microscopic features are not preserved in burnt remains.

A final procedure, when others fail, for species identification may be the biomolecular approach. The goal is to be able to extract species-specific biomolecules. Obviously, DNA is the primary choice nowadays, but even though the sensitivity of polymerase chain reaction (PCR) techniques and sequencing has greatly increased, the problems of extraction, contamination, and degradation must not be overlooked. DNA procedures in forensics have been treated in other sections, but a brief comment on other biomolecular techniques is certainly necessary.

Since the 1950s, the notion of species-specific proteins exists and there are innumerous techniques used to this effect. Most are based on the antibody–antigen reaction, which can be used in several kinds of immunoassays such as immunoprecipitation, agglutination, enzyme-linked immunosorbent assays (ELISA), and radio-immunoassay (RIA). The advantage of the search for epitopes or antigens of a protein nature lies in their relative abundance and resistance to degradation. This is true particularly of simple proteins such as albumin. Albumin has been studied in thousands of species in order to reconstruct genealogical trees and is extremely species-specific. Of all seroproteins, albumin seems to be the best adapted, more than collagen, IgG, and hemoglobin for species determination: it is present in higher concentrations in blood and tissues, it is a simple molecule, and is an easy immunogen.

Extraction of species-specific proteins from bone and tooth powder can be obtained by incubation in chelating substances such as ethylenediammine tetracetic acid (EDTA).

Once extracted, several techniques are available. Naturally, there are less-sensitive techniques such as RIA and immunodiffusion, or safer and more sophisticated ones such as ELISA. In practice, the antigen (in this case, a species-specific protein such as albumin) is blocked onto a plastic surface, and a monoclonal antibody to that specific antigen is then used to detect that specific protein via the antigen–antibody reaction (Figure 8). Specificity of the test is an absolute prerequisite. This is guaranteed by assays performed, for example, with the use of monoclonal antibodies, which decreases the risk of false-positive results.

The advent of DNA analysis has certainly slowed down other investigative techniques and research in this area, such as protein analysis or the study of the microscopic structure of bone, all useful tools in verifying whether the fragment is human or not. Such confidence in DNA may be dangerous. In many instances, for example, with charred or very dry bone, DNA may be very difficult to extract. Thus protein analysis or the study of the microscopic structure of the fragment may be useful, and, at times, the only applicable method. However, research in this area is relatively scarce and has stopped at the late 1990s with few exceptions. More research needs to be performed in this area. It is known that proteins are more resistant than DNA to many environmental factors and thus methods for detection of species-specific protein in bone should be encouraged, such as immuno-PCR, a technique that amplifies the sensitivity of immunological ELISA tests by tagging a DNA marker to one of the ELISA reagents and subsequently performing PCR.

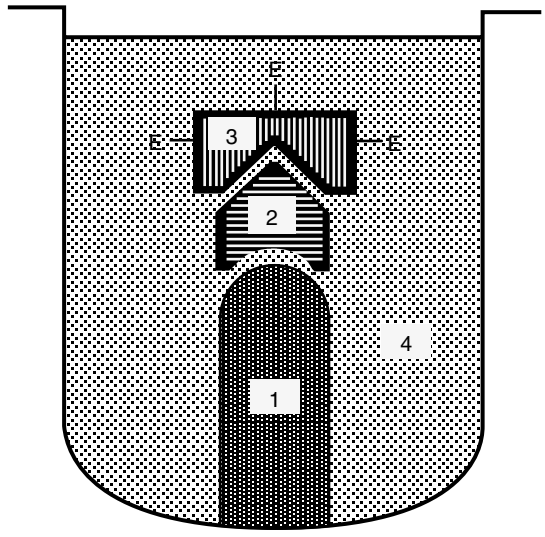


Figure 8 Schematization of an immunoassay (ELISA). In the figure, (1) is the molecule one is looking for, which should be contained in the bone extract, and which has been adsorbed – if present – onto the test plate; (2) is the antibody, which recognizes (1), if present, and "clings" onto it; (3) is a molecule, which is used to link (2) to E, which is a coloring agent that will change from a transparent color to brown or green (4) usually (depending on the enzyme used). This chain sequence that ends with a coloration of the plate proves the presence of a specific molecule in the bone extract

References

[1] Byard, R.W., James, R.A. & Zuccollo, J. (2001). Potential confusion arising from materials presenting as possible human remains, *American Journal of Forensic Medicine and Pathology* **22**(4), 391–394.

[2] Cattaneo, C., Di Martino, S., Scali, S., Craig, O.E., Grandi, M. & Sokol, R.J. (1999). Determining the human origin of fragments of burnt bone. A comparative study of histological, immunological and DNA techniques, *Forensic Science International* **102**, 182–191.

[3] Hillier, M.L. & Bell, L.S. (2007). Differentiating human bone from animal bone: a review of histological methods, *Journal of Forensic Science* **52**(2), 249–263.

[4] Hillson, S. (1986). *Teeth*, Cambridge University Press.

[5] Lowenstein, J.M., Reuther, J.D., Hood, D.G., Scheuenstuhl, G., Gerlach, S.C. & Ubelaker, D.H. (2006). Identification of animal species by protein radioimmunoassay of bone fragments and bloodstained stone tools, *Forensic Science International* **159**(2–3), 182–188.

[6] Martiniakova, M., Grosskopf, B., Omelka, R., Vondrakova, M. & Bauerova, M. (2006). Differences among species in compact bone tissue microstructure of mammalian skeleton: use of a discriminant function analysis for species identification, *Journal of Forensic Science* **51**(6), 1235–1239.

[7] Mulhern, D.M. & Ubelaker, D.H. (2001). Differences in osteon banding human and nonhuman bone, *Journal of Forensic Science* **46**(2), 220–222.

[8] Owsley, D.W., Mires, A.M. & Keith, M.S. (1985). Case involving differentiation of deer and human bone fragments, *Journal of Forensic Science* **30**(2), 572–578.

[9] Ubelaker, D.H., Ward, D.C., Braz, V.S. & Stewart, J. (2002). The use of SEM/EDS analysis to distinguish dental and osseous tissue from other materials, *Journal of Forensic Science* **47**(5), 940–943.

[10] Ubelaker, D.H., Lowenstein, J.M. & Hood, D.G. (2004). Use of solid-phase double-antibody radioimmunoassay to identify species from small skeletal fragments, *Journal of Forensic Science* **49**(5), 924–929.

[11] Cattaneo, C., Porta, D., Gibelli, D. & Gamba, C. (2008). Histological determination of the human origin of bone fragments – technical note *Journal of Forensic Sciences*, in press.

Related Articles

Anthropology

DNA: Degraded Samples

CRISTINA CATTANEO AND DAVIDE PORTA

Specific Intent *see* Behavioral Science Evidence

Sperm: Differential Extraction *see* Differential Extraction

Spree Killer *see* Homicide: Multiple (Behavior)

Stains: Biological *see* Biological Stains

Stalking

The crime of stalking contains three elements: a pattern of unwanted following or harassment; a credible threat; and the induction of fear in the victim. Without victim awareness, there is no crime of stalking. This eliminates the inappropriate application of the term *stalking* to describe violent crimes in which some form of surveillance precedes an attack, such as rape or robbery. Unwanted pursuit and sustained fear are the *sine qua non* of stalking.

Large-scale representative studies of stalking across three continents indicate that 2–13% of males and 8–32% of females will be victimized by a stalker at some point in their adult lives [1]. These figures encompass clinical, forensic, general population, and college samples. Females are the target in 80% of the cases, and half of all stalking cases involve a prior sexual intimate as the perpetrator. The average

duration of stalking is 2 years, but in one large study the modal duration was 1 month [2]. Most stalkers have known their victim in some capacity before the stalking begins. Unfortunately, stalking is a crime that has only received attention in Westernized developed countries, although other areas of the world are beginning to take notice [3].

The most validated typology of stalking is called *RECON* (relationship and context), and identifies four groups of stalkers: those who target prior intimates, prior acquaintances, public figures, and private strangers. This typology was developed utilizing a large nonrandom sample of stalkers ($N = 1005$) in North America, and showed both excellent interrater reliability and discriminate validity [2].

Stalking behavior predictably includes a number of tactics: (i) hyperintimacy – rapidly trying to accelerate intimacy, which induces fear or anxiety in the victim; (ii) proximity/surveillance – watching or following the victim; (iii) invasion – violating the privacy of the victim; (iv) proxy pursuit – using third parties to follow the victim; (v) intimidation or harassment – threatening or psychologically manipulating the victim; (vi) coercion/constraint – forcefully restricting the behavior of the victim; and (vii) aggression – being violent toward self, the victim, third parties, or property [1].

One-third of stalkers will be physically violent toward their victim during the course of their pursuit [1]. The frequency of violence substantially increases when the stalker is a prior sexual intimate of the victim, and exceeds 50%. This is a replicated finding in major urban areas on three continents [4], and suggests that sexual intimacy intensifies attraction, attachment, and emotional reactivity when a bond is threatened.

The nature of the violence in such cases is *affective*, characterized by heightened autonomic arousal, anger or fear, the presence of a threat (usually fear of abandonment), and a lack of planning. It is impulsive, reactive, and immediate. Such violence contrasts with the nature of violence when a public figure is stalked and attacked, which is typically *predatory*: planned, purposeful, and emotionless. These two modes of violence are psychobiologically distinctive in mammals, including humans [5]. Risk of homicide in all stalking cases involving prior sexual intimates is estimated to be 0.25% [6], and data suggest that stalking is a risk factor in the prediction of spousal homicide [7].

The prediction of violence research has found a number of variables that are significantly related to stalking violence, including the presence of threats, substance abuse, a prior sexually intimate relationship, personality disorder, a history of violent behavior, and the absence of psychosis [8]. The most advanced work in the prediction of stalking violence has utilized a regression tree approach, which generates subgroups that have different probability estimates of violence through the interaction effects of the predictor variables [9]. The most useful model contains nine variables: age under 30 years, education less than high school, threats toward the victim, prior intimate relationship, revenge motivation, psychotic disorder, personality disorder, substance abuse history, and criminal history [10]. The directionality of some of these variables depends on their interaction with other variables. Such sophisticated models will likely pave the way for actuarial software programs, which will fairly accurately assess the risk of violence over a particular period of time.

The paradox of stalking – an individual is pursued that is continuously rejecting – is best understood as a pathology of attachment. Attachment is a biologically rooted, species-specific behavioral system that is either secure or insecure, and in cases of stalking, the attachment pathology is insecure and often *preoccupied*. Numerous studies have tested and confirmed this hypothesis [11]. Kienlen found in a small sample of imprisoned stalkers that the majority had lost a primary caretaker in childhood and had also suffered a major loss within a year prior to the onset of stalking. These two findings suggest both predisposing and precipitating events that may contribute to the onset of stalking [12].

Adult stalkers are typically males in their fourth decade of life with prior psychiatric, drug abuse, and criminal histories. They have both a major mental illness and a personality disorder [13], often necessitating a psychiatric and psychological evaluation to discern the best approach to treatment and risk management. Psychosis occurs in a minority of stalkers, but is more likely among stalkers of strangers. Questions remain concerning the psychopathology of college students who engage in "obsessional relational intrusion," a subcriminal form of stalking [14].

Stalkers are preoccupied with thoughts of their object of pursuit. If an obsession is functionally defined as an abnormal frequency of preoccupation with an object, substantial data indicate that

obsessional thinking is the most common cognitive trait of the stalker [15]. The content of the stalker's conscious thoughts varies from case to case, but thinking is generally characterized by paradox and contradiction. Examples include the juxtaposition of statements that both idealize and devalue the victim; the wish for complete freedom for the victim alongside demands for complete control; or statements of rage commingled with yearnings for love and affection.

The preoccupations and contradictions that characterize the stalker's thinking may be unconsciously rooted in "narcissistic linking fantasies," recurrent thoughts of a special and unique relationship with the love object [16]. Such fantasies are normal in the initial stages of romance or infatuation, yet in the case of stalking, they are met with rejection when acted upon, and they usher in emotions of intense humiliation and rage that the stalker will express toward the victim. In normal men and women, romantic rejection often triggers feelings of grief, anger, and sadness, and the search for a new love object. When pathological narcissism predominates, such as one sees among stalkers, the intensity of their fury is a measure of their degree of ego deflation, and the pursuit begins.

The two most prominent emotions among stalkers are anger and jealousy [17]. Such feelings are often consciously felt and acknowledged by the stalker, but often defend against other more vulnerable feelings outside his awareness, such as shame, loneliness, isolation, and social inadequacy. Anger often fuels the pursuit of the victim, and may be further motivated by envy to damage or destroy that which cannot be possessed, or a wish to inflict pain upon the one who has inflicted pain, the primitive impulse of *lex talionis*, an eye for an eye. Anger can also repair narcissistic wounds through a fantasized sense of omnipotent control over the victim. Victim surveys have noted that the most common perceived motivation of the stalker is a desire to control [18].

Jealousy is a complex emotion and is defined interpersonally as competition for the love of the object. Pathological or morbid jealousy may be apparent in some stalkers and may reach delusional proportions, sometimes seen in cases of celebrity stalking [19]. Jealousy also may motivate behavior to dominate and isolate the victim, and has predicted stalking in one study [20].

The psychological defenses used to manage such intense emotions include minimization, denial, projection of blame, and projective identification [21]. Defenses serve to protect the stalker's inflated sense of self, but at a price. He remains vulnerable when confronted by his behavior, which paradoxically can escalate his rage. He inhabits an intrapsychic world populated by his own persecutory objects, but his sense of being persecuted can be confirmed by third parties, such as a husband, an attorney, or a police officer, who are trying to stop his incessant pursuit.

Although most stalkers are male, 15–20% are female. Female stalkers are less likely to have a history of violent criminality or substance abuse, and less likely to stalk a stranger. They are more likely than male stalkers to pursue a professional contact and to pursue a victim of the same gender. Threat and assault rates, on average, are the same as men, but women are less likely to threaten and then assault [22]. In one large study of female stalkers [23], the women were in their 30s, single, educated, and intelligent. A minority were prior sexual intimates of the victims, who were slightly older men. Unlike male stalkers, female stalkers are motivated to *establish* intimacy with the victim, whereas their male counterparts are attempting to *maintain* intimacy with their victim [22, 23].

However, the more intimate the relationship had been prior to the stalking, the greater the risk of violence. The most common documented emotion among the women was anger, particularly abandonment rage. The most common personality diagnosis was borderline personality disorder. Pathological narcissism appears to be less frequent among female stalkers, who are more intent on forming a relationship to assuage feelings of loneliness, dependency, and anger.

Victim impact in stalking cases is severe and chronic. More than one-third of stalking victims will incur a psychiatric diagnosis that will persist long after the stalking has ceased [24]. Many victims have their personal and professional lives seriously disrupted. Risk management of stalking cases is long term, difficult, and complex due to the reluctance of law enforcement to prosecute such behavior and the impressive tenacity and intelligence of many stalkers. Effective management encompasses many principles, which have been elucidated elsewhere [25, 26], but generally focuses upon four recommendations:

First, stalking victims often minimize or deny the crime being committed against them for weeks or months. It is imperative that any unwanted pursuit that induces fear should be recognized as a serious problem, if not a criminal activity, and be treated as such.

Second, stalking victims often attempt to resolve the problem alone. It is imperative that professional help be sought, including contact with both law enforcement and mental health. Stalkers are typically psychiatrically impaired, and efforts to dissuade a stalker often involve simultaneous police and psychiatric interventions.

Third, stalking victims often destroy evidence. It is imperative that all evidence of unwanted pursuit, including notes, letters, e-mails, objects, gifts, audio and video recordings, text messages, and other means of communication be kept in a safe place for eventual prosecution. Stalking laws often require the establishment of a *continuity of purpose* by the stalker, and such evidence is critical in proving a case.

And fourth, stalking victims often decide to initiate contact with the stalker to reason with him. It is imperative that all contact with the stalker be avoided, especially contact initiated by the victim. If the victim attempts to meet with the stalker, it will positively reinforce the stalker's behavior, and increase the frequency of pursuit [23]. One cannot reason with someone who is behaving unreasonably and, in many cases, dangerously.

References

[1] Spitzberg, B. (2007). The state of the art of stalking: taking stock of the emerging literature, *Aggression and Violent Behavior* **12**, 64–86.

[2] Mohandie, K., Meloy, J.R., McGowan, M. & Williams, J. (2006). The RECON typology of stalking: reliability and validity based upon a large sample of North American stalkers, *Journal of Forensic Sciences* **51**, 147–155.

[3] Kasantikul, D. (1998). Erotomania in Thai patients: a study of 20 cases, *Journal of the Medical Association of Thailand* **81**, 852–856.

[4] Meloy, J.R. (2007). Stalking: the state of the science, *Criminal Behavior and Mental Health* **17**, 1–7.

[5] Meloy, J.R. (2006). Empirical basis and forensic application of affective and predatory violence, *Australian and New Zealand Journal of Psychiatry* **40**, 539–547.

[6] Meloy, J.R. (2002). Stalking and violence, in *Stalking and Psychosexual Obsession*, J. Boon & L. Sheridan, eds, Wiley, London.

[7] McFarlane, J., Campbell, J. & Watson, K. (2002). Intimate partner stalking and femicide: urgent implications for women's safety, *Behavioral Sciences and the Law* **20**, 51–68.

[8] Rosenfeld, B. (2004). Violence risk factors in stalking and obsessional harassment: a review and preliminary meta-analysis, *Criminal Justice and Behavior* **31**, 9–36.

[9] Gardner, W., Lidz, C., Mulvey, E. & Shaw, E. (1996). A comparison of actuarial methods for identifying repetitively violent patients with mental illnesses, *Law and Human Behavior* **20**, 35–48.

[10] Rosenfeld, B. & Lewis, C. (2005). Assessing violence risk in stalking cases: a regression tree approach, *Law and Human Behavior* **29**, 343–357.

[11] Meloy, J.R. (2003). Pathologies of attachment, violence, and criminality, in *Handbook of Psychology*, *Forensic Psychology*, A. Goldstein, ed, Wiley, New York, Vol. 11, pp. 509–526.

[12] Kienlen, K., Birmingham, D., Solberg, K., O'Regan, J. & Meloy, J.R. (1997). A comparative study of psychotic and nonpsychotic stalking, *Journal of American Academy of Psychiatry and the Law* **25**, 317–334.

[13] Meloy, J.R. (1999). Stalking: an old behavior, a new crime, *Psychiatric Clinics of North America* **22**, 85–99.

[14] Cupach, W. & Spitzberg, B. (1994). *The Dark Side of Relationship Pursuit: From Attraction to Obsession and Stalking*, Lawrence Erlbaum Associates, Mahwah.

[15] Meloy, J.R., Rivers, L., Siegel, L., Gothard, S., Naimark, D. & Nicolini, R. (2000). A replication study of obsessional followers and offenders with mental disorders, *Journal of Forensic Sciences* **45**, 147–152.

[16] Meloy, J.R. (ed) (1998). *The Psychology of Stalking: Clinical and Forensic Perspectives*, Academic Press, San Diego.

[17] Mullen, P., Pathé, M. & Purcell, R. (2000). *Stalkers and Their Victims*, Cambridge University Press.

[18] Tjaden, P. & Thoennes, N. (1997). *Stalking in America: Findings from the National Violence Against Women Survey*, Center for Policy Research, Denver.

[19] Meloy, J.R., Sheridan, L. & Hoffmann, J. (eds) (2008). *Stalking, Threatening, and Attacking Public Figures*, Oxford University Press, New York.

[20] Mechanic, M., Weaver, T. & Resick, P. (2000). Intimate partner violence and stalking behavior: exploration of patterns and correlates in a sample of acutely battered women, *Violence and Victims* **15**, 55–72.

[21] Meloy, J.R. & Gothard, S. (1995). Demographic and clinical comparison of obsessional followers and offenders with mental disorders, *American Journal of Psychiatry* **152**, 258–263.

[22] Purcell, R., Pathé, M. & Mullen, P. (2001). A study of women who stalk, *American Journal of Psychiatry* **158**, 2056–2060.

[23] Meloy, J.R. & Boyd, C. (2003). Female stalkers and their victims, *Journal of American Academy of Psychiatry and the Law* **31**, 211–219.

[24] Purcell, R., Pathé, M. & Mullen, P. (2005). Association between stalking victimization and psychiatric morbidity

in a random community sample, *The British Journal of Psychiatry* **187**, 416–420.
[25] Meloy, J.R. (1997). The clinical risk management of stalking: someone is watching over me, *American Journal of Psychotherapy* **51**, 174–184.
[26] Pathé, M. (2002). *Surviving Stalking*, Cambridge University Press, London.

J. Reid Meloy

START *see* Risk Assessment: Patient and Detainee

Static-99 *see* Dangerousness: Risk of

Statistical Evidence in Court

When a criminalist testifies about an evidentiary "match" in a trial, there is usually much discussion of the criminalist's qualifications, methods, and conclusions. These are important matters. However, an issue that receives far less attention is the wording that the criminalist uses to describe that match. This is an important omission because research shows that even very subtle modifications of the forensic scientist's words can be the difference between a guilty verdict and an acquittal in a close case [1].

The first section of this article examines the different ways that a forensic science match may be characterized at trial. The emphasis is on statistical characterizations because these are more likely to convey information about the probative value of a match than more qualitative characterizations. The importance of error rates is noted and various common but erroneous match characterizations are also identified. The second section examines the issue of how jurors view and use statistical forensic science evidence. This section shows that jurors struggle to understand statistical evidence and that their judgments are heavily influenced by the particular ways in which this evidence is communicated. The third section is a brief concluding section. This section counsels forensic scientists to take care to avoid mischaracterizing the strength of match evidence and to be up front with fact finders about the assumptions that lie behind the statistics that they offer.

Statistical Characterizations of a Match

Many criminalists are accustomed to testifying that, in their expert opinion, a particular item of physical evidence came from a particular source. Such testimony, if it is believed, could be dispositive in a host of criminal cases. For example, if a forensic scientist testifies that the ransom note was written by Mr Smith, or that the bite mark on the victim's wrist was created by Mr Smith's upper incisors, factfinders will probably believe that Mr Smith was involved in the crime. However, such source identifications are generally inappropriate.[a] Instead, forensic scientists may offer qualitative or quantitative reports that describe their findings.

Qualitative Descriptions

When a forensic scientist cannot find any distinguishing characteristics between an evidentiary marking and a potential source, he may describe the correspondence as an "association", a "consistency", an "inclusion", or a "match". These words are synonyms and they do not by themselves provide any indication of evidentiary strength. When a forensic scientist does not find an association between a marking and a potential source, the relationship is commonly classified as exclusion (e.g., this hammer is excluded as a potential source of this indentation). If the forensic scientist is unable to determine whether particular characteristics are or are not shared between a marking and a potential source, he may describe the correspondence as inconclusive.

Quantitative Descriptions

Qualitative descriptions of a match may be misunderstood. For example, some fact finders may not

realize that an impression that is "consistent with" a particular shoe may also be consistent with the other, unmentioned, shoes. In order to give fact finders a better sense of the probative value of a match, witnesses should provide information about how rare the identifiable characteristics are in the potential source population.[b]

Exclusion Probability. The exclusion probability conveys more information about the probative value of a match. The exclusion probability is the proportion of potential sources (i.e., people or objects) in a reference population[c] that can be excluded as a source of the forensic science marking. Alternatively, one may think of the exclusion probability as the chance that a randomly selected element of the reference population will not match the forensic science marking. Like all the probabilities, exclusion probabilities range from 0 to 1. Larger exclusion probabilities reflect greater probative value than the smaller exclusion probabilities because large values indicate that matches are less likely to be the result of mere chance. In most forensic science work, exclusion probabilities are relatively high (98, 99.86%, etc.).

Inclusion Probabilities: 1 – Exclusion Probability = RMP = Frequency. The inclusion probability (or profile frequency) is, simply, 1 – the exclusion probability. The inclusion probability identifies the proportion of the potential source population that might be the source of a marking. In DNA typing, the inclusion probability is commonly described as the random match probability (RMP). The RMP identifies the frequency of a genetic profile in a reference population. Alternatively, the RMP is the probability that a randomly selected person in a population would happen to match the genetic profile of the forensic science evidence. Lower RMPs (i.e., low inclusion probabilities) indicate that the match is unlikely to be coincidental.

It is important to bear in mind that exclusion and inclusion probabilities do not, by themselves, capture the probative value of a match report. They do not, for example, account for the possibility of error, fraud, or misinterpretation. If any of these possibilities are substantially more likely than the chance of a coincidental match, then the chance of these events, rather than the probabilities described above, controls the probative value of a reported match. This point is elaborated in the section "Error Rates Matter".

Likelihood Ratios. Some scholars have argued that likelihood ratios are also an appropriate way of describing the value of a match for some forensic techniques, such as DNA typing [2, 3]. A likelihood ratio is the ratio of the probability that an item of evidence E would arise if hypothesis H was true to the probability that this same item of evidence would arise if alternative hypothesis $\overline{H}$ was true. In symbols,

$$LR = \frac{P(E|H)}{P(E|\overline{H})} \tag{1}$$

In theory, larger likelihood ratios values convey stronger evidentiary matches than smaller ones.

Despite their theoretical appeal, likelihood ratios often suffer from several significant practical shortcomings. First, there is no such thing as "the" likelihood ratio [4–7]. The value of a likelihood ratio depends critically on the choices one makes for describing the hypothesis H. Meester and Sjerps [7] give an example where there are two evidentiary stains (one on a pillow and the other on a sheet) and the suspect matches the pillow stain. The likelihood ratio in this example depends crucially on whether hypothesis H is "the suspect is one of the crime stain donors", or "the suspect is the donor of the pillow stain". Even in simpler one-stain situations, the value of the likelihood ratio will depend on which of the several hypotheses are used. For example, Koehler [6] showed that the hypotheses "Simpson is the source of the blood", "Simpson had contact with the crime scene", and "Simpson is guilty of the crime" would produce different likelihood ratio values.

Furthermore, most forensic likelihood ratios implicitly assume a zero laboratory error rate.[d] Even if error rates are small, they will often be larger – and perhaps substantially larger – than the RMP. In such cases, an upper bound on the probative value of the reported match is provided by the laboratory rate rather than by the RMP. Thus, computations that assume a zero laboratory error rate may yield a likelihood ratio of, say, 1 000 000 000 : 1 (depending on the frequency of the relevant profile), whereas computations on the identical evidence that incorporate a false-positive error rate[e] of, say, 0.001 will yield a likelihood ratio closer to 1000 : 1.

Error Rates Matter

All of the statistical characterizations of matching evidence discussed above are legitimate ways to describe forensic science evidence. However, as this discussion shows, each characterization is predicated on assumptions. Failure to reveal those assumptions in written reports or testimony is unscientific and potentially misleading.

In the previous section, it is noted that most forensic likelihood ratios implicitly assume a laboratory error rate of zero. Similarly, exclusion probabilities, RMPs, and their equivalent representations generally do not take account of the possibility of that a reported match may not be a true match. This is important because one cannot represent the probative value of a reported forensic science match without taking the error rate into account. Indeed, the false-positive error rate may be so important that the probative value of a match report is controlled almost entirely by this number rather than by the RMP [8, 9]. This point has not been disputed. What is more controversial, however, is how one goes about estimating the chance that a particular analyst committed a false-positive error in a particular instance.

One of the most vexing aspects of any discussion of forensic science error rates is that, although everyone knows that errors sometimes occur, no one really knows the rate of error in ordinary case work. This knowledge gap exists because there is rarely any ground truth against which to compare a criminalist's opinion. Some researchers have argued that proficiency test data can provide an indication of the rates at which errors occur. According to these data, error rates in the forensic sciences are sometimes surprisingly high [10].

However, some forensic science scholars counter that these estimated error rates should be disregarded. One popular argument is that error rates computed from historical data overstate the risk of error because they do not take into account the corrective steps that laboratories and analysts take to ensure that those errors do not happen again. A second argument is that general error rates do not take into account the procedures used in a specific case to guard against the possibility of error.

These two arguments contain elements of truth (error rates are imperfect), but they do not support the claim that data-based error rate estimates are irrelevant. Instead, these arguments suggest that historical error rates alone may not provide a reliable estimate of the chance of an error in a particular case. Regarding the first argument, if factors that produced false-positive errors were corrected, then we would expect to see dramatic declines in the rate of errors across forensic science subfields each year. However, not only is such evidence lacking but also misidentification rates in some fields do not even show a clear pattern of improvement [10]. Regarding the second argument, no one would suggest that the specific and unusual measures that a laboratories or analysts take to reduce the risk of error should be ignored. Where such data exist, they should be used to modify general error rate estimates [6]. But care must be taken to ensure that these more specific considerations are not used selectively, and that they are, in fact, distinctive and diagnostic. After all, nearly every laboratory and analyst follows a detailed protocol designed to reduce the risk of error, and the success of these protocols is largely reflected in the general error rate estimates.

Erroneous Characterizations

In recent years, forensic scientists have come under fire for exaggerating the strength of the evidence they provide at trial. In light of the underlying probative force of this type of evidence, it may seem surprising that some forensic scientists make unjustifiable claims about the strength of their evidence. However, some of the exaggerations appear to spring more from ignorance of the statistical character of forensic science evidence than from an intentional desire to mislead jurors. Below, a series of faulty characterizations of forensic science evidence that tend to exaggerate the strength of reported matches is identified.

Individualization Exaggerations. Forensic scientists often claim object individualization as both a goal and an achievement [11]. That is, they commonly claim to have identified the one and only possible source of a marking to the exclusion of all other possible sources in the world. However, there is little scientific basis for such claims, their frequency in textbooks and courtrooms notwithstanding. The fact is that criminalists cannot know that a marking is unique to a particular source, particularly when they have only examined a small fraction of possible sources. As a rule, source certainty statements are inappropriate and only obfuscate the probabilistic nature of the forensic science enterprise.

Prosecutor's Fallacy. The so-called prosecutor's fallacy [12] is the most well-known erroneous match characterization. The prosecutor's fallacy is committed when one equates the RMP with the probability that a matching defendant is not guilty. It is a fallacy to characterize the match frequency this way because probabilities of guilt and innocence require consideration of *all* evidence, including prior odds of guilt or innocence.[f] If RMPs alone could reveal probabilities of innocence or guilt, there would be no need to consider alibis, eyewitness testimony, or evidence implicating other suspects.

Source Probability Error. A more common error, albeit one that is rarely acknowledged, is the source probability error [13]. This error occurs when one equates the RMP with the probability that the defendant is not the source of the forensic evidence. Just as one cannot identify the probability of guilt or innocence without detailed consideration of the nonforensic evidence, one cannot identify the probability that a potential person or object is the source of a stain or marking based on forensic science evidence alone. One also needs to know, for example, how many others could have been the source of the forensic evidence *prior to* the discovery that the suspect matches. This fact may come as a surprise to forensic scientists who routinely provide source probability estimates in court and are aware of others who do the same. However, the pervasiveness of this error does not make it any less fallacious than its more easily understood cousin, the prosecutor's fallacy.

P (Another Match) Error. The P (another match) error also finds its way into the courtroom with a disturbing degree of regularity (see [13]). This error occurs when one equates the RMP with the probability that there exists another person who matches the defendant's DNA profile. But a moment's reflection reveals why these two probabilities are not identical. If the RMP = 1 in 1 000 000, then hundreds of other people in a country the size of the United States (300 000 000) would match. Consequently, one could not obtain an RMP of 1 in 1 000 000 and conclude that the chances are just 1 in 1 000 000 that there exists another person who would match. In contrary, it is a near certainty that *many* other people match this profile.

The Numerical Conversion Error. Finally, the numerical conversion error occurs when experts describe the significance of the RMP in terms of the number of people who would have to be tested before one should expect another match to occur [13]. This computation is straightforward, though it is not, as some forensic scientists have testified, X in an RMP of one in X. A conclusion that 1 in every 1 000 000 randomly selected people match is *not* equivalent to a conclusion that 1 000 000 people would need to be tested before another match might be expected. The number is smaller; in this case, it is 693 147.[g]

Most of the errors identified above tend to exaggerate rather than minimize the strength of statistical forensic science evidence. Such exaggeration, though, is hardly necessary because statistical forensic science evidence can be quite compelling when presented accurately.

How Do Jurors Think About and Use Match Statistics?

Even when a forensic scientist studiously avoids exaggerating or otherwise mischaracterizing the strength of an evidentiary match, there is still a danger that jurors will misunderstand the technical evidence that they hear. Some misunderstandings arise because jurors have preconceived ideas about the strength of match evidence or because they oversimplify the evidence in an effort to understand it. Strictly speaking, such misunderstandings are not the fault of the forensic scientist. However, forensic scientists should know something about the psychology of how people process statistical evidence. Such knowledge could help forensic scientists prepare written reports and testify in ways that are geared to help jurors appreciate the probative value of the evidence.

The Mock Jury Study

How can we know how much jurors value forensic science evidence? After all, cases that involve scientific evidence are usually complex affairs, and it may be impossible to determine posttrial exactly how jurors valued any particular item of evidence. Verdicts provide only a very gross measure of how impressed jurors were with the totality of evidence in a case. Posttrial interviews with individual jurors may provide additional insight, although research shows that people have limited insight into factors that affect their judgments and choices [14].

A better method for assessing how jurors use statistical evidence is the mock jury study. Mock jury studies are controlled behavioral experiments that test causal relationships between predictor variables (e.g., different types of jury instructions) and outcome variables (e.g., jurors' verdicts). In the typical experiment, jury-eligible subjects assume the role of jurors and watch a videotape (or read a description) of a trial or part of a trial. Jurors are randomly assigned to groups that differ only in terms of the predictor variables of interest. For example, one version of the videotaped trial might include testimony from an expert who commits The prosecutor's fallacy, whereas another version might not. A well-run mock jury study enables researchers to isolate the direction and magnitude of various influences on jurors' judgments and verdicts.

The obvious and central disadvantage of the mock jury study is its artificiality. Mock jury studies often use students rather than real jurors, use written case summaries rather than videotaped or live testimony, or collect verdicts from individuals rather than from juries that have deliberated to unanimity. It is only natural to question the usefulness of data derived from such research.

Nevertheless, there is good reason to believe that mock jury studies provide a window into the minds of actual jurors. Studies that have compared student jurors to actual jurors generally report similar patterns of decisions Casper and Benedict [15, p. 78] found no difference in the verdict patterns displayed by college students and jurors selected from the Cook County jury roll in Illinois Bornstein [16] reviewed dozens of mock jury studies and concluded that "[There is] strong evidence that factors at trial affect students and non-students in the same way" [16, p. 80]. Bornstein [16] also reported that it makes little difference whether mock jury studies employ written summaries, transcripts, audiotape, or videotape: "[R]esearch on the trial medium tends not to find many differences." (p. 82, 84). Of course, even if we accept that different types of mock jury studies yield similar results, it is possible that some of the conclusions reached in those studies may not hold when the stakes are real. But as Hastie [17] has observed, "there are so many converging results concerning juror behavior from laboratories and from surveys of actual jurors that most of the conclusions from the studies reported are surely accurate descriptions of juror behavior in real trials." (p. 28). In short, although mock jury studies are imperfect representations of actual trials, they may provide insight into how jurors think and respond at trial.

The Results of Behavioral Experiments

Defense Attorney Fallacy. Thompson and Schumann [12] coined the phrase "the defense attorney fallacy" to describe the erroneous belief that nonunique match evidence is not probative. For example, suppose that a partial DNA profile from an evidentiary sample occurs with a frequency of 1 in 1000 and that the suspect population includes 30 000 adult males in a city. In this scenario, the defense attorney fallacy would occur if one argued that a DNA match on a particular suspect is worthless because approximately 30 adult males in the city would match. This argument is fallacious because it fails to distinguish between probative evidence and dispositive evidence. Non-unique evidence may be highly probative if the size of the inclusion group is small relative to that of the exclusion group. In our example, the partial DNA match is highly probative because 99.9% of people who were not the source would have been excluded, yet the defendant happens to be a member of the very small group of non-excluded people. Holding aside error rate considerations, the discovery of this match means that a juror should believe that the defendant is now about 1000 times more likely to be the source of the DNA evidence than he was prior to the discovery of the match.

Several studies have shown that significant minorities of jurors are susceptible to this fallacy [12, 18], particularly when it is explicitly offered as an argument by a defense attorney Hans *et al.* [19] found that fully 40% of mock jurors agreed with the statement "The mtDNA evidence in this case is completely irrelevant because a substantial number of other people could also be the source of the hairs." (p. 17).

Conditional Probability Confusion. Mock jury studies have also demonstrated that jurors are prone to inverting conditional probabilities [20]. These inversion errors can lead to source probability errors and prosecutor's fallacies. In Bayesian terms, inversion errors sometimes lead jurors to confuse likelihood ratio components with posterior odds components Koehler [5] showed that mock jurors do not distinguish between likelihood ratio statements (e.g., "It is approximately 1000 times more likely we would

see this DNA match if the defendant is the source of the semen than if the defendant is not the source of the semen".) and posterior odds ratio statements (e.g., "Given that we see this DNA match, it is approximately 1000 times more likely that the defendant is the source of the semen than that he is not the source of the semen".). This finding supports other research that shows that most people verbally confuse likelihood ratios with posteriors in nonlegal contexts [21], and that expert witnesses frequently describe likelihood ratios as posterior probabilities [13].

The confusion between likelihood ratios and posteriors in the forensic context is important because it creates the mistaken impression that match evidence alone can identify the probability that a suspect committed a crime or is the source of trace evidence. But in light of the linguistically subtle differences between likelihoods and posteriors, it is hard to prevent this particular error. In June, 1995, Court TV commentator Dan Abrams noted that likelihood ratios were discussed so frequently during various legal motions and arguments in the O. J. Simpson trial that it was important to explain to the viewing public what exactly a likelihood ratio is. Abrams then committed an inversion fallacy when he explained that a likelihood ratio refers to how likely one hypothesis is true *versus* how likely a competing hypothesis is true.

Failure to Appreciate the Role of Error Rates. Jurors have a hard time understanding how to combine error rates with RMPs. This is a serious problem because the probative value of a reported match is determined almost exclusively by the false-positive error rate in cases where the RMP is very small Koehler *et al.* [8] were the first to examine how mock jurors intuitively combined error rates and RMPs. They found that jurors simply ignored error rates (e.g., 1 in 1000 and 2 in 100) when the RMP was very small (e.g., 1 in 1 000 000 000) Schklar and Diamond [22] (hypothesis 2) found that jurors who received separate estimates for the RMP and error rate were significantly more likely to return a guilty verdict than were jurors who received a single statistic that properly aggregated these two statistics for them. Instruction on how to combine the RMP and error rate had little impact [18, 22]. This result suggests that it is no simple matter to get jurors to appreciate the role that error rates play in the interpretation of match reports.

Reasoning by Exemplars. People tend to overweight low-probability events [23]. This phenomenon partially explains why people buy insurance and lottery tickets even though both purchases have negative expected values. RMPs are almost always low-probability events and sometimes they are extremely low-probability events. One might therefore expect that jurors will tend to assign too much weight to the remote possibility of a coincidental match and thereby underweight a reported DNA match. Research generally supports this position. Most studies that have looked at this issue find that jurors attach less weight to statistical match evidence than the evidence would seem to deserve (for reviews, see [19, 20]). However, caution is needed here. Although people tend to overweight low-probability events, *extremely* low-probability events may be *underweighted* if people view those probabilities as negligible.

Research shows that the weight that jurors assign to very low RMPs depends critically on how those probabilities are presented. When RMPs are presented in ways that make it easy for jurors to imagine others who might match the profile, then a reported DNA match seems less impressive and jurors give less weight to the match evidence. But when RMPs are presented in ways that make it hard to think about others who might match, jurors give significantly more weight to the match evidence.

In one study [1], jurors who were told "The probability that Mr. Clinton would match the semen stain if he were not its source is 0.1%", assigned higher probabilities to the hypothesis that Mr Clinton was the source of the stain than did jurors who were told "1 in 1000 people in Washington, DC, who are not the source would also match the semen stain". Although the RMP is identical in the two cases, the latter description makes it easier to imagine examples of others who would also match. These imagined examples weaken the perceived strength of the match evidence. In another study [24], jurors who heard a DNA RMP described as "0.1 in 100" were more likely to believe that the person who matched the evidence was its source than were jurors who heard this same RMP described as "2 in 2000". Whereas the fractional numerator in the former format conveys a sense that others are unlikely to match, the latter format encourages people to think about a second matchee.

How Should a Forensic Scientist Communicate Match Statistics?

This article examined the various ways in which forensic scientists characterize a forensic science match and how jurors are likely to interpret match statistics. Whereas some ways of characterizing a match are more informative than the others from a strictly mathematical standpoint, it is unclear whether untrained lay jurors understand the meaning of those characterizations. So what should a forensic scientist do?

He should begin by educating himself about the foundational principles of probability and statistics. Science, after all, is a probabilistic enterprise and scientists speak the language of statistics. The forensic scientist should pay particular attention to Bayes theorem (odds form) in his studies. Once he understands the difference between prior probabilities, likelihood ratios, and posterior odds ratios, the forensic scientist will be less likely to commit many of thestatistical misstatements discussed in the first section. He will understand why he should abandon traditional source identity claims in favor of carefully worded probabilistic claims. And he will be able to explain to jurors why his science does not enable him to identify the probability that this bullet came from this gun or that this toolmark was produced by this hammer.[h]

He will also be in a better position to correct and explain the misstatements implicit in the questions that attorneys and judges ask of him. For example, when an attorney attempts to confirm a 1 in 1 000 000 RMP by asking "In other words, you're saying that there is only one chance in a million that the blood could be of anyone other than the defendant's, correct?" the forensic scientist should not agree. Instead, he should explain why this statement mischaracterizes the RMP. In the context of rapid-fire courtroom exchanges, this may be a challenge. But the forensic scientist, like any scientist, is duty-bound to correct misstatements of his conclusions and to educate those who use his data to make important decisions.

In addition to avoiding and correcting misstatements, the forensic scientist should disclose all relevant assumptions. For example, he should note that the RMP requires assumption of independence across characteristics, and that most of the forensic probabilities implicitly assume a zero error rate.

The forensic scientist should also be sensitive to the underlying psychology associated with how people think about statistical evidence, in general, and very small probabilities in particular. After all, the primary purpose of forensic science testimony at trial is to convey probative information to the legal fact finder. If that information is conveyed in ways that make it hard for the trier of fact to understand, then adjustments are needed. One possibility is to place some of the burden on jurors by suggesting that they receive training to help them understand scientific and statistical evidence [25]. However, the practical realities are such that, for now, those who offer technical testimony must find simple and clear ways to explain their evidence, erring on the side of caution and conservatism.

An undesirable side effect of scientific conservatism is that recipients of the scientific message may give insufficient weight to the message. Scientific evidence that is accompanied by cautions about what the evidence does not represent, and cautions about the assumptions and limits of the accompanying statistics may discourage some fact finders from attaching sufficient weight to the evidence. In the case of forensic match evidence, jurors may be disappointed to learn, for example, that a handwriting match does not necessarily mean that the matchee authored the questioned document. When jurors realize that the match merely places the matchee in a small group of people who might have authored the document, they may see the evidence as weak or irrelevant because the forensic scientist was not able to individualize the handwriting. Indeed, as noted earlier, 40% of mock jurors in one recent study said they believed that statistical mtDNA evidence was "completely irrelevant" because more than one person could match [19]. Such results indicate that jurors need substantial help understanding how to value forensic science evidence. Assuming that jurors remain untrained, the forensic scientist will need to provide that help. On the one hand, he must identify critical assumptions and avoid overstating his conclusions. On the other hand, he must teach jurors that a reported match can be powerful evidence of identity even when uniqueness and error-free methods cannot be assumed. In the end, the forensic scientist should be truthful, candid, clear, helpful and, above all else, scientific.

End Notes

[a.] A source identification may be appropriate in rare cases in which all the potential sources of a marking are available for examination and only one of those potential sources matches the target marking.

[b.] The phrase *potential source population* here in much the same way as others uses "suspect population" in cases that include human genetic material. The former phrase applies to a broader array of forensic cases (e.g., cases involving toolmarks) and does not attempt to make the forensic scientist the arbiter of which possible human sources should and should not be included in the groups of potential perpetrators. After all, some people may be members of the source population (e.g., they might be the source of the blood), but they might not be members of the suspect population (e.g., if they were in prison at the time the crime occurred).

[c.] Ideally, the reference population should be the potential source population. In practice, data for this population may be hard to construct. If so, then convenient alternatives such as the general population or an ethnic grouping within the general population are commonly used. The extent to which these alternative reference groups are adequate varies from case to case.

[d.] If the evidence E is described as "a true match", then there is no place for considerations of error in the likelihood ratio because E equates the *report* of a match with a true match. For this reason, it is more appropriate to describe E as "a reported match".

[e.] Discussions of "errors" in forensic science sometimes get bogged down by concerns about what exactly is meant by an error. For purposes of this paper, the words "error" and "error rate" are used as synonyms for false-positive error and false-positive error rate, respectively. A false-positive error arises when an analyst affirmatively claims that two markings match when, in fact, they do not share a common source.

[f.] In Bayesian terms,

$$\frac{P(\textit{suspect is guilty} \mid \textit{evidence})}{P(\textit{suspect is not guilty} \mid \textit{evidence})} = \frac{P(\textit{suspect is guilty})}{P(\textit{suspect is not guilty})} \times \frac{P(\textit{evidence} \mid \textit{suspect is guilty})}{P(\textit{evidence} \mid \textit{suspect is not guilty})}$$

Even if we simplify this formula by assuming that $P(\textit{evidence} \mid \textit{suspect is not guilty}) = \text{RMP}$, we may not conclude that $\text{RMP} = P(\textit{suspect is not guilty} \mid \textit{evidence})$ The latter conditional probability also depends on the prior probability that the suspect is guilty (i.e., the probability that the suspect is guilty based on nonforensic considerations).

[g.] In order to estimate the number of people who would need to be tested before there would be more than a 50% chance to find a match on a profile common to one in X people, one must compute the smallest N that makes this equation true: $(1 - 1/X)N < 0.50$. For computational simplicity, one may solve for N, where $N = [\ln(.50)]/[\ln(1 - 1/X)]$.

[h.] These are posterior probabilities and therefore inaccessible to the forensic scientist absent an estimate of the prior probabilities. Estimates of prior probabilities are matters for the legal fact finder rather than the expert witness [26].

References

[1] Koehler, J.J. (2001). When are people persuaded by DNA match statistics? *Law and Human Behavior* **25**, 493–513.

[2] National Research Council, Committee on DNA Forensic Science (1996). *The Evaluation of Forensic DNA Evidence*, National Academy Press, Washington, DC.

[3] Evett, I.W. & Weir, B.S. (1998). *Interpreting DNA Evidence*, Sinauer, Sunderland.

[4] Dawid, A.P. (2004). Which likelihood ratio? (Comment on: why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence), *Law, Probability and Risk* **3**, 65–71.

[5] Koehler, J.J. (1996). On conveying the probative value of DNA evidence: frequencies, likelihood ratios and error rates, *University of Colorado Law Review* **67**, 859–886.

[6] Koehler, J.J. (1997). Why DNA likelihood ratios should account for error (even when a National Research Council report says they should not), *Jurimetrics Journal* **37**, 425–437.

[7] Meester, R. & Sjerps, M. (2004). Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence, *Law, Probability and Risk* **3**, 51–62.

[8] Koehler, J.J., Chia, A. & Lindsey, J.S. (1995). The random match probability (RMP) in DNA evidence: irrelevant and prejudicial? *Jurimetrics Journal* **35**, 201–219.

[9] Lempert, R. (1991). Some caveats concerning DNA as criminal identification evidence: with thanks to the Reverend Bayes, *Cardozo Law Review* **13**, 303–341.

[10] Saks, M.J. & Koehler, J.J. (2005). The coming paradigm shift in forensic identification science, *Science* **309**, 892–895.

[11] Osterburg, J.W. (1969). The evaluation of physical evidence in criminalistics: subjective or objective process? *Journal of Criminal Law and Criminology* **60**, 97–101.
[12] Thompson, W.C. & Schumann, E.L. (1987). Interpretation of statistical evidence in criminal trials: the Prosecutor's Fallacy and the Defense Attorney's Fallacy, *Law and Human Behavior* **11**, 167–187.
[13] Koehler, J.J. (1993). Error and exaggeration in the presentation of DNA evidence, *Jurimetrics Journal* **34**, 21–39.
[14] Nisbett, R. & Wilson, T. (1977). Telling more than we know: verbal reports on mental processes, *Psychological Review* **84**, 231–257.
[15] Casper, J.D. & Benedict, K.M. (1994). The influence of outcome information and attitudes on juror decision making in search and seizure cases, in *Inside the Juror: The Psychology of Juror Decision Making*, R. Hastie, ed, Cambridge University Press, Cambridge, pp. 65–83.
[16] Bornstein, B.H. (1999). The ecological validity of jury simulations: is the jury still out? *Law and Human Behavior* **23**, 75–91.
[17] Hastie, R. (1994). Introduction, in *Inside The Juror: The Psychology of Juror Decision Making*, R. Hastie, ed, Cambridge University Press, Cambridge, pp. 3–41.
[18] Nance, D.A. & Morris, S.B. (2005). Juror understanding of DNA evidence: an empirical assessment presentation formats for trace evidence with a relatively small random match probability, *Journal of Legal Studies* **34**, 395–444.
[19] Kaye, D.H., Hans, V.P., Dann, B.M., Farley, E. & Albertson, S. (2007). Statistics in the jury box: Haw jurors respond to mitochondrial DNA match probabilities, *Journal of Empirical Legal Studies* **4**, 797–834.
[20] Kaye, D.H. & Koehler, J.J. (1991). Can jurors understand probabilistic evidence? *Journal of the Royal Statistical Society, Series A* **154**(part 1), 75–81.
[21] Wolfe, C.R. (1995). Information seeking on Bayesian conditional probability problems: a fuzzy-trace theory account, *Journal of Behavioral Decision Making* **8**, 85–108.
[22] Schklar, J. & Diamond, S. (1999). Juror reactions to DNA evidence: errors and expectancies, *Law and Human Behavior* **23**, 159–184.
[23] Kahneman, D. & Tversky, A. (1982). The psychology of preferences, *Scientific American* **146**, 160–173.
[24] Koehler, J.J. & Macchi, L. (2004). Thinking about low-probability events: an exemplar cuing theory, *Psychological Science* **15**, 540–546.
[25] Koehler, J.J. (2006). Train our jurors, in *Heuristics and The Law*, Dahlem Workshop Report 94, G. Gigerenzer & C. Engel, eds, The MIT Press, Cambridge.
[26] Wagenaar, W.A. (1988). The proper seat: a Bayesian discussion of the position of expert witnesses, *Law and Human Behavior* **12**, 499–510.

JONATHAN J. KOEHLER

Statistical Interpretation: Low Template DNA *see* Interpretation: Low Template DNA

Statistical Interpretation: Mixtures *see* Mixture Interpretation: DNA

Stature Determination from Skeletal Remains *see* Anthropology: Ancestry and Stature Determination

Stockholm Syndrome

Incidents of Stockholm Syndrome in the Media

Stockholm syndrome, also referred to as the *phenomenon of survival identification* or *hostage identification syndrome*, is a concept that has been extensively covered by the media and seen in several highly publicized cases of hostage rescue [1]. The term *Stockholm syndrome* was first used to describe the experiences of victims of a bank robbery that occurred at the Sveriges Kreditbank, bank in Stockholm, Sweden on August 23, 1973. Four bank employees were held hostage by bank robbers for 131 h. Following their release, the hostages relayed to reporters feelings that seemed opposite of what was expected. They described fearing the police more than the bank robbers. In fact, they felt grateful for the generosity of the hostage takers in giving them back

their lives [2]. The psychological term, *Stockholm syndrome*, came about as a way to explain the positive feelings the hostages had for their captors and the way they identified with their political goals, if not the means they chose to achieve them.

The syndrome gained additional media attention following Patty Hearst's kidnapping and subsequent bank robbery and trial. Patricia Hearst, granddaughter of media mogul, William Randolph Hearst, was a 19-year-old undergraduate student in California when she was kidnapped by the Symbionese Liberation Army (SLA) on February 4, 1974. Her kidnapping was described by the SLA in a communiqué as an "arrest and protective custody; and if resistance execution" of the hostage to draw attention to their cause of ending racial and economic discrimination and opposing capitalism [3]. Fifty-nine days after her capture Patty Hearst sent a tape to a radio station indicating that she had chosen to join the SLA and "fight for my freedom and the freedom of all oppressed people" [4]. She went on to join the organization and participate in a bank robbery and other illegal activities only to be arrested on September 18, 1975. Her actions were defended by many as the result of the stress of her being held against her will and fearing for her life. Nonetheless, she was convicted.

Role of the Hostage-taker's Intentions

There are different scenarios in which hostage taking can occur. Whether a hostage is taken intentionally, or as part of another crime "gone wrong", plays a role in the way a victim identifies with their captor. Intentional hostage taking is frequently undertaken as a means of drawing attention to a political or social cause. It is often a tactic employed by individuals who believe they lack power and have no other way to draw attention or get their demands met. In these situations it is possible for those taken captive to begin to understand the desperation their captors have been feeling, and to mistake their own feelings of fear and helplessness with the experiences of the cause for which the hostage-taker is fighting.

Other times, hostage taking is not planned but an accident of a primary crime such as a robbery which has gone wrong. In this scenario, when a criminal feels trapped by law enforcement, he may react by taking hostages without fully thinking through the consequences of this action. The situation then worsens and may involve a dangerous stand-off that compounds the original offense with a kidnapping charge. The criminal finds himself in an unexpected situation with his own freedom on the line much like that of the hostages he has taken. The hostage-taker may blame the police for forcing him into taking hostages and make statements alleging that he would allow the hostages to go free if only the police would give in to his demands. In this way, the hostages and hostage-taker may feel they are together in a situation not of their own making and begin to form a bond and share negative feelings toward the police.

Theories Regarding the Formation of Stockholm Syndrome

The term *Stockholm syndrome* has evolved into a catch-all to describe the different psychological explanations for why some captives feel no ill will toward their captors but form a bond with the person who took away their freedom. This unconscious reaction to the stress of being held hostage is a way for the individual to both deal with the fear of the situation as well as try to adapt in order to stay alive [5]. The Stockholm syndrome has been described as a combination of feelings on the part of the hostage which includes positive feelings toward the captor and negative feelings toward authority as well as reciprocal positive feelings toward the hostage on the part of the captor [6]. In a study of the effects of kidnapping in Italy, 50% of the victims described some form of positive bond with at least one of the kidnappers [7]. The following is a discussion of some of the more prominent psychological explanations for the Stockholm syndrome.

Denial

Most explanations for the formation of the Stockholm syndrome are based on the idea that the victim uses defense mechanisms to deal with the terror of the situation [8]. The concept of defense mechanism is rooted in Sigmund Freud's theory that the ego, or the part of a person's psyche that attempts to maintain balance between one's instinctual drives and external reality, attempts to control anxiety by employing coping strategies, defense mechanisms, to reduce it [9]. In hostage situations, the hostage

quickly realizes that they are completely at the mercy of another person. They are psychologically reduced to an early stage of development where they are reliant on another person for their very life, much like an infant.

This promotes the use of the more basic coping strategies. One of the most basic defense mechanisms is that of denial in which a person deals with an anxiety-provoking situation by denying its reality. Kidnapping victims may try to convince themselves that the situation is not real and use distracting techniques such as busy work to avoid being confronted with the reality of the situation [2]. It is only after these strategies fail, and the victim must accept their predicament, that the Stockholm Syndrome begins.

Identification with the Aggressor

One of the most commonly cited psychological reasons for the Stockholm syndrome is a psychological phenomenon first identified by Sigmund Freud's daughter, Anna Freud, who expanded on her concepts of defense mechanisms and introduced the concept of "identification with the aggressor". She described this to occur when an individual experiences fear of an aggressive person and subconsciously deals with this uncomfortable emotion by putting himself in the same category, and thereby identifying the aggressor [10]. This theory helps to explain why captives who have no reason to feel anything but anger toward their captors begin to feel an unexpected allegiance to them.

Learned Helplessness

Hostage-takers may alternate between violence toward the hostage and moments of unexpected sympathy and kindness. This inconsistent pattern of interaction could result in the hostages developing learned helplessness. When a person feels they have no ability to control the outcome of a situation (i.e., become helpless), they begin to give up hope [11]. As the pattern of their captors' responses varies without an obvious reason, the victims realize they have little power to affect the situation and stop trying.

They become increasingly dependent on the person they perceive to have control over the situation and feel grateful to the hostage-taker for providing for their basic needs such as food and shelter. Even the smallest act of kindness coming from someone who has previously been indifferent or cruel carries great weight. Patty Hearst herself described "once they don't kill you, you start to think they're nice. They get nicer every day that they don't kill you" [12].

Factors Contributing to the Development of Stockholm Syndrome

Different variables on the part of both hostage and hostage-taker play a role in the formation of the Stockholm syndrome. According to Thomas Strentz [2] one of the factors which promotes the formation of the Stockholm syndrome is the amount of exposure the hostages and their keepers have to one another. The more time a hostage and hostage-taker are together the greater the likelihood that they will begin to see the other as a person, rather than political tools or faceless captors. This effect, however, is negated if the victim is being abused or if the victim and hostage-taker do not share a common language. Other factors which increase the chance that the Stockholm syndrome will develop include the degree to which the captive depends on his keeper for survival as well as the intensity and duration of the experience [5].

Factors which make the formation of Stockholm syndrome less likely include physical abuse by the captors, intentional emotional distancing on the part of the captors [5], and a strong sense of self and purpose on the part of the victim. Giebels *et al.* interviewed 11 kidnapping victims and found that all but 3 developed positive feelings toward their captors. The three who did not develop positive feelings were all physically abused [13]. Turner [14] added that cultural value systems, pre-existing stereotypes, and the intellectual sophistication of the individual also play a role in Stockholm syndrome's development.

Stockholm Syndrome in Cases of Spousal Abuse

While Stockholm syndrome began as a way to explain the reactions of hostages to their captors, it has also been applied to other situations in which one person exerts power over the other and controls the weaker person's freedom. Domestic abuse is such a situation. Some victims of domestic abuse choose not only to stay with their abusive husbands, but also idealize them and blame themselves for the abuse [15]. In such a situation the victims "deny both their terror and the captor's abuse and *bond* to the kind side of their captor" [16]. This seemingly

contradictory reaction is theorized to share many of the same psychological underpinnings as the Stockholm syndrome including identification with the aggressor and denial. This prolonged psychologically traumatic experience can result in a woman seeing her abusive husband as all powerful and that she has no hope for escape from him (*see* **Battered Spouse Syndrome**).

The Stockholm Syndrome in Legal Proceedings

The combination of positive feelings toward one's captors and a mistrust of authorities that is found with the Stockholm syndrome can have an effect on law enforcement efforts and legal prosecution of hostage-takers. The Stockholm syndrome can drive captives to interfere with police negotiations to end the hostage situation by not giving accurate information or refusing to cooperate [17]. Another consequence of the Stockholm syndrome is the reluctance that a victim may feel to testify against their captors [18]. The feelings of gratitude the victims have for being allowed to live and returned to freedom can lead them to believe they should protect the hostage-taker in return.

The Stockholm syndrome will, therefore, likely continue to be a topic of fascination and media attention. Although there is no single universally accepted explanation for its formation, the primary theories involve the use of defense mechanisms as ways to shield the individual from facing the full force of the trauma. As more victims are willing to discuss their experiences, the understanding of the complex psychological reaction and the role it plays in hostage negotiations will grow.

References

[1] Turco, R.M. (1987). Psychiatric contributions to the understanding of international terrorism, *International Journal of Offender Therapy and Comparative Criminology* **31**(2), 153–161.

[2] Strentz, T. (1980). The Stockholm syndrome: law enforcement policy and ego defenses of the hostage, *Annals of the New York Academy of Sciences* **347**, 137–150.

[3] Communique #3 of the Symbionese Liberation Army Western Regional Adult Unit dated February 4, 1974.

[4] *Transcript of Patty Hearst Tape to the Media, Day 59*, (1974) at http://www.law.umkc.edu/faculty/projects/ftrials/hearst/hearstaudiotranscripts.html, (accessed on, 2008).

[5] Strentz, T. (1982). The Stockholm syndrome: law enforcement policy and hostage behavior, in *Victims of Terrorism*, F.M. Ochberg & D.A. Soskis, eds, Westview Press, Colarado, pp. 149–163.

[6] Harkis, B. (1986). The psychopathology of the hostage experience – a review, *Medical Science and Law* **26**(1), 48–52.

[7] Favaro, A., Degortes, D., Colombo, G. & Santonastaso, P. (2000). The effects of trauma among kidnap victims in Sardinia, Italy, *Psychological Medicine* **30**, 975–980.

[8] Tinklenberg, J. (1982). Coping with terrorist victimization, in *Victims of Terrorism*, F.M. Ochberg & D.A. Soskis, eds, Westview Press, Colarado, pp. 59–72.

[9] Shill, M.A. (2004). Signal anxiety, defense, and the pleasure principle, *Psychoanalytic Psychology* **21**(1), 116–133.

[10] Freud, A. (1966). *The Writings of Anna Freud II. The Ego and the Mechanism of Defense*, Revised edition, International Universities Press, New York.

[11] Abramson, L.Y., Seligman, M.E.P. & Teasdale, J.D. (1978). Learned helplessness in humans: critique and reformulation, *Journal of Abnormal Psychology* **87**(1), 49–74.

[12] *Patricia Hearst interview with Larry King on Larry King Live: Patricia Hearst Discusses Her Presidential Pardon*, Aired January 31, 2001 – 9:00. p.m. ET. http://transcripts.cnn.com/TRANSCRIPTS/0101/31/lkl.00.html.

[13] Giebels, E., Neolanders, S. & Varvaeke, G. (2005). The hostage experience: implications for negotiation strategies, *Clinical Psychology and Psychotherapy* **12**, 241–253.

[14] Turner, J.T. (1985). Factors influencing the development of the hostage identification syndrome, *Political Psychology* **6**(4), 705–711.

[15] Cantor, C. & Price, J. (2007). Traumatic entrapement, appeasement and complex post-traumatic stress disorder: evolutionary perspectives of hostage reactions, domestic abuse and the Stockholm syndrome, *The Royal Australian and New Zeland College of Psychiatrists* **41**, 377–384.

[16] Graham, D.L., Rawlings, E.I., Ihms, K., Latimer, D., Foliano, J. Thompson, A., Suttman, K., Farrington, M. & Hacker, R. (1995). A scale for identifying "Stockholm syndrome" reactions in young dating women: factor structure, reliability, and validity, *Violence and Victims* **10**(1), 3–22.

[17] Faure, G.U. (2003). Negotiating with terrorists: the hostage case, *International Negations* **8**, 469–494.

[18] Fuselier, G.D. (1988). Hostage negotiation consultation: emerging role for the clinical psychologist, *Professional Psychology. Research and Practice* **19**(2), 175–179.

Related Articles

Syndromes: Psychological

DELANEY M. SMITH

Stop and Frisk ***see*** Profiles: Psychological and Behavioral

STR: Interpretation ***see*** Short Tandem Repeats: Interpretation

STR, Y- ***see*** Y-Chromosome Short Tandem Repeats

Stress Disorder: Acute ***see*** Posttraumatic Stress Disorder

Structured Professional Judgment ***see*** Dangerousness: Risk of

Student Threat against School ***see*** Threat Assessment: School

Subdural Hematoma ***see*** Battered Child Syndrome

Substance Abuse

Substance abuse is defined in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) primarily by social consequences resulting from substance use [1] (*contrast* **Addictions**). Substance abuse is generally recognized as being less severe than substance dependence, and an individual cannot be diagnosed with both disorders concurrently. The International Classification of Diseases (ICD) categorizes a broad range of morbidity and mortality causes [2, 3]. Currently in its 10th edition, the ICD is published by the World Health Organization. Substance dependence and harmful use are the ICD-10 classifications used to diagnose substance use disorders. Harmful use is defined as use resulting in health problems; social consequences can be included as well.

The association between crime and substance abuse is multifaceted. Criminal activity related to drug use can be a result of the illegal nature of most substances of abuse [4–6]. Since the business of buying and selling illicit drugs is not regulated, some people resort to crime to resolve disputes or gain access to drugs. Alternately, individuals might offend for the same reasons they use substances (e.g., sensation seeking, attempts to alleviate symptoms of mental health disorders), but the behaviors are not directly related to one another. In addition, some substances can cause physiological changes leading to aggression, which may be manifested criminally.

According to the US Department of Justice, nearly one-third of US inmates committed offenses while they were under the influence of a substance, and between 14 and 19% are incarcerated for crimes related to substances [6]. Among US jail inmates in 2002, 23% met criteria for substance abuse and not substance dependence. Substance abuse without dependence was most common in men and those under the age of 25, whereas substance dependence was most common in women and inmates between ages 35 and 44. Offenders with substance

abuse diagnoses were less likely than those with substance dependence to participate in substance abuse treatment programs before, during, and after incarceration. However, the Department of Justice [6] warns that caution must be used in the interpretation of any data reporting the relationship between substance abuse and crime for the following reasons:

- Crime is rarely caused by one isolated factor but is the culmination of a number of biopsychosocial characteristics and events.
- Studies reporting that offenses are related to drugs may be defining this relationship differently from one another. Calling a crime drug-related could refer to involvement of drugs in any fashion, or the definition may be more stricter, such as requiring the offender to be intoxicated at the time of the offense.
- Many studies rely on self-report, which could be inaccurate because of factors including exaggeration, poor memory, or reluctance to admit experience with substances.

In recognition of the common overlap between substance abuse and involvement with the legal system, the Substance Abuse and Mental Health Services Administration (SAMHSA) released a Treatment Improvement Protocol designed to guide professionals working with offenders who abuse substances [7]. In this protocol the conflicting goals of substance abuse treatment (which typically aims to rehabilitate) and the criminal justice system (which is largely punitive in nature) are addressed. SAMHSA discusses considerations substance abuse treatment professionals may not be accustomed to making, such as assessing psychopathy and navigating complex confidentiality laws dictating what must be communicated to probation officers.

In some cases, a drug use history is associated with greater rates of criminal activity even when crimes are committed while the offender is not under the influence of substances [8, 9]. For example, a sample of patients with functional psychosis who had received mental health services in the United Kingdom participated in a study examining their substance use histories and criminal activities [8]. Individuals who had ever used illicit substances in their lives and those whose general alcohol consumption reached at least "occasional intoxication" were significantly more likely to have committed a nonsubstance misuse-related offense. This study also found that offending commonly precedes substance abuse. These results provide evidence that criminal offenses in substances abusers have a complex etiology and cannot be completely explained by altered mental status because of intoxication during criminal activity. In another study similar results were found and the authors posited that relationships between substance use and criminal offenses might be representative of one underlying behavior pattern (e.g., antisocial or deviant behavior) rather than two exclusive elements in a cause and effect relationship (e.g., drug use causes crime) [9].

Substance abuse can complicate the process of litigation. For example, with individuals involved in court proceedings who claim mild traumatic brain injuries have resulted in significant problems, difficulties may be due to substance abuse rather than brain injury [10]. Forensic neuropsychological assessment indicating cognitive deficits in people with brain injuries can show similar deficits in substance abusers, rendering such assessment tools unreliable in differentiating between cognitive impairments caused by mild traumatic brain injury and those resulting from drug abuse (*see* **Head Injury: Neuropsychological Assessment**; **Neuropsychological Assessment**).

Adolescents

Adolescents are at risk for developing maladaptive behavior patterns that can involve illegal activity, including substance abuse (*see* **Juvenile Justice: Adolescent Development**). Several studies have examined these behaviors in youth. A five-year study that followed 80 adolescent substance abusers on probation into young adulthood found both alcohol use and marijuana use in young adulthood were associated with greater involvement in criminal behavior [11]. Further, with increased substance use, the frequency of criminal behavior was higher as well. Another study evaluated 1122 adolescents receiving substance abuse treatment throughout the US to identify patterns in their behavior prior to entering treatment [12]. The majority of youths already had a history of involvement in the criminal justice system, aside from charges related to substance use and possession. Many were ordered to treatment or chose treatment as an alternative to more restrictive consequences of their criminal behavior. Another study

of adolescents in treatment found that while male adolescents had more days of parole or probation upon treatment entry, female adolescents had more severe substance abuse and mental health symptoms at intake [13]. Involvement with the juvenile justice system may not be a reliable indicant of criminal activity, such as illicit substance use, for both male and female adolescents.

Mental Health Disorders

In offenders with substance abuse disorders, the presence of other mental health disorders can lead to additional difficulties. Bipolar disorder is an example of a mental health disorder that can relate to increased rates of criminality in people with substance use disorders. In a study investigating criminal histories of individuals with rapid-cycling bipolar disorder who abused substances, differences between men and women were assessed [14]. Of 132 participants, the majority (68%) had been charged with legal offenses, with a significantly higher percentage of men charged compared to women (78 and 53%, respectively). These rates were compared to a sample of 33 people with rapid-cycling bipolar disorder and similar demographic characteristics who did not have a comorbid substance abuse disorder. Legal charges occurred at significantly lower rates in the comparison group (8% of men and 17% of women).

A study evaluating bipolar arrestees with a history of psychiatric treatment housed in a jail's psychiatric system compared inmates ($n = 66$) with inpatient bipolar patients who had no arrest history ($n = 54$) [15]. Most inmates were manic and psychotic when they were arrested, and most had recently left inpatient psychiatric treatment. Arrestees had significantly greater rates of substance abuse (76%) than hospital patients (19%). This study suggests further treatment (e.g., aftercare) to help manage psychiatric and substance abuse symptoms may be particularly important in similar populations. Another investigation of comorbid mental health disorders found high rates of criminality and substance abuse in patients with schizophrenia. At a psychiatric clinic in Sweden, assessment of patients with schizophrenia found that nearly half reported a substance abuse history [16]. Rates of criminal history were significantly higher for participants who abused substances (50% *vs.* 11%).

Substance Abuse and Aggression

Violence and substance abuse can be associated for varying reasons [4, 5] (*see* **Aggression**). At times people use violent behavior as a way of gaining the resources necessary to use drugs, or violence can be a result of arguments related to the drug trade that cannot be resolved through legal means because of the illicit nature of the product around which the conflict is centered. Further, as with the relationship between substance abuse and criminal offending in general, both behaviors may be caused by a third factor that leads individuals to act out in a deviant manner. Additionally, some substances can cause physiological changes that lead to aggression. While there are individual differences in how people respond to substances, the potential for alcohol to elicit violence has been demonstrated through a number of pathways [17]. Alcohol intoxication changes the processes of intrinsic reward and punishment systems as well as natural physiological responses to stress and threatening situations. Higher order cognitive functions also are affected and it can lead to impairment in self monitoring, risk appraisal, planning, and motivation [18]. Mixed results have been found in investigations of the relationship between aggression and drugs other than alcohol; a consensus has not been reached in many cases [4].

A study examining female perpetrators of violence found victims to be primarily men who were strangers or acquaintances, and the most common reason for violence was reported to be engaging in a quarrel while drinking [19]. Most women were under the influence of alcohol or drugs when they committed the crimes. In addition, the majority of participants had a history of drug and/or alcohol abuse. The lack of victimization as a motive for violent behavior was speculated as being related to the cultural and political environment in Finland, where the study was conducted, which promoted gender equality.

The role of substance abuse in domestic violence offenses has been examined. In a comparison of men and women arrested for domestic violence, while men had greater numbers of prior arrests, participants of both genders were equally likely to have been using substances when they were arrested [20]. Sexual offending has been shown to relate to substance abuse in many cases as well. For example, an assessment of sexual offenders found that substance abuse predicted higher numbers of sexual assault victims [21].

Treatment

Views regarding substance abuse treatment for offenders have varied over time, and the ideal balance between treatment and punishment within the criminal justice system is still debated [22]. However, even contingents that have historically doubted the usefulness of providing treatment to substance abusing offenders have recognized the scientific evidence showing reduced relapse to substance abuse and recidivism following treatment [23].

Treatment options available to offenders range from brief motivational interviews to more lengthy participation in therapeutic communities (see [24] for a review). Motivational interviewing (MI) is a brief intervention that helps people to explore the pros and cons of changing harmful substance abuse behavior [25]. MI is a therapeutic method that can be particularly useful for offenders whose substance abuse is related to the reason for their incarceration [26]. Legal problems experienced as a result of substance abuse can be highlighted as one of the consequences resulting from abusing alcohol or drugs. In a study investigating the effectiveness of MI for offenders charged with crimes related to alcohol use, results showed that most participants (55 of 63) reported positive changes in their drinking behaviors six months after the MI intervention. Offenders remarked that they were surprised about the nonconfrontational nature of the brief intervention as well as the nonjudgmental style of the counselors.

Therapeutic communities are a form of long-term residential treatment focusing on the holistic needs of residents by using their interactions with other residents and staff as a key component of therapy (see [27] for a description). This type of intensive treatment is best suited to offenders with severe substance use disorders and maladaptive behavior patterns, such as antisocial behaviors. Drug courts are another form of treatment that can be used to promote rehabilitation in offenders charged with substance-related crimes (*see* **Mental Health Courts**). In this modality, the judge serves as a case manager of sorts who helps the offender become involved in therapeutic programs and activities [28]. Meditation is an additional method through which substance abuse can be treated, and assessment of offenders who participated in a meditation-based intervention has shown promising results including decreases in substance use, consequences of substance use, and psychiatric symptoms [29, 30].

Conclusion

Alcohol and drug abuse are disorders that can be related to forensic science because of acute physiological effects, the illegality of possession and use of many substances, as well as the potential for underlying disorders or behavior patterns of which substance abuse and criminality may be symptoms. Data indicate differences between inmate populations who abuse substances and those with substance dependence disorders [31], suggesting that diverse strategies must be used in working with these populations. Special consideration is warranted with other groups as well, such as adolescents, individuals with mental health disorders comorbid with substance abuse, and violent offenders. Several treatment options exist that can be used to achieve the desirable outcomes of reducing rates of substance abuse and recidivism.

References

[1] American Psychiatric Association (1994). *The Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition, American Psychiatric Association, Washington, DC.

[2] World Health Organization (1992). *The ICD-10 Classification of Mental and Behavioural Disorders. Clinical Descriptions and Diagnostic Guidelines*, World Health Organization, Geneva.

[3] Sartorius, N., Kaelber, C.T., Cooper, J.E., Roper, N.T., Rae, D.S., Gulbinat, W., Ustün, T.B. & Regier, D.A. (1993). Progress toward achieving a common language in psychiatry. The clinical guidelines accompanying the WHO Classification of Mental and Behavioural Disorders in ICD-10, *Archives of General Psychiatry* **50**, 115–124.

[4] Hoaken, P.N.S. & Stewart S.H. (2008). Drugs of abuse and the elicitation of human aggressive behavior, *Addictive Behaviors* **28**, 1533–1554.

[5] Easton, C. (2005). Commentary: substance abuse and criminality in the mentally disordered defendant, *The Journal of the American Academy of Psychiatry and the Law* **33**, 196–198.

[6] U.S. Department of Justice, Bureau of Justice correctional surveys, 2005, retrieved October 19, (2008). from http://www.ojp.gov/bjs/.

[7] Center for Substance Abuse Treatment (2005). *Substance Abuse Treatment for Adults in the Criminal Justice*

System, *Treatment Improvement Protocol (TIP) Series, Vol. 44*, DHHS Publication No. (SMA) 05–4056, Substance Abuse and Mental Health Services Administration, Rockville.

[8] Wright, S., Gournay, K., Glorney, E. & Thornicroft, G. (2002). Mental illness, substance abuse, demographics and offending: dual diagnosis in the suburbs, *Journal of Forensic Psychiatry* **13**, 35–52.

[9] Scott, H., Johnson, S., Menezes, P., Thornicroft, G., Marshall, J., Bindman, J., Bebbington, P. & Kuipers, E. (1998). Substance misuse and risk of aggression and offending among the severely mentally ill, *British Journal of Psychiatry* **172**, 345–350.

[10] Iverson, G.L., Lange, R.T. & Franzen, M.D. (2005). Effects of mild traumatic brain injury cannot be differentiated from substance abuse, *Brain Injury* **19**, 15–25.

[11] Clingempeel, W.G., Henggeler, S.W., Pickrel, S.G., Brondino, M.J. & Randall, J. (2005). Beyond treatment effects: predicting emerging adult alcohol and marijuana use among substance-abusing delinquents, *American Journal of Orthopsychiatry* **75**, 540–552.

[12] Kinlock, T.W., Gordon, M.S. & Battjes, R.J. (2004). Pretreatment illegal activities of a nationwide sample of adolescent substance abuse clients, *Journal of Psychoactive Drugs* **36**, 5–12.

[13] Stevens, S.J., Estrada, B., Murphy, B.S., McKnight, K.M. & Tims, F. (2004). Gender differences in substance use, mental health, and criminal justice involvement of adolescents at treatment entry and at three, six, twelve and thirty month follow-up, *Journal of Psychoactive Drugs* **36**, 13–25.

[14] Friedman, S.H., Shelton, M.D., Elhaj, O., Youngstrom, E.A., Rapport, D.J., Packer, K.A., Bilali, S.R., Jackson, K.S., Sakai, H.E., Resnick, P.J., Findling, R.L. & Calabrese, J.R. (2005). Gender differences in criminality: bipolar disorder with co-occurring substance abuse, *The Journal of the American Academy of Psychiatry and the Law* **33**, 188–295.

[15] Quanbeck, C.D., Stone, D.C., Scott, C.L., McDermott, B.E., Altshuler, L.L. & Frye, M.A. (2004). Clinical and legal correlates of inmates with bipolar disorder at time of criminal arrest, *The Journal of Clinical Psychiatry* **65**, 198–203.

[16] Cantor-Graae, E., Nordström, L.G. & McNeil, T.F. (2001). Substance abuse in schizophrenia: a review of the literature and a study of correlates in Sweden, *Schizophrenia Research* **48**, 69–82.

[17] Hoaken, P.N.S., Campbell, T., Stewart, S. & Pihl, R.O. (2003). Effects of alcohol on cardiovascular reactivity and the mediation of aggressive behaviour in adult men and women, *Alcohol and Alcoholism* **38**, 84–92.

[18] Giancola, P.R. (2000). Executive functioning: a conceptual framework for alcohol-related aggression, *Experimental and Clinical Psychopharmacology* **8**, 576–597.

[19] Weizmann-Henelius, G., Viemerö, V. & Eronen, M. (2003). The violent female perpetrator and her victim, *Forensic Science International* **133**, 197–203.

[20] Busch, A.L. & Rosenberg, M.S. (2004). Comparing women and men arrested for domestic violence: a preliminary report, *Journal of Family Violence* **19**, 49–57.

[21] Abracen, J.A., Mailloux, D.L., Serin, R.C., Cousineau, C., Malcolm, P.B. & Looman, J. (2004). A model for the assessment of static and dynamic factors in sexual offenders, *Journal of Sex Research* **41**, 321–328.

[22] Pallone, N.J. & Hennessy, J.J. (2003). To punish or to treat: substance abuse within the context of oscillating attitudes toward correctional rehabilitation, *Journal of Offender Rehabilitation* **37**, 1–25.

[23] Lipton, D.S. (1994). The correctional opportunity: pathways to drug treatment for offenders, *Journal of Drug Issues* **24**, 331–348.

[24] Resor, M.R. & Blume, A.W. (2008). Treating substance use disorders in offenders, *Journal of Behavior Analysis of Offender and Victim Treatment and Prevention* **1**, 20–35.

[25] Miller, W.R. & Rollnick, S. (2002). *Motivational Interviewing*, 2nd Edition, Guilford Press, New York.

[26] Sharp, D. & Atherton, S.R. (2006). Out on the town: an evaluation of brief motivational interventions to address the risks associated with problematic alcohol use, *International Journal of Offender Therapy and Comparative Criminology* **50**, 540–558.

[27] Blume, A.W. & Resor, M.R. (2007). Therapeutic communities for the treatment of substance abuse, in *Encyclopedia of Psychology and Law*, B.L. Cutler, ed, Sage Press Thousand Oaks.

[28] Belenko, S., DeMatteo, D. & Patapis, N. (2007). Drug Courts, in *Handbook of Forensic Mental Health with Victims and Offenders: Assessment, Treatment, and Research*, D.W. Springer & A.R. Roberts, eds, Springer Publishing, New York, pp. 385–423.

[29] Parks, G.A., Marlatt, G.A., Bowen, S.H., Dillworth, T.M., Witkiewitz, K., Larimer, M., Blume, A.W., Simpson, T.L., Lanczak, H., MacPherson, L.M., Murphy, D. & Meijer, L. (2003). The University of Washington Vipassana Meditation research project at the Northwest Rehabilitation Facility, *American Jails* **17**, 13–17.

[30] Bowen, S., Witkiewitz, K., Dillworth, T.M., Chawla, N., Simpson, T.L., Ostafin, B.D., Larimer, M.E., Blume, A.W., Parks, G.A. & Marlatt, G.A. (2006). Mindfulness meditation and substance use in an incarcerated population, *Psychology of Addictive Behaviors* **20**, 343–347.

[31] Kjelsberg, E. (2005). Conduct disordered adolescents hospitalised 1963–1990, *European Child & Adolescent Psychiatry* **14**, 191–199.

MICHELLE R. RESOR

Substance Addiction *see* Addictions

Sudden Death Due to Natural Causes: Cardiac ***see*** Cardiac and Natural Causes of Sudden Death

Suggestibility ***see*** Deception: Truth Serum; Elderly in Court; Interrogative Suggestibility; Memory: Repressed

Suggestibility in Interrogations ***see*** Interrogative Suggestibility

Suggestibility of Adults' Eyewitness Memory ***see*** Eyewitness: Suggestibility of

Suggestibility of Children ***see*** Children: Suggestibility of

Suicide: Accidental Autoerotic ***see*** Autoerotic Deaths

Suicide (Behavior)

Introduction

Homicide–suicide is the occurrence of suicide preceded by homicide within a short period of time; typically 24 h. It is an event that occurs rarely at a rate of approximately 0.2 per 100 000 individuals [1]. This incidence rate has remained stable over time and is relatively consistent across nations [2]. In the United States, homicide–suicide accounts for approximately 1200–1300 deaths annually as compared with approximately 16 000 annual murders and 30 000 suicides [3–5]. During 2005, over 10 homicide– suicides occurred every week [3].

The infrequent incidence of homicide–suicide renders it a difficult subject for empirical examination. As a result of the limited research focusing on homicide–suicide, there remains some inconsistency in the language used to describe it. The terms *murder–suicide, extended suicide*, and *dyadic death* are also used to describe the unusual phenomenon [6].

In most homicide–suicide events, the perpetrator is a male (95% of US cases) and the victim is female (85% of US cases). In more than half of all homicide–suicide cases, a jealous male murders a wife or girlfriend who has expressed a desire to divorce or separate from him. Although there is usually one perpetrator and one victim, there may be multiple victims [2]. Firearms were the method used in approximately 95% of the homicide–suicides reported between January 1 and June 30, 2001 [3].

Although homicide–suicide is a psychological phenomenon distinct from either homicide or suicide, it may be more similar to suicide than homicide. Similar to individuals who complete suicide, perpetrators of homicide–suicide are more likely to suffer from depression, feel hopeless, and have prior severe suicide attempts. Unlike individuals who commit homicide, perpetrators of homicide–suicide do not typically have histories of violence or impulsivity [2].

There are four types of homicide–suicide: spousal homicide–suicide, filicide–suicide, familicide–suicide, and extrafamilial homicide–suicide. Spousal homicide–suicide is further divided into jealous type and declining health type. Comprising more than half of all homicide–suicide cases, spousal homicide–suicide jealous type is the most common type of homicide–suicide while spousal homicide– suicide declining health type is less common. In the typical spousal homicide–suicide declining health type scenario an older male murders his physically ill spouse before committing suicide. Often, the older male has experienced feelings of hopelessness and depression [2].

Filicide–Suicide

Filicide–suicide is a specific type of homicide–suicide wherein a parent commits suicide after killing his or her child(ren) [7]. Sixty percent of children under the age of five who die by homicide are killed by a parent [8]. Parents kill their children for five reasons: altruism, acute psychoticism, accident, because the child is unwanted, or as an act of revenge against a spouse. Parents who are acutely psychotic or who have an altruistic motive for killing their child are more likely to also kill themselves than parents who murder a child by accident, because the child is unwanted, or for revenge against a spouse. Most filicide–suicides (70%) are of the altruistic type; a "mercy" killing in which a parent murders his or her child(ren) as a means of removing the child(ren) from real or imaged danger or suffering. Fathers commit filicide–suicide twice as often as mothers [9]. Often, parents use the same means to kill themselves as they used to kill their children. Guns are the most common method.

Familicide–suicide

Familicide–suicide is an event during which a parent kills his or her children and spouse (and sometimes pets and other relatives) prior to committing suicide [2]. Familicide–suicide is less common than filicide–suicide [7]. Older men who suffer from depression are most likely to commit familicide–suicide [2].

Extrafamilial homicide–suicide

Extrafamilial homicide–suicide is an event during which an individual who perceives to have been treated unjustly murders those individuals he or she believes to be responsible prior to committing suicide. Although extrafamilial homicide–suicide is most often portrayed in the media, it is the least common type of homicide–suicide [2].

References

[1] Bossarte, R., Simon, T. & Barker, L. (2006). Characteristics of homicide followed by suicide incidents in multiple states, 2003–04, *Injury Prevention* **12**, ii33–ii38.

[2] Hillbrand, M. (2001). Homicide-suicide and other forms of co-occurring aggression against self and against others, *Professional Psychology* **32**, 626–635.

[3] Violence Policy Center (2006). *American Roulette: The Untold Story of Homicide-Suicide in the United States*, http://www.vpc.org (accessed Jun 2007).

[4] Federal Bureau of Investigation (n.d.) (2004). *Crime in the United States 2004*, http://www.fbi.gov/ucr/cius_04/offenses_reported/violent_crime/murder.html (accessed Jun 2007).

[5] Centers for Disease Control and Prevention (n.d.) (2007). *Suicide Fact Sheet*, http://www.cdc.gov/ncipc/factsheets/suifacts.htm (accessed Jun 2007).

[6] Saleva, O., Putkonen, H., Kiviruusu, O. & Lönnqvist, J. (2007). Homicide-suicide – an event hard to prevent and separate from homicide or suicide, *Forensic Science International* **166**, 204–208.

[7] Shackelford, T., Weekes-Shackelford, V. & Beasley, S. (2005). An exploratory analysis or the contexts and circumstances of filicide-suicide in Chicago, 1965–1994, *Aggressive Behavior* **31I**, 399–406.

[8] Bureau of Justice Statistics, U.S. Department of Justice (n.d.) (2007). *Homicide Trends in the United States: Infanticide*, updated 2006, http://www.ojp.usdoj.gov/bjs/ (accessed Jul 2007).

[9] Hatters-Friedman, S., Hrouda, D., Holden, C., Noffsinger, S. & Resnick, P. (2005). Filicide-suicide: common factors in parents who kill their children and themselves, *Journal of the American Academy of Psychiatry and the Law* **33**, 496–504.

KATRINA MCCOY, TRACY A. THOMAS AND WILLIAM J. FREMOUW

Suicide: by Elderly ***see*** Elder Abuse: Policy

Suicide: Following Murder ***see*** Suicide (Behavior)

Supreme Court (US) Cases on Admissibility of Expert Opinion Testimony ***see*** *Daubert v. Merrell Dow Pharmaceuticals*

Survival Identification by Hostages ***see*** Stockholm Syndrome

Swabs: Biological ***see*** Biological Swabs

Swabs: Differential Extraction ***see*** Differential Extraction

Swabs: Intimate ***see*** Biological Swabs

SWAT Teams ***see*** Policing and Critical Incident Teams

Sweat: Drugs ***see*** Sweat: Toxicology

Sweat: Toxicology

Physiological Aspects of Drug Distribution in Sweat

Theoretical Background

Sweat secretion is an important mechanism for maintaining a constant core body temperature [1]. Following sympathetic nerve stimulation, sweat is excreted onto the surface of the skin and evaporated to release body heat. Sweat is secreted from two types of sweat glands, eccrine and apocrine glands (Figures 1 and 2). These glands originate deep within the skin dermis and terminate in excretory ducts emptying onto the

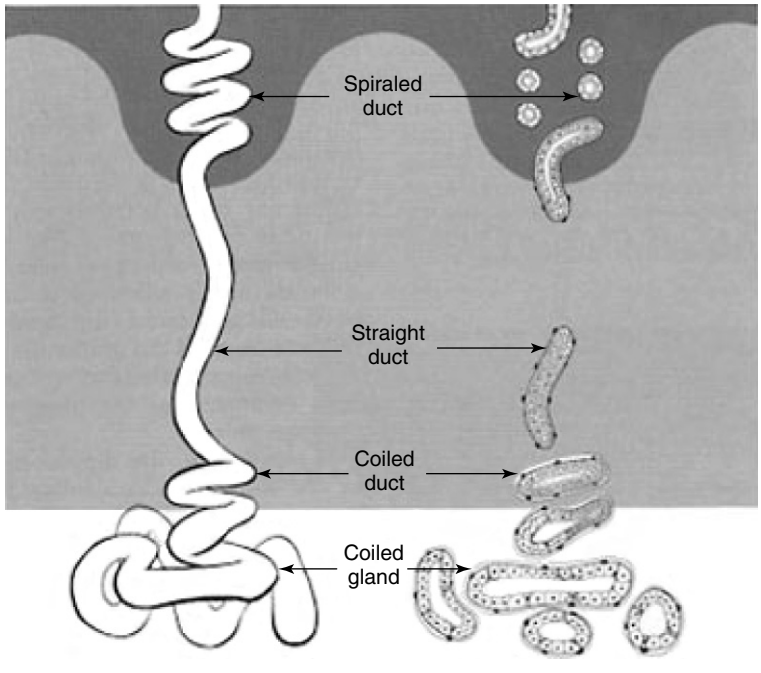


Figure 1 The eccrine gland: eccrine units consist of three portions: (a) the acrosyringium or intraepidermal spiraled duct; (b) the coiled and straight intradermal duct; and (c) the secretory-coiled gland

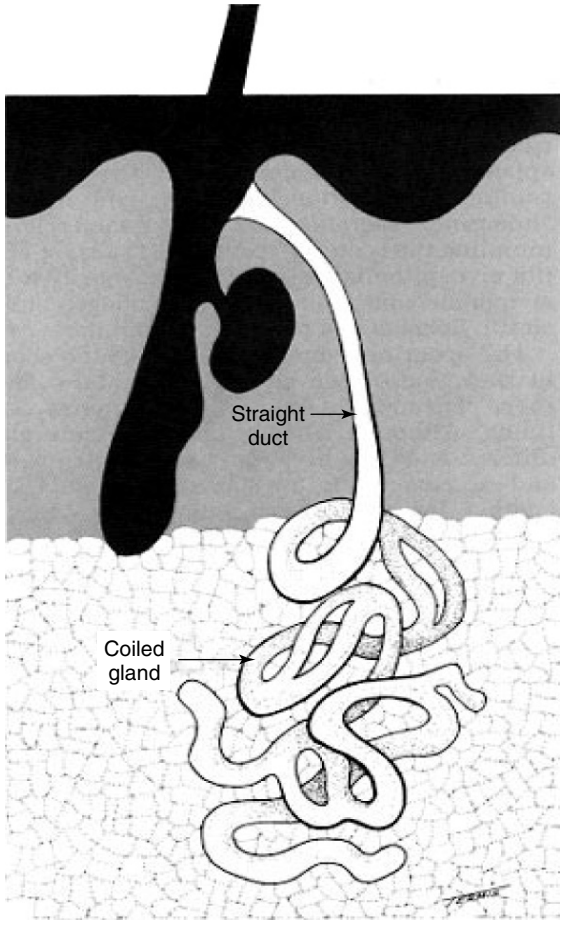


Figure 2 The apocrine unit: apocrine units consist of a secretory-coiled gland and a straight duct that traverses the dermis and empties into the isthmus (uppermost portion) of a hair follicle

skin or the developing hair follicles. Eccrine glands are located on most skin surfaces (not necessarily close to hair follicles), with the highest density on the palm of the hand, the sole of the foot, and on the forehead. These glands are innervated by sympathetic nerves. An odorless, sterile, aqueous hypotonic solution of weak acid pH is produced, derived from plasma by passive filtration. Water is the primary constituent approximately 99%, and sodium chloride is the most concentrated solute. Amino acids, biogenic amines, and vitamins are present only in trace amounts. The pH (4–6) is strongly associated with the amount of lactic acid excreted [2]. Apocrine glands are larger glands located in the shaft of the hair follicles from the armpit, genitalia, and anus. These glands are controlled by a cholinergic innervation but can be also stimulated by adrenergic agents. They secrete a more viscous, cloudy, yellow-white liquid, which is primarily sterile and odorless, rich in cholesterol (75%), triglycerides, and fatty acids (20%) [3].

Sweat also contains albumin, gamma globulins, waste products, trace elements, drugs, and many other substances found in blood.

The rate of sweating is highly dependent upon environmental temperature. Above 31 °C, humans begin to sweat and may excrete as much as $3\,l\,h^{-1}$ over short periods of time [4]. The average sweat pH of resting individuals was reported to be 5.82. Following exercise, the pH was found to increase with the increase in flow rate and was reported to be between 6.1 and 6.7. Approximately, 50% of sweat is generated by the trunk of the body, 25% from the legs, and the remaining 25% from the head and the upper extremities [4].

Multiple mechanisms including passive diffusion and transdermal migration have been suggested for the incorporation of drugs into sweat [5, 6]. Passive diffusion of drugs from blood to sweat is favored for lipid-soluble substances. As distinct from urine, where only drug metabolites are usually present, and from hair, where mainly parent drugs are incorporated, both parent drugs and their metabolites can be detected in sweat, similar to blood. Nonionized basic drugs diffuse into sweat and become ionized as a result of the lower pH of this biological fluid. Basic substances may accumulate in sweat as compared to blood due to the difference in pH between the two matrices. Some drugs may migrate across the dermal and epidermal layers into the stratum corneum. It has to be reminded that drugs are not only excreted by sweat glands but also by sebaceous glands or transdermal liquid transport. Sebaceous glands are associated with hair follicles and they are particularly abundant on the scalp and the forehead. They produce a viscous, yellow-white liquid oily sebum that consists of triglycerides (60%) and wax esters (23%). Excreted sweat and sebum cannot be examined separately. They form an emulsion on the skin surface, which facilitates intercellular diffusion along the cell membrane and a transcellular diffusion and transportation by the keratinocytes [3, 7].

Sweat testing is relatively noninvasive, and identification of drug in sweat may serve as a means

of monitoring recent drug use with a window of detection that can be somewhat wider than that provided by urine testing [8]. Although the use of sweat for drug testing has been hampered by difficulties in sample recovery and sensitivity of analytical methods [9], there are some distinct advantages in using this matrix:

1. Matrix collection is noninvasive.
2. Rapid qualitative on-site tests may be applied to this matrix.
3. Sweat testing can provide a means to obtain information on recent drug use and a cumulative estimate of drug exposure over a period of several weeks, depending on the device applied for collection.
4. The process of sample adulteration is quite difficult.

The disadvantages include the following:

1. Collected volume is generally small (from few μL in normal conditions to 1 – 5 mL when sweating is induced by physical exercise.
2. Production of a quantitative result in drug testing is difficult, since the volume of sweat collected is not measurable when traditional collections systems (cotton wipes, sweat patches) are used.
3. Possibility of environmental contamination should be taken into account.

Collection Devices for Sweat

One of the major difficulties associated with sweat has been its collection and the measurement of the volume excreted in a defined time period. The amount of sweat excreted considerably varies among individuals and even within a single person, and is dependent upon their daily activities, emotional state, and environment.

Systematic collection of specimens is difficult because of the unequal distribution of sweat glands. Indeed, because of regional differences in sweat gland density, the site of sweat collection has a striking impact on the collected quantity of drugs. Furthermore, since apocrine sweat glands are always proximal to sebaceous glands, mixed secretions of sweat and sebum, which are incorrectly referred to as sweat, are generally collected. Furthermore, different perspiration rates and volumes of sweat between individuals could also explain the large intersubject variability in sweat excretion of drugs as shown in controlled administration studies with 3,4 methylenedioxymethamphetamine (MDMA), codeine, and cocaine [10–12].

Sweat can be collected noninvasively by absorption with gauze, cotton wipes, or filter paper, or by collection of liquid sweat in rubber gloves or plastic body bags [8, 13, 14]. To increase sweat production and uptake, physical exercise is required during collection and pads are saturated with sodium chloride solution.

Thermal [15–17] or pharmacological stimulations using pilocarpine [18] have been proposed to secrete large amounts of sweat. In particular, pilocarpine stimulation has been used as a test for the diagnosis of cystic fibrosis via the determination of chloride [8]. In recent years, a sophisticated device "Macroduct® sweat stimulation by pilocarpine iontophoresis and sweat collection system" (Wescor Biomedical, Logan, UT, United States) has been developed and successfully applied in clinical studies using sweat testing in the diagnosis of cystic fibrosis [19] and drug testing at a pilot scale [20]. A modified model of the above-reported apparatus, the "Nanoduct Neonatal Sweat Analysis System®" (Wescor Biomedical, Logan, UT, Unites States) has been designed for the early laboratory diagnosis of cystic fibrosis in neonates [21].

In the 1980s, Phillips reported for the first time the development of an adhesive patch for the long-term collection of sweat (10 days) [22]. This patch was occlusive in design, trapping both the solute and water components, thus permitting the determination of analyte concentration in sweat. A disadvantage of the occlusive patch design was the limited time the patch could be worn because it caused a change in the matrix environment and skin irritation for the wearer [23].

A linear uptake rate of 18 to 47 mg/day of sweat was observed. It was suggested that it might be possible to monitor drug-taking behavior through the use of this patch. This sweat patch was later utilized in the validation of self-reports concerning alcohol use [24].

In the last decade, a new nonocclusive sweat collection device, the PharmChem® Sweat Patch from PharmChem Laboratories (Menlo Park, CA, United States), was developed and successfully applied to sweat testing for drugs of abuse [11, 25–27]. Typically, this type of sweat patch is worn during a

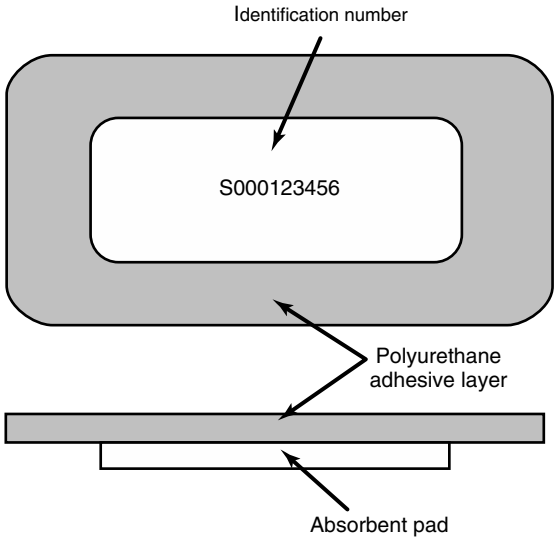


Figure 3 Scheme of PharmChek type patches used for sweat collection

period of one week, but can be also worn for shorter time periods (from several hours to some days) [10, 28].

This patch can be worn for an extended period and concentrates solutes on a collection pad while allowing water to evaporate from the patch. The device consists of an adhesive layer on a thin transparent film of surgical dressing to which a rectangular absorbent pad is attached (Figure 3). Sweat concentrates on the absorbent pad while oxygen, carbon dioxide, and water vapor escape through the transparent film. Larger molecules are excluded by the molecular pore structure of the plastic membrane. The skin must be thoroughly cleaned with isopropyl alcohol prior to affixing the patch to prevent contamination of the patch or interference with the deposition or detection of analytes on the patch. Attempts to remove the patch prematurely or tamper with the device are readily visible to persons trained to remove the sweat patch. Care must be taken not to contaminate the absorbent pad when removing and storing the patch prior to analysis [1].

A potential problem with the PharmChek patch is the absence of a layer between the skin and the absorptive pad to prevent bacterial transfer into the pad and, therefore, the possibility of bacterial growth and drug degradation. Careful preparation of the skin prior to application of the patch should kill or remove bacteria and prevent these problems. Also the absence of substantial moisture in the pad decreases the possibility of bacterial growth [17]. Several patches have been applied (in the back and shoulders) to subjects participating in controlled administration studies to folow the kinetics of drug distribution in sweat. Nevertheless in a situation of sweat testing of illicit drugs in parolees, probationers, or pretrial individuals, a single patch is applied in the forearm of subjects.

Fast Patch devices were developed in the 1990s as an alternative to PharmChek type sweat patches, in order to allow rapid collection of sweat specimens. The Handheld Fast Patch applied to the palm of the hand and the Torso Fast Patch applied to the abdomen or the sides of the trunk were used in controlled trials for cocaine and codeine, but they are no longer commercially available. Both patches employed heat-induced sweat stimulation and a larger cellulose pad for increased drug collection [14].

In Table 1, amounts of recovered substance when applying different sweat collection devices are reported. Cocaine has been used as a model substance to illustrate the performance of collection devices under different experimental conditions (controlled studies vs. field trials).

In this concern, Kidwell *et al.* in conjunction with Jet Propulsion Laboratory developed a system for the monitoring of parolees, probationers, or pretrial individuals. This system both collects sweat and tests it remotely [8]. Such monitoring may be more invasive than that used in the general population because the individuals are under a court order to abstain from drug use. In one embodiment of this device, labeled antibodies are bound to an immobilized drug layer. Drugs in sweat displace a small amount of these antibodies that are then trapped in a superabsorbent polymer layer. An optical system then detects the presence of the label and the readout may be transferred to a remote location using cellular phone technology. The superabsorbent polymer layer facilitates the absorption of substantial amounts of sweat before it becomes saturated, which improves the sensitivity of the device. Two polycarbonate membranes control fluid flow into and out of the sweat badge and improve user comfort. Similar to the polyurethane covering in the PharmChek system, the outer polycarbonate membrane prevents backdiffusion of liquid from the external environment, yet allows evaporation of moisture [8].

Table 1 Concentrations of cocaine in sweat as a function of sample procedure procurement

Sweat collection procedure	Number of subjects	Dose	Concentrations	Reference
PharmChek® Sweat patches	9	3 × 75 mg/kg (every 24 h in 1 week) 3 × 150 mg/kg (every 24 h in 1 week)	First week COC 3.7–197.1 ng/patch; mean(SD) BE 2.6–62.3 ng/patch; mean (SD) EME 3.6–29.2 ng/patch; mean (SD) Second week COC 2.6–59.7 ng/patch; range BE 4.3–162.2 ng/patch; range EME 3.9–41.9 ng/patch; range Third week COC 5.0–13.8 ng/patch; range ND EME 5–6.1 ng/patch; range	Kacinko *et al.* 2005
PharmChek® Sweat Patches	22	field trial	COC First day 325.8 (644.4) ng/patch; mean (SD) First week 1195.0 (338.1) ng/patch; mean (SD) Second week 1664.3 (2330.3)ng/patch; mean (SD) EME First day 29.9 (63.9) ng/patch; mean (SD) First week 117.0 (131.5) ng/patch; mean (SD) Second week 210.9 (279.5) ng/patch; mean (SD) BE First day 59.2 (370.3) ng/patch; mean (SD) First week 181.9 (348.2) ng/patch; mean (SD) Second week 723.6 (748.6) ng/patch; mean (SD)	Liberty *et al.* 2004
Torso fast patches	4	3 × 75 mg/kg (every 24 h in 1 week) 3 × 150 mg/kg (every 24 h in 1 week)	COC First day 22–2085 ng/patch; range BE First day 3–44 ng/patch; range COC First day 40–1463 ng/patch; range BE First day 18–60 ng/patch; range	Huestis *et al.* 1999
Handheld fast patches.	4	3 × 75 mg/kg (every 24 h in 1 week)	COC First day 33–739 ng/patch; range BE First day 11–121 ng/patch; range	

(*continued overleaf*)

Table 1 (*continued*)

Sweat collection procedure	Number of subjects	Dose	Concentrations	Reference
		3 × 150 mg/kg (every 24 h in 1 week)	COC First day 150–3579 ng/patch; range BE First day 15–127 ng/patch; range	
Skin swab	18	Field trial	COC 12.3 (21.7) ng/swab; mean (SD) 2.2–91.3; range (a reported value >940 excluded) BE 2,1 (1,5) ng/swab; mean (SD) 0.9–5.2; range (a reported value 141 excluded)	Kidwell *et al.* 1997

ND = non detected, COC = cocaine, BE = benzoylecgonine, EME = ecgonine methylester

In none of the collection tools, volume of collected sweat can be known and thus, amount of detected drug has to be considered "per patch" or "per cotton wipe" unless the quantity of drug can be related to either sodium or lactate concentration, both substances excreted relatively constantly in sweat and entrapped in patches and in wipes.

Considering the described devices, two different methodological approaches can be considered for drug testing in sweat [29].

The first is aimed at the detection of recent drugs use (less than 24 h) and involves only collection of sweat at a point in time. A punctual sweat collection coupled to an immunochromatographic test of the sample provides a qualitative result [10, 30, 31], or drugs in sweat collected on a cotton wipe can be extracted and subjected to confirmatory analysis [32]. This approach is mainly oriented to identify individuals who are under the influence of drugs.

The second approach is based on patch technology and allows monitoring of illicit drug use for time windows wider than those provided by urine testing. This is because the patches can be worn for up to one week. Drugs accumulate in the collection device, and little or no drug degradation seems to occur during this time interval [33]. Patch technology is used mainly for the follow-up of drug addicts under treatment to verify abstinence. Both approaches benefit from not being invasive and pose fewer ethical problems for sample collection compared to blood or urine testing.

However, there are several issues that still require a thorough investigation as they may limit the application of sweat testing using patch technology or at least the interpretation of obtained results: (i) environmental skin contamination, (ii) drug absorption/loss through patch membrane, and (iii) drug reabsorption from patches.

Sweat patches are sealed to the skin and are designed to exclude environmental contamination. Nonetheless, an *in vitro* study showed that several drugs of abuse directly applied to the skin of drugfree subjects may persist there for several days [34]. Therefore, cleaning procedures (e.g., by applying a swab impregnated with isopropanol to the skin) are recommended before application of the sweat patch to completely remove drugs deposited on skin. Thus, it might be argued that an environmental contamination prior to patch application is a possible occurrence, especially in the case of drugs of abuse that are usually smoked, such as cocaine or cannabis. Experiments proved that cocaine and its metabolite benzoylecgonine directly applied to the skin are more persistent than delta-9-tetrahydrocannabinol (THC) and 11-nor-delta-9-tetrahydrocannabinol-9-carboxylic acid (THC-COOH), but in any case these substances can remain detectable for up to three days after normal hygiene [34]. For this reason, the recommended procedure is to "clean" the skin with 70% isopropanol swabbing before application of the patch. However, isopropanol, being less polar than water, would neither hydrate the skin layer nor assist in breaking the ionic bonding of the drugs to skin

protein. Therefore, once bound to the skin, drugs could remain for several days owing to ionic bonding, not be removed by "cleansing" with the isopropanol swab, and appear in the patch after sweat- induced transfer, with the occurrence of greater concentrations when sweating is greater.

Hence, studies involving drugfree subjects exposed to heavily contaminated environments with cocaine or cannabis smoke should be performed in order to evaluate the relevance of these observations.

There are some reports suggesting that patch membrane is not only permeable to gases and water vapor but also to drugs to some extent[34, 35]. Drugs of abuse (e.g., cocaine, methamphetamine, heroin) applied on the surface of patch membrane can be absorbed, and drugs excreted in sweat may diffuse through the membrane to the outer environment. These findings suggest that membranes are more permeable than expected, although the rate of substance able to diffuse to the outer side is lower than 15% for the drugs studied [34, 35].

Finally, several studies suggest that there is a time-dependent loss of drugs during patch wearing over time. The loss of drugs of abuse from skin patches limits one of the goals of sweat patch testing: cumulative drug detection. Drugs may not persist inside the patch for two reasons: (i) they may be degraded by enzymes present in sweat, bacteria inside the patch, or the humid environment or (ii) they may equilibrate with the skin, passing from the patch back into the skin, eventually through the skin into the body, and be distributed and eliminated.

The reabsorption back to the skin has been recently hypothesized for two basic drugs, cocaine and MDMA [10, 35].

In both cases, the amount of drug spiked in the patches and applied to the back of volunteers declined over time, showing an inflexion in the kinetics at 24 and 10 h postadministration. Interestingly, in case of MDMA, a higher reduction of drug concentration in the 10 h postapplication patch was observed in a volunteer exercising in the gymnasium. It is known that temperature and physical exercise can change sweat flow and pH. It is known that a rise in body temperature or physical exercise could increase sweat flow producing an increase in sweat pH and the volume of sweat in contact with the patch, thereby facilitating drug transfer to the skin [34]. Indeed, in the above-reported study, the subject with a likely higher percentage of skin reabsorption participated in gymnasium exercise session during the experiment. This observation may be of relevance considering the consumption patterns of drugs such as cannabis cocaine or MDMA (parties, overcrowded, with high ambient temperature and physical exercise associated to dancing). Nevertheless, the impact of this observation in sweat testing performed under 'normal' conditions (i.e., seven days wearing patch) is unclear, as well as the rate of reabsorption for different drugs of abuse, especially in case of neutral or acidic drugs, poorly excreted through skin. Nevertheless, it is concluded that the final amount of a drug collected in the patch is the result of a more dynamic process (excretion vs. reabsorption) than expected. In this sense, chronic heavy users would be eventually detected, whereas this may not be the case for recreational users, giving rise to false-negative results.

Skin environmental contamination and drug reabsorption from patches that may lead to false-positive and false-negative results in the interpretation of analytical findings have to be considered, and experimental studies addressing these issues should be performed before drug testing in sweat may be considered routine.

Analytical Approaches

On-site Testing versus Instrumented Immunoassays

Drugwipe from Securetec (Ottobrunn, Germany) is the only non-instrument-based on-site immunochemical assay available for drug testing in sweat. It was designed for the detection of traces of illegal drug residues in body surfaces and liquid media. Different devices can be used for the detection of each class of drugs of abuse (opiates, cannabis, cocaine, amphetamines), although devices combining more than one drug class are also available. A pink color appearing in the test window should be compared with the color-check triangle on the Drugwipe cover. Sweat is collected with the wiping pad of the test kit moistened with 20 ml tap water. The other part of the detection element is an adsorbent pad, which had to be dipped into tap water to initiate the immunochromatographic reaction. In the presence of drug, the detection field of Drugwipe changes its color from white to pink, depending on the amount of drug collected. According to the manufacturer, the coloration of the whole test window is not necessary, but a line

on the edge could be sufficient to qualify a positive result. Coloration remains stable for at least 1 h. Results are classified as positive or negative either by qualitative observation of an analyst or by applying a reading device known as Drugread. Limits of detection of this device are in the nanogram range and vary from 10 ng for methamphetamine, methylenedioxymethamphetamine (MDMA, "ecstasy"), and cocaine, to 25 ng for opiates and 100 ng for cannabis. Drugwipe has been applied in different settings for the detection of MDMA, amphetamines, cocaine, opiates [10, 30–32, 36, 37]. While performance of Drugwipe when monitoring psychostimulants consumption is adequate, this is not the case for other drug classes like cannabinoids, where sensitivity is not adequate for the purpose of drug testing.

Different immunoassays including radioimmunoassay (RIA) and enzyme-linked immunosorbent assay (ELISA) have been used for the detection of drugs in sweat [10, 12, 25–27, 38–41]. They differ from the current methods used for drug testing in urine since they are usually directed at the detection of the parent drugs and more lipophilic metabolites excreted in sweat, and they have an increased sensitivity because of the lower concentrations of drugs in this biological matrix or the lower amount of collected matrix. The most popular immunoassay used for drug screening in sweat has been microplate ELISA marketed by STC Diagnostics (Bethlehem, PA). This ELISA technique is a solid-phase competitive immunoassay involving horseradish peroxidase labeled with a drug derivative. The sample (or a calibrator or control) is allowed to incubate for 30 min with the labeled enzyme. After competition to bind to an antibody linked to the surface of the well where reaction takes place, wells are washed with water, a substrate (3,3′,5,5′-tetramethylbenzidine) is added, and the color produced after 30 min incubation is measured at 450 and 630 nm. Absorbance is inversely proportional to the amount of drug present in sweat. These ELISA techniques were originally designed for the detection of drugs in plasma and some modifications were needed for sweat testing. The main modification is the need for extraction of drugs from the patch/skin swab. Typically, cellulose is extracted with an elution buffer (e.g., 2.5 ml methanol/acetate buffer 0.2 M pH 5.0, 75 : 25) [25, 40]. The composition of the elution buffer may vary among authors depending on the drug targeted [27, 42], and a further solid-phase extraction (SPE) of the buffer extract has been also used to improve the performance of ELISA [10]. After shaking cellulose for 30 min with the elution buffer an aliquot of the eluate is analyzed by immunoassay. It has been suggested that, after collection of sweat, the patches should be stored at $-20\,^{\circ}C$ until analysis for an improved recovery of drugs, particularly for delta-9-tetrahydrocannabinol. Indeed, for this drug it has been observed that the recovery from cellulose is not quantitative, as part of the drug is bound to the cellulose [43].

Mass Spectrometry Coupled to Chromatographic Techniques

As stated earlier, there are no available on-site tests for sweat testing with the sufficient sensitivity to detect drugs at concentrations typically found in this matrix (i.e., cannabinoids) [31]. Therefore, samples are analyzed by laboratories after a preliminary screening by ELISA tests and further confirmation by mass spectrometry coupled to chromatographic techniques or directly by the later analytical techniques. Drugs from patches or skin swab have to be eluted by procedures previously discussed for the analysis of sweat by immunoassays. The eluate is further subjected to solid-phase extraction following procedures very similar to those applied in the analysis of drugs in urine or oral fluid with a specific requirement of increased sensitivity [44]. These sensitivity requirements have been defined by the United States Substance Abuse and Mental Health Services Administration (SAMHSA) mandatory guidelines for workplace drug testing programs when sweat is used as biological matrix for drug testing [45]. Cutoff concentrations for reporting positive findings are defined in terms of concentration per 2.5 ml of eluate. Some authors postulate that results should be reported as concentration per patch or skin swab [46].

Sweat Testing (Patch Technology) in Substance Abuse Treatment Programs as an Alternative to Urinalysis or Oral Fluid Testing

Sweat testing (patch technology) has been postulated as an alternative to urine testing in treatment and probation programs. Some of the advantages include a wider detection window, easier sample collection, and reduced opportunities for sample tampering. As an example, an added value in methadone

maintenance programs is the increased ability of detecting heroin and/or 6-acetylmorphine in sweat when compared to urine due to specific distribution patterns affecting both matrices, and was discussed previously [27]. This fact helps the differentiation between heroin misuse and other sources of opiates like therapeutic codeine prescription or the consumption of poppy seeds of dietary origin [27]. Overall, most studies agree that there is an advantage in sweat testing over urine testing, in particular, when urinalysis is performed twice or less weekly for monitoring opiate consumption [27, 47–50]. In a study comparing sweat testing (patch worn once in a week) and urine testing (thrice weekly) in a methadone maintenance program, results obtained for opiates by ELISA (sweat) and enzyme multiplied immunoassay technique (EMIT) (urine, enzyme multiplied immunoassay) were compared. Taking EMIT results as a reference for the comparison, an apparent rate of 13.5% of false-negative results and 7.9% of false-positive results was attributed to ELISA results in sweat. The sensitivity, specificity, and efficiency of sweat patch results to urine results for opiates were 68.6, 86.1, and 78.6%, respectively [27]. The clinical sensitivity and specificity for detecting drug use by analyzing sweat collected from human subjects following known doses of codeine (0, 30, and 60 mg orally) or heroin (20 mg intravenously) were 76 and 100%, respectively in an study promoted by sweat patch providers [26].

Some of the limitations of the sweat testing approach are the lack of correlation between drug exposure and patch concentrations and the subsequent qualitative nature of results, lack of analysis automation, and lack of controlled studies (except for cocaine and codeine) providing references for the interpretation of results [51, 52]. Further limitations lie on the potential of generating false-positive results because drug release from adipose tissue diffuse to skin in former heavy users once drug consumption is discontinued as it has been postulated in case reports [53] or environmental contamination [54]. A typical example is the possibility of methadone monitoring in sweat (including its chiral analysis) but because of the lack of correlation between doses and drug concentrations, this detection is just limited to demonstrate drug exposure [30].

Sweat patch technology testing has also been applied to monitor cocaine use in treatment programs and clinical trials. In the same study where 44 methadone maintenance patients were monitored for opiates misuse, cocaine consumption was also controlled by sweat testing [39]. ELISA techniques were used for the analysis of cocaine in sweat and results compared to those obtained by GC/MS in the same matrix and with EMIT results in urine. Cocaine was detected by GC/MS in 99% of presumptive findings by ELISA. The sensitivity, specificity and efficiency of ELISA when compared to GC/MS were 93.6, 91.3, and 93.2% respectively. When comparing sweat ELISA with urinary results by EMIT these figures were 97.6, 60.5, and 77.7%. Overall these results supported the use of sweat patch technology for monitoring cocaine use, confirming previous studies where cocaine was administered in a controlled setting [12]. In the context of a clinical trial in cocaine-dependent patients, sweat patches were compared with urine testing and participants' self-report of drug misuse [55]. Patches were worn weekly or per-visit and urine was controlled thrice weekly. There was a good correlation (0.96) between quantitative measurements in sweat patches (weekly vs. per-visit) and between sweat patches and urine analysis of benzoylecgonine (0.76 and 0.73 respectively). The per-visit sweat patch provided cocaine use data in 80.5% in study days ($n = 70$) while urinalysis and weekly sweat patch provided 77.4 and 76.1% respectively. The authors of this study concluded that sweat patch could be an alternative to urine testing as an outcome measure in cocaine clinical trials and that cost incurred in drug testing in sweat compared favorably with that in urinalysis.

Sweat Testing (Wipe Approach) in the Identification of DUID

Sweat could be a matrix of interest for monitoring drivers presumed to be under the influence of drugs (DUID), particularly if rapid drug tests would be available for on-site testing at the roadside [56]. Some studies already show that a combined approach of drug testing in different matrices (oral fluid, sweat and urine) can confirm approximately 98% of policemen suspicions at the roadside concerning drug use by drivers [57].

Two approaches have been used for monitoring drug use in sweat among impaired drivers presumed to be under its influence (DUID). In the first approach, sweat has been obtained with cosmetic pads moistened with isopropanol from forehead

wipes. Alternatively a noninstrumented immunoassay (Drugwipe, Securetec) has been applied for on-site testing after applying a collection pad included in the analytical device to the skin of subjects (e.g., forehead, neck, or back). Overall the skin swab approach provides results similar to oral fluid or urine drug testing but after collection the sample needs to be transferred to analytical laboratories for drug testing [31, 32, 42]. The most favorable situation for police officers that prefer sweat collection over other biological matrices would be the availability of on-site drug tests for sweat at the roadside. Nevertheless, available analytical devices lack the sufficient low sensitivity to detect drugs at concentrations present in this fluid, in particular for cannabinoids and benzodiazepines [42]. The best correlations between on-site Drugwipe results and plasma findings after applying chromatographic techniques in a series of 180 drivers who failed the field sobriety tests at police roadblocks are observed for basic drugs like psychostimulants (95% for amphetamines), and a poorer performance is observed for cocaine and opiates (76 and 75% respectively) [32].

References

[1] Huestis, M.A. & Cone, E.J. (1998). Alternative testing matrices, in *Drug Abuse Handbook*, S.B. Karch, ed, CRC Press, Boca Raton, FL, Chaper 11, pp. 799–814.

[2] Robinson, S. & Robinson, A.H. (1954). Chemical composition of sweat, *Psychological Review* **34**, 202–220.

[3] Sunshine, I. & Sutliff, J.P. (1997). Sweat it out, in *Handbook of Analytical Therapeutic Drug Monitoring and Toxicology*, S.H. Wong & I. Sunshine, eds, CRC Press, Boca Raton, FL, 253–264.

[4] Randall, W.C. (1953). The physiology of sweating, *American Journal of Physical Medicine* **32**, 292–318.

[5] Nielsen, P. & Rasmussen, F. (1975). Relationships between molecular structure and excretion of drugs, *Life Sciences* **17**, 1495–1512.

[6] Pichini, S., Altieri, I., Zuccaro, P. & Pacifici, R. (1996). Drug monitoring in nonconventional biological fluids and matrices, *Clinical Pharmacokinetics* **30**, 211–228.

[7] Levisky, J.A., Bowerman, D.L., Jenkins, W.W. & Karch, S.B. (2000). Drug deposition in adipose tissue and skin: evidence for alternative source of positive sweat patch test, *Forensic Science International* **110**, 35–46.

[8] Kidwell, D.A., Holland, J.C. & Athanaselis, S. (1998). Testing for drugs of abuse in saliva and sweat, *Journal of Chromatography. B, Biomedical Sciences and Applications* **713**, 111–135.

[9] Haeckel, R. & Hanecke, P. (1993). The application of saliva, sweat and tear fluid for diagnostic purposes, *Annales de Biologie Clinique* **51**, 903–910.

[10] Pichini, S., Navarro, M., Pacifici, R., Zuccaro, P., Ortuno, J., Farre, M., Roset, P.N., Segura, J. & de la Torre, R. (2003). Usefulness of sweat testing for the detection of MDMA after a single-dose administration, *Journal of Analytical Toxicology* **27**, 294–303.

[11] Kintz, P., Tracqui, A., Jamey, C. & Mangin, P. (1996). Detection of codeine and phenobarbital in sweat collected with a sweat patch, *Journal of Analytical Toxicology* **20**, 197–201.

[12] Burns, M. & Baselt, R.C. (1995). Monitoring drug use with a sweat patch: an experiment with cocaine, *Journal of Analytical Toxicology* **19**, 41–48.

[13] Inoue, T. & Seta, S. (1992). Analysis of drugs in unconventional samples, *Forensic Science Review* **4**, 90–102.

[14] Huestis, M.A., Oyler, J.M., Cone, E.J., Wstadik, A.T., Schoendorfer, D. & Joseph, R.E. (1999). Sweat testing for cocaine, codeine and metabolites by gas chromatography-mass spectrometry, *Journal of Chromatography. B, Biomedical Sciences and Applications* **733**, 247–264.

[15] Fox, R.H., Goldsmith, R., Hampton, L.F.G. & Lewis, H.E. (1964). The nature of the increase in sweating capacity produced by heat acclimatization, *Journal of Physiology* **171**, 368–376.

[16] Barlow, W.K. & Gibbons, W.A. (1981). Electrically Heated Sweat Collection Device and Method, U.S. Patent 4,266,556, May 12, 1981.

[17] Schoendorfer, D.W. (1995). *Energy-Assisted Transdermal Collection Patch for Accelerated Analyte Collection and Method of Use*, U.S. Patent 5,465,713, November 14, 1995.

[18] Balabanova, S., Schneider, E., Wepler, R., Hermann, B., Boscheck, H.J. & Scheitler, H. (1992). Die Bedeutung der Drogenbestimmung in Pilocarpinschweiss fiir den Nachweis eines zurlickliegendes Drogenkonsum, *Beiträge zur Gerichtlichen Medizin* **50**, 111–115.

[19] Mastella, G., Di Cesare, G., Borruso, A., Menin, L. & Zanolla, L. (2000). Reliability of sweat-testing by the Macroduct collection method combined with conductivity analysis in comparison with the classic Gibson and Cooke technique, *Acta Paediatrica* **89**, 933–937.

[20] Crouch, D.J., Cook, R.F., Trudeau, J.V., Dove, D.C., Robinson, J.J., Webster, H.L. & Fatah, A.A. (2001). The detection of drugs of abuse in liquid perspiration, *Journal of Analytical Toxicology* **25**(7), 625–627.

[21] Barben, J., Ammann, R.A., Metlagel, A., Schoeni, M.H., Swiss Paediatric Respiratory Research Group (2005). Conductivity determined by a new sweat analyzer compared with chloride concentrations for the diagnosis of cystic fibrosis. *The Journal of Pediatrics* **146**, 183–188.

[22] Phillips, M. (1980). An improved adhesive patch for long-term collection of sweat, *Biomaterials, Medical Devices, and Artificial Organs* **8**, 13–21.

[23] Aly, R., Shirley, C., Cunio, C. & Maibach, H.I. (1978). Effect of prolonged occlusion on the microbial flora, pH, carbon dioxide and transepidermal water loss on human skin, *The Journal of Investigative Dermatology* **71**, 378–381.

[24] Phillips, M. (1984). Sweat-patch testing detects inaccurate self-reports of alcohol consumption, *Alcoholism, Clinical and Experimental Research* **8**, 51–53.
[25] Fay, J., Fogerson, R., Schoeendofer, D., Niedbala, R.S. & Spiehler, V. (1996). Detection of methamfetamine in sweat by EIA and GC-MS, *Journal of Analytical Toxicology* **20**, 398–403.
[26] Fogerson, R., Schoendorfer, D., Fay, J. & Spiehler, V. (1997). Qualitative detection of opiates in sweat by EIA and GC-MS, *Journal of Analytical Toxicology* **21**(6), 451–458.
[27] Huestis, M.A., Cone, E.J., Wong, C.J., Umbricht, A. & Preston, K.L. (2000). Monitoring opiate use in substance abuse treatment patients with sweat and urine drug testing, *Journal of Analytical Toxicology* **24**, 509–521.
[28] Liberty, H.J., Johnson, B.D. & Fortner, N. (2004). Detecting cocaine use through sweat testing: multilevel modeling of sweat patch length-of-wear data, *Journal of Analytical Toxicology* **28**(8), 667–673.
[29] de la Torre, R. & Pichini, S. (2004). Usefulness of sweat testing for the detection of cannabis smoke, *Clinical Chemistry* **50**, 1961–1962.
[30] Kintz, P., Cirimele, V. & Ludes, B. (1998). Codeine testing in sweat and saliva with the Drugwipe, *International Journal of Legal Medicine* **111**, 82–84.
[31] Samyn, N. & Van Haeren, C. (2000). Onsite testing of saliva and sweat with Drugwipe and determination of concentrations of drugs of abuse in saliva, plasma and urine of suspected users, *International Journal of Legal Medicine* **113**, 150–154.
[32] Samyn, N., De Boeck, G. & Verstraete, A.G. (2002). The use of oral fluid and sweat wipes for the detection of drugs of abuse in drivers, *Journal of Forensic Sciences* **47**, 1380–1387.
[33] Caplan, Y.H. & Goldberger, B.A. (2001). Alternative specimens for workplace drug testing, *Journal of Analytical Toxicology* **25**, 396–399.
[34] Kidwell, D.A. & Smith, F.P. (2001). Susceptibility of PharmChek drugs of abuse patch to environmental contamination, *Forensic Science International* **116**, 89–106.
[35] Uemura, N., Nath, R.P., Harkey, M.R., Henderson, G.L., Mendelson, J. & Jones, R.T. (2004). Cocaine levels in sweat collection patches vary by location of patch placement and decline over time, *Journal of Analytical Toxicology* **28**, 253–259.
[36] Mura, P., Kintz, P., Papet, Y., Ruesch, G. & Piriou, A. (1999). Evaluation de six tests rapides pour le depistage du cannabis dans la sueur, la salive et les urines, *Acta Clinica Belgica. Supplementum* **1**, 35–38.
[37] Pacifici, R., Farre, M., Pichini, S., Ortuno, J., Roset, P.N., Zuccaro, P., Segura, J. & de la Torre, R. (2001). Sweat testing of MDMA with the drugwipe analytical device: a controlled study with two volunteers, *Journal of Analytical Toxicology* **25**(2), 144–146.
[38] Spiehler, V., Fay, J., Fogerson, R., Schoendorfer, D. & Niedbala, R.S. (1996). Enzyme immunoassay validation for qualitative detection of cocaine in sweat, *Clinical Chemistry* **42**(1), 34–38.
[39] Preston, K.L., Huestis, M.A., Wong, C.J., Umbricht, A., Goldberger, B.A. & Cone, E.J. (1999). Monitoring cocaine use in substance-abuse-treatment patients by sweat and urine testing, *Journal of Analytical Toxicology* **23**(5), 313–322.
[40] Moody, D.E. & Cheever, M.L. (2001). Evaluation of immunoassays for semiquantitative detection of cocaine and metabolites or heroin and metabolites in extracts of sweat patches, *Journal of Analytical Toxicology* **25**(3), 190–197.
[41] Moody, D.E., Spanbauer, A.C., Taccogno, J.L. & Smith, E.K. (2004). Comparative analysis of sweat patches for cocaine (and metabolites) by radioimmunoassay and gas chromatography-positive ion chemical ionization-mass spectrometry, *Journal of Analytical Toxicology* **28**(2), 86–93.
[42] Kintz, P., Cirimele, V. & Ludes, B. (2000). Detection of cannabis in oral fluid (saliva) and forehead wipes (sweat) from impaired drivers, *Journal of Analytical Toxicology* **24**(7), 557–561.
[43] Saito, T., Wtsadik, A., Scheidweiler, K.B., Fortner, N., Takeichi, S. & Huestis, M.A. (2004). Validated gas chromatographic-negative ion chemical ionization mass spectrometric method for delta(9)-tetrahydrocannabinol in sweat patches, *Clinical Chemistry* **50**(11), 2083–2090.
[44] Follador, M.J., Yonamine, M., de Moraes Moreau, R.L. & Silva, O.A. (2004). Detection of cocaine and cocaethylene in sweat by solid-phase microextraction and gas chromatography/mass spectrometry, *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences* **811**(1), 37–40.
[45] Bush, D.M. (2008). The U.S. mandatory guidelines for federal workplace drug testing programs: current status and future considerations, *Forensic Science International* **174**(2–3), 111–119 April 12, 2007 [Epub ahead of print].
[46] Moody, D.E. (2000). Units for sweat patch results, *Journal of Analytical Toxicology* **24**(8), 733.
[47] Kintz, P., Tracqui, A., Mangin, P. & Edel, Y. (1996). Sweat testing in opioid users with a sweat patch, *Journal of Analytical Toxicology* **20**, 393–397.
[48] Kintz, P., Tracqui, A., Marzullo, C., Darreye, A., Tremeau, F., Greth, P. & Ludes, B. (1998). Enantioselective analysis of methadone in sweat as monitored by liquid chromatography/ion spray-mass spectrometry, *Therapeutic Drug Monitoring* **20**(1), 35–40.
[49] Taylor, J.R., Watson, I.D., Tames, F.J. & Lowe, D. (1998). Detection of drug use in a methadone maintenance clinic: sweat patches versus urine testing, *Addiction* **93**(6), 847–853.
[50] Fucci, N., De Giovanni, N. & Scarlata, S. Sweat testing in addicts under methadone treatment: an Italian experience, (2007). *Forensic Science International* **174**(2–3), 107–110 [Epub ahead of print].
[51] Schwilke, E.W., Barnes, A.J., Kacinko, S.L., Cone, E.J., Moolchan, E.T. & Huestis, M.A. (2006). Opioid disposition in human sweat after controlled oral codeine administration, *Clinical Chemistry* **52**(8), 1539–1545.

[52] Kacinko, S.L., Barnes, A.J., Schwilke, E.W., Cone, E.J., Moolchan, E.T. & Huestis, M.A. (2005). Disposition of cocaine and its metabolites in human sweat after controlled cocaine administration, *Clinical Chemistry* **51**(11), 2085–2094.
[53] Levisky, J.A., Bowerman, D.L., Jenkins, W.W., Johnson, D.G., Levisky, J.S. & Karch, S.B. (2001). Comparison of urine to sweat patch test results in court ordered testing, *Forensic Science International* **122**(1), 65–68.
[54] Kidwell, D.A., Blanco, M.A. & Smith, F.P. (1997). Cocaine detection in a university population by hair analysis and skin swab testing, *Forensic Science International* **84**(1–3), 75–86.
[55] Winhusen, T.M., Somoza, C., Singal, B., Kim, S., Horn, P.S. & Rotrosen, J. (2003). Measuring outcome in cocaine clinical trials: a comparison of sweat patches with urine toxicology and participant self-report, *Addiction* **98**(3), 317–324.
[56] Walsh, J.M., de Gier, J.J., Christopherson, A.S. & Verstraete, A.G. (2004). Drugs and driving, *Traffic Injury Prevention* **5**(3), 241–253.
[57] Steinmeyer, S., Ohr, H., Maurer, H.J. & Moeller, M.R. (2001). Practical aspects of roadside tests for administrative traffic offences in Germany, *Forensic Science International* **121**(1–2), 33–36.

RAFAEL DE LA TORRE AND SIMONA PICHINI

Symptoms *see* Psychopathology: Terms and Trends

Syndromes: Psychological

In medicine, the term *syndrome* is used to describe a group of signs and symptoms that collectively characterize or indicate a particular disease or abnormal condition: the sum of signs associated with any pathological process. For example, cardio-auditory syndrome [1] consists of the combination of sensorineural deafness and prolongation of the part of the standard electrocardiogram known as the *QT interval*. While there is no obvious connection between these two seemingly unrelated symptoms, research has determined that their co-occurrence is associated with other systemic characteristics that suggest a common etiology. An inherited autosomal recessive condition, the cluster is a distinct biodynamic entity and not simply a coincidental co-occurrence. Knowledge of the syndrome, moreover, is important because it alerts a clinician detecting a hearing problem to rule out the possibility of an associated heart condition that, if left undetected and untreated, may be fatal. The concept of syndromes takes on additional significance in medical–legal conditions such as "shaken baby syndrome" [2].

In the behavioral sciences, although underlying relationships between disparate symptoms is certainly known to exist, their fundamental biodynamic processes are not yet fully understood [3]. Mental health professionals, therefore, seldom use the term *syndrome*, preferring instead "disorder", which is unequivocally descriptive yet implies no specific etiology. In medicine, the disorder first described by Tourette is understood well enough to be classified as a syndrome, yet the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*), consistent with its accepted nomenclature, is only beginning to discuss the possibility of changing the condition's name from its more conservative label of "Tourette's disorder" [4].

The use of the official and scientific-sounding term *syndrome* as a description of behavioral science concepts developed specifically for use in the courts has and continues to create widespread concern not only in the United States [5] but also worldwide [6]. This has created extensive debate as to whether syndrome evidence meets admissibility standards as well as social and policy questions related to the use of behavioral concepts in ways that may influence the relationships between gender and racial groups. These issues can be best considered in the context of the evolution of syndrome evidence, and the legal conventions that have emerged to deal with its abuse.

Syndromes and Behavioral Evidence

American trial law is unique in that it does not rely simply on whether a defendant did or did not do a particular deed, but looks beyond the obvious facts to motive, intent, and circumstance. State law typically considers mental state broadly as a question of fact to be answered by the jury at trial. Courts in some jurisdictions may therefore grant a *presumption* of sanity and requisite mental circumstances, but if a

defendant asserts a defense based on mental criteria, the burden of proof may shift to the prosecution to disprove the facts of the defense's assertion (*see* **Behavioral Science Evidence**).

All of this makes for subtle procedural distinctions, and it is therefore not surprising that jurors and also judges and lawyers can find the issues confusing. Such a misadventure occurred in the 1988 *Mulica* case [7]. No one disputed that John Mulica had made off with and squandered more than $200 000 of someone else's money. The problem confronting his defense was finding a way to keep John out of jail. The "compulsive gambler defense" had been tried before in other jurisdictions and by the late 1980s had been thoroughly discredited [8]. Posttraumatic stress disorder (PTSD) had only recently been defined and, amidst continuing debate, incorporated into the then-current *Diagnostic and Statistical Manual* [9] (*see* **Posttraumatic Stress Disorder**; **Malingering: Forensic Evaluations**).

The defense asserted PTSD, and the jury had little trouble sorting out the facts of the case, but everyone became confused when the jurors were about to start their deliberation and the judge explained the law. The jury could have accepted that John was a compulsive gambler, but that particular mental state would have gained the defendant nothing in the way of leniency. The judge then asked the jury to base its ruling on PTSD, but when the jury convicted and appeared to reject that argument, the defense contended that it should be able to assert a different theory.

The appellate courts agreed, holding that the prosecution is not relieved of its burden of disproving a defendant's mental incapacity simply because a defendant pursues a focused trial strategy centered on a diagnosis that the jury rejects. The decision was not intended to issue a fishing license for the murky waters of diminished responsibility, but to warn prosecutors to shut all of the doors to these "novel" defenses the first time around. Nevertheless, the prospect of a potential "mini" mental status defense – a defense that could get a defendant out from under a criminal charge without meeting all of the formal requirements and probable confinement reinstated as a part of the actual insanity defense – was too great a temptation for defense attorneys to ignore (*see* **Insanity: Defense**; **Temporary Insanity**).

There is little doubt that stress can change lives. As international traumatic stress expert Bessel van der Kolk said, summarizing a career of research, "overwhelming social experience can become indelibly etched in people's memories and set up a cascade of disturbances that can permanently alter their capacity to regulate their biological systems" [10]. Stress, however, is a fact of life. The same events can have very different impacts on different people, and people can and do recover from extraordinary experiences and never turn to these experiences as justifications for failing to observe reasonable standards of human conduct.

Perpetrators as Victims

For centuries, men killed and were set free by juries who applied the "Unwritten Law" that men are justified in taking out their anger against an unfaithful spouse or her lover. The fact that even more women acted on the same anger and benefited from unwritten social mores did little to change the fact that other women without a convenient excuse for murder endured desperately unhappy marriages. In the late 1970s, psychologist Lenore Walker was called in as an expert witness at several trials of women who had killed their husbands. To help explain the common threads of these cases, Walker came up with a theory of "learned helplessness". A husband may beat his wife for overcooking an egg, or for undercooking it; for turning on the TV or for turning it off; for talking or for keeping still. Eventually, she finds that there is no connection between what she does and what happens to her. Why go on trying?

Learned helplessness is, of course, not unique to women; but it did seem to fit the stories reported by many women who had murdered their mates. Walker's book, *The Battered Woman* [11], struck a resonant cord in the feminist movement, and became well known to lawyers representing women charged with the murder of their mates.

For seven years, Ernest Kelly drank and beat his wife Gladys. She would move out, he would promise to change, and she would return to him. Their rows were ongoing, but usually took place out of public view; but on May 24, 1980, as the typically drunk Ernie and his bride of seven years walked down the street, Gladys stabbed him to death. According to prosecution testimony, Gladys started a scuffle. Quickly separated and restrained by passersby, Gladys shouted that she would kill her

husband. Breaking away, she chased after him, pulled a pair of scissors from her pocketbook, and took his life.

The defense tried unsuccessfully to introduce expert testimony on battered woman's syndrome, but the Supreme Court of New Jersey, in an opinion that draws extensively from Walker's work, overturned the conviction, holding that the research on battered woman's syndrome was "sufficiently reliable" to meet the standards for "scientific testimony" [12].

Battered woman's syndrome, and psychological syndromes in general, would remain a subject of controversy in the criminal law for the remainder of the decade [13, 14].

Human behavior in the context of the law, however, is not only subject to all of the vagaries inherent in mental health, but is compounded by the obvious vested interests of parties to litigation in which not only vast sums of money but freedom and life itself are often at stake. To what extent is it appropriate to lend the imprimatur of "science" to opinions about human behavior and motivation in the context of a criminal trial? Was the judge at Gladys Kelly's trial correct in believing that such testimony was little more than an attempt to "explain and justify" as opposed to sharing scientific findings beyond the experience of the jury?

There can be no clear, bright line between the point at which behavioral science ends and advocacy begins. Inherent in American justice is an acceptance that justice requires consideration of both actions and intent. In a secular culture, our source of guidance in that gray zone, which forms the buffer between worthy and unworthy intent, has fallen, by default, to the behavioral professionals [15]. However, responsibility for acting appropriately on that guidance, like the responsibility for distinguishing good from evil, is one which cannot be delegated [16].

While it is the jury that must define justice, it is the courts that must define the rules of evidence. This was the duty confronted by the Supreme Court of Ohio in *State v Koss*, [17] and either course open to that court brought with it a potential for hindering the administration of justice. At the heart of *Koss*, however, was not a simple act of murder but the seemingly inexplicable behavior of the woman accused of the crime. Where Gladys Kelly simply stabbed her husband, Brenda Koss spent the next day following a routine that seemed inconsistent with her alleged conduct. In the years following *Kelly*, the battered woman's movement had grown increasingly sophisticated [18].

In the process, social psychologist Angela Browne made an observation that the *Koss* court found directly on point. "A history of physical abuse alone does not justify the killing of the abuser," she wrote. "Having been physically assaulted by the abuser in the past is pertinent to such cases only as it contributes to the defendant's state of mind at the time the killing occurred" [19].

The significance of the "syndrome", thus suggested, was *not* to offer a legal excuse for murder. If in fact expert testimony could help the jury make sense of Brenda Koss's unusual behavior in the hours after the crime, then there should be no reason why the jury need be denied its benefit. Obviously, the defense attorney could argue the theory to the jury, but lawyers are not behavioral scientists and are not generally subject to cross-examination. Let the jury hear the theory, the Supreme Court of Ohio concluded in its landmark *Koss* decision that proved a turning point for judicial acceptance of syndrome testimony nationally – the jury will be free to make of it what they may.

Once opened, however, the door to mental health "exotica" proved difficult to monitor. Within a year, the courts were flooded with expert witnesses of dubious distinction hawking plausible theories as implausible defenses. A new definition of syndrome had been created specifically for trial law: A syndrome (legal variety) = (i) socially unacceptable behavior + (ii) a sympathetic antecedent combined with (iii) a behavioral dynamic explanation [20].

In 1985, Lisa Becker Grimshaw arranged for two male friends to waylay her husband in a wooded area where he was bludgeoned to death with baseball bats. She used the battered woman defense and was convicted of the lesser charge of manslaughter. By the early 1990s, however, when the case reached appeal, even the highest courts in the nation were growing impatient with "expert" witnesses, and the Supreme Judicial Court of Massachusetts concluded that the prosecutor's use of the term *hired gun* to describe Lenore Walker, while "to be disapproved of", was not a reversible error in this case since the characterization did not create a "substantial risk of miscarriage of justice" [21].

Syndromes became *the* behavioral science tool of the defense. The conduct that most frequently placed defendants in need of legal refuge, not surprisingly,

centered on primal forces of sex, drugs, greed, and ambition. In the late 1980s, with the arrival of widespread cocaine abuse and particularly crack cocaine, also known as *freebase* or *rock cocaine*, the justice system and society itself was confronted by a tidal wave of crime and health problems. Cocaine abuse by pregnant women ranged from 1% to over 10% in urban areas, with resulting birth defects that will affect the health and functioning of their children throughout their lives.

Many drug abusers have preexisting psychiatric conditions. Others develop symptoms as the consequences of abuse begin to take their toll. By the early 1990s, courts were faced with the potential of countless defendants claiming lack of responsibility for the consequences of events, which the defendants themselves had set in motion. *Commonwealth v. Herd* [22] was a turning point in the resolution of that issue. Reginald Herd had been using cocaine when he beat his girlfriend's three-year-old son to death with 277 blows of an electric cord. He pleaded lack of criminal responsibility, but a noteworthy appellate decision provided the first clear guidelines on the role of drugs in diminished capacity. While *Herd* does not preclude a defense of legal insanity by drug abusers, it does preclude a diminished responsibility defense for symptoms produced by drug abuse, even by a person who begins or continues drug use knowing that it will or may arouse symptoms of a preexisting mental disease or defect.

By the late 1980s, PTSD, first officially created in large part not only in response to patient demands for an officially recognized diagnosis that would fit their symptoms but also in response to shifting attitudes to Vietnam veterans, had become in the words of many mental health and legal professionals the "black hole" of litigation [23]. The diagnosis remains controversial, and many experts would prefer to see it eliminated or at least replaced with something more objective [24–27]. Everyone does agree, however, that no other diagnosis in American psychiatry has had as profound an influence on civil and criminal law.

Charles Delaney and his estranged wife were in the midst of an ongoing disagreement over child custody when he strangled her with a garrote. The grave that he had dug two days earlier hardly suggested that the crime was spontaneous and unpremeditated. Yet like many other veterans, when charged with his crime, Delaney blamed his conduct on the trauma of military service and arrived at trial prepared to impress the jury with war stories.

Just as physicians make it a practice to be among the first to read each month's issue of the *Reader's Digest* medical column to be prepared for the expected symptoms of the month, psychiatrists had learned to follow the current films. Movies like *The Deer Hunter* and *The Burning Bed* – in themselves valuable contributions to a nation's awareness of the unseen scars of trauma and abuse – made excellent guidebooks for anyone searching for excuses to avoid criminal conviction [28]. Military veterans themselves have been among the most active and successful at routing out false claims of military stress "syndrome", which they rightly consider an affront to every serviceman who has returned to civilian life and managed to avoid taking out their frustrations on innocent bystanders.

Delaney claimed, among other defenses, that he had recovered bodies from the October 1983 bombing of the Beirut Marine Barracks. His commanding officer, however, had no knowledge of such an experience, and recalled that the tank mechanic had spent most of his Lebanon tour in a hotel and had never been exposed to combat or even hostile fire. A psychiatrist who examined Delaney prior to trial noted that Delaney brought attention to a twitch, similar to that seen in movies about traumatized veterans, which Delaney affected on initial interview but that vanished when the defendant was alone. Suspecting that he was being led to a self-serving diagnosis, the psychiatrist asked Delaney if he had the nightmare that occurred every night in PTSD victims. Delaney said that he did, but was unable to describe their content and observers were unable to note any indications of actual sleep disturbance.

Delaney was appealed all the way to the highest court in the state, and although each review devoted careful attention to assuring that all of his due process rights had been fully protected, Charles Delaney's murder conviction was upheld [29].

By the time Norma Roman came up with her multiple personality disorder (MPD) defense in the early 1990s, arguing that while one of her personalities may have been caught dealing drugs but her "core" personality was law abiding, the diagnosis itself had begun to wear thin. The courts found that at least one of the "personalities" had criminal intent, and upheld her conviction [30].

No one had ever heard of "multiple personalities" until the 1950s when Chris Sizemore presented with the symptoms that became the basis for the celebrated book and 1957 Academy Award *tour de force* movie *The Three Faces of Eve* starring Joanne Woodward as a woman with separate identities and personalities. Patients, well coached in their performance by *Eve* and its imitators, soon learned to oblige their doctors with similar manifestations, and suddenly America had an epidemic of "split personalities". It was not until the 1980s that clinicians responsible for treating patients at public institutions began to compare notes and discovered that none of them had ever seen a case of MPD. The condition, it seems, had been restricted largely to private patients who had the time, money, and imagination to keep up with the popular arts (*see* **Deception: Truth Serum** for a discussion of the condition that succeeded MPD).

Chris Sizemore, in the meantime, had gone on to become a celebrity and dynamic lecturer. MPD enjoyed a brief vogue as a criminal defense, but the lawyers soon figured out that whichever of their client's personality had done the dastardly deed, they would all hang by the same neck and therefore that particular courtroom charade really was not worth the effort [31].

In a sense, the concept of multiple personalities actually had some validity, and anticipated, by decades, the discovery of the dispersed operation of the human mind [32]. It was the notion that each of the dispersed components would have their own names, wardrobes, and identities that was pure fabrication; or at best, an analogized externalization of an inventive but marginally integrated personality. Chris Sizemore became as active as anyone in correcting the false impression that "multiple personalities" could exculpate a criminal defendant.

Only rarely has a court accorded any significance to a criminal defense plea based on multiple personalities, which makes the 1993 decision in *United States v. Denny-Shaffer* [33] worthy of note. A New Mexico delivery nurse, 37-year-old Bridget Denny-Shaffer, had a history of depression, abusive relationships, and two failed marriages when her pregnancy by a boyfriend ended in miscarriage. Wearing the uniform and false credentials of a medical student, she later walked into the nursery of another hospital, examined several of the babies, selected one of the infants, and walked away with the child undetected. Arriving at the home of her boyfriend in visible (yet false) pregnancy, she spent the night and the next day when the boyfriend returned home from work Bridget greeted him from a blood soaked bed, held out the missing infant, and said "This is your little one."

Thrown out by the boyfriend, Denny-Shaffer and the infant began a cross-country odyssey that ended in an Federal Bureau of Investigation (FBI) stop and a federal conviction for interstate transport of a child unlawfully kidnapped. Her counsel had chosen to assert a multiple personality defense that crumbled in the realization that the diagnosis was being discontinued and in any case one that could not protect her under the post-Hinckley federal insanity guidelines. A sad and sympathetic figure by any definition, her case was taken up by the US Court of Appeals that reversed the conviction and sent the case back for retrial. The decision, however, had little to do with psychiatry and a great deal to do with justice.

Even in early England, with its formalistic legal codes that offered prosecutors little discretionary authority, justice often prevailed despite the law. At a time when the "gentlemen" of the jury were truly gentlemen in the sense of rank and privilege, the early English courts would often look the other way and allow a jury to hand down decisions contrary to law [34]. Known as *jury nullification* in the legal vernacular, America's reconceptualization of justice administration made the practice largely unnecessary. In fact, America's obsession with fairness has led it from justice by judicial discretion being denounced as too prejudicial to this group or that, followed by its replacement with mandatory sentencing guidelines, which are then denounced as too inflexible to serve the interests of justice – a cyclical gyration with often paradoxical consequences [35].

In the case of Denny-Shaffer, the Court of Appeals concluded that a jury could not do any worse by Bridget than what she had already endured, and asked the trial court to reconsider the "jury's right to determine credibility, to weigh evidence, and to draw justifiable inferences". In short, with all due respect to the insanity guidelines, perhaps this was one best left to the common sense of a jury.

No one has yet tried to palm off a bogus syndrome defense under *Denny-Shaffer*, but the case has come to represent a small milestone for juror discretion. Perhaps this reflects, in part, a measure of faith in the axiom that anyone who is able to consider a mitigating defense before committing a crime is unlikely to see that defense accepted by an actual

jury. Less visible to the public, however, was an equally insidious potential for behavioral forensic abuse by the prosecution.

Effect as Proof of Cause

Generally, we think of legal "syndromes" as a tool of the defense. In reality, they have been used to support anything and everything that may be of help to either side.

In *State v. Allewalt* [36], prosecutors sought to convict a 17-year-old of the rape of his girl friend's mother despite testimony by both the defendant and the girlfriend that the mother had initiated the encounter. In this case, the prosecution reached out for expert behavioral syndrome testimony; not simply in an effort to bolster the credibility of and sympathy for the mother, but to use that credibility and sympathy to convict the defendant. As with so many legal "syndromes", the scientific credibility of the testimony drew upon PTSD. Although there have been endless court battles over the admissibility of testimony on everything from "battered spouse syndrome" (**Battered Spouse Syndrome**) and "Vietnam survivor syndrome" to "rape trauma syndrome" [37] and "child sexual abuse accommodation syndrome" (**Child Sexual Abuse Accommodation**) [38], all draw their scientific credibility not from any legal "syndrome" but from the *Diagnostic and Statistical Manual of Mental Disorder*'s inclusion of PTSD as a recognized diagnostic entity.

The PTSD diagnosis, however, was never intended to support litigation. PTSD was included in the *DSM* upon the insistence of *clinicians* who had been confronted by people with symptoms of often-unknown etiology. Often these individuals had difficulty being taken seriously because family, friends, and officials doubted whether they had experienced anything at all which would justify all the fuss. People know, from their real-life experiences, that secondary gain is a powerful motivation. The world is filled with unhappy, vindictive, or envious folks who want recognition, support, and an explanation for their emotions. While this is by no means the explanation of all or even most requests for emotional help, much of the work of behavioral therapy is an investment of human effort in search of a meaning for the events of daily life.

Every day, people have disputes in which it is impossible to sort out the true facts and establish who is right and who is wrong. That is why we have courts. Courts settle disputes by considering all of the evidence and taking a side – this is how we *define* justice.

Clinicians have traditionally been uninterested in objective truth or official blame. Clinicians confront people with problems, and it is their job to address those problems. If a person seeks help for an emotional issue, it really does not matter *how* that issue arose. Since everybody faces disputes every day, and most of these disputes are resolved in one way or the other, everyone by that definition is wrong half of the time. At best, most people learn to compromise and accommodate; nevertheless, most still have some positions on which they are reluctant to go against what they consider right. Others are unhappy because of an entire host of problems that most observers would consider to be of their own making. Others always try to do the right thing yet have exceptionally bad luck. Still others are unhappy because they got what they thought they wanted but are having difficulty living with the price.

Psychotherapists recognized that quibbling over blame would not help – that prolonging dispute, by itself, would just make matters worse. Therefore, they learned to be *nonjudgmental*: to accept what the patient said and move on from there. Some therapists even took a further step and tried to teach the world to take people at their word. Others, quite correctly, pointed out that regardless of fault, stress can cause people to do strange things. Blame is often just a further complication. In a great many cases, however, people can and do have mixed feelings, and these conflicting emotions contribute to distress as much if not more than being a completely blameless victim of uncontrollable forces.

PTSD, as a *DSM* diagnosis, was ultimately accepted simply as a name for powerful emotions, and a diagnostic label through which these emotions can be explored in the context of an environmental explanation. It has four criteria. Three of these relate to emotions: persistent reexperience; the avoiding of people, places, and feelings; and increased arousal. The fourth, placed first on the list, is a past confrontation with a threat that generated intense fear, helplessness, or horror. The diagnosis *assumes* that the environmental trigger is the *cause* of the emotions because the person in treatment *reports* reexperiencing and avoiding things that are related to that experience.

It is significant to consider that the association of experiences and powerful emotions predates the inclusion of this diagnosis in the psychiatric lexicon. In fact, recognition of the association predates Freud and the birth of modern psychodynamic insights. Prior to Freud, professionals who dealt with emotional issues were quick to assume very simplistic relationships between objective events and subsequent emotions – what behaviorists termed *stimulus and response*. Even Freud's early work is filled with just such assumptions. However, as Freud and those who followed progressed in their research, they came to the realization that few causal links are all that simple. The basic defense mechanisms are, in fact, explanations as to just why simple assumptions are generally wrong.

Yet such discredited assumptions are implicit in the acceptance of PTSD as a diagnostic entity. To make the diagnosis of PTSD is to say that there *was* a traumatic event and that the event *caused* the response. This is why PTSD encountered so much resistance in the professional community, and why it took so long for it to be included in the *DSM*. In the end, it was included because the clinicians mustered greater political power than the researchers and scholars. The clinicians wanted to be able to meet a market, and activists who wanted to be able to assert that identifiable life problems were the obvious source of common emotional conditions were a major force in developing the *DSM*.

The original logic behind the legal acceptance of behavioral "syndromes" was grounded in solid psychodynamic theory. People who have been through traumas do inexplicable things. They may be slow to report victimization, fail to take obvious defensive measures, and even maintain a seemingly pleasant relationship with their victimizer. People, who must nevertheless unravel the truth of criminal assertions, including juries, need to understand this. The original purpose of expert testimony was to provide that knowledge.

By testifying that a person has PTSD, however, an expert *by definition* appears to be testifying to the *truth* of the alleged charge. That expert, moreover, lends full weight and authority not only of their personal presence and qualifications but also of all the mental health professions as well. This conclusion is hard to avoid in any such testimony, and is further complicated in *Allewalt* by the psychiatrist's *explicit* statement of the causal conclusion.

In a case like *Allewalt*, there is no way of ever knowing whose story is correct. Both accounts may be completely accurate from the point of view of the person recounting the events. Alternatively, for all we know, the psychiatrist may have gotten it right [39]. At best, Bill Allewalt cannot be commended for his judgment or restraint. Nevertheless, there is another way to look at the facts. Clearly, the deed was done, and clearly, the mom was very unhappy afterwards. Many of her symptoms, however, were consistent with behavior she had been exhibiting for at least three months before. Her testimony also includes references to the fact that following her encounter with young Allewalt she wondered if neighbors, friends, and acquaintances thought she was "cheap", or if she was "dressing cheap", or if they were looking at her out of the windows when she came out of her home. She was no longer comfortable around young men, and was withdrawing into herself.

She was a woman, married for 16 years, and dealing with the reality of living alone. She had been thrown into daily close contact with a healthy young man who was living with her daughter. It does not take a psychiatrist to come up with dozens of alternative scenarios; each more or less favorable to one side or the other. Most jurors could come up with an even longer list. This is the point in fact, and why it is so helpful to be able to call upon a credible "expert" to cast one's own preferred spin to a jury.

Allewalt is still law in Maryland (although its reasoning has been rejected elsewhere) [40], and was cited as precedent in one case that made it all the way to the US Supreme Court [41]. However, one lone voice on the seven man court which decided *Allewalt* had an interesting suggestion. "I agree," the Judge said, "that testimony explaining posttraumatic stress disorder (PTSD) is admissible ... [but] I do not agree that opinion testimony should be received on the question of whether the complainant actually suffered PTSD" [42]. The remainder of the court dismissed the observation as irrelevant to the question placed before it, but the following year the state of Washington took a broader view:

> The courts which have admitted rape trauma syndrome testimony believe it sufficient that the myriad of symptoms encompassed therein are 'generally accepted to be a common reaction to sexual assault.' We find, however, that this is not the relevant question. The issue is not whether rape victims may display certain symptoms; the issue is whether the

> presence of various symptoms, denominated together as 'rape trauma syndrome', is a scientifically reliable method admissible in evidence and probative of the issue of whether an alleged victim was raped. The literature on the subject demonstrates that it is not [43].

Perhaps the most serious problem with syndromes is that in their propensity to depict victims, they create indelible images of alleged perpetrators. This is illustrated vividly in the *Banks* case in which an offender profile was misused to convict Sgt. Russell Banks (*see* **Profiles: Psychological and Behavioral**). The problem came to world attention through the "recovered memory" cases in which dissociative disorders were cited as proof not only that a crime had taken place but also to identify and convict perpetrators who were simply inferred from the nature of the "syndrome" (*see* **Deception: Truth Serum**). Any syndrome or behavioral concept that is advanced to not only describe an effect but also establish a cause must be viewed with special concern.

Evidentiary Standards

Behavioral dynamics are integral to human conduct, and in jurisdictions including the United States they are inextricably linked to the tasks confronting the triers of fact (*see* **Behavioral Science Evidence**). In recent years, judges have become increasingly familiar with behavioral science evidence and expert testimony [44], although their understanding and application of the discipline continues to be questioned [45].

In any trial in which behavioral dynamics have a bearing upon the questions that must be addressed by the trier of fact, the court cannot assert that the topic is irrelevant. If behavioral science evidence or expert testimony is proffered, its admission can only be excluded to the extent that it is cloaked in any inappropriate imprimatur of science or proposes to address issues that are beyond the appropriate scope of the relevant issues of fact. Usually, the question becomes one of credibility rather than admissibility. The court, and opposing parties, must nevertheless assure that scientific standards are upheld, and that the evidence is kept within appropriate boundaries.

Research on actual judicial practice confirms that the courts are admitting behavioral science evidence, and that the unique properties of that evidence are not in themselves viewed as grounds for their prohibition under *Daubert* or other emerging evidentiary standards [46].

Syndromes and Public Policy

In medical malpractice cases, for example, the facts in dispute are generally of sufficient complexity and beyond the daily experience of the court that the law may *require* that expert testimony be introduced to explain the issues and to help frame the questions that the jury must answer. In this unique role, the expert is allowed to speak from general experience and knowledge, and to apply that expertise to facts in dispute even though the expert may have little or no direct knowledge of the facts in the specific case. Between direct testimony and in cross-examination by opposing counsel, it is assumed that the jury will have the opportunity to weigh the expert's opinion and the extent to which that opinion should be relied upon in deciding the case.

In legal cases that involve human behavior, however, it is more difficult to define the appropriate role of the expert. While most nonphysician jurors can easily see how an expert can help them understand a medical issue, most human behavior is not considered beyond the expertise of the general public. People deal with behavior, their own and that of the people around them, every day. Knowledge of the appropriate standards for human behavior – the distinction between right and wrong – is something everyone is taught and then tests against the realities of life from the day they are born.

To understand that abused women may chose to freely remain with their abuser rather than take steps to change their situation is not the same as concluding that this common behavior justifies the subsequent murder of the abuser. To cite the statistics that young black men frequently kill other human beings does make it any more likely that a jury will then conclude that this justifies a dispensation for future black murders – even if you dress up the argument as "Black Rage Syndrome".

The core of the scientific evidence concerning battered women, and black offenders, and virtually every other identified special population which comes before the courts, is that people can and do assess their options; and that often people see their options

as far more limited than they may well be in actuality. This is the "learned helplessness", which Lenore Walker saw in the battered women who went on to kill their abusers, and a similar inability to see more reasonable solutions to common problems has influenced both the characterization of defendant populations and the rationale for sharing this insight with the triers of fact who must determine their fate.

Radical activists, on the other hand, responding to their own agendas, have extended this logic to conclude that all injustice could be eliminated if only the courts would always take their side in any dispute. When there are no witnesses, and we come down to the word of one person against another, we always assume that the woman is right, or the racial minority is right, or that whoever the special interest group represents must be and always will be right.

One problem with this advocacy logic is that it does not stop at simply accepting scientific observations noted in some members of a given population, but elevates these observations to the stature of an inevitable truism. This has the paradoxical consequence of turning members of minority groups into apologists for the very negative stereotypes that most members of that minority group abhor.

Women's legal syndromes, therefore, have come to paint women as indecisive clinging predators who mean no when they say yes, who cannot be trusted to make up their own mind, and who are unable to take responsibility for their own actions and decisions [47]. It has taken decades for lawyers to correct stereotypes created by Lenore Walker and her "battered woman" syndrome; laying a foundation under law for women to save their marriages or have them end in something other than murder or perpetual entrapment [48, 49].

Minority syndromes have similarly painted blacks as violent and impulsive [50, 51], Hispanics as playing by their own rules [52], white men as unable to tolerate the normal vicissitudes of daily life [53], and most foreigners as unable to live in our land of opportunity without welfare and our toleration of their behaviors, which we would consider unacceptable in our own offspring [54].

The American judicial system does not condone and will actively oppose any attempt to use the courts as a forum in which to propagate stereotypes about the relative worth of one group as opposed to another. The justice system, however, does share with all of the responsible institutions in our society a stake in providing citizens with a maximum opportunity to fully understand their options and obligations, and in helping people to make informed choices with which they are both willing and able to abide. This is a desire, however, which the justice system must always balance against its unique duty to preserve, for each responsible American, the ability to make free choices and to live with the fruits of those decisions, both good and bad.

If a dispute involves issues that are truly beyond the experience of a jury, for example the conduct of a person held hostage for a prolonged period, the jury is much more likely to defer to professional expertise. Furthermore, if expert testimony is reasonably consistent with the natural instincts of the jury, they are more likely to accord weight to the testimony and perhaps modify their judgment accordingly. A case in point is the trail of newspaper heiress Patricia Hearst for her part in the April 15, 1974 armed robbery of the Hibernia Bank in San Francisco. Hearst had been taken captive by the radical Symbionese Liberation Army, but later joined her captors as a gun-wielding member of their gang. At trial, experts testified that such behavior was not uncommon among hostages who come to depend upon and identify with their oppressors (*see* **Stockholm Syndrome**).

Captive or "Stockholm" syndrome testimony was novel at the time [55], and led to a major battle between experts as to its applicability. In the end, however, the jury convicted Hearst; not necessarily because they failed to believe the experts who testified that allying with captors was a fairly common phenomenon, or even because they believed that being a hostage was sufficiently within their own experience so that behavioral expertise was unnecessary, but because it was their judgment that regardless of how common the behavior may be it should not excuse or justify armed robbery.

The courts have grown increasingly wary of the use of any legal "syndrome", ascribed to a complainant in a criminal trial, to support an inference that the accused is, in fact, guilty of the crime [56]. On the other hand, courts continue to recognize that expert testimony as to how victims of the *type* of crime alleged by the prosecution typically conduct themselves can assist the trier of fact when such testimony does *not* carry with it any opinion as to whether the complainant is telling the truth about

either the crime itself or an alleged offender [57]. In addition, expert testimony is generally held to be admissible when it is used to refute suggestions that a prosecution witness may be unreliable due to behaviors that can be better appreciated by the trier of fact when placed in context by expert testimony [58].

Use of complainant syndromes, however, is a two-edged sword. Not only are prosecution experts open to attack under cross-examination and their testimony potentially turned to the benefit of the defense but the defense can also introduce their own experts who may characterize the complainant to the advantage of the defendant. Georgia, for example, experienced a spate of cases in which accused child molesters sought to introduce expert testimony on the "Lying child syndrome" in hope of discrediting complainants [59]. The "syndrome" detailed the propensity of a child to relate and to repeat untruthful statements about a person who is an authority figure in their life in order to manipulate that child's environment to advantage. The courts, which characterized the phenomenon as not "unique as a mysterious area of human response", said, thanks but no thanks – we know about that one and if the jury doesn't, it can figure it out for itself.

Conclusions

After a decade of debate and publicity, the legal status of the most advanced behavioral science contributions to the criminal justice process remain wedded to fundamental concepts that would be easily recognized by the twelfth century jurist Henry de Bracton were he with us today. In order for syndrome evidence to be admissible, the physical, emotional, or mental condition addressed by the evidence must be presented in a manner that meets the scientific requirements for legal evidence [60]. Expert testimony concerning a trait of an accused may only be used as evidence that the accused possesses such a trait. It must be left to the jury to determine whether and how such a trait may influence its view of the facts of the case [61]. Although the term *syndrome* may appear to give behavioral evidence an aura of scientific respectability, such labels themselves do nothing to enhance the stature of the substantive underlying observations. In fact, if anything, good science and credible observation are more readily accepted without them [62].

References

[1] Buyse, M.L. (ed) (1990). *Birth Defects Encyclopedia*, Blackwell Scientific Publications, Cambridge, pp. 281–282.

[2] Matter of Andre E. (E.E.). (1996). *New York Law Journal*, **215**, 30.

[3] Kulynych, J. (1996). Brain, mind, and criminal behavior, *Jurimetrics* **36**, 235–244.

[4] American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition, American Psychiatric Association, Washington, DC, pp. 101–103. (DSM 307.23).

[5] Brodin, M.S. (2005). Behavioral science evidence in the age of daubert: reflections of a skeptic, *University of Cincinnati Law Review* **73**, 867–943.

[6] Sutherland, E.E. (2006). Undue deference to experts syndrome? *Indiana International and Comparative Law Review* **16**, 375–421.

[7] Commonwealth v. Mulica, 520 N.E.2d 134 (Mass. 1988).

[8] United States v. Shorter, 809 F.2d 54 (D.C. Cir., 1987); For a review of the law related to behavioral defense and gambling crimes see Buchhandler-Raphael, M. (2008). The relationship between pathological gambling disorder and substantive criminal law and sentencing, *Gaming Law Review & Economics* **12**, 109–125.

[9] American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition, American Psychiatric Association, Washington, DC, pp. 424–429. (DSM 309.81).

[10] van der Kolk, B.A., McFarlane, A.C. & Weisaeth, L. (eds) (1996). *Traumatic Stress*, Guilford Press, New York, p. 560.

[11] Walker, L.E. (1979). *The Battered Woman*, Harper & Row, New York.

[12] State v. Kelly, 478 A.2d 364 (N.J. 1984).

[13] Wallace, D. (1985). The syndrome syndrome, *University of Florida Law Review* **37**, 1035–1058.

[14] McCord, D. (1987). Syndromes, profiles and other mental exotica, *Oregon Law Review* **66**, 19–108.

[15] For a discussion of the role of behavioral science in law, and possible consequences, see Piar, D.F. (2008). A welfare state of civil rights: the triumph of the therapeutic in American constitutional law, *The William and Mary Bill of Rights Journal* **16**, 649–684.

[16] The perpetrator as victim defense became widely know as the "abuse excuse". Dershowitz, A.M. (2000). *The Abuse Excuse*, Little, Brown, Boston.

[17] State v. Koss, 551 N.E.2d 970 (Ohio, 1990).

[18] Jones, A. (1996). *Women Who Kill*, Beacon Press, Boston.

[19] Browne, A. (1987). *When Battered Women Kill*, Free Press, New York, p. 175.

[20] Edwards, C.N. (1998). Behavior and the law reconsidered: psychological syndromes and profiles, *Journal of Forensic Sciences* **43**(1), 141–150.

[21] Commonwealth v. Grimshaw, 590 N.E.2d 681 (Mass. 1992).
[22] Commonwealth v. Herd, 604 N.E.2d 1294 (Mass. 1992).
[23] Simon, R.I. (ed) (1995). *Posttraumatic Stress Disorder in Litigation*, American Psychiatric Press, Washington, DC.
[24] Sparr, L.F. (1995). Post-traumatic stress disorder, *Neurologic Clinics* **13**, 413–429.
[25] Slovenko, R. (1994). Legal aspects of post traumatic stress disorder, *Psychiatric Clinics of North America* **17**, 439–446.
[26] Stone, A.A. (1993). Post-traumatic stress disorder and the law, *The Bulletin of the American Academy of Psychiatry and the Law* **21**, 23–36.
[27] Orr, S.P. & Pitman, R.K. (1993). Psychophysiologic assessment of attempts to simulate posttraumatic stress disorder, *Biological Psychiatry* **33**, 127–129.
[28] Rosenberg, J.E. & Resnick, P.J. (1995). The detection of malingered post-traumatic stress disorder, *Proceedings of the American Academy of Forensic Sciences*, Seattle, February 13–18.
[29] Commonwealth v. Delaney, 616 N.E.2d 111 (Mass. App. Ct. 1993); affirmed, Commonwealth v. Delaney, 639 N.E.2d 710 (Mass. 1994).
[30] Commonwealth v. Roman, 606 N.E.2d 1333 (Mass. 1993).
[31] Slovenko, R. (1995). Multiple personality, *Medicine and Law* **14**, 623–629.
[32] Gazzaniga, M.S. (1992). *Nature's Mind*, Basic Books, New York.
[33] United States v. Denny-Shaffer, 2 F.3d 999 (10th Cir., 1993).
[34] Green, T.A. (1985). *Verdict According to Conscience*, University of Chicago Press, Chicago.
[35] Toobin, J. (1994). The man who kept going free, *The New Yorker*, March 7; 38–53.
[36] State v. Allewalt, 517 A.2d 741 (Md. 1986).
[37] Frazier, P.A. & Borgida, E. (1992). Rape trauma syndrome, *Law and Human Behavior* **16**, 293–311.
[38] State v. W.L., 650 A.2d 1035 (N.J. Super 1995).
[39] For an argument that behavioral evidence should be admitted so as to weigh in favor of defendants or seeming victims, see Findley, K.A. (2008). Innocents at risk: adversary imbalance, forensic science, and the search for truth, *Seton Hall Law Review* **38**, 893–973.
[40] State v. Alberico, 861 P.2d 219 (N.M. App. 1991).
[41] Maryland v. Craig, 497 U.S. 836 (1990).
[42] McAuliffe, J., concurring, 517 A.2d 741, 759 (Md., 1986).
[43] State v. Black, 745 P.2d 12, 17–18 (Wash. 1987).
[44] Dahir, V.B., Richardson, J.T., Ginsburg, G.P., *et al.* (2005). Judicial application of daubert to psychological syndrome and profile evidence: a research note, *Psychology, Public Policy, and Law* **11**, 62–82.
[45] Gatowski, S.I., Dobbin, S.A., Richardson, J.T., *et al.* (2001). Asking the gatekeepers: a national survey of judges on judging expert evidence in a post-daubert world, *Law and Human Behavior* **25**, 433–458.
[46] Fradella, H.F., Fogarty, A. & O'Neill, L. (2003). The impact of daubert on the admissibility of behavioral science testimony, *Pepperdine Law Review* **30**, 403–444.
[47] Coughlin, A.M. (1994). Excusing women, *California Law Review* **82**, 1–93.
[48] Dutton, M.A. (1993). Understanding women's response to domestic violence: a redefinition of battered woman syndrome, *Hofstra Law Review* **21**, 1191–1242.
[49] Goldfarb, S.F. (2008). Reconceiving civil protection order for domestic violence: can law help end abuse without ending the relationship? *Cardozo Law Review* **29**, 1487–1550.
[50] Copp, K.M. (1995). Black rage, *The John Marshall Law Review* **29**, 205–238.
[51] Falk, P.J. (1996). Novel theories of criminal defense based upon the toxicity of the social environment, *North Carolina Law Review* **74**, 731–811.
[52] Maguigan, H. (1995). Cultural evidence and male violence, *New York University Law Review* **70**, 36–99.
[53] Tesner, M.A. (1991). Racial paranoia as a defense to crimes of violence, *Boston College Third World Law Journal* **11**, 307–333.
[54] Coleman, D.L. (1996). Individualizing justice through multiculturalism, *Columbia Law Review* **96**, 1093–1167.
[55] United States v. Chancey, 715 F.2d 543, 547 (Fla. 1983); United States v. Peralta, 941 F.2d 1003 (Cal. 1991).
[56] People v. Peterson, 537 N.W.2d 857 (Mich. 1995); State v. W.L., 650 A.2d 1035 (N.J. Super. 1995); People v. Patino, 32 Cal.Rptr.2d 345 (Ct. App. 1994); State v. Davis, 581 N.E.2d 604 (Ohio Ct. App. 1989).
[57] State v. Freeney, 637 A.2d 1088 (Conn. 1994).
[58] Commonwealth v. Hudson, 631 N.E.2d 50 (Mass. 1994); People v. Taylor, 536 N.Y.S.2d 825 (1988).
[59] Gilstrap v. State, 450 S.E.2d 436 (Ga. App. 1994); Knight v. State, 426 S.E.2d 1 (Ga. App. 1992); Jennette v. State, 398 S.E.2d 734 (Ga. App. 1990).
[60] State v. Percy, 507 A.2d 955 (Vt. 1986).
[61] State v. Hicks, 649 P.2d 267 (Ariz. 1982).
[62] Hadden v. State, 670 So.2d 77 (Fla. Ct. App. 1996).

CARL N. EDWARDS



Tarasoff Doctrine ***see*** Duty to Warn

Tarasoff v. Regents ***see*** Duty to Warn, Violence Risk Assessment for Mental Health Professionals

Teeth ***see*** Odontology

Telephone: Cell, Network Analysis ***see*** GSM Analysis and PDAs

Temporary Insanity

Insanity is a legal, not medical, term (*see* **Insanity: Defense**). Although the term has wide popular usage as a way of referring to any type of mental illness, it carries no specific medical meaning. Broadly speaking, psychiatric illness consists of symptoms that cause impairment of cognition and/or mood. An illness is diagnosed when a practitioner can match groups of signs and symptoms to a particular diagnostic label. In psychiatry, these diagnostic labels are included in various standardized handbooks. In the United States, it is the Diagnostic and Statistical Manual-IV-TR (DSM-IV-TR) and in Europe, it is the International Classification of Diseases-9 (ICD-9). A psychiatrist using the DSM-IV-TR would therefore put a patient's symptoms of visual hallucinations (e.g., hearing voices) and delusions (e.g., paranoia) under the diagnosis of schizophrenia, provided the symptoms have been present for at least six months and cause the patient distress and/or functional impairment. The term "insanity" describes neither a symptom nor a diagnosis recognized in either handbook.

Insanity is a legal construct used to justify a defendant's lack of criminal responsibility for an act. To prove that a crime has occurred, a prosecutor must show the presence of both *actus reus* (the "forbidden act") and *mens rea* (criminal intent). In asserting an insanity plea, a defendant acknowledges *actus reus,* but denies the presence of *mens rea*, by reason of mental disease or defect. Although the defendant may have committed the illegal act, they did not do so with criminal intent, and therefore are not punished with conviction or imprisonment.

History of the Insanity Plea

Since the time of Aristotle [1], society has agreed that the mentally ill should not be punished for acts caused by illness or ignorance. This general concept was

more explicitly stated in the trial of *Rex v. Arnold*, with the resulting "wild beast" test of insanity: a man would be exempted from punishment if he was "totally deprived of his understanding and memory and doth not know what he is doing, no more than an infant, than a brute, or a wild beast" [2].

M'Naghten

The "wild beast" test was used until the 1843 trial of Daniel M'Naghten, a Scottish woodcutter who shot and killed Lord Peel, secretary of the British Prime Minister, in the belief that the Prime Minister was conspiring against him. Medical experts testified (based on their observation of M'Naghten, not examination) that he was insane, and he was acquitted. The public was outraged. At the direction of Queen Victoria, the Law Lords established the following standard of insanity, commonly known as the M'Naghten test: sanity would be presumed unless the defense proved "that at the time of the committing of the act, the party accused was laboring under such defect of reason, from disease of the mind, as not to know the nature and quality of the act he was doing, or if he did know it, that he did not know he was doing what was wrong" [3]. This rule became the standard for insanity in the United States and Britain and is still used in 25 states [4].

Durham or "Product" Rule

Critics of M'Naghten stated that this cognitive test was too limiting; it would not include persons who knew right from wrong, but were unable to apply this knowledge rationally. The 1954 trial of *Durham v. United States* liberalized the M'Naghten rule by establishing the "product test": a defendant is not criminally responsible "if his unlawful act was the product of mental disease or defect".[a] This rule, however, was soon felt to be too broad. Moreover, it allowed testifying psychiatrists to go beyond expressing an opinion and reach conclusions that many felt should be left to the trier of fact.

Irresistible Impulse

Some states added a volitional component to the cognitive M'Naghten test: "irresistible impulse". This test, which was first used successfully in the 1840 trial *Regina v. Oxford*, stated that regardless of knowledge of right and wrong, a defendant could not be held criminally responsible if mental illness caused them to lack ability to conform their conduct to the law. "If some controlling disease was . . . the acting power within him which he could not resist, then he will not be responsible" [5].

Model Penal Code

In 1972, the American Law Institute (ALI) developed a new test for insanity as part of the Model Penal Code. This test included both the cognitive M'Naghten test, and the volitional irresistible impulse test, by stating a defendant is not criminally responsible if they "lack substantial capacity either to appreciate the criminality of his conduct or to conform his conduct to the requirement of law." This broad "substantial capacity" test is at present used by 19 US states [4].

Insanity Defense Reform Act

Insanity as a defense came under sharp criticism in the United States after John Hinckley was found not guilty by reason of insanity (NGRI) in the 1981 attempted assassination of President Ronald Reagan. The public was outraged and many demanded the abolition of the insanity defense altogether. In 1984, the federal insanity defense act was established, and placed the burden of proof on the defendant to prove, by clear and convincing evidence, that "at the time of the commission of the acts constituting the offense, the defendant, as a result of a severe mental disease or defect, was unable to appreciate the nature and quality or the wrongfulness of his acts." This Act is used in all the federal courts and branches of the US military [4].

Since then, various states have abolished the insanity defense. Others have limited the defense to the cognitive, and not volitional, test. Others have placed the burden of proof away from the prosecution and on the defense.

Temporary Insanity

Definition

The insanity plea has long been controversial. The defendant admits to commission of an illegal act, but

seeks to excuse their behavior by reason of mental illness. Especially in the case of a violent crime, the public seeks punishment and perceives that a criminal may be literally "getting away with murder". When John Hinckley was found NGRI, the public was outraged. After shooting President Reagan and Press Secretary James Brady, it did not seem fair to many that Mr Hinckley was hospitalized instead of imprisoned. Although he required hospitalization for treatment of his ongoing mental illness, he was technically acquitted of his crime. Studies show that most persons acquitted by reason of mental illness spend more time institutionalized in a hospital than they would have in a prison if found guilty and sentenced, yet the public perceives that such institutionalization is easier than prison, and not just.

The public has even greater difficulty accepting an insanity plea in the case of a defendant who is no longer insane and does not require any type of institutionalization. This is the case in temporary insanity: the defendant was insane at the time of the act, and is now sane. A defendant found to be not criminally responsible under such a plea is subject to release into the community without psychiatric hospitalization or treatment. Imagine the American public's outrage if John Hinckley had been acquitted of his crime by reason of insanity, and then released into the community. Despite this controversy, the plea of temporary insanity has continued to be used, with a variety of outcomes.

Permanent Insanity Versus Temporary Insanity

At present, some argue that law should not distinguish temporary insanity from chronic insanity. If a defendant is found to be not criminally responsible at the time of the act due to insanity, "it makes no difference whether the period of insanity lasted several months or merely a number of hours" [6], so long as the insanity was "fixed and stable for a reasonable duration" [4].

Some psychiatric illnesses can be temporary in nature. For example, a person under extreme stress might experience transient hallucinations or delusions. Moreover, even chronic psychiatric illnesses can wax and wane in symptom severity. A person with chronic schizophrenia may suffer an exacerbation of psychosis and temporarily experience hallucinations or paranoia.

Of course, the difficulty with such temporary illnesses or exacerbations is proof. Since a psychiatric exam is typically necessary to establish a diagnosis of mental illness, a psychiatrist should examine the defendant as soon as possible.

Historical Use of the Temporary Insanity Plea

United States: Congressman Daniel Sickles. The temporary insanity defense was first used in the United States by Congressman Daniel Sickles of New York in 1859. Sickles, a public figure known for personal scandals, learned that Phillip Key (son of American composer Francis Scott Key) was having an affair with his wife. A few days later, Sickles murdered Key in the middle of the day, in a park across from the White House. At the highly publicized trial, Sickle's team of attorneys evoked a plea of temporary insanity brought on by a "brainstorm" and "uncontrollable frenzy." The all-male jury sympathized with Sickle's rage at being cuckolded and voted to acquit. Many authors consider this verdict to be less a statement on the concept of temporary insanity than a reflection of the morals of the nineteenth century. Jurors told reporters, "In the absence of any adequate punishment by law for adultery, the man who violates the honor and desolates the home of his neighbor, does so at the peril of his life, and if he falls into the outraged husband's hands, deserves his doom" [7].

Ireland: Mary Rielly. In 1887, 30-year-old Mary Rielly was a widow with four children to support. In April, she was employed by the family of Michael Dillon, a 35-year-old man with fever, to act as nurse. According to the Royal Irish Constabulary police report, early in the morning of April 23, members of the Dillon household awoke to find Michael Dillon "lying dead on a fire which had been lit in the centre of the floor", with Mary Rielly "in a state of wild excitement, throwing burning coals upon him". Ms Rielly was arrested, found "temporarily insane from excessive drinking", and institutionalized at Dundrum Lunatic Prison in Dublin [8].

Like the case of Congressman Sickles, the case of Mary Rielly is an interesting comment on society in Ireland in the nineteenth century, explored in detail in Prior's article, "Roasting a Man Alive: the Case of Mary Rielly, Criminal Lunatic". At the time, few women were convicted of murder. Of

those who were, the majority were found guilty of infanticide. Society assumed that this action could only stem from the hormonal irregularities of women in the postpartum, and therefore such women were put in asylums for the insane rather than prisons.

Mary Rielly, who had apparently killed a grown man, was unique. Although she did not have a known history of alcohol abuse, most people involved with her case assumed that her actions stemmed from delusions caused by "demon whiskey." At the time of her arrest and trial, her manner was violent and bizarre, and she did not vigorously or coherently defend herself. Moreover, it appeared that some measure of whiskey was missing from the premises. Almost everyone, including the court, concluded that Mary had been drunk. This conclusion was further bolstered by two facts: first, murder as a result of drink was not uncommon and second, the medical literature of the day described the ability of alcohol to cause "transient frenzy" and delusional ideas. Society, in an effort to make sense of an unusual situation – a women committing murder other than postpartum infanticide – was quick to attribute the incident to alcohol.

What did Mary say of the event? Mary did not recall many details of the night, as she was exhausted from having cared for Mr Dillon night and day for over a week. An 1890 report by doctors at Dundrum to the Inspector of Lunatics noted Mary "states, as she has always done" that Mr Dillon was sick and restless, "and must have got out of bed and fallen on the fire", while the exhausted Mary was in a "state of profound slumber by his bedside." When she awoke, she was indeed in a "transient frenzy" and she attempted to put out the fire and remove coals from under Mr Dillon. The official postmortem examination offered support to her version, noting that Mr Dillon had been "very seriously ill" and had likely died from shock resulting from one initial burn to the scalp, such as might occur if a person fell onto a coal fire.

In 1891, after four years of being locked up as a criminal lunatic, Mary was judged to have regained her sanity and released on condition she emigrate to the United States. In the end, it is unknown whether she was truly "temporarily insane", or whether she was a victim of the legal system and political climate of the time.

Current Use of Temporary Insanity

Examples of mental illnesses that are severe enough to interfere with *mens rea* at the time of the crime, and then spontaneously remit, are sparse, but exist. Examples include acute stress disorder and shared psychotic disorder (or folie a deux). Acute stress disorder can occur from situations such as wife battering. In some jurisdictions, premenstrual syndrome (PMS) has been accepted as a means of reducing criminal responsibility for a defendant. In contrast, some conditions currently classified as mental illnesses – notably, substance abuse, or dependence – has not been accepted as a successful insanity defense in most jurisdictions. Likewise, conditions not currently classified as mental or medical disorders (such as homosexual panic) have not been successful insanity defenses.

Shared Psychotic Disorder. In this disorder, a delusion develops in an individual or individuals who are in a close relationship with a person who already has established a similar delusion. Joshi *et al.* published a recent account of a shared psychotic disorder among three sisters (folie a trois). In this case, two sisters without prior psycohsis developed religious delusions that mimicked those of their schizophrenic sister. The sisters, after a period of becoming increasingly enmeshed, broke into a home they believed was theirs by divine decree, and vigorously assaulted both the homeowners and officers who attempted to stop them. In jail, the sisters were initially housed together, and their delusions continued. Upon separation from their schizophrenic sister, however, the delusions of the two sisters spontaneously resolved. After a four-month hospitalization mandated by South Carolina law, the sisters were discharged to the community without incident [4].

In this situation, although the two non-schizophrenic sisters did not suffer from chronic mental illness, their extreme closeness to each other, and to their schizophrenic sister, produced a temporary delusion. At the time the sisters committed the actus rhea, therefore, they lacked *mens rea*. Specifically, they believed their actions were ordained by god, and therefore were not criminal or wrong. That the sister's delusions were temporary does not change the fact that their actions stemmed from mental illness, and therefore did not constitute a crime.

Premenstrual Syndrome (PMS). Although the defense of PMS to support claims of temporary insanity has not been particularly successful in courts, it has been successfully used to support diminished capacity [9]. In diminished capacity, a charge of murder is reduced to manslaughter on the grounds that the defendant was mentally impaired at the time of the crime, and is therefore not wholly responsible for his or her actions. This is illustrated in several landmark cases that occurred in these countries in the 1980s.

In the United Kingdom, cases involving barmaid Smith (nee Craddock) received wide media attention. In 1980, Craddock, a barmaid with a long criminal history of theft, arson, and assault was charged with murdering a coworker. Investigation of her violent behavior revealed it had a cyclical nature: PMS turned her into a "raging animal each month". On the basis of this defense, in *R v Craddock*, she was found guilty of manslaughter based on a plea of diminished responsibility. She was sentenced to probation and court-ordered progesterone treatment. As long as she remained on progesterone, she did not commit acts of violence. In 1981, her progesterone dose was lowered, and shortly thereafter Ms Smith threw a brick through a window, attempted suicide, wrote a threatening letter to a police officer, and was found hiding outside of the police station with a concealed weapon. In *R v Smith*, she was again sentenced to probation and court-ordered progesterone [10].

In another UK trial, *R v. English*, the defendant had no prior criminal record, but still successfully used the defense of PMS to reduce responsibility. After a fight with her lover, she got in her car, and ran him over. A medical expert testified that her "extreme" PMS, coupled with the fact that she had not eaten for 9 h prior to the event, produced "exceptional circumstances" that resulted in diminished responsibility. In Canada, several women accused of theft were sentenced to probation and psychiatric care due to the mitigating factor of PMS.

Although lawyers have continued to present PMS as a defense to criminal charges, reaction by the courts and public is mixed. The courts have most often rejected that PMS is a strong enough factor in a crime to completely eliminate *mens rea*, but have accepted that PMS can diminish capacity and therefore criminal responsibility. Many argue that this approach is inconsistent: if the mental condition of PMS can diminish capacity, why is it not a valid defense for temporary insanity? The public's reaction is similarly controversial. While a woman suffering from a genuine mental disorder should not be held guilty of a crime if no criminal intent existed, the "excuse" of PMS should not be used to stigmatize women.

Intoxication. The use of drugs or alcohol can significantly and transiently alter a person's mental status (*see* **Alcohol: Behavioral and Medical Effects**; **Substance Abuse**). Drugs, in particular, can cause hallucinations and delusions that mirror a state of psychosis. Traditionally, however, courts have not accepted such intoxication as grounds for diminished or absent criminal responsibility under the insanity defense, based on the fact that the intoxication was voluntary. In addition, a person who is intoxicated "is not insane, for insanity requires a 'disease of the mind' or a mental disease or defect; this requirement is one that 'mere drunkenness' cannot satisfy." An intoxicated person is therefore presumed sane; and a sane person cannot plead insanity [11].

Homosexual Panic. In 1998 in the United States, Aaron McKinney was convicted of the brutal murder of student Matthew Shepherd. Shepherd, a homosexual young man, was beaten to death by a group of men who included McKinney. McKinney's attorneys tried to argue a "gay panic defense", stating that Shepherd made a pass at McKinney, causing him to fly into murderous rage. The court rejected the defense, on the grounds that it was in essence a plea of temporary insanity under the irresistible impulse test, which is not allowed in the state of Wyoming. In addition, the court stated that "homosexual rage panic" did not have precedent in psychiatry, and would "mislead and confuse the jury" [12] (*see* **Homosexual Panic**).

Wife Battering. In a sensational United States 1993 case, Virginia the wife Lorena Bobbitt cut off her sleeping husband's penis with a kitchen knife and flung it into the woods. She explained that her husband John had repeatedly raped and abused her. The state brought criminal charges against Mr Bobbit, but he was found not guilty in criminal trial. In 1994, Ms Bobbitt faced criminal charges of malicious wounding and was ultimately found not guilty based on the jury's conclusion that she was temporarily insane from stress over her husband's battering. Ms Bobbitt was sent to a psychiatric

hospital for treatment and released after 45 days [13] (*see* **Battered Spouse Syndrome**; **Battered Woman's Reality**).

Temporary Insanity and the Media

Pleas of temporary insanity increased after the movie *Taxi Driver*, and, more recently, *The Matrix*. In the latter film, characters live not in reality, but in a computerized version of it called "the matrix". After the movie, defendants in several countries claimed they were not responsible for their actions because they also believed they were living in "the matrix". In Canada, Tonda Lynn Ansley was found NGRI of murdering her landlady. She reported that she committed the crime because she believed her landlady was part of a computerized conspiracy to brainwash and kill her. In the United States, a California man was similarly found NGRI of murdering his landlady because he had believed she caused him to be drawn into "the matrix". In Virginia, the attorneys of Joshua Cooke, who murdered his parents, stated he did so believing he was living in "the matrix" [14].

Such cases increase the controversy of the temporary insanity defense. Proponents argue that psychotic persons often center their delusions on real-world events. For example, a schizophrenic man might be paranoid if he is being watched by police, if an officer lives next door. The same man might similarly develop delusions about a popular movie, especially if that movie focuses on conspiracy and danger. Alternately, opponents of the insanity defense argue that criminals are smart, and use popular movies as an excuse to avoid responsibility for their actions.

In summary, temporary insanity, like insanity itself, is a defense that is frequently mired in controversy. The most successful use of this defense occurs in the context of a previously established mental illness (such as an exacerbation of an underlying disorder), or a new but recognized illness (such as folie a deux). Even in such cases, however, public perception and contemporaneous politics play a role in affecting the decision of the court.

End Notes

[a.] Durham v. United States, 214 F.2d at 862.

References

[1] Ross W., translator. Nicomachean Ethics by Aristotle. 350 B.C.E.

[2] Rex v. Arnold, 16 How. St. Tr. 695 (1724).

[3] M'Naughten's case, 8 Eng Rep. 718 (1843).

[4] Joshi, K., Frierson, R. & Gunter, T. (2006). Shared psychotic disorder and criminal responsibility: a review and case report of folie a trois, *Journal of the American Academy of Psychiatry and the Law* **34**(4), 511–517.

[5] Regina v. Oxford, 9 Car. P. 525, 546 (1840).

[6] Malo, A., Barach, M. & Levin, J. *The Temporary Insanity Defense in California*, Public Law Research Institute report.

[7] Brandt, N. (1991). *The Congressman Who Got Away With Murder*, Syracuse University Press.

[8] Prior, P. (2006). Roasting a man alive: the case of Mary Rielly, criminal lunatic, *Eire-Ireland* **41**(1–2), 169–191. Also available in UKPMC 2006 July 27.

[9] Downs, L.L. (2002). PMS, psychosis and culpability: sound or misguided defense? *Journal of Forensic Sciences* **47**(5), 1083–1089.

[10] Easteal, P. *Premenstrual Syndrome (PMS) in the Courtroom*, *Trends and Issues*, Vol. 37, Australian Institute of Criminology, Canberra.

[11] Hawaii legislative Reference Bureau (1998). *Reports: Drugs, Alcohol and the Insanity Defense: the Debate Over Settled Insanity*, Report No. 7, Legislative Reference Bureau, http://www.state.hi.us/lrb/rep;orts/1998.html.

[12] Tuma, C. (2000). *Matthew Shephard Murder Trial*, Courtroom Television Network, http://www.courttv.com/archive/trials/mckinney/110399_verdict_ctv.html (accessed Jul 2006).

[13] Lorena Bobbit Goes Free (1994). www.select.nytimes.com (accessed March 1, 1994).

[14] http://edition.cnn.com/2003/LAW/05/21/ctv.matrix.insanity/

JOY E. STANKOWSKI

Terry Stops ***see*** Profiles: Psychological and Behavioral

Testimony ***see*** Eyewitness Testimony

Testimony: Cross-Examination Impact on ***see*** Cross-Examination: Impact on Testimony

Testimony: Hypnosis Enhanced ***see*** Hypnosis and Memory

Testimony: Opinion (USA) ***see*** *Daubert v. Merrell Dow Pharmaceuticals*

Testimony: Reliability as Impacted by Cross-Examination ***see*** Cross-Examination: Impact on Testimony

Testing: Neuropsychological ***see*** Neuropsychological Assessment

Testing: Psychological ***see*** Psychological Testing

Therapeutic Jurisprudence

A New Idea

In 1990, David Wexler published an anthology entitled *Therapeutic Jurisprudence: The Law as a Therapeutic Agent* [1], highlighting the potential for the fields of law and mental health care to become more streamlined. He supported the use of the law as another tool in the arsenal of treatment options available for the mentally ill. Wexler emphasized the notion that any outcome in the justice system has the potential to act in either a therapeutic or nontherapeutic way, and he encouraged revisions to the system that could yield in therapeutic results. Since this time, many authors, including Wexler, have continued to expand on this idea.

Therapeutic Jurisprudence

Therapeutic jurisprudence "is the study of the role of the law as a therapeutic agent" [1]. Any legal decision and the participants in the process can have either a positive or a negative impact on the individual involved. Participants who are aware of the potential psychological ramifications of their actions may be able to avoid unintended consequences. Attorneys must value not only their clients' legal rights but also their psychological well being. The concept of therapeutic jurisprudence is most often referenced in the context of mental illness but is applicable to many other types of law as well, such as domestic violence and family law [2].

Early writings on the topic stress four main areas where the application of therapeutic jurisprudence could be further studied. The first area concerns the psychological dysfunction that the law can cause. This could refer specifically to such unintended consequences as diminished self-esteem or negative thoughts of hopelessness, helplessness, and worthlessness. The second involves investigating potential therapeutic aspects of legal rules. For example, this could include sentencing patients to a treatment facility rather than a prison. The third area entails examining the therapeutic aspect of legal procedures; individuals may reclaim some control in their lives

by taking responsibility for their wrongdoing. The fourth calls for the examination of the therapeutic aspects of judicial and legal roles. For example, attorneys may be encouraged to more vigorously represent their clients in civil commitment hearings [3]. Many authors have expanded on these ideas and offered creative solutions.

While it is important to consider the potential therapeutic advantages of legal outcomes, it is emphasized that these benefits should not stand in the way of appropriate adjudication of court cases. Therapeutic jurisprudence is but one of the many elements to consider when determining the resolution of legal matters. Ultimately, mental health laws can serve to both protect the civil liberties and assist in meeting the therapeutic needs of the mentally ill [4].

The Development of Therapeutic Jurisprudence

In the first half of the twentieth century, American psychiatric practice was governed by the theory of paternalism, in which the physician provided the treatment that he believed was in the patient's best interest. Patients could be committed to long term facilities based on the recommendation of a physician alone in some cases. Procedures were performed and medications were given without necessarily obtaining the patient's consent [5]. In the later half of the century, in conjunction with factors including deinstitutionalization and a national focus on civil liberties, the rights of the mentally ill were closely examined [6].

With the subsequent development of a rights driven approach to mental health law, Wexler contended that the field became too regimented, ignoring the uniqueness of each individual [7]. That the theory of therapeutic jurisprudence is not limited to a rights orientation allows for a less biased analysis of whether changes in the legal system actually produce their intended outcome [8]. While recognizing the importance of these rights, others favoring therapeutic jurisprudence proposed that, if these rights have antitherapeutic consequences, perhaps they are of secondary importance to an appropriate therapeutic outcome [9, 10]. This different perspective may create the ability to reconceptualize the field of mental health law [11]; instead of changing the laws themselves, the way that the law is practiced could be reformed [12].

Though therapeutic jurisprudence may initially appear paternalistic, proponents believe that there are many applications with outcomes that foster self-determination, and it is these results that are felt to be more therapeutic than rules dictating the fate of the individual [4, 13]. Winick, a legal scholar with many publications on therapeutic jurisprudence, noted that his work "is animated by the insight that such paternalism is often antitherapeutic, and that legal protection for individual autonomy can have positive therapeutic value" [14].

Applications of Therapeutic Jurisprudence

Wexler identified "psycholegal soft spots", which he defined as areas where the opportunity arises for the law to intervene in a therapeutic way [15]. Specialized courts, civil commitment, mandatory treatment, and informed consent are processes that can each lend themselves to the application of therapeutic jurisprudence.

Specialized Courts

One of the clearest examples that incorporate the tenets of therapeutic jurisprudence is the existence of specialized courts (*see* **Mental Health Courts**). Participants in the proceedings of specialized courts are provided specific training pertinent to their area of practice. The overall purpose of these courts is to shift away from the traditionally adversarial legal process in order to address the underlying causes of crime.

Drugs courts were developed in the 1980s following a tremendous rise in drug-related cases. In these courts, addiction treatment may be viewed as an option for decreasing recidivism when used in lieu of punishment. The power of the law allows for supervision to support offenders' participation in and completion of substance use treatment programs. The court's involvement in the process assists in clearly defining expectations and affirming offenders' agreements to take part in the program. The law also provides the ability to sanction offenders when requirements are not fulfilled. Finally, the court offers offenders the opportunity to begin taking personal responsibility for their addictions by making choices concerning their treatment [13].

The first mental health court (*see* **Mental Health Courts**) was established in the United States in 1997,

and since that time, others have been developed nationally [16, 17] and internationally [18]. These courts serve to better utilize therapeutic opportunities created through legal interventions. In addition to many of the tenets of drug courts, mental health courts can

1. expedite case processing for early identification of individual treatment needs;
2. encourage participation of the individual and his or her family; and
3. employ a liaison to facilitate interactions between the individual, their family, the court and mental health services [13].

Other examples of specialized courts include domestic violence courts, family courts, and community focused courts [13]. While they do not necessarily deal directly with addiction or mental illness, the principles of therapeutic jurisprudence are still evident in their focus on providing appropriately specialized services to those individuals processed through their dockets.

The existence of specialized courts is evidence that, rather than merely being applied by individual judges or attorneys, the principles of therapeutic jurisprudence are being utilized on a systemic level. There are, however, some notable drawbacks in maintaining these specialized courts. It may be difficult to define the spectrum of the courts' practice in order to effectively process the maximum number of cases. It is important that the courts maintain their specialization without overly narrowing their focus such that there are not enough cases to warrant a separate court. Professionals practicing in these courts may suffer burnout due to overexposure to the increased psychological and social complexity of their caseload. In addition, some practitioners may be unwilling to participate due to perception of less prestigious cases and fewer opportunities for professional advancement [19].

Civil Commitment

A number of authors have examined the issue of civil commitment (*see* for example, **Civil Commitment**) in relation to the principles of therapeutic jurisprudence. Though there has been significant change over the last several decades, some still feel that civil commitment hearings exist only as a formality [20]. Given their clients' diagnoses of mental illness, attorneys may advocate less vigorously for their rights, instead of feeling that they should do what they deem to be in the best interest of their clients [21, 22]. This may cause the hearing to become a nonadversarial event, in which the judge simply "rubber stamps" the opinion of the expert without giving fair consideration to the patient's opposing position [20, 23]. Such a scenario could demoralize and discourage patients, who may anticipate active participation in what they assume will be fair and balanced proceedings. Subsequently, patients may lose confidence in the legal system, their attorneys, judges, and mental health care providers [24, 25]. This may ultimately lead to increased difficulties with hospitalizations and treatment compliance in the future [8].

According to the principles of therapeutic jurisprudence, civil commitment hearings offer an opportunity for the court to intervene in a therapeutic manner. Those involved may use this occasion to explain to the patient the meaning and process of civil commitment, including the benefits and repercussions that can be associated with it. In particular, this may include the potential improvement in symptoms or increased likelihood of safety that could be associated with an extended hospitalization. On the other hand, patients should also be aware of the potential stigma associated with civil commitment. Collectively, this knowledge may assist the patients in making good choices, encourage them to advocate for themselves and decrease their concerns about the proceedings. If patients feel that their concerns have at least been recognized, they may be more willing to accept the decision, even if they feel that it is unfavorable to them [20].

The release of a patient from civil commitment also creates a situation in which the theory of therapeutic jurisprudence may apply. According to Wexler, this offers an opportunity for the law to intercede, smoothing the transition between life within the confines of the hospital and what awaits outside. This could be achieved through such measures as outpatient civil commitment. The therapeutic goals of this easier transition would include less psychological stress on the individual and ultimately the decreased likelihood of recommitment [15].

Mandatory Treatment

When a mentally ill individual no longer requires an inpatient setting but will most likely decompensate

in the community, the court can intervene in the form of outpatient civil commitment. Alternatively, when a hospitalized patient refuses medications, and for example, appears to be a threat to themselves or others, after being petitioned the court may require the patient to accept the proposed treatment regimen.

Given that medication compliance can be made mandatory, there exists a high likelihood that individuals may feel coerced, which could undermine the effectiveness of the treatment [26]. Feelings of coercion may be minimized by acting out of concern for the individual, being respectful, fair, and honest, and providing a forum for the individual to express concerns [27]. The experience may also be improved for individuals if they are offered some degree of decision making ability, such as choosing between medications [28].

Informed Consent

Informed consent requires explaining to a patient the nature and purpose of treatment as well as its risks, benefits and alternatives [29]. Failure to do so may be a violation of tort law [30]. In addition to protecting oneself from malpractice, advocates of therapeutic jurisprudence believe that the process of informed consent can be used to strengthen the therapeutic alliance between care providers and patients [28]. Defining goals and negotiating about different treatment options allows patients to take ownership of their own health care. Inherent in the process of informed consent is offering the option of no treatment. Recognizing and discussing this right to refuse treatment (*see* **Treatment, Right to Refuse: Mental Health**) may also assist in enhancing the therapeutic relationship between care providers and patients [31].

Therapeutic Jurisprudence and the Future

Therapeutic jurisprudence is a young field that continues to evolve as more investigators apply its tenets to areas in which law and psychiatry overlap. Given the variety of people with input into judicial proceedings, the future of research in this field requires interdisciplinary collaboration, including participation from judges, attorneys, parole officers, law enforcement, social workers, and mental health care providers.

Preventative Law

While the judges may practice therapeutic jurisprudence in their courtrooms by taking advantage of therapeutic moments as they arise, Wexler proposed that the future lies in preventative law [32]. This concept, espoused by Stolle in the context of elder law [33], highlights the importance of "legal checkups" to identify "soft spots" in order to minimize potential legal problems [34]. When combined with therapeutic jurisprudence, these practices would aid in the identification of legal issues that may generate anxiety, distress, depression, or resentment. Subsequent to the recognition of these scenarios, lawyers may implement solutions to lessen these effects. Wexler provided the example of an estate planning attorney who assists clients with the potential emotional repercussions of choosing to leave someone out of their will.

Advance Directives

Another potential future oriented application of therapeutic jurisprudence involves the creation of mental health care advance directives. When patients are competent, lawyers and mental health care providers may suggest to them that they consider drafting a document that describes their wishes should they become incompetent. These advance directives empower patients and provide clinicians with some insight into patients' desires concerning their treatment [35].

Conclusion

Advocates of therapeutic jurisprudence note that, with continued collaboration between the fields of law and mental health care, each would become more familiar and comfortable with the workings of the other. This may serve to streamline the efforts of both and create better therapeutic outcomes for patients by opening up more options for treatment and encouraging better usage of the modalities already in existence. According to its proponents, this becomes evident when the tenets of therapeutic jurisprudence are applied to situations such as specialized courts, civil commitment, mandatory treatment and informed consent. Future applications of therapeutic jurisprudence seem to exist in prevention of unanticipated and distressing legal consequences.

References

[1] Wexler, D. (ed) (1990). *Therapeutic Jurisprudence: The Law as a Therapeutic Agent*, Academic Press, Durham.

[2] Wexler, D. (1995). Reflections on the scope of therapeutic jurisprudence, *Psychology, Public Policy and Law* **1**, 220–236.

[3] Wexler, D. & Schopp, R. (1992). Therapeutic jurisprudence: a new approach to mental health law, in *Handbook of Psychology and Law*, D. Kagehiro & W. Laufer, eds, Springer-Verlag, New York.

[4] Wexler, D. (1992). Putting mental health into mental health law: therapeutic jurisprudence, *Law and Human Behavior* **16**(1), 27–38.

[5] Melton, G.B., Petrila, J., Poythress, N.G. & Slobogin, C. (eds) (1997). Civil competencies, in *Psychological Evaluations for the Courts: A Handbook for Mental Health Professionals and Lawyers*, 2nd Edition, The Guilford Press, New York, pp. 337–362.

[6] Melton, G.B., Petrila, J., Poythress, N.G. & Slobogin, C. (eds) (1997). Civil commitment, in *Psychological Evaluations for the Courts: A Handbook for Mental Health Professionals and Lawyers*, 2nd Edition, The Guilford Press, New York, pp. 297–336.

[7] Wexler, D. (1993). Therapeutic jurisprudence and changing conceptions of legal scholarship, *Behavioral Sciences and the Law* **16**, 27–38.

[8] Wexler D. (1991). An introduction to therapeutic jurisprudence, in *Essays in Therapeutic Jurisprudence*, D. Wexler & B. Winick, eds, Carolina Academic Press, Durham.

[9] Schopp, R. (1993). Therapeutic jurisprudence and conflicts among values in the mental health system, *Behavioral Sciences and the Law* **11**, 31–45.

[10] Slobogin, C. (1995). Therapeutic jurisprudence: five dilemmas to ponder, *Psychology, Public Policy and the Law* **1**, 193–219.

[11] Petrilla, J. (1995). Who will pay for involuntary civil commitment under capitated managed care? An emerging dilemma, *Psychiatric Services* **46**(10), 1045–1048.

[12] Wexler, D. (1996). Applying the law therapeutically, in *Law in a Therapeutic Key*, B. Wexler & B.J. Winick, eds, Carolina Academic, Durham.

[13] Casey, P. & Rottman, D.B. (2000). Therapeutic jurisprudence in the courts, *Behavioural Sciences and the Law* **18**(4), 445–457.

[14] Winick, B.J. (1996). The jurisprudence of therapeutic jurisprudence, in *Law in a Therapeutic Key*, B. Wexler & B.J. Winick, eds, Carolina Academic, Durham.

[15] Wexler, D. (1998). Practicing therapeutic jurisprudence: psycholegal softspots and strategies, *Revista Juridica Universidad de Puerto Rico* **67**, 317–342.

[16] Goldkamp, J.S. & Irons-Guynn, C., Bureau of Justice Assistance (2000). *Emerging Judicial Strategies for the Mentally Ill in the Criminal Caseload: Mental Health Courts in Fort Lauderdale, Seattle, San Bernardino, and Anchorage*, U.S. Department of Justice, Washington, DC.

[17] Kuehn, B.M. (2007). Mental health courts show promise, *Journal of the American Medical Association* **297**(15), 1641–1643.

[18] Lawrence, J.M. (2004). Queensland's mental health court: the first 12 months, *Health Law in Canada* **24**(3), 47–59.

[19] Rottman, D.B. & Casey, P. (1999). Therapeutic jurisprudence and the emergence of problem-solving courts, *National Institute of Justice Journal* **summer**, 12–19.

[20] Winick, B.J. (2001). The civil commitment hearing: applying the law therapeutically, in *The Evolution of Mental Health Law*, L.E. Frost & R.J. Bonnie, eds, American Psychological Associations, Washington, DC.

[21] Abisch, J.B. (1995). Mediational lawyering in the civil commitment context: a therapeutic jurisprudence solution to the counsel role dilemma, *Psychology, Public Policy and the Law* **1**, 120–141.

[22] Perlin, M.L. & Sadoff, R.L. (1982). Ethical issues in the representation of individuals in the commitment process, *Law and Contemporary Problems* **45**, 161–192.

[23] Poythress, N.G. (1977). Mental health expert testimony: current problems, *Journal of Psychiatry and the Law* **5**, 201–222.

[24] Tyler, T.R. (1992). The psychological consequences of judicial procedures: implications for civil commitment hearings, *Southern Methodist University Law Review* **46**, 433–445.

[25] Winick, B.J. (1997). Coercion and mental health treatment, *Psychology, Public Policy and the Law* **3**, 184–206.

[26] Bennett, N.S., Lidz, C.W., Monahan, J., Mulvey, E.P., Hoge, S.K., Roth, L.H. & Gardner, W. (1993). Inclusion, motivation, and good faith: the morality of coercion in mental hospital admission, *Behavioral Sciences and the Law* **11**(3), 295–306.

[27] Monahan, J., Hoge, S., Lidz, C., Eisenberg, M., Bennett, N., Gardner, W., Mulvey, E. & Roth, L. (1996). Coercion to inpatient treatment: initial results and implications for assertive treatment in the community, in *Coercion and Aggressive Community Treatment: A New Frontier in Mental Health Law*, D. Dennis & J. Monahan, eds, Plenum, New York.

[28] Winick, B. (1997). Mandatory treatment: an examination of therapeutic jurisprudence, *New Directions for Mental Health Services* **75**, 27–34.

[29] Kleinke, C. (1994). *Common Principles of Psychotherapy*, Brooks/Cole, Pacific Grove.

[30] Fenwick, P. & Beran, R.G. (1997). Informed consent – should Bolam be rejected? *Medicine and Law* **16**(2), 215–223.

[31] Elbogen, E.B. & Tomkin, A.J. (1999). The psychiatric hospital and therapeutic jurisprudence: applying the law to promote mental health law, *New Directions for Mental Health Services* **84**, 71–84.

[32] Wexler, D. (2001). The development of therapeutic jurisprudence: from theory to practice, in *The Evolution*

of Mental Health Law, L.E. Frost & R.J. Bonnie, eds, American Psychological Associations, Washington, DC.
[33] Stolle, D.P. (1996). Professional responsibility in elder law: a synthesis of preventive law and therapeutic jurisprudence, *Behavioral Sciences and the Law* **14**(4), 459–478.
[34] Hardaway, R.M. (1997). *Preventative Law: Materials on a Nonadversarial Legal Process*, Anderson, Cincinnati.
[35] Winick, B.J. (1996). Advance directive instruments for those with mental illness, *University of Miami Law Review* **51**, 57–95.

Further Reading

Website: www.therapeuticjurisprudence.org (last visited September 8, 2007).

Related Articles

Treatment, Right to: Mental Health

SARA G. WEST AND SUSAN HATTERS-FRIEDMAN

Threat Assessment *see* Risk Assessment

Threat Assessment: Mental Health Settings *see* Violence Risk Assessment for Mental Health Professionals

Threat Assessment: School

Rationale for Threat Assessment in Schools

Threat assessment (*see* **Threat Assessment: Workplace**) became an important school safety practice in the United States following a series of highly publicized school shootings in the late 1990s. These tragic events stimulated the perception that violence was rampant in American schools. In response, both the U.S. Secret Service [1] and Federal Bureau of Investigation (FBI) [2] conducted studies of school shootings and observed that the student perpetrators were often victims of bullying who had become angry and depressed, and were influenced by a variety of social, familial, and psychological factors. However, the studies concluded that, because similar characteristics are found in so many students, it would not be possible to develop a profile or checklist that could be used to pinpoint the small number of truly violent students among them. As a result, both the FBI and Secret Service cautioned schools against a profiling approach.

Nevertheless, the FBI and Secret Service did point out that almost all of these students communicated their intentions to attack through threats and warnings to their peers. Had these threats been reported to authorities and investigated, the shootings might have been prevented. In its study, the FBI identified a number of potential school shootings that were prevented because students reported a threat to authorities that was investigated and determined to be serious. On the basis of these observations, the FBI and Secret Service both recommended that schools adopt a threat assessment approach [2, 3].

School threat assessment is a form of risk assessment designed to prevent targeted violence, i.e., incidents in which an identified individual targets a particular victim prior to attack [4]. In the case of school shootings, there may be multiple targeted victims or the general school population that are subject to attack. Threat assessment was developed by the Secret Service as a means of evaluating the risk posed by individuals who communicated threats to public figures [5], but was soon applied to broader contexts such as workplace violence (*see* **Threat Assessment: Workplace**).

Threat assessment begins when an individual is identified as having made a threat, which is defined broadly as any expression of intent to harm someone, including indirect or ambiguous threats. Threat assessment is concerned with determining whether the individual who has "made" a threat actually "poses" a threat [1]. An individual is more likely to pose a threat if he or she has engaged in planning or preparation to carry out the threat, such as obtaining weapons, making detailed plans, rehearsing an

attack, recruiting accomplices, or inviting others to observe the attack. It follows that a threat assessment is focused on behaviors that are linked to carrying out a threat, as distinguished from broader characteristics such as the individual's personality or demographic background.

Threat assessment can be distinguished from another form of risk assessment: criminal profiling [2, 4]. Profiling originated as a means of identifying the likely perpetrator of a crime by examining crime scene evidence – which is a retrospective process, but evolved to include prospective attempts to infer the identity of potential perpetrators by matching them to the characteristics of individuals who have committed similar crimes. Prospective profiling is dependent on the existence of a set of defining characteristics of individuals who commit a particular kind of crime. There are two main weaknesses to this procedure that are common to all prediction efforts: the indicators must be sensitive enough to detect all or nearly all perpetrators and they must be sufficiently specific that they do not identify innocent individuals as perpetrators [6] (*see* **Serial Homicide**; **Profiles: Psychological and Behavioral**). The FBI and Secret Service concluded that the common characteristics of students who engaged in rampage school shootings lacked both sensitivity and specificity, i.e., there was no characteristic or set of characteristics that were common to all offenders and the most promising characteristics that they shared were not specific to offenders.

The weakness of profiling can be observed in the checklist of warning signs that the US Department of Education included in its 1998 publication, "Early warning, timely response: A guide to safe schools" [7]. Among the 16 warning signs are items such as "history of discipline problems", "drug use and alcohol use", and "feelings of being picked on and persecuted", and "excessive feelings of rejection" that could be found among many adolescents. Similarly, the National School Safety Center [8] developed a 20-item "checklist of characteristics of youth who have caused school-associated violent deaths" that included items such as "has been previously truant, suspended, or expelled from school", "has little or no supervision from parents or a caring adult", and "tends to blame others for difficulties she or he causes". Because the base rate of severe violence is low, checklists like these will produce large numbers of false-positive identifications of students who are not violent [6]. As the authors of the federal warning signs report cautioned, "Unfortunately, there is a real danger that early warning signs will be misinterpreted" [7], p. 7.

Perceptions of Risk

Forensic consultants must be aware of the influence of public perceptions of risk on school policies and administrative decision-making. In response to the school shootings of the 1990s, school administrators began to impose zero tolerance sanctions on students for seemingly minor transgressions such as bringing a plastic knife to school in a lunchbox, pointing a finger like a gun, and shooting a paper clip with a rubber band [9]. The public response reflected a similar fear of school violence: a Gallup poll conducted after the 1999 Columbine shooting found that two-thirds of Americans believed that a similar incident could occur at schools in their community [10]. Ironically, in the year of the Columbine shooting, 17 students were killed at school, but over 2500 young people (ages 5–19) were murdered outside of school, and more than 9700 were killed in accidents [11].

Heightened fears often follow a highly publicized school shooting and generate new calls for security measures. For example, after a man invaded a one-room Pennsylvania Amish school and killed five girls in 2006, there were renewed recommendations to arm teachers with guns [12] and a call to issue Kevlar-coated textbooks to students for use as bullet shields [13]. A Fort Worth suburban school division hired a former military officer to train students to make a mass attack and attempt to subdue an armed gunman [14]. After the 2007 shooting at Virginia Tech, colleges across the country began to implement new safety and security measures [15].

Public fears may be encouraged by media speculations about "new trends". The Pennsylvania Amish school shooting generated news reports of "a new trend of adults killing children in schools" [16] and "a pattern in rural school shootings" with "girls as targets" [17]. Pseudo-trends can be generated by random patterns in the large numbers of violent acts that occur each day in the United States. In a nation of 300 million people, there are approximately 30 000 shooting fatalities (suicide, homicide, and accident) each year [18], which means an average of 82 fatalities every day.

In making recommendations to school officials and legal authorities, forensic consultants may need to correct misperceptions of the risk of violence at school. Student-perpetrated homicides are rare events in the nation's 119 000 schools. There were 103 such cases during the 12 school years from 1992–1993 to 2003–2004, which means an average of 8.58 incidents per year [7]. This rate will mean that the public could hear about a fatal school shooting by a student approximately once a month throughout the school year, fueling the perception that such events are commonplace. Although even one school homicide is too many, an event that affects an average of 8.58 schools out of 119 000 means that the average school can expect a student-perpetrated homicide about once every 13 870 years (119 000 divided by 8.58). These figures are estimates that will vary over time, but illustrate the very low base rate of student-perpetrated school homicides.

After the Virginia Tech shooting in 2007, there was a public perception that colleges were dangerous places, too. According to the latest available data from the US Department of Education [19], there were 95 murders on college campuses in the six years from 1999 to 2004, an average of 16 per year. Since there are approximately 4200 colleges in the United States, this means that an average college can expect to experience a murder on campus about once every 265 years. If we include all 2808 murders that occurred in the surrounding community – off campus as well as on campus – the rate is much higher: about once every 9 years. This is a reflection of the much higher rate of violence in the general community.

Although there are limitations to any single source of information about violent crime, the evidence from multiple sources is consistent in indicating an overall decline in youth violence in communities and schools from the early 1990s to recent years.

- According to FBI arrest statistics, juvenile homicide declined more than 80% from 1993 (3284 arrests) to 2006 (956 arrests; [20]).
- Homicides on the grounds of primary and secondary schools during the school day dropped from 42 during the 1992–1993 school year to 11 during the 2006–2007 school year [21].
- Serious violent crimes, as measured by victim reports on the National Crime Victimization Survey dropped from 13 per 1000 students in 1994 to 5 per 1000 in 2005 [22].
- According to the Youth Risk Behavior Surveys administered nationally to over 10 000 students each year, there have been declines of 21% in physical fighting and 48% in weapon carrying at school from 1993 to 2003 [23].

The attention given to school homicides can overshadow the more pervasive problems with fighting, bullying, and other forms of aggressive behavior at school. According to data compiled by the National Center for Education Statistics [22], during the 2005–2006 school year, 78% of public schools reported at least one violent crime at their school. In 2005, 8% of high school students reported being threatened or injured with a weapon in the previous 12 months, and 28% of students, ages 12–18, reported being bullied at school in the previous six months.

Schools authorities take disciplinary action in response to student aggression, and frequently make referrals for risk assessments of those students. According to the national School Survey on Crime and Safety [22], nearly half (48%) of all public schools took serious disciplinary action against one or more students during the 2005–2006 school year. This included suspensions of 5 days or more (74%), expulsions without services (5%), and transfers to special schools (20%).

The Virginia Model for Student Threat Assessment

Many school systems have developed in-house methods for conducting risk assessments, such as the Dallas Threat of Violence Risk Assessment [24], which uses a checklist of 19 risk factors to classify a student as low, medium, or high risk. There are also several guidebooks on school security that advocate principles of threat assessment [25, 26]. Researchers at the University of Virginia developed and field-tested a comprehensive set of threat assessment guidelines [27] derived from the recommendations of the FBI [2] and Secret Service reports [3]. The Virginia model conceptualizes threat assessment as a process involving threat management and risk reduction rather than the more traditional prediction of violence [28]. The threat assessment team attempts to understand why a student has made a threat in order to address the underlying conflict or problem that triggered it. This

model places an emphasis on determining the seriousness of the threat in order to avoid the complementary errors of overreacting to threats that are not serious as well as failing to take decisive action in the face of a serious threat.

According to the Virginia model, each school has a threat assessment team led by the principal or assistant principal and staffed by representatives from mental health and law enforcement. Teams follow a seven-step decision tree. The initial stages of a threat assessment are typically handled by the team leader (principal), and many cases can be readily resolved. In more complex or ambiguous cases, the team leader brings in additional team members for more extensive assessment. The seven steps are reviewed briefly as follows:

Step 1. The leader of the threat assessment team interviews the student who made the threat and witnesses using a standard set of questions. The focus is not only on determining what the student said or did but also the context in which the threat was made and what the student intended by making the threat.

Step 2. The team determines whether the threat is "transient" or "substantive". Transient threats can be readily identified as expressions of anger or frustration (or perhaps inappropriate attempts at humor) that dissipate quickly when the student reflects on the meaning of what he or she has said. In contrast, substantive threats represent a sustained intent to harm someone beyond the immediate incident. If there is doubt whether a threat is transient or substantive, the threat is regarded as substantive. One way to identify a threat as substantive is to look for certain characteristics derived from the FBI report [2] that suggest that the threat is likely to be serious:

- the threat includes plausible details, such as a specific victim, time, place, and method of assault;
- the threat has been repeated over time or communicated to multiple persons;
- the threat is reported as a plan, or planning has taken place;
- the student has accomplices, or has attempted to recruit accomplices;
- the student has invited an audience of peers to watch the threatened event; and
- there is physical evidence of intent to carry out the threat, such as a weapon or bomb materials.

Step 3. A transient threat identified at step 2 can be resolved without engaging a comprehensive evaluation. The student may make an explanation and apologize to those affected by the threat, or take other action to make amends for his or her behavior. There may be a reprimand or other disciplinary consequence if the behavior was disruptive or violated the school's discipline code. If a transient threat was sparked by an argument or conflict, the principal can involve other team members in helping to address or resolve the problem.

Step 4. If the threat is not transient, a substantive threat is determined to be "serious" or "very serious", based on the intended severity of injury. A "serious" threat is a threat to assault, strike, or beat someone up. A "very serious" threat is a threat to kill, sexually assault, or severely injure someone.

Step 5. In the case of a serious substantive threat, the team takes actions to protect potential victims. Protective actions depend on the circumstances of the threat, as well as how soon and where the threat might be carried out. Immediate protective actions include cautioning the student about the consequences of carrying out the threat and contacting the student's parents. The team also has the responsibility of notifying the intended victim of the threat.

Step 6. Very serious substantive threats require the most extensive action by the team. In addition to any immediate action to assure that the threat is not carried out, the student is suspended from school, pending a complete assessment of the threat and determination of the most appropriate school placement. The team conducts a more comprehensive safety evaluation, which includes both a mental health assessment conducted by the school psychologist or another suitably trained mental health professional and a law enforcement investigation conducted by the school resource officer.

Step 7. The team integrates findings from the safety evaluation into a written safety plan, which is designed both to protect potential victims and to address the student's educational needs. At this point, the principal decides whether the student can return to school or should be placed in an alternative setting. If the student is permitted to return to school, the plan describes the conditions that must be met and the procedures in place to monitor the student when he or she returns.

This threat assessment procedure was field-tested in 35 schools for one school year [9, 29, 30]. The 35 threat assessment teams responded to a total of 188 threats. The most common threat was a threat to hit or beat up someone (77 cases, 41%). There were also 27 threats to kill, 24 threats to shoot, and 18 threats to cut or stab. There were 32 cases in which the threat was vague or nonspecific (e.g., "I'm going to get you"), and 10 miscellaneous threats, such as to set a fire or detonate a bomb. All types of threats were seen at all school levels, although threats to kill or shoot occurred more frequently in elementary school than in middle and high school combined.

The majority (70%) of threats were easily resolved as transient threats. Of the remaining 30% that were substantive threats, 22% (42 cases) were "serious" substantive threats to fight or assault someone, and 8% (15 cases) were "very serious" substantive threats to kill or severely injure someone. A threat assessment approach gives school authorities flexibility in choosing the disciplinary consequences for students who make threats. Under a zero tolerance policy, many students would have been expelled for making threats to kill or injure someone. Using threat assessment guidelines, only three of the 188 threat cases resulted in expulsion. Half of the students who made threats (94 cases) were given a short-term suspension, either an in-school suspension or a suspension outside of school. The modal suspension (32 cases) was one day, with a range of 1–10 days. Only six students were arrested; three students had made a false bomb threat, two students had assaulted a school staff member, and the final case involved a student found with a knife.

At the end of the school year, principals were interviewed to obtain follow-up information on each of their cases. In three cases, the principal was not sure whether a student's threat to hit another student was carried out, but in all other cases, the principals reported that the threat was not carried out. When asked to assess the student's overall behavior after the threat, the principals rated 43% of the students as demonstrating improved behavior during the remainder of the school year, 39% as about the same, and only 18% as worse in their behavior.

Limitations to Threat Assessment

There is a need for randomized controlled studies examining the effects of threat assessment (and other risk assessment procedures) on school safety and student outcomes. Research on threat assessment is limited in part because school systems are understandably reluctant to engage in experimental procedures about a highly sensitive issue.

Another limitation is that threat assessment is designed for persons identified as making threats, so it is critically important that students (as well as parents and teachers) be encouraged to report threats. Many students are reluctant to report threats because they regard it as a form of "snitching" or they fear that they will become victims of retaliation. Accordingly, schools are advised to teach students to distinguish snitching for personal gain from seeking help to prevent someone from being hurt, and to offer anonymous as well as direct means of reporting threats.

Because threat assessment is concerned with targeted violence, it is not a useful approach for preventing less specific forms of violence. To prevent instrumental crimes such as robbery, schools must rely primarily on security measures and close supervision of students. To prevent fighting and bullying, there are a variety of effective school-based programs ranging from individual counseling to schoolwide systems of positive behavior support [9, 31–33]. A meta-analysis of 249 controlled studies confirmed the effectiveness of school-based interventions for aggressive and disruptive behavior [34].

References

[1] Vossekuil, B., Fein, R.A., Reddy, M., Borum, R. & Modzeleski, W. (2002). *The Final Report and Findings of the Safe School Initiative: Implications for the Prevention of School Attacks in the United States*, U.S. Secret Service and U.S. Department of Education, Washington, DC.

[2] O'Toole, M.E. (2000). *The School Shooter: A Threat Assessment Perspective*, National Center for the Analysis of Violent Crime, Federal Bureau of Investigation, Quantico.

[3] Fein, R., Vossekuil, B., Pollack, W., Borum, R., Modzeleski, W. & Reddy, M. (2002). *Threat Assessment in Schools: A Guide to Managing Threatening Situations and to Creating Safe School Climates*, U.S. Secret Service and Department of Education, Washington, DC.

[4] Reddy, M., Borum, R., Berglund, J., Vossekuil, B., Fein, R. & Modzeleski, W. (2001). Evaluating risk for targeted violence in schools: Comparing risk assessment, threat assessment, and other approaches, *Psychology in the Schools* **38**, 157–172.

[5] Fein, R.A., Vossekuil, F. & Holden, G.A. (1995). *Threat Assessment: An Approach to Prevent Targeted Violence*, National Institute of Justice: Research in Action, *1–7 (NCJ 155000)*, http://www.secretservice.gov/ntac.htm. (accessed August 2, 2004).

[6] Sewell, K.W. & Mendelsohn, M. (2000). Profiling potentially violent youth: Statistical and conceptual problems, *Children's Services: Social Policy, Research, and Practice* **3**, 147–169.

[7] Dwyer, K., Osher, D. & Warger, C. (1998). *Early Warning, Timely Response: A Guide to Safe Schools*, U.S. Department of Education, Washington, DC.

[8] National School Safety Center (1998). *Checklist of Characteristics of Youth Who Have Caused School-Associated Violent Deaths*, Westlake Village, http://www.nssc1.org (accessed Mar 2008).

[9] Cornell, D. (2006). *School Violence: Fears Versus Facts*, Lawrence Erlbaum, Mahwah.

[10] Saad, L. (1999). *Public Views Littleton Tragedy As Sign of Deeper Problems In Country*, http://www.gallup.com/poll/content/login.aspx?ci=3898 (accessed Apr 2005).

[11] Anderson, R.N. (2001). Deaths: leading causes for 1999, *National Vital Statistics Reports* **49**, 1–88.

[12] Associated Press (2006). *Wisconsin Lawmaker Urges Arming Teachers*, http://www.usatoday.com/news/nation/2006-10-05-arming-teachers_x.htm (accessed November 12, 2006).

[13] Associated Press (2006). *Candidate Proposes Using Textbooks As Shields*, http://www.cnn.com/2006/EDUCATION/10/20/school.shootings.textbooks.ap/index.html (accessed November 12, 2006).

[14] *Dallas Morning News* (2006). *Burleson Changes Stance On Student Attack Training*, October 25, 2006, http://www.dallasnews.com/sharedcontent/ dws/dn/latestnews/stories/102606dnmet burleson736f1085.html (accessed November 30, 2006).

[15] Lenckus, D. (2008). Shootings heighten college safety concerns, *Business Insurance*, http://www.businessisurance.com/cgi-bin/article.pl?articleId=24212 (accessed March 4, 2008).

[16] Thomas, P. (2006). *Why the Spike in School Shootings?* http://abcnews.go.com/GMA/story?id=2521025 (accessed March 6, 2008).

[17] Chaddock, G.R. & Clayton, M. (2006). *A Pattern in Rural School Shootings: Girls as Targets* (October 4, 2006), The Christian Science Monitor, http://www.csmonitor.com/2006/1004/p01s01-usgn.html (accessed March 6, 2008).

[18] Minino, A.M., Anderson, R.N., Fingerhut, L.A., Boudreault, M.A. & Warner, M. (2006). Deaths: Injuries, 2002, *National Vital Statistics Reports* **54**(10), 1–125.

[19] U.S. Department of Education (2001–2004). *Summary Campus Crime and Security Statistics – Criminal Offenses*, Washington, DC, http://www.ed.gov/admins/lead/safety/crime/criminaloffenses/index.html (accessed March 6, 2008).

[20] Federal Bureau of Investigation (1994–2007). *Uniform Crime Reports: Crime in the United States*, U.S. Printing Office, Washington, DC.

[21] National School Safety Center (2008). *School Associated Violent Deaths*, Westlake Village, Retrieved http://www.schoolsafety.us/pubfiles/savd.pdf.

[22] Dinkes, R., Cataldi, E.F., Lin-Kelly, W. & Snyder, T.D. (2007). *Indicators of School Crime and Safety: 2007 (NCES 2008-021, NCJ 219553)*, U.S. Departments of Education and Justice, U.S. Government Printing Office, Washington, DC.

[23] Brener, N., Lowry, R., Barrios, L., Simon, T. & Eaton, D. (2004). Violence-related behaviors among high school students – United States, 1991–2003, *Morbidity and Mortality Weekly Report* **53**, 651–655.

[24] Van Dyke, R.B. & Schroeder, J.L. (2006). Implementation of the Dallas Threat of Violence Risk Assessment, in *The Handbook of School Violence and School Safety: From Research to Practice*, S.R. Jimerson & M.J. Furlong, Eds, Erlbaum, Mahwah, pp. 603–616.

[25] McCann, J.T. (2002). *Threats in Schools: A Practical Guide for Managing Violence*, The Haworth Press, Binghamton.

[26] Trump, K.S. (2000). *Classroom Killers? Hallway Hostages? How Schools can Prevent and Manage School Crises*, Corwin Press, Thousand Oaks.

[27] Cornell, D. & Sheras, P. (2006). *Guidelines for Responding to Student Threats of Violence*, Sopris West, Longmont.

[28] Heilbrun, K. (1997). Prediction versus management models relevant to risk assessment: The importance of legal decision–making context, *Law and Human Behavior* **21**, 347–360.

[29] Cornell, D., Sheras, P., Kaplan, S., McConville, D., Douglass, J., Elkon, A., McKnight, L., Branson, C. & Cole, J. (2004). Guidelines for student threat assessment: Field-test findings, *School Psychology Review* **33**, 527–546.

[30] Kaplan, S. & Cornell, D. (2005). Threats of violence by students in special education, *Behavioral Disorders* **31**, 107–119.

[31] Mayer, G.R. (1995). Preventing antisocial behavior in the schools, *Journal of Applied Behavior Analysis* **28**, 467–478.

[32] Osher, D., Dwyer, K. & Jackson, S. (2004). *Safe, Supportive and Successful Schools: Step by Step*, Sopris West, Longmont.
[33] Sprague, J.R. & Horner, R.H. (2006). Schoolwide positive behavioral supports, in *The Handbook of School Violence and School Safety: From Research to Practice*, S.R. Jimerson & M.J. Furlong, Eds, Erlbaum, Mahwah, pp. 413–428.
[34] Wilson, S.J. & Lipsey, M.W. (2007). School-based interventions for aggressive and disruptive behavior: Update of a meta-analysis, *American Journal of Preventive Medicine* **33**(Suppl 2), S130–S143.

DEWEY G. CORNELL

Threat Assessment: Workplace

History

The concept of assessing the validity or meaning of threatening communications must be as old as threats themselves, but the effort to develop a scientific approach to threat assessment is a more recent phenomenon. Threat assessment arose in the second half of the twentieth century as an effort to evaluate the risk posed by individuals who were communicating threats to others and, where the person making the threat had not been identified, to narrow the search for the identification of that person. Threat assessment could not have emerged as a field of endeavor without the emergence in the nineteenth century of psychiatry, psychology, criminology, professional policing, and professional security – the professional disciples from which threat assessment practice continues to draw with greater or lesser fidelity depending on the credentials, training, and experience of those who undertake such work.

Early pioneers whose work forms the basis of interest in what would become known as *threat assessment* are as follows:

- Brussel, a psychiatrist, whose analysis of the anonymous writings and deeds of the "Mad Bomber" in 1956–1957 helped the law enforcement officials to predict his behavior, identify the offender, and arrest him, preventing further bombings [1].
- Macdonald, a psychiatrist, whose monograph on mental patients civilly committed to a mental hospital after making homicidal threats was the first empirical study of death threats [2].
- Miron, a psycholinguist, whose work first achieved national attention in the Patricia Hearst case through the successful analysis, on behalf of the Federal Bureau of Investigation (FBI), of transcripts of seven tape recordings sent by the Symbionese Liberation Army [3, 4].
- De Becker, a private security specialist, was the first to recognize the need to preserve threatening and other unusual communications directed to public figures in the private sector and to urge the government to fund scientific research of such communications [5].
- Park Dietz, a forensic psychiatrist and criminologist, whose interdisciplinary team conducted the first scientific study of threatening and other inappropriate communications to public figures on behalf of the National Institute of Justice [6–8].
- Threat Assessment Group, Inc., of Newport Beach, CA, founded in 1987, which was the first organization specializing in the prevention of workplace violence and the assessment of threats on behalf of employers.

The US Secret Service, the Federal Bureau of Investigation, the US Capitol Police, the US Marshall's Service, the Central Intelligence Agency, and other federal, state, and local law enforcement agencies have developed varying internal capacities to monitor threats communicated to their protectees and to others within their jurisdiction. The primary responsibilities of these public entities are dictated by law and by agency policy, but from time to time representatives of these organizations assess threats on behalf of employers in the private sector, and the federal agencies are often called upon by the local law enforcement agencies seeking guidance on what to tell employers. Because law enforcement must neutrally enforce the law – even where doing so carries risk to one or more of the concerned parties – and because law enforcement has no legal duty to prevent crime or to protect the public, much less any particular employer or employee, the guidance and actions of law enforcement do not always work to

the benefit of the employer. Examples of such problems, from the standpoint of employers, are urging victims of domestic violence to leave their abusers and obtain restraining orders, worsening the odds of an attack at the workplace; interviews of threatening employees by uniformed officers in the presence of their coworkers, families, or neighbors, thus adding stress to the threatening employee and possibly compromising the identity of the informant; or, in some of the more egregious cases, reacting with excessive force to the unfounded rumor that a threat has been made.

In part, not only because of these issues but also because employers value privacy, immediacy, and control, the needs of employers for threat assessment services are today served primarily by small, private businesses that offer fee-for-service consultation. These businesses vary greatly in approach, philosophy, credentials, qualifications, technique, experience, and emphasis, and the consumer has little guidance available to select among them. Private investigators, security consultants, retired police officers, psychologists, psychiatrists, psycholinguists, people with diploma mill credentials, and others with no relevant expertise whatsoever, all offer threat assessment services to the unwary consumer.

Varying Meanings of "Threat Assessment"

The concepts of "threat assessment" and even "threat" are used inconsistently, in part, because many writers use "threat" as a synonym for "danger" or "risk", whereas others use "threat" to refer only to communications that contain some verbal reference to the possibility of interpersonal violence. Moreover, the term *threat assessment* has been used in the corporate security community to describe certain activities that might be more precisely described under other names, which are as follows:

1. *Vulnerability assessment* consists of efforts to determine the vulnerability of various sites, industries, procedures, or equipment to criminal action or other disruptive events. Thus, for example, some security and law enforcement professionals use the term *threat assessment* to refer (i) surveys of which of a corporation's multiple plants are most vulnerable to labor disputes, natural disaster, or other challenges, (ii) evaluations of the vulnerable points in an executive's itinerary, or (iii) studies of the vulnerabilities in an overall security plan.
2. *Intelligence collection and analysis* is actually the development and analysis of information about potential adversaries. Many corporations find it useful – and some find it critical – to develop and analyze intelligence about potentially disruptive or destructive elements (such as terrorist groups, animal rights activists, environmentalists, hackers, or corporate spies). Corporate espionage has proved an enormous drain on the US economy that is beginning to receive the attention it deserves from businesses of all sizes [9] and the FBI's Counterintelligence Domain Program.
3. *Emergency triage* is the rapid determination of the severity of an urgent situation. For example, Cadwell [10] used the terms *threat analysis* and *determination of threat level* to refer to procedures whereby the guards at a nuclear facility decide how to respond to various types of alarm signals that have been activated. Tactical and self-defense instructors sometimes use the term *threat assessment* to describe the process of deciding whether to shoot or not to shoot at an assailant or the process of deciding the sequence in which to shoot multiple assailants. In the information technology community, "threats" refer to various hazards to networks and computers, and the procedures for triaging these threats is sometimes referred to as *threat assessment*.

In the mental health community and elsewhere, there is also confusion between the concepts of "threat assessment" and "violence risk assessment". Traditionally, *violence risk assessment* referred to a mental health professional's evaluation of an individual for the purpose of assessing future violence risk (*see* **Dangerousness: Risk of**; **Bomb Scene Management**). Although psychiatrists and psychologists have long endeavored to evaluate the risk of future violence among those they examine, in recent decades, some have begun to refer to this process as *threat assessment* if the person was referred for evaluation after uttering a threat. Mental health professionals commonly rely on a face-to-face examination as one of their primary assessment tools, so many assume that this is a necessary or desirable component of a violence risk assessment.

Likewise, attorneys and employers unfamiliar with the advances of recent decades may assume that the way to determine whether an employee will become violent is to have the employee examined by a mental health professional as part of a disability evaluation or fitness for duty evaluation. The employer expects to receive a report that provides a valid indication of violence risk, perhaps with recommendations for treatment or other interventions that might reduce the risk. The results of this approach are often disappointing to the employer, who does not realize what information is needed to conduct such an assessment, what information may be shared with the evaluee, or the limits of the mental health disciplines or the particular provider.

Of potentially greater importance is the observation that such evaluations often prolong the employment relationship unnecessarily and may worsen the risks of litigation or violence. In many instances, an equally valid determination of violence and other risks is possible without an interview of the subject, thereby avoiding these hazards.

In practice, many violence risk assessments are conducted by mental health professionals and others without conducting a face-to-face examination or interview of the person whose behavior is of concern. The practice of *indirect assessment* has its roots in such techniques as psychological and psychiatric autopsy (*see* **Psychological Autopsy**), the forensic evaluation of mental state at some time in the past (e.g., testamentary competence or criminal insanity), profiling or criminal behavior analysis by law enforcement, psychobiography, and personality assessment of foreign leaders by intelligence agencies (*see* **Profiles: Psychological and Behavioral**). All these applications of indirect assessment draw on whatever information can be developed about the subject of the analysis and the context of his behavior, which may include personal writings; voice mail messages; videotaped statements or interviews; employment, military, business, criminal, correctional, medical, mental health, and other records; interviews with third parties; published descriptions; historical documents; and other sources of information. The depth of an analyst's experience with such assignments may contribute greatly to the value of an indirect assessment of threats in the workplace.

In reference to interpersonal communications, the term *threat* is most often used to mean an expression of an intention to harm, but this common meaning is the source of great mischief. "Threat assessment" can only have maximal value if applied to a much broader class of inappropriate communications and behaviors than the subset of communications that include a clearly articulated intent to harm. Threat assessment is best conceptualized as the analysis of communicative behavior for four specific purposes:

- *Identification*: the specific identity and location of the author of the communication.
- *Description*: the characteristics and motives of the communicator.
- *Prediction*: the probable future behavior of the communicator.
- *Prevention*: the options available for minimizing the risks of adverse outcomes.

Threatening and other inappropriate communications pose dilemmas to their recipients and to those responsible for protecting the recipients. These messages may be an end in themselves or may be warnings of future harm. In threat assessment, the analyst attempts to draw specific inferences concerning the author of a communication from an examination of the text of messages and all other available data, to make predictions about the subject; and on the basis of these inferences and predictions, to suggest options to reduce the risk of harmful outcomes. Violence is not the only harmful outcome of concern to thoughtful employers, who do not limit their use of threat assessment to articulated threats to behave violently.

An annual survey of *Fortune 1000* corporate security directors has identified a variety of "security threats" of recurrent concern to large US employers. These include workplace violence prevention, business continuity planning, information security, employee selection, employee theft, property crime, terrorism, fraud, and other issues. The priority of these and other concerns varies from year to year, but in 2008 workplace violence prevention was ranked first, as it was from 1999 to 2003 [11].

Threatening Communications in the Workplace

Within the United States, in all large workplaces and in all industries, the primary types of naturally occurring threatening communications are as follows:

- verbal threats (direct, indirect or veiled, or contingent);

- threatening behavior (e.g., a raised fist, a threatening stance, or an outright battery);
- frightening behavior (e.g., stalking, harassment, vandalism, throwing objects, or a psychotic rant); and
- other inappropriate communications or behavior.

Contrary to popular belief, obvious and direct verbal threats are not the most significant warnings of future harm. Other kinds of communications, interactions, and misconduct may carry far more significance than mere verbal promises to harm another, yet clear spoken threats to harm another are more likely to be reported, investigated, and acted upon than the myriad other ways in which those who will become violent signal this risk. Indeed, in some contexts (e.g., letters to celebrities and politicians), the presence of a threat is unrelated to risk or a significant indicator of risk reduction [6].

Moreover, the threatened action is not always interpersonal violence. Other commonly threatened actions include harming someone's reputation or career (e.g., by making false accusations, exposing infidelity or other indiscretions, conveying unfavorable information to management, assigning undesirable shifts or tasks); damaging production or property (e.g., labor strikes, sabotage, and vandalism); damaging the company as a whole (e.g., litigation, unfavorable publicity, and complaints to regulatory authorities); or causing both casualties and economic damage (e.g., arson or bomb threats, threats to tamper with food or drugs, and threats to create nuclear disasters).

In US workplaces today, a majority of injurious assaults and overt verbal threats are reported to management, at least in nonunion environments. In many union environments, the tradition of "shop talk" tolerates large numbers of threats used in daily discourse, making it more difficult to recognize those that violate cultural norms, and there is often great reluctance to inform management of any misconduct. In both union and nonunion environments, a majority of threats that are neither injurious nor obvious goes unreported until the subject does something more serious; at which time old and undocumented stories emerge about past misdeeds for which the employee should have been counseled or disciplined. This is particularly true for misconduct that bears no obvious relationship to violence risk, and which employees may only recognize and report when properly and frequently trained and reminded.

Kidnappings, telephonic bomb threats, aggravated assaults, brandishing of weapons, sexual assaults, threats to use weapons of mass destruction, and extortionate death threats are typically and properly reported to law enforcement. These relatively uncommon events comprise less than 5% of the threats reported to large corporate security departments.

The behaviors constituting 95% or more of workplace threats are either misdemeanor offenses (e.g., terroristic threats, harassment, stalking, disorderly conduct, or trespass) or not crimes at all but merely violations of the rules of the company or of common civility, making it impossible for law enforcement to provide definitive interventions. Many small businesses, lacking access to professional security staff or consultants, turn to the local police department for guidance, where they encounter a highly variable response, ranging from no response at all, to helpful guidance, to action or guidance that worsens the risk of violence. Others turn for guidance to their employment lawyer, who may or may not be sensitive to the risks inherent in these situations when using discipline, restraining orders and similar injunctions, or fitness for duty evaluations.

Threats in the workplace may be communicated by any medium, but in settings other than call centers, the entertainment industry, or the offices of government leaders and other public figures, the majority is communicated in person. In one large manufacturing company keeping good data on reported threats, 67% were communicated through direct, in-person contact, 21% by telephone, 9% through a third party, and 3% in writing.

As one might expect, males threaten other males and females more often than females threaten others. Offender/victim relationships vary by industry, depending in part on the frequency of customer contact and the sex ratio of employees. For example, retailers and banks experience a higher prevalence of customer/employee and intimate partner/employee threats than manufacturers.

A large proportion of the more serious cases handled by large corporations arises from intimate partner violence, a jealous member of a love triangle, or frustrated romantic pursuit. The offender in these cases is often not an employee, but rather an intimate partner of the victim or a customer. Cases with these origins span the full range of verbal abuse,

harassment, intimidation, stalking, overt threats to harm, and attacks of all kinds.

Another large share of cases arises internally, most often involving not only peers threatening each other but also involving threats from employees to supervisors, from supervisors to employees, and from former employees to others. Although epidemiologists and other researchers sometimes label these events as "disputes" in efforts to classify the incidents, that is not often an accurate characterization of what occurred, nor is "dispute resolution" the ideal response in many cases. Most often, an employee reacted in a highly inappropriate manner to a perceived slight, indignity, or frustration. In some cases, there was objective provocation in the form of bullying, baiting, or gruff management to which the employee reacted inappropriately. In others, the perceived provocation was an innocuous event or comment that was misperceived by someone impaired by a personality disorder, another mental disorder, intoxication, or mounting life stressors.

Threat Assessment and Workplace Violence Prevention

Employers undertaking development of a comprehensive program of workplace violence prevention require threat assessment as one part of a larger program that includes

- policies regarding violence, threats, misconduct, weapons, intimate partner violence, and bomb threats;
- a system for reporting, investigating, assessing, and managing all threats, misconduct, and inappropriate behavior;
- access controls;
- a security plan and procedures;
- preemployment screening procedures;
- a drug-free workplace program;
- an employee assistance program;
- critical incident response procedures; and
- training of all employees in their roles in each of the above.

Threats and other inappropriate behaviors and communications must be brought to the attention of those who are capable of properly investigating, assessing, and managing them. The first effort to achieve this occurred in the 1980s after one large employer experienced two mass murders. The affected company implemented a policy in which all threatening employees would be referred to the medical director's office for a fitness for duty evaluation. The company did not then realize that fitness for duty evaluations does not necessarily lead to valid assessments of risk, that the causes of workplace threats are usually not medical, and that other approaches lead to a higher rate of successful outcomes. Unfortunately, this outmoded and ineffective approach has been incorporated into the policies of a number of government agencies and government-regulated industries, such as the nuclear power industry.

Over the past 20 years, a majority of *Fortune 1000* companies has implemented many of the components of a comprehensive program, but there remain many that have not yet devoted resources to appropriate training of employees, supervisors, managers, human resources, and others who ultimately manage threats and other misconduct. Moreover, all of the same problems arise in smaller companies, yet few businesses outside of the *Fortune 1000* have succeeded in developing comprehensive programs.

Employers have used three main approaches to incorporate threat assessment services into their workplace violence prevention programs as follows:

- **Completely internal**

A few large employers have undertaken to develop an internal capacity to assess threats of all kinds, usually by having some members of the security or loss prevention department attend various courses, workshops, and professional meetings. Some of the incentives for this approach are short-term cost containment, reducing the risk of confidentiality breaches, maintaining independence, and perhaps a bid for job security for those with these responsibilities. Problems with this approach include the low odds that sufficient expertise exists internally in all of the relevant disciplines (particularly expertise in indirect assessment and, now that there are so few internal employee assistance programs, expertise in psychopathology and personality) and the fact that this approach keeps all of the liability risk internal, too. In the event that litigation ensues, it would be difficult to defend the failure to consult a highly credentialed expert.

- **Internal with external support**

The most common approach for the largest US employers is to develop internal procedures for

screening and managing most cases, using an outside vendor to provide training for an internal, cross-disciplinary management team; training materials for various audiences in the company; and consultation in selected high-risk, complicated, or unusual cases. This is the model developed collaboratively by 3M Company and Threat Assessment Group, Inc., during the 1990s, and it has been widely benchmarked, implemented, and imitated.

- **Completely outsourced**

Many employers seek guidance from an external vendor whenever they become aware of behavior they regard as threatening. This is particularly common among those with fewer than 10 000 employees or those who do not have internal security or human resources professionals. If an appropriate vendor is selected, which is not a simple task in the current market, the employer may receive excellent guidance at a reasonable cost. Unless coupled with appropriate internal policies and training, however, this approach tends to insure that many threats are missed entirely and that most that receive attention do so at a late stage in the escalation of risk.

Although public attention focuses largely on the highly publicized shooting incidents that capture the headlines from time to time, the problem of workplace violence is much broader, encompassing events such as verbal abuse, bullying, harassment, fears related to intimate partner violence, and threats that occur at a rate in excess of 10% of employees per year. Many large employers have taken important steps to reduce the productivity losses that routinely accompany these events. Nevertheless, even in the United States which has been a leader in addressing these problems, there is a long way to go before maximum benefit can be achieved. Much of the world has yet to even address the challenges or systematically consider solutions.

References

[1] Brussel, J.A. (1968). *Casebook of a Crime Psychiatrist*, Bernard Geis Associates, New York.

[2] Macdonald, J.M. (1968). *Homicidal Threats*, Charles C Thomas, Springfield.

[3] Miron, M.S. & Pasquale, A. (1978). Psycholinguistic analyses of coercive communications, *Journal of Psycholinguistic Research* **7**(2), 95–120.

[4] Miron, M.S. & Douglas, J.E. (1979). Threat analysis: the psycholinguistic approach, *FBI Law Enforcement Bulletin* **48**(9), 5–9.

[5] De Becker, G. (1997). *The Gift of Fear: Survival Signals that Protect Us from Violence*, Little, Brown, Boston.

[6] Dietz, P.E. & Martell, D.A. (1989). *Mentally Disordered Offenders in Pursuit of Celebrities and Politicians*, Report to the National Institute of Justice, Washington, DC.

[7] Dietz, P.E., Matthews, D.B., Martell, D.A., Stewart, T., Hrouda, D.R. & Warren, J. (1991). Threatening and otherwise inappropriate letters to members of the United States Congress, *Journal of Forensic Sciences* **36**, 1445–1468.

[8] Dietz, P.E., Matthews, D.B., Van Duyne, C., Martell, D.A., Parry, C.D.H., Stewart, T., Warren, J. & Crowder, J.D. (1991). Threatening and otherwise inappropriate letters to Hollywood celebrities, *Journal of Forensic Sciences* **36**, 185–209.

[9] Kingstone, B. (2005). *The Real War Against America*, Specialty Publishing, Carol Stream.

[10] Cadwell, J.J. (1983). *Nuclear Facility Threat Analysis and Tactical Response Procedures*, Charles C Thomas, Springfield.

[11] Securitas Security Services USA (2008). *Top Security Threats and Management Issues Facing Corporate America*, Securitas, Parsippany.

PARK DIETZ

Threat Level ***see*** Threat Assessment: Workplace

Three-dimensional Modeling ***see*** Visual Recognition Systems in Identification

Three-dimensional Reconstruction ***see*** Reconstruction: Three Dimensional

Thrill Killing *see* Homicide: Multiple (Behavior)

Time of Death: Insects *see* Entomology

Time of Death Determinations

General Considerations

Estimation of the time since death is a practical task in daily forensic casework. Research on estimating the time since death should always bear this practical task in mind. However, a great number of articles on "estimating the time since death" just describe the time dependence of an analyte or parameter postmortem without being of any value in practice [1–3].

The main principle of determination of the time since death is a calculation of a measurable parameter along a time-dependent curve back to the starting point, the time of death. The characteristics of the curve (e.g., the slope) and the starting point are influenced by internal and external, antemortem and postmortem conditions (Figure 1). Since the initial value and the slope are influenced by so many factors, e.g., the ambient temperature or local temperature at the site of measurement and the estimation of the time since death can only suggest a range and not a precise and accurate point in time.

The current parameters used for estimating the time since death are completely different in nature [3, 4]:

- e.g., body cooling and postmortem lividity are predominantly physical processes;
- rigor mortis and supravital reagibility of skeletal muscle are physicochemical processes;
- putrefaction is mainly based on bacterial processes;
- chemical methods to estimate the time since death are based on metabolic processes, autolysis and diffusion according to the concentration gradient.

The methods used for estimating the time since death are not only different in nature but their scientific value, the underlying scientific background,

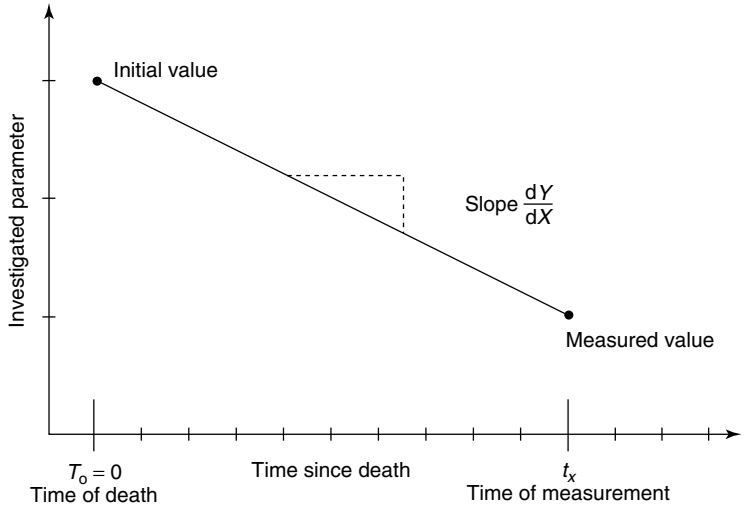


Figure 1 Main principle of the determination of the time since death (calculation from a measured value along a curve back to the initial value)

and the amount of validation of a method differ widely.

- While for some methods of estimating the time since death extensive quantitative measurements with mathematical description of a time change have been carried out taking into account the influencing factors quantitatively and clear data on the precision of estimating the time since death with proof of the precision on independent case material (example: body cooling) are available, other methods are based on a subjective grading of the postmortem change; for further methods, time estimations are based just on empiric data (empiric knowledge, no controlled longitudinal, or cross-sectional studies) instead of statistically evaluated reference values.
- The range of validity of different methods of estimating the time since death concerning experimental background, study type, influencing factors, and statistical evaluation have to be kept in mind in casework. Of highest evidential value are the methods that can show.
- Longitudinal studies on well-characterized reference samples.
- Mathematical description of postmortem changes.
- Clear data on precision of death time estimation.
- Precision of death time estimation proved on independent control sample.
- Field studies on the applicability of the method in practice.

These should replace the other, less reliable, methods of estimating the time since death.

Objectives of Death Time Estimation

Methods of estimating the time since death use two different approaches:

- which premortal changes, either physiological or pathological, can be detected on the body and allow together with investigations by the police conclusions on the time since death (survival time); and
- which postmortem changes allow a conclusion on the time since death.

The first approach comprises methods like wound age estimation or death time estimation from gastric content, normally only very rough estimations are possible.

The second approach is the more important one and comprises the spectrum of methods of death time estimation based on progressive postmortem changes. These are usually used at the scene of crime. However, if time of death and time of assault are not identical, it is necessary to combine both the approaches in order to get an impression of how long an injury was survived and in which interval death occurred after the assault.

The police may use further criminalistic or scene markers that have of course of no scientific background like

- date of last mail or newspapers;
- when was the individual last seen alive by neighbors; and time and date of an opened TV guide.

The medical estimation of the time since death at the scene of crime may have different purposes.

- To give the police a preliminary idea of the time of the assault. However, the time since death will give information on the time of assault only in those instances with a short survival period. This has to be checked on the autopsy findings.
- To check whether the time since death is consistent or inconsistent with the alibi of the suspect.

Only in rare cases timing of death plays a major role at court as the only evidence for or against the guilt of a suspect.

The estimation of the time since death should of course be as precise as possible but even more important is a reliable result. A very precise but unreliable result may mislead police investigations for a considerable time.

Practical Application of Different Methods for Estimating the Time Since Death

As no single method has been shown to produce a precise and accurate result, practitioners generally use a combination of methods to derive an opinion. From all the methods of estimating the time since death, the nomogram method by Henssge based on the body cooling after death is the most intensively investigated, most precise and reliable procedure.[a] Therefore, it should be used in all the cases if

requirements are met and it should be used as the leading method [1–3, 5–19]. However, even in the most favorable case the resulting time since death comprises an interval of ±2.8 h about the mean value. By combining the temperature method with other methods of estimating the time since death efforts should be made to narrow down this range of at least ±2.8 h further.

Methods that can be combined with the nomogram method as a compound method are time-dependent criteria of rigor mortis (beginning, maximum, and reestablishment) [20, 21] and hypostasis (beginning, confluence, maximum, disappearance on thumb pressure, and complete/incomplete shifting) [20, 21], electrical excitability of facial muscles (musculus orbicularis oculi with six degrees, musculus orbicularis oris) [3, 22], mechanical excitability of skeletal muscle (idiomuscular pad) [3, 23], pharmacological excitability of the iris [22]. The issue is how to use the data of these time-dependent classical signs of death with supravital reactions?

The mean values for any criterion of these classical methods do not represent the time interval, in which a positive reaction occurs. For reliable estimations derived from these criteria only the upper or lower limits of these criteria should be used (Figure 2). In Figure 2 mean values, 95% limits of confidence (electrical excitability of orbicularis occuli muscle), or variation (lividity, mechanical excitability, rigor, and chemical excitability) of the different criteria are arranged chronologically over the time since death (x-axis) according to the increasing mean values. As can be seen, the mean value does not really truly represent the time interval in which a positive reaction occurs; a positive reaction may occur at any point between the upper and lower limits of confidence (right and left end of the bar). Therefore, for reliable estimations derived from these criteria only the upper or lower limits of confidence of these criteria should be used.

For casework, the data of the described signs of death and supravital reactions were together with the time of death estimation based on the nomogram rearranged into a special chart which facilitates the choice of the subsequent helpful criteria in an actual case (Figure 3).

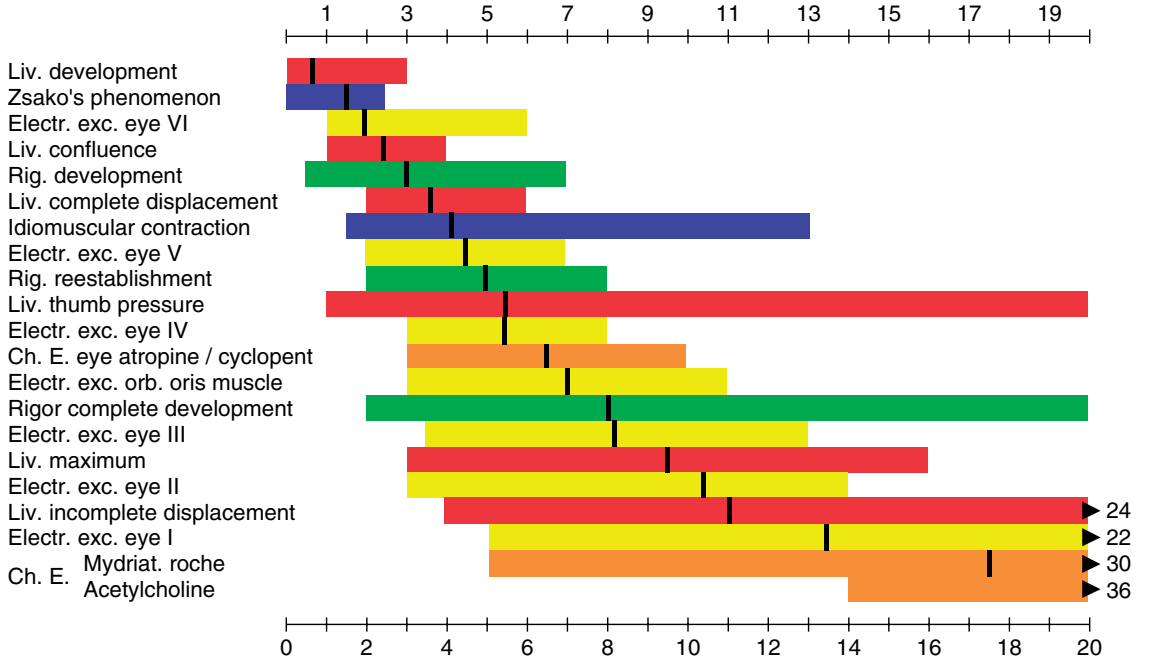


Figure 2 Mean values (black bars) and variations of different time of death dependent criteria. Red, lividity; blue, mechanical excitability of skeletal muscle; yellow, electrical excitability; green, rigor mortis; and orange, pharmacological excitability

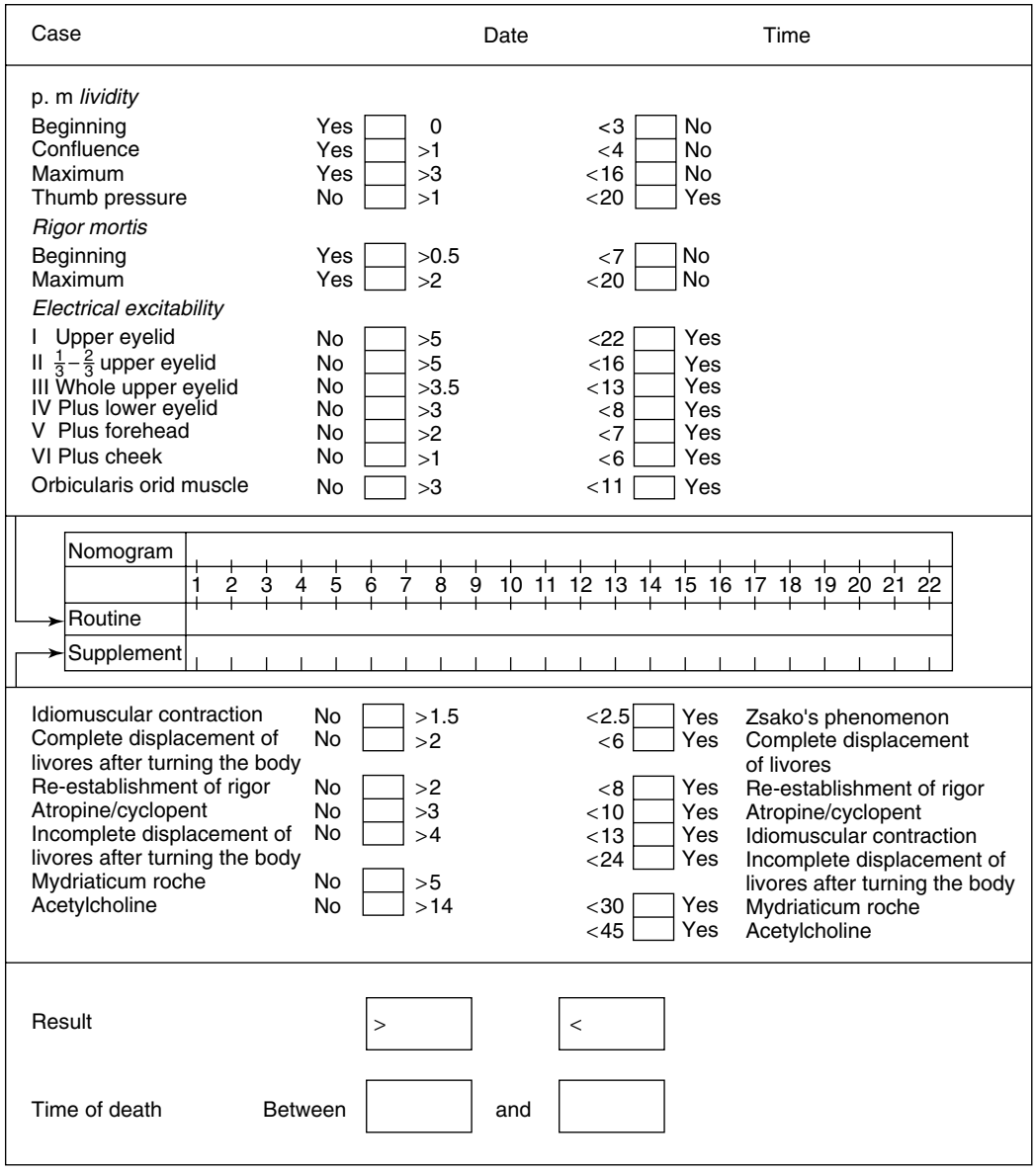


Figure 3 Integrating chart for casework at a scene of crime

In casework, the examination begins with estimating the time since death by the nomogram method after taking temperatures and choosing the appropriate corrective factor. Dependent on the range given by the nomogram method specifically those criteria are examined which are suitable to narrow down the lower and upper margin given by the temperature method. Using this chart at the scene of death will minimize the risk of error, and the inspection and examination of the body should be efficient concerning the time of death estimation.

In Figure 4, a practical example is given. At the scene of crime, the examination began with the measurement of the rectal temperature. Using the nomogram, the estimated time of death was between 4.5 h (lower limit) and 10.1 h (upper limit). The

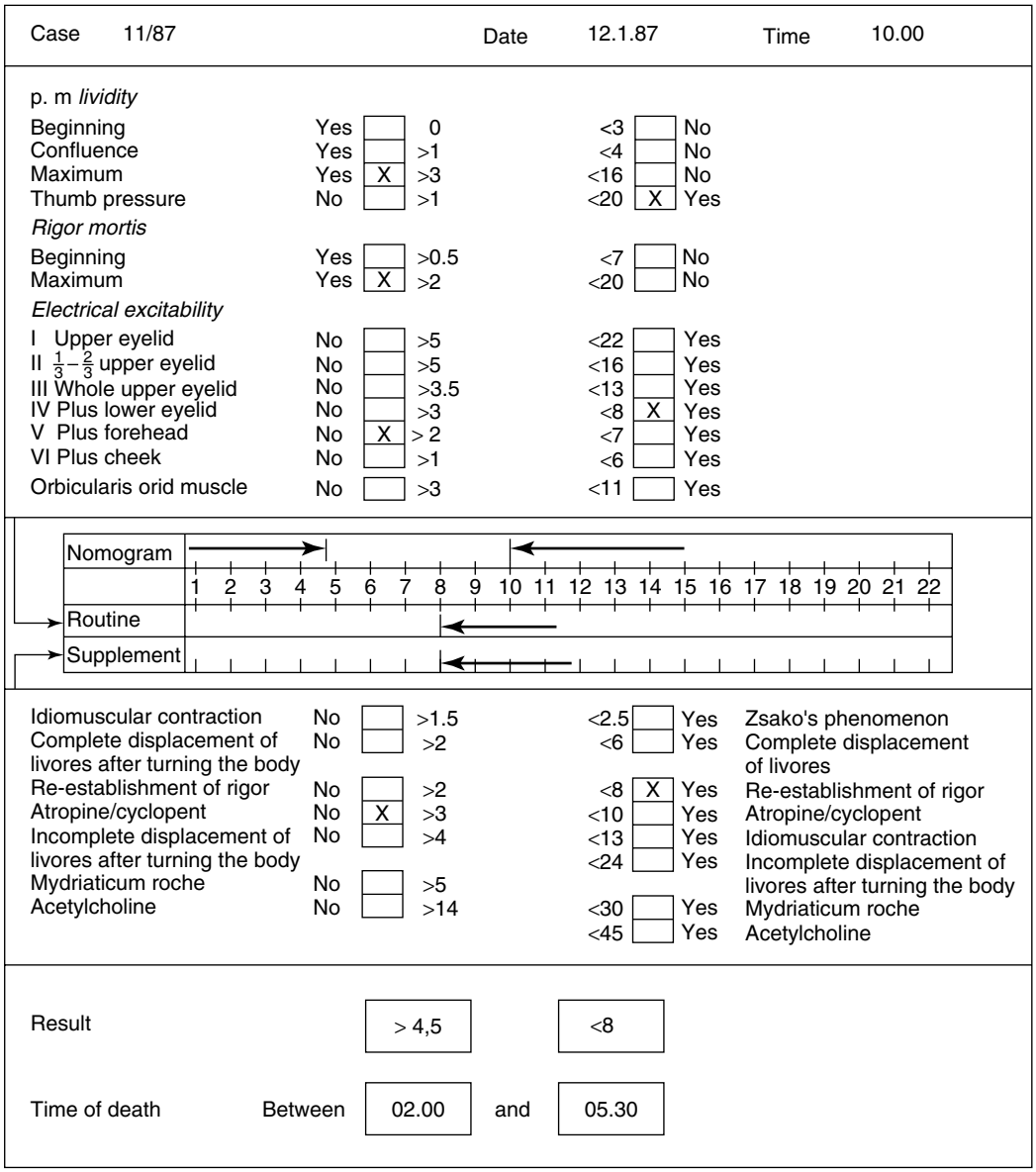


Figure 4 Integrating chart for casework at a scene of crime with an example

lower limit can be confirmed or improved only by a criterion with a higher value than 4.5 h; the upper limit can be reduced only by a criterion with a lower value than 10.1 h. The electrical excitability of facial muscles showed a positive reaction according to degree IV, which means time since death was below 8 h. Furthermore, rigor mortis reestablished after breaking can also only be observed in time intervals up to 8-h postmortem. Therefore, the upper limit of death time estimation (10.1 h) could be reduced to 8 h.

In cases in which temperature methods must not be used alone with this chart (e.g., hypothermia, fire in building where the bodies were found) valuable information on the time since death can still be obtained (Figure 5). In the case of Figure 5,

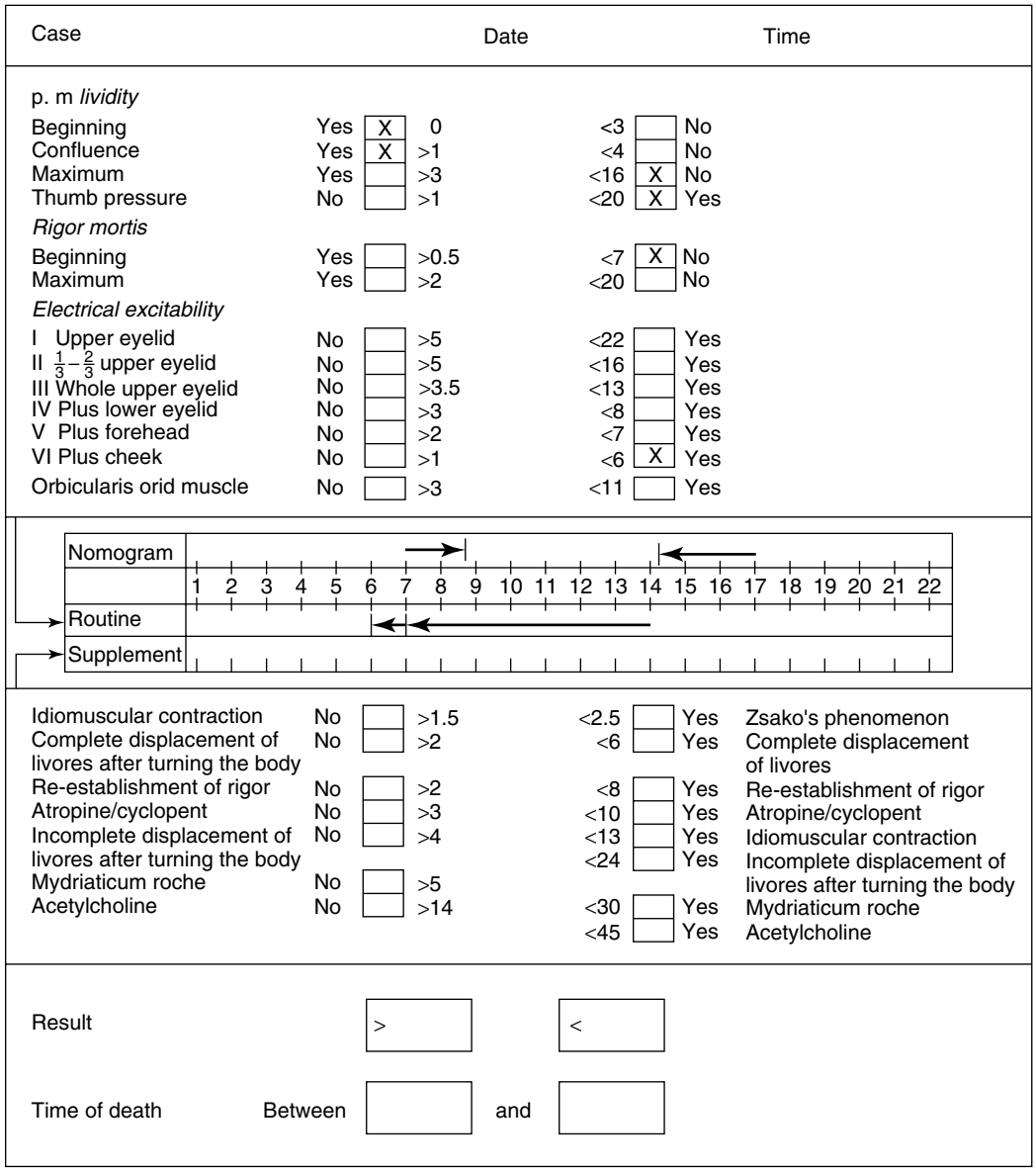


Figure 5 Integrating chart with contradictory results of death time estimation based on body cooling (nomogram), rigor mortis and electrical excitability (case of fatal hypothermia)

the temperature-based nomogram method revealed a death time estimation from 8.8 to 14.3 h based on the rectal temperature of 26.6 °C, ambient temperature 10 °C and body weight of 72 kg. Rigor mortis had not yet established and electrical excitability showed a full reaction (degree VI) resulting in a death time estimation of less than 7 or 6 h, respectively. These results were mutually exclusive. Autopsy revealed hypothermia as the cause of death and this decrease of body temperature during life simulated a long time since death. Therefore, the body temperature must not be used in cases of fatal hypothermia for estimating the time since death, but the other parameters give reliable results.

Our own experiences using electrical excitability besides the nomogram method revealed already that, especially, in the time interval from 3- to 8-h postmortem the combination of both the methods results in a more precise and accurate death time estimation than using one method alone (Figure 6) [24]. Disparate results between the death time estimation derived from the degree of electrical excitability and nomogram may be a hint of special circumstances in the case (for instance hypothermia).

A recent field study on the compound method on 72 consecutive cases over a long postmortem interval revealed that in 49 cases the limits of the period since death estimated by the temperature method were improved by the nontemperature methods [18]. The degree of electrical excitability was the most valuable additional method, but the classical signs may improve the nomogram result as well. Even if the temperature result is not improved but just confirmed this may increase the confidence of the investigator on the time since death, because the conclusion is based on independent methods.

Further Methods

Potassium in Vitreous Humor

Although there is a vast literature on chemical methods for estimating the time since death they do not play any role in practice.

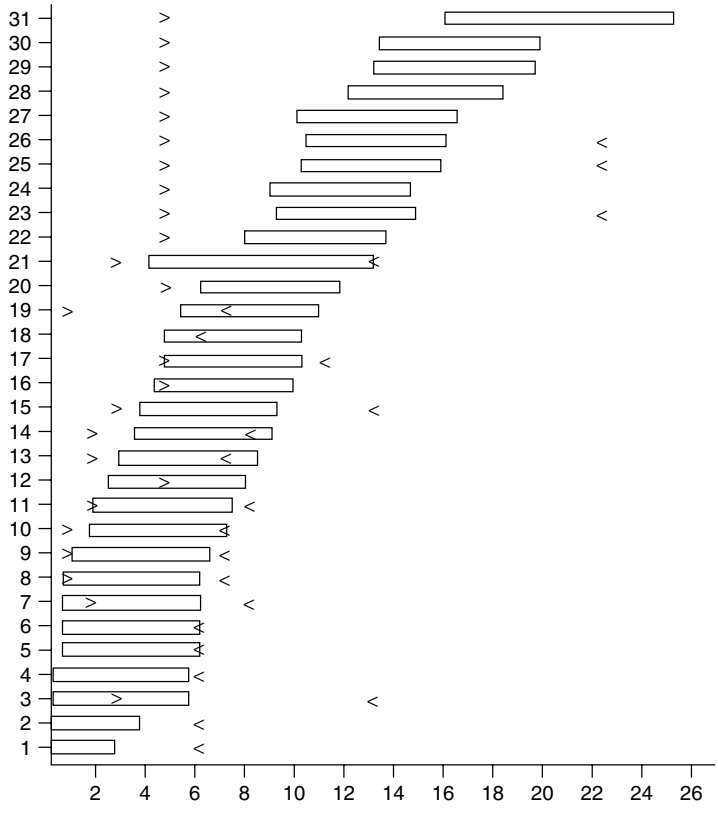


Figure 6 Combined application of the temperature-based nomogram and the electrical excitability for determining the time since death. On the *y*-axis, case numbers. Boxes, death time estimation by the nomogram method (lower and upper 95% limits of confidence). ><, death time estimation based on the degree of electrical excitability. For instance, degree III positive (case 15) not only provides the information that the time since death is below 13 h postmortem but also means degree IV negative; therefore, the time since death is >3 h postmortem. Especially, during the time period from 3–8 h postmortem, combining electrical excitability with the nomogram method allows more precise death time estimation than using one method alone. The real time since death was always within the calculated time since death

After the original description of the postmortem increase of vitreous potassium with very narrow 95% limits of confidence of death time estimation authors found a much wider range, especially when cases with abnormal electrolytes prior to death were included and different statistical parameters of the regression line were described [3, 25–31]. The literature on vitreous potassium has recently been reviewed again [29, 32, 33].

Factors that govern the postmortem rise of vitreous potassium are temperature, chronic illness, and urea retention. Beside temperature rise and scatter of vitreous potassium are determined by the state of health or chronic illness preceding death. Potassium values from individuals dying after chronic illness are much more erratic than those derived from individuals dying of acute trauma. Our own investigations revealed that urea nitrogen is a very suitable indicator of disturbed homeostasis of electrolyte metabolism and depending on the urea values, the corresponding reference sample and formula can be chosen to extrapolate time since death with different 95% limits of confidence (Table 1). In an entire sample including clinical and forensic pathology cases the 95% limits of confidence were ±34 h. By eliminating cases with urea values $>100\,\mathrm{mg\,dl^{-1}}$ and in a second step additionally cases with a terminal episode >6 h, the 95% limits of confidence could be reduced to ±22 and ±20 h. Other factors which are of importance for estimating the time since death by vitreous potassium are of course new methods, e.g., capillary zone electrophoresis or the methodology of pretreatment of vitreous humor before biochemical analysis (e.g., centrifugation, homogenization using enzymes like hyaluronidase, etc.). Statistical aspects for death time estimation have also recently been reviewed [3, 31, 32].

Gastric Contents and Time Since Death

The gastric content alone allows only a rough estimation of the interval between last meal and death. State of digestion and the distribution of the last meal in stomach and upper intestine have long been proposed as a method to estimate the time since death. For the estimation of the time since death the volume of the stomach content compared to the volume of the last meal and transportation distance into the small intestine must be known. Even if the volume of the last meal is not known the type of the meal (breakfast, lunch, and dinner) may allow rough estimations of the daytime when death occurred. Gastric emptying has been studied and quantified in the last decades using different methods (radiological, intubation-aspiration, radioisotopes, ultrasound, absorption-kinetics of orally administered solutes, and ferromagnetic traces) [2, 35, 36], review in [33, 37]. Liquids leave the stomach much faster than solids. While gastric emptying follows obviously for liquids an exponential function for solids show a linear emptying pattern. The following gastric emptying times are given in the literature: 1–3 h for a light small volume meal; 3–5 h for a medium sized meal; and 5–8 h for a large meal. However, it must be kept in mind that different anatomical and functional disorders cause delayed or rapid gastric emptying [33] (Table 2). According to Horowitz and Pounder [35] only the solid compartments of a mixed solid and

Table 1 Precision of time of death estimation using vitreous potassium in different random samples[(a),(b)]

	Urea as inner standard		
	Entire sample	Urea $<100\,\mathrm{mg\,dl^{-1}}$	Change (%)
n	270 (170)	288 (138)	−15.5
Intercept	6.10 (5.99)	6.02 (5.88)	−1.3
Slope	0.20 (0.2033)	0.18 (0.1877)	−10
Correlation coefficient	0.89 (0.86)	0.91 (0.89)	+2.2
Variance S^2	8.57	5.09	−40.6
Standard deviation S_{yx}	2.93 (3.42)	2.25 (2.62)	−23.2
95% limits of confidence (*h*)	±25.51 (±34)	±21.78 (±22)	−14.6

[(a)] Urea was used as inner standard. Statistical parameters of the entire sample and subgroups
[(b)] The data in brackets are from Madea *et al.* [34]

liquid meal should be considered and the weight of the stomach content should be compared with an estimated weight of the last meal and reference made to the known 50% emptying times for the solid components of meals of various sizes.

Tröger *et al.* [37] compared the gastric content (volume) found at autopsy to time and volume of last meal on an autopsy collective of 47 cases (sudden and unexpected death, exclusion of brain tumors, operations of gastrointestinal tract, intoxication, and alcoholization). Gastric contents in percentage of the volume of the last meal was plotted against the survival time. Only a gastric content with a weight over 10 g was considered. Regression line and 90 and 98% confidence limits were calculated (Figure 7). From this graph, the following conclusions can be derived: if at autopsy 50% of the volume of the last meal is found the last food intake was about 3–4 h prior to death with 98% confidence limits not shorter than 1 and not longer than 10 h.

If 90% of the last meal is found in the stomach the last ingestion took place probably within the last hour prior to death with 98% limits of confidence of not more than 3–4 h.

However, conditions for a delayed and rapid gastric emptying have to be kept in mind (Table 2).

Table 2 Etiology of delayed and rapid gastric emptying[(a)]

Transient delayed gastric emptying
Postoperative illness
Acute viral gastroenteritis
Hyperglycemia
Drugs: morphine, anticholinergics, levodopa, beta-adrenergic agonists, and nicotine
Stress: labyrinthine stimulation, cold, pain, and pectin supplementation
Chronic gastric stasis
Diabetes mellitus
Postsurgical – truncal vagotomy with pyleroplasty and antrectomy
Gastro-oesophageal reflux
Anorexia nervosa
Progressive systemic sclerosis
Chronic idiopathic intestinal pseudo-obstruction
Amyloidosis
Myotonia dystrophica
Familial dysautonomia
Dermatomyositis
Tachygastria
Paraplegia
Idiopathic myocardial infarction
Acute abdomen
Laparotomia
Physiological: liquids, acid, lipids, and left-side position
Rapid gastric emptying
After gastric surgery
Vagotomy
Antrectomy/subtotal gastrectomy
Zollinger–Ellison syndrome
Duodenal ulcer disease
Reserpine
Physiological: liquids and hunger

Source:
[(a)] Modified according to Horowitz and Pounder [35] and Tröger *et al.* [37]

Putrefaction

Putrefaction is a bacterial process predominantly influenced by environmental factors, mainly ambient temperature, but by underlying diseases and body proportions as well; advanced stages of putrefaction may be seen within a few hours after death, which are in moderate or cold climates not seen after weeks. Even in relatively constant ambient temperatures the progression of putrefaction varies considerably (Table 3). Recent H-magnetic resonance spectroscopic investigations on metabolites emerging during decomposition of brain tissue have already offered promising results that even decomposition may be used as reliable method of estimating the time since death. However, these longitudinal studies on postmortem changes are still on an experimental level and decomposition under different ambient temperatures has to be studied. A weak point may be that decomposition has been studied in isolated brains and skulls; thus, an invasion of bacteria from the gastrointestinal tract is not possible. There is no sound approach to use the stages of putrefaction as shown in Table 3 for estimating the time since death while the body is lying in air. For bodies recovered from water, a quite good and reliable method for estimating the minimum and maximum water time has been developed based on putrefactive changes, which are visible at external examination or at the dissection of the body [37]. This is mainly due to the fact that water temperature is relatively constant over a longer period of time and during day and night, while air temperature differs not only from day to night but from one day to the other. Morphological findings that are taken into consideration for the estimation of duration of immersion in immersed bodies are as follows:

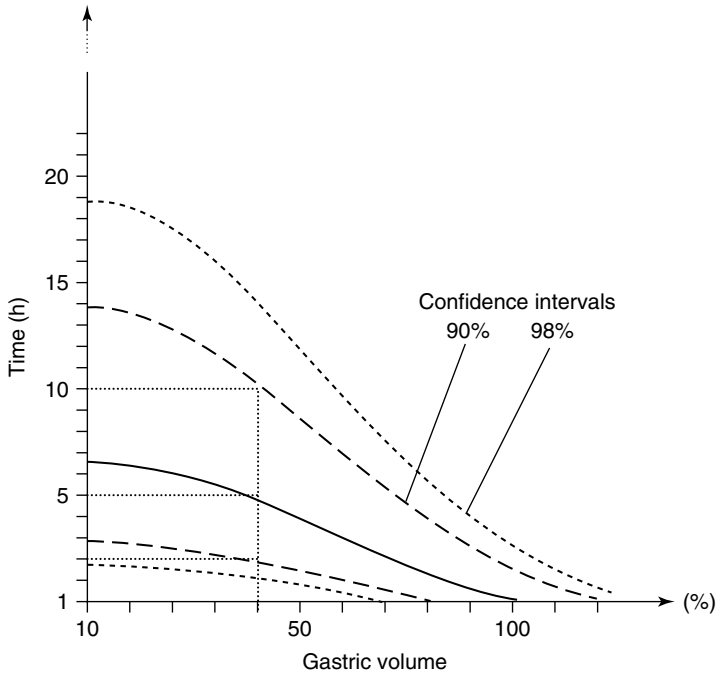


Figure 7 Relation between gastric volume and a mixed meal in percentage of ingested volume and time after ingestion with regression line, 90 and 98% limits of confidence (from Tröger *et al.* [37])

Table 3 Progression of putrefaction of bodies in air temperatures of about 20 °C[(a)]

After 1–2 days
Green discoloration of abdominal wall, softening, and decreased tension of eyeballs
After 3–5 days
Dark-green discoloration of great parts of the abdominal wall
Some patchy green discoloration of the skin, of other body regions
Body fluid leaking out from mouth and nostrils
After 8–12 days
Whole body surface dark green
Face, neck, and thoracic wall partly reddish green
Bloating of abdomen, scrotum, and face
Fingernails still fixed
Hair loose, begins to peel
After 4–20 days
Whole body green or reddish brown
Bloating of the whole body
Blisters, partly filled with putrefactive fluid, partly burst with desiccation of the dermis
Eyes (iris, pupil, and sclera) discolored red brown
Fingernails peeling

[(a)] According to Ref. [38]

External findings

1. rigor mortis
2. lividity
3. marbling
4. bloating of face, scrotum, and subcutaneous tissue
5. discoloration of skin (green, black, and reddish)
6. loss of epidermis
7. loss of hairs
8. hands
   (a) washerwomen's skin
   (b) loosening of nails
   (c) peeling of skin
   (d) loss of nails
9. feet
   (a) washerwomen's skin
   (b) loosening of nails
   (c) peeling of skin
   (d) loss of nail.

Table 4 Chart to estimate the minimum time interval of immersion[a]

	Month	January	February	March	April	May	June	July	August	September	October	November	December
Ø	Median water temperature (°C)	3.5°	3.9°	5.8°	9.9°	13.0°	17.4°	18.6°	18.6°	17.3°	13.2°	8.8	4.7°
1.	Marbling	32	25	16 (23)	9–10	4–5	2	1–2	2	3	4–5	10	17
2.	Distension of tissues by gas	35	25	16 (23)	10	4–5	2–3	2	3	3–4	7	10	17
3.	Discoloration of the body	35	25	16 (23)	(14)	4–5	2	2	3	3–4	7	10	17
4.	Peeling of the epidermis	35	25	16 (23)	(16)	4–5	3	2	3	3–4	7	10	17
5.	Hair lost	35	25	16 (23)	10–12	4–5	2–3	2–3	3	3–4	7	10	17
6.	Hands: beginning of wrinkling	(1)	(1)	(12 h)			(6 h)			2 h		2 h	(1)
7.	Nails become loose	Over 35	28–30 (40)	23	16	5	2–3	3	3	3–4	11	17	28
8.	Peeling of skin in glove form	35	30–32 (45)	23	16	10	3	3	3–4	4	7	20	28
9.	Nails lost	Over 53	45	30 (40)	21	14	8	3	4	10	Over 11	20	Over 35

10.	Feet: beginning of wrinkling	(1)	(1)	(12 h)	(1)		(6 h)	0.30 h		2 h		2 h	(1)
11.	Nails become loose	Over 53	40	26 (35)	17	10	5	3	4	8	12	17	28
12.	Peeling of skin	Over 53	60	35	16	10	5	3	5–6	8–9	Over 11 (14)	20	28
13.	Nails lost	Over 53	Over 60	53	Over 35	Over 28	Over 10	3	Over 10	Over 10	Over 11	Over 20	Over 35
14.	Transsudate in pleural cavity[b]	35	25 (40)	18 (35)	10	5	3–4	3	3	11	5	Over 20	
15.	Heart without blood	Over 39	32–34 (40)	23	14–15	9	4	3	3	5	11	20	28
16.	Brain liquefied	35	30 (40)	(23)	14–16	5	3–4	3	3	6	10	17	28

[a] First line: month; second line: median water temperature for the month; left column: signs of putrefaction and Maceration; following columns: minimal time interval in days. If, e.g., in July marbling, distension of tissues by gas, discoloration of the body, peeling of the epidermis, loosening of finger and feet nails, peeling of the skin of hands and feet, and a liquefied brain are observed, the minimum time interval of immersion would be about 2–3 days
Minimum time of duration of immension. Data in () indicates maximum time
[b] >500 ml in adults

Internal findings

1. volume of transudate in pleural cavity
2. heart without blood
3. liquefaction of brain.

The warmer the water the sooner a definite stage of putrefaction is achieved. From the mean water temperature for each month and the stages of decomposition, the German forensic pathologist Reh [37] developed a chart with the minimum time intervals of immersion. Considering all the 16 parameters for estimating the minimum time of immersion the following chart can be used (Table 4). On the left side the useful criteria, in the first line the months, in the second line the mean water temperature, and in the following line the minimum time interval in days.

As many criteria as possible should be used for estimating the minimum time interval since death. With more than only one or two criteria the result will become more reliable. For estimating the minimum time interval, the mean water temperature that is nearest to the actual water temperature at the time of recovery should be used. With this chart not only the minimum time interval of immersion but also the maximal interval can be estimated, by considering those criteria that have not yet developed. If in June marbling, bloating and discoloration of the body have developed, the nails are loose but not lost, it may be concluded that the interval of immersion is over three days but below eight days. Our personal experience with this chart is quite good since it is much better than the old rules of thumb because it takes the actual temperature for the progression of putrefaction into consideration [32].

End Notes

[a.] This method is described in detail in **Death: Time of**.

References

[1] Henssge, C. & Madea, B. (2007). Estimation of time since death, *Forensic Science International* **165**, 182–184.

[2] Horowitz, M., Maddern, G.J., Chatterton, B.E., Collins, P.J., Harding, P.E. & Sherman, D.J.C. (1984). Changes in gastric emptying rates with age, *Clinical Science* **67**, 213–218.

[3] Madea, B. & Rödig, A. (2006). Time of death dependent criteria in vitreous humor – precision of estimating the time since death, *Forensic Science International* **164**, 87–92.

[4] Madea, B. & Henssge, C. (1985). Historisches zur Todeszeitbestimmung, *Zeitschrift fur Rechtsmedizin* **95**, 19–25.

[5] Albrecht, A., Gerling, I. & Henssge, C. (1990). Zur Anwendung des rektaltemperatur-todeszeit-nomogramms am leichenfundort, *Zeitschrift fur Rechtsmedizin* **103**, 257–278.

[6] Althaus, L. & Henssge, C. (1999). Rectal temperature time of death nomogram: sudden change of ambient temperature, *Forensic Science International* **99**, 171–178.

[7] Henssge, C. (1992). Rectal temperature time of death nomogram: dependence of corrective factors on the body weight under stronger thermic insulation conditions, *Forensic Science International* **54**, 51–56.

[8] Henssge, C. (1979). Precision of estimating the time of death by mathematical expression of rectal body cooling, *Zeitschrift fur Rechtsmedizin* **83**, 49–67.

[9] Henssge, C. (1988). Death time estimation in case work I. The rectal temperature time of death nomogram, *Forensic Science International* **38**, 209–236.

[10] Henssge, C. (2002). Temperature based methods II, in *The Estimation of the Time Since Death in the Early Postmortem Period*, 2nd Edition, C. Henssge, B. Knight, T. Krompecher, B., Madea & L. Nokes, eds, Edward Arnold, London.

[11] Henssge, C. & Madea, B. (1988). *Methoden zur Bestimmung der Todeszeit an Leichen*, Schmidt- Römhild -Verlag, Lübeck.

[12] Henssge, C., Beckmann, E.R., Wischhusen, F. & Brinkmann, B. (1984). A determination of time of death by measuring central brain temperature, *Zeitschrift fur Rechtsmedizin* **93**, 1–22.

[13] Henssge, C., Frekers, R., Reinhardt, S. & Beckmann, E.R. (1984). Determination of time of death on the basis of simultaneous measurement of brain and rectal temperature, *Zeitschrift fur Rechtsmedizin* **93**, 123–133.

[14] Henssge, C., Hahn, S. & Madea, B. (1986). Praktische Erfahrungen mit einem Abkühlungsdummy, *Beitrage zur gerichtlichen Medizin* **XLIV**, 123–126.

[15] Henssge, C., Madea, B., Schaar, U. & Pitzken, C. (1987). Die Abkühlung eines dummy unter verschiedenen bedingungen im vergleich zur leichenabkühlung, *Beitrage zur gerichtlichen Medizin* **XLV**, 145–149.

[16] Henssge, C., Madea, B. & Gallenkemper, E. (1988). Death time estimation in case work II. Integration of different methods, *Forensic Science International* **39**, 77–87.

[17] Henssge, C., Althaus, L., Bolt, J., Freislederer, A., Haffner, H.T., Henssge, C.A., Hoppe, B. & Schneider, V. (2000). Experiences with a compound method for estimating the time since death. I. Rectal temperature nomogram for time since death. II. Integration of non-temperature-based methods, *International Journal of Legal Medicine* **6**, 303–319, 320–331.

[18] Henssge, C., Madea, B. (2004). Frühe Leichenerscheinungen und Todeszeitbestimmung im frühpostmortalen Intervall, in *Handbuch Rechtsmedizin Bd. I*, B. Brinkmann & B. (Hrsg) Madea, Springer, Heidelberg, New York.
[19] Henssge, C. & Madea, B. (2004). Estimation of the time since death in the early postmortem period, *Forensic Science International* **144**(2–3), 167–175.
[20] Mallach, H.J. & Mittmeyer, H.J. (1971). Totenstarre und Totenflecke, *Zeitschrift fur Rechtsmedizin* **69**, 70–78.
[21] Reh, H. (1969). *Diagnostik des Ertrinkungstodes und Bestimmung der Wasserzeit*, Triltsch, Düsseldorf.
[22] Knight, B. & Nokes, L. (2002). Temperature based methods I., in *The Estimation of the Time Since Death in the Early Postmortem Period*, 2nd Edition, C. Henssge, B. Knight, T. Krompecher, B. Madea & L. Nokes, eds, Edward Arnold, London.
[23] Dotzauer, G. (1958). Idiomuskulärer wulst und postmortale blutung, *Deutsche Zeitschrift fur die Gesamte Gerichtliche Medizin* **46**, 761–771.
[24] Madea, B., Herrmann, N. & Henssge, C. (1990). Precision of estimating the time since death by vitreous potassium – comparison of two different equations, *Forensic Science International* **46**, 277–284.
[25] Madea, B. (1990). Längsschnittuntersuchungen zur supravitalen elektrischen Erregbarkeit der Skelettmuskulatur. I. Objektivierung der supravitalen Muskelkontraktion, *Rechtsmedizin* **2**, 107–121.
[26] Madea, B. & Henssge, C. (1990). Electrical excitability of skeletal muscle postmortem in casework, *Forensic Science International* **47**, 207–227.
[27] Madea, B., Käferstein, H., Herrmann, N. & Sticht, G. (1994). Hypoxanthine in vitreous humour and cerebrospinal fluid – a marker of postmortem interval and prolonged (vital) hypoxia? Remarks also on hypoxanthine in SIDS, *Forensic Science International* **65**, 19–31.
[28] Madea, B., Kreuser, C. & Banaschak, S. (2001). Postmortem biochemical examination of synovial fluid – a preliminary study, *Forensic Science International* **118**(1), 29–35.
[29] Madea, B. & Henssge, C. (2002). Eye changes after death, in *The Estimation of the Time Since Death in the Early Postmortem Period*, 2nd Edition, C. Henssge, B. Knight, T. Krompecher, B. Madea & L. Nokes, eds, Edward Arnold, London.
[30] Madea, B., Krompecher, T. & Knight, B. (2002). Muscles and tissue changes after death, in *The Estimation of the Time Since Death in the Early Postmortem Period*, 2nd Edition, C. Henssge, B. Knight, T. Krompecher, B. Madea & L. Nokes, eds, Edward Arnold, London.
[31] Madea, B. & Rödig, A. (2006). Precision of estimating the time since death using different criteria of excitability, *Forensic Science, Medicine, and Pathology* **2**(2), 127–133.
[32] Madea, B. (2005). Is there recent progress in the estimation of the postmortem interval by means of thanatochemistry? *Forensic Science International* **151**, 139–149.
[33] Madea, B. & Henssge, C. (2003). Timing of death, in *Forensic Medicine: Clinical and Pathological Aspects*, J. Payne-James & A. Busuttil, eds. Greenwich Medical Media, London, 91–114.
[34] Madea, B., Henssge, C., Hönig, W. & Gerbracht, A. (1989). References for determining the time of death by potassium in the vitreous humour, *Forensic Science International* **8**, 231–243.
[35] Horowitz, M. & Pounder, D.J. (1985). Gastric emptying – forensic implications of current concepts, *Medicine, Science, and the Law* **25**, 201–214.
[36] Madea, B., Oehmichen, M. & Henssge, C. (1986). Postmortaler transport von mageninhalt, *Zeitschrift fur Rechtsmedizin* **97**, 201–206.
[37] Tröger, H.D., Baur, C. & Spann, K.W. (1987). *Mageninhalt und Todeszeitbestimmung*, Schmidt- Römhild, Lübeck.
[38] Naeve, W. (1978). *Gerichtliche Medizin für Polizeibeamte*, Kriminalistik, Heidelberg.

Further Reading

Lange, N., Swearer, S. & Sturner, W.Q. (1994). Human postmortem interval estimation from vitreous potassium: an analysis of original data from six different studies, *Forensic Science International* **66**, 159–174.
Madea, B. (1992). Estimating time of death from measurement of electrical excitability of skeletal muscle, *Journal - Forensic Science Society* **32**, 117–129.
De Saram, G.S.W., Webster, G. & Kathirgamatamby, N. (1955). Post-mortem temperature and the time of death, *The Journal of Criminal Law and Criminology* **46**, 562–577.
Madea, B. (1994). Importance of supravitality in forensic medicine, *Forensic Science International* **69**, 221–241.
Mallach, H.J. (1964). Zur Frage der Todeszeitbestimmung, *Berliner Medizin* **18**, 577–582.
Madea, B. (1993). Längsschnittuntersuchungen zur supravitalen elektrischen Erregbarkeit der Skelettmuskulatur. II. Quantifizierung der supravitalen Muskelkontraktion, *Rechtsmedizin* 1993, **3**, 44–50.
Klein, A. & Klein, S. (1978). *Todeszeitbestimmung am Menschlichen Auge*, MD Thesis, Dresden University.
Madea, B. (2001). Estimation of duration of immersion, *Scandinavian Journal of Forensic Medicine* **8**(1), 4–9.

Related Articles

Postmortem Interval: Anthropology

BURKHARD MADEA

Time Since Death *see* Human Remains and Identity

Tire Impressions

Introduction

Locard's exchange principle contends that every contact, no matter how slight, will leave a trace. Thus when two objects come into contact with one another, trace materials are exchanged between the two materials. This principle can be applied to a variety of objects at a crime scene. In this article, we will look at the implications on tire marks in particular. These marks can potentially be significant because vehicles play and important role in 75% of all major crimes in today's society [6]. Tire marks are found in cases involving the movement of suspects away from the crime scene and/or a change of vehicle and/or in cases that occur far from housing areas. Tire marks have been used in may cases to either implicate or more importantly exonerate suspects related to crime scenes. Tire marks mostly occur when a vehicle proceeds to drive over a pliable surface like mud, sand, snow, dirt roads, etc. These marks are called tri-dimensional marks.

Bi-dimensional marks, usually occurring on smooth surfaces, are often found in a garage or shed and are often incomplete, following the path of the vehicle on a pool of oil, paint, or blood [7, 8]. They can be composed of a variety of materials as the tire passes through the material and then deposits it on the smooth surface as a two dimensional mark [9].

The way in which tire marks are looked for, detected, protected, recovered, examined, and interpreted are very similar to the forensic examination of shoemarks.

Scene Investigation

As with any forensic evidence, tire mark examination starts at the crime scene. For an investigator it is extremely important to process and document the marks thoroughly. Sketches, photographs, and casts should be obtained in order to reconstruct events and place the vehicle in context of the crime that took place. It is important to determine the direction the car was moving in as well as differentiating between the rear and front wheels and to take note of the required measurements [7, 8]. Nothing should be overlooked and all mark evidence should be treated as valuable pieces of information [10].

When entering a crime scene and in particular surrounding areas, great caution should be taken in order to detect and protect such marks because they are extremely fragile. They should be recorded photographically immediately after they have been discovered using different light sources in order to maximize contrast. They should then be lifted using gelatin lifters by experience technicians. When processing a crime scene many vehicles are active around the scene and it is thus of utmost importance to know which marks are known and which marks are under further investigation [11].

The lifting of the mark is a critical part of processing the evidence. Because marks may be found in a variety of mediums and thus different media holding the mark together, different recovery methods may be used. It is very important that a trained technician makes the decision as to what method will be used to make an accurate lift [4]. Many of the lifting methods and theories can be taught but a large amount of information can only be obtained by hands on experience. Many new methods for lifting is available in the literature. Wilson proposed using expandable polyurethane foam for producing 3-D test prints of tires in question [12]. Buck *et al.* demonstrated that the non destructive method of 3-D optical surface scanning by GOM ATOS 2 system, delivers more detailed results and higher accuracy than the conventional casting techniques [13]. The most commonly used method for casting is still using dental stone. This particular method is discussed in detail by Bodziak and Hammer [14] (Figures 1 and 2).

The tire evidence is mostly used for comparison purposes. The comparison is made between an evidentiary tire imprint and a tire of known source [7, 8]. They are also valuable in obtaining information regarding the sequence of events at the crime scene. Information regarding the vehicle and the actions of that particular vehicle at the scene can be obtained. As a wheel rolls forward, a layer of ground on which the wheel is moving is compacted. As the wheel moves



Figure 1 Dental stone being applied to a mark [15] [Reproduced with permission from Ref. 15. © Canadian Society of Forensic Science.]



Figure 2 Cast of tire mark using dental stone [15] [Reproduced with permission from Ref. 15. © Canadian Society of Forensic Science.]

forward, this dirt is then lifted in the direction the wheel is going [16]. The use of physics and logical thinking helps in establishing the direction the vehicle was moving. For example, as a vehicle moves through water, the tracks will slowly diminish in the direction the vehicle is moving.

Examination and Value of Tire Marks

Tire marks are very important as the tire contains elemental pattern features that enable manufacture and acquired characteristics to be identified. Of particular interest are " sipes" that distinguish design elements. Sipes are very small raised areas located on the design elements [11]. The design of these components is very important when studying a suspect's tire mark.

The elemental tread patterns illustrated in Figure 3 are very important manufacture (class) characteristics that could enable the investigating officer to determine what type of tire created the mark. Tire marks can lead to the identification of a brand name and possibly the model of the tire [7, 8]. It is possible to obtain a highly detailed three-dimensional representation of the tire by casting the tire mark. This will demonstrate the dimensions, design and wear of the tire. If the quality of the tire mark is high, many manufacture and even acquired characteristics (some of which can be truly individual) will be visible [17]. Manufacture characteristics enable us to link the tire mark to a certain model produced by a certain manufacturer.

The investigative value of tire marks has been clearly stated in [17]: "The ultimate goal of tire track investigation is the identification of the vehicle producing the track. If sufficient tire track information is present, careful and accurate measurements can yield both the manufacturer and the production year of the suspects vehicle."

Over the years many different databases have been created to help identify the brand of tire that a particular tire mark originated from. These include *Tread Design Guide* by Tire Guides, Inc. and *Who Makes It and Where released* on CD-Rom by Pearl Communications. This particular CD contains 18 000 photographs of tires and information regarding the manufacturers [4]. Another very useful tool is a database created by the Michigan State Police and the Royal Canadian Mounted Police [11].

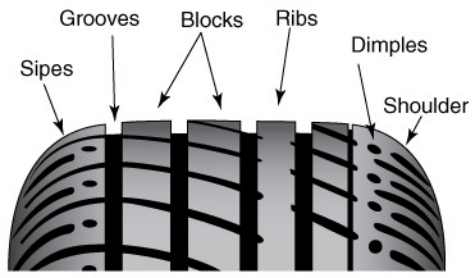


Figure 3 Elemental patterns visible on a tire [18] [Reproduced from Ref. 18. © Elsevier, 2003.]

(c) 2011 John Wiley & Sons, Inc. All Rights Reserved.

Acquired characteristics are those that are randomly produced during the life of the tire, and, in their totality, they enable the investigator to link a tire mark to a particular tire [11]. The terms manufacture and acquired characteristics are increasingly used instead of respectively class and individual characteristics to reflect the actual source of these characteristics and also to depict their identifying value in a more accurate manner (some acquired features, e.g., wear pattern, may be very valuable in the identification process without being truly individual).

Overall, the identification process and scientific underlying theories are similar to those presented in the series of article about shoemarks:

- The tire mark under investigation and the test tire print or cast present manufacture characteristics that are indistinguishable from each other and there are insufficient acquired characteristics to reach a definite conclusion regarding the identification. In this case, the mark could have been left from a particular tire (i.e., nonexclusion).
- The tire mark under investigation and the test tire print or cast present manufacture and acquired characteristics that are indistinguishable from each other. Depending on the quality of the mark and on the quantity and randomness of the acquired features, it may be possible to conclude that the particular tire did leave the mark found at the scene (i.e., identification). In some cases, a qualified opinion may be given, where the value of the mark will depend on its quality and on the rarity of the tread pattern, construction and/or size of the tire, combined with the presence, or not, as well as the number of random acquired features.
- The tire mark under investigation and the test tire print or cast present at least one meaningful (or unexplainable) difference and the mark could not have been left from a particular tire (i.e., exclusion). In this case, it is important to determine that the highlighted difference or differences cannot be due to differences in the life history of the mark compared to that of the suspected tire (e.g., major cut acquired by a tire after it had left an evidential mark).

Tire Tracks

The tracks that tires make are also of considerable value to an investigator. A standard tire track is composed of at least one set of tracks from the same vehicle. They have the potential to show many details of a particular vehicle that a single tire mark cannot. The advantage of tire tracks lies in their ability to give the investigator information about the vehicle itself by displaying the dimensions between the tracks.

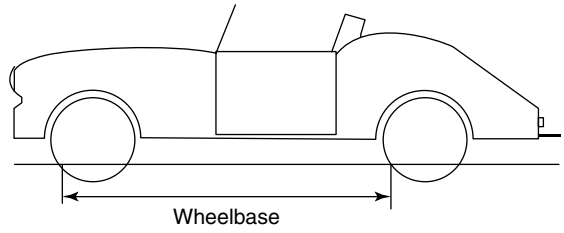


Figure 4 Wheelbase of a vehicle [7, 8] [Reproduced from Ref. 7. © Taylor and Francis Group, 2005.]

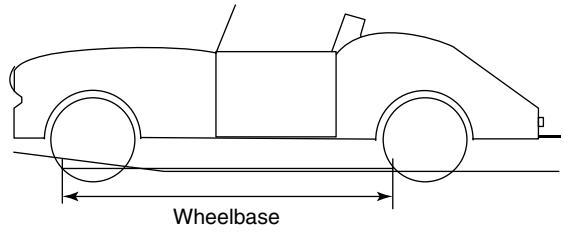


Figure 5 Wheelbase of a vehicle [7, 8] [Reproduced from Ref. 7. © Taylor and Francis Group, 2005.]

Tire marks have the ability to give the investigator critical information regarding the actual tire that left the mark while the tire tracks give the investigator critical information regarding the vehicle that left the tracks. A skilled tire technician can use information obtained through the study of tire tracks to determine the vehicle's turning radius, track width, size type and wheelbase [11] (Figures 4 and 5).

The total length and the relative position of the mark in relation to the rest of the crime scene are also important. Multiple marks can enable the investigator to reconstruct events and draw a conclusion as to the direction the car moved [7, 8] (Figure 6).

Skid Marks

Another extremely important mark that involves tires is a skid mark. These marks offer valuable

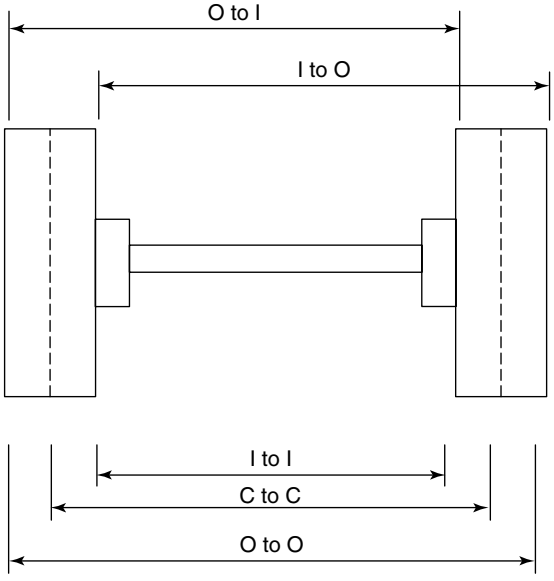


Figure 6 Different methods of obtaining tire track stance [7, 8]. I, inside; O, outside; and C, center [Reproduced from Ref. 7. © Taylor and Francis Group, 2005.]

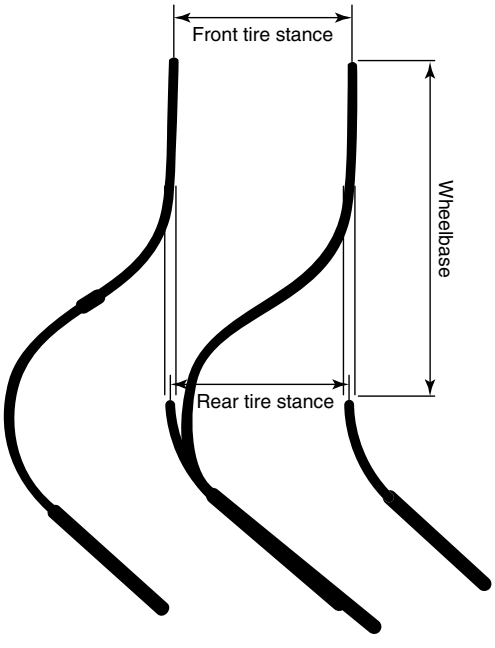


Figure 7 One method for obtaining wheelbase and tire track stance measurements [7, 8] [Reproduced from Ref. 7. © Taylor and Francis Group, 2005.]

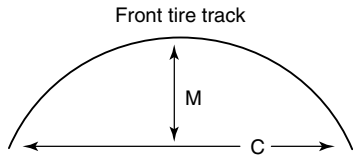


Figure 8 Measurement of turning diameter [4] [Reproduced with permission from Ref. 4. © Perason Education Ltd, 2007.]

information regarding speed and direction of travel. These marks are routinely analyzed at hit and runs as well as homicides and burglaries [6]. At the moment these marks do not offer much information regarding the vehicle but is more useful in reconstructing the event. Research however has been conducted into the analysis of tires and tire traces using chemical analysis; e.g., fourier transform infrared spectroscopy (FTIR) and pyrolysis- gas ghromatography mass spectrometry (Py-GC/MS) [2, 14, 19]. It is important to obtain particular information by using the skid marks. For example, circumference of the tire, tread width and turning diameter [4]. If one knows both the tread and the width, the manufacturers can determine the size of the tire [17]. The turning diameter can also be determined.

The chord (C) is measured followed by the distance from the centre of the chord to the closest point on the art [4].

Skid marks are also commonly used in vehicular accident reconstruction because their size and shape assist the investigators to estimate the vehicle speed and forces of acceleration or deceleration [17] (Figures 7 and 8).

References

[1] www.worldwatch.org/node/1537 Homepage: www.worldwatch.org (last accessed 18 Oct 2008).

[2] Bessman, C.W. & Schmeiser, A. (2001). Survey of tire tread design and tire size as mounted on vehicles in Central Iowa, *Journal of Forensic Identification* **51**(6), 587–596.

[3] http://www.bridgestone.eu/bfe/v/index.jsp?vgnextoid=00000000000000000000000000000001368RCRD Homepage: http://www.bridgestone.eu/bfe/index.jsp (last accessed 18 Oct 2008).

[4] Keppel, R.D., Brown, K.M. & Welch, K. (2007). *Forensic Pattern Recognition*, Pearson Education Ltd, pp. 49–62.

[5] McDonald, P. (1989). *Tyre Imprint Evidence*, Elsevier North Holland, New York, pp. 10–11.
[6] Fisher, B., Svennson, A. & Wendel, O. (1981). *Techniques of Crime Investigation*, 3rd Edition, Elsevier North Holland, New York.
[7] Bodziak, W.J. (2005). Forensic footwear evidence, in *Forensic Science: An Introduction to Scientific and Investigative Techniques*, 2nd Edition, S. James & J. Nordby, eds, Taylor and Francis, New York, pp. 361–375.
[8] Bodziak, W.J. (2000). *Footwear Impression Evidence: Detection, Recovery, and Examination*, 2nd Edition, CRC Press, New York.
[9] McDonald, P. (1989). *Tyre Imprint Evidence*. Elsevier North Holland, New York.
[10] Bessman, C.W. & Scmeiser, A. (2001). Surbey of tyre tread design and tyre size as mounted on vehicles in Central Iowa, *Journal of Forensic Identification* **51**(6), 587–596.
[11] www.raceandroad.com/images/ns-tyrediagram.gif http://www.raceandroad.com/. Homepage: http://www.raceandroad.com/ (last accessed 18 Oct 2008).
[12] Buck, U., Albertini, N., Naether, S. & Thali, M.J. (2007). 3D documentation of footwear impressions and tyre tracks in snow with high resolution optical surface scanning, *Forensic Science International* (In Press).
[13] Bodziak, W.J. & Hammer, L. (2006). An evaluation of dental stone, traxtone, and crime-cast, *Journal of Forensic Identification* **56**(5), 769–787.
[14] Hsiu-Jung, H., Jung-Nan, O., Tseng-Long, K., Suh-Huey, W & Shiea, J. (2007). Using electrospray-assisted pyrolysis ionization/mass spectrometry for the rapid characterization of trace polar components in crude oil, amber, humic substances, and rubber samples, *Rapid Communications in Mass Spectrometry* **21**(3), 375–384.
[15] The analysis of tyres and tyre traces using FTIR and Py-GC/MS, *Canadian Society of Forensic Science Journal*.
[16] Wilson, J.D. (2004). Casting tyres with expandable polyurethane foam and other materials, *Journal of Forensic Identification* **54**(2), 158–169.
[17] Given, B.W., Nehrich, B.R. & Shields, J. (1977). *Tyre Tracks and Tread Marks*, Gulf Publishing, Houston, pp. 18–19.
[18] Abbas, A. & Rutty, G.N. (2003). Forensic web watch, *Journal of Clinical Forensic Medicine* **10**(2), 129–131.
[19] Park, S., Kim, T., Choi, Y. & Chae, S. (2006). Analysis of two tyre marks on the head and clothing, *Japanese Journal of Forensic Science and Technology* **11**(1), 125–129.

Further Reading

Clarke, P.D.B. (1972). Forensic tyre examinations – an analysis, *Journal of the Forensic Science Society* **12**(4), 559–566.
Aspegren, B., Carlsson, K. (1999). Shoe/tire impression casting, *Minutiae*, **56**, 8–9.
Bodziak, W.J. (1986a). *Shoe and Tire Impression Evidence, FBI Law Enforcement Bulletin*, U.S. Department of Justice, Federal Bureau of Investigation, July, 1984 (Revised September, 1986).
Brennan, J.S. (1993). Dental stone for casting depressed shoemark s and tyremarks, *Journal of the Forensic Science Society* **23**, 275–286.
Byrd, M. (2000). *Other Impression Evidence*. http://police2.ucredu/otherimpressionevidence.html.
Gualtieri, M., Andrioletti, M., Mantecca, P., Vismaraand, C. & Camatini, M. (2005). Impact of tire debris on *in vitro* and *in vivo* systems, *Particle and Fibre Toxicology* **2**(1), 1.
Hébrard, J. & Du Pasquier, E. (1993). Experimental and comparative study o f new casting materials. *International Symposium on Forensic Aspect of Footwear and Tire Inpression Evidence*, FBI Academy, Quantico.
International Association of Identification. (2000). *Recommended Course of Study for Footwear and Tire Track Examiners*, International Association of Identification, Mendota Heights.
Lambourn, R.F. (1989). The calculation of motor car speeds from curved tire m arks, *Journal of the Forensic Science Society* **29**, 371–371.
McDonald, P. (1993). *Tire Imprint Evidence*. CRC Press, Boca Raton, ISBN: 0-8493-9515-1.
Rathman, G.A. (1992). Tires and toolmarks, *Association of Firearms and Tool Mark Examiners Journal* **24**(2), 146–159.
www.jeffreycrend.com/impressions/impressions (accessed on, (2008).

Related Articles

Footwear and Foot Impressions: Comparison and Identification
Footwear and Foot Impressions: Databases
Footwear and Foot Impressions: Foot Impressions and Linking Foot to Shoe
Footwear and Foot Impressions: Intelligence
Footwear and Foot Impressions: Overview
Toolmarks
Traffic Fatalities

CHARL DU PREEZ AND CLAUDE ROUX

Tolerance to Alcohol and Drugs

see Alcohol: Use, Abuse, Tolerance, and Dependency

Toolmarks

Introduction

Toolmarks examinations are performed by most of the forensic science institutes. Such examinations are closely related to the following subjects: mechanical properties and wear characteristics of condensed matter, manufacturing methods of materials, tools, and machines [1].

Samples of Toolmarks Examinations

Surface marks on various materials are the main objects of toolmarks examinations. They occur as striation and impression marks. Different manufacturing processes of materials result in characteristic volume structures inside the materials, and especially in specific characteristics of the surface of the engineering materials. Examples of surface marks are structures produced by grinding, turning, milling, electro-discharge erosion, laser-beam machining, casting, powder metallurgy, extrusion, and foam molding [1, 2].

Examination Processes

A goal of toolmark examinations is the identification of a questioned sample to the exclusion of all other samples. However, this goal is often not achieved at all or at least not in the form of a clear evidence (then formulated as a "probability statement" in the expertise opinion). Sometimes merely class characteristics can be determined (determination of the type of a tool). In other cases, the determination of the manufacturer, the origin, the use or the function of the questioned sample is desired, or a damage analysis is requested [3]. Table 1 gives a survey of the types of examinations of mechanically induced surfaces (toolmarks).

In general, the sequence of the examination process is as follows:

1. Inspection of the marks and the tool
   (a) Visual inspection of the questioned surface.
   (b) Surface imaging of the working surface of the questioned tool using light microscopy or scanning electron microscopy (SEM) and a subsequent classification of the surface topography into class and identifying characteristics.
2. Verification of the relationship between the marks and the questioned tool.
   Verification of a possible relationship between the marks and the tool by conducting tests with the tool in question under defined conditions; examinations using the comparison microscope or the scanning electron comparison microscope, if necessary, also with the aid of image analysis methods.
3. Reporting the results
   (a) Writing a report of the results (conclusions) of the forensic tests (examinations): expert opinion (court statement).

Table 1 Types of examinations of toolmarks

	Comparative examination of test and evidence marks	
Examination of a single questioned sample	Pattern fit (physical match)	Striated/impression marks (pattern transfer)
Interpretation of surface structures: manufacturing processes, fracture mechanisms, kind of toolmarks *Microstructures*: class and identifying characteristics	Fractures or cracks in condensed matter (brittle, ductile materials), which can be joined unambiguously or with some probability	*Striated marks*: screwdrivers, milling machines, etc. *Impression marks*: punches, etc. *Aim*: establish a link with a specific source (tool)

(b) Considering of at least two hypotheses in the expert opinion:
- The evidence mark was produced by the working surface of the questioned tool (hypothesis).
- The evidence mark was not produced by the working surface of the questioned tool (alternative hypothesis).

(c) Using a "multisteps expert opinion scale" (conclusion scale) including a guideline for the evaluation of the different steps (levels).

Individual Characteristics of the Working Surface of a Tool

Saying a toolmark is caused by a tool may sound trivial. Among the tools submitted to a forensic laboratory are pliers, screwdrivers, hammers, chisels, coining dies, tablet press machines, and cameras. To link a toolmark with a seized tool, however, only a small part of the tool, the working surface, is relevant. The specific type of tool is therefore of only minor importance for forensic examinations. To evaluate the evidence toolmarks and the test marks, the condition of the tool's working surface is the decisive factor.

Technical and Natural Surfaces

The working surface of a tool is a technical surface, originating from technical processes designed by man. In contrast to this, there are natural surfaces formed without the influence of man. Examples of the latter are the surface of a natural stone or a crystal, a shell, a peppercorn, or the bark or leaf of a tree. The state of such surfaces changes permanently. Water, air, wind, and dust as well as temperature influences change their appearance and destroy them gradually, even without the influence of man. As scanning electron micrographs prove, natural surfaces both exhibit very regular and extremely irregular shapes.

Within the scope of technology, "bodies" with a well-defined geometry surface – such as plane or curved surfaces – are produced. However, "surface" in the technical sense has another meaning than the term *surface* in the mathematical sense. It shall be explained using the following example. For a technician the "cylindrical surface of a workpiece" means a more or less rough body with form errors, while the mathematical term *cylindrical surface* merely evokes the idea of a form which can be exactly defined with respect to its geometry and which does not even have to exist in a physical sense.

Random Parameters of a Technical Surface due to the Manufacturing Process

Many materials would have undergone a long manufacturing process before they are in the semifinished stage and they can then be given their final form. Sometimes the original surface of such a semifinished product remains unchanged during the final working stages. In this case, it must already have the geometric characteristics required of the finished product. However, usually an additional forming or cutting cycle is necessary, which is frequently followed by a surface finishing process. All these surfaces that were produced by technical means and that also serve a technical purpose are referred as *technical surfaces*.

Technical surfaces are the surfaces of machines and equipment parts, tools, and facilities that are subjected to certain mechanical, chemical, thermal, or other stresses that finally lead to wear, corrosion, or other types of destruction. In addition, such surfaces may undergo chemical and physical changes, already at the moment of their formation, which may even affect deep structural layers, even if such changes will hardly be visible on, for example, a freshly lathe-cut workpiece. Thus, technical surfaces differ from natural surfaces only in that they are formed under the laws of nature under conditions that are not or only rarely found in nature.

A manufacturing method will always result in a specific workpiece surface that is characteristic of the respective method. Given the wide range of variations, even in one manufacturing method, it is obvious how unforeseeable and manifold the resulting surface condition may be. The generation of the surface profile – especially the generation of the geometric texture of the roughness and of the surface character as a result of a vast number of microscopic random parameters forms – the basis for the forensic evaluation of toolmark examinations.

Examples of microscopic random parameters are the statistical variations of local material characteristics of the workpiece of the tool, the chemical composition, the crystalline texture including the structural characteristics (crystal defects and crystal anisotropy), and the mechanical–technical characteristics (hardness, tensile strength, and elongation).

Random parameters attached to the machining process are the dynamic characteristics of the machine tool, type and sequence of motion, speed, normal force, temperature, cutting thickness, rake angle, and chip dimensions.

All these characteristics must be considered as random parameters, because they are subject to instantaneous variations and are thus continuously influencing and changing the surface topology to a varying extent. In addition, stochastic wear mechanisms, which dull the cutting edge, play a decisive role in the generation of the respective surface roughness and character. In general, the term *wear* as used in materials science describes the stresses that are acting on the surface of a condensed matter through the contact and relative motion of a second matter, including the related material losses (small material particles breaking off in a way unforeseeable in time and space, which are permanently causing various changes of the surface contour).

For this reason, measurement and control engineering as part of the engineering sciences describes and analyzes the roughness of technical surfaces as a stochastic (random) process using statistical methods. When calculating statistical degrees of dependency (correlation analyses) of surface profiles with irregular textures (stochastic profiles, e.g., of machined surfaces), the correlation function attains a very small value (tending toward zero). Such a value – obtained on condition that the examined surface profiles were produced by one and the same manufacturing method – characterizes individual structures.

This is confirmed by forensic "recurrence models" of toolmarks, based on probabilistic approaches [4, 5]. The individual character of the surface profiles of the working surfaces of tools is substantiated on materials science and probabilistic grounds. If the patterns of the evidence toolmark and the test mark cannot be differentiated and scientifically founded individualizing characteristics are present, the evidence toolmark can then frequently be unambiguously linked with the tool (working surface) producing the mark to the exclusion of all other tools.

Application of Probability Theory Models on the Comparison of Striations

Since the results of toolmarks examinations are based on "geographic features", efforts have been made to obtain objective criteria as to whether the degree of similarity is sufficient for identification [1, 4, 5]. The mentioned articles study several probability theory models, based on certain assumptions, e.g., the patterns must be generated by random processes.

Very often the relevant forensic information related to the toolmarks is given by the position, intensity, and the shape of the "curve diagram" of the striation. In [1, 4, 5] models are described mainly based on the position information of the striation (model of intervals or linear cells).

Experimental Methods

Toolmark examinations require methods in which exhibits are analyzed, compared, and finally evaluated. The applied techniques are mainly based on microscopic principles. In many cases, the use of reflected oblique lighting (stereo binocular microscope and comparison light microscope) will be the major requirements. In plastic bag cases, the application of incident and transmitted light microscopy is often necessary, sometimes in combination of shadowgraph and Schlieren photography including "crossed polarizing filters".

Comparison Light Microscopy

Light comparison microscopes are the instruments usually used for toolmark examinations [1]. The images obtained with this method are clear and convincing as they allow side-by-side comparison of the evidence toolmark and the test mark. All information gathered in the laboratory must be presented in such manner that nontechnical people at court will be ready to accept concise and detailed information – yet sophisticated presentations ought to be avoided wherever feasible. Comparison light microscopes (Figure 1) are well-established tools in forensic science institutes all over the world, and jurisprudence will readily accept photographic recordings of this kind. It is expected, that the light-optical comparison microscopy will keep its place in the majority of cases in the future. Figures 2 and 3 show images (examination of vehicle license plates [6]; fracture matching of two broken pieces of plastic [3]), received with such kind of instrument.

However, there are many cases where the question whether or not two specimens are of common origin cannot be answered to full satisfaction. Also, photographic recording often leads to difficult problems.



Figure 1 Comparison light microscope "Leica DMR" allowing direct visual comparison of test and evidence marks (oblique and transmitted lighting). The magnification range is between 16 and 500. By using polarization effects, surface reflections can be eliminated and traces on plastic surfaces – such as adhesive tape – which are hardly visible with oblique lighting can be imaged with high contrast. For comparison of transparent objects, transmitted light device gives homogenous illumination. Nowadays, instead of a video camera a digital camera is used. The microscope is connected to an image analysis system to provide higher magnification and a greater amount of contrast

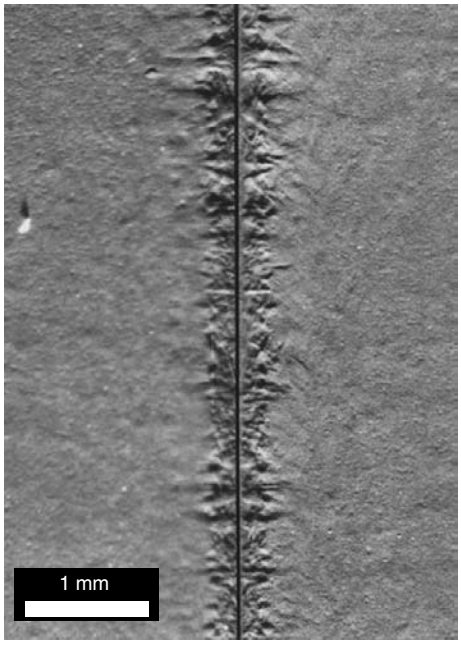


Figure 2 Comparison light microscopy: vehicle plate examination of test and evidence marks. During the embossing process, when the die hits the plate it creates a deformation of the plate of the same shape. If the die bears small defect, it will be reproduced on every plate that was embossed by the particular die. The proper observation of this characteristic and its comparison with the original die may allow for its identification as the source of the impression



Figure 3 Fracture matching (PMMA-plastic; PMMA, polymethyl methacrylate) done with a comparison light microscope showed "a very good match". Some broken pieces were found at the scene of crime (traffic accident). Some parts of the blinker were found at a suspect car. The question was to prove whether pieces of the broken material were attached to another

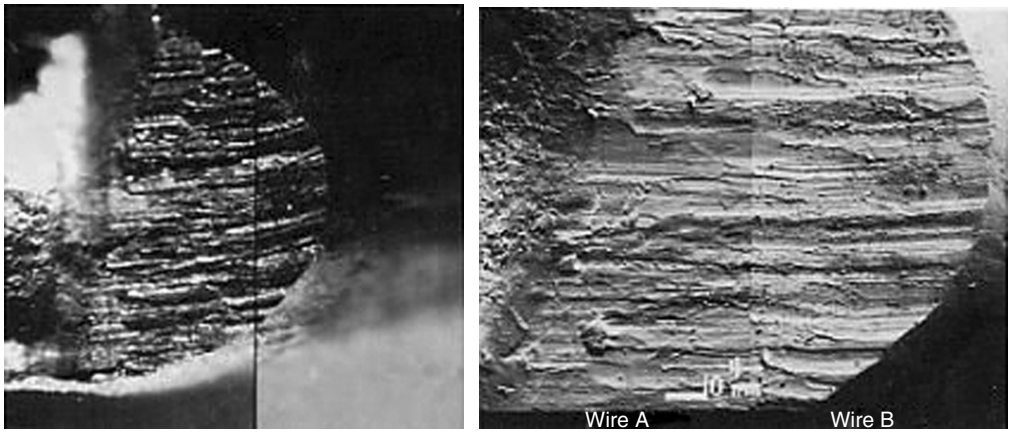


Figure 4 Light microscope (LM) and scanning electron microscope (SEM). The depth of focus and the resolution of the SEM are greater than of the LM. Here are shown the cut surfaces of a copper wire, imaged with a LM (a) and with a SEM (b)

For example, bullets may be severely deformed or their characteristic topography is too faint for recording. Moreover, highly three-dimensional structures or deeply undercut toolmarks are nearly impossible to investigate by light-optical methods. Here, a scanning electron microscope proves to be a very powerful tool, producing images of great clarity with plenty of depth of focus [1, 3] (Figure 4).

Comparison Scanning Electron Microscopy

For objects with a rugged surface or with structures of 1 μm or smaller, the light comparison microscopic method yields only insufficient results. In these cases, a scanning electron microscope with its high depth of focus and high resolution is capable of providing useful images of the structure of the questioned surface [1, 3, 7]. The disadvantage of a single scanning electron microscope is that two specimens which are to be compared cannot be shown at the same time. Instead, each specimen must be mounted and observed individually. After taking micrographs of each one, these images must then be trimmed to size and mounted in such way that comparison is feasible. It is not a simple task to mount two specimens and than record their surface details with similar illumination and at exactly the same magnification. Disregarding the fact that this takes a lot of time, experience also has shown that often important information is simply being overlooked. For the operator, it is a tedious and hard job.

These considerations were brought together, keeping in mind the questions arising in light-optical comparison microscopy. This gave rise to the concept of comparison scanning electron microscopy. Two scanning electron microscopes were electronically interconnected so that the image information from two specimens, mounted in each of the two specimen chambers, could be combined in a variety of presentation modes on a single video screen. Now, the advantages of scanning electron microscopy can be fused together with the above mentioned well established comparison microscopical methods [1, 3, 7, 8]. In Figure 5, a total view of the instrument is shown, and Figures 6–8 show images received with a comparison scanning electron microscope: striation marks of a pair of nippers on two pieces of a copper wire; fracture matching of two broken pieces of a key [3]; and impression marks on Polaroid photographs [9].

Casting Techniques

Casting techniques are used to produce true-to-shape and dimensionally stable reproductions for the examination of surface marks in forensic science. These techniques are used in all cases where an examination of the original surface is impossible or

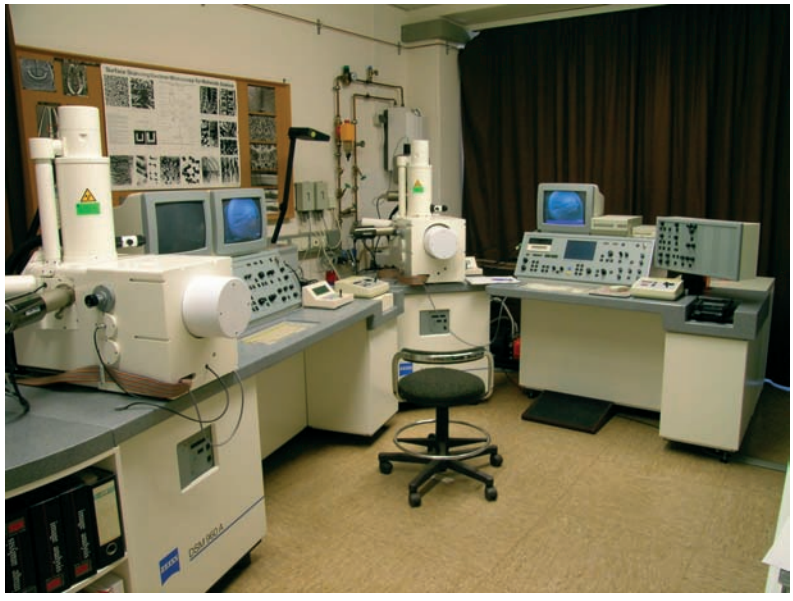


Figure 5 Comparison scanning electron microscope (Zeiss DSM 962/960A). The instrument is equipped with several electron detection systems, which produce special contrast effects and with an EDX-system (EDR 288 W Röntec). A number of interchangeable specimen holders have been designed

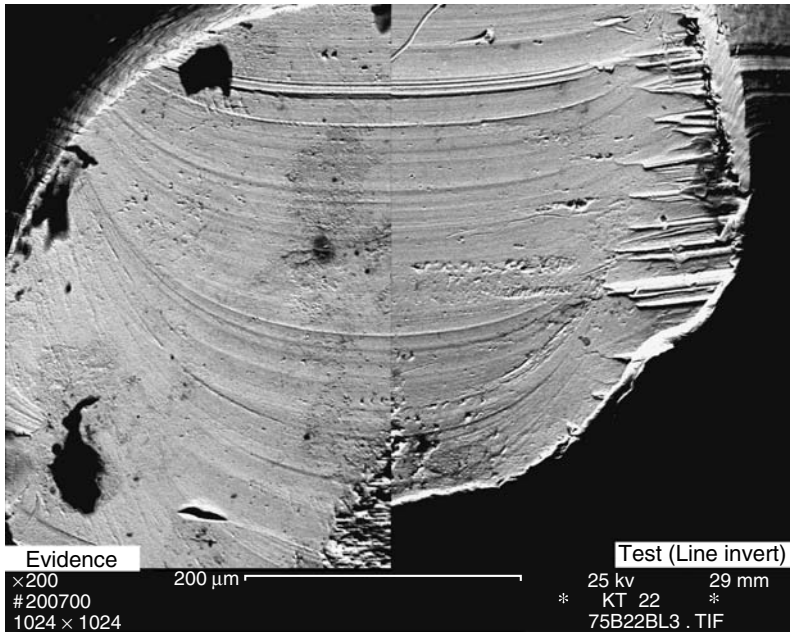


Figure 6 Corresponding striation marks on two cutting surfaces of copper wires produced by the same tool (a pair of nippers); comparison of SEM image

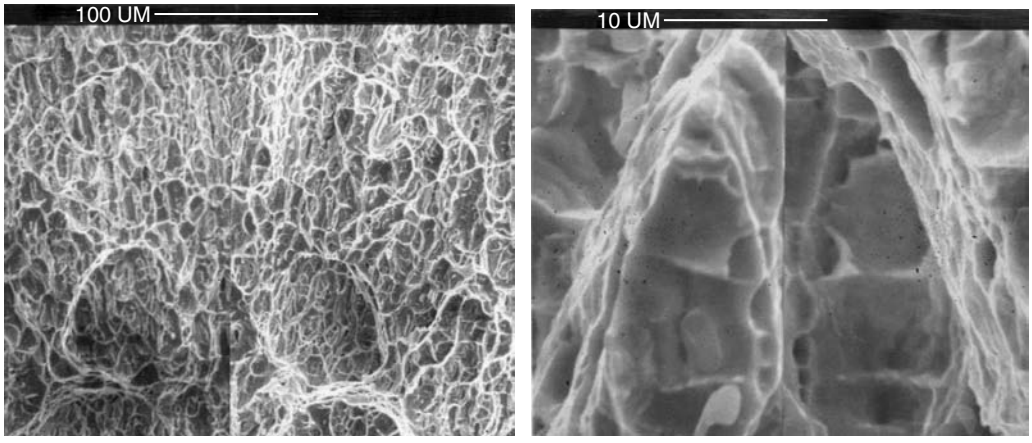


Figure 7 Fracture matching: two broken pieces of a key. The two comparison SEM images were produced with different magnifications. Matching of microstructural surface details of a rupture with ductile features (dimples)

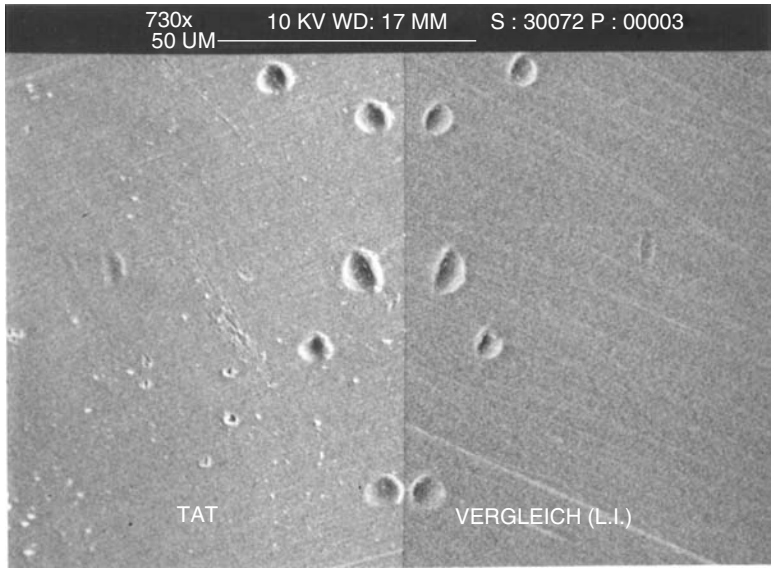


Figure 8 Impression marks on Polaroid photographs. The impressions are made by the steel rollers of the camera. The largest tips of the stalagmitic drops (produced by an electrical-discharge machining process) are pressed in the surfaces of the photographs. The random events of the producing process – shape, size, and location – are in very good agreement to each other. Having all those characteristics in combination, one can say that the recurrence of such pattern constellation would be extremely rare (identifying characteristics). In the report of the result of this kind of forensic test you would write: It was determined that the known impressions in the Polaroid pictures were produced by the questioned camera

where they can simplify or improve examination and documentation. They are also used for collecting and archiving marks, if the original piece of evidence does not lend itself to these purposes. In the field of firearm and tool marks castings are made of

- highly reflecting (shiny) objects to improve the photographic documentation of evidence and
- areas bearing marks that are difficult to examine.

The ability to produce castings of complex three-dimensional structures can be of tremendous

importance for proficiency tests or for additional forensic comparative examinations at other laboratories.

Castings materials used in forensic science and, especially, for light microscope and scanning electron microscope examinations should have the following properties [1]:

- reproduction of the finest details;
- no or only slight shrinkage;
- no inherent structure;
- chemical neutrality toward the surface to be cast;
- easy release from the surface to be cast;
- opacity and high reflectivity;
- high contrast for microscopic examinations;
- high-dimensional stability combined with sufficient elasticity;
- low or no propensity for the formation of bubbles;
- variable processing and curing times;
- low heat buildup during curing;
- freedom from tack; and
- long-term storage suitability.

When introducing a new material for making single and/or double castings, first you have to check the material if the required properties are fulfilled. The reproduction of fine details with no dimensional change is a property of special importance. Therefore, this property of manufactured castings should be tested by using a calibration standard and measured by means of a laser profilometer and a comparison scanning electron microscope. A summary of some results of different casting materials is given in [1]. For making "mass-produced" castings by means of "nontime-consuming techniques" special casting devices were developed [10].

Aspects to the Expert Opinions

A question arising frequently is whether the agreement between two surface profiles found by examination is sufficient to establish a doubtless, unambiguous "physical match", or an unambiguous link with one and the same source of the mark (working surface of the questioned tool). Would another event be able to produce the same surface structures? This problem can be scientifically discussed with the aid of probabilistic methods (Bayes theorem, likelihood ratio (LR), errors of type I and type II, and occurrence models). To estimate an error of type II in the sense of a "recurrence quantity", there are probabilistic models, which assume that the characteristic features used in the calculation can be attributed to random processes. The following discussion describes a concept for the forensic examination of marks based on probabilistic methods [1, 11]. Another section of this encyclopedia provides a thorough coverage of probabilities and statistics in forensic science.

Probabilistic Causality Concept

The question frequently asked by the requesting agency (police, public prosecutor's office, and court) is whether or to what degree of probability a seized tool has caused the toolmark at the scene of crime. Before this question is answered here, some related aspects shall be discussed first.

We are used to think in categories of cause and effect. We note that it starts to rain and expect the road to become wet. We burn dry leaves and expect smoke to rise. We hit a coconut with a stone and the coconut breaks (sometimes). Speaking in general terms: by observing an event C we conclude that a second event E will follow, i.e., a cause C results in an effect E (principle of causality). There is a relationship (a correlation) between C and E.

In mathematical terms, the correlation of two events may also be expressed as a conditional probability $P(E/C)$, put into words: the probability P of the event E in the presence of C. Examples: P(wet road/rain); P(smoke/fire); P(nut broken/hit on the nut); P(toolmark/tool); P(shoeprint/shoe); or – in general, – P(effect/cause) = $P(E/C)$ [direct causal relation]. These examples prove that the conditional probability may be interpreted as the "probability concept of causality". The examples mentioned represent "directly causal" probabilities.

Reverse Causality when Examining the Toolmarks in the Laboratory

We will now look at a scene of crime, where the working surface C of a tool produced a toolmark E : $P(E/C)$. Subsequently, the toolmark E is preserved. A tool is seized from a suspect. The examination request to the forensic laboratory may state: "What is the probability that the toolmark E at the scene of crime was produced by the working surface C of the seized tool?" This question put to the expert deals with "conclusions" or "posteriors", i.e.,

the conclusion from "effect E" to "cause C", i.e., $P(C/E)$. The expert shall, so to speak, draw a backward (retrograde, retrocausal) conclusion from a present effect E to a cause C: The conditioned and the conditioning event are reversed ("reverse causality"), i.e., we have a "retrocausal probability".

Application of Bayes Theorem to Retrocausal and Directly Causal Probabilities

A mathematical formula for the interrelationship between retrocausal and directly causal probabilities is provided by Bayes theorem. For the application of Bayes theorem within the framework of forensic expert opinions, the very elegant "odds form" of Bayes' rule is often used [1, 11–14]:

$$\frac{p(C/E)}{p(\overline{C}/E)} = LR \cdot \frac{p(C)}{p(\overline{C})} \qquad (1)$$

where $LR = P(E/C)/P(E/\overline{C})$.

The left-hand side of the equation is known as the *posterior odds*; the last quotient of the right-hand side as the *prior odds*. The factor that converts prior odds to posterior odds is the *LR*. The *LR* characterizes in a brief form the result of the forensic science examination. If the *LR* has a value greater than one, it lends support to the hypothesis C, and a value less than one lends support to the alternative hypothesis $\overline{C}$.

For toolmark examinations, the "posterior odds" can be interpreted as follows: What is the probability of a toolmark E having been produced by the recovered tool C or a different tool (tool $\overline{C}$)? To make a statement on this, both the "prior odds" and the "likelihood ratio" must be known [15]. Some forensic experts say that it is the role of the forensic scientist to evaluate the LR, and it is the role of the judge and jury to assess the prior odds and posterior odds [13]. However, the experience of many forensic scientists in court is that the judge wants an answer from the forensic expert to questions like:

> *"What is the probability that the submitted tool made the tool mark?"*

That means that the forensic scientist has to present to the court a posterior probability in addition to a likelihood ratio. Therefore, the expert has to make an estimation of the prior probability in order to be able to apply Bayes rule. For the estimation of the posterior probability, an estimation of the LR and an estimation of the prior probabilities are necessary.

This demand was also made during a Marks Conference about "Interpreting Evidence" in Berlin in 2003 by a Supreme Judge. The Supreme Judge requires the expert to make a scientific assumption of the prior probability. Arguments for this requirement are that only the forensic expert has knowledge on scientific or technical details that are needed to estimate a scientific prior probability. In cases of "lack of information" the expert can proceed – as a "pragmatic way" and not as an absolute role – according to the "principle of maximum entropy" in information theory [15] with prior odds equal 1.

If assumptions of the prior probabilities are made, it is important that this happens in a transparent way. The expert has to name explicitly the value for the prior probability. He/she should also be able to explain why he/she has chosen this value and what information led him/her to his choice.

Further on, he/she should name the value of the LR. Thus the judge has the opinion of modifying the expert's opinion and of repeating his/her calculation by using a different value for the prior probability. Sometimes it is worthwhile for the expert to make up a table relating different prior probabilities (up to four values) to posterior probabilities by considering the same value for the LR [14].

Quality Management

Quality management (QM) in forensic science laboratories encompasses everything from the equipment maintenance to training of the staff and to proficiency testing. The aim of QM is to maintain the integrity and accuracy of all laboratory testing systems over time. A guideline for QM programs is the EN 17025, formerly EN 45001 [1]. This international standard is used by various national accreditation bodies. An accreditation to the standard EN 17025 has an international recognition. A thorough discussion of QM in forensic science can be seen in **Accreditation: Organizational**.

The EN 17025 was originally meant for testing laboratories. Testing laboratories are laboratories that perform repetitive tests (such as "drinking water laboratories") and the test results are data with a well-defined precision (such as the concentration of

sodium in water). This type of testing is called *objective testing*.

According to this definition, a forensic laboratory is only a testing laboratory for a part of its activities, because laboratories usually do not end their reports with the above-mentioned data from laboratory analysis. The result of most of the reports is a concluding statement (conclusion). This part of the job is very important and gives a forensic laboratory its surplus value.

So, it appears that a forensic laboratory cannot be accredited according to EN 17025, because it does not fit within the definition of a testing laboratory. However, there are two reasons for fitting the definition of a testing laboratory [1]. In practice, the definition of a testing laboratory is interpreted widely and second a special standard for forensic laboratories, the "ILAC-G 19: Guidelines for Forensic Science Laboratories" (International Laboratory Accreditation Cooperation (ILAC)) was issued.

According to ILAC-G 19, the definition of an "objective test" is as follows: a test which having been documented and validated is under control so that it can be demonstrated that all appropriately trained staff will obtain the same results within defined limits. These defined limits relate to expressions of degrees of probability as well as numerical values: for example, "The result of the tool mark examination is more probable if the suspected tool generated the mark on the box than if another tool did generate it." Furthermore, according to ILAC-G 19, an objective test will be controlled by: documentation of the test, validation of the test, training of staff, and maintenance of the equipment. An objective test will obtain by a provision of guidance for interpretation and by checking of the results by an independent checker. Visual inspection, qualitative examinations, and computer simulations are included in the definition of an objective test.

Does the tool mark examination satisfy this definition of an objective test? It is obvious that the general aspects for tool mark examination do not differ much from other types of forensic examination. Graphically speaking, a tool mark expert can be considered as a black box with an input of questioned tool marks and test tool marks to compare and an output of a limited number of possible statements. If it can be demonstrated that several tool mark experts examined the same case and all these experts obtained the same conclusion with not more than one step difference on a conclusion scale, then there is an objective test within the scope of this definition. Before any statement (court statement, written report) leaves the laboratory, the examination must be peer reviewed. The checker may not actually rework the whole case but he/she is shown all the relevant materials that have been used for the examination.

Harmonized Conclusion Scale for Marks

The way experts express their opinions may differ depending on the country, jurisdiction, or simply their education and training. Over the last 10 years, in Europe, the situation moved toward some harmonization. During an European Network of Forensic Science Institutes (ENFSI) Expert Working Group (EWG) Marks Conference in Stockholm in 1999 a "Scale Committee" was established. It was agreed that the goal of the Committee would be to harmonize all the conclusions for the European examiners in many countries. Therefore, at the end of this project a "Conclusion Scale for Interpreting Findings in Proficiency Tests of the EWG Marks" should be reached, enabling practitioners to understand what was meant by conclusions formed by examiners across Europe.

The project had many parts: discussions about interpreting evidence, guidelines for interpreting patterns, discussion with a judge of the Superior Court in Berlin. This project produced many discussions, often controversial discussions in the "Bayesian community". The proposed harmonized conclusion scale is arguably a step toward a more practical use of a probabilistic framework to express opinions of identity of source in forensic marks examination. All the members of the Conclusion Scale Committee reached an agreement of a "six-step scale", considering two mutually exclusive hypotheses, such as: "the unknown mark was produced by the known object" and "the unknown mark was produced by another object but the known object" (see Table 2).

The harmonized conclusion scale of the EWG marks is an official ENFSI document, approved by the ENFSI "Quality and Competence Committee" [16]. A similar project was simultaneously carried out by an expert team in Germany [17]. The main result of this project was a six-step conclusion scale too.

Table 2 Conclusion scale of the ENFSI EWG Marks: assumed prior odds $P(C)/P(\overline{C}) = 1$

Step	Likelihood ratio	Probability (posterior)
1	Identification	Identification
2	Very strong support for (C)	Very probably (C)
3	Moderate support	Probably
4	Inconclusive (support for neither (C) nor ($\overline{C}$)	Inconclusive
5	Strong support for ($\overline{C}$)	Likely ($\overline{C}$)
6	Elimination	Elimination

References

[1] Katterwe, H. (1996). Modern approaches for the examination of toolmarks and other surface marks, *Forensic Science Review* **8**, 45–72 For example, traces/toolmarks and their evidentiary value (Stauffer), restoration of serial numbers (Katterwe), examination of steering columns and ignition locks (Mangine), examinations of vehicle keys (Fivaz and Bonfanti), examination of vehicle license plates (Brossier).

[2] Stauffer, E. & Bonfanti, M.S. (2006). *Forensic Investigation of Stolen-Recovered and Other Crime-Related Vehicles*, Elsevier.

[3] Katterwe, H. (2005). Fracture matching and repetitive experiments: a contribution of validation, *AFTE Journal* **37**, 229–241.

[4] Deinet, W. (1981). Studies of models of striated marks generated by random processes, *Journal of Forensic Science* **26**, 35–50.

[5] Katterwe, H. & Deinet, W. (1983). Probability models and tool marks, *Archiv für Kriminologie* **171**, 78–88.

[6] Brossier, D. (2006). Examination of vehicle license plates, in *Forensic Investigation of Stolen-Recovered and other Crime-Related Vehicles*, E. Stauffer, M.S. Bonfanti, eds, Elsevier, pp. 367–387.

[7] Voss-de Haan, P., Katterwe, H. & Simmross, U. (2003). Physics in forensic science, *Physik Journal* **2**, 35–41.

[8] Katterwe, H., Göbel, R. & Groos, K.D. (1983). The comparison scanning electron microscope within the field of forensic science, *AFTE Journal* **15**, 14–25.

[9] Göbel, E., Katterwe, H. & Körschgen, A. (2000). Camera impression evidence in polaroid photographs, *Information Bulletin for Shoeprint/Toolmark Examiners* **1**, 29–46.

[10] Koch, A. & Katterwe, H. (2007). Castings of complex stereometric samples for proficiency tests in firearm and tool mark examinations, *AFTE Journal* **39**, 299–306.

[11] Katterwe, H. (2006). *Toolmarks and Restoration of Erased Numbers*, G. Widmaier, Münchener Anwaltshandbuch Strafverteidigung, Beck München, pp. 2394–2407.

[12] Evett, I.W. & Buckleton, J.S. (1989). Some aspects of the Bayesian approach to evidence evaluation, *Journal of the Forensic Science Society* **29**, 317–324.

[13] Champod, C., Baldwin, D., Taroni, F. & Buckleton, J.S. (2003). Firearms and tool marks identification: the Bayesian approach, *AFTE Journal* **35**, 307–316.

[14] Deinet, W. & Katterwe, H. (2007). Comments on the application of theoretical probability models including Bayes theorem in forensic science relating firearms and tool marks, *AFTE Journal* **39**, 4–7.

[15] Jaynes, E.T. (1957). Information theory and statistical mechanics, *Physical Review* **106**, 620–628.

[16] ENFSI Expert Working Group Marks "Conclusion Scale Committee" (2006). Conclusion scale for shoeprint and toolmarks examinations, *Journal of Forensic Identification* **56**, 255–280.

[17] Katterwe, H., Brandes, G., Eisgruber, R., Grimmer, W., Küppers, W., Marquardt, W. & Pohl, K.H. (2007). Harmonisierte Befundbewertungsskala für kriminaltechnische Formspurenuntersuchungen, *Zeitschrift für Kriminalistik* **12**, 745–750.

HORST KATTERWE

Tort *see* Behavioral Science Evidence

Toxicology: Analysis

Introduction

The choice of which postmortem specimens to be used for toxicology examinations in a death investigation is at least partly historical. Prior to the development of spectrophotometric- and

chromatographic-based assays, testing relied on bulk distillation of a specimen such as the brain (for alcohol and other volatiles), or laborious extraction and isolation techniques for plant-derived drugs such as alkaloids, where the best specimen was stomach contents or one of the large organs such as the liver or lungs [1–4]. However, the development of instrumental analytical techniques allowed the use of smaller sample volumes and simpler extraction and isolation methods. Around this time, and especially after the development of pharmacokinetic theories and therapeutic drug monitoring, postmortem blood (in place of serum or plasma) became the specimen of choice for postmortem toxicology testing. This is because, in theory, blood is the specimen that best reflects the drug concentration at the site(s) of action of the drug or substance (i.e., the most easily accessible fluid that circulates closest to the site of action – the target receptors).

However, because postmortem blood is usually not a homogeneous fluid and because postmortem blood drug concentrations may change upward or downward after death, other specimens are often analyzed. Hence, a thorough postmortem toxicological investigation may involve multiple tissues or fluids. No single postmortem specimen should be relied upon for toxicological interpretation in all circumstances.

Specimens

Postmortem Blood

Blood is the most common postmortem specimen that is collected. It can be drawn from various locations in the body, generally without any form of dissection. It is usually liquid, typically with small clots that are easily avoided or drawn through a large-bore needle. Larger clots may exist in the major pulmonary vessels or the major chambers of the heart. Blood can be collected from the major vessels during external examination of the body or during the autopsy (dissection). It is good practice to collect blood from a peripheral site such as the femoral vein and a central site such as one of the cardiac or adjacent vessels. The larger volumes available from the central vessels (e.g., 30–80 ml) should be used to screen for the presence of drugs and toxicants, whereas the smaller volumes available from peripheral vessels (e.g., 3–10 ml) should be used for drug quantification. A list of suggested postmortem specimens and uses is given in Table 1.

Wherever the blood is obtained, it should be taken from a discrete vessel and not by a "blind stick" into the chest cavity in the hope of hitting one of the cardiac vessels or chambers. A "blind stick" carries the significant risk of collecting pooled blood or other fluid from the pleural or chest cavity, rather than from an intact vessel. The consequence may be a grossly elevated concentration of alcohol or drug if the stomach has been ruptured or lungs punctured [5].

Collection and sometimes measurement of drug concentrations in blood collected from two different sources (e.g., peripheral and central) is important because postmortem blood is rarely entirely homogenous and may contain vastly different concentrations for some drugs. This can be because of incomplete distribution of a drug prior to death or redistribution of a drug after death. Although this is not necessarily critical for a potent toxin such as strychnine, which ought not to be present at any concentration, it can affect interpretation of drugs the deceased person was prescribed or known to be taking.

The distinction between peripheral and central blood is critical from the perspective of interpretation, but not absolute. In other words, blood from these sources should not be regarded as separate and unvarying. For example, "central" blood can contain vastly different concentrations of certain drugs subject to postmortem redistribution depending on whether the blood actually comes from the aorta, inferior vena cava (IVC) or the pulmonary vein or aorta, or a combination [6]. Similarly, a large volume of blood drawn from an unligated femoral vein may be contaminated with blood from the iliac vein and IVC, carrying an increased risk of being contaminated with "central" blood subject to postmortem redistribution.

Ideally, the vessel from which the blood is drawn from should be ligated (i.e., tied off) to prevent blood being drawn from another part of the body [7]. For example, if a needle is inserted into the femoral vein, drawing a large volume (e.g., 30–80 ml) may result in much of that blood siphoning down from the IVC and other central vessels that are subject to postmortem redistribution. Applying a tourniquet above, the point of collection from the femoral vein is an alternate method of minimizing the siphoning of blood from central vessels.

Table 1 List of postmortem specimens, suggested amount, and uses for toxicological analysis

Specimen	Amount	Use
Central blood	20–50 ml	General drug screening and quantitation, including volatiles
Peripheral blood	5–10 ml	Drug/toxicant quantitation, including volatiles
Vitreous humor	3–6 ml	Quantitation of ethanol and other volatile, glucose, electrolytes (except potassium), quantitation of specific drugs
Liver	20–100 g	Secondary specimen for drug screening in the case of insufficient blood; additional specimen for drug quantitation to assist interpretation
Gastric fluid	All available	Quantitation of drugs
Lungs	Whole lung (tied off)	Sometimes useful for the detection of acutely inhaled gases if the death is very sudden, with little opportunity to exhale the inspired gas
Brain	20–100 g	Sometimes useful for the detection and measurement of drugs and/or metabolites that are more stable in brain than in blood (e.g., heroin/6-acetylmorphine; cocaine)
Spleen	20–100 g	Source of blood, where otherwise not available, for the determination of carboxyhemoglobin (carbon monoxide) or cyanide
Kidney	20–100 g	Sometimes useful for the determination of heavy metals, as these tend to concentrate in renal tissue
Bile	3–10 ml	Useful for the detection of some drugs that concentrate in bile, but where blood levels are very low
Cerebrospinal fluid	2–10 ml	Useful alternate specimen for the measurement of ethanol
Other organ tissues (e.g., lungs, heart, brain, spleen, kidneys, and skeletal muscle)	20–100 g	Usage is similar to that of liver. Tissue from multiple organs sometimes used for the determination of drug distribution in a homicidal poisoning or other potentially difficult cases
Hair	~20–100 hairs, pulled or shave at level of scalp	Can be useful in special cases to help differentiate acute from chronic exposure, e.g., abuse cases, homicidal poisoning, and sexual assault
Bone and nails	1–20 g	Can be used for qualitative determination of drugs where other specimens are not available. Quantitative data useful for arsenic or heavy metals

In special circumstances, an otherwise good specimen may be affected by external factors. For example, intravenous pumps (e.g., patient controlled anesthesia devices) or even simple intravenous lines may continue to run, even for a short time after death, resulting in locally elevated concentrations of the drug being administered – potentially resulting in misinterpretation of a high blood concentration.

The toxicologist should also be wary of specimens labeled as "blood" taken from moderately decomposed bodies. Such specimens may not always be collected from discrete vessels and in fact may be pooled bloody fluid from the chest cavity containing drug concentrations that will be difficult to interpret (*see* **Postmortem Toxicology: Artifacts**).

Urine

Urine is typically a clear fluid, ranging in color from nearly colorless to almost brown, containing

98–99% water. Urine has long been preferred as one of the best fluids for screening for drugs and other toxicants. The reason is the lack of protein and relative lack of lipids that can interfere with (gas chromatography) GC-based screening tests, and since urine contains concentrations of "basic" drugs (e.g., alkaloids) and other substances that can be 10–100 times the corresponding blood concentration. In addition, as a clear fluid, urine is easy to pipette and amenable to automated or semiautomated screening systems such as immunoassay and is easy to use with solid-phase extraction columns. By its nature, urine should be completely homogeneous at the time of collection.

There are, however, disadvantages in using urine as a postmortem specimen. First, there is very little correlation between the concentrations of most drugs in urine and blood. Second, urine is not always available as a specimen, either because of trauma or because of loss of bladder control at the time of or after death. Third, many drugs are so extensively metabolized in the body that little if any of the parent drug is excreted into the urine. That stated, many toxicologists find the detection of known metabolites to be equally useful.

Vitreous Humor

Vitreous humor is a clear colorless, slightly viscous fluid found in the center of the eyeball. Approximately 3 ml can be recovered from each eye. The main use of vitreous fluid is for the measurement of alcohol (ethanol), or more specifically as a fluid that can be used to corroborate postmortem blood alcohol findings. Alcohol can be produced in blood after death due to postmortem fermentation (i.e., due to metabolism of glucose and other substrates in blood contaminated with exogenous microorganisms or those naturally present) [8]. Fermentation in postmortem blood can produce alcohol concentrations approaching or in rare cases considerably exceeding the legal limit for driving of 80–100 mg/100 ml. However, the vitreous fluid in the eye is sterile and therefore not subject to postmortem fermentation [9, 10].

Vitreous humor is also useful for confirming the presence of a hyperglycemic state (abnormally high glucose) and therefore also diabetic ketoacidosis at the time of death. "Normal" glucose concentrations rapidly decrease to almost zero after death due to continued cellular metabolism. Conversely, if blood is collected close to the liver (e.g., from the IVC), glucose concentrations can be erroneously high due to breakdown of glycogen in the liver. Vitreous humor is remote from the liver and therefore not affected by glycogenolysis. Although glucose concentrations are also known to fall in the vitreous humor after death, this decrease in concentration appears to be limited, such that in a severely hyperglycemic individual, vitreous humor concentrations of glucose will invariably remain high after death. If the person is also in a state of ketosis, vitreous acetone concentrations will also be elevated [11, 12].

In addition to glucose, vitreous biochemistry (e.g., urea and creatinine) and electrolytes (e.g., sodium and chloride) are sometimes measured [12]. Elevated concentrations of these substances, particularly urea and creatinine, can indicate renal insufficiency or failure at the time of death. Elevated sodium and chloride can indicate dehydration of the body. Measurement of potassium is generally not performed because it increases with time after death due to diffusion from surrounding tissue. Although measurement of potassium has been studied as a possible indicator of time since death, the actual increases are too unpredictable to give a reliable estimate in any given case [13].

Vitreous humor is also increasingly being used for the measurement of drugs. Such use was originally limited by the relatively small volume of fluid available in each eye. However, increases in sensitivity of analytical instrumentation and methods now permit a wide range of drugs to be measured. In this respect, the use of vitreous humor as a specimen offers two advantages. First, vitreous humor lacks the enzymes that are in blood, which rapidly hydrolyze certain drugs or metabolites such as cocaine and 6-acetylmorphine (the intermediate and diagnostic metabolite of heroin) [12, 14]. Second, the specimen is relatively protected from postmortem redistribution and diffusion; it is relatively remote from the major organs such as liver, lungs, stomach, and heart. Measurement of digoxin in vitreous humor is a classical example [15]. However, for some drugs postmortem redistribution from the brain is a possibility.

Besides the relatively small volume of fluid available, the other disadvantage of measuring drug

concentrations in the vitreous is the lack of a comprehensive database. Although the forensic literature contains innumerable articles that report the distribution of drugs in many different tissues and fluids, including vitreous humor, the data are not easy to locate for any specific drug, and even then may only be available for a very small number of cases. Direct interpretation by comparison to blood databases cannot be done because drug distribution into the vitreous humor is often different than for blood. For example, drugs that are highly lipid soluble or strongly protein bound tend to have significantly lower concentrations in vitreous fluid [16].

Liver

Other than the skin, muscle, and fat, the liver is the largest organ in the body, weighing 1.5–2 kg in an adult. It has been used in the past for drug detection and measurement because it was easy to collect and offered a large mass to work with – especially if trying to isolate a pure substance for crystallography or animal studies. Now, however, the main advantage of liver tissue is as a secondary specimen for the measurement of drug concentrations. It is a tissue that is relatively easy to macerate (to form a homogenate). Drug concentrations of many drugs are 10- to 50-fold higher than in blood, enabling easier detection and measurement. Most drug concentrations are also relatively stable in the liver, undergoing relatively little postmortem distribution compared with that in the blood [17, 18]. Measurement of drugs that can undergo extensive postmortem redistribution, such as tricyclic antidepressants, has been well published [19–21].

However, as for many "secondary" tissues, a disadvantage of measuring drug concentrations in liver tissue is that comprehensive databases to assist with interpretation do not exist; notwithstanding that apart from blood (and in the living, serum or plasma), liver is the organ in which drug concentrations have been most reported. The forensic toxicologist should also be aware of the potential problems of analyzing liver tissue that has been collected from the lobe proximal to the stomach. In cases of overdose, or even sometimes recent therapeutic doses, postmortem diffusion from the stomach (across the stomach wall) can occur and may elevate drug concentrations in the adjacent liver tissue [22]. Despite these potential problems, the liver can be a useful secondary specimen that, when properly considered in the context of the case, can provide useful quantitative information.

Bile

Bile is the straw to green-colored fluid that collects in the bile duct with a volume of 5–10 ml. Bile has traditionally been used for the detection and measurement of some opiates – in particular morphine – because of the high concentrations found. Similarly many other drugs that undergo enterohepatic circulation also tend to concentrate in bile fluid. The use of bile has been advocated by some as a useful specimen for postmortem work [23, 24]. It is potentially useful since some drugs concentrate in bile and can be detected in bile for a longer period of time than that in blood, e.g., morphine and colchicine. However, there is not a well-established quantitative relationship between the concentrations of drugs in blood and bile, limiting its use for interpretation. This is not surprising. Most tissues are highly perfused with blood, and therefore drugs are expected to develop and maintain equilibrium between that tissue and blood. As the blood concentration rises, the tissue concentration of the drug in that organ should also rise accordingly (although the absolute concentration of the drug in the tissue may be quite different than that in the blood). However, bile is a fluid that is secreted by the liver over time and stored in the bile duct until required. As for urine, bile fluid is not in constant equilibrium with circulating blood and therefore cannot be a good predictor of blood drug concentration, nor drug concentration at the target organ or receptor. Bile drug concentrations may also be influenced by the liver and in the case of an overdose, by postmortem diffusion from the stomach.

Attempts have been made to use bile as a secondary specimen for the measurement and interpretation of postmortem alcohol [25, 26]. However, although the correlation is reasonable in a fresh cadaver, bile alcohol concentrations will certainly be unreliable where the possibility of postmortem fermentation exists, due to its proximity to the liver and gastrointestinal tract, and the massive source of glycogen and microorganisms, respectively.

Gastric Contents

The nature of gastric contents can vary anywhere from almost clear, sometimes mucous fluid excreted

by the stomach, to barely digested half-macerated food. The measurement of drugs and other toxicants in gastric contents is, or should be, standard practice in the case of suspected oral overdose. Homogenization of the stomach contents prior to sampling for drug measurement, or dissolution in a water-miscible solvent, can be important for quantitative accuracy, especially if unabsorbed medication is present.

It is important to recognize that, unlike for blood and tissue, interpretation of drug residue in the gastric contents should be made on the basis of the *total amount* present (e.g., milligrams or grams in the total volume) and not the concentration (e.g., milligrams per liter). In particular, gastric fluid drug concentrations should never be compared directly with blood concentrations. For example, a very high concentration can result if a person swallows a single, therapeutic dose of a medication on a nearly empty stomach. The same dose will result in a much lower concentration if the same dose of medication was consumed after a meal and a drink. It is the total amount of drug in the stomach that is important – not the concentration.

Obviously, if a drug is consumed recently and orally, its detection in the gastric contents can be expected. However, it should be borne in mind that merely detecting a drug in the gastric contents does not prove an oral route of administration. Any drugs circulating in the blood will, at least in theory, occur at some concentration in the gastric fluid for no other reason than that most drugs are widely distributed to virtually every tissue and fluid in the body. Drugs can therefore enter the stomach by secretion in gastric juice. Some drugs will also enter the stomach through the bile. Drugs that are "basic" (or alkaline) in character, such as alkaloids, may tend to occur in the gastric fluid more than neutral or acidic drugs, because of the low pH of gastric fluid.

For similar reasons, caution should be expressed about drawing conclusions regarding the time of ingestion of a drug from its presence or concentration in the stomach. The origin of low concentrations has already been discussed. "Stomach emptying" is reported to occur within an hour or so. However, stomach emptying is highly variable and is known to be delayed by an unusually large meal. (How many of us have woken up in the middle of the night with a bloated stomach, hours after the meal was consumed?) It is known that irritant substances can delay gastric emptying. In addition, after a drug overdose, some medications can form a semisolid mass in the stomach that persists largely unabsorbed for many hours or even days [27, 28]. Similarly, the presence of "ghost pills" in the stomach has been reported [29]. This occurs because the tablet excipient is poorly digestible, while the more easily dissolvable drug diffuses out and passes into the small intestine, leaving the tablet "shell" behind in the stomach. This typically occurs with slow-release tablets that have a waxy matrix, e.g., slow-release oxycodone, theophylline, and acetylsalicylic acid. Furthermore, some drugs affect gut motility and will therefore delay gastric emptying and the passage of food and drugs through the gut (e.g., opiates, plus drugs with cholinergic action including many antipsychotic medications).

Other Tissues/Fluids

Virtually any tissue can be used for the measurement of drugs, including brain, lungs, kidney, spleen, and muscle. The main consideration is availability, the work entailed in the measurement and ability to interpret the result obtained. For example, in some types of death (severe decomposition; animal activity of a body outside) other than bone, muscle may be the only tissue available. Interpretation of a drug concentration may be difficult, but simple detection and confirmation of the presence of the drug or toxicant may be all that is required (e.g., suicide or poisoning by strychnine).

Brain tissue is frequently used for the measurement of cocaine concentrations because they are typically higher than that in blood, due to greater stability in that organ [30, 31]. Drug concentrations in the brain should not be affected by postmortem redistribution or diffusion. However, for at least some drugs, concentrations may be dependent on the area of the brain the tissue is sampled from [32].

Spleen can be useful for the measurement of carbon monoxide and cyanide because the organ is rich in red blood cells.

Kidney has been classically used for the measurement of heavy metals and there is a good database of "normal" concentrations for this tissue.

Cerebrospinal fluid (CSF) can be used for the same purpose as vitreous humor. However, although the volume of CSF is generally larger than vitreous fluid, it is more difficult to collect and may not always be available.

Injection Sites

Pathologists sometimes submit suspected injection sites for the purpose of determining whether a drug or toxicant was injected. Certainly, if the drug was injected into muscle just before death, the tissue should contain elevated concentrations of that drug. However, the simple detection of a drug at an injection site cannot be used as proof of the route of administration. The reason is that virtually all drugs are widely distributed throughout the body and certainly into all muscle tissue. Therefore, regardless of the route of administration, some drug is expected to be found in muscle. The only way to demonstrate a likely point of injection is to also collect and analyze a similar portion of tissue from the other site of the body and compare the drug concentrations. For intravenous injections, proof of the site of injection may be difficult because the drug will be swept into the general circulation quickly, with little remaining near the injection site.

Hair

Hair has been used for the detection of potential poisons for decades, albeit mostly arsenic and heavy metals. Most drugs and poisons with significant stability will be taken up into the structure of hair as hair growth occurs. Therefore, if external contamination of the hair sample can be excluded, the presence of a toxicant indicates administration when alive. Segmentation and sequential analysis of those segments can be used to differentiate acute from chronic administration (for example, this has been used for arsenic) [33]. This type of analysis has also been useful in corroborating criminal administration of a drug as part of drug-facilitated sexual assault [34]. Analysis of hair is also widely applied to demonstrate illicit drug use. Hair analysis for this purpose has the advantage that it can demonstrate drug use for a much longer period (e.g., weeks or months) than does urine testing (e.g., hours or days). (*see* **Hair: Toxicology**).

Nails and Bone

Although nails and bone are chemically different (nails are cartilage and bone is not), both take up drugs and can be used to demonstrate exposure to drugs and other poisons [35]. The application of analysis of these specimens is limited to the detection of substances that should not be there. Interpretation of quantitative data is difficult or impossible because of the lack of reference values. However, detection of a substance such as strychnine or an unusually high concentration of arsenic or a heavy metal may be sufficient to prove that poisoning has occurred [36, 37].

Other Exhibits

Analysis of drug paraphernalia, such as papers and foils containing powders, residues in syringes, spoons, etc., may provide useful information. For example, heroin is rapidly broken down in the body to morphine via 6-acetylmorphine – neither heroin or 6-acetylmorphine may be detected. Finding heroin in a nearby syringe or spoon, therefore, provides useful corroboration of heroin use in a person where only morphine was found in blood postmortem. Similarly, other drugs may be detected at the higher concentrations often present in the paraphernalia that are analyzed compared with the very low concentrations often present in blood (even in an overdose).

Antemortem/Perimortem Specimens

One of the most difficult problems encountered in postmortem toxicology is the postmortem redistribution of drugs and other changes such as continued metabolism in a delayed death. Therefore, the availability of a blood or serum specimen that was collected close to death (for example, during emergency medical treatment or during resuscitation attempts), or immediately after death, can provide invaluable information for interpretation that cannot be obtained in any other way. The simplest example is the determination of a drug concentration in a blood sample drawn close to the time of death. For drugs subject to postmortem redistribution, an antemortem specimen will provide the most valuable data.

Where a person ingests or injects a potentially fatal overdose, but is found alive and resuscitated, death may eventually occur from complications of the drug intoxication days or even weeks after the initial event (e.g., hypoxic brain damage, pneumonia, and organ failure). In such circumstances, the original drugs would have been completely cleared from the body, leaving no chemical evidence of the original

intoxication. Availability of a blood sample collected soon after the original admission may therefore provide the only evidence of the original intoxication.

Premortem specimens may also provide invaluable information about the ability of the person to metabolize and clear drugs when alive. When alive, some drugs accumulate in the body due to impaired metabolism or excretion and may ultimately cause death [38]. If the drug concentration is high at the time of death, then the toxicologist, pathologist, and medical examiner/coroner have to try and determine whether it was high due to a natural process or acute ingestion. That may still be difficult to determine, but at least the occurrence of elevated drug concentrations due to postmortem redistribution can be ruled out.

Summary

In summary, most drugs and poisons are widely distributed throughout the human body. Therefore, virtually any biological specimen can be used to screen for, identify and measure their presence. The governing fact is usually the availability of a specimen (for example, blood, vitreous, and urine not may be available in a severely decomposed body). Thereafter, the usefulness of analysis of a particular fluid or tissue will be dictated by the ability to interpret the analytical finding and its value in determining the cause and manner of death.

References

[1] Daubney, C.G. & Nickolls, L.C. (1937). An investigation into the methods of toxicological analysis of viscera. Part I, *The Analyst* **62**, 851–859.

[2] Daubney, C.G. & Nickolls, L.C. (1938). An investigation into the methods of toxicological analysis of viscera. Part II, *The Analyst* **63**, 560–566.

[3] Feldstein, M. & Klendshoj, N.C. (1954). The analysis of "general unknowns" in toxicology, *Canadian Journal of Medical Technology* **16**, 48–52.

[4] Curry, A.S. (1963). *Poison Detection in Human Organs*, Charles C. Thomas, Spinglfield.

[5] Logan, B.K. & Lindholm, G. (1996). Gastric contamination of postmortem blood samples during blind-stick sample collection, *The American Journal of Forensic Medicine and Pathology* **17**, 109–111.

[6] Jones, G.R. & Pounder, D.J. (1987). Site dependence of drug concentrations in postmortem blood – a case study, *Journal of Analytical Toxicology* **11**, 186–190.

[7] Druid, H. & Holmgren, P. (1997). A compilation of fatal and control concentrations of drugs in postmortem femoral blood, *Journal of Forensic Sciences* **42**, 79–87.

[8] Corry, J.E. (1978). A review. Possible sources of ethanol ante- and post-mortem: its relationship to the biochemistry and microbiology of decomposition, *The Journal of Applied Bacteriology* **44**, 1–56.

[9] Backer, R.C., Pisano, R.V. & Sopher, I.M. (1980). The comparison of alcohol concentrations in postmortem fluids and tissues, *Journal of Forensic Sciences* **25**, 327–331.

[10] Caplan, Y.H. & Levine, B. (1990). Vitreous humor in the evaluation of postmortem blood ethanol concentrations, *Journal of Analytical Toxicology* **14**, 305–307.

[11] DiMaio, V.J., Sturner, W.Q. & Coe, J.I. (1977). Sudden and unexpected deaths after the acute onset of diabetes mellitus, *Journal of Forensic Sciences* **22**, 147–151.

[12] Coe, J.I. (1993). Postmortem chemistry update. Emphasis on forensic application, *The American Journal of Forensic Medicine and Pathology* **14**, 91–117.

[13] Madea, B., Herrmann, N. & Henbge, C. (1990). Precision of estimating the time since death by vitreous potassium–comparison of two different equations, *Forensic Science International* **46**, 277–284.

[14] McKinney, P.E., Phillips, S., Gomez, H.F., Brent, J., MacIntyre, M. & Watson, W.A. (1995). Vitreous humor cocaine and metabolite concentrations: do postmortem specimens reflect blood levels at the time of death? *Journal of Forensic Sciences* **40**, 102–107.

[15] Vorpahl, T.E. & Coe, J.I. (1978). Correlation of antemortem and postmortem digoxin levels, *Journal of Forensic Sciences* **23**, 329–334.

[16] Scott, K.S. & Oliver, J.S. (2001). The use of vitreous humor as an alternative to whole blood for the analysis of benzodiazepines, *Journal of Forensic Sciences* **46**, 694–697.

[17] Hilberg, T., Bugge, A., Beylich, K.M., Ingum, J., Bjorneboe, A. & Morland, J. (1993). An animal model of postmortem amitriptyline redistribution, *Journal of Forensic Sciences* **38**, 81–90.

[18] Hilberg, T., Morland, J. & Bjorneboe, A. (1994) Postmortem release of amitriptyline from the lungs; a mechanism of postmortem drug redistribution,. *Forensic Science International* **64**, 47–55.

[19] Bailey, D.N. & Shaw, R.F. (1979). Tricyclic antidepressants: interpretation of blood and tissue levels in fatal overdose, *Journal of Analytical Toxicology* **3**, 43–46.

[20] Bailey, D.N. & Shaw, R.F. (1980). Interpretation of blood and tissue concentrations in fatal self-ingested overdose involving amitryptyline: an update (1978–1979), *Journal of Analytical Toxicology* **4**, 232–236.

[21] Apple, F.S. & Bandt, C.M. (1988). Liver and blood postmortem tricyclic antidepressant concentrations, *American Journal of Clinical Pathology* **89**, 794–796.

[22] Pounder, D.J., Fuke, C., Cox, D.E., Smith, D. & Kuroda, N. (1996). Postmortem diffusion of drugs from

gastric residue: an experimental study, *The American Journal of Forensic Medicine and Pathology* **17**, 1–7.
[23] Tracqui, A., Kintz, P. & Ludes, B. (1998). Buprenor phine-related deaths among drug addicts in France: a report on 20 fatalities, *Journal of Analytical Toxicology* **22**, 430–434.
[24] Vanbinst, R., Koenig, J., Di Fazio, V. & Hassoun, A. (2002). Bile analysis of drugs in postmortem cases, *Forensic Science International* **128**, 35–40.
[25] Winek, C.L. & Esposito, F.M. (1981). Comparative study of ethanol levels in blood versus bone marrow, vitreous humor, bile and urine, *Forensic Science International* **17**, 27–36.
[26] Winek, C.L., Henry, D. & Kirkpatrick, L. (1983). The influence of physical properties and lipid content of bile on the human blood/bile ethanol ratio, *Forensic Science International* **22**, 171–178.
[27] Baum, J. (1984). Enteric-coated aspirin and the problem of gastric retention, *The Journal of Rheumatology* **11**, 250–251.
[28] Cereda, J.M., Scott, J. & Quigley, E.M. (1986). Endoscopic removal of pharmacobezoar of slow release theophylline, *British Medical Journal (Clinical Research Ed.)* **293**, 1143.
[29] Anderson, D.T., Fritz, K.L. & Muto, J.J. (2002). Oxycontin: the concept of a "ghost pill" and the postmortem tissue distribution of oxycodone in 36 cases, *Journal of Analytical Toxicology* **26**, 448–459.
[30] Spiehler, V.R. & Reed, D. (1985). Brain concentrations of cocaine and benzoylecgonine in fatal cases, *Journal of Forensic Sciences* **30**, 1003–1011.
[31] Moriya, F. & Hashimoto, Y. (1996). Postmortem stability of cocaine and cocaethylene in blood and tissues of humans and rabbits, *Journal of Forensic Sciences* **41**, 612–616.
[32] Spiehler, V.R., Sedgwick, P. & Richards, R.G. (1981). The use of brain digoxin concentrations to confirm blood digoxin concentrations, *Journal of Forensic Sciences* **26**, 645–650.
[33] Poklis, A. & Saady, J.J. (1990). Arsenic poisoning: acute or chronic? Suicide or murder? *The American Journal of Forensic Medicine and Pathology* **11**, 226–232.
[34] Kintz, P., Cirimele, V., Jamey, C. & Ludes, B. (2003). Testing for GHB in hair by GC/MS/MS after a single exposure. Application to document sexual assault, *Journal of Forensic Sciences* **48**, 195–200.
[35] McIntyre, I.M., King, C.V., Boratto, M. & Drummer, O.H. (2000). Post-mortem drug analysis in bone and bone marrow, *Therapeutic Drug Monitoring* **22**, 79–83.
[36] McBay, A.J., Hudson, R.P. & Boling, V.R. (1979). Arsenic in exhumed bodies, *Journal of Analytical Toxicology* **3**, 222–223.
[37] Hann, S., Latkoczy, C., Bereuter, T.L., Prohaska, T., Stingeder, G. & Reiter, C. (2005). Reconstruction of a case of thallium poisoning using LA-ICP-SFMS, *International Journal of Legal Medicine* **119**, 35–39.
[38] Swanson, J.R., Jones, G.R., Krasselt, W., Denmark, L.N. & Ratti, F. (1997). Death of two subjects due to imipramine and desipramine metabolite accumulation during chronic therapy: a review of the literature and possible mechanisms, *Journal of Forensic Sciences* **42**, 335–339.

GRAHAM R. JONES

Toxicology: Behavioral *see*
Behavioral Toxicology

Toxicology: Driving *see*
Drug-Impaired Driving

Toxicology: Forensic Applications of

Introduction

The purpose of forensic toxicology is to produce scientifically valid toxicological information for the administration of justice and to improve the legal protection of individuals and society. This goal sets a high standard for sampling, handling, and analyzing evidence; ensuring the chain of custody is intact. Laboratory analysis requires the use of validated chromatographic and spectrometric techniques in order to obtain robust and unambiguous results [1]. This distinguishes forensic toxicology with clinical toxicology, where the use of rapid immunoassay techniques for drug classes is often justified and is sufficient for diagnosis – but is not suitable for legal purposes alone.

The origin of forensic toxicology is related to the work of the Spanish-born French toxicologist and chemist M. J. B. Orfila (1787–1853), later professor at the Paris Faculty of Medicine. He contributed

fundamentally to the scientific basis of toxicology and established the principles on giving expert opinion in the court. In the famous trial of Marie Lafarge, who was suspected for the murder of her husband, Orfila was able to prove, using the Marsh test, that there was arsenic in the body. This made it possible for Lafarge to be found guilty [2]. Since then, the discovery of homicide by poisoning has been the ultimate duty of forensic toxicologists.

In forensic toxicology, suspicion is the driving force and the general unknown is the main challenge. The suspicion arises usually from the background information of the circumstances of the case: what has happened and what is out of the ordinary. In postmortem cases, this kind of information is most often collected from the scene of death and gathered by the police or coroner. Sufficient background information is in every case an essential starting point to a successful laboratory investigation and is needed to widen the analytical scope in the right direction. For instance, if an insulin pen is found at the scene, the body should be assayed at least for insulin; the sudden death of a goldsmith warrants an analysis of cyanide; and aged people are more often prescribed digoxin than younger people.

The first phase of the analytical procedure is usually screening (the initial test) analysis. Depending on the purpose of the investigation, the screening can be comprehensive, as is the case of death investigations, where all possible findings are of interest. It can also be limited, as in drugs and driving cases, where only the drugs that have an effect on driving performance are included. The focus can also be set by agreement, for example, when conducting urine testing for drugs in the workplace (*see also* **Toxicology: Initial Testing**).

The scope of forensic toxicology is vast: it includes investigation of poisonings in humans and animals, screening for alcohol and drug abuse, monitoring pollution of the environment, detecting tampering of pharmaceutical and chemical products, and also uncovering the substances used in a terrorist attack. The wide range of applications explains the multitude of analytical techniques that are required to solve these tasks. Many laboratories have also received certification or accreditation from an authoritative body.

In the analysis of human samples, forensic toxicology answers two key questions: (i) has the person under investigation been exposed to alcohol, drugs, or other xenobiotics and (ii) has he been under the influence of any foreign substances [3]. Exposure to a substance is verified by detecting the presence of a substance in a biological sample, not only in urine but also in blood, breath, hair, oral fluid, sweat, or tissues. The mere presence of the parent substance, e.g., cocaine in the hair, is sometimes not sufficient alone to prove systemic use, but the presence of metabolites may also need to be demonstrated to reliably interpret the finding. In any case, a positive finding does not mean that the person has actually taken the substance by himself, because exposure can take place deliberately or unwittingly.

An accurate measurement of the blood concentration of the substance is a prerequisite for evaluating whether a person is under the influence. With living persons, a representative specimen of venous blood can easily be collected by venepuncture; however, sampling is more challenging with cadavers. Blood from central compartments, such as heart blood, may contain drugs in up to 10 times higher concentrations compared to peripheral blood [4] (*see* **Postmortem Toxicology: Artifacts** for further information).

Considering all the shortcomings in collecting postmortem blood, femoral venous blood represents antemortem blood more closely than does any other blood sample [5]. In some countries, breath testing with evidential breath alcohol analyzers can be used *per se* as evidence of being under the influence of alcohol (*see* **Alcohol: Analysis** for examples and further information).

Forensic toxicology is often associated with a forensic medicine department at universities. Research typically includes development and validation of analytical methods, pharmacokinetics of drugs and poisons, pharmacogenetics, case studies, and epidemiology of poisonings, and social studies such as evaluation of the consequences of drug abuse and research on drug safety.

Postmortem Toxicology

In postmortem forensic toxicology, analytical results are utilized to determine the cause and manner of death. Poisoning can be the underlying cause of death (WHO Ic), which means that poisoning "initiated the train of morbid events leading directly to death", or the immediate cause of death (WHO Ia), defined as a "condition with the symptoms of which the

deceased died". Poisoning can also be a contributing cause of death (WHO II) [6]. It is possible for Ia and Ic to be the same, as in the case of death by alcohol poisoning after heavy drinking if no other cause is operating simultaneously. The substances that frequently cause fatal poisonings include opioids, tricyclic antidepressants, neuroleptics, alcohols, carbon monoxide, and, especially in developing countries, pesticides. In traffic and workplace accidents, the economic consequences can be considerable and the question of culpability has significance from the viewpoint of criminal and civil law. Applications of postmortem toxicological investigations are listed in Table 1.

The question of poisoning arises typically at the scene of death: medicine bottles or packets, pills, tablets, syringes, or needles are found in the vicinity of the body. Alternatively, prescribed drugs are missing or the consumption of drugs has surpassed the amount prescribed. In suicidal poisoning, a suicide note may also be found. External examination of the body can reveal dried drug stains around the mouth or powder around the nose, or needle marks may be detected in the skin. Skin color and lividity can suggest poisoning, for example, in carbon monoxide and cyanide poisoning the lividity is typically reddish. The internal examination, or autopsy, includes opening of the skull and investigation of the brain. On this occasion, an odor suggesting alcohol, cyanide, or other volatiles may be noticed. In the alimentary tract, irritation, corrosive changes, or tablet remains are indicative of oral poisonings. In other organ systems, congestion and edema are typically present in poisonings. The blood can be flowing and appear more reddish than usual. However, these changes are not necessarily pathognomonic and can be caused by nonspecific changes found in sudden death due to asphyxia or exposure to a cold environment.

Table 1 Applications of postmortem forensic toxicological investigation

Autopsy findings or circumstances suggest poisoning
Homicides and other suspicious cases
Suspicion of malpractice
Need to confirm an apparent cause of death
Traffic accident
Occupational or workplace accident
Check of compliance with pharmacotherapy
Obscure death

Another application of postmortem toxicology is a suspicion that the death is due to pharmacotherapy. These include cases related to drug injection given by a member of the nursing staff or that the death has otherwise an obvious temporal connection to a dose of drug. Postmortem toxicology can be used to assess compliance with pharmacotherapy, e.g., the use of anticonvulsant drugs, antidepressants, insulin, oral hypoglycemic drugs, etc.

Sometimes, such as in suicides committed with car exhaust gas, the cause of death is obvious, but the toxicological determination of carboxyhemoglobin (caused by binding carbon monoxide with hemoglobin) is still justified. Currently, this is even more important since new cars are equipped with catalytic converters and the death may be due to suffocation (from depletion of oxygen) instead of carbon monoxide. According to good practice, the diagnosis of obscure or unascertained death (*Mortuus inventus, causa ignota*) as the underlying cause of death can be issued only if a postmortem toxicological investigation has been performed and did not reveal poisoning.

Sudden and unexpected death can also be due to potentially criminal matters such as homicide or negligent medical treatment, and suspected cases ought to be subjected to forensic cause of death investigations, including full forensic toxicology. A striking example of negligence in this respect is the Shipman case. Dr Harold Shipman was a private practicing doctor who was also an active serial killer in 1970–1998 in England. He was convicted for 15 murders by poisoning with diamorphine (heroin) or morphine, but the actual number of victims possibly amounted to at least 218 [7].

Traffic Medicine

Impairment of drivers is the main cause of traffic crashes and accidents. Typical reasons for impairment are the use of alcohol and drugs including in particular drugs of abuse. This is why forensic toxicological investigation is needed for evaluating whether the driver has been under the influence of a substance. According to the legislation of some countries, police must first perform a roadside evaluation of impairment before subjecting the driver to a breath or blood test, whereas in other countries they are allowed to perform random stop checks. Punishable blood–alcohol concentration varies from

zero in Hungary through 20 mg/100 ml (Sweden and Norway) to 50 mg/100 ml (Australia and most European Union countries) and 80 mg/100 ml or higher (United States and Canada). If the driver is able to blow sufficient breath and no other adverse substances are suspected, the presence of alcohol is typically assessed by using an evidential breath analyzer. Otherwise, laboratory determination of blood–alcohol concentration by headspace gas chromatography is performed (*see* **Alcohol: Analysis**; **Alcohol** for further information).

Drugs of abuse are more frequently the cause of impairment in driving. This is why new methods to reveal drug-positive drivers have been developed. Much effort has been focused on oral fluid drug testing, but a present survey of oral fluid testing indicates that the sensitivity of the tests for cannabis and benzodiazepines is still unsatisfactory. However, random drug testing has been used successfully as a deterrence. Positive roadside tests by immunoassay do not authorize the imposition of legal penalties unless a confirmation analysis has been carried out in a forensic laboratory equipped with specific chromatography-mass spectrometry methods.

There are differences in the legislations of different countries regarding how to handle a positive test result. In some countries, such as the United Kingdom, Norway, and Denmark, the law stipulates that a driver's impairment must be proven. Some countries, e.g., Germany, France, Switzerland, Belgium, Sweden, Finland, and many Australian States, apply the so-called zero tolerance, which means that any time a banned drug or its metabolite is detected in a blood sample, violation of the law occurs. In practice, an analytical method has a limit of detection and a limit of quantification, and consequently a cutoff limit is applied. Any result exceeding the cutoff is interpreted as positive and conversely, any result below the cutoff is interpreted as negative. Cutoff limits for different drugs vary between countries, and they have also changed over the years. In some countries, such as France and Finland, the cutoff is equal to the limit of quantification of the laboratory method (*see* **Drug-Impaired Driving** for more information).

Clinical Forensic Toxicology

Clinical forensic toxicology is applied to a multitude of purposes in society, some of which are listed in Table 2. Drugs may precipitate violent behavior, including psychotic effects of amphetamines or disinhibitory effects of alcohol or benzodiazepines, leading to an assault, even homicide. Sampling in such cases may take place in ordinary healthcare units, often under the surveillance of a police officer or other authority, or at a police medical department. Extreme care should be taken to ensure that the appropriate procedures are followed in collection of any specimens.

A current topic in clinical forensic toxicology is drug-facilitated crime, such as drug-facilitated sexual assault, theft, child abuse, or the use of force or coercion to forge legal documents. The substances used in these crimes typically cause anterograde amnesia, have a rapid onset of effects, leave no taste, and can easily be mixed with drinks or food. Drugs used in these cases are short-acting benzodiazepines, other hypnotics such as zopiclone or zolpidem, or psychedelics and empathogenics such as ecstasy, cannabis, or lysergic acid diethylamide (LSD). Other potential substances include γ-hydroxybutyrate (GHB), clonidine, chloral hydrate, scopolamine, tetryzoline, fentanyl, or neuroleptic drugs.

The investigation of drug-facilitated crime is a challenge to the laboratory since typically several hours elapse before the victim seeks help. Collection of both blood and urine is always advisable, and a hair sample may provide some useful additional information. Discriminating between voluntary and involuntary drug use is a major problem and requires consultation between the toxicologist, the police, and the clinician to maximize the interpretation of any

Table 2 Indications of forensic toxicological investigation in living persons

Driving under the influence of alcohol or drugs
Clinical forensic toxicology
Victims and offenders of assault
Drug-facilitated crime
Drug-running
Child welfare
Drunkenness in office
Drug testing programs
Workplace
Schools
Armed forces
Correctional facilities
Treatment of drug dependent persons
Doping Control

results (*see* **Drug-Facilitated Sexual Assault** for further information).

Workplace Drug Testing

Workplace drug testing started in the United States after some conspicuous accidents in transportation and aviation in the 1980s. As early as 1986, all federal employees working in critical positions concerning safety and security were tested, according to an executive order. Additionally, several private companies began to test employees in risk areas such as transportation and the nuclear industry. As some of these companies were operating overseas, testing programs were extended internationally. In Europe, testing has been in force since the 1990s. The typical drugs tested are included in the Mandatory Guidelines for Federal Workplace Drug Testing Programs of SAMHSA (Substance Abuse and Mental Health Services Administration, USA): amphetamines, cannabis, cocaine, opiates, and phencyclidine. In Australia and Europe, phencyclidine is not routinely tested but benzodiazepines are usually included.

Drug testing is performed in most cases by urine immunoassay. Positive test results are confirmed by a specific technique, such as gas chromatography–mass spectrometry (GC-MS), and the final laboratory result is interpreted by a medical review officer. Thus, possible confounding causes such as cough medicines can be excluded. Cutoff limits for screening and confirmation are consistent among the European Union, United Kingdom, and Australia; however, in the United States cutoffs are higher for amphetamines and opiates [8].

Alternatives to urine samples are oral fluid and hair samples. See the respective articles on these specimens.

Workplace drug testing can be preemployment or random testing, and the practice can be mandatory or voluntary. Ireland and Finland are the first countries to have a special law regulating workplace drug testing. In Finland, an employee has the right to refuse participating in a test, but in such a case the medical review officer cannot make a statement on the suitability of the employee for a given task, which may result in rejection. Although it is easy to understand and accept the rationale behind drug testing at a workplace, in schools, in the armed forces, etc., there is a balance with the individual's civil liberties.

Doping

Doping refers to the unfair use of drugs or methods to enhance performance in sports. The prohibited list was first published in 1963 under the leadership of the International Olympic Committee. Since 2004, the World Anti-Doping Agency (WADA) is responsible for the preparation and publication of the banned substances list. The prohibited substances are divided into three main categories: prohibited in-competition, out-of-competition, and in particular sports. Substances that are prohibited at all times include anabolic agents, hormones and related substances, β-2 agonists, agents with antiestrogenic activity, and diuretics and other masking agents. The prohibited methods include enhancement of oxygen transfer, chemical and physical manipulation, and gene doping. Substances that are prohibited in-competition include stimulants, narcotics, cannabinoids, and glucocorticosteroids. Substances prohibited in certain sports include alcohol and β-blockers. WADA gives laboratories accreditation to perform doping analysis and follows the performance of the laboratories. The accredited laboratories should also comply with the requirements of the international standard, ISO/IEC 17025. It is noteworthy that doping is not only a violation against rules of fair play but also, in some countries, a criminal act. As the list of prohibited substances is extensive, a doping laboratory has to utilize several screening and confirmation methods to cover all common drugs [9]. Some of the substances are also endogenous, e.g., testosterone and erythropoietin, and in these cases special techniques, such as carbon isotope ratio or metabolite ratio determination, are required.

Terrorism

Forensic toxicological expertise is crucial in preventing and investigating terrorist attacks. In a well-known case in Japan, a deviant religious group released the warfare agent sarin. In 1994, in Matsumoto close to 600 inhabitants were suddenly suffering from strange symptoms and seven victims died. The next year in a Tokyo subway, up to 5500 passengers were exposed to sarin and 12 died [10].

In both the cases, the acetylcholine esterase activity of the victims was lowered and sarin was detected. Another tragedy took place in 2002 in Moscow, when Chechen terrorists captured Dubrovka theater with about 800 spectators. On the third day, the Russian special forces attacked the theater and made unconscious both the terrorists and the victims by using aerosol. Rescue forces were delayed and 129 hostages were killed. After the incident, the Russian health minister admitted that the gas was a fentanyl derivative. The symptoms also suggested exposure to an opioid-type knock-out gas or aerosol.

A difficult and unparalleled toxicological case was the poisoning of the ex-Bulgarian citizen Georgi Markov in 1978 in London, when he was assassinated with a pellet containing ricin. During an election campaign in 2004 in Ukraine, the presidential candidate Viktor Yushchenko was suddenly taken seriously ill. Among other symptoms he developed facial eruption, which was finally recognized as chloracne, and the diseased condition was found to be poisoning with dioxin. In the above-mentioned cases of terrorism, the substances used are not on the ordinary analysis repertoire of forensic laboratories, and even all the reference materials required for analysis are not freely available.

The Tylenol tampering incident in 1982 in Chicago, USA, revealed a new pattern of crime: someone had switched Extra Strength Tylenol (acetaminophen) capsules with similar capsules containing cyanide, resulting in the death of seven people. Since then, comparable tampering misdeeds have occurred, and Tylenol itself was again tampered with in 1986. These incidents have promoted the use of tamper resistant packages, protective seals, and labels.

Environmental Toxicology

Pollution has resulted in severe environmental disasters and caused diseases and deaths. Perhaps the worst manmade global problem is air pollution due to gaseous and particle emission. In a classic case, the London smog of 1952, it is known that the excess mortality during the period was 3500–4000 people. Death came typically to the elderly, who were the first to suffer from cardiac or respiratory diseases. In these types of cases, identification of a single victim is a difficult task, as is also the identification of the culprit because of collective responsibility. However, there are many other more local environmental disasters, where individual victims and offenders can be identified. Such cases are cadmium pollution (Itai-itai disease), methylmercury poisoning (Minamata disease), and PCBs (polychlorinated biphenyls)/PCDFs (polychlorinated dibenzofurans) poisoning (Yusho disease), all of which were first discovered in Japan; the dioxin disaster in Seveso, Italy, and the methylisocyanate accident in Bhopal, India.

For legal proceedings, exposure must be verified by toxicological analysis. Certainly pollution of the environment is *per se* a criminal offence, and laboratory results are needed to verify the incident. An exceptional series of poisonings happened in 1996 in Caruaru, Brazil, where a private dialysis center used defectively purified water: 117 out of 136 patients in the treatment became ill and 50 died. The cause of the disaster was an algal hepatotoxin, a microcystin, which was present in the water used in the treatment.

Database Research and Epidemiology

Conventionally, forensic toxicological investigation has been primarily used to resolve individual cases. However, a lot of information that is valuable to society and is otherwise difficult or impossible to obtain can be extracted from casework. To give statistics additional epidemiological value, the size and structure of the target population, sampling procedure, and coverage of analytical processes must be known thoroughly. An example of the possibility of utilizing a cumulative database of individual cases is the statistical approach that is used when compiling the fatal and control concentrations of drugs found in postmortem femoral blood [11]. Generally, prospective studies produce more well-defined information than retrospective since there is better information on the circumstances, the presence of other contributing drugs, and pathology processes (*see also* **Postmortem Toxicology: Interpretation**).

When reporting fatal poisonings within an area, the number of cases related to a particular drug or drug category is usually presented on an annual basis. However, these kinds of statistics do not contribute to the evaluation of drug safety, and a more informative picture of fatal poisonings is obtained if the fatalities are related to drug consumption. A fatal toxicity index, where the number of deaths associated with

a particular drug is divided by the number of prescriptions for the drug, can be used to rank drugs or drug categories [12]. As the number of prescriptions is not always available, the consumption of a drug can be estimated by the quantity of sales using the concept of Defined Daily Dose per 1000 inhabitants per day [13]. This approach can reveal differences in intrinsic toxicities, prescription practices, and abuse potentials of therapeutic drugs. It has been shown that, among antidepressants, the risk of death related to older tricyclic drugs is much higher than with newer selective serotonin reuptake inhibitors, while other newer antidepressantsmoclobemide, mirtazapine, and venlafaxine fall in between.

References

[1] Peters, F.T. & Maurer, H.H. (2002). Bioanalytical method validation and its implications for forensic and clinical toxicology: a review, *Accreditation and Quality Assurance* **7**, 441–449.

[2] Bertomeu-Sánchez, J.R. & Nieto-Galan, A. (eds) (2006). *Chemistry, Medicine, and Crime*, Watson Publishing International, Sagamore Beach.

[3] Flanagan, R.J. & Connally, G. (2005). Interpretation of analytical toxicology results in life and at post-mortem, *Toxicology Review* **24**, 51–62.

[4] Pounder, D.J. & Jones, G.R. (1990). Post-mortem drug redistribution – a toxicological nightmare, *Forensic Science International* **45**, 253–263.

[5] Prouty, R.W. & Anderson, W.H. (1990). The forensic science implications of site and temporal influences on post-mortem blood – drug concentrations, *Journal of Forensic Sciences* **35**, 243–270.

[6] World Health Organization (1993). *ICD-10 International Statistical Classification of Diseases and Related Health Problems*, tenth revision, Part 1 and 2, Geneva.

[7] Esmail, A. (2005). Physician as serial killer, the Shipman case, *The New England Journal of Medicine* **352**, 1843–1844.

[8] Penders, J. & Verstraete, A. (2006). Laboratory guidelines and standards in clinical and forensic toxicology, *Accreditation and Quality Assurance* **11**, 284–290.

[9] Trout, G.J. & Kazlauskas, R. (2004). Sports drug testing – an analyst's perspective, *Chemical Society Reviews* **33**, 1–13.

[10] Asai, Y. & Arnold, J.L. (2003). Terrorism in Japan, *Prehospital and Disaster Medicine* **18**, 106–114.

[11] Druid, H. & Holmgren, P. (1997). A compilation of fatal and control concentrations of drugs in postmortem femoral blood, *Forensic Science* **42**, 79–87.

[12] Cassidy, S. & Henry, J. (1987). Fatal toxicity of antidepressant drugs in overdose, *British Medical Journal* **295**, 1021–1024.

[13] Vuori, E., Ruohonen, A., Penttilä, A., Klaukka, T. & Lahti, T. (1989). Fatal poisonings with antidepressants in Finland 1985–1987, *Acta Psychiatrica Scandinavica* **80**(Suppl 354), S55–S60.

ERKKI VUORI AND ILKKA OJANPERÄ

Toxicology: Initial Testing

Introduction

Toxicological analysis, be it forensic, clinical, environmental, workplace, drug abuse, or doping control, principally involves the detection of chemical substances potentially harmful to living organisms. Chemical analysis is used to detect the presence of these substances, measure their concentrations, and interpret this in relation to their relative toxicity. Paracelsus (1493–1541) said, "All things are poison and nothing is without poison, only the dose permits something not to be poisonous."

Our society is surrounded by a large number of poisons (including chemicals, drugs, pesticides, common household products, etc.), and the task of the toxicologist is a formidable one to attempt to detect as many of these substances as possible in a toxicology screen.

Forensic toxicology requires a number of analytical techniques to exclude the presence of any poisons. For most laboratories, that may be analyzing hundreds or thousands of samples per year, it is not practical to try and cover all of these compounds simultaneously and an appropriate approach is to screen for the common drugs of abuse, prescription and nonprescription drugs in the first instance, and then perform any specific analyses that may be indicated by the case history. In most forensic toxicology laboratories, a drug screen will include a range of immunoassay tests, alcohol analysis (either by direct liquid injection or by headspace gas chromatography), headspace analysis for volatile organic compounds and a broad-based gas chromatography (GC) or high-performance liquid chromatography

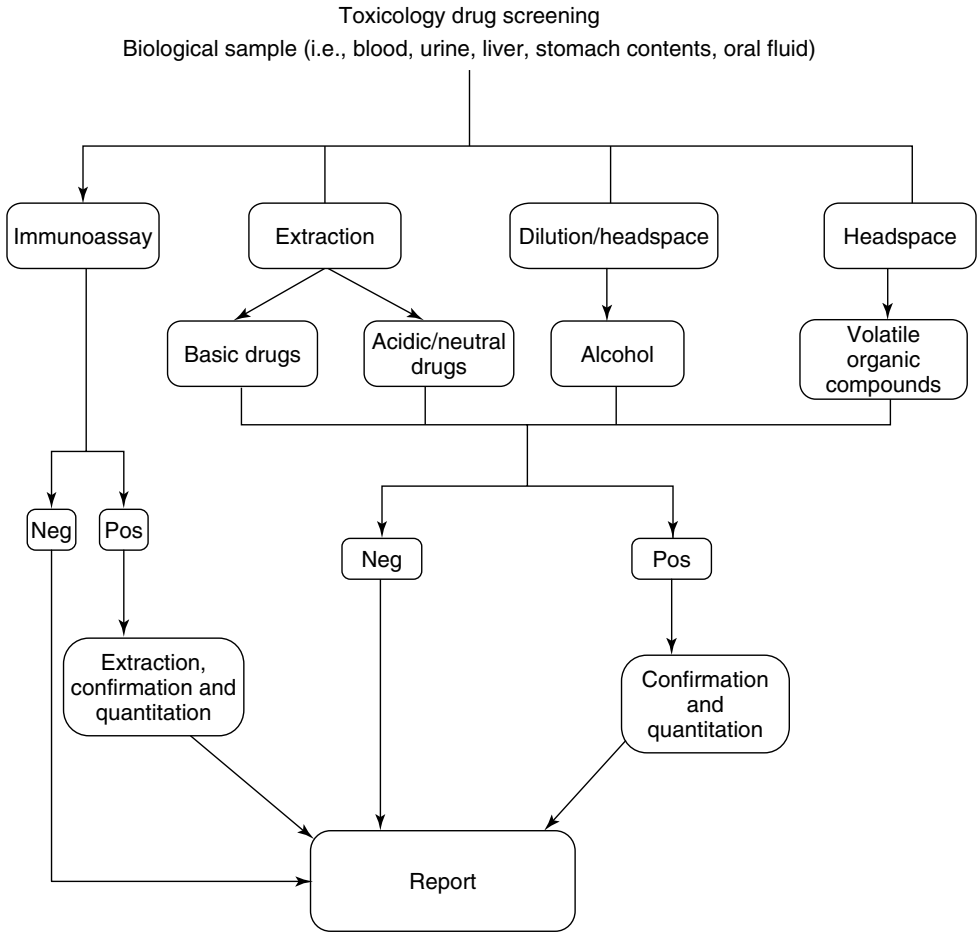


Figure 1 Flowchart showing common screening procedures

(HPLC) procedure (Figure 1). Thin layer chromatography (TLC) may also be used in some laboratories. Nitrogen–phosphorus detectors (NPD) are commonly used in GC screening as the large majority of drugs contain nitrogen, whereas many endogenous compounds (fatty acids, cholesterol, and other lipids) are nonnitrogenous. Electron capture detectors (ECD) have also been used for the screening of halogenated compounds (e.g., many benzodiazepines contain halogens).

Similarly, HPLC systems are often coupled to an ultraviolet (UV) detector or to a photodiode array detector (DAD). With the improvement in technology and the decrease in costs, mass spectrometer (MS) detectors are now increasingly replacing the NPD and ECD in GC systems and UV and DAD in HPLC systems. There are a number of articles in the literature relating to systematic toxicology analysis (STA) [1–4].

Sample Selection

It is important in toxicology that the appropriate sample is selected for screening, as this will have significant relevance for the interpretation of the results. In cases where it is necessary to only show that a subject has used a drug, it would be more suitable to analyze urine or hair as these samples are more likely to show past exposure to a drug. However, if the presence of a drug is necessary to show the involvement of that drug in

some actions of the subject, i.e., cause of death, driving under the influence (DUI), or criminal activity then the best specimens for analysis would be blood or oral fluid. In postmortem cases, it is also preferable to analyze peripheral blood as this is the least likely site to experience any postmortem changes in drug concentrations. Some laboratories may choose to screen blood from the heart, as this will contain higher concentrations of the drug; however, it is important that any further quantitative analysis is carried out on peripheral blood for interpretation purposes. As an example if 11-nor-9-carboxytetrahydrocannabinol (carboxy-THC) was detected in a urine sample from a person suspected of DUI, it could not be used to support the charge of DUI (unless there is other evidence available, e.g., sobriety tests) as this metabolite of tetrahydrocannabinol (THC) can be detected in the urine of frequent users for several weeks after use. It also may be found from exposure to passive smoking. However, determining the concentration of THC in blood can be used to show the recent use of cannabis. A simple guide for sample selection is shown in Table 1.

For selection of specimens in a postmortem environment, (*see* **Toxicology: Analysis**).

Immunoassays

Immunoassay in forensic toxicology is primarily used to screen biological samples for the presence of drugs or drug classes. They offer a rapid and convenient way of screening large numbers of samples from a variety of different matrices (blood, urine, oral fluid, etc.) for a number of drugs and drug classes. A negative immunoassay result avoids the need for more complicated and expensive analytical procedures. They require little or no sample preparation, although some techniques require pretreatment of blood or tissue specimens. Any laboratory can use immunoassay with methods that range from on-site testing for the analysis of a single sample to fully automated systems capable of handling thousands of samples per day.

Immunoassays are based on the interaction of the target compound (antigen) with a corresponding antibody. For drug testing, the immunoassay uses an antibody specific for the nominated drug or drug class and a labeled form of the same drug or of the antibody to produce a measurable signal.

Although immunoassays are extremely useful in screening large numbers of samples, it must be remembered that positive results should always be confirmed by a more specific technique (e.g., gas chromatography–mass spectrometry (GC–MS) or liquid chromatography–mass spectrometry (LC–MS)). Immunoassays will only give an indication of a drug or drug group being present and should only be used for screening purposes and cannot be used as confirmation of the drug or drug group.

The basic immunoassay technique involves the competition of a fixed amount of labeled drug with an aliquot of the sample for the specific binding sites on a fixed quantity of antibody. At equilibrium, the proportion of labeled drug molecules bound is inversely proportional to the number of unlabeled drug molecules (Figure 2).

Immunoassays can be divided into two types.

1. Homogeneous immunoassays

These are assays that do not require the separation of the antibody bound drug from the unbound (free) drug before measurement of the signal. They include assays that rely on optical change (e.g., UV absorption, fluorescence, or luminescence) where there is a difference between the signals from the bound and the unbound labeled drug. Examples of this type of immunoassay are EIA (enzyme immunoassay), CEDIA (cloned enzyme donor immunoassay),

Table 1 Time frame for drugs in different biological specimens

Specimen	Drug(s) present	Time frame	Interpretation
Blood	Parent/metabolites	Hours	Recent use
Oral fluid	Parent drug	Hours	Recent use
Urine	Metabolites	Hours/days	Use
Hair	Parent drug	Months	Use

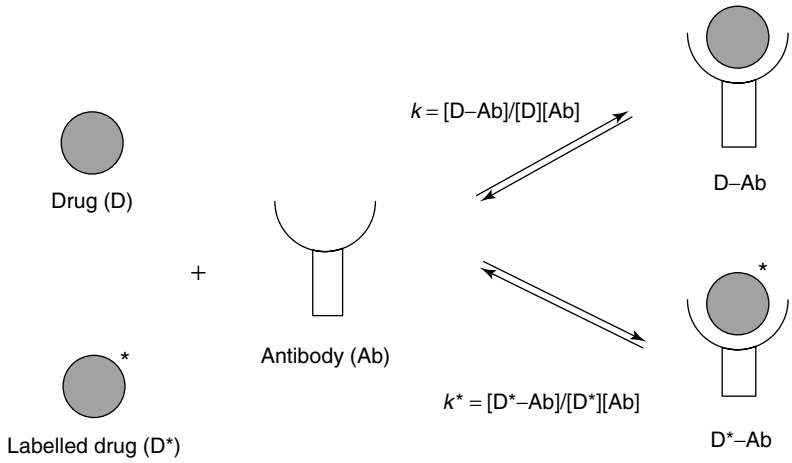


Figure 2 Competitive binding process in immunoassay

FPIA (fluorescence polarization immunoassay), and microparticle methods.

2. Heterogeneous immunoassays

These assays require the separation of the antibody bound drug and the unbound drug before measurement of the signal. This is required as there is no difference in the signals from the antibody-bound and the unbound drugs. Examples of this type of immunoassay are ELISA (enzyme-linked immunosorbent assay), RIA (radioimmunoassay), and chemiluminescence immunoassays.

There are two distinct advantages of the heterogeneous immunoassays over the homogeneous type. First, using a wash step any potential endogenous interfering compounds can be removed, which may be produced by whole blood or highly discolored urine. This also means that there are no preliminary sample extraction steps required. The second advantage is that these assays have lower limits of detection. This is best illustrated by being able to use whole blood in heterogeneous assays, whereas the extraction of the blood sample is often required for homogeneous assays.

Specificity and Cross-reactivity

It is important with any immunoassay technique to determine how the assay responds to other drugs relative to the drug used as a calibrator. The cross-reactivity of the assay is important, as it will determine the suitability of the test. For example, determining the use of *Cannabis* by the analysis of a urine sample requires the immunoassay to have good cross-reactivity with the major urinary metabolite, carboxy-THC. However, by contrast, the assay needs to have good cross-reactivity to THC if the samples to be analyzed are oral fluid or hair, as the relative parent drug concentration is greater than metabolite in these specimens.

There are many immunoassay kits available commercially and specifications as to their sensitivity and specificity should be available for each kit. Table 2 shows cross-reactivity data for an opiate ELISA assay.

Table 2 Typical cross-reactivity of common drugs in an opiate ELISA assay

Analyte	% Cross-reactivity
Morphine	100
Codeine	100
Hydrocodone	93
Morphine-3-glucuronide	90
6-acetylmorphine	88
Dihydrocodeine	85
Diacetylmorphine (heroin)	60
Morphine-6-glucuronide	47
Hydromorphone	6.7
Oxycodone	5.2
Oxymorphone	1.4
Meperidine	<0.1
Imipramine	<0.1

Basic Principles of Immunoassay

All immunoassays use antibodies and the way they are produced determines the performance of the system. Antibodies are immunoglobins (Ig), which are produced by beta-lymphocytes in response to an immunogen. Drug molecules are not large enough to produce an immune response so drug immunogens are made by conjugation of the drug of interest to a large carrier protein (such as bovine sera albumin (BSA)) to form an immunogen, this process is known as *haptenization*. Animals such as rabbits, sheep, or goats are used as hosts for antibody production of polyclonal antisera and mice for the production of monoclonal antibodies. The conjugation is usually via a $-NH_3$ or $-COOH$ group on the drug.

Specificity of the antibodies is determined by the position on the drug molecule at which the protein is conjugated. A good example of this is the morphine molecule.

If the immunogen is produced at the 3-position, the resultant antibody will display cross-reactivity to all metabolites of heroin (diacetylmorphine) and morphine (e.g., morphine-3-glocuronide). It will also cross-react with codeine (3-*O*-methylmorphine). Thus, the antibody produced by the haptenization at the 3-position will produce a broad cross-reacting antibody of all opiates.

If the haptenization occurs at the 6-position then antibodies with greater specificity to morphine relative to codeine and morphine-3-glucuronide would be expected. These antibodies should also display good cross-reactivity with the active metabolite morphine-6-glucuronide and 6-acetylmorphine. Conjugating via the nitrogen following derivitization would produce a more specific assay for morphine (Figure 3).

All antibodies have an affinity constant that is the measure of the strength of the binding between the antigen and the antibody. This binding is reversible, noncovalent, and will reach equilibrium. Antibodies with a higher affinity will bind faster than those with lower and will produce better performance in immunoassays. Polyclonal antisera contain a complex mixture of higher affinity antibodies raised through the immunization process, whereas a monoclonal antibody is a singe entity produced from a single antibody cell with lower affinity.

Enzyme Immunoassay (EIA)

EIAs can be either homogeneous (enzyme multiplied immunoassay technique (EMIT) [5–8]) or heterogenous immunoassays (ELISA [9–12]).

EMIT is a liquid phase assay based on the enzyme activity of a drug-labeled enzyme. The enzyme attached to the labeled drug is glucose-6-phosphate

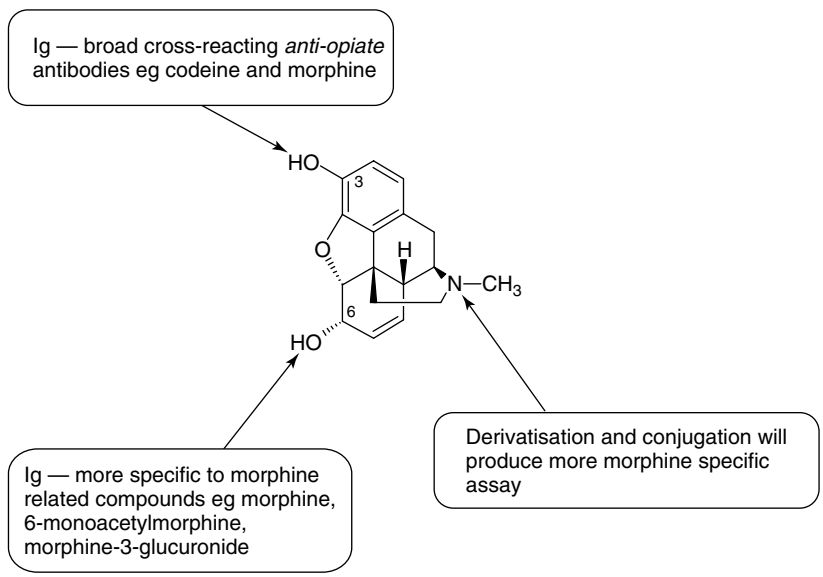


Figure 3 Different antibody binding site on a morphine molecule

dehydrogenase (G6PDH) and antibodies raised against the drug of interest modulate the activity of the enzyme. In the presence of glucose-6-phosphate, the enzyme reduces nicotine adenine dinucleotide (NAD) to NADH and the resulting increase in absorbance is measured spectophotometrically at 340 nm. If the drug of interest is present in the sample, it will compete with the drug-labeled enzyme for binding to the antibody and allow unbound enzyme to become active. This will result in an increase in absorbance at 340 nm. Thus, as the amount of drug in the sample increases then there will be a corresponding increase in enzyme activity and an increase in the rate of NADH production, which will be measured spectrophotometrically by monitoring the λ_{max} at 340 nm.

As EMIT relies upon a color change in the solution, it is not suitable for samples such as blood or highly discolored urine without sample pretreatment. Another limitation of the technique is that, as well as interference for other cross-reacting compounds, any compounds in the matrix that may inhibit the enzyme activity will also affect the result.

ELISA is probably the most commonly used immunoassay in most forensic toxicology laboratories. Commercial ELISA kits are supplied as a dry 96-well microplate with each well precoated with the antidrug antibodies. Table 3 gives a list of common commercial drugs of abuse kits.

ELISA is a heterogeneous, solid phase assay that requires the separation of reagents. There are two techniques available for antigen measurement, competitive immunoassay, and sandwich or double-antibody technique.

The most common ELISA technique is the competitive immunoassay (Figure 4). In antigen competitive assays, the antibody is bound to the well and the sample and labeled antigen are added, the labeled and unlabelled antigen competes for the limited number of antibody binding sites. After an incubation period, an immune complex is formed from the antigen–antibody binding and any unbound antigen is removed by washing. A substrate and chromogen is then added to the immune complex and a reaction between the enzyme and the substrate causes the chromogen to become colored. The more antigen (drug) present in the sample the less color is produced. The reaction is stopped after a preset time by the addition of dilute acid and the color measured at the λ_{max} at 450 nm. ELISA is a more sensitive technique than EMIT and is less subject to matrix effects. It is regularly used to analyze postmortem blood samples.

Radioimmunoassay (RIA)

Radioimmunoassay (RIA) was one of the first heterogeneous immunoassays, but problems associated with the disposal of radioactive waste and rapid decaying of the radioactive isotopes have seen it being replaced with nonisotopic EIAs [7, 8, 13, 14]. The most common type is the antibody-coated tube utilizing γ-emitting I^{125}-labeled antigen. Separation of the bound and free antigen is accomplished by simply decanting the liquid leaving the bound fraction coated to the tube. The technique is extremely sensitive and specific and tends to be less susceptible to matrix effects, especially for analysis of whole blood samples, but when radioactive substances are involved it requires special precautions and sophisticated apparatus (γ-radiation counter). The technique has been largely replaced by the ELISA method.

Cloned Enzyme Donor Immunoassay (CEDIA)

Cloned enzyme donor immunoassay (CEDIA) is a more recent homogeneous immunoassay that uses

Table 3 Common commercial ELISA drug screening kits

Buprenorphine	Barbiturates	Benzodiazepines
Fluoxetine	Cocaine/benzoylecgonine	Fentanyl
Methadone	Ketamine	Lysergic acid diethylamide (LSD)
Opiates	Oxycodone/oxymorphine	Morphine
Propoxyphene	Haloperidol	Phencyclidine
Tramadol	Tricyclic antidepressants	Hydromorphone
Amphetamine/ methylenedioxyamphetamine (MDA)	Methamphetamine/ methylenedioxymethamphetamine (MDMA)	Cannabinoids (tetrahydrocannabinol/carboxy-THC)

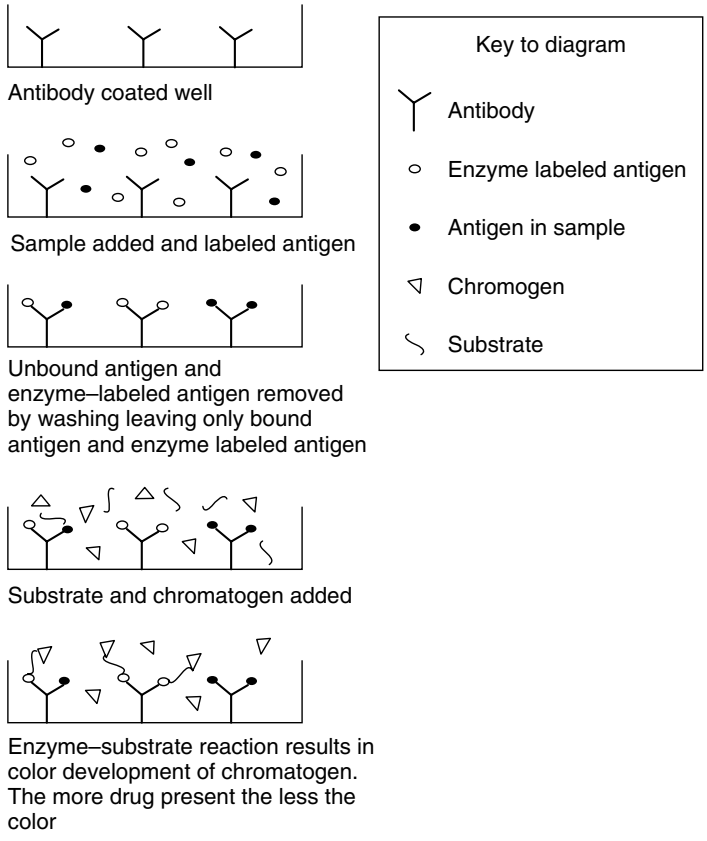


Figure 4 Competitive immunoassay process

the binding of an antibody to change the activity of an enzyme [7, 8, 15–18]. CEDIA uses genetically engineered fragments of β-galactosidase from *Escherichia coli* as the enzyme label. The enzyme is present as two inactive fragments known as the *enzyme acceptor* (*EA*) and *enzyme donor* (*ED*). The ED contains the small portion of enzyme missing from the larger EA fragment. Antibodies that bind to the hapten that is conjugated to the ED fragment prevent the reassociation of the enzyme and therefore the enzyme activity. As the amount of drug in the sample increases, the amount of bound antibody to the ED fragment decreases, resulting in an increased enzyme activity due to the reassociation of EA and ED. The enzyme hydrolyzes chlorophenolred-β-galactoside (CPGR) to chlorophenolred (CPR) and galactoside and the enzyme activity can be measured spectrophotometrically.

Fluorescence Polarization Immunoassay (FPIA)

Fluorescence polarization immunoassay (FPIA) is also a heterogeneous immunoassay technique that uses fluorescein-labeled antigen [8, 19]. The labeled antigen can rotate freely when not bound to an antibody; however, when it is bound, its ability to rotate is markedly reduced. Light at the excitation wavelength of fluorescein is passed through a polarized filter. The amount of polarized light emitted is reduced by freely rotating fluorescein but there is little loss of polarization by the relatively stationary antibody bound fluorescein. As any drug present in the sample competes with the fluorescein-labeled antigen for the antibody, there is a reduction in the amount of fluorescein bound and therefore a reduction in the amount of fluorescence through the polarized filter.

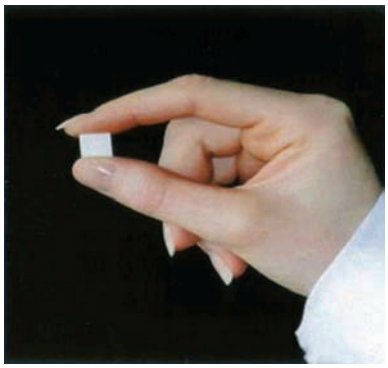


Figure 5 Biochip – actual size (Photograph courtesy of Randox) [Reproduced with permission from Randox.]

Chemiluminescence Immunoassay

The principle used in this immunoassay technique is similar to other heterogeneous assays but offers greater sensitivity using chemiluminescence. The signal produced by chemiluminescence immunoassays is generated by the emission of light during a chemical reaction. The glow of light over a short period is captured by a charged coupled device (CCD) camera, which also enables detection of separate assays. This technique has been developed to assay up to 10 different drugs, 8 drug classes, in the 1 sample analysis using Randox evidence Biochip Array Technology. This is achieved by attaching specific drug antibodies to discrete test regions on the biochip. The CCD will simultaneously record the light emission from each of the biochips and the image is processed to quantify and validate the signal. Figure 5 shows the actual size of a biochip and Figure 6 the discrete test regions on the chip.

Kinetic Interaction of Microparticles in Solution (KIMS)

Kinetic interaction of microparticles in solution (KIMS) is a homogeneous immunoassay in which the labeled compound is a microparticle with several drug molecules attached [8, 17, 20, 21]. If the sample being analyzed is drug free, then the microparticle with drug molecules attached conjugates with several antibody molecules and aggregates to form a larger particle. This aggregate will scatter transmitted light and as the reaction proceeds the abundance will increase. If drug molecules are present in the sample, they will compete with the conjugates bound to antibodies and result in a decrease in the rate of absorbance increase in proportion to the concentration of drug present.

Sample Extraction

There are two major methods for sample extraction:

Solvent Extraction (Liquid–Liquid Extraction)

This is the most frequently used separation technique employed in toxicology laboratories worldwide. Most

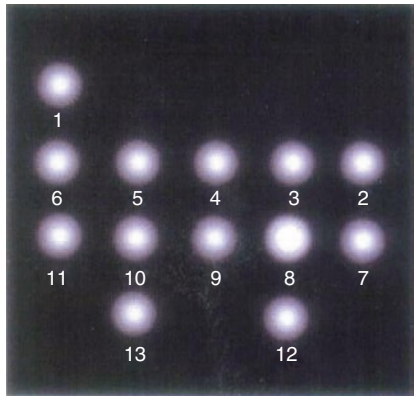


Figure 6 Magnified image showing luminescing discrete test regions of the processed drugs of abuse biochip (Randox evidence) (Photographs courtesy of Randox) [Reproduced with permission from Randox.]

drugs have some degree of polarity so it is important to select a solvent that will maximize the extraction of the target drugs but minimize the extraction of endogenous compounds. By adjusting the pH of the specimen drugs can be separated into basic and acidic drugs.

Making the specimen basic (pH 9) by the addition of ammonia or sodium borate buffer will favor the extraction of weakly basic drugs. Buffering the specimen to pH 4 by the addition of an acidic buffer (potassium dihydrogen phosphate) will allow the extraction of acidic and neutral drugs.

Basic drugs comprise the majority of drugs found in forensic cases and are often present at low concentrations in blood. To improve the detection of these drugs, it may be necessary to incorporate a clean-up step (or back extraction) in the extraction scheme to remove more of the endogenous compounds and still retain the drugs of interest. This can be particularly important when analyzing postmortem blood specimens that can often contain a multitude of endogenous compounds formed postmortem. Once the specimen has been extracted at basic pH into an appropriate solvent (such as butyl chloride), it is back-extracted into aqueous acid solution (usually 0.1 M sulfuric acid), the acidic and neutral compounds will remain in the organic layer. The acidic layer is separated from the organic layer and is then made alkaline (sodium hydroxide) and re-extracted into organic solvent, which is then separated from the aqueous layer and concentrated, usually under a stream of nitrogen. Care should be exercised in the evaporation step, as some of the more volatile compounds can be lost if subjected to heat; amphetamine and methylamphetamine are prime examples.

The concentrated extract can now be analyzed using the appropriate chromatographic technique. A typical basic/neutral drug extraction scheme is shown in Figure 7 and a back-extraction for basic drugs in Figure 8.

Acidic drugs of forensic interest tend to be present in much higher concentrations than basic drugs and

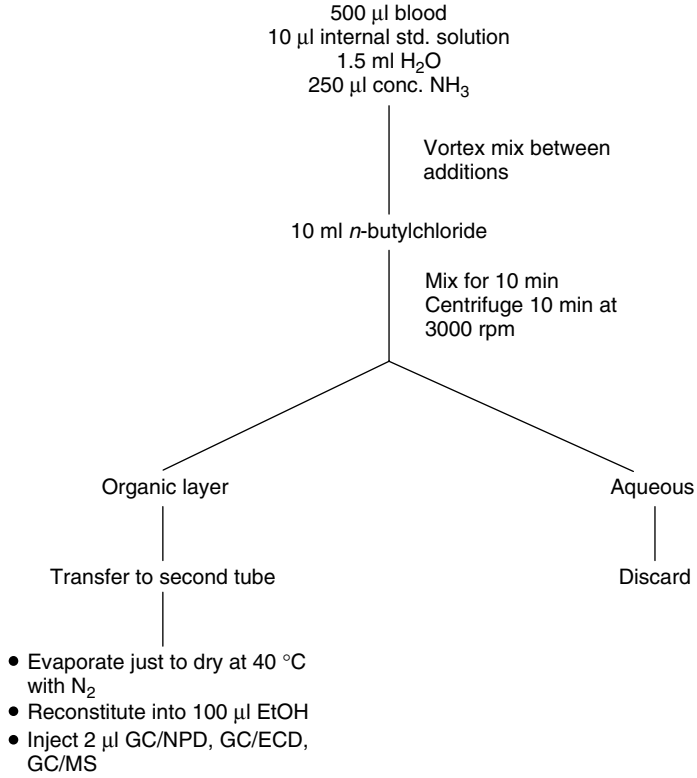


Figure 7 Extraction scheme for basic and neutral drugs

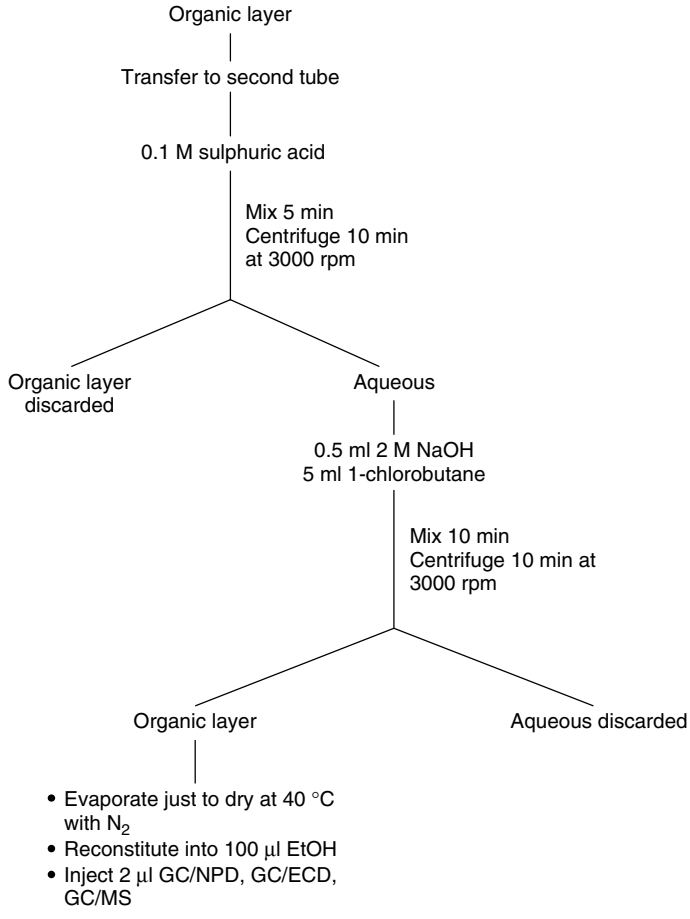


Figure 8 Back-extraction scheme for basic drugs

the need for a cleanup step in the extraction is generally not required. If a cleanup step is required it can be achieved by the partitioning between immiscible solvents of differing polarities. The more lipid soluble endogenous compounds (sterols and fatty acids) will partition into the less polar solvent (e.g., petroleum ether), whereas the more polar drugs will partition into the more polar solvent (e.g., acetonitrile). A typical extraction scheme for acidic and neutral drugs is shown in Figure 9.

Solid Phase Extraction (SPE)

Solid phase extraction (SPE) techniques are becoming more popular in forensic toxicology laboratories because of their better extraction efficiencies, particularly with more polar drugs, and the ability to automate the extraction using liquid handling systems. Initial difficulties with postmortem samples, which often contain large amounts of particulate material and are often clotted, which in the past could easily plug the fine SPE absorbent packing, have now been overcome by better sample preparation techniques (sample dilution and centrifugation) and improved SPE column technology. There are several SPE column manufacturers and many published screening techniques in the literature [22–26]. Most commonly used sorbents are porous silica particles with octadecyl (C_{18}) or other hydrophobic alkyl groups surface bonded. SPE sorbents have evolved from liquid chromatography phases and have developed over time to include a number of different functionalities e.g., $-C_2$, –CH, –phenyl, –CN, –diol, and mixed-mode bonded phases.

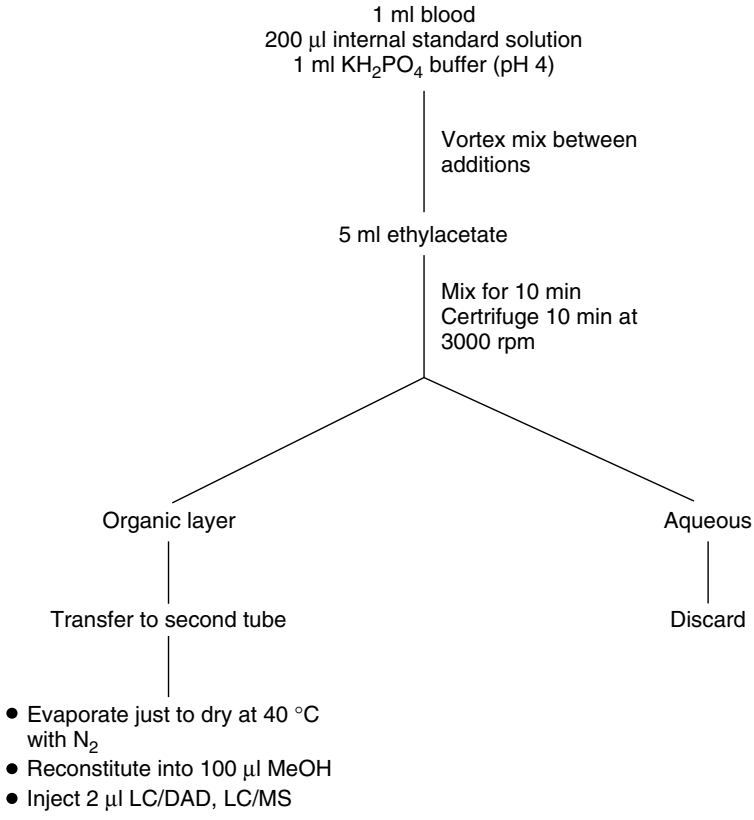


Figure 9 Extraction scheme for acidic/neutral drugs

There are predominantly three different mechanisms for the retention of compounds by the sorbent:

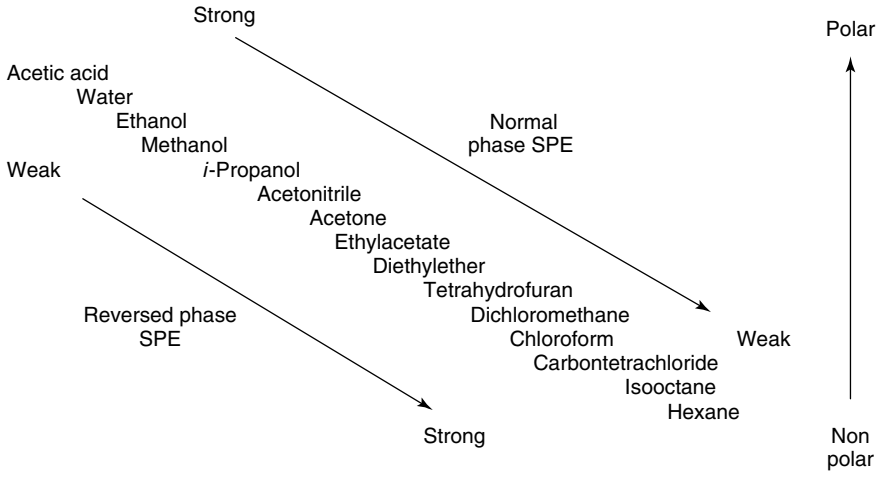
1. Reverse phase (polar liquid phase and nonpolar solid phase): Compounds are retained by hydrophobic interactions (nonpolar–nonpolar interactions and dispersion forces).
2. Normal phase (nonpolar liquid phase and polar solid phase): Compounds are retained by hydrophilic interactions (polar–polar interactions and hydrogen bonding).
3. Ion exchange: It is electrostatic attraction of charged group on compound and charged group on the sorbent.

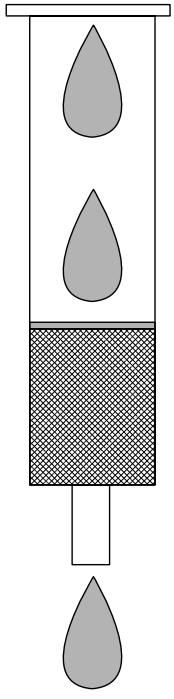
Table 4 shows the characteristics of solvents commonly used in SPE.

The basic procedure for SPE is a five-step process as shown in Figure 10, and a typical SPE using a combination of a strong cation exchange and a C8 sorbent (UCT XTRACT®, 200 mg 3 ml^{-1} P/N XRDAH203) for the analysis of opiates in blood is shown in Figure 11.

An alternative technique to liquid–liquid extraction and SPE is solid phase micro extraction (SPME) that has been developed for the analysis of volatile and some nonvolatile compounds [27–32]. This technique incorporates sample extraction, concentration, and sample introduction in one step and is fast, requires no solvents, is reusable, and can be adapted to any GC. The apparatus consists of a fused silica fiber coated with a stationary phase that is attached to a stainless steel plunger. After inserting the plunger through a septum into a vial containing the sample the fiber is exposed to either the headspace or the liquid for 20–30 min. The equilibration time can be reduced by increasing the ionic strength of the sample, stirring, or by heating the sample. After equilibration, the fiber is retracted back

Table 4 Characteristics of solvents commonly used in SPE





1. Select the appropriate SPE cartridge – capacity, bed weight, and sorbent type
2. Condition the cartridge. Use an appropriate solvent or aqueous buffer to wet the sorbent material and allow the sample to come into contact with the packing. Do not allow the SPE packing to dry between conditioning and sample addition as there will be uneven contact between sample and packing resulting in non-reproducible efficiency and drug recoveries.
3. Add the sample. Pass the sample solution slowly through the column using either vacuum or positive pressure at a flow rate of ~2 ml min^{-1}.
4. Wash the column. Remove unretained and unwanted compounds from the column using the same solution that the sample was dissolved in or solvents of sufficient polarity that they will not remove the compound(s) of interest.
5. Elute the compound(s). The compound(s) of interest are eluted using a solvent of sufficient polarity to recover the analyte(s) but leave any unwanted compounds that have not been removed by the wash step

Figure 10 Solid phase extraction sequence

into the plunger and inserted into the GC injection port where it is desorbed by exposing the fiber again (Figure 12). This also serves to recondition the fiber, which can be reused for up to 100 further analyses. The fiber may also be re-extracted into an organic solvent and analyzed by HPLC. There are number of different stationary phases available to suit the type of analysis. For less polar compounds, a poly-dimethoxysiloxane phase is used and depending upon the molecular weights (mw) of the compounds of interest the film thickness will vary. Generally, as the thickness of the film increases the lower the

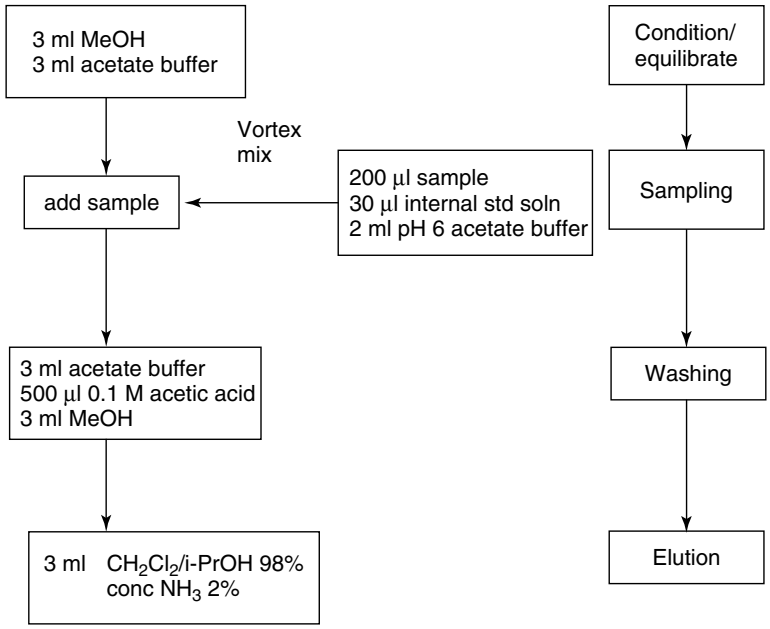


Figure 11 Analysis of opiates using mixed-mode SPE

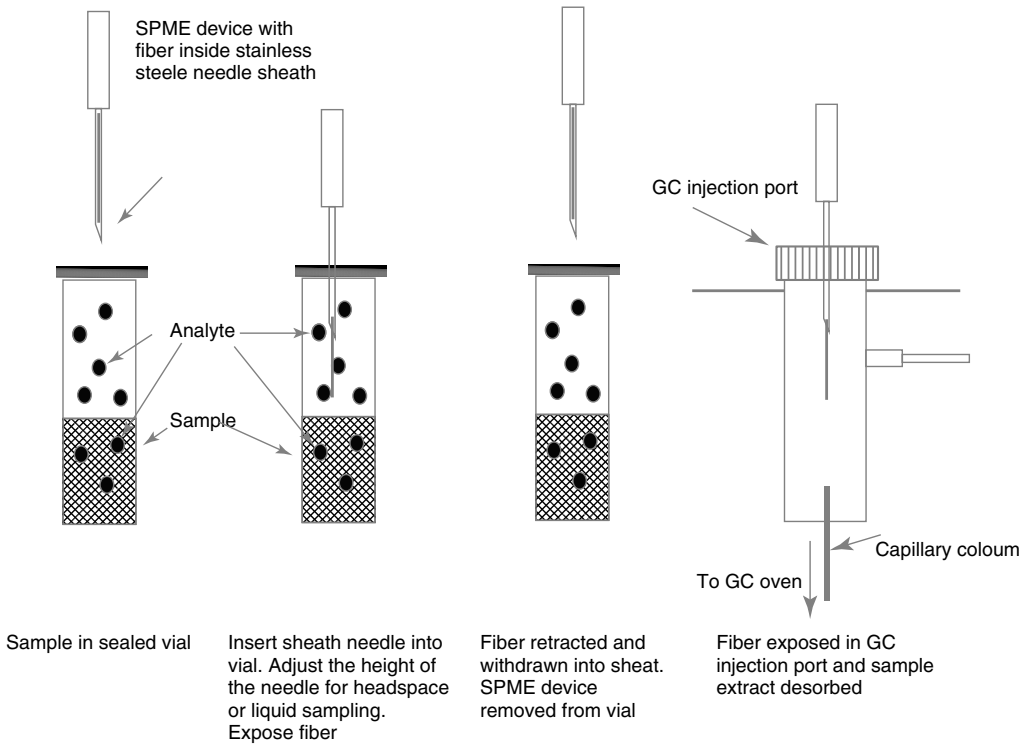


Figure 12 SPME sample headspace adsorption and GC desorption process

molecular weight (MW) of the compounds retained. Polyacrylate films are used for polar compounds (*see also* **Postmortem Toxicology: Laboratory Analysis**).

Screening by Gas Chromatography

GC has been the technique most widely used in forensic toxicology laboratories. If a drug has sufficient volatility for it to exist in the vapor phase at temperatures up to 350 °C and does not decompose at these temperatures, then there is a high probability that it can be analyzed by GC. Compounds in a specimen extract can be separated by introducing an aliquot of the specimen extract (sample injection) into a column that contains a stationary phase that has a continuous flow of an inert gas (e.g., helium) through it. The column is contained in an oven with controlled temperature. The components within the extraction mixture will spend differing periods of time in the stationary phase, depending upon their affinity for it and the time for the molecules to reach the end of the column will vary (retention time). The greater the affinity of the molecule to the stationary phase the longer they will be retained on the column and hence the greater the retention time. At the end of the column, there is a detector that produces a signal that is proportional to the amount of the compound passing through it. This signal will be processed and recorded. Each compound that elutes from the column will have a characteristic retention time, which can be defined, as the time interval between sample injection and detector response.

Stationary Phases and Column Supports

Few solid stationary phases are used for screening in forensic toxicology; however, blood–alcohol analysis is an exception. The majority of forensic laboratories analyze blood and other liquid biological specimens for alcohol concentration using either graphitized carbon black or polymer conventional packed columns. These phases are extremely good at separating low molecular weight hydrocarbons (C1–C10). Some examples of these are Carbopak C with 0.2% Carbowax 20M, Chromosorb 101–108, Porapak, and Tenax. Blood–alcohol analysis can be achieved by dilution of the specimen with in internal standard (e.g., *n*-propanol) and direct injection of an aliquot of the aqueous mix (Figure 13).

To prolong the life of the column and reduce the frequency of changing injection port liners, tungstic acid can be added and the diluted solution centrifuged to precipitate any protein in the sample. Another technique is to add an internal standard to an aliquot of the sample and seal the vial. After heating the sample in the vial to ~60 °C, the headspace is then

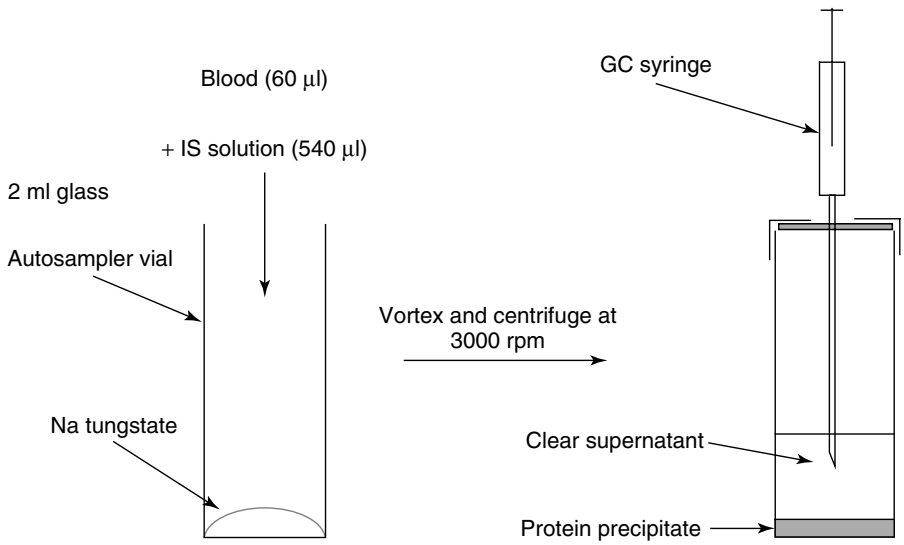


Figure 13 Blood–alcohol analysis – sample preparation for direct injection technique

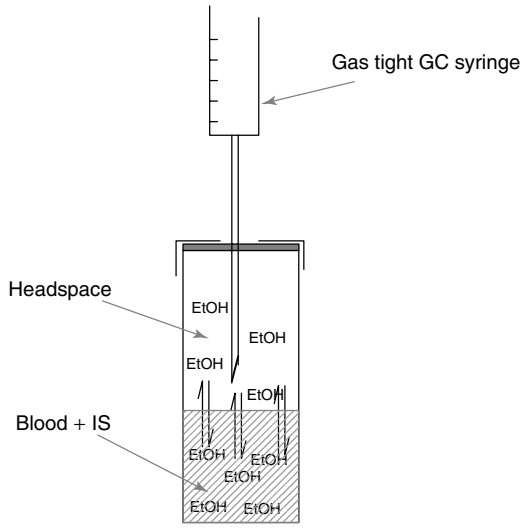


Figure 14 Blood–alcohol analysis: headspace technique

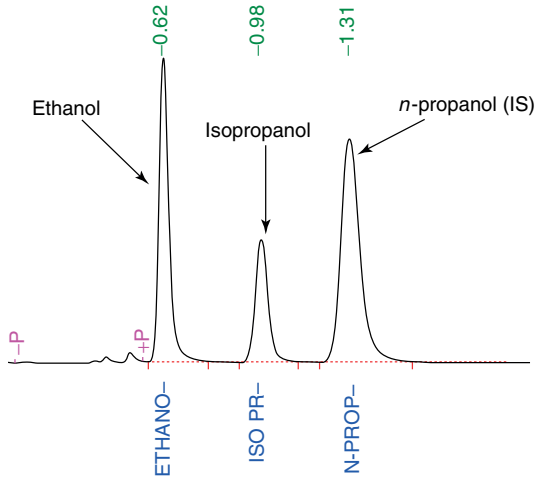


Figure 15 Chromatogram of ethanol analysis in blood: 2 m × 2 mm i.d. glass column packed with Carbopak C with 0.2% Carbowax 20 M. Oven 100 °C, injection port 220 °C, FID 220 °C, and He carrier gas flow 30 ml min^{-1}

injected onto the GC column via the heated injection port (Figure 14).

A chromatogram of ethanol analysis using a 2 m × 2 mm internal diameter (i.d.) glass column packed with Carbopak C with 0.2% Carbowax 20 M is shown in Figure 15.

The most commonly used stationary phases for drug screening are the liquid stationary phases comprising polysiloxanes. These stationary phases are liquids coated to the inside of a fused silica capillary column. The dimensions of the capillary column can vary between 0.1 and 0.32 mm i.d. and can be from 10 to 60 m long.

Polysiloxanes are characterized by a repeating siloxane backbone and each silicon (Si) atom can have two functional groups attached to it. The type and number of these functional groups will distinguish the properties of the stationary phase. The least polar polysiloxane is when it is 100% substituted with methyl groups, and this is generally designated as *XX-1* (where XX is usually the manufacturers identifier e.g., DB-1 and HP-1). By substituting more polar functional groups such as phenyl and cyano groups the polarity of the column can be changed (Figure 16). If 5% of the Si atoms contain two phenyl groups this will be known as *XX-5*; if 50% contain phenyl groups this is known as *XX-17*. Some polysiloxanes have two different functional groups present on the same Si atom, e.g., cyanopropyl and phenyl to increase the polarity, an example of this is the XX-1701 column (14% cyanopropyl-phenyl-methylpolysiloxane).

In total, 100% methyl-substituted polysiloxanes have been used extensively in producing retention time databases and are commonly used for screening purposes. In more recent times, columns that have up to 5% of the methyl groups substituted with phenyl

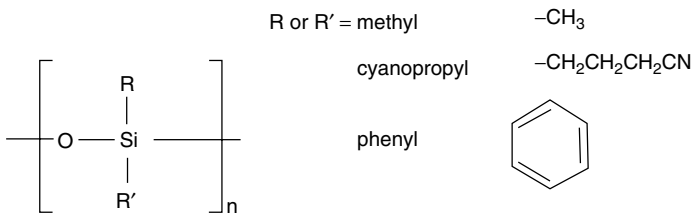


Figure 16 Polysiloxane backbone with differing functional groups

have produced essentially a nonpolar stationary phase that have improved thermal stability and as a result produce much less column bleed, which is especially useful in systems that use MS as the method of detection.

Split/Splitless Injectors

An inlet splitter allows a high flow of carrier gas into the injector, while still allowing a low flow ($1-2\,\mathrm{ml\,min^{-1}}$) through the capillary column. The excess carrier gas is vented via a split line to the atmosphere. The ratio of the inlet splitter flow and the column flow is known as the *split ratio* and determines what proportion of the injected sample passes on to the column. The lower the split ratio, the more sample is introduced onto the column.

When the injector is operated in the splitless mode it allows the entire injected sample to pass onto the column. This is particularly useful for extracts that contain very low concentrations of drugs. After a fixed time, the split vent is opened to allow a larger flow of carrier to flow through the injector, which will purge any remaining sample from the injection port. By keeping the temperature of the column oven below the boiling point of the solvent, the sample can be focused into a narrow band and reduce any peak broadening. Auto injectors allow the analysis of large numbers of samples, usually overnight when the laboratory is unattended.

Headspace Analysis

The headspace sampling technique can be used to screen biological samples for the presence of volatile hydrocarbons (e.g., petroleum products, propane, and butane) [33]. To increase the sensitivity of this technique, it is usual to sample a large volume of headspace (50 ml) and pass it through an adsorbent (e.g., Porapak Q) to trap the volatiles, which are then thermally desorbed onto a GC column. For complex mixtures, it is usual to analyze the headspace by capillary GC/MS. Figure 17 shows a typical total ion trace for a headspace sample of petrol in blood.

Detectors

For example, some detectors like the flame ionization detectors (FID) are nonspecific in nature and will respond to almost all compounds that pass through the GC column, while there are others that will specifically respond to compounds that contain particular atoms. For example, the NPD will detect compounds that contain nitrogen or phosphorus and ECD will detect compounds that have electron capturing atoms or functional groups (e.g., halogens, nitro groups and carbonyl groups). The NPD is extremely useful in drug screening since most drugs contain nitrogen while extraction solvents and many of the coextracted endogenous compounds in biological samples do not (e.g., fatty acids, lipids, and sterols). There are a number of relative retention databases that use standard capillary columns that are available for compound identification [34]. The ECD is a highly sensitive selective detector particularly useful to detect benzodiazepines and halogenated pesticides and herbicides. These detectors also offer greater sensitivity and selectivity over nonspecific detectors. This greater sensitivity can be utilized by derivatizing the

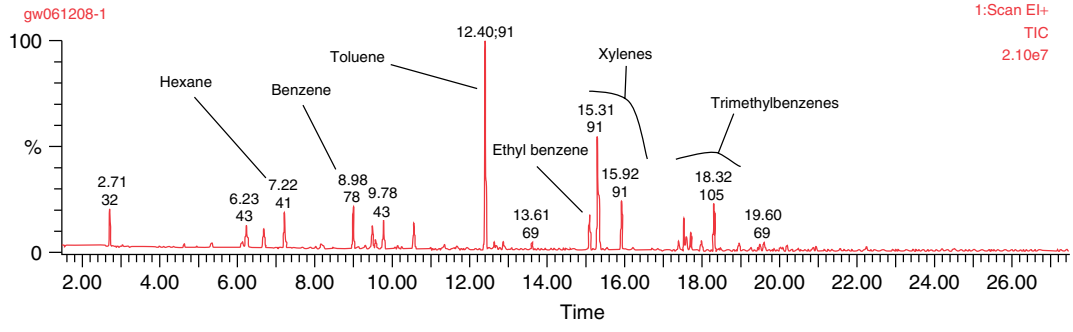


Figure 17 Total ion chromatogram of a headspace sample: 60 m × 0.25 mm i.d. BP-1 capillary (film thickness 1 μm), oven temperature 45 °C (6 min) to 260 °C at 10 °C min^{-1}

compounds with functional groups containing halogen, e.g., trifluoroacetic, pentafluoropropionic, and heptafluorobutyric anhydrides.

MS detectors have become increasingly popular as they provide the opportunity for unique identification of compounds. The most commonly used MS detector uses an electron impact technique to ionize the compounds eluting from the capillary column, which then fragment in a characteristic and reproducible way. These fragments are focused and accelerated into a mass filter (quadrupole or ion-trap), which rapidly scans for masses usually in the range up to about 1000 amu. The abundance of each mass at a given scan time is measured and results in a mass spectrum. This characteristic mass spectrum can be searched against a number of commercial mass spectral libraries (e.g., PMW (Pfleger, Maurer, and Weber), NIST, and Wiley libraries). The advances in computer technology allow the searching of vast library data extremely fast.

For added sensitivity, the quadrupole MS can be operated in the selected ion mode (SIM) where only preselected masses for the compounds of interest are collected. This is a useful technique for targeted screening of drugs but is limited for general unknown screening.

The ion-trap detector offers similar sensitivity when operated in either the full scan or SIM mode due to its longer ion collection period. Mass spectra produced by the two different mass detectors are sometimes quite different as they are characterized by the conditions under which they are run, this can make comparison between different instruments difficult.

Dual Detector Systems

It is not uncommon in forensic toxicology laboratories to use a dual detector screening system. There are several different combinations that can be used and the sample can be split at the injection port and two columns of differing polarity used to separate the drugs in the extract (Figure 18) or the effluent from the column can be split between two different detectors (Figure 19).

When using the single injection dual column approach the effluent from both the columns will be detected by the same type of detector (usually NPD or ECD) and a relative retention time (RRT) database can be set up for both the columns (e.g., DB-1 and DB-17). Searching the RRT of unknown peaks in the two chromatograms will provide better discrimination

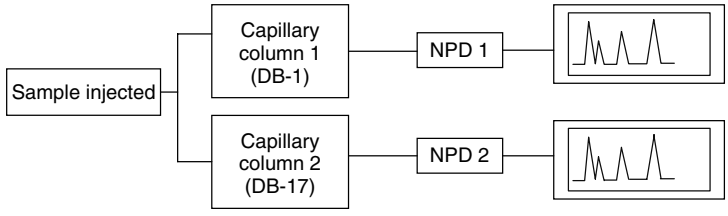


Figure 18 Sample injection split between two different columns

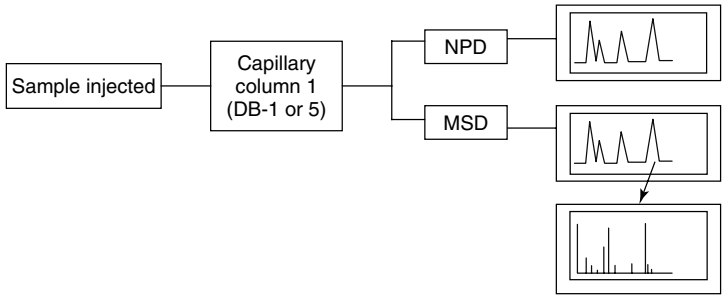


Figure 19 Column effluent split between two detectors

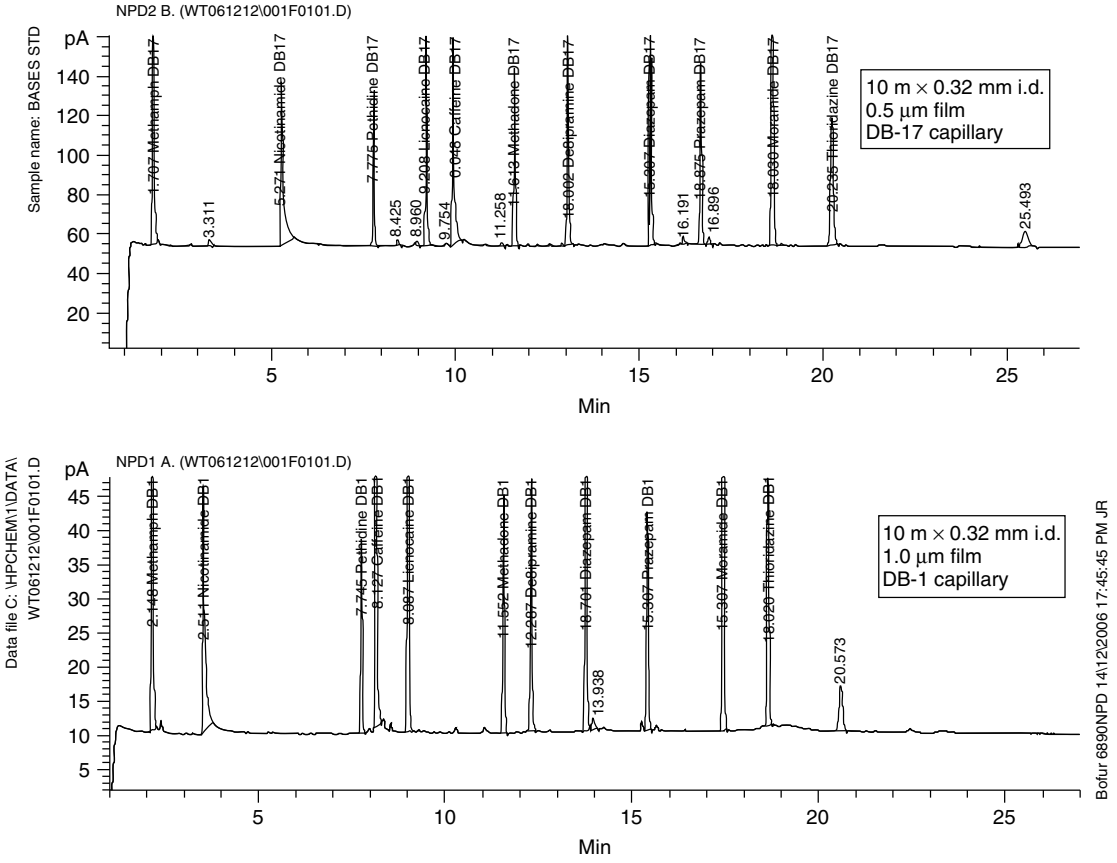


Figure 20 Dual column NPD chromatogram. Column oven 100 °C (0.5 min) to 280 °C (8 min) at 10 °C/min, injection port 250 °C, detector 325 °C, He carrier 4.9 ml min^{-1}, and splitless injection

as to the identity of the unknown than if data from a single chromatogram was used (Figure 20).

A much more common configuration is to split the effluent from the end of the column to two detectors, one being an NPD and the other an MSD. By using the NPD chromatogram to identify peaks of interest (many endogenous peaks will not produce a NPD response) the mass spectrum of the peak at the corresponding retention time can be obtained from the MS total ion chromatogram (Figure 21).

Gas Chromatography–Mass Spectrometry (GC–MS)

The use of GC–MS to screen biological samples for drugs is also a popular technique [26, 35–37]. However, because extracts from biological samples often contain many endogenous compounds the technique requires extensive sample preparation before GC–MS analysis if a general drug screen is required. Total ion chromatograms (TIC) containing endogenous compounds can often mask low concentrations of drugs. An alternative technique is to use selected ion extraction for characteristic ions of drugs of interest, these extracted ion chromatograms can show the presence of drugs even if the TIC does not. Unfortunately, this technique can limit the number of drugs that can be screened for, but is very good for selective drug screening. Figure 22 shows a TIC and extracted ion chromatogram of a blood extract.

High-Performance Liquid Chromatography (HPLC)

At one point of time HPLC was only used for the analysis of compounds that could not be analyzed

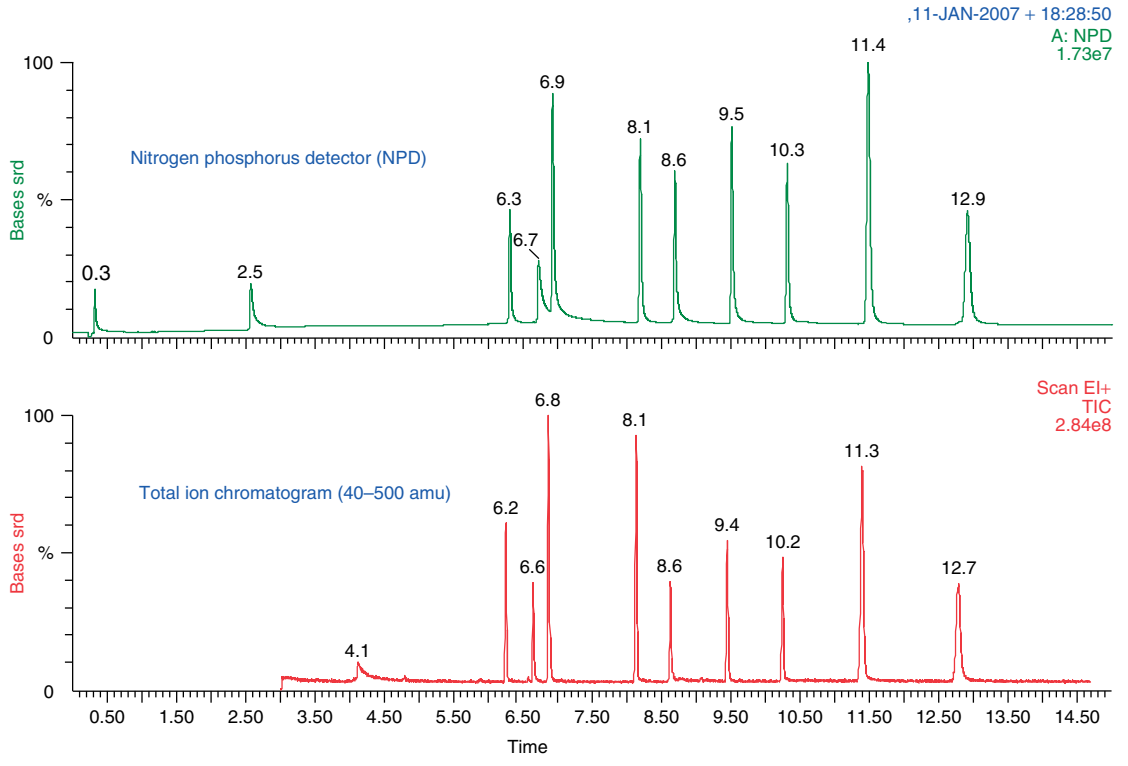


Figure 21 Dual detector (MSD and NPD). Column: 15 m × 0.32 mm id DB-5 ms 0.5-μm film thickness: 70 °C (0 min) to 280 °C (4.5 min) at 20 °C min^{-1}, He carrier 4 ml min^{-1}, injection port 250 °C, NPD 250 °C, and MSD transfer line 260 °C

by GC (nonvolatile thermally unstable compounds). Since the introduction of the DAD in the mid-1980s the ability to produce a UV spectrum has dramatically increased the use of HPLC systems for routine drug screening [38–41]. The identification of drugs is based on the combination of retention time and UV spectrum. Figure 23 shows the HPLC-DAD trace from an acid drug extract.

HPLC is the separation of components in a mixture based on the selective partitioning of compounds between a liquid-mobile phase and a stationary phase. The sample is introduced through an injection port into the mobile phase stream, delivered by the high-pressure pump, and moves through the column where the separation takes place. As the compounds are separated they are detected at the column outlet with a flow through detector or are transferred into the MS.

The selection of the stationary phase and mobile phase is important as alterations to one or the other can lead to improved compound separation particularly when there are endogenous compounds present.

Columns and Mobile Phases

Typical HPLC columns are 10–30 cm long with internal diameters of 2–4.6 mm. The columns are filled with small diameter particles (3–10 μm in diameter). For drug screening, the most common type of packing is based on bonded organic polymers; examples of these are styrene–divinylbenzene copolymers, used in reversed phase LC and octadecylsilane silica columns. Polymeric materials have an advantage over silica-based ones as they are more stable over a larger pH range.

Adsorption is the major mechanism in silica normal phase column/nonpolar mobile phase where the separation is achieved by the competition between analyte molecules and mobile phase molecules for the adsorption sites on the silica. Polar groups are retained more strongly than nonpolar ones and

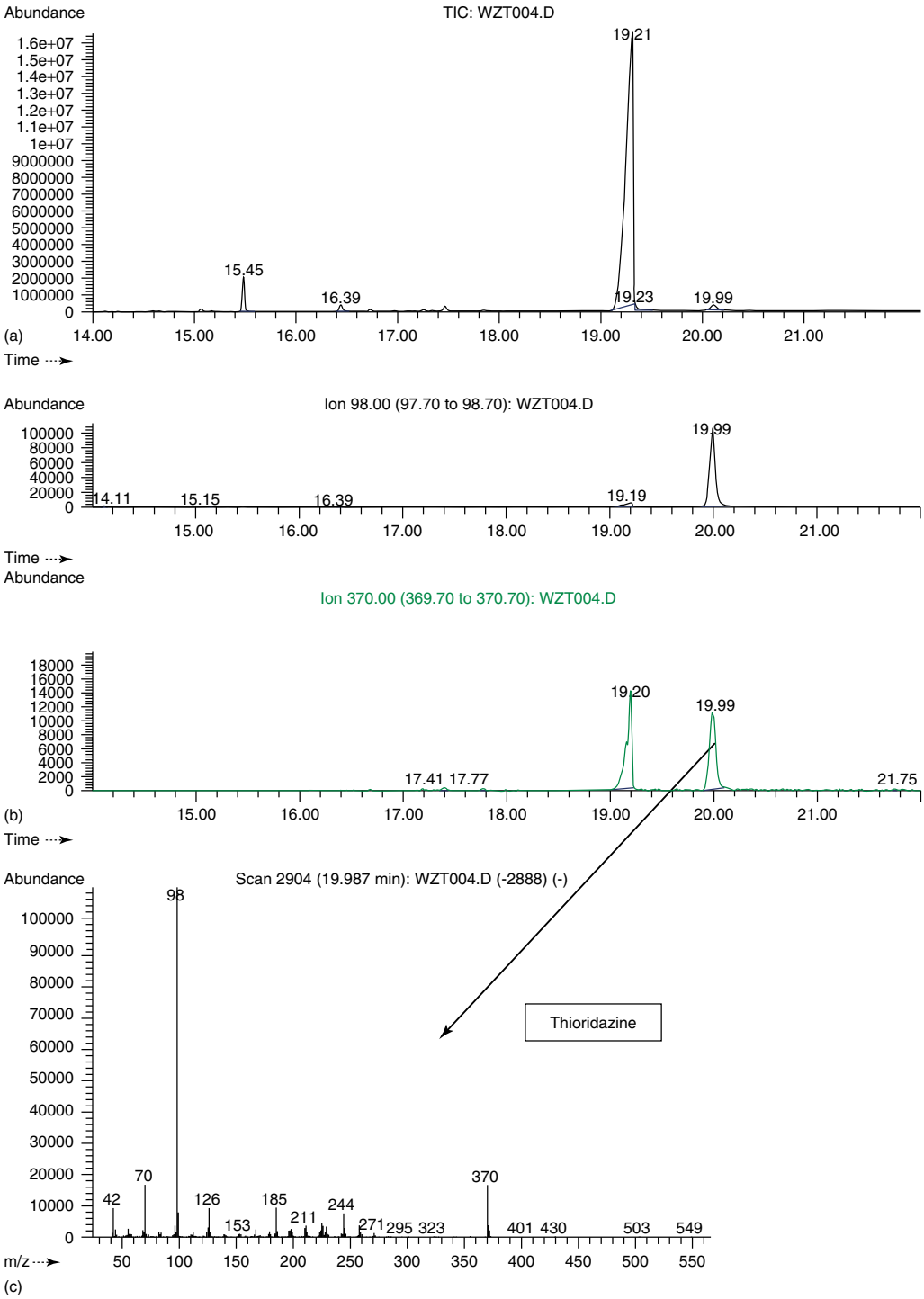


Figure 22 (a) Total ion chromatogram of blood extract. (b) extracted ion (m/z 98 and 370), and (c) mass spectrum – thioridazine

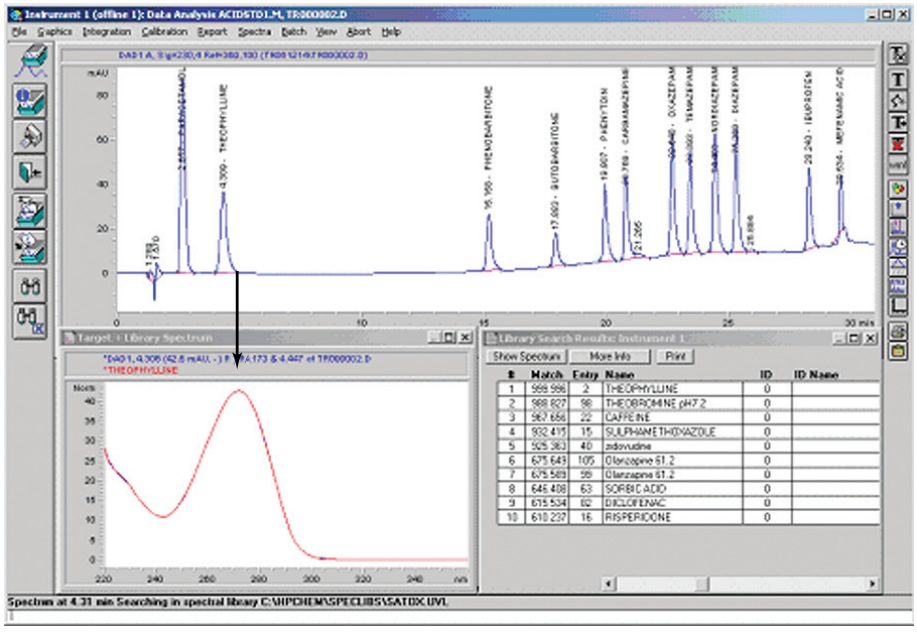


Figure 23 HPLC-DAD chromatogram of a drug mix showing the UV spectra and library search of theophylline. Agilent Zorbax SB-C18 (5 µm, 2.1 × 150 mm) with XDBC8 guard column (5 µm, 4.6 × 12.5 mm)

hence retention times increase as the polarity of the compound increases. Water is very strongly bound to silica and care should be taken to minimize the amount of water in the eluent solvent. The elution strength of the mobile phase can be varied by changing the proportion of miscible solvents of differing polarity.

By increasing the polarity of the elution solvent, the adsorption mechanism ceases to be the major mechanism, leading to difficulties in predicting retention behaviors. Often mobile phases consist of predominantly methanol modified with an aqueous buffer (sodium hydrogen phosphate or ammonium nitrate) and retention of the compounds can be controlled by pH, ionic strength, or methanol: aqueous ratio.

With reversed phase systems, the retention of compounds is controlled by the hydrophobic interactions between the drugs and alkyl groups bonded to the packing material. As the polarity of the analyte increases the retention will decrease. Polar mobile phases (usually acetonitrile or methanol) are modified with an aqueous buffer. Increasing the polarity of the mobile phase will result in an increase in retention time. The chain length of the alkyl group bonded to the packing will also affect the retention of compounds with an increase in retention and an increase in chain length ($C_{18} > C_8 > C_2$).

Liquid Chromatography–Mass Spectrometry (LC–MS) and Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS)

In recent years, the use of liquid chromatography–mass spectrometry (LC–MS) [42–48] and liquid chromatography–tandem mass spectrometry (LC–MS/MS) [49–53] in drug screening has developed rapidly. There are two main interfaces that allow the introduction of a liquid effluent from the HPLC column at atmospheric pressure into the high-vacuum MS system: electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI). In ESI, the column effluent is formed into charged droplets using a strong electrical field with ions produced from evaporating solution droplets. In APCI, the effluent is first nebulized and then heated with ions produced by gas-phase ion–molecule reactions. A series of low-pressure chambers and ion focusing lenses is used to optimally transport the ions into the MS. Both the ESI and APCI produce limited spectral information compared to electron impact mass spectra and

for identification purposes it is often necessary to perform secondary fragmentation. Figure 24 shows the separation of benzodiazepines and metabolites in blood using LC-APCI/MS/MS. Table 5 shows the selected precursor ions $[M + H]^+$ and the resultant product ions used for the identification of the compound.

One problem that is commonly encountered in ESI LC-MS is ion suppression. This results from a reduction of ionization due to the presence of coeluting compounds. These compounds can effect the formation of the charged ions in the liquid phase. If there is sufficient ion suppression of the drug(s) in the extract then they would not be detected and

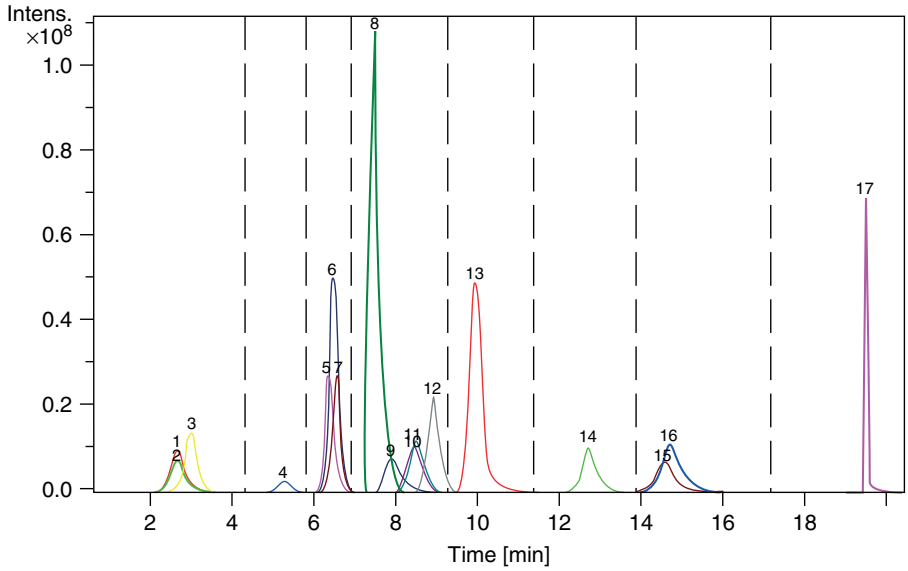


Figure 24 APCI product ion chromatogram for separation and detection of benzodiazepines: Agilent Zorbax XDB C8 column (4.6 × 150 mm, 5 µm particle size). See Table 5 for details of drugs

Table 5 Precursor ions and major product ions for each of the benzodiazepines and metabolites in Figure 24

Peak number	Drug/metabolite	Pseudo-molecular ion $[M + H]^+$	Product ion (m/z)
1	7-Aminonitrazepam	252	224
2	7-Aminoclonazepam	286	250
3	7-Aminoflunitrazepam	284	264
4	Bromazepam	316	288
5	Clonazepam	316	270
6	Nitrazepam	282	236
7	Flunitrazepam	314	268
8	Clobazam	301	259
9	Triazolam	343	308
10	Alprazolam	309	281
11	Lorazepam	321	275
12	Oxazepam	287	241
13	Temazepam	301	255
14	Nordiazepam	271	243
15	Diazepam	285	257
16	Midazolam	326	291
17	Prazepam	325	271

this could lead to incorrect results being reported [46, 54]. For this reason, the sample preparation should reduce the amount of endogenous compounds in the matrix extract that can lead to ion suppression and the effect of other compounds used in the analysis (salts, drugs, metabolites, and deuterated internal standards) should be evaluated before using this technique routinely.

References

[1] Maurer, H.H. (2000). Screening Procedures for Simultaneous Detection of Several Drug Classes Used for High Throughput Toxicological Analyses and Doping Control. A review, *Combinatorial Chemistry & High Throughput Screening* **3**, 467–480.

[2] de Zeeuw, R.A. (1997). Drug screening in biological fluids. The need for a systematic approach, *Journal of Chromatography B* **689**, 71–79.

[3] Maurer, H.H. (1999). Systematic toxicological analysis procedures for acidic drugs and/or metabolites relevant to clinical and forensic toxicology and/or doping control, *Journal of Chromatography B* **733**, 3–25.

[4] Lillsunde, P., Michelson, L., Forsström, T., Korte, T., Schultz, E., Ariniema, K., Portman, M., Sihvonen, M. & Seppäla, T. (1996). Comprehensive drug screening in blood for detecting abused drugs or drugs potentially hazardous for traffic safety, *Forensic Science International* **77**, 191–210.

[5] Holler, J.M., Bosy, T.Z., Klette, K.L., Wiegand, R., Jemionek, J. & Jacobs, J. (2004). Comparison of the Microgenics CEDIA® Heroin Metabolite (6-AM) and the Roche Abuscreen® ONLINE Opiate Immunoassay for the Detection of Heroin Use in Forensic Urine Samples, *Journal of Analytical Toxicology* **28**, 489–493.

[6] Way, B.A., Walton, K.G., Koenig, J.W., Eveland, B.J. & Scott, M.G. (1998). Comparison between the CEDIA and EMIT II immunoassays for the determination of benzodiazepines, *Clinica Chimica Acta* **271**(1), 1–9.

[7] Stout, P.R., Gehlhausen, J.M., Horn, C.K. & Klette, K.L. (2002). Comparison of EMIT registered II, CEDIA registered, and DPC registered RIA Assays for the Detection of Lysergic Acid Diethylamide in Forensic Urine Samples, *Journal of Analytical Toxicology* **26**, 519–523.

[8] Armbruster, D.A., Schwarzhoff, R.H., Hubster, E.C. & Liserio, M.K. (1993). Enzyme immunoassay, kinetic microparticle immunoassay, radioimmunoassay, and fluorescence polarization immunoassay compared for drugs-of-abuse screening, *Clinical Chemistry* **39**, 2137–2146.

[9] Moore, K.A., Werner, C., Zannelli, R.M., Levine, B. & Smith, M.L. (1999). Screening post-mortem blood and tissues for nine cases of drugs of abuse using automated microplate immunoassay, *Forensic Science International* **106**, 93–102.

[10] Laurie, D., Mason, A.J., Piggott, N.H., Rowell, F.J., Seviour, J., Strachan, D. & Tyson, J.D. (1996). Enzyme linked immunosorbent assay for detecting benzodiazepines in urine, *Analyst* **121**(7), 951–954.

[11] Kemp, P., Sneed, G., Kupiec, T. & Spieler, V. (2002). Validation of a Microtiter Plate ELISA for Screening of Postmortem Blood for Opiates and Benzodiazepines, *Journal of Analytical Toxicology* **26**, 504–512.

[12] Kerrigan, S. & Phillips, W.H. (2001). Comparison of ELISAs for Opiates, Benzodiazepines, Phencyclidine, and Cannabinoids in Whole Blood and Urine, *Clinical Chemistry* **47**, 540–547.

[13] Peskar, B. & Spector, S. (1973). Quantitative Determination of Diazepam in Blood by Radioimmunoassay, *Journal of Pharmacology and Experimental Therapeutics* **186**, 167–172.

[14] Owens, S.M., McBay, A.J., Reisner, H.M. & Perez-Reyes, M. (1981). 125I radioimmunoassay of delta-9-tetrahydrocannabinol in blood and plasma with a solid-phase second-antibody separation method, *Clinical Chemistry* **27**, 619–624.

[15] Armbruster, D.A., Hubster, E.C., Kaufman, M.S. & Ramon, M.K. (1995). Cloned enzyme donor immunoassay (CEDIA) for drugs-of-abuse screening, *Clinical Chemistry* **41**, 92–98.

[16] Wu, A.H., Forte, E., Casella, G., Sun, K., Hemphill, G., Foery, R. & Schanzenbach, H. (1995). CEDIA for screening drugs of abuse in urine and the effect of adulterants, *Journal of Forensic Sciences* **40**(4), 614–618.

[17] Schwettmann, L., Kulpmann, W.R. & Vidal, C. (2006). Drug screening in urine by cloned enzyme donor immunoassay (CEDIA) and kinetic interaction of microparticles in solution (KIMS): a comparative study, *Clinical Chemistry* **44**(4), 479–487.

[18] Kottenhahn, B., Meyer, L., Drasch, G., Roider, G. & Hofbauer, B. (2002). Direct Detection of Drugs of Abuse in Whole Hemolysed Postmortem Blood and Qualitative Measurement in EDTA–Plasma using the CEDIA DAU Urine Assays, *Toxichem + Krimtech* **69**, 62–72.

[19] Verstraete, A.G. & Heyden, F.V. (2005). Comparison of the Sensitivity and Specificity of Six Immunoassays for the Detection of Amphetamines in Urine, *Journal of Analytical Toxicology* **5**, 359–364.

[20] Lyons, T.P., Okano, C.K., Kuhnle, J.A., Bruins, M.R., Darwin, W.D., Moolchan, E.T. & Huestis, M.A. (2001). A Comparison of Roche Kinetic Interaction of Microparticles in Solution (KIMS®) Assay for Cannabinoids and GC-MS Analysis for 11-nor-9-Carboxy-δ^9-Tetrahydrocannabinol, *Journal of Analytical Toxicology* **25**, 559–564.

[21] Klette, K.L., Wiegand, R.F., Horn, C.K., Stout, P.R. & Magluilo Jr, J. (2005). Urine Benzodiazepines Screening using Roche Online® KIMS Immunoassay with β-Glucuronidase Hydrolysis and Confirmation by Gas Chromatography-Mass Spectrometry, *Journal of Analytical Toxicology* **29**, 193–200.

[22] Soriano, T., Jurado, C., Menéndez, M. & Repetto, M. (2001). Improved Solid-Phase Extraction Method for Systematic Toxicological Analysis in Biological Fluids, *Journal of Analytical Toxicology* **25**, 137–143.

[23] Stimpfl, T., Jurenitsch, J. & Vycudilik, W. (2001). General Unknown Screening in Postmortem Tissue and Blood Samples: A Semi-Automatic Solid-Phase Extraction using Polystyrene Resins Followed by Liquid-Liquid Extraction, *Journal of Analytical Toxicology* **25**, 125–129.

[24] Boland, D.M., Burke, M.F., Mitchell, T. & Madley, P. (2001). Development of a Generic Method to the Solid-Phase Extraction of Acidic Compounds from Complex Matrices, *Journal of Analytical Toxicology* **25**, 602–606.

[25] Yawney, J., Treacy, S., Hindmarsh, K.W. & Burczynski, F.J. (2002). A General Screening Method for Acidic, Neutral, and Basic Drugs in Whole Blood using the Oasis MCX® Column , *Journal of Analytical Toxicology* **26**, 325–332.

[26] Solans, A., Carnicero, M., de la Torre, R. & Segura, J. (1995). Comprehensive Screening Procedure for Detection of Stimulants, Narcotics, Adrenergic Drugs, and Their Metabolites in Human Urine, *Journal of Analytical Toxicology* **19**, 104–114.

[27] Snow, N.H. (2000). Solid-phase micro-extraction of drugs from biological matrices, *Journal of Chromatography A* **885**, 445–455.

[28] Kraemer, H.-J., Breithaupt, H., Junting, L., Peng, C. & Suzuki, O. (1998). Solid-phase microextraction (SPME) of drugs and poisons from biological samples , *Forensic Science International* **97**, 93–100.

[29] Mills, G.A. & Walker, V. (2000). Headspace solid-phase microextraction procedures for gas chromatographic analysis of biological fluids and materials, *Journal of Chromatography A* **902**, 267–287.

[30] Mosaddegh, M.H., Richardson, T., Stoddart, R.W. & McClure, J. (2001). Application of solid-phase microextraction technology to drug screening and identification, *Annals of Clinical Biochemistry* **38**, 541–547.

[31] Kumazawa, T., Lee, X.-P., Sato, K. & Suzuki, O. (2003). Solid-phase microextraction and liquid chromatography/mass spectrometry in drug analysis, *Analytica Chimica Acta* **492**, 49–67.

[32] Walles, M., Mullett, W.M. & Pawliszyn, J. (2004). Monitoring of drugs and metabolites in whole blood by restricted-access solid-phase microextraction coupled to liquid chromatography-mass spectrometry, *Journal of Chromatography A* **1025**, 85–92.

[33] Sharp, M.-E.,E. (2001). A Comprehensive Screen for Volatile Organic Compounds in Biological Fluids, *Journal of Analytical Toxicology* **25**, 631–636.

[34] Moffat, A.C., Osselton, M.D. & Widdop, B. (2004). *Clarke's Analysis of Drugs and Poisons*, 3rd Edition, Pharmaceutical Press, Vol. 1.

[35] Polettini, A. (1996). A Simple Automated Procedure for the Detection and Identification of Peaks in Gas Chromatography-Continuous Scan Mass Spectrometry. Application to Systematic Toxicological Analysis of drugs in Whole Human Blood, *Journal of Analytical Toxicology* **20**, 579–586.

[36] Maurer, H.H. & Joachim, W.A. (1999). Screening Procedure for detection of dihydropyridine calcium channel Blocker Metabolites in urine as Part of a Systematic Toxicological analysis procedure for acidic compounds by gas chromatography-mass spectrometry after Extractive Methylation, *Journal of Analytical Toxicology* **23**, 73–80.

[37] Maurer, H.H. & Bickeboeller-Friedrich, J. (2000). Screening Procedure for Detection of Antidepressants of the Selective Serotonin Reuptake Inhibitor Type and their Metabolites in Urine as pert of a Modified Systematic Toxicological Analysis Procedure using Gas Chromatography-Mass Spectrometry, *Journal of Analytical Toxicology* **24**, 340–347.

[38] Drummer, O.H., Kotsos, A. & McIntyre, I.M. (1993). A Class-Independent Drug Screen in Forensic Toxicology Using a Photodiode Array Detector, *Journal of Analytical Toxicology* **17**, 225–229.

[39] Maier, R.D. & Bogusz, M. (1995). Identification Power of a Standardizer HPLC-DAD System for Systematic Toxicological Analysis, *Journal of Analytical Toxicology* **19**, 79–83.

[40] Gaillard, Y. & Pépin, G. (1997). Use of high-performance liquid chromatography with photodiode array UV detection for the creation of a 600-compound library. Application to forensic toxicology, *Journal of Chromatography A* **763**, 149–163.

[41] Herzler, M., Herre, S. & Pragst, F. (2003). Selectivity of Substance Identification by HPLC-DAD in Toxicological Analysis using a UV Spectra Library of 2682 Compounds, *Journal of Analytical Toxicology* **27**, 233–242.

[42] Rittner, M., Pragst, F., Bork, W.-R. & Neumann, J. (2001). Screening Method for Seventy Psychoactive Drugs or Drug Metabolites in Serum Based on High-Performance Liquid Chromatography-Electrospray Ionization Mass Spectrometry, *Journal of Analytical Toxicology* **25**, 115–124.

[43] Kerskes, C.H.M., Lusthof, K.J., Zweipfenning, P.G.M. & Franke, J.P. (2002). The Detection and identification of Quaternary Nitrogen Muscle Relaxants in Biological Fluids and Tissues by Ion-Trap LC-ESI-MS, *Journal of Analytical Toxicology* **26**, 29–34.

[44] Niessen, W.M. (2003). Progress in Liquid chromatography-mass spectrometry instrumentation and its impact on high-throughput screening, *Journal of Chromatography A* **1000**, 413–436.

[45] Venisse, N., Marquet, P., Duchoslav, E., Dupuy, J.L. & Lachâtre, G. (2003). A General Unknown Screening Procedure for Drugs and Toxic Compounds in Serum using Liquid Chromatography-Electrospray-Single Quadrupole Mass Spectrometry, *Journal of Analytical Toxicology* **27**, 7–14.

[46] Maurer, H.H. (2005). Advances in analytical toxicology: the current role of liquid chromatography-mass spectrometry in drug quantification in blood and oral fluid, *Analytical and Bioanalytical Chemistry* **381**, 110–118.

[47] Politi, L., Groppi, A. & Polettini, A. (2005). Applications of Liquid Chromatography-Mass Spectrometry in Doping Control, *Journal of Analytical Toxicology* **29**, 1–14.
[48] Miyaguchi, H., Kuwayama, K., Tsujikawa, K., Tatsuyuki, K., Iwata, Y.T., Inoue, H. & Kishi, T. (2006). A method for screening for various sedative-hypnotics in serum by liquid chromatography/single quadrupole mass spectrometry, *Forensic Science International* **157**, 57–70.
[49] Weinmann, W., Gergov, M. & Goerner, M. (2000). MS/MS-libraries with triple quadrupole-tandem mass spectrometers for drug identification and drug screening, *Analusis* **28**, 934–941.
[50] Thieme, D. & Sachs, H. (2003). Improved screening capabilities in forensic toxicology by application of liquid chromatography-tandem mass spectrometry, *Analytica Chimica Acta* **492**, 171–186.
[51] Herrin, G.L., McCurdy, H.H. & Wall, W.H. (2005). Investigation of an LC-MS-MS (QTrap®) Method for the Rapid Screening and Identification of Drugs in Postmortem Toxicology Whole Blood Samples, *Journal of Analytical Toxicology* **29**, 599–606.
[52] Sauvage, F.-L., Saint-Marcoux, F., Duretz, B., Deporte, D., Lachatre, G. & Marquet, P. (2006). Screening of Drugs and Toxic Compounds with Liquid Chromatography-Linear Ion Trap Tandem Mass Spectrometry, *Clinical Chemistry* **52**, 1735–1741.
[53] Rivera, H.M., Walker, G.S., Sims, D.N. & Stockham, P.C. (2003). Application of liquid chromatography-tandem mass spectrometry to the analysis of benzodiazepines in blood, *European Journal of Mass Spectrometry* **9**, 599–607.
[54] Dams, R., Huestis, M.A., Lambert, W.E. & Murphy, C.M. (2003). Matrix Effects in Bio-Analysis of Illicit Drugs with LC-MS/MS: Influence of Ionization Type, Sample Preparation, and Biofluid, *Journal of the American Society for Mass Spectrometry* **14**, 1290–1294.

Further Reading

Baselt, R.C. (2004). *Disposition of Toxic Drugs and Chemicals in Man*, 7th Edition, Biomedical Publications, Foster City.
Cole, R.B. (ed) (1997). *Electrospray Ionisation Mass Spectrometry – Fundamentals, Instrumentation and Applications*, John Wiley & Sons, New York.
Curry, A.S. (ed) (1984). *Analytical Methods in Human Toxicology, Part 1*, The Macmillan Press, London.
Curry, A.S. (ed) (1986). *Analytical Methods in Human Toxicology, Part 2*, The Macmillan Press, London.
Levine, B. (ed) (1999). *Principles of Forensic Toxicology*, American Association for Clinical Chemistry.
Moffat, A.C., Osselton, M.D. & Widdop, B. (eds) (2004). *Clarke's Analysis of Drugs and Poisons*, 3rd Edition, Pharmaceutical Press, Vols. 1 and 2.
Niessen, W.M.A. (1999). *Liquid Chromatography – Mass Spectrometry*, 2nd Edition, Marcel Dekker, New York.
Telepchak, M.J., August, T.F. & Chaney, G (2004). *Forensic and Clinical Applications of Solid Phase Extraction*, Humana Press, Totowa.
Watson, J.T. (1997). *Introduction to Mass Spectrometry*, 3rd Edition, Lippincott-Raven, Philadelphia.

PETER FELGATE

Toxicology: Postmortem ***see*** Postmortem Toxicology: Interpretation

Toxicology: Quality Assurance ***see*** Quality Systems: Toxicology

Toxicology: Saliva ***see*** Oral Fluid Toxicology

Toxicology and Biochemical Postmortem Examinations ***see*** Postmortem Biochemical Examinations

Trace Evidence: Sampling ***see*** Sampling Trace Evidence

Trace Evidence: Transfer, Persistence, and Value

Introduction

A list of the types of material that might be classed as trace evidences would be endless as the material encountered could range from the classical items such as glass and fibers to more obscure examples such as pollen, cosmetics, and dust. While the range of material is broad, the mechanics of transfer and persistence is generally the same.

Transfer of material from one item (donor) to another (recipient) is generally associated with contact between the two items. It may also be associated with proximity between the two items, such as glass transfer close to a breaking window or paint transfer by splashing. Direct contact on its own may not be sufficient for transfer to occur. There must also usually be some sort of action or force.

Factors that affect the degree of transfer of material include

- the duration of contact;
- physical strength of the contact;
- the nature and condition of the donor material; and
- the nature and condition of the recipient material.

Persistence over time of the transferred material on the recipient item is vital if forensic science makes any contribution in a criminal case. Rarely, the suspect in a crime is apprehended at time zero (immediately). Even if this happens, still there would be a time lag between the apprehension of the suspect and the seizure of the relevant items for laboratory examination. Normally, a suspect is apprehended minutes, hours, or days after the crime. The amount of transferred material remaining i.e., persistence, depends on a number of factors including

- the amount of material transferred initially;
- the time between the initial transfer and the recovery of transferred material;
- the nature of the transferred material; and
- the activities of the recipient item subsequent to transfer.

Transfer and persistence studies have been important to forensic science for many decades beginning with the writings of Dr. Edmond Locard in France in 1929 [1–3]. Initially, the interest stemmed from a need to understand the mechanism by which evidence may be generated and our need to understand the science behind the concept of "every contact leaves a trace". It soon became apparent that the results of such studies feed our expectations in the approach to casework. The emphasis in forensic science on the interpretation of results in the context of the background circumstances of the case in question took on a different angle through the use of a Bayesian approach [4–6]. Transfer and persistence studies have been valuable sources of data, which the forensic scientist can use to estimate probabilities and likelihood ratios. Information gained from such studies is also valuable to determine whether or not it is appropriate to seize certain material from a crime scene or to conduct certain laboratory examinations.

The underlying principles of transfer and persistence outlined above are applicable across the full spectrum of cases in forensic science; however, in light of the fact that the nature of the material in question affects both transfer and persistence, it is appropriate to examine the various evidence types individually. There is more published data on the individual commonly encountered types of trace than on the overall concept.

Fibers

One of the most important studies on the transfer of fibers was undertaken by Pounds and Smalldon [7] Ụsing wool and acrylic as donor items, a number of recipient items were examined. Their general conclusions were that the number of fibers transferred increased with the area of contact and pressure of contact, but decreased with consecutive contact passes and considerably varied depending on the recipient garment. The mechanism for the transfer was considered by the same authors [8] (*see also* **Examination of Fibers and Textiles**). They suggested that there were a number of mechanical processes taking place including

- transfer of loose fibers that are already on the surface of the fabric;
- loosely bound fibers are easily pulled out by friction; and

- new fiber fragments are created by the contact and are transferred.

Other studies confirmed the conclusions of the earlier studies and examined the transfer of various fiber types to different fabric types and to different surfaces [9–11]. Where a mixture of fibers exists in the donor garment, preferential shedding is frequently encountered and the minor component may be recovered disproportionably to its amount in the donor garment [12, 13].

The persistence of fibers was also studied by Pounds and Smalldon [14]. Further, the fibers being studied were wool and acrylic and the authors found no significant difference in persistence between the two fibers. The most significant factor was the time of wear of the recipient garment after transfer of fibers. Fiber loss is initially rapid with approximately 80% of fibers being lost after only 4 h. However, they recovered transferred fibers up to 34 h after initial contact. Subsequent studies [15–19] confirmed their results and posed other reasons for loss of transferred fibers, including other garments being worn over recipient garment, contact of recipient garment with other garments/objects, and the surface features of the recipient garment. Shorter fibers (<2.5 mm) persisted longer than longer fibers but that finding was not supported by a further study, which found no difference in persistence between short and long fibers. In addition to the loss of transferred fibers from the recipient garment, transferred fibers may be redistributed on the recipient garment. This can cause problems if significance is attached to the specific location of the recovered fibers.

The persistence of fibers in head hair was studied [20, 21]. Wool and acrylic ski masks were used as the donors. Initial rapid loss of transferred fibers was again a feature of the decay pattern. Fibers were recovered from the hair up to 6 days, provided the hair was not washed. No significant difference was observed between wool and acrylic. Differences in hair type did affect the number of fibers transferred and their persistence, fibers persisted more readily in "short coarse" hair than in "long fine"/"straight" hair.

Hair

Irrespective of the significance that scientists place on the evidential value of hair, it is a commonly encountered evidence type and it is important to understand the factors affecting the transfer and persistence of hair. Hair is lost from the scalp naturally each day. Hair transfer from the donor would normally occur by direct contact with the donor's head or by contact with the donor's clothing, which has shed hairs on the surface. The persistence of hair on clothing was examined [22] and found to display the same decay pattern as fibers with only a small number of hairs surviving a day's wearing of the garment. Other studies [23, 24] showed that the rate of loss of hair from the fabric differed depending on the fabric. Hair persisted longer on garments made from wool than on nonwoollen garments. It is suggested that the scales on hair and wool facilitate the hair to "hold on" to the woolen fibers (*see also* **Hair: Microscopic Analysis**).

Glass

Glass is a material that is frequently encountered in forensic science laboratories in the course of normal casework. In contrast with fibers and hair transfer, which normally involve contact, glass transfer will not occur if fabric comes in contact with a nonbroken glass surface. Particles of glass are not lost from the surface of the glass. The normal transfer situation is where the person actually breaks the glass (kicks, uses elbow, etc.) or is standing nearby when the glass is broken. Standing nearby could refer to that person using an object to break the glass, throwing an object to break the glass, or being an innocent observer to any of the other activities, which lead to the glass being broken. Pounds and Smalldon [25] examined the distribution of glass fragments on the floor and on individuals in front of a breaking window. They showed that the number of particles (>1 mm) per unit area falls off extremely rapidly with distance, decreasing by half for each 20 cm increase in distance. The effect was less marked for smaller glass fragments. For individuals standing nearby, the results were in line with expectations, most fragments were found on the individual who stood closest to the window. Very few large fragments (>1 mm) were found on clothing even when there were significant numbers on the ground surrounding the person. This suggested that the larger particles are not retained on clothing even for relatively short periods. One of the problems associated with transfer experiments using glass is the fact that the number

of fragments produced and the distance travelled will be affected by how the window is broken and with what force. Several studies [26–28] employed a pendulum to standardise the impact. Glass fragments were projected as far as 3.3 m. Another factor, which affects the transfer mechanism, is the window type and size. In practice, the exact circumstances of how a given window is broken in casework cannot be known and it may be impossible to estimate the number of fragments likely to have been transferred. It is suggested that it is possible to estimate the distance between the offender and the window (*see also* **Glass**). This is based on three likely positions of the offender:

- breaking the window with body (shoulder/elbow), distance would be 0.0 m;
- breaking by striking with tool or foot, distance would be 0.6–0.9 m;
- breaking by throwing an object, distance would exceed 1.5 m.

Glass fragments may also be transferred to the hair or headwear of the person who is suspected of breaking the window [29].

The persistence of glass has also been studied. An early study [30] used an air rifle to project glass particles of known size onto two fabric types, wool/acrylic and cotton (denim). The persistence of transferred glass was determined at time intervals from 30 min to 6 h. The results show that the size of the particles influence the percentage retained on the fabric. Particles of the order of 1 mm are retained longest. Particles less than 0.5 mm were lost earlier and it was found that cotton retained them better than wool/acrylic. In one experiment, the persistence after 5 days was examined and a small number of particles were found even after this time gap. Further studies on glass persistence were done by breaking a window and observing how many fragments were retained after a certain period of time [31–33]. Initial rapid loss was further seen and then a slowed loss over a period of time. Even 8 h after the breakage, as many as seven fragments of glass were recovered from clothing.

Apart from time, the main factors influencing persistence of glass are post transfer activity, nature of recipient garment, and the condition of the garment. Damp or wet would appear to be more retentive than dry.

Paint

The most frequently encountered mechanism of paint transfer in the context of forensic science is, contact together with some degree of force. This can be caused by impact or abrasion and the amount of material transferred will depend on the degree of force applied. In traffic accident situations, it is not unusual to have multiplayer transfer from one vehicle to the other and vice versa. In some cases, the paint is loosely attached to the substrate and contact with little or no force may be required for transfer to occur. This is more frequently encountered in household paint transfer, which is common in burglary cases. In most cases, however, the paint is firmly attached to the substrate and force is required for transfer to take place (*see also* **Paint**).

In the case of wet paint, transfer can occur by means of simple contact. In some instances, paint will transfer through the air by splashing or dripping.

The persistence of paint flakes on clothes has received little attention in the literature, but some surveys were conducted on paint flakes found on clothing [34, 35]. It has been suggested [36] that the persistence of paint particles would resemble that of glass, but the number of paint fragments likely to transfer to clothing would be very small in comparison to glass. When wet paint is transferred and dried, then this will persist indefinitely until the item is washed and even then may not be removed depending on the nature of the paint.

Gunshot residue

Gunshot residue or firearm residue is produced when a firearm is discharged. The mechanism consists of a number of chemical reactions and decompositions produced at the second the firing pin strikes the primer of the cartridge. The metal vapors from the projectile mix with the decomposition products and the resultant vapor ejects from all possible openings in the firearm and condense/deposit on items nearby, including the hands and clothing of the person discharging the firearm (*see also* **Firearm Discharge Residue: Analysis of**).

Once deposited, the gunshot residue will be lost continuously, depending on the activity of the individual. Laboratory tests indicate that gunshot residue will remain on the hands only for a short time period

[37, 38] but in casework detectable amounts were detected several hours later. It is suggested that the probability of finding gunshot residue on hands more than 3 h after firing is very low.

Hydrocarbon Fire Accelerants

Hydrocarbon fire accelerants such as petrol can be important when crimes such as arson and petrol bombings are being investigated. Hydrocarbon fire accelerants are volatile liquids and will transfer to clothing and skin by direct contact with the liquid or the vapor. Unlike the other materials referred to earlier, hydrocarbon fire accelerants will naturally decay because of their volatile nature; therefore, persistence is less dependent on post transfer activity but is highly dependent on time and environmental conditions, especially temperature. Little or no study has been done on the persistence of hydrocarbon fire accelerants. One study examined the persistence of petrol on clothing and shoes after pouring petrol around a room [39]. Evaporation rates from shoes, clothing, wood, and carpet material were studied [40]. They found that petrol could still be detected on carpet even after 7 days. The transfer and persistence of petrol to car carpets was studied [41]. The study showed that petrol can transfer from the soles of shoes to carpet and can be detected for up to 24 h. In the same study, petrol was deposited directly onto carpet and depending on the amount deposited, petrol could be detected after 7 days (*see also* **Fire Debris: Laboratory Analysis of**).

Recent studies on persistence on hands suggest that petrol can be detected for longer periods than was traditionally considered. This may be associated with improved detection as well as improved recovery.

Other Materials

Other materials, which can be transferred, include pollen, soil, vegetation, dust, and other particles. Few of these have received attention. Pollen transfer and persistence was examined [42], and the distance-decay pattern previously observed for glass was further observed for pollen in a room containing cut flowers. The greatest number of grains was found in close proximity to the flowers. Once pollen is transferred to clothing, the persistence depends on the level of activity after transfer.

Metal particles on clothing in forensic casework can arise in crimes such as safe breaking or burglary. Metal particles will be deposited by the stream of particles produced by the action of the abrasive disc of the angle – grinder on the metal in question. One study [43] showed that after transfer, larger metal particles are lost more easily than smaller particles and that particles could be detected on the recipient garment even after 17 h.

Biological Traces

Body fluids are a powerful source of information to the forensic scientist because of DNA (*see also* **DNA**). Increasingly, the key questions with the detection of a DNA profile on an item are, how did the DNA reach there (transfer) and how long was the DNA there (persistence)? For saliva, the method of transfer may be direct, by licking or indirect, by spitting. Blood transfer mechanisms have become an important part of the work of the forensic biologist because of bloodstain pattern analysis. Generally, transfer will be by contact or by airborne transfer. Semen is normally transferred by sexual intercourse or external ejaculation onto clothing.

The persistence of saliva on skin and clothing has had limited attention [44, 45]. A recent study shows that saliva will be detected up to 4 days on the skin of living persons [46]. Saliva has been detected on clothes up to 28 days [47].

The persistence of semen on vaginal and other swabs has been studied [48] and semen has been detected on vaginal swabs up to 7 days after intercourse. It is believed that if semen is deposited on clothing it will survive indefinitely providing the conditions are not such as to encourage biodegradation. Washing of clothing will not necessarily remove the semen but will render the detection difficult.

Once blood is deposited on a surface, it will remain there indefinitely unless removed by cleaning or washing. If the conditions are damp or musty, then degradation can occur. The question of aging of blood is occasionally encountered, but there are difficulties in addressing the issue and generally it is not considered to be possible.

Problems with Transfer and Persistence

Transfer and persistence studies can be beneficial to forensic science in trying to understand how

something may have entered into an item originally and how long it is likely to remain on the item. There are dangers associated with drawing conclusions based on these studies. In most cases, we do not know the exact circumstances of the action, which lead to the transfer. How long exactly were the two people struggling? Was it a strong/very strong interaction? What was the exact surface area of the contact? How far away exactly was the person standing who threw the brick at the window? What exact force was used to throw the brick? What exactly has the suspect been doing since the crime that could have lead to loss of material from his/her clothing?

In the absence of answers to the above questions, we are in a difficult situation in estimating expected amounts of material persisting after a given time period as we do not know the amount transferred at time zero. The presence of CCTV footage and some eyewitness accounts can help inform our decision making in the area. It does not mean that we cannot make a reasonable estimate and especially if we have done some reconstruction experiments.

Another complicating factor is the phenomenon of secondary (or higher order) transfer. This arises where there is transfer from the primary donor to recipient number 1 and recipient number 1 subsequently transfers to recipient 2. The first recipient may be a person or some object such as a chair/car seat.

Secondary transfer of fibers has been studied [49, 50]. In general, the likelihood of secondary transfer of fibers is higher if the initial recipient is a chair/car seat rather than items of clothing. If the number of fibers initially transferred is high then this provides a large pool of fibers with the possibility of further transfer. When small numbers of fibers are recovered, the interpretation may be difficult and the scientist must examine the background circumstances to establish if primary or secondary transfer is in operation. Secondary transfer may also be beneficial to the investigation of crime. One study examined the secondary transfer of fibers from ski masks via head hair to pillowcases [50]. While the fibers from the ski mask would be lost from the head hair relatively quickly, a link between a suspect and the ski mask could be established indirectly via the fibers recovered from the pillowcase. Secondary transfer may also lead to a redistribution of fibers once transferred, and this phenomenon was recently examined using ski masks [51]. Fibers, which were seeded onto the inside, were recovered from the outside and *vice versa*.

Secondary transfer of hair has also been examined and this is a common occurrence, as the head hair is frequently shed to the persons' clothing as the initial transfer and subsequently transferred to recipient number 2. Some of the factors which favor secondary transfer of hair were also examined [22].

Secondary transfer of glass has also been studied [52, 53]. The studies examined scenarios where a person has broken a window and then travels in a car with another individual. Very little secondary transfer was observed. If a person carries an item, which has broken glass on the surface, then the conditions favor secondary transfer.

In most cases, secondary transfer of paint does not occur. A vehicle colliding with one vehicle and subsequently colliding with another vehicle could give rise to secondary transfer. A tool used to cause damage to a painted surface could transfer paint from that surface to the scene of the next burglary. This could in fact be used to link both scenes.

Secondary transfer can occur in relation to DNA evidence also. Transfer of DNA from one individual to another and subsequently to an object is possible under specific laboratory conditions [54, 55].

Secondary transfer is essentially the mechanism whereby contamination of items can take place. Whether it is the crime scene examiner or forensic scientist, anti-contamination measures must be established to ensure that they are not the intermediaries that facilitate unwanted transfers.

Conclusion

There are essentially two methods of transfer of trace material, by direct contact such as with paint and fibers or by a contactless transfer such as with breaking glass or gunshot residue vapor. Persistence studies are generally influenced by the same factors where time since transfer has a major effect on the amount of material recovered. Decay patterns over time for most materials will have a similar profile.

Transfer and persistence studies help to increase our knowledge on the basic mechanics involved in trace evidence exchanges. The information gained can help decide whether trace evidence is likely to

be present on a suspect's clothing after a certain time lapse. In addition the results can be used to inform our expectations on how much of a certain type of evidence we would expect to find based on the background circumstances of the case. These expectations can then be used in our case assessment and interpretation framework.

Persistence studies invariably depend on our ability to detect small amounts of trace evidence, and advances in searching and analytical processes will push out the boundaries of detection.

Much of the work on transfer and persistence was carried some time ago and there is scope for the area to be revisited specifically with respect to epithelial cells/touch DNA.

References

[1] Locard, E. (1928). Dust and its analysis, *Police Journal* **1**, 177.

[2] Locard, E. (1929). L'analyse des poussieres en criminalistique, *Revenue Internationale de Criminalistique* **1**, 176–249.

[3] Kirk, P.L. (1953). *Crime Investigation*, Interscience, New York.

[4] Evett, I.W. (1986). A Bayesian approach to the problem of interpreting glass evidence in forensic science casework, *Journal of Forensic Science Society* **26**, 3–18.

[5] Evett, I.W. & Buckleton, J.S. (1988). Some aspects of the Bayesian approach to evidence evaluation, *Journal of Forensic Science Society* **29**, 317–324.

[6] Cook, R., Evett, I.W., Jackson, G., Jones, P.J. & Lambert, J. (1998). A model for case assessment and interpretation, *Science and Justice* **38**(3), 151–156.

[7] Pounds, C.A. & Smalldon, K.W. (1975). The transfer of fibres between clothing materials during simulated contacts and their persistence during wear. Part 1-Fibre transfer, *Journal of Forensic Science Society* **15**, 17–27.

[8] Pounds, C.A. & Smalldon, K.W. (1975). The transfer of fibres between clothing materials during simulated contacts and their persistence during wear. Part 3-A preliminary investigation of the mechanism involved, *Journal of Forensic Science Society* **15**, 197–207.

[9] Kidd, C.B.M. & Robertson, J. (1982). The transfer of textile fibres during simulated contacts, *Journal of Forensic Science Society* **22**, 301–308.

[10] Lowrie, C.N. & Jackson, G. (1991). Recovery of transferred fibres, *Forensic Science International* **50**, 111–119.

[11] Grieve, M.C., Dunlop, J. & Haddock, P.S. (1989). Transfer experiments with acrylic fibres, *Forensic Science International* **40**, 267–277.

[12] Parbyk, A.E. & Lokan, R.J. (1985). A study of the numerical distribution of fibres transferred from blended fabrics, *Journal of Forensic Science Society* **26**, 61–68.

[13] Salter, M.T. & Cook, R. (1987). Differential shedding from blended fabrics, *Forensic Science International* **33**, 155–164.

[14] Pounds, C.A. & Smalldon, K.W. (1975). The transfer of fibres between clothing materials during simulated contacts and their persistence during wear. Part 2-Fibre persistence, *Journal of Forensic Science Society* **15**, 17–27.

[15] Robertson, J., Kidd, C.M.B. & Parkinson, H.M.P. (1982). The persistence of textile fibres transferred during simulated contacts, *Journal of Forensic Science Society* **22**, 353–360.

[16] Akulova, V., Vassiliauskiene, D. & Talaliene, D. (2002). Further insights into the persistence of transferred fibres on outdoor clothing, *Science and Justice* **42**, 165–171.

[17] Scott, H.G. (1985). The persistence of fibres transferred during contact of automobile carpets and clothing fabrics, *Canadian Society of Forensic Science Journal* **18**(4), 186–199.

[18] Grieve, M.C. (1990). Fibres and their examination in forensic science, *Forensic Science Progress* **4**, 40–55.

[19] Siegal, J.A. (1997). Evidential value of textile fibre-transfer and persistence of fibres, *Forensic Science Review* **9**(2), 82–96.

[20] Ashcroft, C.M., Evans, S. & Tebbett, I.R. (1988). The persistence of fibres in head hair, *Journal of Forensic Science Society* **28**, 289–293.

[21] Salter, M.T. & Cook, R. (1996). Transfer of fibres to head hair, their persistence and retrieval, *Forensic Science International* **81**, 211–221.

[22] Dachs, J., McNaught, I.J. & Robertson, J. (2003). The persistence of human scalp hair on clothing fabrics, *Forensic Science International* **138**, 27–36.

[23] Robertson, J., Harding, H. & Somerset, H. (1987). The persistence of hairs on clothing, *Canadian Society of Forensic Science Journal* **20**, 240–246.

[24] Gaudette, B.D. & Tessarolo, A.A. (1987). Secondary transfer of human scalp hair, *Journal of Forensic Sciences* **32**(5), 1241–1253.

[25] Pounds, C.A. & Smalldon, K.W. (1978). The distribution of glass fragments in front of broken window and the transfer of fragments to individuals standing nearby, *Journal of Forensic Science Society* **18**, 197–203.

[26] Locke, J. & Unikowski, J.A. (1991). Breaking of flat glass part 1: size and distribution from plain glass windows, *Forensic Science International* **51**, 251–262.

[27] Locke, J. & Unikowski, J.A. (1992). Breaking of flat glass part 2: effect of pane parameters on particle distribution, *Forensic Science International* **56**, 95–102.

[28] Locke, J. & Scranage, J.K. (1992). Breaking of flat glass part 3: surface particles from windows, *Forensic Science International* **57**, 73–80.

[29] Curran, J.M., Hicks, T.N. & Buckleton, J.S. (eds) (2000). *Forensic Interpretation of Glass Evidence*, CRC Press.

[30] Brewster, F., Thorpe, J.W., Gettingby, G. & Caddy, B. (1985). The retention of glass particles on woven fabrics, *Journal of Forensic Sciences* **30**(3), 798–805.
[31] Hicks, T., Vanina, R. & Margot, P. (1996). Transfer and persistence of glass fragments on garments, *Science and Justice* **36**, 101–107.
[32] Allen, T.J. & Scranage, J.K. (1998). The transfer of glass, part 1: transfer of glass to individuals at different distances, *Forensic Science International* **93**, 167–174.
[33] Allen, T.J., Hoefler, K. & Rose, S.J. (1998). The transfer of glass, part 2: a study of the transfer of glass to a person by various methods, *Forensic Science International* **93**, 175–193.
[34] Pearson, E.F., May, R.W. & Dabbs, M.G.D. (1971). Glass and paint fragments found in men's outer clothing – a report of a survey, *Journal of Forensic Sciences* **16**, 283–302.
[35] Lau, L., Beveridge, A.D., Callowhill, B.C., Conners, N., foster, K., Groves, R.J., Ohasi, K.N., Sumner, A.M. & Wong, H. (1997). The frequency of occurrence of paint and glass on the clothing of high school students, *Canadian Society of Forensic Science Journal* **30**, 233–240.
[36] Willis, S.M., McCullough, J. & McDermott, S.D. (2001). The interpretation of paint evidence, in *Forensic Examination of Glass and Paint. Analysis and Interpretation*, Taylor and Francis, New York, pp. 273–287.
[37] Kilty, J.W. (1975). Activity after shooting and its effect on the retention of primer residues, *Journal of Forensic Sciences* **20**, 219–230.
[38] Jalanti, T., Henchoz, P., Galluser, A. & Bonfanti, M.S. (1999). The persistence of gunshot residue on shooter's hands, *Science and Justice* **39**, 48–52.
[39] Coulson, S.A. & Morgan-Smith, R.K. (2000). The transfer of petrol on to clothing and shoes while pouring petrol around a room, *Forensic Science International* **122**, 135–141.
[40] Folkman, T.E., Kuehl, A.M., Groves, R.J. & Beveridge, A.D. (1990). Evaporation rates of gasoline from shoes, clothing, wood and carpet materials and kerosene from shoes and clothing, *Canadian Society of Forensic Science Journal* **23**, 49–59.
[41] Cavanagh-Steer, K., Du Pasquier, E., Roux, C. & Lennard, C. (2005). The transfer and persistence of petrol on car carpets, *Forensic Science International* **147**, 71–79.
[42] Morgan, R.M., Allen, E., Lightowler, Z.L., Freudiger-Bonzon, J. & Bull, P.A. (2008). A forensic geoscience framework and practice, *Policing* **2**, 185–195.
[43] McDermott, S.D. (1994). Metal particles as evidence in criminal cases, *Journal of Forensic Sciences* **39**(6), 1552–1559.
[44] Sweet, D., Lorente, J.A., Valenzuela, A., Lorente, M. & Villanueva, E. (1997). PCR-based DNA typing of saliva stains recovered from human skin, *Journal of Forensic Sciences* **42**(3), 447–451.
[45] Rees, B. & Baxter, S.J. (1975). The identification of saliva in stains in forensic casework, *Medicine Science and the Law* **15**(1), 37–41.
[46] Kenna, J., Smyth, M., McKenna, L., Dockery, C. & McDermott, S.D. The recovery and persistence of salivary DNA on human skin, *Journal of Forensic Sciences* (personal communication).
[47] Tsutsumi, H., Higashide, K., Mizuno, Y., Tamaki, K. & Katsumata, Y. (1991;). Identification of saliva stains by determination of the specific action of amylase, *Forensic Science International* **50**(1), 37–42.
[48] Willott, G.M. & Allard, J.E. (1982). Spermatozoa – their persistence after sexual intercourse, *Forensic Science International* **19**, 135–154.
[49] Grieve, M.C. & Biermann, T.W. (1997). Wool fibres-transfer to vinyl and leather vehicle seats and some observations on their secondary transfer, *Science and Justice* **37**(1), 31–38.
[50] Palmer, R. & Banks, M. (2005). The secondary transfer of fibres from head hair, *Science and Justice* **45**, 123–128.
[51] Chewning, D.D., Deaver, K.L. & Christensen, A.M. (2008). Persistence of fibres on ski masks during transit and processing, *Forensic Science Communications* **10**(3).
[52] Allen, T.J., Hoefler, K. & Rose, S.J. (1998). The transfer of glass, part 3: the transfer of glass from a contaminated person to another uncontaminated person during a ride in a car, *Forensic Science International* **93**, 195–200.
[53] Allen, T.J., Hoefler, K. & Rose, S.J. (1998). The transfer of glass, part 4: the transfer of glass fragments from the surface of an item to a person carrying it, *Forensic Science International* **93**, 201–208.
[54] Lowe, A., Murray, C., Whitaker, J., Tully, G. & Gill, P. (2002). The propensity of individuals to deposit DNA and secondary transfer of low level DNA from individuals to inert surfaces, *Forensic Science International* **129**, 25–34.
[55] Phipps, M. & Petricevic, S. (2007). The tendency of individuals to transfer DNA to handled items, *Forensic Science International* **168**, 162–168.

SEAN D. MCDERMOTT

Trace Metal Detection *see* Firearms: Identification of Handling of Firearms/Trace Metal Detection

Traffic Fatalities

Introduction

Globally, the annual number of traffic deaths has risen to more than 1 million after 2000. The rate of traffic fatality during the last decades has decreased in high-income countries but increased in low-income countries. For all countries, the annual cost of road injuries in terms of medical care, disability, and property damage has been estimated to exceed USD 50 billion [1].

To enable the application of the best available preventive measures, better knowledge of the causes of fatal traffic crashes is desired. In the following, we present some data regarding different types of traffic deaths, followed by a section regarding different preventive measures.

Intentional Deaths

Although traffic homicides probably represent an extremely small fraction of all traffic deaths, suicides are estimated to be constitute a few percent of all motor vehicle crash fatalities [2–4]. For obvious reasons, suicides in traffic may, however, be underestimated [5]. Suicide by motor vehicle crash typically involves a single vehicle, single occupant crashes, or a head-on collision of a single occupant vehicle with a heavy goods truck. Many of the victims have attempted suicide on at least one previous occasion and about half of the traffic suicide victims have a mental insufficiency. Inebriation of alcohol is also frequently described in these crashes. For obvious reasons, these characteristics are typical of the cases we recognize as suicides, and the characteristics described may thus be biased. One must also consider the possibility of "hidden suicides" when the victim makes the suicide look like a unintentional traffic crash, e.g., for insurance benefits. It is, however, impossible to have a reasonably safe opinion regarding what the victim actually intended during the seconds before crash, and consequently the circumstances and the previous history of the victim are crucial factors in the classification process. But, even in a "typical" case of head-on collision of a single occupant vehicle with a heavy goods truck, there is still the possibility that the driver fell asleep and did not intend the collision, and the case will be even more difficult to classify if the victim was inebriated. In some cases, the autopsy findings may be indicative in themselves, e.g., when other signs of self-destructive behavior are revealed.

It should be noted that "traffic suicides" are not necessarily restricted to those killed in vehicle crashes, but include also the people who use the traffic environment to commit the suicide, and that also such suicides are subject to preventive measures. For example, suicide by jumping from a bridge represents a method where the potential of prevention is great [6].

Unintentional/Accidental Deaths

These cases form the vast majority of road traffic deaths, and represent probably more than 80% of all traffic deaths. The remaining deaths are distributed as follows: 2–3% suicides, ~2% undetermined manner of death, and some 10% natural deaths at the wheel [4]. Although some preventive strategies such as deformation zones may apply to all types of traffic incidents (cf. [7]), railway, watercraft, and aircraft fatalities will not be further discussed here.

The majority of accidental traffic deaths occur in road traffic, involving passenger cars and/or trucks, and where other traffic elements are more seldomly involved. In rural areas, however, off-road traffic fatalities form a significant percent of all traffic deaths. For example, in Northern Sweden, snowmobile-related deaths form the third most common group of all traffic-related deaths, and the number of deaths involving ATVs (all-terrain vehicles) is increasing (cf. [8]). Both these types of off-road traffic deaths involve mostly younger males, and are significantly related to high speeds and drunken driving [9, 10].

Natural Deaths

Deaths in the traffic environment due to natural causes are estimated to represent a minority of all traffic deaths. Several factors, among others a low autopsy rate and other biases in the statistics might, however, underestimate these types of events. In addition, in severe traffic trauma, natural death might go undetected since the injuries are expected to cause the fatality or, occasionally, since the body is severely mutilated and not allowing the pathologist

to correctly examine crucial inner organs such as the heart. It is clear though, that many natural deaths of passenger car drivers are associated with no or only minor injury and property damage. This indicates that the driver experiences some warning symptoms and consequently reduces the speed or even brings the vehicle to a stop before he becomes unconscious [11]. In the medico-legal investigation, the main findings in these cases are absence of severe injuries and presence of more or less severe pathological findings of disease, most often some form of ischemic heart disease [11].

A closely-related issue is to what extent an (acute) disease can be considered to have "caused" an accident where traumatic injuries have caused the death. In a study of 480 traumatic traffic deaths, disease processes was judged to "probably" have caused the crash in 6% of all deaths, of which in 1.3% the causal relationship was considered "very probable" [12].

Undetermined Manner of Death

In a few percent of traffic deaths, there are some – but not enough – indications of a suicide, and in these cases the forensic pathologist classifies the manner of death as "undetermined". The same is the case in even fewer cases, where there are signs – but not enough – of a natural death, or a homicide.

Prevention – Precollision Phase

During the last decades, vast improvements have been made regarding the traffic environment, including improvements and/or introduction of traffic lights, signs, flyovers, highways, speed reduction, game fences, etc. The vehicle itself is of course also under constant development, and reformed driver's education, information campaigns toward high risk groups, differentiated insurance rates, police surveillance and so forth, has also aided in reducing the risk and effects of crashes. More recently, the importance of separating passenger cars from heavy vehicles has been emphasized [13].

Disease. Preventing people with different diseases to drive a car is a very controversial issue. Also from a medical standpoint, it is in many cases hard to argue in favor of restrictions, since we do not have the instruments to identify the individuals 'at risk'. Of course, in individuals with loss of hearing, loss of eye sight, advanced dementia, etc., this decision is easily made. But it must at present be considered impossible to identify for example, which individuals have a coronary heart disease of such an advanced stage that continued driving must be prohibited because of the danger posed to the driver himself or others. To do this with presently available methods, large sums must be spent on advanced medical examinations, and a large population of drivers must be stopped from driving. In an autopsy study of 480 consecutive traffic deaths, 85% of the drivers over 70 years of age had medical factors that could lead to sudden incapacitation of driving ability [12], but only a small proportion of the victims had sought medical care for the disease.

Drunken Driving. Alcohol is a well-known risk factor in traffic accidents, and as the first country, Sweden passed in 1941 a law prohibiting drunken driving. Mainly, ethanol has a hypnotic and anesthetic effect on the central nervous system, and the seemingly stimulating effect is the result of a primary depression of higher centers in the brain, resulting in an unchecked behavior. Already in low concentration, ethanol has a negative effect upon judgment and motor functions. It has been estimated that at a blood alcohol concentration of $1.5\,g\,l^{-1}$, the risk of a car driver to be involved in a crash causing personal injury has been estimated to be about 300–600 times higher than when driving sober (cf. [14]).

Owing to the measures taken to reduce the number of alcohol-related deaths and injuries in traffic, the percent of alcohol-related crashes has decreased in many high-income countries. The percent of alcohol-related crashes resulting in a fatality is, however, still very high – in the United States, the percent is around 40% [15]. In Sweden, a country with a very restrictive view on drunken driving – both legally and according to the public opinion – the percent is still high, around 20% of traffic deaths, and exceeding 50% in fatal single vehicle crashes (but around 10% in multi vehicle crashes) [16]. Obviously, it seems very difficult to further reduce the percent of alcohol-related incidents. Although inebriation of vehicle drivers is most dangerous, the negative effect of alcohol is also obvious in cyclists, pedestrians, etc. [17, 18]. Another aspect is that fighting against drunken driving might displace this problem to other categories of road users [18].

One explanation for this is most certainly that a considerable proportion, may be 60% or more, of the drunken drivers are chronic alcoholics, who do not care about the legal implications of drunken driving. Autopsy findings in support of chronic alcoholism are, mainly, a high or very high ethanol concentration (on average often around $1.7\,g\,l^{-1}$ in fatally injured drivers) and the finding of liver steatosis without any other medical explanation.

As a consequence of the findings presented above, effective treatment of alcohol abuse would certainly have a beneficial effect. Another measure would be the introduction of a so called alcolock, a device which allows the driver to start the engine only if his breath is free from ethanol. In recent years, alcolock devices have become mandatory in commercial vehicles in some countries.

Drugged Driving. Whereas alcohol has been known for a very long time period to cause or contribute to traffic crashes, other drugs have attracted interest only in more recent times. Legislation regarding driving under the influence of drugs is, however, tricky, since many people are taking prescription drugs which may have an effect on the ability to drive. The drugs used may differ between different countries, but benzodiazepines and central stimulants are two of the most common according to the literature [19].

Young and Old Drivers. The risk of fatal or severe injury is six times greater among 18- to 19-year-old drivers than among 45- to 54-year olds, mostly because of higher speed and lesser driving experience. Also elderly persons are subjected to higher risks of injury and death; the risk is twice as high in the ages 65–74, and six times higher in the ages 75–84 [20, 21]. With an aging population, and successively older vehicle drivers, age-related health problems are expected to play an increasing role in the prevention strategies, e.g., regarding dementia (cf. [22]).

Prevention – Collision Phase

This is the true phase of injury, when the kinetic energy ($mv^2/2$) transforms to heat and energy-consuming deformation of the vehicle and the human body. To reduce the deleterious effects of this energy transformation, which are generally associated with rapid deceleration, it is important to reduce the pressure (i.e., the amount of force per unit area) when the body hits the interior of the vehicle, or reduce the rate of deceleration. This can be achieved by decreasing the speed of deformation by distributing it over a larger area and over a longer time period. Airbags provide one example where the design is such as to provide a large surface area which spreads the force applied to the body over a wider area.

Similar principles may be applied to the vehicle itself by constructing deformation zones. Irrespective of the protection measures taken, forces and consequent risk of injury is proportional to the deceleration.

Safety belts increase deceleration time in the collision phase, distribute the deceleration forces to the pelvis and over the chest, decrease or derate the effect of head injury, and decrease the risk of ejection from the car. The safety belt has a high protective effect and has been reported to decrease the risk of lethal injury by 40% and to reduce personal injuries by 50%.

Air bags represent a passive safety system which at a rapid deceleration will inflate and form a protective balloon between the person and the interior of the car. Air bags were primarily constructed to reduce the risk of serious head injury in frontal collisions, but also have a protective effect regarding neck injuries, and have been reported to reduce both these types of injury by approximately 40%. Air bags should always be used in combination with safety belts.

A variety of *protective devices for children* have been introduced during the last decades. These include a number of different child seats, designed for different ages (body sizes), where the forces are distributed over a larger body area, and where the intergroup movements of torso and head are minimized. Persons with a body height of less than 140 cm are not recommended to sit in front of an air bag since this may increase the risk of head and neck injury.

Rear end collision may cause disabling neck injuries, so called whiplash injuries, and *head-rests* have been shown to significantly reduce this type of injury.

It has been calculated that the use of *helmets* may reduce fatal injuries among bicyclists by 40%, and severe injuries by approximately 70% [23], and similar figures are also applicable to motorcycles. Helmets have, however, probably not a similar protective effect in snowmobile crashes [24].

Prevention – Postcollision Phase

After the crash, effective medical care and rehabilitation are most important to reduce the long-term effects of traffic injuries. Alarm and ambulance systems and highly specialized trauma care have gone through a rapid development in high-income countries, including the use of helicopter transports, especially in remote areas [25] to minimize delay. The potential to save lives and to reduce long-term effects of injury through optimized medical care is significant [26, 27].

The Postmortem Examination in Traffic Deaths

In many countries, fatally injured traffic victims are subjected to a medico-legal autopsy, following a police investigation of the crash. The reasons for such an approach is in part to investigate whether further legal actions are motivated from the law enforcement, but also to give a basis for in depth investigations, as already for a long time have been applied in aviation crashes. In this way, data from the traffic environment, the vehicles involved, and the injured victims are gathered and processed by a multi-disciplinary team. To optimize the quality of data from the victims, medicolegal autopsies are routinely performed in many countries. In the postmortem examination, the determination of the cause and manner of death is of great importance, but additional questions to the traffic investigation might also be answered, for example the use of seat belts, position in the vehicle, and time of survival. The medico–legal investigation includes a careful external and internal examination of the body including histopathological investigations in order to document injuries and signs of disease. Toxicological analyses including alcohol and drugs are also of paramount importance for the understanding of the fatal crash.

The following are the main tasks of the forensic pathologist in the investigation of traffic deaths:

- To reveal the cause and manner of death – for the cause-of-death statistics, for the next-of-kin, and for accident preventive work.
- To assist in any reconstruction of the event after analysis of the case, including autopsy findings and the results of toxicological analyses.
- To diagnose any disease – including also previously unknown disease – which may have initiated the crash, or contributed to death.
- To reveal if any medical treatment before (and after) the crash has been effective and relevant, since imperfections in these respects may have caused the crash and/or have influenced upon the chances of survival.
- To act as a "whistle blower" in cases where extraordinary findings are made, e.g., new drugs, changes in drug panorama, and new injury patterns (cf. [28]).

The final aim is of course to find and introduce preventive measures, in order to reduce the traffic fatalities and injuries.

References

[1] WHO (2004). *World Report on Road Traffic Injury Prevention*, WHO, http://www.who.int/violence_injury_prevention/publications/road_traffic/world_report/en/index.html.

[2] Imajo, T. (1983). Suicide by motor vehicle, *Journal of Forensic Sciences* **28**, 83–89.

[3] Hernetkoski, K. & Keskinen, E. (1998). Self-destruction in Finnish motor traffic accident in 1974–1992, *Accident Analysis and Prevention* **30**, 697–704.

[4] Ahlm, K., Eriksson, A., Lekander, T. & Björnstig, U. (2001). Alla dödsfall i trafiken är inte "dödsolyckor"–en analys av officiell statistik över dödsfall i svensk vägtrafik år 1999. [All traffic related deaths are not "accidents"–an analysis of Swedish official statistics], *Läkartidningen* **98**, 2016–2022.

[5] Öhberg, A., Penttilä, A. & Lönnqvist, J. (1997). Driver suicides, *British Journal of Psychiatry* **171**, 468–472.

[6] Lindqvist, P., Jonsson, A., Eriksson, A., Hedelin, A. & Björnstig, U. (2004). Are suicides by jumping off bridges preventable? An analysis of 50 cases from Sweden, *Accident, Analysis and Prevention* **36**, 691–694.

[7] Eriksson, A., Ericsson, D., Lundström, N.-G. & Thorson, J. (1984). Mechanisms of derailment injuries to railway carriage occupants and engine drivers, *Travel Medicine International* **2**, 85–89.

[8] Hansson, S., Bylund, P.-O., Ahlm, K. & Eriksson, A. (2006). All-terrain vehicle fatalities in Sweden, 1992–2004, *Scandinavian Journal of Forensic Science* **2**, 58–61.

[9] Eriksson, A. & Björnstig, U. (1982). Fatal snowmobile accidents in northern Sweden, *Journal of Trauma* **22**, 977–982.

[10] Öström, M. & Eriksson, A. (2002). Snowmobile fatalities. Aspects on preventive measures from a 25-year review, *Accident Analysis and Prevention* **34**, 563–568.

[11] Öström, M. & Eriksson, A. (1987). Natural death while driving, *Journal of Forensic Sciences* **32**, 988–998.
[12] Sjögren, H., Eriksson, A. & Öström, M. (1996). Role of disease in initiating the crashes of fatally injured drivers, *Accident Analysis and Prevention* **28**, 307–314.
[13] Björnstig, U., Björnstig, J. & Eriksson, A. (2008). Passenger car collision fatalities – with special emphasis on collisions with heavy vehicles, *Accident, Analysis and Prevention.* **40**, 158–166.
[14] Timby, N., Sjögren, H., Björnstig, U. & Eriksson, A. (1998). Crash responsibility versus drug and alcohol use among fatally injured and hospitalized motor vehicle drivers in Sweden, *Alcoholism: Clinical Experimental Research* **22**, 1838–1841.
[15] NHTSA (2007). *Traffic Safety Facts. 2006 Traffic Safety Annual Assessment. Alcohol Related Fatalities*, NHTSA's National Center for Statistics and Analysis, Washington, DC.
[16] Öström, M. & Eriksson, A. (1993). Single-vehicle crashes and alcohol: a retrospective study of passenger car fatalities in northern Sweden, *Accident Analysis and Prevention* **25**, 171–176.
[17] Öström, M., Björnstig, U., Näslund, K. & Eriksson, A. (1993). Pedal cycling fatalities in Northern Sweden, *International Journal of Epidemiology* **22**, 483–488.
[18] Öström, M. & Eriksson, A. (2001). Pedestrian fatalities and alcohol, *Accident, Analysis and Prevention* **33**, 173–180.
[19] Sjögren, H., Björnstig, U., Eriksson, A., Öhman, U. & Solarz, A. (1997). Drug and alcohol use among injured motor vehicle drivers in Sweden: prevalence, driver, crash and injury characteristics, *Alcoholism: Clinical and Experimental Research* **21**, 968–973.
[20] Sjögren, H., Björnstig, U., Eriksson, A., Sonntag-Öström, E. & Öström, M. (1993). Elderly in the traffic environment: analysis of fatal crashes in Northern Sweden, *Accident Analysis and Prevention* **25**, 177–188.
[21] Sjögren, H., Björnstig, U., Eriksson, A. & Öström, M. (1996). Differences between older and younger drivers; characteristics of fatal car crashes and driver injuries, *Safety Science* **23**, 63–77.
[22] Viitanen, M., Johansson, K., Bogdanovic, N., Berkowicz, A., Druid, H., Eriksson, A., Krantz, P., Laaksonen, H., Sandler, H., Saukko, P., Thiblin, I., Winblad, B. & Kalimo, H. (1998). Alzheimer changes are common in aged drivers killed in single car crashes and at intersections, *Forensic Science International* **96**, 115–127.
[23] Björnstig, U., Öström, M., Eriksson, A. & Sonntag-Öström, E. (1992). Head and face injuries in bicyclists – with special reference to possible effects of helmet use, *Journal of Trauma* **33**, 887–893.
[24] Björnstig, U., Öström, M. & Eriksson, A. (1994). Would a helmet law for snowmobile riders reduce head injuries? *Arctic Medical Research* **53**, 196–199.
[25] Vesterbacka, J. & Eriksson, A. (2001). A rural ambulance helicopter system in Northern Sweden, *Air Medical Journal* **20**, 28–31.
[26] Boman, H., Björnstig, U., Hedelin, A. & Eriksson, A. (1999). "Avoidable" deaths in two areas of Sweden – analysis of deaths in hospital after injury, *European Journal of Surgery* **165**, 828–833.
[27] Henriksson, E.M., Öström, M. & Eriksson, A. (2001). Preventability of vehicle-related fatalities, *Accident, Analysis and Prevention* **33**, 467–475.
[28] Björnstig, U., Eriksson, A., Thorson, J. & Bylund, P.-O. (1986). Collisions with passenger cars and moose, Sweden, *American Journal of Public Health* **76**, 460–462.

ANDERS F. ERIKSSON AND MATS G. ÖSTRÖM

Training and Certification (in Criminalistics)

Introduction

Criminalistics is composed of many different sciences and covers a large spectrum of disciplines. As a result, training[a] in criminalistics must be strongly based on natural sciences (physics, chemistry, and biology) as well as emphasize topics specific to criminalistics and the criminal justice field. To obtain his/her basic training, the future criminalist can choose to attend one of the rare university programs offering a comprehensive coverage of forensic sciences, to attend a university offering a natural science degree with an emphasis in forensic sciences, or to pursue a regular natural science degree [1, 2]. Well-known university degrees in forensic sciences (or with a forensic emphasis) include the ones offered by the School of Criminal Sciences of the University of Lausanne in Switzerland, the Centre for Forensic Science of the University of Strathclyde in Scotland, and the Centre for Forensic Science of the University of Technology Sydney in Australia. When attending universities offering a natural science degree with an emphasis on forensic sciences, one must be careful in choosing a program. Owing to a recent global interest generated by the production of idealized TV shows, a plethora of university programs have been developed with the claim of forming forensic scientists. Extreme caution is required with some programs that offer no more than a curriculum that merely includes a few criminal justice and forensic classes taught by professors who have little or no experience in forensic sciences.

Another problem is the existence of "broad brush" courses taught by people with previous forensic experience but who rely solely on their experiences and do not include a robust underpinning science component. As a result, students graduating from such programs may falsely expect to become forensic scientists; however, they have acquired little if no notion of the basic principles of forensic sciences and are no more prepared to work in the forensic science field than a regular chemist, biologist, or physicist. Overall, a balance must be struck.

In the United States, where this problem is prominent, the Technical Working Group on Education and Training in Forensic Science (TWGED) issued a document defining the minimum requirements for the core curriculum of a forensic sciences degree [3]. Another very interesting and more informal approach was developed by the National Institute of Forensic Science (NIFS) in Australia: a list of questions that a prospective student, or parents, can ask about a university program was defined, along with desirable answers from the forensic science industry perspective [4]. This allows the student to carry out a relatively good assessment of the seriousness of the university program and to evaluate its content vis-à-vis the intended career.

Table 1 shows the requirements set forth by TWGED for some specific forensic scientists [3]. Generally, it is often recognized that a minimum of a bachelor's degree in a natural science, criminalistics, or a closely related field is necessary to work as a forensic scientist [5, 6]. However, academic education is not a criterion sufficient to start working in a crime laboratory on real cases, although it is a solid and necessary starting point.

Figure 1 shows the path followed by a forensic scientist during his/her career. After academic education, the forensic scientist must undergo a period of training, which will lead to his/her full operability at the crime laboratory and at crime scenes. After this induction training, a continuing education program is also devised to ensure that the criminalist is up-to-date with evolving science and that the most appropriate and valid science is applied in casework.

Induction Training Programs

A newly hired, inexperienced criminalist cannot immediately work on real cases, but spends a certain amount of time getting field-specific education by learning from his/her colleagues and working on mock cases. Before they perform casework, all newly hired forensic scientists will have to follow an initial training program, also known as an *induction training program*, which can vary significantly in length based on the person's previous education and on the operational institution's resources, sometimes up to two years. This initial training can be defined as [3] "the formal, structured process through which a forensic scientist reaches a level of scientific knowledge and expertise required to conduct specific forensic analyses." If the university degree conferred did not include a major forensic component, the training will likely be extensive and must include the basic principles of forensic sciences. By contrast, the student who followed a valid forensic program will typically require less training before he/she can work on real cases.

The program includes on-the-job training as well as specialized education. Even when the candidate has a comprehensive forensic degree, a certain level of specialization may be required when working in some particular domains. This training may be obtained through designated courses or university programs, or from the laboratory where he/she works in the form of on-the-job training. Responsibility is gradually increased as the criminalist starts working on real cases under the very close supervision of more experienced colleagues or supervisors. After the induction training period, which may conclude with successfully completing an examination, the forensic scientist may undertake his/her own casework independently.

Continuing Training

The training of a forensic scientist does not stop at the end of the initial training period. Forensic scientists face a constantly evolving criminal element combined with ongoing technological advances and, as a result, have to develop new techniques and methods. Keeping current with modern criminalistics is a necessary process referred to as *continuing training*, and can be defined as [3] "the mechanism through which a forensic scientist remains current or advances to a higher level of expertise, specialization, or responsibility". This can be achieved through continuing education as well as practical experience.

Table 1 Sample curriculum for undergraduate forensic sciences degree proposed by TWGED [3]. These examples are based on a minimum of 120 semester hours to obtain a degree. Credit hours as described above are meant to indicate semester credit hours[(a)]

	Firearms Impression evidence Questioned documents Prints	Chemistry evidence Controlled substances	Biology
University general education (36–40 h)	Courses required by the university, which may include language, humanities, social sciences, technical writing, mathematics, computer science, and public speaking. Credit hours required will vary from one university to another. Some forensic degree coursework may count toward fulfilling these requirements		
Natural science core (34–38 h)	Biology I	Biology I	Biology I, II
	Calculus	Calculus	Calculus
	General chemistry I, II	General chemistry I, II	General chemistry I, II
	Organic chemistry I, II	Organic chemistry I, II	Organic chemistry I, II
	Physics I, II	Physics I, II	Physics I, II
	Statistics	Statistics	Statistics
Specialized core (12 h)	Inorganic chemistry	Analytic chemistry quantitative	Biochemistry
	Instrumental analysis	Inorganic chemistry	Genetics
	Optics/lasers	Instrumental analysis	Instrumental analysis
	Physical chemistry	Physical chemistry	Molecular biology
Forensic science core (6 h)		Forensic science survey	
		Forensic professional practice	
Forensic laboratory science (9 h)	Internship	Forensic chemistry internship	Forensic biology internship
	Microscopy	Microscopy	Microscopy
	Physical methods	Physical methods	Physical methods
Additional courses (19 h)	Crime scene	Advanced instrumental analysis	Cell biology
	Image analysis	Drugs	Introduction to criminal justice
	Introduction to criminal justice	Introduction to criminal justice	Legal evidence
	Legal evidence	Legal evidence	Microbiology
	Materials science	Analytical toxicology	Population genetics
		Materials science	Immunology
		Pharmacology	Public speaking
		Public speaking	

[(a)] Reproduced from Ref. 4. US Department of Justice, 2004

Continuing education may be obtained through two main categories: scientific literature (autodidactic manner) and attendance at courses and conferences.

Scientific literature: The forensic scientist should regularly peruse scientific periodicals, such as *Forensic Science International, Science & Justice, Journal of Forensic Sciences*, the *Canadian Society for Forensic Science Journal*, and the *AFTE Journal*. Articles relevant to his/her domain of competency should be carefully and thoroughly read and understood. The criminalist should also be acquainted with additional scientific journals more focused on some specific topics of a natural science or other relevant disciplines that publish articles pertinent to forensic sciences on a regular basis. Most crime laboratories have subscriptions to some of the most important journals addressing forensic sciences topics.

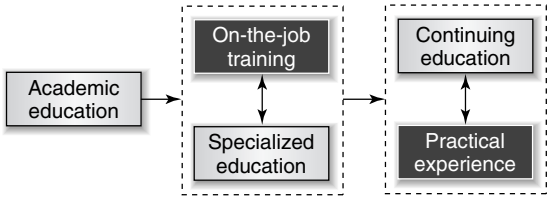


Figure 1 The career path of a forensic scientist

Courses and conferences: Attending conferences is also extremely important. The scientific value of the presentations is a serious asset to continuing education, as is the possibility of sharing information with other practicing colleagues. Many professional organizations conduct conferences on a regular basis, such as the triennial meeting of the European Academy of Forensic Sciences, the annual meeting of the American Academy of Forensic Sciences, the biennial symposium of the Australian & New Zealand Forensic Science Society, and the triennial meeting of the International Association of Forensic Sciences.
Different associations around the world also offer topic-specific courses. These classes usually provide excellent continuing education. The basic and advanced fire debris analysis courses organized by the Bureau of Alcohol, Tobacco, Firearms and Explosives with the National Center for Forensic Sciences (USA) are some examples.

Furthermore, the experience gained from handling practical casework is also a valuable addition to the constant evolution of the criminalist's skills. This is particularly useful when combined with other activities, such as scientific research and teaching. Although teaching is a valuable asset to continuing education, criminalists also must keep learning from others and reevaluating themselves from time to time.

Certification

The goal of certification for personnel is to provide a formal and objective guarantee that a candidate has a minimal level of knowledge, skills, and abilities (KSA) in a given discipline. The KSA constitute the (defined) competency of a person. This competency is formally recognized through the delivery of a certificate and, in many instances, becomes visible with the addition of initials after the person's name. The process of certification should be voluntary and must include an objective peer review of one's education and experience as well as the successful completion of a series of examinations.

Criminalistics encompasses a wide variety of different areas, and thus personnel must meet a large number of skills and competencies. This renders the certification process quite complex and, as a result, it may be quite difficult for a small organization to have its personnel certified (and/or to maintain certifications) in all different areas.

Program Development

In order to set up a certification program, the following three conditions must be met [7]:

1. The KSA of the criminalist must be clearly defined.
2. An efficient manner of evaluating the criminalist's competency must be developed.
3. The certification must be recognized by professionals in the field (both providers and clients) as proof that the criminalist reached the level of competency defined under (1).

The definition of the level of competency of forensic practitioners is the responsibility of professional organizations. As such, it is one of the goals of the California Association of Criminalistics (CAC) and the Forensic Science Society (FSS) [8]. As professional organizations are constituted of members active in the forensic community, they are the decision makers that will act in the best interests of the profession.

Although it is obvious that there should be a minimum level of competency common to all forensic scientists – fundamental forensic concepts, principles of identification and individualization, chain of custody, ethical constraint – each discipline's certification will have its own definition of KSA. The European Network of Forensic Science Institutes (ENFSI) published a comprehensive document defining the baseline competency for forensic practitioners [9]. This document exposes the general concepts that all forensic scientists should acquire, and may be used as a baseline to develop other certification programs in different specialties. Nevertheless, other professional organizations have defined competency levels for the certification in different specialties. For example, the Scientific Working Group for Fire and Explosives (SWGFEX) produced the *Training guidelines*

for the fire debris analyst, which defines the minimum KSA for a person to become a fire debris analyst [10] (*see also* **Fire Investigator: Standardization, Accreditation, and Certification**), just as the Scientific Working Group for the Analysis of Seized Drugs (SWGDRUG) produced a *Training and Education* document defining the minimum KSA criterion for a drug analyst working in a forensic laboratory [11]. Other organizations have published documents in their respective disciplines [12–14].

Once the KSA is defined, it is necessary to develop a method of evaluating a candidate and determining whether or not the competency level has been reached. There are several possibilities to test a candidate's competency. In general, an efficient certification program will include a written examination, an oral examination, and a practical examination or a proficiency test. It is only through the combination of these three exams that the criminalist's competency can be comprehensively tested. A written examination tests the entirety of the knowledge of the discipline. Such an examination should not be exclusively constituted of multiple-choice questions, but should also include some development questions. An oral examination tests one's capacity to reason with the acquired knowledge on a particular topic. As an alternative to a regular oral examination, the candidate could be asked to defend a certain number of case reports. This tests one's aptitude to justify one's actions and to reason through a particular scenario, a very important quality. Finally, a practical examination, while expensive and cumbersome to organize, is the best manner of observing a forensic practitioner in his/her everyday medium and evaluating his/her ability to apply his/her knowledge in practice. For example, a fingerprint examiner would be asked to make a few fingerprint comparisons and identifications or a drug analyst would be required to analyze several unknown samples. Alternatively, a proficiency test can be used *in lieu of* a practical examination. In this case, it is important to understand that there will be no direct control over how the test is performed.

All of this effort would be useless if the professional community did not recognize the certification program. This goal is automatically (but not instantaneously) achieved when the first two conditions are irreproachably fulfilled. However, the reputation of a certification program takes a long time to be established and is highly influenced by the behavior of its members. Also, the independence of the certification organization is highly influential to the well being of the program. For example, an organization should not offer both the certification program and the training to take the certification. Thus, a university can offer training, but should not certify its graduates. This lack of independence would damage the professional community's recognition of the certification. When a particular discipline benefits from several organizations, it may not be wise for each organization to develop its own certification program(s). Besides the obvious redundancy, this contributes to a dilution of the programs and a diminished recognition of the value of the certification. As a good example, the American Board of Criminalistics offers several certifications; however, they do not offer any training.

Certification Process

Each certification organization has its own process. However, a generic process is described here (Figure 2). The process can be divided into three main steps: application, evaluation, and recertification.

First, the criminalist who desires to obtain certification must complete an application, for which a certain number of criterion or prerequisites must be met. These prerequisites usually concern education (and possibly training) and experience. Other requirements, such as citizenship in the country in which the certification organization is registered, a minimum age, or no criminal records, also may be necessary. In some rare circumstances, the organization may require a particular training. This is the case with the National Association of Fire Investigators (NAFI), which requires each candidate for the vehicle fire investigator certificate to attend their four-day *NAFI/CVFI vehicle fire investigation course* [15]. A more elegant solution would be for an organization to require that the candidate follow a class or course provided by a third-party organization.

Once the application is received by the organization and all requirements are successfully met, the candidate may be evaluated through examinations. The registration fee for the examination is usually a couple of hundred US dollars. Whenever the candidate successfully passes all examinations, a certificate is issued for a given period of time.

Since the validity of the certification is limited in time, a recertification process ensues. In general, most

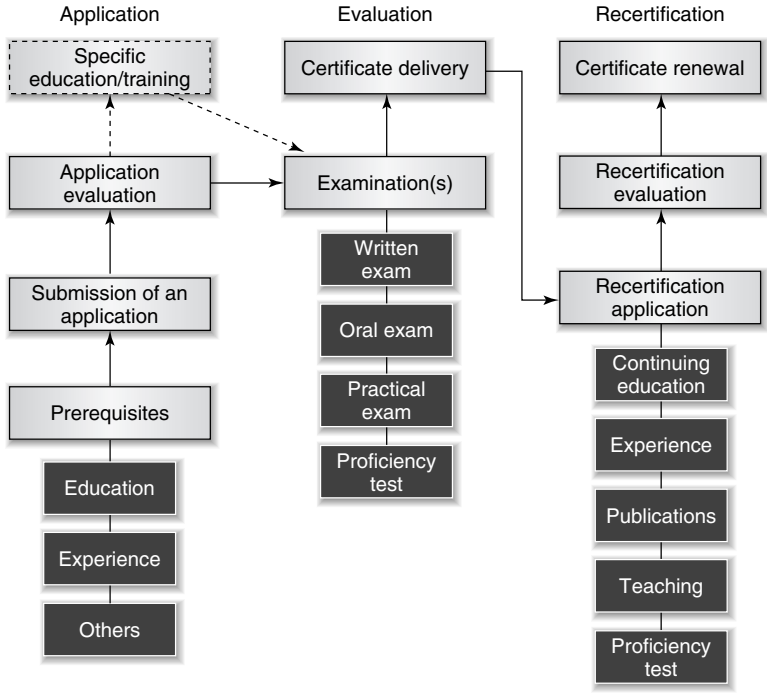


Figure 2 The generic process of certification

certificates are valid for a period ranging from two to five years. The recertification process usually consists in completing paperwork and demonstrating that the requirements necessary to maintain certification were fulfilled. These requirements typically consist of continuing training – practical experience and attendance at courses or conferences – as well as the publication of scientific articles and successful completion of proficiency tests. A system with points for each category is developed and, upon the acquisition of a given number of points, the certificate is renewed for a new period of validity.

Advantages and Drawbacks

It is important to understand the limitations of a certification program. A perfect certification program defines rigorous competency criteria and guarantees an infallible evaluation of a criminalist, and, as a result, certification should be a synonym of competency. This is the idea behind a certification program: To certify the criminalist's competency with full guarantee. Unfortunately, in practice, it is not possible to consider all certificates as proof of competency [7]. The reasons behind the gap between competency and certification are multiple; however, the main one lies in the evaluation of candidates. To date, there are no flawless examination processes that will guarantee the competency of the successful candidate. This problem has been recognized by the American Board of Criminalistics, which states that [16] "No single examination is a total measure of an individual's ability to do the work. Examinations measure knowledge and reasoning" and that "knowledge measured on an examination only reflects the understanding at that point in time. Continuing professional education is needed."

Another fundamental question also comes into play. In the adversary system, an expert witness can be called to testify and must first qualify to do so. During this procedure, referred to as *voir dire* in the United States, the certified criminalist will use his/her certifications as credentials to qualify as an expert witness. As such, if the certification is simply obtained through the payment of a fee, it should not allow someone to qualify as an expert. This will lead to a system in which the certified incompetent is recognized while the noncertified

Table 2 Some certification programs with some of their characteristics

Organization	Name	Education	Prerequisite	Required experience	Examination type
American Board of Criminalistics (ABC)[a]	Diplomate (D-ABC)	Bachelor, or equivalent, in a natural science or an appropriately related field from an accredited institution	Two letters of recommendation	Two years full-time experience[b]	Written (MCQ)
	General criminalistics Fellow (F-ABC)	Bachelor, or equivalent, in a natural science or an appropriately related field from an accredited institution	D-ABC certification – proficiency test – two letters of recommendation	Two years experience in the specialty[b]	Written (MCQ)
	Forensic biology Drug analysis Fire debris analysis Trace evidence				
	Technical specialist (S-ABC)	Bachelor, or equivalent, in a natural science or an appropriately related field from an accredited institution	Proficiency test – two letters of recommendation	Two years experience in the specialty[b]	Written (MCQ)
	Drug analysis Forensic molecular biology				
International Association for Identification (IAI)	Certified crime scene investigator (CCSI)	None	Two courses in crime scene investigation – two letters of recommendation	One year	Written (MCQ)
	Certified crime scene analyst (CCSA)	None	Four courses in crime scene investigation – two letters of recommendation	Three years	Written (MCQ)
	Certified senior crime scene analyst (CSCSA)	None	Six courses in crime scene investigation – two letters of recommendation	Three years	Written (MCQ)

(*continued overleaf*)

Table 2 (*continued*)

Organization	Name	Education	Prerequisite	Required experience	Examination type
	Certified footwear examiner	High school diploma	Education specific to the discipline – two letters of recommendation	Three to eight years[b,c]	Written (MCQ)
	Bloodstain pattern examiner	None	240 h of discipline-specific education – two letters of recommendation	Three years following the 40-h basic training	Written
	Forensic art certification[d]	None	80 h of discipline-specific education – submission of two drawings, which led to the identification of a suspect – two letters of recommendation	One year and 25 drawings	Written (MCQ) and practical (1 drawing)
	Forensic photography certification	None	40 h of discipline-specific education – two letters of recommendation	Three years[b]	Written (MCQ) and practical (10 photographs)
	Certified latent-print examiner (CPLE)	None	80 h of discipline-specific education – two letters of recommendation	Two to four years[c]	Written, practical (50 fingerprints), and oral defense
	Tenprint fingerprint certification	None	96 h of discipline-specific education – two letters of recommendation	Two to four years[c]	Written, practical, and oral defense
Association of Firearm and Tool Mark Examiners (AFTE)	Firearm evidence examination and identification	None	Membership in the association	Three years experience in the discipline	Written and practical
	Toolmark evidence examination and identification	None	Membership in the association	Three years experience in the discipline	Written and practical
	Gunshot residue evidence examination and identification	None	Membership in the association	Three years experience in the discipline	Written and practical

The Forensic Science Society (FSS)	Crime scene investigation	None	Three referees	Three years experience in the discipline	Written, practical, and oral
	Document examination	None	Three referees	Five years experience in the discipline	Written, practical, and oral
	Fire investigation	None	Three referees – case files submission	Five years experience in the discipline	Written and oral
	Firearms examination	None	Three referees	Five years experience in the discipline	Written, practical, and oral
	Forensic imaging	None	Three referees – case files submission	Five years experience in the discipline	Written, practical, and oral
	Document examination	None	Three referees	Five years experience in the discipline	Written, practical, and oral
	Fire investigation	None	Three referees – case files submission	Five years experience in the discipline	Written and oral
	Firearms examination	None	Three referees	Five years experience in the discipline	Written, practical, and oral
	Forensic imaging	None	Three referees – case files submission	Five years experience in the discipline	Written, practical, and oral

MCQ, multiple-choice questions

[(a)] At the time of writing of this article, ABC was changing their certification programs. Please refer to their website http://www.criminalistics.com for the most current information

[(b)] In these cases, the candidate must also be actively working in the discipline of certification

[(c)] The required experience varies according to the candidate's education

[(d)] This certification is under revision and the new one will include more stringent training and experience requirements

competent criminalist is not. Thus, it is necessary to ensure that the gap between certification and competency is as small as possible.

The advantages of certification lie at the level of the employer, the employee, and clients. For the employer, it is a first step toward a quality management system. By paying particular attention to the qualification of criminalists, the employer places the criminalist's competency before the techniques used, an important characteristic too often neglected. For the practitioner, it provides a formal recognition of one's achievements, a particularly important fact since criminalistics is not a professional activity ruled by a special law, such as with professional engineers. Finally, for the clients of a crime laboratory, it provides a certain assurance that the personnel are qualified. This confidence is quite important and should not be undermined.

From a critical point of view, the certification may provide a false impression of permanent quality coming from a laboratory. One must remember that certification only concerns personnel and neither equipment nor procedures (*see* **Accreditation: Organizational**; **Accreditation: Laboratory**). As such, it does not attest to the quality of the work performed in a crime laboratory.

Certification Programs

There are several certification programs available in forensic sciences in general and a few that concern criminalistics. Table 2 presents some of the most reputable programs closely connected to criminalistics, along with some of their characteristics.

End Notes

[a.] There is a distinction between education – which refers more broadly to cognitive development, such as the formal one obtained through school studies – and training – which refers more to a vocationally oriented approach providing specific information regarding a particular task [17]. While attending a university program is education rather than training, some training can be provided during such a curriculum, particularly if practical laboratories are taught. In the scope of this article, the term training is to be understood as encompassing education.

Acknowledgment

The authors would like to thank Dr. Sarah D. Brown for her editorial review of this article.

References

[1] Brightman, R. (2005). *Australian Forensic Science – Education and Training for the Future*, National Institute of Forensic Science, Melbourne.

[2] American Academy of Forensic Sciences (2007). *Education Opportunities*, available at: http://www.aafs.org/default.asp?section_id=resources&page_id=colleges_and_universities.

[3] National Institute of Forensic Science (2001). *Tertiary Institutes Offering Forensic Courses*, available at: http://www.nifs.com.au/F_S_A/Courses.asp.

[4] Technical Working Group for Education and Training in Forensic Science (2004). *Education and Training in Forensic Science: A Guide for Forensic Science Laboratories, Educational Institutions, and Students*, National Institute of Justice, US Department of Justice, Washington, DC.

[5] National Institute of Forensic Science (2001). *Student Questions*, available at: http://www.nifs.com.au/F_S_A/StudentQuestions.pdf.

[6] American Society of Crime Laboratory Directors/Laboratory Accreditation Board (2004). *ASCLD/LAB-International Accreditation Program*, American Society of Crime Laboratory Directors/Laboratory Accreditation Board, Garner.

[7] American Society of Crime Laboratory Directors/Laboratory Accreditation Board (2005). *ASCLD/LAB Legacy Accreditation Manual*, American Society of Crime Laboratory Directors/Laboratory Accreditation Board, Garner.

[8] Stauffer, E. & Schiffer, B. La certification de forensicien: concepts de base et regard critique sur l'expérience américaine. *Revue Internationale de Criminologie et de Police Technique et Scientifique 2007 LX(octobre-décembre)*, 461–478.

[9] Barnett, P.D. (2000). The role of forensic science professional organizations in the new millennium of accreditation, certification, registration and standardization, *Science & Justice*, **40**(2), 138–142.

[10] ENFSI Standing Committee for Quality and Competence (2004). *Performance Based Standards for Forensic Science Practitioners*, European Network of Forensic Science Institutes.

[11] Technical and Scientific Working Group for Fire and Explosives (2006). *Training Guidelines for the Fire Debris Analyst*.

[12] Scientific Working Group for the Analysis of Seized Drugs (2006). *Training and Education*.

[13] Scientific Working Group on Friction Ridge Analysis, Study and Technology (2002). *Training to Competency for Latent Print Examiners*.

[14] Scientific Working Group for Firearms and Toolmarks (2001). *Recommended Guidelines for Developing a Training Manual for Forensic Laboratories*.
[15] Scientific Working Group on Digital Evidence and Imaging Technology (2004). *SWGDE/SWGIT Guidelines & Recommendations for Training in Digital & Multimedia Evidence*.
[16] National Certification Board (2002). *Certified Vehicle Fire Investigator – Certification Program Guidelines & Handbook*, National Association of Fire Investigators, Sarasota.
[17] American Board of Criminalistics (2004). *American Board of Criminalistics History*, available at: http://www.criminalistics.com/history.cfm.

ERIC STAUFFER AND BEATRICE SCHIFFER

Transfer: DNA

DNA is skin cells and in other cells contained in body fluids, such as blood, saliva, urine, and semen. Each of us has about 10^{14} cells in our body (100 000 000 000 000), almost all with a full DNA profile packed inside them. We lose a number of these cells every minute of every day. A typical fingerprint contains many dozens of cells and one should expect to leave one's DNA everywhere that one goes. This is one of the main differences between DNA and fingerprints; DNA can be transferred from one person to another and from that other person to some other place where the first person may have never been. Most contacts between people and objects are expected to result in the transfer of cells, and hence DNA, from one to the other and vice versa.

A large number of parameters (such as, but not limited to moisture levels, texture of the surfaces involved, strength, and duration of contact) need to be considered when evaluating the potential for DNA transfer in any given situation. Consequently, the amount of DNA transferred, and the ability for subsequent transfer to other people and objects, has not yet been thoroughly explored in research studies. Most of the limited research that has been conducted has used small numbers of individuals (about 4–10) in a very limited range of scenarios. In summary, no clear picture has emerged that could enable a confident interpretation of the mechanism that caused DNA to be present when the finding of DNA is not associated with any particular body fluid (e.g., blood or semen) on an object.

This discussion highlights some of the comments and conclusions that feature in the published work. Operational scientists may attempt to make reference to their casework and experience as a source of informed data on these points, but it should be obvious that it is impossible to use this as a substitute for controlled experiments or to draw confident conclusions, no matter how many profiles are observed or objects examined, without knowledge of what *actually* caused the DNA to be present. Within this limited context, it can be reported simply how many objects submitted for such testing can yield profiles; even then, this is a biased sample of objects, since objects are unlikely to have been submitted unless there was an initial expectation that someone's DNA would be associated with them.

The presence of a DNA profile matching that found on an item does not necessarily show that a particular person ever had direct contact with the item. "It has also been shown that a full profile can be recovered from secondary transfer of epithelial cells (from one individual to another and subsequently to an object) at 28 cycles" [1].

"The full DNA profile of one individual was recovered from an item that they had not touched while the profile of the person having contact with that item was not observed. This profile was also detected using standard 28-cycle amplification" [2].

These statements were reporting experimental results performed in controlled conditions and so the extrapolation to casework is difficult other than to establish that such transfer is possible.

The authors state, "When 30 min and 1 h delays were incorporated between human contact and contact with the object, profiles from both individuals were detected. This suggests that for most casework situations where secondary transfer is concerned, mixed DNA profiles can be expected" [2].

This remains speculation.

They also state, "In the context of other cases or scenarios where the conditions are different to those tested within this work, specific experimentation would be required to generate the appropriate data on secondary transfer. It would be important to hear what the 'suspect' says in the context of each case."

One author concludes, "a detectable secondary transfer is very unlikely" [3] although there is no

quantitation of "unlikely", the level at which detection can be reliably accomplished, nor supportive evidence from controlled trials. While this paper asserts that "case experience has found that the handled object bears the profile of the most recent handler", several other works contradict this claim. While the DNA of the last person to touch the item may, or may not, be on the item, the DNA of others who may or who may not have touched the item may also be present. There is no reliable way of ascribing a particular profile to a particular time.

The factors affecting transfer of DNA are not fully known, but there is work to suggest that, *inter alia*; the duration of contact (longer – more transfer), force (higher – more), friction (more – more), and movement (more – more) are important factors. None of these are unexpected.

The results obtained in some of the papers discussed so far did not replicate (agree) the work of another group who had published in 1999 [4]. The improving sensitivity of current techniques may explain this discrepancy. Another may be the variable influence of personal factors in the ability of the participants, of which there are usually few (less than 10), to shed DNA.

There is work that has discovered that people vary greatly in their ability to shed DNA:

> "The observation made by other authors that individuals differ in their tendency to deposit DNA when in contact with an object has been confirmed through specific experimentation ... The reason for this difference in shedding ability is as yet unknown."
>
> "It is well known that there are significant differences among individuals in terms of the amount of DNA that they deposit upon touching an object. The reasons for these differences remain unclear and need further investigation" [5].

> "We have also determined that the quantity and quality of DNA profiles recovered is dependent upon the particular individuals involved in the transfer process. The findings reported here are preliminary and further investigations are underway in order to further add to understanding of the issues of DNA transfer and persistence" [2].

The literature contains preliminary studies or case histories on the possibility of recovering DNA from fingerprints left on the skin or in rope, cord, wire, etc., used for strangling, gloves, knives, solid parts of cars, and other objects, and on the interference of substances used to highlight fingerprints during later genetic analysis.

These works report on isolated experiments, dictated by the need to resolve definite cases, but "*systematic studies on recovery techniques, interference by contaminants, the influence of individual and exogenous factors in the number of cells left ... , and the percentage of success in PCR analysis of the genetic impression from nuclear DNA, have not been exhaustively carried out*" [6] (the quoted text has been italicized here).

Hence, there is a lack of definitive work on the transfer of DNA, but there are, nevertheless, studies that show that DNA can persist for some time after contact.

The highly variable results obtained by experimenters in this area merely add to the uncertainty about *any* conclusion based on the finding of low amounts of DNA on an object.

All of the work cited refers to person to person, person to object, or person to person to object transfers. We can find no work referring to object to object transfer, although this is clearly a possibility; given the precautions taken by laboratories and other agencies in avoiding such, there is one study, undertaken as part of a court case, that has demonstrated tertiary transfer [7].

Primary transfer (i.e., person A to person B or item) is always going to occur more frequently than secondary transfer (i.e., person B or item to someone or something else), and is therefore inherently more likely. It is usually impossible to quantify the actual or relative probabilities of these.

References

[1] Peel, C. & Gill, P. (2004). Attribution of DNA profiles to body fluid stains, *International Congress Series* **1261**, 53–55.

[2] Lowe, A., Murray, C., Whitaker, J., Tully, P. & Gill, P. (2002). The propensity of individuals to deposit DNA and secondary transfer of low level DNA from individuals to inert surfaces, *Forensic Science International* **129**, 25–34.

[3] Wickenheiser, R.A. (2002). Trace DNA: A Review, Discussion of theory, and application of the transfer of trace quantities of DNA through skin contact, *Journal of Forensic Sciences* **47**(3), 442–450.

[4] Ladd, C., Adamowicz, M.M., Bourke, M.T., Scherezinger, C.A. & Lee, H.C. (1999). A systematic analysis of secondary DNA transfer, *Journal of Forensic Sciences* **44**(6), 1270–1272.

[5] van Oorschot, R.A.H., Phelan, D.G., Furlong, S., Scarfo, G.M., Holding, N.L. & Cummins, M.J. (2003). Are you collecting all the available DNA from touched objects? *International Congress Series* **1239**, 803–807.
[6] Pesari, M., Buscemi, F., Alessandrini, F., Cecati, M. & Tagliabracci, A. (2003). Qualitative and quantitative analysis of DNA recovered from fingerprints, *International Congress Series* **1239**, 947–951.
[7] http://www.bioforensics.com/conference07/Transfer/Taylor&Johnson%20Study.pdf, (accessed on, 2008).

ALLAN JAMIESON

Transfer of Trace Evidence *see* Trace Evidence: Transfer, Persistence, and Value

Trauma *see* Rape Trauma Syndrome

Trauma Analysis of Skeletal Remains

Introduction

The role of the forensic anthropologist, when dealing with human remains is not only to establish the biological profile, time of death, and personal identification, but also in helping the pathologist in interpreting signs and therefore possible cause of death on skeletal remains. In some ways, the forensic anthropologist is for the skeleton what the forensic pathologist is for an intact cadaver. It is frequently difficult to determine cause of death when examining skeletons – most deaths by natural causes and many accidental or violent deaths do not leave signs on the skeleton. However, when studying skeletal remains that may belong to a murder victim, the anthropologist must look for the smallest signs of trauma that may indicate the application, for example, of blunt or sharp force injury. In these cases, the experience of an expert osteologist is crucial.

Signs visible on bone are generally of those produced by mechanical injury or any biodynamic reaction to a mechanical insult. Trauma modalities are the same as for soft tissues. Therefore details of different types of lesions are not discussed (since they have been discussed elsewhere). However, forensic anthropology has studied peculiar aspects of trauma specifically on bone [1–42], which are briefly summarized in the following paragraphs.

Bone Reaction and the Issue of Aging a Lesion

In the forensic context, it is necessary to distinguish when a particular lesion was inflicted with respect to the time of death. When one is examining a well-preserved cadaver (or a living person for that matter), it is fairly easy to distinguish if the lesion was produced during life or after. If a lesion is vital, tissues around the lesion (a bruise, a cut) will be a red or purplish color, and if the person survived a good deal after a specific lesion was inflicted, one can also, according to chromatic variations of the wound, more or less determine how much time has gone by. It is common knowledge that a bruise is at first blue, becomes green, brown, and then yellow. If, for example, the wound is inflicted already on a dead body, no such vital reaction will be seen on the skin.

The situation is very different for bone, and it is frequently impossible to determine whether a bone fracture found on a skeleton was inflicted just before or just after death (for reasons that are clarified below). Therefore, forensic anthropology, when dealing with trauma, has developed a different type of nomenclature that divides skeletal trauma into three different types: antemortem trauma, perimortem trauma, and postmortem trauma.

Antemortem Trauma

Bone, when vital, can react only in two different ways to trauma (or any type of insult, for that matter: mechanical, chemical, or microbiological). The two

main responses are the lytic response, i.e., a loss of tissue or a reaction that "take away" bone tissue; or a proliferative response. Generally, in cases in which the person survives the trauma, mechanical insults give a proliferative reaction. Therefore, if a person dies immediately after a bone fracture, for example, there will be no time for the proliferative reaction to develop. If the person survives, vital reactions will be visible – this is the case, for example, of subjects that have been subjected to torture before dying. On the skeletons of these individuals one can still detect, if the survival time was long enough, signs of a vital reaction.

Bone trauma inflicted weeks or months or years before death of the individual can frequently be diagnosed on skeletal remains. When a bone is lesioned, fractured for example, a series of events takes place: Initially, in the living person, vascular damage produces a hematoma around the region involved. Within days a fibrous matrix produces a support for a successive formation of the bone callus.

These reactions occur also in the presence of bone contusions or other types of trauma, even if a fracture is not present.

However, when a fracture is present on healthy bone, the sequence of reactions is the following: (i) there is a circulatory phase that is related to the loss of blood due to the rupture of vessels, which seeps into the tissues and calls in white blood elements and platelets – this occurs between the first hours and the first two weeks after trauma; (ii) the first metabolic phase, which consists of soft callus formation. The cells called in by the inflammatory reaction become cartilage cells, which can create a soft callus that unites the two fracture edges; (iii) the second metabolic response, which implies the maturation of the initial callus with osseous apposition. At this stage, periostitis begins to become visible. Final remodeling will follow.

So initially, the callus consists of woven or young bone (an immature version with a completely different aspect from mature bone). Immature bone has the appearance of a porous "lacey" structure; mature bone looks like normal adult bone, with a very smooth surface (Figures 1 and 2).

These lesions, where a bone reaction can be seen, are called *antemortem*. However, the time necessary for signs of survival to trauma to be evident depends on many factors. These include the site of the lesion or the health status of the person, or

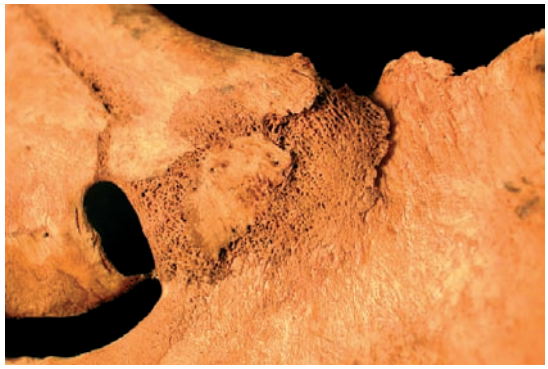


Figure 1 A close-up of the superior part of the scapula where one can notice the presence of porous, slightly darker bone between two fracture edges. This is woven bone and has been aged at around four weeks before death



Figure 2 Tibia and fibula, at the proximal end (next to the knee) of the leg. One can appreciate on the fibula (the thinner bone) a bulge that indicates the presence of callus formation. The bone of the callus is less porous and more robust – the lesion occurred one year before death

even environmental and genetic factors. Sledzik and Kelly studied 257 war victims who had survived skeletal lesions (with different times of survival). These authors found that bone remodeling appeared already after 13 days from trauma; some individuals even showed osteological bone reactions at seven days from trauma. The first appearance of woven bone can be seen as "periostitis", i.e., immature bone, deposited on the margins of the fracture and slowly closing it. Minimum time since death when periostitis is visible is one week. Other clinical sources also indicate that new bone formation after a fracture may be macroscopically visible between 4 and 10 days, but generally occurs between 10 and 14 days from

the lesion. The formation of the callus (first made of soft tissue and then of bone) occurs between 14 and 40 days, sometimes 90, while final remodeling of the callus into adult bone usually occurs between one and two years.

When a visible callus has been formed, it is convenient to perform a radiological analysis of the bone – in fact this may show the internal state of bone remodeling and help to age the lesion.

It is obvious that detection of antemortem trauma is useful for verifying trauma inflicted some time before death, which may be related to maltreatment or torture; however, it may be related to more "normal" events, such as accidents, and therefore be useful in building the biological profile.

Perimortem Trauma

All lesions called *perimortem* occur when the bone is still elastic: the margins of the fracture will therefore have an elastic aspect (green fractures) and will often be squashed. One can imagine a thin branch, when it is still green and not dry (Figures 3 and 4). If one tries to break it, it will bend and finally break, but in an "elastic" fashion. This peculiar mode of breaking can occur, however, in two different situations: when the person is alive, i.e., just before death, and therefore the trauma has something to do with the cause or moment of death; or just after death, when some sort of early postmortem trauma (maneuvers on the cadaver, geological events such as rocks falling on the body) has fractured the bone. In this last case, the macroscopic reaction of bone will be the same as when the trauma is inflicted just before death, since tissues will still be elastic. Usually, perimortem fractures show bone spicules that are slightly intro- or extroflexed. When the skeleton is buried, an indication that the trauma was perimortem also comes from the fact the fracture edges are "dirty", i.e., they have in time absorbed soil and therefore have become colored.

Thus, in forensic anthropology, the term perimortem means either shortly before or shortly after death – traditional methods cannot go further; however, the adoption of histopathology techniques may indeed solve this issue: forensic histopathology (microscopic) techniques may, in theory, be able to still find some sign of vital reaction in a specific fracture (i.e., the presence of early bone remodeling (Figure 5) or of those soft tissue factors that



Figure 3 "Green" perimortem fracture of the distal third of a femur. The elastic appearance as well as the dark edges of the fracture indicate its perimortem origin



Figure 4 The posterior spinous processes of a vertebra is broken in a perimortem fashion

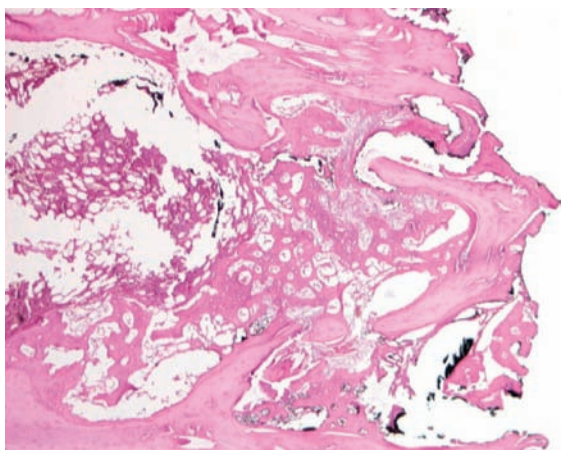


Figure 5 Microscopic analysis of the edge of a perimortem fracture of clean bone showing the microscopic presence of early bone remodeling, which indicates a brief time of survival. No signs of vital reaction were visible macroscopically in the bone

may remain microscopically visible even after most soft tissue has decomposed). Regardless of the fact that histopathology techniques could improve forensic anthropology diagnostic capabilities, no one has ever looked at the presence of vital reactions in "clean", dry or macerated bone. As mentioned above, the diagnosis of the vitality of a skin wound, for example, (whether it has been produced antemortem or postmortem) as well as determination of the time elapsed between the production of the wound and death (in other words the age of the lesion) is a crucial issue in forensic pathology. It is a well-known fact in forensic medicine that the inflammatory reaction that follows wound production will mark the vitality of a peculiar lesion. In fresh skin, in fact, the red-purplish coloration (hemorrhagic infiltration) of a cut or bruise will reveal its vitality, and therefore antemortem formation, whereas the change in coloration, from a macroscopic perspective (from blue to brown to yellow, etc.) will reveal the time of survival. In more difficult cases (e.g., when hemorrhagic infiltration is not so evident or when a more accurate time of survival is needed), microscopic analyses can be performed. A plethora of data from classical histological techniques and newer immunohistochemical techniques provides more accurate statements of survival time, although this depends on the stain used, time from death, etc., new research may shed some light in the future on this important issue and help verify if a "perimortem" lesion is actually antemortem or postmortem.

Postmortem Lesions

It is also possible to verify if the bone fracture has been inflicted on material that was already skeletonized, and therefore certainly classifiable as postmortem. In these cases, since the majority of soft and therefore elastic tissues has disappeared, fractures have a "drier" appearance. Once again one should imagine the example of a branch. In this case, the branch is dry, and it breaks in a very clean fashion (Figure 6). An elastic dry bone behaves in the same manner. These are bone lesions that are frequently determined by the actual excavation or manipulation of skeletons. Usually, they have margins that are white, or at least lighter in color than the surrounding bone, since the fracture edges, having been produced recently, have not been colored by soil pigments.



Figure 6 Example of typical postmortem trauma on a humerus. One can notice that the margin is clean cut, does not have an elastic appearance, and is lighter in color

Also, there exist what are called *taphonomic lesions*: these are due to the destiny encountered by the human remains. Environmental, geological, climatic, animal (Figure 7) or human events (Figure 8) may indeed create lesions on bone, which must be distinguished from the lesions due to the action of a weapon.

Most Commonly Found Traumatic Lesions of Forensic Interest on Bone

Blunt Trauma

Blunt trauma includes all types of impacts of the bone with a resistant surface without points or sharp



Figure 7 The distal extremity of this humerus shows a loss of the lateral part of the bone with adjacent puncture marks. These are due to fauna

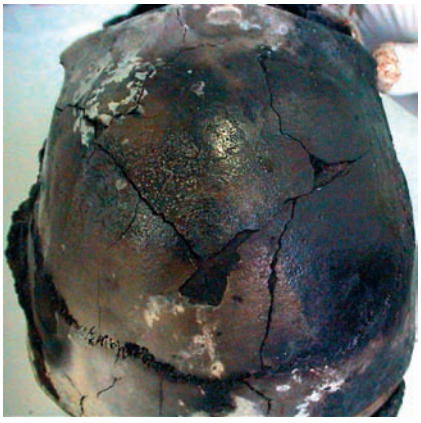


Figure 8 The top of a charred cranium. Fractures on burnt bone on the other hand are extremely difficult to interpret and at times it is impossible to distinguish fractures due to perimortem trauma from fractures caused by the effect of fire



Figure 9 Example of a formed fracture on the cranial vault where a plug of bone of similar dimensions to the weapon (a hammer) was forced inward

edges. Lesions on bone tissues are characterized mainly by fractures, presenting different patterns and therefore modes of production. In forensic pathology and anthropology, analysis of cranial fracture patterns is of utmost importance; the fracture pattern can provide information on lesion production modality and weapon characteristics. Different hypotheses have been developed in order to explain different fracture patterns.

According to the type of blunt surface area, cranial fractures can be classified into formed "impact site" fractures, when a plug of bone is forced inward (Figure 9), and linear or radiating fractures.

The cranium in fact can be considered as a plasto-elastic material; when applied forces are feeble, bone structure undergoes a transient deformation, which disappears after trauma removal. With higher forces, bone structure is permanently deformed. At highest forces, bone structure breaks, with production of fractures.

Lesions in the remaining skeletal districts can be classified into open (when there is also an interruption of skin) or closed fractures; bending, twisting, compression, or tearing fractures according to their shape and mode of production; diaphyseal, metaphyseal, or epiphyseal fractures according to the site of lesion; complete, incomplete, or comminuted in conformity with lesion severity.

In conclusion, the analysis of the bone fracture patterns on the skeleton may help reconstruct a part of the dynamics of the traumatic event.

Sharp Force Injury

These lesions are produced by pointed or sharp-edged tools. Penetrating lesions caused by tools such as needles, nails, and screw drivers can sometimes be visible on bone and a detailed study of the lesion, especially if performed with a cast, can help reproduce tool shape and size and allow at times to obtain very precise information. Cutmarks are produced by the sliding of a sharp edge on the bone surface and, once again, can give some information on the instrument used (Figures 10 and 11).

Sharp tools and saws are usually utilized in cases of body dismemberment, a peculiar modality of



Figure 10 Cutmark on the upper surface of the middle third of the left 11th rib due to the passage of the blade of a kitchen knife, while the victim was stabbed



Figure 11 Once the anthropologist has performed the osteological analysis of the lesion, it is then up to the pathologist to verify which soft tissues may have been involved and to hypothesize if and how that specific lesion may have compromised vital functions. Here is a diagram of the anatomical distribution in vivo of the viscera around the lesioned rib (circle)

corpse concealment or mutilation. As in other cases of penetrating or cutting wounds on bone, the in-depth study of the cut can give detailed information on the instrument used, particularly in the case of saws (Figure 12). As saw teeth cut bone a kerf is formed. The corners formed in the initial cut are called *initial corners of the kerf*. As the cut develops, these develop into kerf walls and kerf floor. Kerf floors frequently provide the most information about the points of each tooth and the relation of the teeth to each other in the blade.

Gunshot Lesions

Gunshot lesions are produced by the projection of a bullet by gas expansion, and are classified into single bullet lesions and multiple bullet lesions; the first type is carried out by short gunshot weapons such as revolvers and handguns, the second type by long gunshot weapons such as rifles and shotguns. The single bullet, when it hits a body, produces an entry wound, a path inside the anatomical structure it crosses, and an exit wound (if it does not impact and get trapped inside internal structures, especially

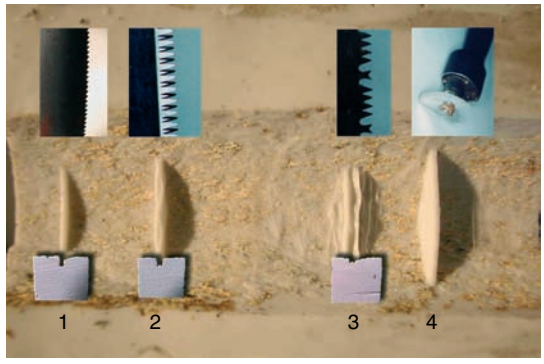


Figure 12 In the figure at the top one can see, from left to right, (1) a saw for metal with small teeth positioned on the same line; (2) a hard wood saw with multifaceted teeth; (3) a green wood saw with teeth laterally inclined with respect to the blade; and (4) a Stryker saw. These saws were used to make four incomplete cuts in the cortical bone of a femoral diaphysis. In the middle of the figure, one can notice the negative cast, made in silicone, of the lesion made by each saw. At the very bottom, the positive cast made in chalk. One can notice, particularly in the silicone negative cast, the different morphology of the kerf floor, which reflects the different characteristics of the saw blade

bone). Entrance wounds in bone, particularly the cranium, are usually (but not always) round shaped and regular; on cranial flat bones, the lesion is characterized by a typical kind of beveling. Entrance wounds are beveled with larger cratering on the interior surface (*tabula interna*) than on the *tabula externa*, probably because of the spreading of forces and lack of support within the *diploe* (Figure 13a, b).

Gunshot fractures are useful in distinguishing an entrance wound from an exit lesion; and fracture lines from an exit wound will break off on preexistent fractures originating from entrance wound.

Where gunshot reaches the cranial surface tangentially, the detachment of a bone fragment produces the typical keyhole lesion, which indicates the gunshot direction

Multiple bullet lesions have different morphological appearances according to gunshot distance; at short distance pellets tend to travel bound together; at long distance instead pellets tend to separate and spread, and each bullet will produce its own entrance lesion.

Mechanical Asphyxia and Drowning

Sometimes the forensic anthropologist may be confronted with the hypothesis of mechanical asphyxia



(a)



(b)

Figure 13 (a) An entrance wound seen from the external surface of the cranial vault: round with a relatively regular margin. (b) An exit wound seen from the external surface of the cranial vault: notice the obvious outwards beveling with cratering of the tabula externa

since suspicious signs may be present on the skeleton. Moreover, among different forms of asphyxia, three modalities may produce bone: manual and ligature strangulation, and hanging. Manual strangulation (throttling) occurs when the neck is constricted by hands. Ligature strangulation occurs by the use of a string or rope, for example, coiled around the neck. Hanging occurs when the coil(s) around the neck is/are tightened by the weight of the body. Manual and ligature strangulation and hanging may produce fracture of the lateral horns of the hyoid, which will be visible on the skeletonized remains but must be distinguished from taphonomic effects.

Death by drowning is a form of mechanical asphyxia due to penetration of water in the respiratory

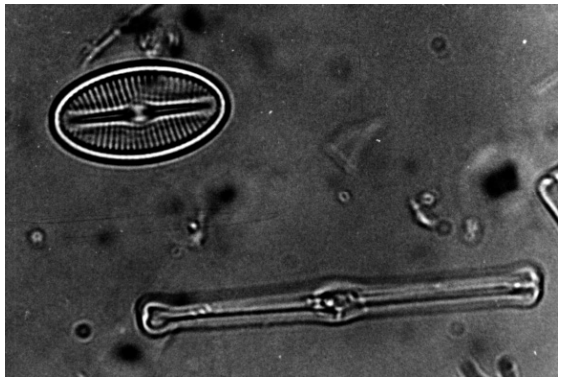


Figure 14 Example of diatoms extracted from human tissues and observed under the light microscope

airways, with replacement of alveolar air. In cases of skeletons found in or nearby lakes, rivers or the sea, the anthropologist should know of the potential of diatom analysis. Diatoms are a major group of eukaryotic algae, and are one of the most common types of phytoplankton. A characteristic feature of diatom cells is that they are encased within a unique cell wall made of silica (hydrated silicon dioxide) called a *frustule* (Figure 14). This makes them particularly resistant. During drowning, diatoms penetrate in the airways with water, enter the blood circulation, and then spread into different tissues and bone structures as well. Thus the presence of diatoms in bone may be an indication of drowning, although much caution must be adopted in the interpretation of such data, since there may be contamination from other sources.

References

[1] Alunni-Peret, V., Muller-Bolla, M., Laugier, J.P., Lupi-Pegurier, L., Bertrand, M.F., Staccini, P., Bolla, M. & Quatrehomme, G. (2005). Scanning electron microscopy analysis of experimental bone hacking trauma, *Journal of Forensic Sciences* **50**(4), 796–801.

[2] Bartelink, E.J., Wiersema, J.M. & Demaree, R.S. (2001). Quantitative analysis of sharp-force trauma: an application of scanning electron microscopy in forensic anthropology, *Journal of Forensic Sciences* **46**(6), 1288–1293.

[3] Berryman, H.E., Smith, O.C. & Symes, S.A. (1995). Diameter of cranial gunshot as a function of bullet caliber, *Journal of Forensic Sciences* **40**(5), 751–754.

[4] Berryman, H.E. & Symes, S. (1998). Recognising gunshot and blunt cranial trauma through fracture interpretation, in *Forensic Osteology: Advances in the Identification of Human Remains*, K.J. Reichs, eds, Charles C. Thomas Publisher, Springfield.

[5] Bohnert, M. (1997). Fractures of the base of the skull in charred bodies – postmortem heat injuries or signes of mechanical traumatisation? *Journal of Forensic Sciences* **87**(1), 55–62.

[6] Bohnert, M., Rost, T. & Pollak, S. (1998). The degree of destruction of human bodies in relation to the duration of the fire, *Journal of Forensic Sciences* **95**(1), 11–21.

[7] Byard, R.W., James, R.A. & Gilbert, J.D. (2002). Diagnostic problems associated with cadaveric trauma from animal activity, *American Journal of Forensic Medicine and Pathology* **23**(3), 238–244.

[8] Calce, S.E. & Rogers, T.L. (2007). Taphonomic changes to blunt force trauma: a preliminary study, *Journal of Forensic Sciences* **52**(3), 519–527.

[9] De Gruchy, S. & Rogers, T.L. (2002). Identifying chop marks on cremated bone: a preliminary study, *Journal of Forensic Sciences* **47**(5), 933–936.

[10] DeHaan, J.D. & Nurbakhsh, S. (2001). Sustained combustion of an animal carcass and its implications for the consumption of human bodies in fires, *Journal of Forensic Sciences* **46**(5), 1076–1081.

[11] Gruspier, K.L. & Pollanen, M.S. (2000). Limbs found in water: investigation using anthropological analysis and the diatom test, *Forensic Science International* **112**(1), 1–9.

[12] Herrmann, N.P. & Bennett, J.L. (1999). The differentiation of traumatic and heat-related fractures in burned bone, *Journal of Forensic Sciences* **44**(3), 461–469.

[13] Holden, J.L., Phakey, P.P. & Clement, J.G. (1995). Scanning electron microscope observations of heat-treated human bone, *Forensic Science International* **74**(1–2), 29–45.

[14] Houck, M. (1998). Skeletal trauma and the individualisation of knife marks in bone, in *Forensic Osteology: Advances in the Identification of Human Remains*, K.J. Reichs, ed, Charles C. Thomas Publisher, Springfield.

[15] Humphrey, J.H. & Hutchinson, D.L. (2001). Macroscopic characteristics of hacking trauma, *Journal of Forensic Sciences* **46**(2), 228–233.

[16] Iwase, H., Yamada, Y., Ootani, S., Sasaki, Y., Nagao, M., Iwadate, K. & Takatori, T. (1998). Evidence for an antemortem injury of a burned head dissected from a burned body, *Forensic Science International* **94**(1–2), 9–14.

[17] Klotzbach, H. Delling, G., Richter, E., Sperhake, J.P., Putschel, K. (2003). Post-mortem diagnosis and age estimation of infants' fractures, *International Journal of Legal Medicine* **117**, 82–89.

[18.] Love, J.C. & Symes, S.A. (2004). Understanding rib fracture patterns: incomplete and buckle fractures, *Journal of Forensic Sciences* **49**(6), 1153–1158.

[19] Myers, J.C., Okoye, M.I., Kiple, D., Kimmerle, E.H. & Reinhard, K.J. (1999). Three-dimensional (3-D) imaging in post-mortem examinations: elucidation and identification of cranial and facial fractures in victims of homicide utilizing 3-D computerized imaging reconstruction techniques, *International Journal of Legal Medicine* **113**(1), 33–37.

[20] Maples, W. (1986). Trauma analysis by the forensic anthropologist, in *Forensic Osteology: Advances in the Identification of Human Remains*, K.J. Reichs, ed, Charles C. Thomas Publisher, Springfield.

[21] Nelson, R. (1992). A microscopic comparison of fresh and burned bone, *Journal of Forensic Sciences* **37**(4), 1055–1060.

[22] Cattaneo, C., Andreola, S., Marinelli, E., Poppa, P., Porta, D. & Grandi, M. (2007). The detection of microscopic markers of haemorrhaging and wound age on dry bone: a pilot study, *American Journal of Forensic Medicine and Pathology* in press.

[23] Pope, E.J. & Smith, O.C. (2004). Identification of traumatic injury in burned cranial bone: an experimental approach, *Journal of Forensic Sciences* **49**(3), 431–440.

[24] Quatrehomme, G. & Iscan, M.Y. (1998). Analysis of bveling in gunshot entrance wounds, *Forensic Science International* **93**(1), 45–60.

[25] Quatrehomme, G. & Iscan, M.Y. (1997). Beveling in exit gunshot wounds in bones, *Forensic Science International* **89**(1–2), 93–101.

[26] Quatrehomme, G. & Iscan, M.Y. (1999). Characteristics of gunshot wounds in the skull, *Journal of Forensic Sciences* **44**(3), 568–576.

[27] Quatrehomme, G. & Iscan, M.Y. (1998). Gunshot wounds to the skull: comparison of entries and exits, *Forensic Science International* **94**(1–2), 141–146.

[28] Quatrehomme, G. & Iscan, M.Y. (1997). Postmortem skeletal lesions, *Journal of Forensic Sciences* **89**(3), 155–165.

[29] Rao, V.J. & Hart, R. (1983). Tool mark determination in cartilage of stabbing victim, *Journal of Forensic Sciences* **28**(3), 794–799.

[30] Reichs, K.J. (1998). Postmortem dismemberment: recovery, analysis and interpretation, in *Forensic Osteology: Advances in the Identification of Human Remains*, K.J. Reichs, ed, Charles C. Thomas Publisher, Springfield.

[31] Rhine, J.S. & Curran, B.K. (1990). Multiple gunshot wounds of the head, an anthropological view, *Journal of Forensic Sciences* **35**(5), 1236–1245.

[32] Sauer, N. (1984). Manner of death: skeletal evidence of blunt and sharp instrument wounds, in *Human Identification: Case Studies in Forensic Anthropology*, T.A. Rathbun & J.E. Buikstra, eds, Charles C. Thomas Publisher, Springfield.

[33] Sauer, N. (1998). The timing of injuries and manner of death: distinguishing among antemortem, perimortem and postmortem trauma, in *Forensic Osteology: Advances in the Identification of Human Remains*, K.J. Reichs, ed, Charles C. Thomas Publisher, Springfield.

[34] Sitiene, R., Zakaras, A., Pauliukevicius, A. & Kisielius, G. (2007). Morphologic, experimental-comparative investigation as an identification of the injuring instrument method, *Forensic Science International* **167**(2–3), 255–260.

[35] Stubblefield, P.R. (1999). Homicide or accident off the coast of Florida: trauma analysis of mutilated human remains, *Journal of Forensic Sciences* **44**(4), 716–719.

[36] Symes, S.A., Berryman, H.E. & Smith, O.C. (1998). Saw marks in bone: introduction and examination of residual kerf contour, in *Forensic Osteology: Advances in the Identification of Human Remains*, K.J. Reichs, ed, Charles C. Thomas Publisher, Springfield.
[37] Thompson, T.J.U. (2005). Heat-induced dimensional changes in bone and their consequences for forensic anthropology, *Journal of Forensic Sciences* **50**(5), 1008–1015.
[38] Tomczak, P.D. & Buikstra, J.E. (1999). Analysis of blunt trauma injuries: vertical deceleration versus horizontal deceleration injuries, *Journal of Forensic Sciences* **44**(2), 253–262.
[39] Tucker, B.K., Hutchinson, D.L., Gilliland, M.F., Charles, T.M., Daniel, H.J. & Wolfe, L.D. (2001). Microscopic characteristics of hacking trauma, *Journal of Forensic Sciences* **46**(2), 234–240.
[40] Ubelaker, D.H. & Adams, B.J. (1995). Differentiation of perimortem and postmortem trauma using taphonomic indicators, *Journal of Forensic Sciences* **40**(3), 509–512.
[41] Villa, P. & Mahieu, E. (1991). Breakage patterns of human long bones, *Journal of Human Evolution* **21**, 27.
[42] Walsh-Haney, H.A. (1999). Sharp-force trauma analysis and the forensic anthropologist: techniques advocated by William R. Maples, Ph.D., *Journal of Forensic Sciences* **44**(4), 720–723.

Related Articles

Blunt Force Trauma
Diatoms
Gunshot Wounds
Length Measurement
Wounds: Sharp Injury

CRISTINA CATTANEO AND DAVIDE PORTA

Trauma Causation: Analysis of Automotive

Objective and Introduction

Humans have developed refined mental activity and physical adeptness but have not evolved the capacity to withstand impacts at high speeds.

In a high-velocity incident, analysis of physical evidence on the human body may assist in identifying the cause of impact trauma. Biomedical engineering integrates engineering, biological principles, and physical evidence for forensic analysis and can therefore provide useful input for the consideration of the cause of the incident, in conjunction with other disciplines including accident reconstruction, forensic pathology, orthopaedic surgery, and neuroradiology.

What is Biomedical Engineering?

Biomedical engineering applies engineering principles to biological systems, holistically utilizing knowledge in engineering, biology, and medicine. The cross-disciplinary field integrates engineering (e.g., external force applied) with biomedical sciences (e.g., tissue tolerance) and clinical input from health care providers (e.g., medical diagnosis).

Biomedical engineering differs from other fields such as mechanical engineering, as biomedical engineers have esoteric knowledge and training about the kinematic and kinetic behavior of biological systems, while mechanical engineers have mechanical system expertise, for example. The key difference between a biomedical engineer and a medical doctor is that diagnosis, prognosis, and treatment are outside the scope of the engineer. The biomedical engineer focuses on the forensic elements or the cause of the incident or trauma on the basis of the engineering aspects. On the basis of a clinician's diagnosis, the biomedical engineer can mathematically model the system, which is often the mechanics/kinematics of the trauma sustained. The next step is to derive quantitative comparisons with the human tolerances for the specific tissues that sustained failure in the subject incident.

In forensic biomedical engineering, physical evidence on the body can be examined and interpreted relative to the body's physical interaction with the environment. The biomedical engineer applies engineering principles to better understand the effect that forces have on the body.

Biomedical engineering may have first begun more than 3000 years ago with the oldest known limb prosthesis. In the year 2000, a 3000-year-old mummy from Thebes bearing a wooden prosthetic tied to its foot, serving as a big toe, was uncovered by German archaeologists.

Historically, biomedical engineering education programs focused on prosthetic design. In the last

50 years the field has also contributed to the study of injury mechanics in accident analysis. Today, there are more than 90 university programs in Biomedical Engineering. In forensic applications, the field focuses on the kinematics and the mechanics of trauma.

Biomedical Engineering Methodological Approach to Trauma Causation

There is no single formula to assess the causation of trauma. There are, however, proven foundations for the biomedical engineering analysis of trauma.

The engineer considers multiple variables to determine the probability of each potential cause of injury. These include vehicular damage, change in velocity, the plaintiff's pre-existing medical condition, the temporal relationship to the onset of diagnosis, as well as the expertise of medical doctors, surgeons and/or forensic pathologists to provide the specific diagnosis of trauma. Insights and analysis of the accident investigator may also be used to further understand the mechanism of trauma. Because the biomedical engineering discipline involves training in biology and medicine, biomedical engineers may be called upon to evaluate how external outputs, i.e., forces and accelerations on the body, will affect the internal structures, i.e., load capacity and failure modes.

There are three categories of analytical elements that serve as a foundation for study. These elements are integrated with the available investigative facts in order to evaluate the mechanisms of trauma, and/or determine whether the trauma is consistent with the specific incident. The collective analysis is termed *Mechanical, Analytical, and Medical* (MAM) analysis.

1. Mechanical input–consideration of the static and dynamic forces acting on the body:
   (a) impetus for the trauma/accident reconstruction;
   (b) trauma characteristics;
   (c) kinematics–the study of the positions, angles, velocities, and accelerations of body segments and joints during motion; and
   (d) Preimpact position.
2. Analytical engineering–examination of available evidence in conjunction with accepted scientific principles:
   (a) tissue tolerances and susceptibilities;
   (b) testing;
   (c) computerized mathematical modeling analysis; and
   (d) crash recorder or global positioning system (GPS) data.
3. Medical facts and findings:
   (a) radiological findings;
   (b) clinical facts and findings;
      history
      diagnosis
      contact trauma (bruising, contusions, etc.);
   (c) surgical findings.

The findings of the biomedical engineer are dependent upon the quantity and quality of the available evidence and data. The analysis converges to a net result when all of the elements point clearly to the same causal factor. If the results do not fit consistently toward a single cause, then both consistent and inconsistent findings are identified and/or additional work is recommended.

The Application of Biomedical Engineering to Forensic Disciplines

Biomedical engineering can be applied to a variety of forensic analyses. Here we will focus on the vehicle: for example, use of occupant restraints, pedestrian analysis (contact specifics and injury causation), trauma causation, repetitive stress in the automobile environment, and product defects. A few examples of the types of issues that can be analyzed follow.

Occupant Restraint

The physical evidence of the seat belt condition, the occupant kinematics and trauma, and the vehicle interior may determine whether the occupants were likely restrained, and restrained properly. From this information, other elements can be determined, such as (i) the proportion of the trauma due to the plaintiff's lack of a seat belt versus the defendant's excessive speed or (ii) the likely trauma sustained had the 2-year-old been properly restrained in a child's car seat compared to seated in his mother's lap.

Pedestrian Analysis

Pedestrian analysis may reveal multifaceted data about the incident. For example, (i) the direction the



Figure 1 Pedestrian direction of travel prior to impact

pedestrian was traveling prior to impact (Figure 1), (ii) whether the brain injury occurred from the initial impact or from the subsequent impact, (iii) whether the brain trauma would/would not have been less severe had the vehicle been traveling at the speed limit, and (iv) at the time of impact, whether the boy was on his bike or walking next to his bike in the crosswalk. Had the boy been wearing a helmet, what trauma would he likely have sustained?

Trauma Causation

Biomedical engineering analysis can be utilized to determine whether the mechanics are present, or sufficient force is present, to cause the trauma diagnosed by the medical doctor: for example (i) the statistical likelihood that a brain injury was sustained in a low-force incident; (ii) whether or not the mechanics for shoulder impingement are present in the impact; (iii) whether the boy's arm in the back seat was amputated during ejection, rather than inside the vehicle; (iv) whether the front seat occupant's trauma was due to the kinematics of the unrestrained occupants of the rear seat; and (v) whether, using the principles of fracture mechanics, the plaintiff's elbow implant most likely failed from the unrestrained frontal impact rather than that same day's fall on roller blades (Figure 2).

Repetitive Stress in the Automotive Environment

For example, can the lateral femoral cutaneous nerve damage present in 52 Highway Patrol officers be



Figure 2 Condition before and after elbow implant failure. Is it due to frontal impact or from a roller blade fall?

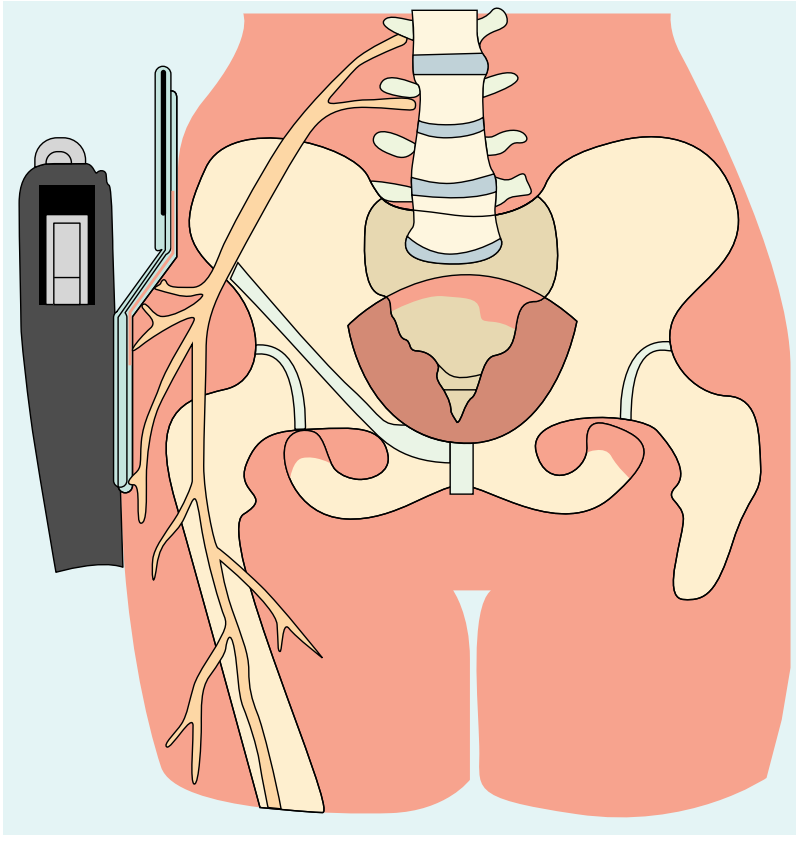


Figure 3 Holster over the pelvis and neuronal structures illustrating compression of the lateral femoral cutaneous nerve

causally traced to a change in holster design that compresses the hip over the lateral femoral cutaneous nerve (Figure 3)?

Product Analysis

Examples include (i) whether the seat anchor failure did or did not cause trauma; (ii) whether or not the injuries were caused by either the ejection or the initial impact itself; (iii) whether the roof should have withstood the force as the commercial truck rolled onto its roof, mechanically asphyxiating the occupant (Figure 4); (iv) whether the globe (eyeball) rupture was caused by the air bag rather than the occupant's diamond wedding ring; and (v) whether the small-stature woman's cervical and facial fractures were caused by the air bag rather than the non-use of seat belts.

The Design and Proper Use of Biomedical Engineering Experiments

The previous section outlined how biomedical engineering may be applied forensically. This section describes the added value of well-executed experiments to determine scientifically the most likely causal factors.

A means of determining causation is to experimentally analyze the issues. These experiments can make use of variants in the environment to compare results. Common questions that experiments may help to answer include:

- Would a seat belt have prevented specific trauma?
- If the vehicle were traveling within the speed limit, what trauma would most likely be sustained?
- Which of the ejected decedents was driving?



Figure 4 Should the roof have withstood the weight of the commercial truck?

- What would have been the probability for brain injury in the subject incident had the seat belt been worn correctly?
- Were the bilateral locked cervical spine facets caused by the first or second impact?

Engineering experimentation can be called upon to assess the biomedical engineering elements of vehicular accidents through two principal methods:

1. Interpolation using existing experiments to "bookend" a specific incident.
2. Custom experimentation, which involves designing a custom experiment to answer a specific question.

Using Existing Experiments

Exhausting the sources of existing experiments is recommended as a first step, even if custom experiments are eventually required. If the subject incident falls into a category of standardized crash tests (in the United States an example is the New Car Assessment Program's frontal impact at 35 mph), existing experiments may be sufficient and more cost/time effective. If not, customized experiments may be necessary. Internationally, there are many sources for crash tests and experimental results. In the United States the most common include the following:

- National Highway Safety Administration (NHTSA)
- Insurance Institute for Highway Safety (IIHS)
- Society of Automotive Engineers (SAE)
- American Association for Automotive Medicine (AAAM)
- Other State or National Agencies
- Automotive Manufacturers.

The Application of Standardized Crash Tests: Interpolation. The key here is how the standardized experimental results may be properly applied to a specific problem. The following short examples will illustrate the key points: *interpolation, not extrapolation*. Interpolation of engineering experiments is the means to explain the expected result or trend between two or more data points. Take the example of two crash tests, one at 10 mph and another at 15 mph. If the vehicles are substantially similar, crashes and occupants may be analyzed. The results of a 12.5 mph crash test could potentially be interpolated from the two experimental bookends of 10 and 15 mph.

Conversely, it may be inappropriate quantitatively to extrapolate, or to go outside the data, to project

30 mph results from tests at 10 and 15 mph; however, sometimes nominal data can be determined if conducted correctly. For example, an appropriate interpretation regarding the 30 mph analysis relative to the 15-mph test is that the forces and accelerations would be expected to be higher. The nominate data may prove helpful in the absence of quantitative data. How much higher, however, may not be reliably projected.

Real World Example of Crash Test Interpolation. On the basis of the forces and accelerations of a specific side impact collision, what is the likelihood that a brain injury was sustained?

A driver is side-impacted by another vehicle to the driver's side of the vehicle. Medical doctors retained by the plaintiff and defendant disagree on whether a brain injury was sustained by the driver in the incident. Is a brain injury more or less likely on the basis of the magnitudes of accelerations and forces?

Mechanical. The side impact was direct to the driver's door by another vehicle of like bumper height and weight. There was no evidence of occupant contact to the vehicle interior, such as window fracture, indentations of the body against the door, and/or blood/tissue evidence. The impact was insufficient to deploy the side air bag. Digital data within the vehicle, specifically the air bag computer and crash data recorder data, were not available since the force of the incident was insufficient to "wake up" the air bag algorithm.

The male driver was of average height and weight, so contact with the roof rail or window frame would not be expected. There were no other occupants or heavy cargo in the vehicle at the time which may have contributed to the driver's trauma.

Medical History and Clinical. The driver was middle aged and in good health with no contributory past medical history. A review of the medical records of the injured driver helps to determine whether there was:

- contact trauma sustained by the injured driver, which might yield important information regarding the motion of the occupant during the collision;
- loss of consciousness (LOC) at the scene (more specifically, Glasgow Coma Score) documented by the first responder.

The medical records are also utilized to review the history of present illness, subjective complaints, objective findings, and diagnoses from the health care providers.

A review of the trauma sustained by the other occupants in the injured driver's vehicle may assist the analysis since other occupant(s) may have experienced a similar force condition.

A biomedical engineering analysis of the medical records is not for diagnostic purposes; that is the expertise of the diagnostic clinicians. The medical records may assist in characterizing any trauma or in anatomically understanding the specific trauma sustained. The presence of contact trauma may increase the statistical likelihood of brain trauma. If there is contact trauma, what is its cutaneous character? Does the contact trauma shape, size, or depth assist in pinpointing the source? Is it consistent with a specific surface and does it rule out other surfaces? Is it consistent with a yielding or unyielding surface? (An unyielding surface has the potential to greatly increase the force experienced by the occupant as a function of the topography of the surface.) In a separate unique analysis, a piece of plastic inside the vehicle broke off and impaled an occupant, so the contact could be matched to a specific piece of plastic in the vehicle and a single unyielding point of contact.

In this analysis, there was no clinical evidence of contact documented by the health care providers or recalled by the injured person during the deposition.

Medical: Radiological. Continuing with the example, a review of the head computed tomography (CT) films taken on the date of the accident also confirmed no evidence of contact. For background, it may prove helpful to look at the actual films themselves with the assistance of a radiologist, if needed, to supplement the radiological report. This is not radiologically to diagnose the trauma, but to utilize the physical evidence in determining its cause. For example, the size and shape of a minor scalp swelling may not be clinically important and therefore not mentioned in the radiological reports. Even if it is included, external contact trauma may not be a diagnostic focus, so the precise description of the shape, size, orientation, and depth should not be assumed to be included in the report.

In this specific case, the radiological evidence of no contact trauma was consistent with the clinical and investigative evidence.

Analysis: Research of Existing Testing. A review of standardized side-impact tests identified a test at a change in velocity of slightly greater than the subject vehicle test. On the basis of a comparison of the crash profiles, the subject incident was slightly less of an impact than the published research test that utilized a similar make and model of the (year) vehicle. Therefore, the crash test would be expected to overestimate the forces sustained. The dummy (in the test) also did not contact any surfaces, as its head traveled laterally. The research test showed that the accelerations measured at the center of the dummy's head were less than required to cause a brain injury compared to human tolerance according to Federal Motor Vehicle Safety Standards. Since the actual impact was of lesser magnitude, the forces and accelerations would be expected to be of lesser magnitudes or smaller compared to the standard test. Hence, if the accelerations measured by the dummy in the standardized test at a faster impact speed did not exceed tolerance, then a similar impact at a slower speed would probably not exceed tolerance unless the trajectory or area of contact was dramatically changed.

In this example, the answer could be derived without the time and expense of customized crash testing. However, when the impact or the seating scenario is unusual, crash test results may not be available for interpolation and customized testing may be required. An example of a custom biomedical engineering experiment is given below.

Biomedical Engineering Experimental Design. The second most common categorical use of biomedical engineering forensic experiments is the development of customized experiments to investigate specific issues in a specific incident. If the analysis undertaken does not fit into common categories, such as the side, rear, or frontal impacts, a detailed review of the literature is required. If there is no similar test, then custom test directions may be considered. Test alternatives include full-scale testing (full-scale vehicle into full-scale vehicle), deceleration sled testing, and/or subsystem testing. Subsystem refers to a part of the system; an example will illustrate an experiment that helped to determine whether or not a brain injury would likely have been sustained.

An Example of Design of a Biomedical Engineering Subsystem Experiment. What is the statistical probability that a minor rear impact would cause a brain injury? This section illustrates the use of subsystem experimentation.

Mechanical Considerations. An undersized pickup truck was at complete stop when it was rear-impacted by a passenger vehicle. The accident reconstruction analysis indicated that the change in velocity was approximately 5–10 mph and the principal direction of force was directly rearward without offset (both vehicles' axes were in line). There was no structural damage to the rear of the pickup truck. The damage to plaintiff's vehicle was limited to the bumper. The defendant's vehicle sustained no damage to the front bumper.

Analysis of the Vehicular Interior Environment. The plaintiff reported to health care providers that a head injury was sustained when his head struck the rear window of the truck cab, causing the glass to crack. The vehicle seated three occupants in a front bench seat and has a configuration with a sliding glass window immediately behind the occupants. Contrary to the plaintiff's history provided to health care providers, the photos of the truck (Figure 5), and the repair estimates for the truck, there was no indication that the back window was damaged.

The plaintiff driver was the only occupant in the pickup. The plaintiff was seated normally and naturally, looking straight ahead. There was no hematoma or laceration or other evidence of contact trauma to his head. The vehicle had a head restraint although it did not fit the plaintiff optimally. The head restraint was too low for the 5'10" driver.

Medical Considerations: Review of Health Care Provider Documents. A review of the past medical history included:

- legal actions including two workers' compensation claims, one slip and fall, and one previous MVA (motor vehicle accident);
- two prior DUI (driving under the influence) arrests;



Figure 5 Damage to plaintiff's vehicle in a rear impact

- employment for five to six days in a job before quitting, stating the job "aggravated (his) hypertension";
- comment from five years earlier: "there is no etiology for (his) pains";
- comment from six months earlier: "I cannot explain his current symptoms on an organic basis";
- comment from the last health care provider 2 months prior to the incident: "wants disability extended".

The first health care provider (two days post incident) was a medical doctor who noted the following history provided by the patient:

- rear-ended at 30 mph;
- head broke glass in the pickup;
- bloody nose; and
- no LOC, no problem with memory/smells/weakness.

The diagnosis of the medical doctor was cervical muscle sprain. Further, there were no radiological findings on the cervical plain films (X-ray) or on the brain MRI conducted three weeks post incident. Post-concussion syndrome was first diagnosed one month post incident.

An independent medical examiner with specialization in neurological surgery did a review of the medical records and found: "it would be considered unlikely that any significant injury resulted from the vehicular accident. At most, he may have suffered a mild cerebral concussion and a cervical strain."

A neurologist/psychiatrist also conducted a medical record review and examination. He concluded that there was insufficient evidence to conclude that the possible mood disorder (major depressive disorder) had been caused by the MVA. Additionally, numerous inconsistencies and credibility issues were raised by the reviewer including the following:

- In the deposition, He left a job at "A" hospital for a better job at "B" hospital. Employment records at "A" hospital indicate that he was dismissed for forging a physician's signature for a medical leave excuse.

Vehicle Digital Data: Crash Data Recorder and Air bag Data. There was no crash data recorder on either the plaintiff or the defendant's vehicle. No air bags were deployed in either vehicle, so no air bag data was available.

Previously Conducted Impact Testing. There was no rear-impact testing available in a similar truck that had been conducted at a low velocity with an instrumented dummy.

Design of Biomedical Engineering Experiment to Quantify Head Acceleration. An inverted pendulum subsystem test was designed to contact the seat at the same juxtaposition and velocity as the plaintiff's head/neck might have contacted the seat in a 7.4-mph rear impact as shown in Figure 6.

The digital data was acquired at 10,000 Hz. Analysis of the data utilized the head injury criterion (HIC),



Figure 6 Experimental juxtaposition of Hybrid III Dummy Head relative to suspended rear window glass

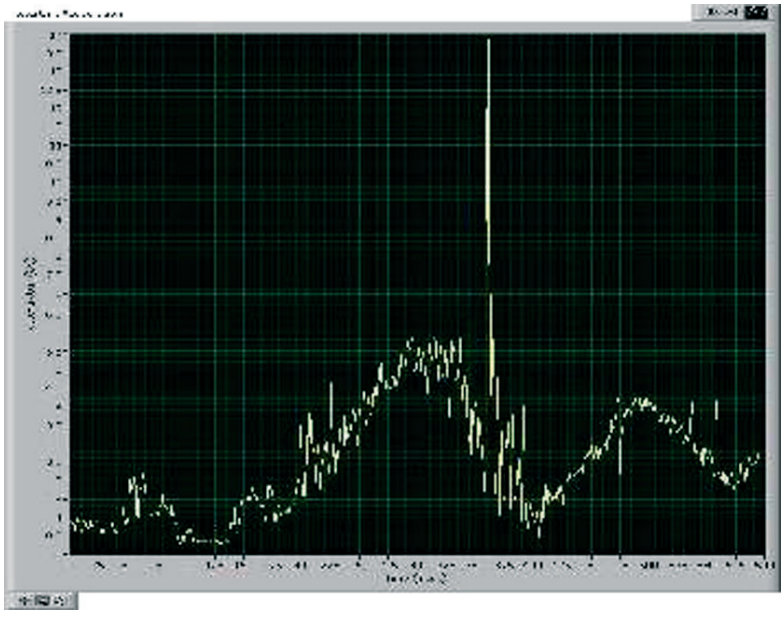


Figure 7 Graph of acceleration in G's (unit of gravity) versus time in milliseconds

or as shown below, where $(t_2 - t_1)$ represents duration of impact and a is acceleration.

$$\left\{ \frac{1}{(t_2 - t_1)} \int_{t_1}^{t_2} a \, dt \right\}^{2.5} (t_2 - t_1) < 1000 \qquad (1)$$

An example of a resultant acceleration versus time graph is shown in Figure 7. The results of the HIC analysis (Figure 8) illustrate that the values are below tolerance. Brain injury would not be likely in this incident on the basis of the HIC results of nine trials.

Findings. In this case, the work by the independent medical examiners consistently indicated that a head injury was unlikely. For example, statements included: "It would be considered unlikely that any significant injury resulted from the vehicular accident", and "At most, he may have suffered a mild cerebral concussion".

The experiments resulted in a means with which to quantify mechanically the likelihood that brain injury tolerances could have been exceeded in the incident. This contribution supplemented the expertise of the health care professionals.

Computerized Mathematical Modeling Analysis

Mathematical modeling allows a scientist to analyze the motion of a human and quantify the forces sustained in an injurious event without exposing human subjects to potential harm. The human body is represented by a number of rigid bodies linked together by springs and dashpots representing the characteristics of human joints and, collectively, the human body. Because of the body's numerous degrees of freedom, the occupant's motions during a collision can be complex. Mathematically, the human body, modeled by connected rigid bodies, can interact with the environment via system inputs or external accelerations. External accelerations include impacts with surfaces, interaction with restraint systems, and/or air bags, for example. Simple systems such as inverted pendulums can be modeled, as well as complex systems such as an unrestrained occupant hitting a complex surface at impact. Note that simulation is different from animation since simulation mathematically defines kinematics and kinetics. In contrast, animation is simply the visualization of motion that may or may not be defined by the simulation or mathematical results. Animation relies upon the quality and validity of the simulation to determine the accuracy of the result.

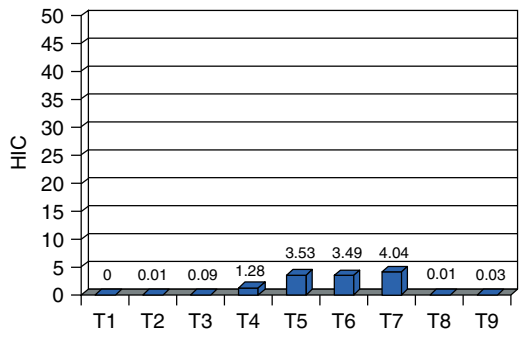


Figure 8 Results of HIC values for nine trials

As an example of this methodology, here we will review a model of a motor vehicle occupant subjected to a same-side impact. The occupant is restrained, and the goal is to study the dynamics of the head; therefore, the lumbar spine, pelvis, and lower extremities are simplified. This is separately determined to be appropriate in this case because the lower body is restrained by the seat belt. The body is modeled in four segments representing the head, neck, thorax, and lumbar spine/pelvis.

The human side-impact (restrained) model with five degrees of freedom is illustrated in Figure 9. The model can be utilized in its two- or three-dimensional (3D) form.

The variables are defined as follows: q_1, pelvic translation; q_2, rotation between the thorax and lumbar spine/pelvis (torsional spring and dashpot kT, bT); q_3, rotation between the neck and thorax (torsional spring and dashpot kN, bN); q_4, rotation between the head and the neck (torsional spring and dashpot kH, bH); q_5, head axial rotation.

Once the model is developed and verified, it can be utilized within those parameters of verification. Examples of model verification include human subjects, cadaver, and anthropometric dummy experiments at impacts that bookend the subject incident (experiments at higher and lower exposures than the subject incident). The model is considered valid as

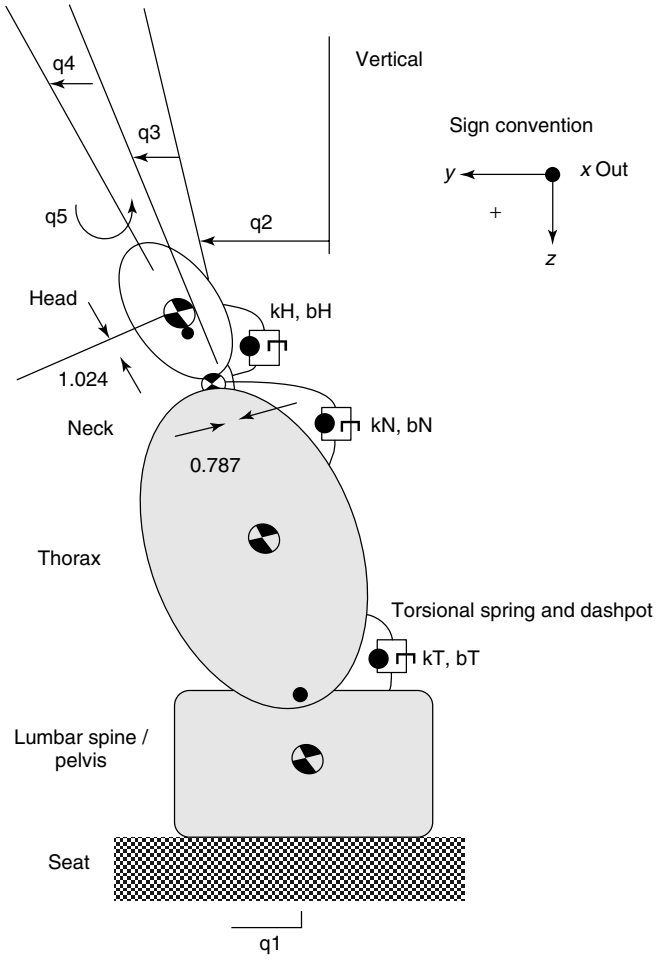


Figure 9 Occupant with five degrees of freedom in a side-impact model

long as the new variables are within the validated parameters.

Properly validated computer mathematical modeling can be an effective biomedical engineering method to quantify the kinematics and kinetics of an impact incident. The correct use of the validated model is interpolation within the experimentally validated parameters. For example, if a model is validated at fall heights of 20 and 25 ft, the model can be used between that range, especially if it is known whether the relationship is linear, or predictably nonlinear and no additional governing variables apply. Conversely, once contact occurs with the automobile dashboard, the dashboard noncontact experiments are often not proper for model validation.

Mathematical modeling can also be utilized when the input variables change. This is especially helpful, for example, when more information becomes available to the accident investigator that alters the known vehicle's change in velocity. The model could be then used to reanalyze the result cost effectively as long as the new variables are within the validated parameters.

References

[1] Austerlitz, H. (1991). *Data Acquisition Technologies Using Personal Computers*, Academic Press, Inc., San Diego.

[2] Avallone, E.A. & Baumeister, T. (eds) (1997). *Marks Standard Handbook for Mechanical Engineers*, 11th Edition, McGraw-Hill, Inc., New York.

[3] Bendat, J.S. & Piersol, A.G. (1986). *Random Data*, 2nd Edition, John Wiley, New York.
[4] Bourne, J.R. (1981). *Laboratory Minicomputing*, Academic Press.
[5] Collins, J.C. (1979). *Accident Reconstruction*, Charles C. Thomas, Illinois.
[6] Colloms, M. (1983). *Computer Controlled Testing and Instrumentation*, Pentec Press, Devon.
[7] Craine, J.F. & Martin, G.R. (1965). *Microcomputers in Engineering and Science*, Addison-Wesley, Reading.
[8] Fraser, C.J. & Milne, J.S. (1990). *Microcomputer Applications in Measurement Systems*, Macmillan, London.
[9] Gupta, S. & Gupta, J.P. (1989). *PC Interfacing for Laboratory Data Acquisition and Process Control*, Instrument Society of America, Research Triangle Park.
[10] Malcolm-Lawes, D.J. (1988). *Microcomputers and Laboratory Instrumentation*, Plenum Press, New York.
[11] Mase, G.E. (1970). *Theory and Problems of Continuum Mechanics*, McGraw-Hill, New York.
[12] Pilkey, W.D. (1994). *Formulas for Stress, Strain, and Structural Matrices*, Wiley, New York.
[13] Sheingold, D.H. (ed.,) (1986). *Analog-Digital Conversion Handbook*, 3rd Edition, Prentice-Hall, Englewood Cliffs.
[14] Stone, H.S. (1982). *Microcomputer Interfacing*, Addison -Wesley.
[15] Van Doren, A.H. (1982). *Data Acquisition Systems*, Reston Publishing.
[16] Van Vlack, L.H. (1985). *Elements of Materials Science and Engineering*, Addison-Wesley, Reading.
[17] Wakerly, J.F. (1981). *Microcomputer Architecture and Programming*, John Wiley & Sons, New York.

LAURA L. LIPTAI

Traumatic Amnesia *see* Recollective Accuracy of Traumatic Memories

Traumatic Brain Injury *see* Head Injury: Neuropsychological Assessment

Traumatic Memories *see* Recollective Accuracy of Traumatic Memories

Treatment of Sex Offenders *see* Sex Offenders: Treatment of

Treatment, Mandated: Mental Health

These risks include a fourfold increase in the risk of suicide, increased risk of violence, arrest, poorer mental functioning, life satisfaction, and long-term prognosis [1]. Far from being an outdated practice, several recent developments in mental health law and treatment has brought the return of mandated treatment to the forefront of mental health law.

In recent years, mandated treatment has been expanded to domestic abusers [2], individuals with severe addictions [3], sexually violent sexual offenders [4], and individuals deemed dangerous as the result of a host of mental illnesses [5]. However, mandated treatment remains controversial given its utilization of legal leverage to accomplish its aim of outpatient community treatment. In cautioning against such incursions against liberty and individual autonomy, Justice Brandeis warned, "The greatest dangers to liberty lurk in insidious encroachment by men of zeal, well intentions, well meaning, but without understanding." However, others argue that the government has the responsibility to prevent harm and provide effective treatment, even at the expense of some individual freedom for those suffering from a severe mental illness that leaves them unable to care for themselves or leaves them disposed to dangerous behaviors within the community. Moreover, many individuals voluntarily agree to participate in treatment thus lessening loss of liberty [6].

Of course, clinicians are faced with issues related to low-base rate, false positives, and false negatives within their clinical assessment of risk. It is with these issues in the forefront that we discuss mandated treatment. This article focuses on three primary issues. First, we discuss important legal precedents concerning applications of mandated treatment. Second, we intend to delineate issues and various definitions of dangerousness, especially as they are related to civil commitment standards. Finally, we discuss recent trends in mandated treatment, with specific emphases

on various populations mandated treatment is applied to. Through exploration of these topics we hope to elucidate issues associated with mandated treatment.

Legal Foundations for Civil Commitment and Community-Based Treatment

Various courts have provided guidance and direction for procedural issues and requirements governing how commitment standards can be applied. Statutes and cases regarding mandated treatment have been wide ranging and each touch upon important constitutional questions related to individual autonomy and self-determination. Such issues include diagnoses related to the civil commitment of sex offenders, conceptualizations of dangerousness, and preserving the delicate balance between preservation of individual rights versus the global concern of protecting the public from potentially dangerous individuals who suffer from mental illnesses. The principle landmark case dealing with civil commitment was the case of Alberta Lessard. Arrested in 1971, Ms Lessard was committed to an inpatient facility on the sole basis of the presence of schizophrenia. In *Lessard v. Schmidt* [7] the Court ruled that mental illness alone was no longer sufficient to justify commitment. Instead, civil commitment was premised on both the presence of a treatable mental disorder *and* linking that disorder to likelihood of violence or aggression. Lessard was a watershed case in mental health law as it signaled a move from *parens patria* to almost exclusive use of the state's police powers in effectuating civil commitments. In voicing an alternative perspective to Lessard, a Wisconsin psychiatrist coined the term "dying with their rights on" to describe the inability to commit seriously ill patients who were in need of mental health care [8]. *Lessard* represented the enduring difficulties inherent in mental health law that often focuses on balancing the individual rights of the patients to be free from coercive treatment and society's demand for safety and treatment related to abnormal behaviors.

The Lessard case was the first in a series of cases leading to changes in civil commitment. The changes resulted in thousands of individuals once committed to hospitals being returned to the community requiring resources to be funneled into community-based treatment programs. For instance, a case management model and assertive community treatment has led the way to treating mentally ill individuals within their home communities rather than reliance on large, institutionalized-based care [9]. According to the National Alliance for the Mentally Ill, assertive community-based programming is available in approximately 19 states throughout the United States. The presence of community-based treatment to assist with mandates is even more essential in light of the recent ruling in ***Olmstead v. LC***, decided in 1999 [10]. In this case, the Supreme Court ruled that the American with Disabilities Act required that states place persons with mental illnesses in community care when deemed clinically appropriate insofar as such placement is a reasonable accommodation given the fiscal ability of the state to provide such care. When combined with the Lessard ruling, the Supreme Court of the United States, through these two decisions, ensured that mentally ill individuals have rights that must be protected, which includes the right to the least restrictive environment while being afforded protections against unlawful and unneeded commitment.

In addition to providing effective treatment, a key consideration of mandated treatment to individuals living in the community with mental disorders is to decrease violence [11]. Moreover, treating and reducing the likelihood of violent behavior is a central consideration in the creation and implementation of policy, especially regarding civil commitment [12]. The data for commitments has shown that the implementation of mandated treatment can have a beneficial effect for both individuals and the community. As noted by Monahan and colleagues [13] mandated treatment falls within the realm of social welfare, yet relies upon the judicial system in successfully monitoring and enforcing mandated treatment. It should be noted that mandated care can take several forms including, representative payeeship, the use of diversion agreements, and brief, involuntary hospitalizations for stabilization of psychiatric symptoms [14]. Not surprisingly, individuals coerced to engage in treatment often feel disenfranchised making continued involvement in treatment difficult.

Changing Conceptualizations of Dangerousness: Mandated Treatment

A recent trend in the implementation of mandated treatment revolves around how dangerousness is

defined. In highlighting this changing conceptualization one only needs to look at variations of civil commitment and increasing such definitions. In discussing specific issues we outline (i) changes to laws in standard civil commitment, (ii) mandated treatment for so-called sexually violent persons, and (iii) the insanity acquittee and community treatment.

Standard Civil Commitment

In a law review article entitled, "Beyond Overt Violence: Wisconsin's Progressive Civil Commitment Statute as a Marker of a New Era in Mental Health Law" Erickson *et al.* [6] outlined changes in many states, with an emphasis on Wisconsin, how overt violence or the threat of it is no longer the only definition considered for civil commitment. The article discussed how legislatures have expanded criteria allowing for civil commitment. This shift in commitment criteria demonstrates a growing understanding that certain individuals are so severely incapacitated by severe mental illness that they are unable to care for themselves. In considering individuals who do not present with overt violence as eligible for mandated treatment, expanded definitions of dangerousness within various state statutes acknowledge the insidious danger poised by untreated mental illnesses to both patients and the public at large. While such polices have a de facto parens patria overtone, consistent with court precedent, they are founded upon the state's police powers and require a finding of dangerousness before treatment can be initiated.

Sex Offender Commitment

One of the most contentious areas of mental health law concerns the civil commitment of sex offenders who already served their correctional time but are deemed to pose a continued risk of offending, typically related to a mental disorder. In *Kansas v. Hendricks* (1997) the Supreme Court of the United States upheld Kansas' right to civilly commit sexual offenders for additional treatment, even after completion of their incarceration [15]. In re Crane (2000), the Supreme Court provided individual states wide latitude in deciding which clinical conditions or diagnoses would warrant an individual eligible for commitment under sexual civil commitment laws [16]. This decision no longer required the presence of a major mental illness and instead provided commitment power to states on the basis of a variety of disorders, including personality disorders. In contrast to most previous decisions concerning criteria for civil commitment, the court expanded upon what could lead to loss of liberty. The court in Crane stated,

> "In recognizing [in *Hendricks*] that [lack of control is required], we did not give to the phrase lack of control a particularly narrow or technical meaning. And we recognize that in cases where lack of control is at issue, inability to control behavior will not be demonstrable with mathematical precision. It is enough to say that there must be proof of serious difficulty in controlling behavior. And this, when viewed in light of such features of the case as the nature of the psychiatric diagnosis, and the severity of the mental abnormality itself, must be sufficient to distinguish the dangerous sexual offender whose serious mental illness, abnormality, or disorder subjects him to civil commitment from the dangerous but typical recidivist convicted in an ordinary criminal case." (p. 413)

This decision has led one scholar [17] to suggest that relying on personality disorders and paraphilia as mental disorders represents a weak link in mental health diagnoses and should not be used as a basis for commitment. Despite substantial criticism these laws have continued to expand throughout many jurisdictions in the United States.

Since the beginning of laws authorizing the civil commitment of sex offenders, 17 states have enacted legislation with the subsequent commitment of over 2500 sex offenders [18]. These laws are based upon the state's responsibility to protect its citizenry from undo harm. In committing sex offenders, it has been argued that the state's rights for safety outweigh the loss of liberty for sex offenders who are likely to commit future sexual crimes. Nonetheless, there have been substantial criticisms, beyond diagnostic issues, leveled against laws authorizing the sex offender commitment. These criticisms have been aimed at three broad aspects of sex offender laws: (i) the inaccuracy of sex offender civil commitment evaluations [19], (ii) the high cost of committing sex offenders to inpatient treatment facilities [20], and (iii) the misuse of the mental health system to commit sex offenders to indeterminate commitment in locked inpatient facilities [21].

In addressing the accuracy of sex offender evaluations, Jackson and colleagues [19] used case studies

with actuarial information to assess if 392 individuals attending a professional conference in psychology and law could accurately predict sexual violence. The results were discouraging in several ways. First, the results found that seasoned clinicians were no more accurate than graduate students at predicting which offenders would be reconvicted of a sex offense. In fact, these "experts" were "accurate in their predictions of future sexual violence approximately one-half of the time" (p. 124).

Another criticism that has gained momentum is considering the cost-benefits for civilly committing sex offenders for indefinite periods of time. The high cost of committing sex offenders has recently come under fire. Janus [20] discussed how the population of civilly committed sex offenders is growing in Minnesota and Wisconsin, which is taking away needed funding from more prevalent forms of violence, such as domestic abuse. Combined with the lack of studies demonstrating accuracy of commitment procedures, this could lead to large amounts of money being allocated to committing individuals to secure facilities who, in reality, pose little risk of sexual reoffense.

Lastly, using mental health laws to commit sex offenders has drawn additional criticisms. Appelbaum [21] paralleled the civil commitment of sex offenders to new legislation being introduced in England to confine dangerous offenders. He stated, "Identification of person's who are likely to commit violent acts in the future is a fraught process. . . . " (p. 398). His concerns could readily be applicable to sexual commitment evaluations, especially in light of recent research [18] questioning the potential issues with the accuracy of such evaluations. The problems with the commitment of sex offenders have led some legal scholars [22, 23] to question the construct of sexual civil commitment laws. Such concerns continue to fuel the controversy associated with the mandated treatment of sexual offenders beyond their prison terms.

Mandated Treatment of Insanity Acquittees

A concept in the treatment of individual found not guilty by reason of insanity (NGRI) is treating individuals in the community. Often, these individuals are released from locked inpatient hospitals and transitioned into the community [24]. However, these individuals, to minimize dangerousness and reduce mental health symptoms, are typically released with conditions. These conditions typically include those associated with probation and parole (e.g., abstinence from alcohol and drug, no firearms); however, conditions often include mandated treatment, including but not limited to medication compliance and treatment designed to decrease violence (e.g., anger management) [25].

Vitacco *et al.* recently studied the efficacy of a statewide program for conditionally released NGRI acquittees [26]. Results demonstrated a low rate of both general and violence recidivism with the largest number of NGRI acquittees being returned to the hospital for rule violations. Several other research studies [27–29] demonstrated similarly low rates of violence and aggression through the use of community-based mandated treatment. Although mandated treatment for NGRI acquittees has been implemented for a long time, recent advancements in how programs are administered and implemented have paved the way for fiscal responsibility and successful treatment.

Conclusions and the Promise of Mandated Treatment

The use of mandated treatment always presents a delicate balancing act between public safety and ensuring the rights of mentally ill individuals. That balance is made more difficult through changing conceptualizations of dangerousness and expansion of mandated treatment to sex offenders and insanity acquittees in communities. However, mandated treatment is an important tool that when used properly can decrease dangerousness and protect the public. With continued advances in neuroscience, mandated treatment may prove effective in keeping mentally ill individuals in the least restrictive environment and ultimately saving lives.

References

[1] Erickson, S.K., Ciccone, R., Schwarzkopf, S., Lamberti, J.S. & Vitacco, M.J. (2007). Legal fallacies of antipsychotic drugs, *International Journal of Law and Psychiatry* **35**, 235–246.

[2] Rosenbaum, A., Gearan, P.J. & Ondovic, C. (2002). Completion and recidivism among court- and self-referred batters in a psychoeducational group treatment program, *Journal of Aggression, Maltreatment, and Trauma* **5**, 199–220.

[3] Wells-Parker, E. (1995). Mandated treatment: lessons from research with drinking and driving offenders, *Alcohol Health and Research World* **18**, 302–306.
[4] Doren, D.M. (2002). *Evaluating sex Offenders: A Manual for Civil Commitments and Beyond*, Thousand Oaks, Sage Publications.
[5] Van Dorn, R.A., Elbogen, E.B., Redlich, A.D., Swanson, J.W., Swartz, M.S. & Mustillo, S. (2006). The relationship between mandated community treatment and perceived barriers to care in persons with severe mental illness, *International Journal of Law and Psychiatry* **29**, 495–506.
[6] Erickson, S.K., Vitacco, M.J. & Van Rybroek, G.J. (2005). Beyond overt violence: Wisconsin's progressive civil commitment statute as a marker of a new era in mental health law, *Marquette Law Review* **89**, 359–405.
[7] Lessard v. Smith, 349 F. Supp. 1078 (E.D. Wis 1972).
[8] Appelbaum, P.S. (1994). *Almost a Revolution: Mental Health Law and the Limits of Change*, Oxford University Press, New York.
[9] Stein, L.I. & Test, M.A. (1980). Alternative to mental hospital treatment: conceptual model, treatment program, and clinical evaluation, *Archives of General Psychiatry* **37**, 392–397.
[10] Olmstead v. L. C. (98–536) 527 U.S. 581 (1999). 138 F.3d 893.
[11] Swanson, J.W., Van Dorn, R.A., Monahan, J. & Swartz, M. (2006). Violence and leveraged community treatment for persons with mental disorders, *American Journal of Psychiatry* **163**, 1404–1411.
[12] Appelbaum, P.S. (2006). Violence and mental disorders: data and public policy, *American Journal of Psychiatry* **163**, 1319–1321.
[13] Monahan, J., Bonnie, R.J., Appelbaum, P.S., Hyde, P.S., Steadman, H.J. & Swartz, M.S. (2001). Mandated community treatment: beyond outpatient commitment, *Psychiatric Services* **52**, 1198–2005.
[14] Monahan, J., Swartz, M. & Bonnie, R.J. (2003). Mandated treatment in the community for people with mental disorders, *Mental Health Law* **22**, 28–38.
[15] Kansas v. Hendricks, 521 U.S. 346 (1997).
[16] Kansas v. Crane, 534 U.S. 407 (2002).
[17] Zander, T. (2005). Civil commitment without psychosis: the law's reliance on the weakest links in psychodiagnosis, *Journal of Sex Offender Civil Commitment: Science and the Law* **1**, 17–82.
[18] Miller, H.A., Amenta, A.E. & Conroy, M.A. (2005). Sexually violent predator evaluations: empirical evidence, strategies for professionals, and research directions, *Law and Human Behavior* **29**, 29–54.
[19] Jackson, R.L., Rogers, R. & Shuman, D.W. (2005). The adequacy and accuracy of sexually violent predator evaluations: contextualized risk assessment in clinical practice, *International Journal of Forensic Mental Health* **3**, 115–129.
[20] Janus, E.W. (2004). Closing Pandora's Box: sexual predators and the politics of sexual violence, *Seton Hall Law Review* **34**, 1233–1253.
[21] Appelbaum, P.S. (2005). Dangerous and severe personality disorders: England's experiment in using psychiatry for public protection, *Psychiatric Services* **56**, 397–399.
[22] Morse, S.J. (1998). Fear of danger, flight from culpability, *Psychology, Public Policy, and Law* **4**, 250–267.
[23] Schopp, R.F. (1998). Civil commitment and sexual predators: competence and condemnation, *Psychology, Public Policy, and Law* **4**, 323–376.
[24] Parker, G.F. (2002). Conditional release and mandated outpatient treatment, *Psychiatric Services* **53**, 483–484.
[25] Junginger, J., Claypoole, K., Laygo, R. & Crisanti, A. (2006). Effects of serious mental illness and substance abuse on criminal offenses, *Psychiatric Services* **57**, 879–882.
[26] Vitacco, M.J., Van Rybroek, G.J., Erickson, S.K., Rogstad, J., Tripp, A., Harris, L. & Miller, R. (2008). Developing services for insanity acquittees released into the community: Maximizing success and minimizing recidivism, *Psychological Services* **5**, 118–125.
[27] Bloom, J.D., Williams, M.H. & Bigelow, D.A. (2001). Monitored conditional release of persons found not guilty by reason of insanity, *American Journal of Psychiatry* **148**, 444–448.
[28] Hartwell, S. & Orr, K. (1999). The Massachusetts forensic transition program for mentally ill offenders re-entering the community, *Psychiatric Services* **50**, 1220–1222.
[29] Monson, C.M., Gunnin, D.D., Fogel, M.H. & Kyle, L. (2001). Stopping or slowing the revolving door: factors related to NGRI acquittees' maintenance of a conditional release, *Law and Human Behavior* **25**, 257–267.

MICHAEL J. VITACCO AND STEVEN K. ERICKSON

Treatment, Right to: Mental Health

Right to Treatment

The right to treatment and its antithesis the right to refuse treatment are cornerstones of therapeutic jurisprudence (TJ). As noted by Perlin, legal decisions regarding treatment rights should be considered from a TJ perspective because they maximize constitutional protections for the mentally disabled [1].

According to David Wexler, TJ is the focus and examination of law on human and emotional factors and transforming the legal system toward a therapeutic benefit [2]. Thus, instead of law for law's sake, TJ places the therapeutic goal ahead of all others, including efficiency, *starie decisis*, and legal formalism. In considering the right to treatment within the framework of TJ, there are several landmark cases harmonious with this point of view.

In reviewing these legal precedents, no case is as important for the right to treatment as *Wyatt v. Stickney* [3]. The court in the Wyatt case was unwavering in declaring that patients "unquestionably have a constitutional right to receive such individual treatment as it will give each of them a realistic opportunity to be cured or to improve his or her mental condition". This decision had a dual impact: on one level, it required the state of Alabama to provide treatment beyond the historical asylum model that often led to lifelong institutionalized care; at a broader level, it hastened the deinstitutionalization of thousands of mentally ill individuals, returning them back to the community. Since the right to treatment announced in *Wyatt* amounted to a judicial mandate to massively increase fiscal spending on an already overburdened public mental health system, the stage was set to end the decades long monolithic asylum model.

Unfortunately, for mental health clinicians, later judicial decisions further muddied the water with mixed rulings which obfuscated long held principles asserting the state's right to care for disabled and mentally ill persons. Much of the confusion was borne by the addition of numerous constitutional protections by the courts that dramatically changed the procedures used to provide institutionalized care. One such example is the decision in *O'Connor v. Donaldson* [4], where the Supreme Court tackled the issue of liberty for the mentally ill. In brief, the uncontested facts were that Mr Donaldson was delusional and believed he was being poisoned. Similar to many persons with severe mental illnesses, he possessed little insight into his illness and denied any need for mental health treatment. However, there were no evidence of manifest dangerousness and Mr Donaldson appeared to be capable of residing outside the state institution with little danger to self or others. The Court agreed with Mr Donaldson, which resulted in a simple and straightforward ruling: The key paragraph in the decision reads: "A finding of 'mental illness' alone cannot justify a State's locking a person up against his will and keeping him indefinitely in simple custodial confinement" and "In short, a state cannot constitutionally confine without more a nondangerous individual who is capable of safely surviving in freedom by himself or with the help of willing and responsible family members or friends." The *Donaldson* decision struck down many involuntary civil commitment laws in favor of expanded procedural protections transposing an informal process to one with most of the rigors of the adversarial process. Nonetheless, this ruling has been roundly criticized by mental health scholars [5, 6] on the bases that the O'Connor decision does not (i) provide adequate guidance regarding treatment and dangerousness, (ii) lacks empirical support, and (iii) provides little to no right for treatment for nondangerous individuals.

Despite the obvious ideological shift of the right to treatment movement from state paternalism to an emphasis on individual autonomy, this movement led to other notable outcomes. Indeed, right to treatment laws have indelibility influenced the public mental health system by ensuring that patients receive less coercive and effective treatment. The benefits often noted include improved quality of care and recognition of valuable constitutional rights among a frequently disfranchised group. However, frequent criticisms of this modern shift observed include increased risks of suicide, substance abuse, and involvement with the criminal justice system. Such issues highlight the need for effective treatment of the mentally ill, even in the absence of manifest dangerousness [7]. Within this framework, this article is designed to discuss issues and complications surrounding the right to treatment and how to navigate right to treatment from within the broader context of TJ. Once we discuss the issues related to the right to treatment, we further focus on three specific populations: adolescents, prisoners, and not guilty by reason of insanity (NGRI) acquittees. These populations have been instrumental in fueling discussions on treatment rights within medical and mental health contexts.

Navigating the Right to Treatment: Setting the Stage

The literature is replete with examples informing clinicians how to deal with right to treatment issues.

In correctional settings, there has been recent case law dictating that a constitutional right to treatment exists [8]. However, in other settings all practitioners too frequently left to their own judgment to navigate how to ensure that rights to treatment are upheld. Of course, difficult issues are prevalent in those decisions. For instance, the right to treatment does not denote which type of treatment patients are legally entitled. For example, what a therapist should do if a patient requests treatment that is less effective than a more empirically based one, yet the client is more comfortable with? This scenario is played out across the county often leaving clinicians to discuss the pros and cons of less effective treatment.

In the case of *Osheroff v Chestnut Lodge* [9], a physician sued Chestnut Lodge in Washington, DC for failing to provide appropriate treatment for his depression as he experienced intense despair during his entire seven-month stay. Instead of medication they offered one to one therapy. During the seven months, Dr Osheroff was declared incompetent, lost his medical practice, and his wife divorced him and took custody of their children. After seven months, he was transferred to another hospital, placed on antidepressant medications, and never needed to be rehospitalized for mental health symptoms. He eventually remarried and continued practicing medicine in New York. The case was eventually settled; however, its influence remains as it alerted mental health clinicians that they have the responsibility to ensure that a selected treatment, even if wanted by their patients, is not provided *in lieu* of treatment that is likely more effective (e.g., medications for depression) and to make sure that the record reflects they presented the client with several known benefits and liabilities of all known treatments.

In considering the right to treatment, Perlin [10] discussed considerations related to the availability of appropriate resources to ensure proper treatment decisions are made. As discussed above, clinicians have an ethical and potentially legal obligation to inform client of alternative, more effective treatments. In extrapolating TJ to issues of treatment in clinical practice, clinicians have a wide variety of issues to consider. Such issues include autonomy, decision-making capabilities, and providing an environment where informed treatment decisions can be made by clients who know their legal rights regarding treatment.

Juvenile Offenders and the Right to Treatment

Adolescents are no longer considered as parental properties incapable of decision making regarding their own treatment; instead, they are allowed autonomy regarding treatment decision. A recent public policy report highlighted the ever increasing practice of allowing adolescents to be responsible for many treatment-related decisions [11]. In many jurisdictions, adolescents are allowed to enter therapy without parental consent. This expansion of rights for adolescents is a recent development showing improved understanding of the capabilities and insight of adolescents regarding their best interests. This understanding is evidenced by changes in philosophy concerning the treatment of adolescents, including those incarcerated.

The right to treatment in adolescents in correctional settings is salient for another reason. If we view the right to treatment from a TJ perspective and we acknowledge that treatment is designed to improve quality of life, then it is a logical step to guarantee incarcerated adolescents with mental health issues treatment aimed at decreasing recidivism and promoting community tenure. Although these theoretical assumptions are laudable, achieving these goals remains elusive in practice. However, as we discuss in the forthcoming paragraph, a balance between incarceration and treatment can be obtained when using cost-effective, empirically based strategies.

In a study of young offenders, Shelton [12] discovered that only 23% of any youth diagnosed with a mental disorder residing in a juvenile detention facility received treatment. Such results are discouraging in light of recent evidence demonstrating that therapeutic interventions are cost-effective over the long term and serve to reduce violent and criminal recidivism [13]. These statistics underscore the need for effective treatment with adolescent offenders and provide a logical foundation for goals of treatment – simply, to minimize the likelihood of future criminal behavior while allowing the greatest opportunity for an incarcerated free life.

Nonetheless, programs for young offenders will continue to face challenges in the future. From a policy perspective, "get tough" programs for adolescent offenders have recently gained partial support due to several highly publicized horrific crimes of violence committed by some young offenders.

Thus, incorporating new treatment approaches often entails overcoming public skepticism, which favors traditional juvenile justice interventions that mirror approaches within the adult corrections system. Practitioners must find ways to integrate right to treatment within this complex system. One such way is to find empirically supported treatment that demonstrates recidivism reduction and long-term cost savings [13, 14].

Adults and the Right to Treatment

Like adolescents, challenges abound in the treatment of adult offenders within the criminal justice system are underscored by recent changes in policy. One such example is the formation of "super max" prisons where the focus is solely on management of offenders exhibiting behavioral problems in other correctional facilities or whose crimes are considered to be so horrendous that rehabilitation is not considered as an appropriate option. Such prisons primarily rely on isolation with the goal of protecting society and other prisoners from a selected group of intractable offenders. The creation of such facilities, which are vastly expensive, is borne from the popular notions that such offenders are untreatable, and thus, incapacitation is the only legitimate aim for this subgroup. Yet, the empirical evidence is mixed as to find whether such predictions are true.

Moreover, in regards to mentally ill offenders, court precedents consistently evince a right to treatment. Such rights are buttressed by significant case law decisions. A prime example is *Estelle v. Gamble* [15] where the US Supreme Court ruled that the Eighth Amendment of the Constitution guaranteed prisoners the right to medical care. Likewise, in ***Vitek v. Jones*** [16] 445 U.S. 480 (1980), the Supreme Court held that a mentally ill inmate was entitled to procedural protections given his impending transfer to a psychiatric facility. The court noted that the enduring stigmatizing effect inmates suffer when they are labeled and segregated based on their mental illnesses.

Not surprisingly, state institutions have been mandated to provide a basic level of services to individuals under their care. In ***Youngberg v. Romeo*** [17] the Supreme Court ruled that a mentally disabled child residing in a state institution in Pennsylvania had a right to several basic necessities including safe conditions within the institution, freedom from restraint, and adequate living conditions.

The right to treatment in insanity acquittees was underscored in *Rouse v. Cameron* [18]. Charles Rouse was committed to St Elizabeth's Hospital as NGRI on weapon charges, for which he was held indefinitely and significantly longer than if he was incarcerated in prison. The Circuit Court of Appeals for the District of Columbia reiterated that commitment is not used *in lieu* of incarceration and that individuals committed under NGRI standards have a right to treatment. As Judge Bazelton alluded to in the Rouse case, mental health practitioners must treat individuals, improve their functioning, and return them to the community when they are no longer ill. This perspective is wholly consistent with TJ, as it affirms the desire to place the person in the least restrictive environment and uses the law as a tool to do so.

Current insanity laws and requirements for continued commitment reflect this model of treatment. In *Foucha v. Louisiana* [19] the US Supreme Court held that continued confinement of Foucha would be unconstitutional as his mental illness was in remission and continued commitment required both the diagnosis of mental illness and dangerousness.

Foucha represented a step forward in the conceptualization of commitment and unlike other standards, it was clearly defined, even if the concepts of mental illness and dangerousness are not.

Conclusions

The right to treatment is a relatively new movement in mental health law. Although it is fairly new, the movement has gained momentum through court cases and its consistency with the concept of TJ. By using treatment to secure earlier release from inpatient facilities and improve overall functioning, the foundations for right to treatment are rooted in a person's rights of autonomy, which underlie the constitution. Yet this right to treatment lies in tandem with the right to effective treatment and release if treatment proves ineffective. The right to treatment presupposes that the individual is able to be treated and that practitioners have some treatment (e.g., medications or therapy) that would benefit the patient. Ultimately, the right to treatment would be most aptly construed as a right to effective treatment.

References

[1] Perlin, M.L. (1995). Therapeutic jurisprudence and the civil rights of the institutionalized mentally disabled persons: hopeless oxymoron or path to redemption, *Psychology, Public Policy, and Law* **1**, 80–119.

[2] Wexler D.B. (2004). Therapeutic jurisprudence: it's not just for problem-solving courts and calendars anymore, in *Future Trends in State Courts*, National Center for State Courts, Vol. 87.

[3] Wyatt v. Stickney, 344 F.Supp. 387 (M.D. Ala. 1972).

[4] Connor v. Donaldson, 422 U.S. 563 (1975).

[5] Bernard, J.L. (1977). The significance for psychology of O'Connor v. Donaldson, *American Psychologist* **32**, 1085–1088.

[6] Siegel, R.A. (1978). The significance for psychology of O'Connor v. Donaldson: a reply to Bernard, *American Psychologist* **33**, 858–861.

[7] Erickson, S.K., Vitacco, M.J. & Van Rybroek, G. (2005). Beyond overt violence: Wisconsin's progressive civil commitment statute as a marker of a new era in mental health law, *Marquette Law Review* **89**, 359–405.

[8] Cohen, F. & Gerbasi, J.B. (2005). Legal issues regarding the provision of mental health care in correctional settings, in *Handbook of Correctional Mental Health*, C. Scott & J. Gerbasi, eds, American Psychiatric Publishing, Washington, D.C.

[9] Osheroff v. Chestnut Lodge, 490 A.2d 720 MD (1985).

[10] Perlin, M.L. (2005). 'May you stay forever young': Robert Sadoff and the history of mental disability law, *Journal of the American Academy of Psychiatry and Law* **33**, 236–244.

[11] Reyna, V.F. & Farely, F. (2006). Risk and rationality in adolescent decision making: implications for theory, practice, and public policy, *Association for Psychological Science* **7**, 1–44.

[12] Shelton, D. (2005). Patterns of treatment services and costs for young offenders with mental disorders, *Journal of Child and Adolescent Psychiatric Nursing* **18**, 103–112.

[13] Caldwell, M.F., Vitacco, M.J. & Van Rybroek, G.J. (2006). Are violent delinquents worth treating? A cost-benefit analysis, *Journal of Research in Crime and Delinquency* **43**, 148–168.

[14] Timmons-Mitchell, J., Bender, M.B., Kishna, M.A. & Mitchell, C.C. (2006). An independent effectiveness trial of multisystemic therapy with juvenile justice youth, *Journal of Child and Adolescent Psychology* **35**, 227–236.

[15] Estelle v. Gamble, 429 U.S. 97 (1976).

[16] Vitek v. Jones, 445 U.S. 480 (1980).

[17] Youngberg v. Romeo, 457 U.S. 307 (1982).

[18] Rouse v. Cameron, 373 F.2d 451 (D.C. 1966).

[19] Foucha v. Louisiana, 499 U.S. 946 (1994).

MICHAEL J. VITACCO AND STEVEN K. ERICKSON

Treatment, Right to Refuse: Mental Health

Consequently, courts have decreed that mentally competent adults may refuse treatment even if such treatment is considered in their best interests. While many constitutional scholars point out that the word privacy is never mentioned in the federal constitution others maintain that privacy, and hence, autonomy is imbued within American legal traditions. As early as 1891, the United States Supreme Court held, "No right is held more sacred, or is more carefully guarded, by the common law, than the right of every individual to the possession and control of his or her own person, free from all restraint or interference of others, unless by clear and unquestioned authority of law [1]." This entry is going to discuss issues, with a specific emphasis on court cases, related to the right to refuse treatment.

The right to refuse treatment enjoys wide international support concurrent with the emergence of the modern privacy rights during the 1960s. The Canadian Supreme Court ruled in 2006 that the "right to refuse unwanted medical treatment is fundamental to a person's dignity and autonomy." This statement provides the ethical backdrop many cite in support of the right to refuse treatment. As noted by psychiatrist and Harvard law professor Stone [2] the right to refuse treatment has assisted in the development of informed consent documents, which have gained general acceptance in the mental health field and serve to inform patients of the benefits and risk of treatments. While mental health practitioners often view informed consent as a means of mitigating liability, their existence is borne from the notion that patients should understand risks entailed by medical procedures being performed on their body.

The increasing use of the right to refuse treatment has been roundly criticized by many groups who argue that the doctrine directly harms patients who are often unable to accurately assess the benefits of treatment because of the cognitive impairments inherent with their mental illnesses. Additionally, some argue that due process invariably requires lengthy procedural matters before courts that are ill equipped in discerning the potential harm patients risk when allowed to forego treatment. Such harm includes an increased risk of suicide, incarceration, homelessness,

victimization, and rehospitalization [3]. To describe cases where patients waited for treatment, a popular Wisconsin psychiatrist coined the phrase "dying with their rights on" to describe the well-intentioned, but problematic quagmire inherent when mentally compromised adults refuse treatment.

The right to refuse treatment has also permeated aspects of criminal procedure leading to incompetent defendants being allowed to refuse treatment to restore their competency [4]. While criminal defendants enjoy this right under a different rubric of the privacy doctrine, the courts have nonetheless recognized a limited right of defendants to refuse treatment when there does not exist a compelling state interest for involuntary treatment. Although complications are numerous when trying to balance the best interests of patients with their right to refuse treatment, practitioners must be cognizant of how the right to refuse treatment can influence the manner in which they practice. To improve our understanding of the right to refuse treatment, we will evaluate court cases and discuss their influence on clinical practice.

Influential Court Cases and the Right to Refuse Treatment

There are wide ranging cases that have been decided at the state and federal level that provide guidance on the right to refuse treatment. A close review of these cases demonstrates wide jurisdictional differences in the refusal of treatment and how such refusal can be overcome by the state. In the following section, we will explore many of the influential cases from the last 50 years and their application to the right to refuse treatment.

In the "Application of the President and Directors of Georgetown College, Inc." [5] the rights of individuals and the medical establishment conflicted and court guidance was sought regarding the potential death of a patient who refused medical treatment. In this instance the discussion of the right to die after a suicide attempt was at the forefront. In this case, the court ruled, "where attempted suicide is illegal by the common law or statute, a person may not be allowed to refuse necessary medical assistance when death is likely to ensure without it." Since suicide has long been considered illegal in most jurisdictions, courts will refuse to recognize a right to an essentially illegal act. However, in cases where death is imminent because of disease, such as cancer, the decision to end life may ultimately rest with the circumstances of the case, although the recent trend has been a broadening recognition of the "right to die" (see generally *Cruzan v. Director* [6]).

The ability to refuse medical treatment has also been extended to mental health patients and incarcerated individuals. The case of *Rennie v. Klein* [7, 8] was a landmark decision in mental health law and began setting parameters for when an individual could refuse mental health treatment. In this case, Mr Rennie was involuntarily hospitalized because of schizophrenia and decompensation after his brother's death; however, he refused to take his prescribed antipsychotic medication. The hospital, believing the treatment was in his best interests and Mr Rennie incompetent to decide the matter, applied in the court for permission to involuntarily treat Mr Rennie. The United States District Court of New Jersey ruled that (i) Mr Rennie could not be forced psychotropic medications in the absence of an emergency situation and (ii) medications might not be forced without offering the patient due process. The importance of this decision is that it required a finding of dangerousness and entitled patients to basic due process rights prior to involuntary administration of treatment.

The Rennie decision set the stage for new refusal rights for mentally ill individuals. Subsequent decisions continued to expand on those newly acquired rights. Notably, in *Rogers v. Orkin* [9], the Massachusetts Superior Court ruled that even individuals deemed appropriate for involuntary commitment had a right to refuse medications. The court in Rogers differentiated the need for commitment versus a finding of incompetence to refuse medications. The bottom line was that a justification for committing an individual to a hospital could not be substituted for the due process necessary to mandate medication. In the absence of due process, the continued right to privacy regarding control over one's body was viewed as superseding the right of the state to medicate without further due process.

Another Massachusetts case *Rogers v. Commissioner* [10] established that a judicial hearing is required for overriding an individual's unwillingness to take medications and competent patients have the right to refuse medication. Such a model is found in a recent decision in the state of Alaska [11]. In *Myers v. Alaska Psychiatric Institute* [11], the Supreme Court of Alaska ruled, "In the absence of emergency, a

court may not authorize the state to administer psychotropic drugs to a nonconsenting medical patient unless the court determines that the medication is of the best interests of the patient and no less intrusive alternative means is available (p. 239)." Formal judicial hearings have become the cornerstone of right to refuse treatment in addition to informed consent.

A patient's consent to treatment has long been considered fundamental to American jurisprudence [12]. Nowhere was this fundamental right more evident than in the case of *Zinermon v. Burch* [13]. In 1981, Mr Darrell Burch was found wandering incoherently and bloodied along a Florida highway and was brought to a private hospital. Once there, Burch was clearly delusional (e.g., reported being in heaven) and appeared confused. He agreed to be admitted to the hospital and signed consent forms to be admitted. Eventually he was transferred to Florida State Hospital (again, he signed informed consent to be admitted) and was successfully discharged after five months at Florida State Hospital.

Three years later, Burch sued his doctor, Marlus Zinermon and the state of Florida alleging civil rights violations that he was subjected to during involuntary confinement, which included treatment. The Supreme Court sided with Burch and ruled that a violation of rights occurs when an incompetent individual is allowed to voluntarily enter a mental health hospital because the person is incapable of providing consent. In these cases, the hospital bears the responsibility of applying for an order of commitment in a court of law within a certain designated period of time, usually 72 h.

The decision in *Burch* raised the bar for hospitals higher in ensuring that an individual they suspect as incompetent be provided a judicial hearing on the matter to protect patients' rights. As noted by the court in an earlier decision [14], involuntary treatment for the mentally ill is unconstitutional if the mentally ill individual can survive safely with the help of friends or family.

As evident by the above cases, the issue of consent and the right to refuse treatment have been clear with individuals with mental illness. Further, the court has also ruled on prisoners. The case of *Washington v. Harper* [15] exemplifies the complex issues involved in right to refuse treatment cases. Harper was a mentally ill inmate whose violence was reduced through compliance with antipsychotic medications. Nonetheless, Harper refused his medications and sued the prison. The United States Supreme Court ruled that the decision to medicate Harper should be left to medical professionals and that in doing so Harper's interests in liberty must be weighed and balanced with the interests of the government to provide a safe and orderly penal institution. In this case, the Supreme Court ruled that an administrative act was sufficient and no formal judicial hearing was constitutionally required to ensure that Harper's rights were protected.

Pretrial issues related to mental state are also influenced by right to refuse treatment decisions. A question that may be posed by a defendant is whether the jury has a right to view the defendant in an unmedicated state when a mental health issue might be the crux of the case. The case of *Commonwealth v. Louraine* [16] grappled with this issue. Mr Louraine was on trial for first-degree murder after stabbing a man 21 times in a halfway house for individuals with mental illness. In this case, the Supreme Court of Massachusetts ruled that the trier of fact is entitled to see the defendant in the unmedicated state, especially as it relates to mental health issues (e.g., mental state at the time of the offense). Allowing the right to refuse treatment to apply to pretrial defendants simply means that the defendant's demeanor is as important as verbal testimony when considering certain mental health issues. Certainly, viewing a defendant in a nonmedicated state when the person is more disorganized and symptomatic may improve the likelihood of a jury accepting a mental health defense.

Right to Refuse Treatment and Therapeutic Jurisprudence

Like its counterpart, the right to treatment, the right to refuse treatment has been viewed from the perspective of therapeutic jurisprudence. As Winick [17] discussed, refusing treatment is controversial, yet remains sanctioned by the law of the United States, even in life threatening situations. Viewing the right to refuse treatment through the lens of therapeutic jurisprudence maximizes constitutional rights to individuals with mental disorders and those experiencing significant medical issues [18]. Relatedly, competent individuals have wide latitude to choose

tailored treatment or refuse treatment altogether and consider a wide range of factors including religious and cultural beliefs. Moreover, as noted in the landmark cases of *Rogers* and *Rennie*, the presence of a mental illness did not equate to incompetence and required separate due process; no longer were incompetence and mental illness equated in the eyes of the court.

Conclusions

In a broad analysis, practitioners and hospitals must consider several factors regarding the right of an individual to refuse treatment. In our conclusion, we provide several key considerations regarding the right to refuse treatment relevant to both practitioners and legal advocates.

1. The right to refuse treatment is fundamentally linked to privacy.
2. The fundamental right to refuse treatment is not absolute and can be overturned if the individual wishing to refuse treatment is either incompetent to refuse treatment or considered dangerous.
3. Be aware of jurisdictional issues regarding the right to refuse treatment. Often, laws vary widely from jurisdiction to jurisdiction concerning how issues related to right to refuse treatment are handled.
4. Informed consent is a critical component of both the right to treatment and the right to refuse treatment; however, informed consent is only valid when completed by a competent individual.
5. The right to refuse treatment has been considered through a lens of therapeutic jurisprudence where the constitutional rights of the disabled are afforded maximal protections.

Considering the right to refuse treatment with these in mind will improve quality of care and allow for the rights of the ill to be cared for while remaining mindful of the responsibilities of practitioners.

References

[1] Union Pacific Rail Road v. Botsford, 141 U.S. 250 (1891).
[2] Stone, A.A. (1981). The right to refuse treatment: why psychiatrists should and can make it work, *Archives of General Psychiatry* **38**, 358–362.
[3] Guthiel, T.G., Simon, R.I. & Hillard, J.T. (2005). The wrong handle: flawed fixes of medicolegal problems in psychiatry and the law, *Journal of the American Academy of Psychiatry and Law* **33**, 432–436.
[4] Bassman, R. (2005). Mental illness and the freedom to refuse treatment: privilege or right, *Professional Psychology: Research and Practice* **36**, 488–497.
[5] 118 United States App. D.C. 80, 331 F.2d 1000 (1964).
[6] Cruzan v. Director, 1990, Missouri Department of Health, 497 U.S. 261 (1990).
[7] Rennie v. Klein, 462 F.Supp. 1131 (1978).
[8] Rennie v. Klein, 720 F.2d 266 (3d Cir. 1983).
[9] Rogers v. Orkin, 478 F. Supp. 1342 (D. Mass. 1979).
[10] Rogers v. Commissioner, 390 Mass. 489, 458 N.E.2d 308 (1983).
[11] Myers v. Alaska Psychiatric Institute, 138 P.3d 238 (2006).
[12] Schloendorff v. Society of New York Hospital, 211 N.Y. 125, 105 N.E. 92 (1914).
[13] Zinermon v. Burch, 494 U.S. 113 (1990).
[14] O'Connor v. Donaldson, 422 U.S. 563 (1975).
[15] Washington v. Harper, 494 U.S. 210, 110 S.Ct. 1028 (1990).
[16] Commonwealth v. Louraine, 453 N.E. 2D, 437 (Mass. 1983).
[17] Winick, B.J. (2000). Right to refuse treatment, *Encyclopedia of Psychology* **7**, 104–106.
[18] Winick, B.J. (1997). A therapeutic jurisprudence analysis of the right to refuse mental health treatment, in *Law and Public Policy*, American Psychological Association, Washington, D.C., pp. 327–344.

MICHAEL J. VITACCO AND STEVEN K. ERICKSON

Trial: Capacity to Stand *see* Capacity to Stand Trial

Truth Serum *see* Deception: Truth Serum



Ultimate Issue Evidence by Experts

A recognized axiom of the common law is that the finder of the facts is the sole authority on what happened during events that gave rise to litigation. Witnesses were traditionally not permitted to speculate on the "ultimate issue" of guilt or innocence, or on whether the defendant was negligent or careful, or whether necessary facts had really occurred. It was also a general rule that an expert testifying as (*see* **Expert Opinion: United States**) to cause and effect from an analysis had to state his conclusion as a matter of expert opinion, and not as "fact". Cases later extended the rule to exclude opinions which embraced the issue that was to be the final resolution of the events on which the litigation was based. The rule came to be known as the *ultimate issue doctrine* [1].

Over the years the rule's prohibition was gradually relaxed because many expert findings unavoidably border on what might be considered the ultimate issue questions. Legal scholars, also, had strongly criticized the barring of experts' ultimate issue opinions. Thus, in recent times, the prohibition all but disappeared and experts are, today, permitted to state opinions based on most though not all facts that relate to the ultimate issue to be resolved.

In the US federal courts, the common law prohibition was abolished when the Federal Rules of Evidence (FRE) (*see* **Federal Rule of Evidence 702**) were adopted in 1975. Rule 704 specifically provides that expert opinions or inferences are not objectionable solely because they embrace the ultimate issue to be decided. However, the rule retains a prohibition against certain behavioral expert witnesses' opinions.[a]

While a clear modern trend of state courts has been to follow the federal model and abolish the ultimate issue rule, that movement is by no means unanimous. Some courts have specifically rejected the relaxation of the prohibition on ultimate issue opinions. Thus, one court stated:

> We are not prepared to reject the ultimate issue prohibition . . . in a criminal case . . . where life or liberty often turns upon inferences raised by circumstantial evidence The process of resolving conflicting inferences, affected as it is by the credibility of the witnesses who supply such evidence, is the historical function of a jury. . . . We are unwilling to entrust that function to experts in the witness box [2].

End Notes

[a.] FRE 704 on opinions relating to the ultimate issue to be decided provides the following:

(i) Except as provided in subdivision (ii), testimony in the form of an opinion or inference otherwise admissible is not objectionable because it embraces the ultimate issue to be decided by the court.

(ii) No expert witness testifying with respect to the mental state or condition of a defendant in a criminal case may state an opinion or inference as to whether the defendant did or did not have the mental state or condition constituting an element

of the crime charge or of a defense thereto. Such ultimate issues are matters for the trier of fact alone.

Subdivision (ii) was added by Congress in 1984, as a result of a celebrated trial wherein multiple psychiatric experts for the government testified that the defendant was sane, and equal number testified for the defense, asserting the defendant's insanity. A behavioral science expert can, of course, testify to a diagnosis with regard to defendant's mental state and motivation, as long as these findings are expressed in clinical terms, rather than as ultimate legal conclusions.

References

[1] Moenssens, A.A., Henderson, C.E. & Portwood, S.G. (2007). *Scientific Evidence in Civil and Criminal Cases*, 5th Edition, Foundation Press, pp. 82–84.
[2] Bond v. Commonwealth, 226 Va. 534, 311 S.E.2d 769 (1984).

ANDRE MOENSSENS

Electron Microscopy *see* Expert Opinion: United States

Urine: Drug Testing *see* Drug Testing: Urine

Use of Knowledge-Based Systems in Forensic Science

Introduction

A knowledge-based system (KBS) is a computer system that is programed to imitate human problem solving by means of artificial intelligence and reference to a database of knowledge on a particular subject [1]. KBSs exist in various guises and specialties in forensic science, but they generally have one common goal – consistency. Many forensic scientists concern themselves with the idea of consistency. This means that if a piece of evidence is measured or evaluated using the same technique, then two scientists should independently arrive at the same result. Inherent in the goal of consistency is the idea that the user of a KBS can draw on and add to the knowledge and information contained within a system. For example, programs such as Computerised Assistance for Glass Evidence (CAGE) use information from previous scientific studies on transfer and persistence of glass, and allow users to see the probability distributions used by other analysts in similar glass cases.

KBSs have become increasingly important to forensic scientists. They allow new scientists to use the results of previous research and casework, and, at the same time, add to the repository of information contained therein. Such systems can reduce errors and improve the quality of casework. At the heart of many KBSs lies a graphical model or Bayesian network [2]. Bayes nets allow scientists to formalize the conditional dependencies that exist between variables in a complex process without necessarily concentrating on the statistical detail that may be defined or refined at "run time". Such systems will become increasingly more commonplace in the forensic community in the future.

The remainder of this article attempts to briefly review the KBSs that are in active use in the forensic community. The list of systems is by no means exhaustive.

The Statistical Evaluation of Forensic Glass Evidence

The evidential value of forensic glass evidence has been heavily statistical since the early 1970s [3, 4]. One of the first problems that scientists in this field encountered was calculating the required quantities from Student's t distribution. Previously, this had been done with published tables, an approach that is slow and error prone. Because they cannot be exhaustive, tables are also subject to rounding and interpolation errors. To counter this problem, and to aid in the process of grouping recovered glass, Evett and Lambert [5, 6] designed the program RUNG [7]. This program evolved with the input of John

Buckleton and Richard Pinchin into a commercial product called Fragment Data System (FDS). This program incorporated the work of Evett, Lambert, Buckleton, and others. RUNG was not a KBS as such because it incorporated no prior knowledge from either the literature or from other casework. However, RUNG is regarded as an important precursor to later work. FDS took a small step toward being a KBS in that it allowed the user to calculate part of a likelihood ratio by using a refractive index (RI) frequency table. CAGE, produced by Richard Pinchin and John Buckleton, was the first real KBS for the evaluation of likelihood ratios with glass evidence. CAGE allowed nonspecialists to evaluate the likelihood ratios espoused in the revolutionary work of Evett and Buckleton [8, 9]. CAGE was revised and altered by Hicks and Wyss and rereleased as CAGE 2000. The aim of CAGE 2000 was to develop user documentation and make the program accessible from the internet. The purpose of the latter was to include as much expertise as possible from the global community into the system [10].

Curran, in his PhD work, produced two programs for glass evaluation. The first was Statistical Analysis of Glass (StAG), which implemented the work of Walsh *et al.* [11] on a continuous approach for glass evidence, and the work of Triggs *et al.* [12] on divisive grouping. The second program produced by Curran is called TRANSFER [13], and it allows the expert to input case details from scene examination in order to estimate the transfer probabilities. TRANSFER (also known as TFER) was one of the early users of a graphical model or Bayes Net, and the first program that allowed the user to alter the model graphically, albeit in a limited way.

RUNG, FDS, and StAG are systems that allow the comparison of RI measurements, and, in the case of the latter two, calculation of the continuous part of the likelihood ratio. TRANSFER is a KBS explicitly for the evaluation of transfer probabilities that are part of the likelihood ratio (LR) calculations for glass. CAGE and CAGE2000, on the other hand, attempt to encompass all of these tasks as a true KBS should.

None of the systems deal with the evaluation of elemental concentration information arising from glass evidence. Curran *et al.* [14], Aitken and Lucy [15], and Aitken *et al.* [16] present likelihood ratio calculations for multivariate evidence that rely on the user having a database of measurements of elemental concentrations from other samples of glass not related to the case at hand.

The Evaluation of DNA Evidence

The evaluation of forensic DNA evidence is a complex, multistage task and can be aided by KBSs at each stage. KBSs exist for the postamplification stages of DNA evidence handling. The first stage is automated allele designation. Modern DNA typing equipment produces a "trace" corresponding to each of the three to five marker dyes that are used to tag different groups of loci. This set of traces is called an electropherogram (or epg), and information about the alleles present in the epg is given by "peaks" on the traces measured in the relative fluorescence units (rfus). The height of the peaks corresponds roughly to the basic amount of DNA present. The location of the peaks along the horizontal axis (size in base pairs) is matched to an allelic ladder, and it is here that the allele is designated. Software that automates this designation process must take into account knowledge of the multiplex in use, the performance of the machine in the particular run in which that evidence was measured, and information about the stochastic behavior of the systems. On the basis of this information, a KBS makes allele designations and usually allows the expert to override those designations. The next stage of the process depends on the circumstances of the case. It may be that no suspect exists, and matching profiles are searched for in a database of known offenders. Similarly, one may look for "matching" relatives of the true offenders. Alternatively, if there are suspect profiles, then an assessment of the relative strength of the evidence, given the purported facts of the case, may be carried out. KBSs can aid in all of these tasks.

The Forensic Science Service (FSS), driven by Ian Evett, Richard Pinchin, John Buckleton, and Peter Gill were early proponents of producing specialist software for the interpretation of DNA evidence. The software produced by the FSS covers a raft of tasks. First, for automated allelic designation, there is STRess and it's predecessor STRipe. Familial provides computer-aided search of DNA databases for profiles of potentially related individuals. Kwikmix I, II and III (Curran, Healy, and Young) provide implementations of traditional nonquantitative DNA mixture interpretation [17, 18]. Kwikmix III may be used

in conjunction with Pendulum (Buckleton, Pinchin, Curran, Healy, and Young) [19, 20] (now known as i-Stream) to provide a computer aided implementation of the binary approach to DNA mixture interpretation. In the binary approach certain allelic combinations may be ruled as infeasible given the quantitative information available from modern genotyping systems. The FSS-i^3 suite of tools (i-STRess, i-STReam and i-ntegrity) are heavily validated and in current use in the United Kingdom and around the world. Kwikmix III is validated and in internal use in the FSS. LoComatioN (Curran and Healy) is a system that allows the nonexpert to statistically evaluate Low Copy Number (LCN, also called low template DNA) evidence, using an extension to the methods of Gill and Buckleton [21, 22]. The program FINEX (also known as HyperDue) written by Robert Cowell [23] allows the evaluation of likelihood ratios for complex pedigree problems. Although heavily validated, LoComatioN [24] and HyperDue are not currently used in casework.

There are also several programs that are not produced by the FSS that are in common use. DNAMIX I, II (John Storey) and III (Gary Beecham) is a package for traditional interpretation of DNA mixtures based on the work of Weir *et al.* [18], Curran *et al.* [17], and Beecham and Weir [25]. Charles Brenner produces a package called DNA-VIEW, which allows, amongst other things, the user to interpret DNA mixtures as well as do paternity and kinship analysis (http://dna-view.com/dnaview.htm). Mark Perlin and Beata Szabady [26] developed a method for the deconvolution of mixtures based on linear analysis that is implemented in the software Trueallele®.

In conjunction with a recent book on the statistical interpretation of DNA evidence [27], and following a series of dedicated papers by Fung and Hu, a software suit is now available to handle computation of mixed DNA profiles and kinship cases (http://www.hku.hk/statistics/EasyDNA/).

The Institute of Environmental Science Research Limited (ESR) has several products for DNA evidence. The program FaSTR is designed to do automated allele designation [28], and SPURS (Curran) estimates the sampling error in DNA mixture calculations using the Bayesian highest posterior density (HPD) method described in Curran *et al.* [29, 30]. SPURS III (Curran, Delme Thomas, Tony Ellis) also evaluates DNA mixtures both symbolically and numerically.

Outlook

KBSs are expected to be developed in the future on other evidence types as well, as they allow for consistency, storage, and transmission of expert knowledge. KBSs on other evidence types are being developed: The FSS produces a system for a Bayesian interpretation of fingerprint evidence. This system implements the work of Neumann *et al.* [31, 32]. DNA will remain an area with strong needs that will be met either through academic research (see e.g., http://people.math.aau.dk/~tvede/dna/) or commercial development (such as GeneMapper® ID-X from Applied Biosystems, the FSS-i^3 suite from the FSS, or TrueAllele® from Cybergenetics).

References

[1] Knowledge Based System (2008). [cited 1/29/2008]; Available from, http://en.wikipedia.org/wiki/Knowledge-based_systems (accessed 2008).

[2] Taroni, F., Aitken, C.G.G., Garbolino, P. & Biedermann, A. (2006). *Bayesian Networks and Probabilistic Inference in Forensic Science*, Springer Berlin, Heidelberg.

[3] Evett, I.W. (1977). The interpretation of refractive index measurements, *Forensic Science* **9**, 209–217.

[4] Evett, I.W. (1978). The interpretation of refractive index measurements II, *Forensic Science International* **12**, 37–47.

[5] Evett, I.W. & Lambert, J.A. (1982). The interpretation of refractive index measurements III, *Forensic Science International* **20**, 237–245.

[6] Evett, I.W. & Lambert, J.A. (1984). The interpretation of refractive index measurements IV, *Forensic Science International* **24**, 149–163.

[7] Evett, I.W. & Lambert, J.A. (1985). The interpretation of refractive index measurements VI, *Forensic Science International* **28**, 251–268.

[8] Evett, I.W. & Buckleton, J.S. (1989). Some aspects of the Bayesian approach to evidence evaluation, *Journal of The Forensic Science Society* **29**(5), 317–324.

[9] Evett, I.W. & Buckleton, J.S. (1990). The interpretation of glass evidence: a practical approach, *Journal of The Forensic Science Society* **30**, 215–223.

[10] Hicks-Champod, T.N. (2003). *De l'interpretation des Fragments de Verrre en Sciences Forensiques*, Ph.D., Universite de Lausanne, Lausanne.

[11] Walsh, K.A.J., Buckleton, J.S. & Triggs, C.M. (1996). A practical example of glass interpretation, *Journal of The Forensic Science Society* **36**, 213–218.

[12] Triggs, C.M., Curran, J.M., Buckleton, J.S. & Walsh, K.A.J. (1997). The grouping problem in forensic glass analysis: a divisive approach, *Forensic Science International* **85**, 1–14.
[13] Curran, J.M., Triggs, C.M., Hicks, T., Buckleton, J.S. & Walsh, K.A.J. (1998). Assessing transfer probabilities in a Bayesian interpretation of forensic glass evidence, *Journal of The Forensic Science Society* **38**(1), 15–22.
[14] Curran, J.M., Triggs, C.M., Almirall, J.R., Buckleton, J.S. & Walsh, K.A.J. (1997). The interpretation of elemental composition measurements from forensic glass evidence II, *Journal of The Forensic Science Society* **37**(4), 245–249.
[15] Aitken, C.G.G. & Lucy, D. (2004). Evaluation of trace evidence in the form of multivariate data, *Journal of the Royal Statistical Society, Series C* **53**, 109–122.
[16] Aitken, C.G.G., Lucy, D., Zadora, G. & Curran, J.M. (2006). Evaluation of transfer evidence for three-level multivariate data with the use of graphical models, *Computational Statistics and Data Analysis* **50**(10), 2571–2588.
[17] Curran, J.M., Triggs, C.M., Buckleton, J.S. & Weir, B.S. (1999). Interpreting DNA mixtures in structured populations, *Journal of Forensic Sciences* **44**(5), 987–995.
[18] Weir, B.S., Triggs, C.M., Starling, L., Stowell, L.I., Walsh, K.A.J. & Buckleton, J.S. (1997). Interpreting DNA mixtures, *Journal of Forensic Sciences* **42**(2), 213–222.
[19] Bill, M.R., Gill, P.D., Curran, J.M., Clayton, T., Pinchin, R., Healy, M. & Buckleton, J. (2005). PENDULUM – a guideline based approach to the interpretation of STR mixtures, *Forensic Science International* **148**, 181–189.
[20] Gill, P., Sparkes, R.L., Pinchin, R., Clayton, T., Whitaker, J.P. & Buckleton, J.S. (1998). Interpreting simple STR mixtures using allelic peak areas, *Forensic Science International* **91**, 41–53.
[21] Curran, J.M., Gill, P. & Bill, M.R. (2005). Interpretation of repeat measurement DNA evidence allowing for multiple contributors and population substructure, *Forensic Science International* **148**(1), 47–53.
[22] Gill, P., Whitaker, J.P., Flaxman, C., Brown, N. & Buckleton, J. (2000). An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA, *Forensic Science International* **112**, 17–40.
[23] Cowell, R.G. (2003). FINEX: a probabilistic expert system for forensic identification, *Forensic Science International* **134**, 196–206.
[24] Gill, P., Kirkham, A. & Curran, J. (2007). LoComatioN: a software tool for the analysis of low copy number DNA profiles, *Forensic Science International* **166**(2–3), 128–138.
[25] Beecham, G.W. (2006). *Statistical Methods for the Analysis of DNA Mixtures*, North Carolina State University, Raleigh.
[26] Perlin, M.W. & Szabady, B. (2001). Linear mixture analysis: a mathematical approach to resolving mixed DNA samples, *Journal of Forensic Sciences* **46**, 1372–1378.
[27] Fung, W.K. & Hu, Y.-Q. (2008). *Statistical DNA Forensics: Theory, Methods and Computation*, John Wiley & Sons, Chichester.
[28] Power, T., McCabe, B. & Harbison, S.A. (2008). FaSTR DNA: a new expert system for forensic DNA analysis, *Forensic Science International: Genetics* **2**(3), 159–165.
[29] Curran, J.M. (2005). An introduction to Bayesian credible intervals for sampling error in DNA profiles, *Law, Probability and Risk* **4**(1–2), 115–126.
[30] Curran, J.M., Buckleton, J.S., Triggs, C.M., Painter, I.S. & Weir, B.S. (2002). Assessing uncertainty in DNA evidence caused by sampling effects, *Journal of The Forensic Science Society* **42**(1), 29–38.
[31] Neumann, C., Champod, C., Puch-Solis, R., Egli, N., Anthonioz, A. & Bromage-Griffiths, A. (2007). Computation of likelihood ratios in fingerprint identification for configurations of any number of minuti&aelig, *Journal of Forensic Sciences* **52**(1), 54–64.
[32] Neumann, C., Champod, C., Puch-Solis, R., Egli, N., Anthonioz, A., Meuwly, D. *et al.* (2006). Computation of likelihood ratios in fingerprint identification for configurations of three minuti&aelig, *Journal of Forensic Sciences* **51**(6), 1255–1266.

Related Articles

Bayesian Networks
Case Assessment and Interpretation
Evidence Interpretation: a Logical Approach
Friction Ridge Examination (Fingerprints): Interpretation of
Glass Evidence: Bayesian Approach to
Mixture Interpretation: DNA

JAMES M. CURRAN



Variable Number Tandem Repeats

Introduction

There are a range of polymorphisms available for scientific analysis, many of which are capable of assisting the assessment of identity in a forensic inquiry. One major variety of polymorphism applied for this purpose are a particular type of repetitive sequence known as *minisatellites*, or variable number of tandem repeats (VNTRs). VNTRs are large fragments of DNA that are composed of sequentially aligned units, each of which has a high degree of sequence homology. Individual units are typically between 20 and 100 bp in length and are repeated consecutively up to 100 times. The overall molecular weight (MW) of a VNTR locus is determined by the number of times that the units are repeated. This is the polymorphic feature of these loci. VNTR loci are highly polymorphic and for that reason are powerful identification markers.

VNTRs were the first loci ever targeted in mainstream forensic DNA techniques. In the technique developed by Sir Alec Jeffreys [1, 2] VNTRs were analyzed using restriction fragment length polymorphism (RFLP) and following electrophoresis were visualized on an autoradiograph via Southern blotting [3]. These methods, originally referred to as "DNA fingerprinting", produced highly discriminating results in criminal [4] and noncriminal [5] cases. Restriction enzymes were specifically chosen to correspond with sites that flanked minisatellite loci. This meant that the minisatellite was excised and the inherent length variation of the loci was assessed (Figure 1). If a single DNA probe was utilized, the method was known as *single locus profiling* (*SLP*) whereas if a collection of probes were used this was termed *multilocus profiling* (MLP).

While the RFLP-VNTR approach was valid and reliable from a forensic perspective, the technique had significant limitations. A large volume (greater than 50 ng) of high-molecular-weight (HMW) high-quality DNA was required for RFLP analysis [6]. The technique was laborious and time consuming and radioactive materials were required for detection. The technique was unable to unequivocally resolve all the alleles of most VNTR loci and as such was criticized for the level of subjectivity required to interpret results [7–9].

A profound extension of forensic capabilities came through the emergence of the technique known as the *polymerase chain reaction* (*PCR*) [10, 11]. Essentially a sample preparation step, the PCR amplifies subanalytical quantities of DNA to a concentration that allows genotyping to be performed by routine analytical methods that vary according to the polymorphism being analyzed. Early PCR-based methods targeted smaller VNTR loci such as the D1S80 locus [12, 13]. These approaches were available in commercially produced DNA profiling kits and proved to be robust and effective forensic tools, providing an average random match probability (RMP) of approximately 10^{-4}, and requiring a template DNA concentration of only 2–5 ng (10–25

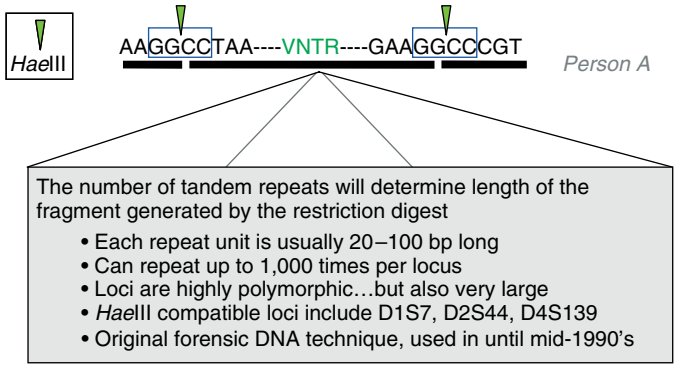


Figure 1 Diagrammatic representation of RFLP-typing of VNTR loci using a commonly applied restriction enzyme, *Hae*III

times less than required for RFLP-based VNTR typing).

References

[1] Jeffreys, A., Wilson, V. & Thein, S.L. (1985). Hypervariable 'minisatellite' regions in human DNA, *Nature* **314**, 67–73.

[2] Jeffreys, A.J., Wilson, V. & Thein, S.L. (1985). Individual-specific 'fingerprints' of human DNA, *Nature* **316**, 76–79.

[3] Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis, *Journal of Molecular Biology* **98**, 503–517.

[4] Gill, P., Jeffreys, A.J. & Werrett, D.J. (1985). Forensic application of DNA 'fingerprints', *Nature* **318**, 577–579.

[5] Jeffreys, A.J., Brookfield, J.F.Y. & Semeonoff, R. (1985). Positive identification of an immigration test-case using human DNA fingerprints, *Nature* **317**, 818–819.

[6] Budowle, B. & Baechtel, F.S. (1990). Modifications to improve the effectiveness of restriction fragment length polymorphism typing, *Applications of Theoretical Electrophoresis* **1**, 181–187.

[7] People v. Castro, 545 NYS 2d 985 (1989).

[8] R v. Tran, 50 A Crim R 233 (1990).

[9] Lander, E.S. (1989). DNA fingerprinting on trial, *Nature* **339**, 501.

[10] Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. & Arnheim, N. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction analysis for diagnosis of sickle cell anemia, *Science* **230**, 1350–1354.

[11] Mullis, K. & Faloona, F. (1987). Specific synthesis of DNA *in vitro* via a polymerase catalyzed chain reaction, *Methods in Enzymology* **155**, 335–350.

[12] Budowle, B., Chakraborty, R., Giusti, A.M., Eisenberg, A.J. & Allen, R.C. (1991). Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE, *American Journal of Human Genetics* **48**, 137–144.

[13] Kasai, K., Nakamura, Y. & White, R. (1990). Amplification of a variable number of tandem repeats (VNTR) locus (pMCT118) by the polymerase chain reaction (PCR) and its application to forensic science, *Journal of Forensic Sciences* **35**, 1196–1200.

Related Articles

Amplified Fragment Length Polymorphism

Autoradiograph

SIMON J. WALSH

Vehicle Accident Reconstruction

see Reconstruction: Accident

Vehicle Airbags *see* Airbags

Vehicle Deaths *see* Traffic Fatalities

Victimization and Decision to Call Police ***see*** Crime Victims' Decision to Report Crime

Video: Length Measurement ***see*** Length Measurement

Vietnam Survivor Syndrome ***see*** Syndromes: Psychological

Violence ***see*** Aggression

Violence: Risk Assessment ***see*** Risk Assessment: Patient and Detainee

Violence: Risk of Committing ***see*** Dangerousness: Risk of

Violence in Females ***see*** Aggression: Gender Differences in

Violence Risk Appraisal Guide (VRAG) ***see*** Dangerousness: Risk of

Violence Risk Assessment for Mental Health Professionals

A Bureau of Justice National Crime Victimization Survey found the annual rate of violent victimization to be 12.6 per 1000 for all workers. In comparison, the annual rate of violent victimization for mental health professionals was 68.2 per 1000 [1]. Compared to the rest of the healthcare field, the rate of violent victimization for mental health professionals was found to be approximately three times higher. Approximately 32.4–56% of psychiatric trainees reported being assaulted sometime during their training [2–5]. The subject of violence committed by patients both in the community and against mental health professionals has become a pressing issue for mental health professionals. When fatalities occur and are highly publicized, both public and professional concerns are raised about the adequacy of mental health assessment and treatment of potentially violent individuals [6]. Additional concerns about risk assessment and prevention have arisen as a result of changes in the public mental health system over the past several decades. The public mental health system is now servicing an increasing number of patients with serious mental illness who have been transferred from correctional institutions or other forensic services [7, 8]. It is understandable that this trend might increase clinician anxiety; however, its actual effects on risk management are still unclear.

The causes and treatment of pathological aggression and violence remain poorly understood despite their substantial costs to society [9]. It is believed that the vast majority of persons suffering from mental illness are not violent; rather, certain untreated symptoms appear to increase the risk of violent behavior [10]. The paucity of evidence-based research on

assessing and managing violence in clinical practice is striking, and has left clinicians with little guidance on which approaches may be best in the acute setting. In particular, there is a major gap in the research addressing how the clinician should approach even basic procedures such as interviewing for violence risk, or responding to a patient's aggressive behavior.

To date, most studies of patient violence have recommended implementing staff training that includes (i) didactic lectures on the biopsychosocial causes of violence, (ii) identifying high risk clients, (iii) training concerning the use of seclusion and restraints, (iv) regular patient risk conferences, (v) conflict management training, (vi) simulated training exercises, and (vii) required reporting of incidents [11, 12]. Clinical risk assessment and management involves identifying patient factors that affect violence risk (both aggravating and protective factors), organizing one's reasoning in the context of the patient's circumstances, and coming up with a risk management plan [12]. (*see also* **Risk Assessment: Patient and Detainee**; **Duty to Warn**).

General Principles

It is possible to conceptualize risk factors as falling into two broad categories – dynamic or static [13]. "Dynamic risk factors" are fluid and potentially modifiable. The clinical importance of dynamic risk factors lies in the clinician's potential ability to target them with interventions. "Static risk factors" do not change (e.g., gender, past violence), and have shown a statistical relationship with violence risk.

The following points are basic, yet important principles of clinical risk assessments:

1. Risk assessments should be contemporaneous (i.e., done immediately after the clinical encounter).
2. Risk assessments should consider both risk enhancing and protective factors.
3. Risk assessments should consider dynamic and situational variables.
4. Risk assessments should be done at clinically relevant or critical times (e.g., removal of restrictions, level changes, clinical worsening/improvement, discharge, etc.).
5. Whenever possible, relevant collateral data should be gathered (e.g., mental health records, police or probation records, etc.). Some clinical scenarios may warrant contacting family or social contacts, which may require the patient's documented consent. In the case of a psychiatric emergency, consent is waived.
6. Documentation of risk assessments should include decisions and reasons for choosing, or not choosing a particular intervention. This serves as a proof that you used careful, thoughtful reasoning consistent with the standard of care – or, what the courts refer to as "reasonable professional judgment".

Risk documentation should include some form of analysis of risk factors, and a general estimate of overall risk level (low, moderate, or high). Some experts believe that due to significant inter-rater reliability problems in the clinical setting with ratings in the "moderate" range, only ratings of low or high can be reasonably considered. This is the reasoning adopted by the Classification of Violence Risk (COVR) software program, designed to assist clinicians with decisions about discharge planning for acutely hospitalized civil patients [14]. The risk level should be followed by a treatment plan that directly addresses each relevant dynamic risk factor, and the clinician's reasoning for choosing or rejecting options.

In the forensic risk assessment literature, numerous actuarial risk assessment instruments (ARIs) have been developed. ARIs attempt to make predictions based on empirically demonstrated relationships between risk factors and outcomes [15]. These tools can be helpful guides for the clinician, but using them alone to determine patient management may run the risk of missing important clinical factors. Most authorities believe that they are best used to "structure" clinical judgment insofar as they remind the clinician to inquire about certain risk factors. Thus, at the present time, the standard of care *does not require* their use in clinical treatment settings [16, 17].

The standard of care requires mental health clinicians to "exercise the skill, knowledge, and care normally possessed and exercised by other members of their profession" [18]. Documentation showing that the clinician (i) performed a reasonable assessment of risk and then (ii) provided some rationale for implementing a reasonable management plan will be

very likely to meet the standard of care. The documentation need not be extensive, but should include the basic elements discussed in the following section on documentation.

The use of an ARI, or an ARI-derived checklist, to help "structure" the clinician's assessment may ultimately represent an "ideal" (above standard) practice. In the forensic mental health field, two of the most commonly used ARIs are the HCR-20 and Violence Risk Assessment Guide (VRAG), which have been shown to significantly predict violence in the community [19, 20]. These tools have enhanced validity when combined with a knowledge of "dynamic" risk factors derived from structured professional judgment [20]. These methods of approaching violence risk assessment are seldom taught to general clinicians and trainees who might benefit from them [21].

Summary of Clinical Violence Risk Factors

A *clinical violence risk assessment* quantifies the level of risk at a particular time using known risk factors, important clinical nuances, and professional judgment. Because violence risk is dynamic and influenced by many variables, periodic assessments over time are often necessary. The following list of risk factors has been culled from the above ARIs, the general literature on clinical violence risk [22–28], and consultation with forensic mental health experts in the field of violence risk assessment. The risk factors below are intended to assist the clinician in structured assessment of risk:

Historical factors

- past violence – must consider the pattern (egosyntonic, affective, and predatory);
- severe or frequent past violence;
- use of weapons during violent acts;
- age – late teens, early 20s
- male gender;
- low IQ;
- unemployed;
- major mental illness;
- criminal record;
- combat training;
- access to and familiarity with weapons;
- juvenile delinquency – especially first arrest before age 18;
- cruelty to animals, fire setting;
- childhood abuse.

Clinical factors

- homicidal or violent thoughts;
- substance use;
- impulsivity;
- poor insight into mental illness or past violent behaviors;
- noncompliance with treatment;
- psychosis: especially command hallucinations of familiar voices, and hallucination-related delusions;
- delusions: especially persecutory, systematized, misidentification syndrome, history of acting on delusions;
- depression – with suicidal ideas, or ideas about committing a homicide-suicide;
- mania (acutely symptomatic);
- organic brain dysfunction: especially traumatic brain injury, frontal lobe syndrome, and intermittent explosive disorder;
- posttraumatic stress disorder (PTSD) (acutely symptomatic);
- lack of empathy, antisocial or psychopathic traits;
- paranoid personality traits;
- accepting or condoning attitudes toward violence.

Acute factors

- homicidal or violent intent or plans;
- intoxication or recent substance use;
- actions taken on plans/threats;
- unconcerned with consequences;
- no alternatives to violence seen;
- intense fear or anger;
- specified victim – consider proximity, likelihood of provocation.

In addition, Table 1 outlines other important factors to consider after a patient has made a threat.

The Risk Management Plan

A risk management plan should be crafted immediately after the clinical risk assessment has been completed. Recall that risk assessments should be done at clinically relevant or critical times, such

Table 1 Factors to consider in assessing threat and risk [29][(a)]

A – *Attitudes that support or facilitate violence*: What is the nature/strength of the patient's attitude toward the behavior? Condoning or accepting? The stronger the perceived justification, the greater the likelihood of action. It may be helpful to assess the patient's appraisals of provocation from others, violent fantasies, and expectations of outcome.
C – *Capacity or means to carry out the violence*: Does the patient have the physical or intellectual capability, access to means, access to the victim, or opportunity to commit the act? How well does the patient know the victim's routines, whereabouts, etc.?
T – *Thresholds crossed*: Has the patient already engaged in behaviors to further the plan? Acts committed in violation of the law suggest a willingness/ability to engage in the ultimate act.
I – *Intent*: Does the patient have mere ideas/fantasies or solid intention? Level of intent may be inferred from the specificity of the plans and thresholds crossed. How committed is the patient to carrying out the act? Does he believe he has "nothing to lose"?
O – *Others' reactions and responses*: What reactions does the patient anticipate from others? Does the social network reduce or enhance the risk? Do social contacts believe the patient is serious?
N – *Noncompliance with risk reduction*: Is the patient willing to participate in risk management interventions? What is the patient's history of compliance/adherence to previous plans? How much insight into the situation does the patient have?

[(a)] Adapted from Borum and Reddy, 2001

as when the patient experiences a significant clinical worsening, significant stressor or upon admission and discharge. At such times, it may be necessary to obtain collateral data from mental health records, family members, or other social contacts. It should be noted that in the case of a psychiatric emergency (e.g., risk of suicide or violence) the need to preserve life supersedes the need to obtain consent. In most circumstances, this means that obtaining the patient's consent to contact family is not necessary.

The basic principle behind the "risk management plan" is to identify all those risk factors that are amenable to treatment interventions (dynamic risk factors), and target them with reasonable treatment interventions.

Documentation

The importance of good documentation cannot be overstated. It is the central piece of evidence in every malpractice trial, and good documentation has stopped many malpractice cases from proceeding. Courts do not expect clinicians to predict the future, prevent all tragedies, and render continuously flawless care. Rather, there is a general expectation that clinicians will use "reasonable professional judgment" based on a thorough consideration of the clinical data.

When clinicians do not document their reasoning, there is no evidence to show that they used thoughtful and reasonable professional judgment. Documenting information received, clinical decisions, and actions taken is an essential exposure-limitation technique [30]. The rule of austerity should be considered when documenting. In other words, the clinician should document the important facts and conclusions in an objective tone. The clinician should avoid waging battles of professional disagreement in the progress notes. Venting emotions into the progress notes rarely serves a useful purpose, yet is often harmful to the clinician who is a defendant in a malpractice case.

Risk assessment documentation should include some form of analysis of risk factors, and a general estimate of the overall level of risk. This should be followed by a treatment plan that directly addresses relevant dynamic risk factors, and the clinician's reasoning for choosing or rejecting options. In the event that instructions and information is given to the patient and the family, this too should be documented, along with whether or not they agree with the treatment decisions. In the event of a Tarasoff type situation, unrecorded warnings to a patient's family member that he or she has been threatened run the risk of being perceived as less credible by a jury.

When documenting interventions as part of the risk management plan (e.g., hospitalization versus intensive outpatient treatment, warning versus not warning a third party), it is advisable to include a statement explaining the rationale for the decision. For example, the clinician should document

that the option of hospitalization was considered, the clinical basis for rejecting that option, and the clinical basis for proceeding with a different option.

General Documentation Principles

First, the clinician should keep in mind at all times during actual documentation the fact that if a lawsuit occurs, the records may be read out loud in court [31]. In some cases, entire sections of the record are photographed, enlarged, and displayed on a poster board for the jury's inspection. Therefore, it is important that the documentation be clear and legible. To provide evidence of competent clinical care, the record should contain objective findings, patient statements, clinical judgments, and clinical decision-making. The most credible documentation is recorded, dated, and timed just after service is rendered. Documentation occurring after an adverse event is likely to be seen as self-serving and vulnerable to accusations of fabrication. After a tragedy has occurred, hindsight bias will often cause others to regard the event as more probable than it really was. Because many aspects of mental health treatment are less than "certain", it is helpful to document what seems tentative along with the reasoning for the clinical decisions. This helps emphasize the reality of the uncertainties inherent in clinical practice.

To counteract erroneous lay perceptions that all psychiatric patients are incompetent, at the relevant times the documentation should reflect the patient's capacity for decision making and ability to understand responsibilities such as reporting side effects, seeking emergency care, or notifying caregivers about changes in thought or mood. Quotations from the patient or family members are often viewed as highly credible evidences. For example, documentation that the patient stated, "I would never kill myself because I love my children too much", will provide important data for clinical decision making, as well as powerful evidence for a jury's consideration. Finally, even the most experienced clinicians regularly consider consultation with colleagues in difficult circumstances. Seeking an outside, objective opinion is the mark of a competent, caring clinician. In addition, documentation of the consultation will make it rather difficult for a plaintiff's attorney to claim that no other reasonably prudent clinician would have made the decision in question when both clinician and consultant have come to the same conclusion.

References

[1] Duhart, D. (1993–1999). *Violence in the Workplace*, National Crime Victimization Survey: Bureau of Justice Statistics Special Report. December, 2001. at http://ojp.usdoj.gov/bjs/pub/pdf/vw99.pdf (accessed on: 4/29/08).

[2] Schwartz, T. & Park, T. (1999). Assaults by patients on psychiatric residents: a survey and training recommendations, *Psychiatric Services* **50**(3), 381–383.

[3] Black, K.J, Compton, W.M., Wetzel, M. Minchin, S., Farber, N.B. & Rastogi-Cruz, D. (1994). Assaults by patients on psychiatric residents at the three training sites, *Psychiatric Services* **45**(7), 706–710.

[4] Dhumad, S., Wijeratne, A. & Treasaden, I. (2007). Violence against psychiatrists by patients: survey in a London mental health trust, *Psychiatric Bulletin* **31**, 371–374.

[5] Pieters, G., Speybrouck, E., De Gucht, V. & Joos, S. (2005). Assaults by patients on psychiatric trainees: frequency and training issues, *Psychiatric Bulletin* **29**, 168–170.

[6] Freedman, R., Ross, R., Michels, R., Appelbaum, P., Siever, L., Binder, R., Carpenter, W., Friedman, S.H., Resnick P. & Rosenbaum, J. (2007). Psychiatrists, mental illness, and violence, *American Journal of Psychiatry* **164**(9), 1315–1317.

[7] Cuddeback, G., Morrissey, J. & Cusack, K. (2008). How many forensic assertive community treatment teams do we need? *Psychiatric Services* **59**, 205–208.

[8] Lamb, H. & Weinberger, L. (2005). The shift of psychiatric inpatient care from hospitals to jails and prisons, *The Journal of the American Academy of Psychiatry and the Law* **33**(4), 529–534.

[9] Siever, L. (2008). Neurobiology of aggression and violence, *The American Journal of Psychiatry* **165**, 429–442.

[10] Friedman, R. (2006). Violence and mental illness – how strong is the link? *The New England Journal of Medicine* **355**(20), 2064–2066.

[11] Flannery, R., Farley, E. & Walker, A. (2007). Characteristics of staff victims of psychiatric patient assaults: 15-year analysis of the assaulted staff action program (ASAP), *Psychiatric Quarterly* **78**, 25–37.

[12] McNiel, D., Chamberlain, J., Weaver, C., Hall, S., Fordwood, S. & Binder, R. (2008). Impact of clinical training on violence risk assessment, *The American Journal of Psychiatry* **165**(2), 195–200.

[13] Meloy, J. (2000). *Violence Risk and Threat Assessment: A Practical Guide for Mental Health and Criminal Justice*, Specialized Training Services, San Diego.

[14] Monahan, J., Steadman, H.J., Robbins, P.C., Appelbaum, P., Banks, S., Grisso, T., Heilbrun, K., Mulvey,

E.P. & Roth, L. (2005). An actuarial model of violence risk assessment for persons with mental disorders, *Psychiatric Services* **56**, 810–815.
[15] Dawes, R., Faust, D. & Meehl, P. (1989). Clinical versus actuarial predictions of violence in patients with mental illness, *Science* **243**, 1668–1674.
[16] Simon, R. (2006). The myth of "imminent" violence in psychiatry and the law, *University of Cincinnati Law Review* **75**, 631–643.
[17] Mossman, D. (2008). Violence risk: is clinical judgment enough? *Current Psychiatry* **7**(6), 66–72.
[18] Dobbs, D. (2000). *The Law of Torts*, West Group, St. Paul, p. 269.
[19] Douglas, K., Ogloff, J. & Hart, S. (2003). Evaluation of a model of violence risk assessment among forensic psychiatric patients, *Psychiatric Services* **54**, 1372–1379.
[20] Doyle, M. & Dolan, M. (2006). Predicting community violence from patients discharged from mental health services, *The British Journal of Psychiatry* **189**, 520–526.
[21] Swanson, J. (2008). Preventing the unpredicted: managing violence risk in mental health care, *Psychiatric Services* **59**(2), 191–193.
[22] Appelbaum, P., Robbins, P. & Monohan, J. (2000). Violence and delusions: data from the macarthur violence risk assessment study, *The American Journal of Psychiatry* **157**, 566–572.
[23] Sirotich, F. (2008). Correlates of crime and violence among persons with mental disorder: an evidence-based review, *Brief Treatment and Crisis Intervention* **8**(2), 171–194.
[24] Pulay, A., Dawson, D., Hasin, D., Goldstein, R., Ruan, W., Pickering, R., Huang, B., Chou, S. & Grant, B. (2008). Violent behavior and DSM-IV psychiatric disorders: results from the national epidemiologic survey on alcohol and related conditions, *The Journal of Clinical Psychiatry* **69**(1), 12–22.
[25] Quanbeck, C., McDermott, B., Lam, J., Eisenstark, H., Sokolov, G. & Scott, C. (2007). Categorization of aggressive acts committed by chronically assaultive state hospital patients, *Psychiatric Services* **58**(4), 521–528.
[26] Swanson, J., Swartz, M., Van Dorn, R., Elbogen, E., Wagner, H., Rosenheck, R., Stroup, T., McEvoy, J. & Lieberman, J. (2006). A national study of violent behavior in persons with schizophrenia, *Archives of General Psychiatry* **63**(5), 490–499.
[27] Elbogen, E., Van Dorn, R., Swanson, J., Swartz, M. & Monahan, J. (2006). Treatment engagement and violence risk in mental disorders, *The British Journal of Psychiatry* **189**, 354–360.
[28] Petit, J. (2005). Management of the acutely violent patient, *The Psychiatric Clinics of North America* **28**, 701–711.
[29] Borum, R. & Reddy, M. (2001). Assessing violence risk in tarasoff situations: a fact-based model of inquiry, *Behavioral Sciences and The Law* **19**, 375–385.
[30] http://www.psychprogram.com/Risk_Management/news_treatingviolentind_p2.htm. (2008).
[31] Mossman, D. (2008). Tips to make documentation easier, faster, more satisfying, *Current Psychiatry* **7**(2), 80–86.

JAMES L. KNOLL, IV

Visitation Rights

Approximately half of the marriages in the United States end in divorce, affecting more than one million children a year. Since about 10% of divorces involve litigation over custody or visitation, ∼100 000 children a year are the subject of legal battles. Mental health professionals (usually a psychologist, psychiatrist, or social worker with child training) aid the courts by providing recommendations for custody and visitation. The methodology for a custody evaluation, the topic of this chapter, has been described by the American Psychological Association [1], the American Academy of Child and Adolescent Psychiatry [2], the National Interdisciplinary Colloquium on Child Custody [3], and Bernet [4].

Brief History of Divorce and Child Custody

Public policy and legal practice in the United States have passed through several stages [5]. Through most of recorded history it was understood that the children of a marriage were the property of the father and he routinely took custody of the children when divorce occurred. During the latter part of the nineteenth century, both the English and American legal systems started to act on behalf of the child and consider the relative moral fitness of the competing parents. The "tender years" doctrine, which was introduced in the late nineteenth century, presumed that young children should be raised by the mother because she had a stronger attachment and would provide better care. Since the 1920s, lawmakers and courts have placed emphasis on "what is best for the interest of the child," as expressed by Justice Benjamin Cardozo in *Finlay v. Finlay* [6]. The concept of best interests

of the child is broad and somewhat elusive, but significant because it implies that the needs of the child are more powerful than the rights of either parent. Starting in the 1970s, our society emphasized equality between the sexes – for example, many women objected that they seemed to have lower status in the workplace and many men objected that they seemed to have lower status in child rearing. As part of this broad social phenomenon, fathers challenged the assumption that the mother should routinely be the custodial parent and the father the noncustodial parent. In 1981, the landmark case of *Ex parte Devine* [7] established that the tender years presumption violated the Fourteenth Amendment to the US Constitution. Thereafter, the courts increasing ordered fathers joint and primary custody.

Reasons for Custody Evaluation

The vast majority of parents who separate and divorce do not need an elaborate and expensive custody evaluation conducted by mental health professionals. In most instances of divorce, parents work these issues out between themselves. In other instances, a mediator helps the parents agree on plans for the child or a judge simply takes the available information and makes the decision for the parents.

A formal custody evaluation by a mental health professional may be helpful to the child, the parents, and the court in circumstances, such as the following:

- one or both of the parents have a mental disorder that may impair the person's parenting skills;
- the child may have mental health needs that the parents and court should consider in developing the custody arrangements or parenting plan;
- the divorce has been unusually hostile and the custody evaluation may be a less adversarial approach to make decisions involving the children;
- the child's relative attachment to the mother and father appears to be an important issue;
- it is suspected that one of the parents has tried to alienate the child from the other parent through suggestion or indoctrination; and
- a parent has been accused of physical or sexual abuse.

Format for the Evaluation

Although practitioners may vary in the details of their evaluation procedures, the following are common components of custody and visitation evaluations.

Initial Conference

It is sometimes helpful at the outset to have a meeting with the mother and father together in order to clarify the purpose and the format of the custody and visitation evaluation that is about to occur. This initial conference can be used to collect basic information and resolve all of the administrative aspects of the evaluation, such as the following:

- the chronology of the marriage, including the births of the children, the separation(s), and the divorce;
- the legal status of the case, such as pending court dates;
- the court order authorizing the evaluation;
- the names and addresses of the attorneys; and
- the current situation, including the current visitation schedule.

The initial conference can be used to schedule all of the testing and interviews that constitute the evaluation. It is also an opportunity to obtain the written authorizations from both parents to obtain additional information from the pediatrician, teachers, therapists, and other pertinent individuals. The evaluator should obtain written permission from both parents to release clinical information at the end of the evaluation to: both of the parents' attorneys; the guardian *ad litem* or attorney for the child, if there is one; and the court. The arrangements for payment should also be established at the initial meeting.

At the outset of the custody evaluation, it is wise to clarify ethical issues such as confidentiality and role definition. Both children and parents need to be aware that this forensic evaluation does not provide total confidentiality. Ultimately, much of this information will be presented in the written report (to be read by a judge, attorneys, and other persons) and perhaps in the testimony of the evaluator. Before accepting this document, the practitioner should clearly explain that any material that she reads in conducting the evaluation might, and in some jurisdictions must, be produced to the opposing parent and attorney.

Regarding role definition, it is important for the parties to understand the difference between a forensic evaluation and psychotherapy. That is, the forensic evaluation is not therapy, although it may be therapeutic in the general sense. Also, practitioners who are already involved with some member of the family as a therapist should not undertake to perform a custody evaluation. Those are mutually exclusive roles that should be performed by different professionals. Throughout the initial meeting, the evaluator observes how the parents speak and cooperate with each other, aiding the evaluator's formation of an opinion on legal custody.

Parent Meetings

The evaluator meets with each parent individually in order to complete a clinical evaluation and to assess that person's parenting attitudes and skills. This may require one long meeting of 2 or 3 h or several shorter meetings. Although evaluators have different priorities as far as what information to collect from each parent, a sample agenda could include the following:

- a brief history of the marriage;
- a brief history of the period of separation and divorce;
- information about each child, such as the parent's opinion of the child's strengths, weaknesses, and reactions to the divorce;
- the past history of the parent, including education, work history, and legal problems, such as arrests;
- the psychiatric history of the parent, including symptoms, episodes of treatment, and how it might have affected that parent's relationship with the child;
- the medical history of the parent, including use of drugs and alcohol;
- the parent's proposal for the child's custody and visitation; and
- the parent's feelings toward the other parent.

In order to assess parenting attitudes, the evaluator may decide to present several hypothetical situations for the parent to assess and resolve. For example, "If you are granted custody of the child, how would you help the child maintain a good relationship with the other parent?" and "If you lose custody, how would you maintain a good relationship with the child?" As with any psychiatric or psychological interview, the evaluator is interested not only in the content of the answers but also in the way the parent approaches the task of the interview and the parent's style of relating to the interviewer.

Child Meetings

The meetings with the child should be used to accomplish these tasks: complete a clinical evaluation of the child; assess the child's attachment to each parent; determine how the parents' separation or divorce has affected the child; and assess whether the child has been indoctrinated in some way. In most circumstances, the evaluator should determine the child's preferences regarding custody and visitation and his reasons for that preference. In some instances, it may not be appropriate or useful to elicit the child's conscious preferences. There should be at least two interviews with the child, so that each parent brings the child for one appointment. The content and format of the interviews depend on the age of the child.

Preschool Children. With preschool children, the evaluator may want to start the interview with the parent and the child together. The evaluator could invite the parent and child to engage in some play activity together, such as drawing pictures or building with blocks, and could ask them to plan a weekend outing together. It should be possible to make observations about how both the child and the parent deal with such an assignment. About halfway through the meeting, the evaluator should ask the parent to leave and explain to the child that the parent will be waiting outside the interview room. The evaluator would have explained this procedure to the parent prior to the interview. When the parent leaves the room, the evaluator should observe how the child and parent deal with separation. During the remainder of the meeting, the evaluator can use play, drawing, or other techniques that are customarily used in assessing young children. At the end of the meeting, the evaluator should observe how the child and parent deal with reunification.

School Age Children. With latency children, the evaluator can use a semistructured interview commonly used in clinical evaluations of children. She can introduce topics or tasks that pertain to school, peers, recreational activities, and family relationships.

Simple projective questions can be helpful, such as the baby bird story; the evaluator may say, "A baby bird that lived in nest with its mother and father. One day there was so much wind that the baby bird was blown out of the nest. The baby bird, could fly, but only a little bit. Tell me what happens next in the story."

The child might express himself through drawings, such as draw-a-person, drawing his family doing something exciting, or drawing a picture of something happy/fun and something that is not so nice. One approach is to make a list of the parents and step-parents and ask the child what he likes and what he does not like about each person. The child's dislikes might turn out to include information about child maltreatment or parental behavior such as alcohol and drug abuse. The evaluator might ask the child to describe the current custody and visitation arrangements and ask whether the child would want them changed in any way. If the evaluator chooses to elicit the child's preferences, it should be done in a way that minimizes the importance of the question. It is useful to mention that it is the judge who is going to decide this issue, not the child or the interviewer. Because of the possibility of parental indoctrination, the interviewer should explore what each parent told the child to say to the evaluator and what the parents say about each other.

Adolescents. With an adolescent, the evaluator may take the approach of asking the youngster to express his opinions about the current situation at home. For instance, an adolescent might relate his perception of the relationship between the parents, what he knows about the reasons for the divorce, and the effects of the divorce on his own life. Most adolescents can discuss the advantages and disadvantages of life with each parent, the merits of particular visitation arrangements, and how to keep their own lives from becoming entangled with their parents' issues. The adolescent may have definite opinions about where he wants to live and what he wants to say to the judge. The evaluator should explore in detail the reasons for the youngster's preference and also other aspects of the case, such as how the adolescent intends to maintain a good relationship with both parents.

Home Visits

Many custody evaluations include home visits at each parents' residences. A home visit could include checking for the following: safety (e.g., "baby proofing"); age-appropriate toys and educational material; adequate sleeping arrangements; suitable food supply; and first aid supplies. Many evaluators interview the child and parent during the home visit, instead of interviewing the child with the parent at the office.

Outside Information

In conducting a custody and visitation evaluation, the evaluator should collect information from certain outside sources after obtaining permission from one or both parents, as appropriate. It is usually helpful to interview the other people who live in the mother's or father's home, such as step-parents, grandparents, nannies, and other siblings who are not directly involved in the custody dispute. In some cases, it is appropriate to speak on the phone with the family's pediatrician, since she may provide unbiased observations about both parents' skills and attitudes. It is often important to speak to previous and current psychotherapists of the child and the parents. When speaking with collaterals, including teachers and therapists, the evaluator must be mindful that they may already be allied with one of the parents and weigh their opinions accordingly. Treating therapists should not make custody recommendations to the court; however, they can discuss their observations of the parties with the individual who is conducting the independent forensic evaluation.

Psychological Testing

Psychological testing can be very useful in many custody and visitation evaluations. Comprehensive personality inventories such as the Minnesota Multiphasic Personality Inventory-2 and the Personality Assessment Inventory can be used to screen for unidentified parental psychopathology and to determine each parent's overall psychological adjustment. Both of these tests also contain scales that indicate the person's openness to providing sensitive information.

In some circumstances, it may be helpful to conduct formal intellectual testing (if there is a question about a parent's cognitive abilities) or a battery of

projective tests (if there is a question about the diagnosis of a parent's mental illness). Projective psychological tests constitute an opportunity to observe the parent's efforts to cope with an unstructured and stressful situation, as well as providing information about psychological resources and characteristics. Other psychological tests, such as the Parenting Stress Index, can provide information about a parent's level of attachment to the child, perceived problem areas in the child, and the perceived helpfulness of the spouse. The Ackerman–Schoendorf Parent Evaluation of Custody Test is designed to assess the fitness of parents for custody. Of course, if psychological testing is used, the same tests should be administered to both parents.

Psychological testing of the child can also provide useful information. Behavior rating scales such as the Achenbach Child Behavior Checklist and the Personality Inventory for Children can identify problem areas and clinical syndromes. The Family Relations Test, although it is not standardized, provides a nonthreatening means of assessing a child's feelings toward each parent, as well as the child's overall style of dealing with unpleasant feelings.

There are several standardized psychological tests that are specifically intended for custody evaluations, including the Bricklin Perceptual Scales (BPS), the perception-of-relationships test (PORT), and the Parent Awareness Skills Survey [8].

Conference with Attorneys and Parents

At the end of a custody and visitation evaluation some practitioners schedule a conference with the attorneys and parents together. This face-to-face conference gives the evaluator an opportunity to explain her conclusions and recommendations and to explain why certain factors were considered more important and other issues less significant. This wrap-up conference gives the parents and attorneys an opportunity to ask questions and for both parties to have access to the evaluator's answers.

Critical Factors

It is understood that, in contested cases, decisions regarding custody and visitation are guided by seeking the best interests of the child. However, parents and attorneys may disagree and argue about exactly what constitutes the best interests of the child in their particular set of circumstances.

There is no standard set of guidelines for what factors should be taken into consideration and what weight should be given to each factor. Each state has its own laws and precedents that spell out the issues for judges to consider in that jurisdiction. It is likely that evaluators and judges are influenced by their personal values when they make recommendations and hand down decisions in these cases. Many legislatures, courts, and mental health professionals consider the following issues important in child custody and visitation determinations.

Factors Associated with the Parents

- **Parental Attitudes and Parenting Skills.**

The evaluator should be able to assess whether the parent truly tunes to the child's emotional and physical needs *or* simply considers the child a narcissistic extension of himself or herself *or* is extremely anxious, passive, and helpless in common childcare situations.

- **History of caretaking.**

An evaluator should assess the level of involvement of each parent in the day-to-day childrearing activities. That is, to what degree has each parent fed and bathed the children, supervised their homework, organized their birthday parties, and taken them to the pediatrician? Being an active parent favors that person to be the custodial parent, although the evaluator must also consider a parent's unavailability due to work obligations.

- **Continuity of placement.**

It is usually presumed preferable to continue the current custody arrangement unless there is a good reason to change it. Usually a parent cannot file for a change in custody unless there has been a change in circumstance since the last time the court decided the issue.

- **Parental alienation.**

The presence of alienation must be assessed. The reader is referred to the section in this encyclopedia for a review of parental alienation phenomena.

- **Physical health.**

It is important to assess whether one of the parents has a serious or chronic illness which would compromise that person's ability to nurture the child.

- **Mental health.**

The clinical evaluation should reveal any serious psychiatric condition or any significant drug or alcohol abuse. The evaluator should comment on whether the psychiatric condition will impair parenting skills. For example, a history of repeated bouts of paranoid schizophrenia would be ominous. The specific diagnosis may not be as important as an assessment of the person's parenting skills in the present and in the future.

- **The households.**

The evaluator should consider the other people that are in the child's life, while in the home of each parent. The evaluator should note the people who are not appropriate for the child to be around, as well as people who are positive influences.

- **Financial considerations.**

This generally is not a factor, since ideally payment of child support tends to equalize financial differences. However, an evaluator might consider finances to be meaningful if it results in a large difference between the households – for instance, if one parent cannot afford an appropriate sleeping arrangement for the child.

- **Allegations of abuse.**

Sometimes one parent alleges that somebody in the other household, either the parent or step-parent or perhaps a boyfriend, has physically, emotionally, or sexually abused the child. If these allegations are substantiated, it would have a bearing on the child's custody and visitation.

- **Parental morals.**

The evaluator should assess whether a parent is exposing the child to immoral conduct. Obvious examples include exposing the child to criminal behavior, pornography, or parental drug use.

Factors Associated with the Child

Child's Attachment to the Parents. An evaluator should assess and consider the level of attachment between the child and each parent. Many children have a solid, positive attachment to both the mother and the father, which is certainly the most desirable circumstance. However, some children have a much stronger and healthier attachment to one of the parents.

Child's Preference. Many children understand that their parents are fighting over their custody, but really do not have a strong opinion or preference as to their living arrangements as long as they continue to spend time with both parents. Other children are able to express a definite preference and can explain their reasons for that preference. For instance, the child might express that one parent is warm and nurturing and the other is cold and aloof. The evaluator should determine whether the child's stated preferences are substantive and consistent with her own observations (e.g., "My stepdad is really mean!"); based on trivialities ("Mommy lets me stay up late on the weekend."); or misinformed ("Daddy told me that Mommy is stupid."). Whether the judge considers the child's preference depends on the child's age and varies from state to state. In general, a court is likely to give more serious consideration to the preference expressed by an adolescent than that of a child. In some states, the court is required by statute to consider the child's preference once the child reaches a particular age.

Outcome: Common Scenarios

Two Competent Parents

After conducting a custody evaluation, perhaps, the most common result is to find two parents who are both competent and nurturing in their own ways. That is, either of the parents would be a more-than-adequate custodial parent. If that is the case, the evaluator can list the various factors mentioned above – and others that might be relevant – and indicate which ones favor the mother and which ones favor the father. The evaluator can also indicate the significance or weight attached to each factor. By presenting the data in this way, both the evaluator and the judge can add up the list and determine whether one parent is preferred over the other.

One Competent, One Deficient Parent

Arriving at a conclusion in this kind of situation seems easier, since the evaluator can recommend

that the competent parent should have custody of the child. In addition, however, the evaluator can make specific suggestions designed to help the noncustodial parent become more capable. The evaluator would also want to discuss whether it is important for the child to continue to have a good relationship with both parents or whether one of the parents is so disturbed that he or she should not have contact with the child.

Two Deficient Parents

In some custody disputes, the evaluator may find that neither parent has the skills and attitudes to be fully satisfactory. The evaluator must be satisfied with the least detrimental alternative. If the evaluator is concerned about the parents' abilities, it may also be an opportunity to be creative and offer additional suggestions that might benefit the children. For example, supportive therapy or parenting classes or ongoing mediation between the parents might be helpful. The evaluator might recommend some continuing involvement by a capable relative or continuing supervision by a social service agency.

Outcome: Less Common Scenarios

Disputes Over Relocation

Americans are very mobile and frequently one or the other of divorced parents moves some distance from their original home, which may create a problem for the noncustodial parent to exercise visitation. These cases involve conflict among three competing interests – the right of the custodial parent to move, the right of the noncustodial parent to have visitation, and the right of the child to have a good relationship with both parents, as well as stability in his life. Because this is a changing area of law, an evaluator must know the current law in his practicing area. The current trend is for state legislatures and courts to make it somewhat easier for the custodial parent to move to another location, especially if it is done because of better job opportunities or a significant change in the family. Needless to say, there are times when it is in the child's interests switch custody to the parent who stays in the original community, because of the child's deep ties to friends, relatives, or school in that community.

Parent Versus Step-parent

When the dispute is between a biological parent and some other individual, such as a step-parent, grandparent, or foster parent, it is generally held that the biological parent has a greater right to the child as long as he or she is considered "fit." At times, this principle seems to contradict the principle of pursuing the best interests of the child, since a child might be removed from a wonderful step-parent or foster home and returned to biological parents who are barely adequate.

Grandparent Visitation

In recent years courts have considered whether grandparents have the right to visit their grandchildren, even without the agreement of the children's parents. Mental health professionals may be asked to make recommendations regarding grandparent visitation after considering the potential pros and cons, i.e., the possible benefit to the child on the one hand and the possible aggravation of family conflict on the other. Typically, that issue would arise in cases that have already come to the attention of the court, such as a custody or visitation dispute between the parents. In a recent ruling, *Troxel v. Granville* [9], the US Supreme Court concluded that the Constitution protects the right of parents to "establish a home and bring up their children". That means that in an intact family, grandparents will not be able to ask the court to override the parents' refusal to allow visitation.

Dispute over Religious Upbringing

Strictly speaking, each parent has the legal right to take his or her child to whatever religious activity the parent desires when the child is in that parent's household. In some cases, both parents want to enroll the child in religious education and to promote actively and assertively their respective religious beliefs. Since some religious beliefs profoundly contradict each other (one religion versus another, one sect versus another, and religious belief versus atheism), the child may become mentally confused by incompatible dogmas and emotionally disturbed by intense loyalty conflicts. Also, the issue of religion might be a consideration if the child has already formed an attachment to a particular faith and it

would be confusing or disruptive to interrupt that attachment. If the parents cannot negotiate a reasonable compromise, the evaluator may recommend that a particular parent be in charge of the child's religious upbringing. Also, an evaluator may be asked to comment on whether a person's religious beliefs has reached the point of becoming fanatical or even delusional, which could affect one's style of parenting.

Parental Kidnapping

One of the tragic outcomes of child custody disputes is that one of the parents may kidnap the child. This may occur if one of the parents concludes that his own circumstances are above the law or if he believes the child may be in danger if allowed to visit the other parent. A parent's intent on abducting a child may be assisted by an underground network. Parents desiring to bring back an abducted child have been known to hire private commando units. Federal laws (the Uniform Child Custody Jurisdiction Act and the Parental Kidnapping Prevention Act) and an international agreement (the International Child Abduction Remedies Act) provide procedures and sanctions to address some aspects of this issue. These laws provide that the ensuing custody trial should take place in the location where the child habitually resided prior to the abduction.

Homosexuality

In some states – for instance, California and New York – it is common for male homosexuals and lesbians to have custody of their children, but it is less usual in other parts of the country. A custody evaluation involving a homosexual parent should be based on objective data and not on stereotypes. The professional literature regarding this issue indicates that homosexuals can provide healthy, nurturing homes, although the children may be affected by social stigmatization in the community.

Possible Recommendations

By the end of the evaluation, the evaluator will be expected to provide the court an opinion regarding physical custody and legal custody. The two recommendations are not dependent on each other. That is, if the evaluator recommends sole physical custody, he may recommend joint legal custody. Alternatively, if the evaluator recommends joint physical custody, he may recommend sole legal custody.

Primary or Sole Physical Custody

With primary or sole physical custody, the child lives primarily with the custodial parent and has visitation with the noncustodial parent. In most instances, the child has regular, predictable visitation with the noncustodial parent. In some cases, the visitation is limited or supervised if the court finds that the noncustodial parent may be irresponsible (such as drinking and driving) or dangerous to the child. Sole physical custody would be appropriate if one parent is clearly competent and the second parent is significantly impaired and would not be reliable in caretaking activities.

Joint Physical Custody

In joint physical custody, the children live about equal time with each parent. For instance, the children might live one week at the mother's household and the next week at the father's household.

Split Custody

Split custody means that the children are divided between the parents rather than primarily living in the same household. It is generally considered advantageous to keep siblings together, mainly because children of divorce feel threatened and insecure and they derive support and consistency from each other. If the brothers and sisters are living together, they have the sense that at least part of the family is still in one piece. The evaluator would need to provide a strong reason for recommending that siblings should live in different households. However, when siblings will be living in different households, visitation can be arranged so that they are together much of the time. Every-other-weekend visitation, for example, can be scheduled in such a way that the siblings are always together every weekend.

Primary or Sole Legal Custody

When one parent has primary or sole legal custody of the child, that parent makes the decisions regarding

the child's education, medical care, and religious upbringing.

Joint Legal Custody

In joint legal custody, neither parent's rights are superior. In this arrangement, both parents have equal rights and responsibilities regarding issues such as the child's education, medical care, and religious upbringing. The joint custody order may include language to the effect that if the parties cannot agree on any particular issue, one party (usually the parent with actual physical custody) will have the final and exclusive decision on that issue. The evaluator may recommend joint legal custody if the parents are able to communicate with each other and are willing to take each other's opinions into consideration. Joint legal custody would not be an appropriate recommendation if the parents seem incapable of cooperating with each other.

Parenting Time

Some states do not use the terminology of "custody" and "visitation," but refer to the "primary residential parent" and the "non-primary residential parent." Instead of having conclusions regarding custody and visitation, the evaluator may make recommendations regarding the duration of the mother's and the father's "parenting time." The evaluator may make recommendations as to whether a particular parent should be responsible for major decisions (regarding education, health, and religion) or that the responsibility should be shared.

The Written Report

The evaluator should be aware that the written report will be seen by several people, including readers who may attach undue significance to isolated sentences and phrases. The best approach is to make the report detailed enough so that the reader fully understands the methodology that was followed and the basis for the conclusions and recommendations, but not so detailed as to include every datum that was collected. Although each evaluator writes reports in widely different ways, a report may include the following headings.

Identifying Information

This includes names and birth dates of the child(ren) and the contesting parties, who are usually the parents.

Referral Information

This usually includes a brief chronology of the marriage; a statement of the current status of the child(ren)'s custody and visitation; and an excerpt from the court order authorizing the evaluation. There should be a statement about the circumstances of the referral and the specific purpose of the evaluation.

Procedure for the Evaluation

This consists of a list of the various meetings that were held, the psychological tests utilized, and the outside information that was collected.

Observations

This part includes a separate section for each family member. Each parent is discussed individually, with a summary of that person's strengths, weaknesses, personality traits, significant medical and psychiatric problems, and whether these factors have a bearing on the person's ability to be a good parent. Each child is also discussed individually, with a summary of his strengths and weaknesses, identification of any psychiatric disorder, and a comment about how the child is coping with the parents' divorce. In addition, the report should address the following: the child's attachment to each parent; whether the child has a preference regarding custody and the reasons for the preference; and whether the child seemed unduly influenced by one of the parents.

Conclusions

This part consists of a list of specific statements that the evaluator believes are supported by her data, i.e., the observations discussed above. For instance, the evaluator might conclude that one parent has had a major mental illness in the past, that is, likely to be a problem again in the future;

or, that a parent has personality traits or a personality disorder that affects his or her parenting abilities; or, that the children are uniformly attached and bonded to one parent more than the other.

Recommendations

This section should follow logically from the conclusions. The evaluator may make recommendations regarding the custody of the children, the visitation schedule, whether the visitation should be supervised, whether any member of the family should be in psychotherapy, whether the parents should attend parenting classes, and other issues that are important in this particular evaluation.

References

[1] American Psychological Association (1994). Guidelines for child custody evaluations in divorce proceedings, *The American Psychologist* **49**, 677–680.

[2] American Academy of Child and Adolescent Psychiatry (1997). Practice parameters for child custody evaluation, *Journal of the American Academy of Child and Adolescent Psychiatry* **36**, 57S–68S.

[3] National Interdisciplinary Colloquium on Child Custody Law (1998). *Legal and Mental Health Perspectives on Child Custody Law: A Deskbook for Judges*, West Group.

[4] Bernet, W. (2002). Child custody evaluations, *Child and Adolescent Psychiatric Clinics of North America* **11**, 781–804.

[5] Derdeyn, A.P. (1976). Child custody contests in historical perspective, *The American Journal of Psychiatry* **133**, 1369–1376.

[6] Finlay v. Finlay, 240 N.Y. 429 (1925).

[7] Devine, Ex parte, 398 So.2d 686 (1981).

[8] Bricklin, B. (1995). *The Custody Evaluation Handbook: Research-Based Solutions and Applications*, Brunner/Mazel, New York.

[9.] Troxel v. Granville, 120 S.Ct. 2054 (2000).

Related Articles

Children: Suggestibility of

Parental Rights and Prerogatives

Parental Alienation

WILLIAM BERNET AND JOSEPH KENAN

Visual Recognition Systems in Identification

There has been a significant rise all over the world in police-reported crime figures particularly due to thefts, assaults, and robberies; consequently, the use of video-surveillance cameras has increased to face and deter this sort of crime. Thus, identification by video-surveillance systems is more frequently becoming the object of forensic investigation; such an investigation must be based on the physiognomic comparison of two persons: the suspect and the person represented in the photographs or film.

The issue of the identification of living criminals was first tackled in Europe at the end of the nineteenth century by Bertillon, who introduced a method of personal identification of the living based on anthropology, i.e., on the study of body measurements for anthropological comparison and classification. This approach has since been used in identification procedures.

With the increasing diffusion of video recording systems as deterring crime devices, it has become a routine procedure to compare images of a suspect with those obtained from surveillance films, paying special attention to the morphological characteristics of some peculiar anatomical parts (such as eye, nose, mouth, and ear) of the head, which are the most discriminant. These methodologies often meet limits of a practical nature: the subject may resort to disguises in order to hide identity and to prevent identification; on the other hand, in many cases, the initial images available have an informative content, which is too poor for applying the procedure in a reliable manner. Nonetheless, sometimes the images are of satisfactory quality and forensic examination can be performed.

Anthropologists are therefore getting more and more involved in the identification of the living [1–20]. Since the 1990s, anthropologists have, in fact, noticed an increase in the number of cases they are called upon as expert witnesses concerning the identification of living individuals from photographs. Nowadays, judges are constantly asking for "anthropometric" or photometric (though the term

is inaccurate) expertise with the request of verifying if a person seen on a specific video or photograph is the same person as the suspect. This is a completely different aspect of forensic anthropology, which deals with human diversity and strives to verify morphological and metric characteristics that make the physiognomy of one person distinct from that of another.

Facial recognition is one of the most common tasks performed in these instances, and one of the most difficult. In the previous articles, we have mentioned the difficulty in identification or individualization, the lack of common standards, and the intrinsic problems in identifying human remains by bone morphology during the comparison of postmortem and antemortem morphology. It may seem paradoxical but the comparison of faces is even more difficult. What our brain automatically does every day (i.e., the recognition of individuals by their face) is extremely difficult to analyze and replicate by scientific procedure. Thus, something that seems intuitive may be difficult to support in court. No common protocols exist at the moment and most experts still use "their own method". Furthermore, at present, perhaps because of the lack of science and of a clear methodology (the old methods are being severely questioned by recent research [7]), "identification" in these cases is frequently performed by nonbiological experts such as video analysts or even police staff, i.e., by experts who are not trained in anthropology or human diversity. Although there are, at the moment, no clear evaluations of how unexpert and expert opinions compare, and satisfactory algorithms still do not exist, the authors' experience has shown that the specialist in biology/anthropology/human diversity, which has a forensic formation, will tend to focus on the errors, difficulties, and pitfalls of this sort of comparison and be more cautious, whereas an "unscientific" eye may be led by instinct supported by pseudoscientific data (measurements, for example), which usually can be made to fit any hypothesis provided one does not consider the innumerous sources of error intrinsic to the method and to biological variability. Thus, it is the authors' opinion that this kind of expertise should be performed by experts trained in human diversity and forensic analysis. Another false impression is that computer experts may be sufficient as appropriate experts. Again, computer sciences may provide excellent instruments with which to perform a comparison and properly deal with and correct images; however, only disciplines such as human biology and anthropology can properly tackle the problems of variation and the study of human morphological traits – for example, of how and why they may change within the same person or be similar in two different persons. What is important for facial identification is still in part a mystery. We do know that human brains identify known faces more accurately and with a different method with respect to unknown faces, and that automated face recognition software has been seen to function almost as well; nonetheless, as in all situations, machines cannot be held the only ones responsible for such an important task.

Traditionally, there are three main approaches to facial identification. The first and perhaps most ancient is that of morphological analysis. Several classifications of forms of facial regions or types exist (from the Interpol schemas to other authors [1, 9]). As can be seen in Figure 1, these systems provide several variants of a facial character (such as face, mouth, and nose – some classifications envisaging up to 40 characters). They aim at meticulously classifying facial traits, for example, facial outline shapes, hairline shapes, mouth, nose, etc. Combinations vary among two different faces as among different populations. These tables are usually "designed to promote a consistent, systematic, and scientifically comparable evaluation of facial features as well as … making individual variation and population differences emerge from a seemingly unremarkable visage and can guide the expert in making a decision about the final match" [1]. Nonetheless, if it is true that verification of similar shapes among two faces can be a first step in the forensic analysis, it must be said that it cannot, however, be the only basis for forensic identification or individualization, i.e., even if two individuals share the classification of all 40 characters in the classification form, one cannot base identification only on this. A main problem is that the appearance of facial figures can be altered by expression, age, photographic angle, lighting, etc. Furthermore, high intra and interobserver error has been reported.

Then the anthropometric approach. Photoanthropometry can be defined as the analysis of anthropometric landmarks, dimensions, ratios, and angles to quantify facial characteristics and proportions from a photograph. Once again, the dangers in comparing faces on 2D images are only too well known: slight, imperceptible differences in orientation, facial

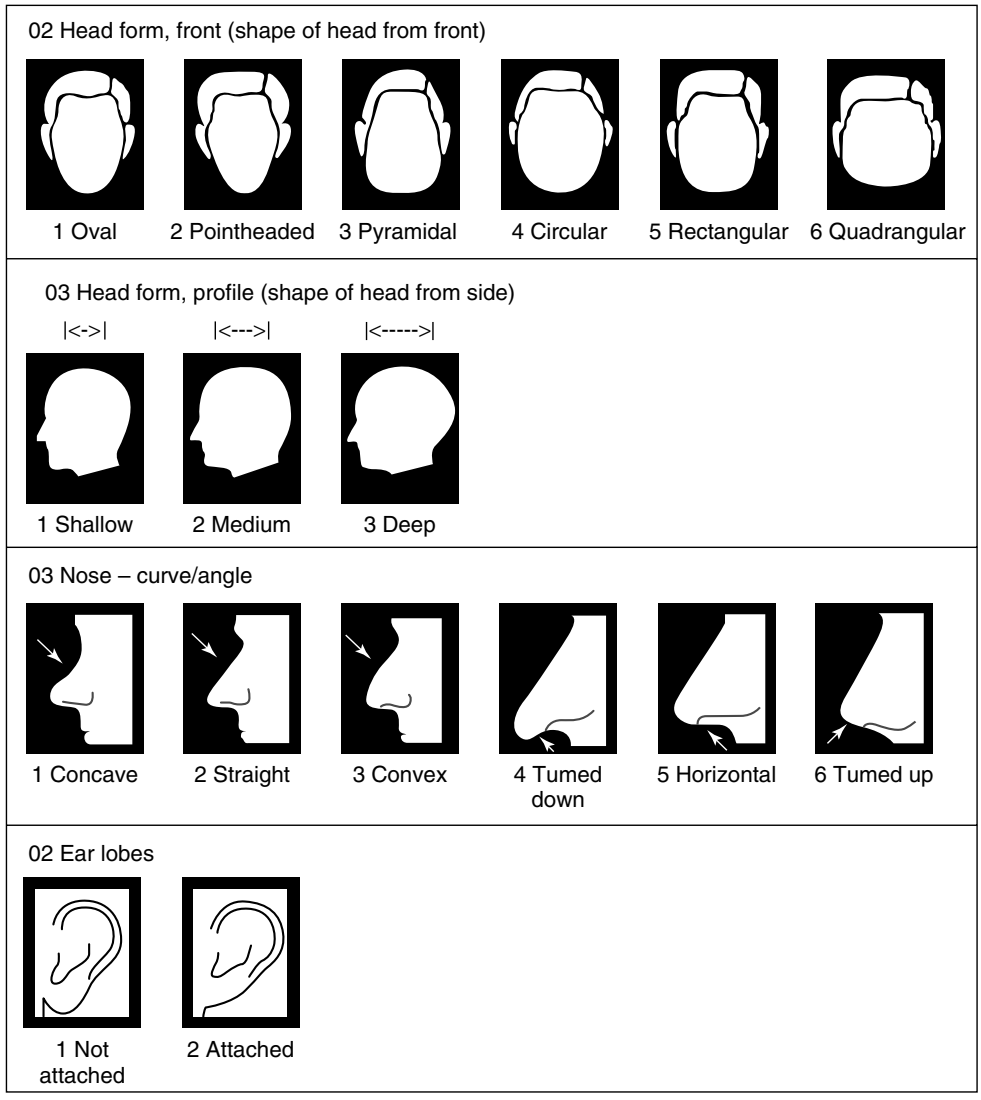


Figure 1 Example of part of the Interpol classification form concerning facial morphology

expression as well as interobserver differences in setting facial landmarks may cause drastic errors in the study of indices. Recent research has proven that the study of facial landmarks and measurements does not give consistent results. In a recent study, in particular [7], specific anthropometric landmarks were chosen and proportions and angle values between these landmarks were measured to compare target with test photos. The method did not generate the consistent results necessary for use as evidence in court.

Finally, "photographic superimposition" was introduced. The aim of this technique was to determine if facial features could be correctly superimposed on one another. Pictures with various degrees of transparency were initially superimposed; then videos and more elaborate computer systems began a novel era of superimposition. For example, one of the more common methods of recognition was (and unfortunately still is) performed by bringing the suspect to the scene of crime, having him take on the same position as the robber,

for example, and taking pictures or a video of the suspect with the closest possible head orientation as the criminal, and then performing a comparison of the facial characteristics of the two subjects by facial superimposition or through a numerical analysis of the two images. This solution, however, is full of pitfalls and requires the same recording conditions (the same camera, with the same focusing, shutter, positioning and orientation, magnification, etc.), which presumes that both camera and camera settings are not changed, usually impractical in actual situations, especially if a long time has elapsed between the criminal event and the expertise.

Furthermore, the problem that still remains and needs to be solved concerns the evaluation of the goodness of fit, whether it is possible to quantify it, as well as the usual problems that need to be corrected concerning image distortion, fuzziness, etc.

If we break free of traditional nomenclature, and focus on more modern technology, one can think in terms of comparison in 2D or 3D. More recently, advances in systems have encouraged attempts toward objective and automated identification approaches. Quantitative identification procedures can be split into two main categories: 2D–2D methods, based on the comparison between the photos of the subject to be identified (the "guilty") and the ones of the subject under investigation (the "suspect") (as mentioned above), and the so-called 3D–2D techniques, where a photo of the criminal and a 3D model of the suspect's face are matched. 3D–3D comparison techniques are more reliable because they overcome some limitations typical of the other two methodologies but they cannot be applied in actual situations until traditional 2D recording systems are substituted with 3D recording devices (such as stereoscopic systems and 3D scanners).

The 2D–2D techniques involve methodologies interested in "facial features" extraction, "geometrical features" extrapolation (eyes, eyebrows, mouth, etc.), or "jets" synthesis (referred to as *facial fiducial points* and texture around them): recognition is derived from the "similarity" degree between such facial elements in the original image and the corresponding ones belonging to the suspect's photo. Two commonly adopted approaches to perform face recognition of a person from a database are principal component analysis (PCA) and linear discriminant analysis (LDA). Both techniques try to simplify the original information content of the original images, by pointing out only the information useful for identification. The advantage is a reduction of computational costs, but they can be applied only to "one-to-many" matches (specifically to find the best match of a subject in a gallery) and not to "one-to-one" matches, that are the most common situations, when a comparison between the filmed subject and only one suspect is demanded. Moreover, all 2D–2D procedures show some intrinsic limitations, since recognition efficiency strongly depends on different environment conditions, (such as illumination, face positioning, and orientation). A number of image preprocessing techniques are requested in order to reduce the environment influence on recognition reliability, such as illumination compensation or pose estimation.

An enhancement in personal identification can be achieved through the use of 3D–2D techniques. Some authors have introduced an identification procedure based on a function, which computes the similarity between an unknown image pair and the 2D projection of a 3D face model of a stored subject. The stored 3D model is obtained starting from a stereoscopic system and extracting some facial fiducial points from the image pair.

More novel computer-aided recognition techniques can be envisaged, as independent as possible of recording conditions (illumination, face positioning, and orientation); for example, the application of the concept of "geometrically compatible images", as a tool capable of synthesizing a virtual camera, with the same optical and geometrical characteristics of the camera that filmed the original scene. The idea is to use the virtual camera to film a 3D facial model of a suspect and so to obtain a 2D image, which has the same size, position, and orientation of the original frame extracted from the video-surveillance sequence. The superimposition between the original 2D facial image and the 3D one is accomplished in correspondence of anthropometrical landmarks, distinctive for the person. The recognition is performed by superimposing the 2D facial image and the 3D one in correspondence of anthropometrical landmarks and an identification judgment is formulated on the basis of the measurements between corresponding landmarks positioned

both on the original photo and on the 2D facial model projection.

The more novel methods therefore strive to project a 3D model of the person under examination (produced with a laser scanner or with two stereoscopic cameras) (Figure 2) to the 2D image of the person to identify, in order to check matching of facial landmarks and contours. This is indeed a very rich area of research and facial identification methods may, once finely tuned, change the type and methods used for surveillance systems in terrorist attacks, bank robberies, and other crimes.

Although much research is being done, this particular field still needs to grow as regards methodology, and, from a forensic point of view, is full of pitfalls. As in the case of personal identification of the dead, the issue of "how much correspondence one needs to perform an individualization" is still an urgent matter to solve. Identification of faces still remains one of the most difficult and challenging tasks for the forensic anthropologist.

A similar type of reasoning can be applied to identification by ear morphology, limited by the same pitfalls as for the face, in general [10–14]. Sometimes, however, it is not possible to study these areas. In such cases, height and gait analysis can be the only elements which will at least allow the expert to exclude a suspect or to verify a mere compatibility.

At present, there is a growing interest for "at a distance" identification systems, such as systems of analysis and recognition of gait and/or the estimation of stature of the person on images. These systems do not require participation of the suspect and one can work on images of poorer quality with respect to that required by an identification system based on the face (which needs images with greater resolution).

The systems based on gait recognition quantify some typical biometric parameters such as the length of "stride" and gait [19, 20], typical of each person, and analyze the way the "silhouette" of a subject taped while walking, changes with time. These systems show some intrinsic limits, such as dependence on the angle of shooting, variations of the environmental conditions, the presence of persisting shades, and everything that can obstruct a correct extraction of the silhouette. Besides being systems that base the recognition on the periodical nature of gait, they entail problems when the subject is moving at varying speeds or on a route other than a straight line. Finally, these systems presuppose the existence of a reference sample of the person's gait, a factor which makes their application difficult as these are rarely available.



Figure 2 To the left, one can appreciate the face of the "perpetrator" with next to it the 3D image produced with a laser scanner from the suspect. To the right, superimposition of the face and 3D model, to check for matching points and profiles

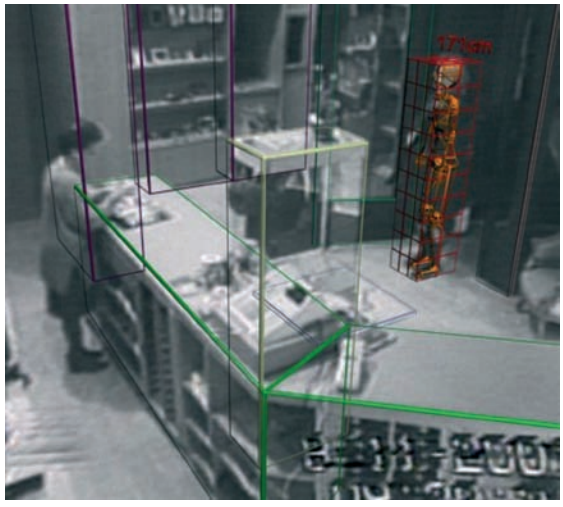


Figure 3 According to some authors, the determination of height is easily obtained by developing a virtual telecamera having the same characteristics of the video-surveillance system with which the images have been shot. The estimate of the height is done by comparison with a probe whose height can be varied in such a way as to be adjusted to that of the subject under investigation. The figure shows the criminal within a colored "box" (which includes the probe)

Height appears to also be an effective parameter (Figure 3), at least in proving the compatibility between the subject on tape and the suspect, or in excluding the suspect as the perpetrator [15–18].

Some have proposed the use of a stereoscopic camera surveillance system mounted on the ceiling. The head, shot by two telecameras, is extracted with a procedure called *head pattern matching*, which models the head on a sphere. The correspondence of the head seen in the two images allows one to estimate the distance between the head and the camera and therefore to deduct the height of the person.

Others demonstrate an increase in the probability of individualization when by including an estimate of height along with quantification of the biometric parameters of gait. Such parameters, however, must be intended as an aid to the individualization as they are not sufficient by themselves to ensure identification, but useful as an additional discriminating factor.

Regardless of the technological advances and applications, identification or individualization of the living by physiognomic traits still remains one of the most difficult and dangerous tasks for forensic anthropology.

References

[1] Iscan, M.Y. (1993). Introduction of techniques for photographic comparison: potential and problems, in *Forensic Analysis of the Skull*, M.Y. Iscan & R.P. Helmer, eds, Wiley Liss, New York, pp. 57–70.

[2] Farkas, L.G. (1994). *Anthropometry of the Head and Face*, Raven Press, New York.

[3] Porter, G. & Doran, G. (2000). An anatomical and photographic technique for forensic facial identification, *Forensic Science International* **114**(2), 97–105.

[4] Peacock, C., Goode, A. & Brett, A. (2004). Automatic forensic face recognition from digital images, *Science and Justice* **44**(1), 29–34.

[5] Burton, A.M., Miller, P., Bruce, V., Hancock, P.J. & Henderson, Z. (2001). Human and automatic face recognition: a comparison across image formats, *Vision Research* **41**(24), 3185–3195.

[6] Goos, M.I. & Alberink, I.B. (2006). Ruifrok AC 2D/3D image (facial) comparison using camera matching, *Forensic Science International* **163**(1–2), 10–17.

[7] Kleinberg, K.F., Vanezis, P. & Burton, A.M. (2007). Failure of anthropometry as a facial identification technique using high-quality photographs, *Journal of Forensic Sciences* **52**(4), 779–783.

[8] Kleinberg, K.F. & Vanezis, P. (2007). Varation in proportion indices and angles between selected facial landmarks with rotation in the Frankfurt plane, *Medicine, Science, and the Law* **47**(2), 107–116.

[9] Vanezis, P., Lu, D., Cockburn, J., Gonzalez, A., McCombe, G., Trujillo, O. & Vanezis, M. (1996). Morphological classification of facial features in adult Caucasian males based on an assessment of photographs of 50 subjects, *Journal of Forensic Sciences* **41**(5), 786–791.

[10] Alberink, I. & Ruifrok, A. (2007). Performance of the FearID earprint identification system, *Forensic Science International* **166**(2–3), 145–154.

[11] Hoogstrate, A.J., Van Den Heuvel, H. & Huyben, E. (2001). Ear identification based on surveillance camera images, *Science and Justice* **41**(3), 167–172.

[12] Swift, B. & Rutty, G.N. (2003). The human ear: its role in forensic practice, *Journal of Forensic Sciences* **48**(1), 153–160.

[13] Warren, N. (1996). Earprints in identification, *The Medico-Legal Journal* **64**(Pt 2), 82.

[14] Purkait, R. & Singh, P. (2008). A test of individuality of human external ear pattern: its application in the field of personal identification, *Forensic Science International* **178**(2–3), 112–118.

[15] BenAbdelkader, C., Cutler, R. & Davis, L. (2002). *Person Identification using Automatic Height and Stride Estimation*, IEEE International Conference on Pattern Recognition.

[16] De Angelis, D., Sala, R., Cantatore, A., Poppa, P., Grandi, M. & Cattaneo, C. (2007). New method for height estimation of subjects represented in photograms taken from video surveillance systems, *International Journal of Legal Medicine* **121**(6), 489–492.

[17] Criminisi, A., Zissermann, A. & Van Gool, L. (1998). A new approach to obtain height measurement for video, *Proceedings of the SPIE*, Boston, Vol. 3576, 1–6 November, 1998.

[18] Alberink, I. & Bolck, A. (2008). Obtaining confidence intervals and likelihood ratios for body height estimations in images, *Forensic Science International* **177**(2–3), 228–237.

[19] Lynnerup, N. & Vedel, J. (2005). Person identification by gait analysis and photogrammetry, *Journal of Forensic Sciences* **50**(1), 112–118.

[20] Larsen, P.K., Simonsen, E.B. & Lynnerup, N. (2008). Gait analysis in forensic medicine, *Journal of Forensic Sciences* **53**(5), 1149–53.

Related Articles

Anthropology: Aging the Living

Anthropology

Anthropology: Age Determination of Remains

Anthropology: Ancestry and Stature Determination

Facial Reconstruction

Human Remains and Identity

Identification of Human Remains

Identification and Individualization

Length Measurement

Sex Determination of Remains

Trauma Analysis of Skeletal Remains

CRISTINA CATTANEO AND DANILO DE ANGELIS

VNTR ***see*** Variable Number Tandem Repeats

Voice Recognition ***see*** Speaker Recognition

Vulnerability Assessment ***see*** Threat Assessment: Workplace



Warrants: No-Knock ***see*** Policing and Critical Incident Teams

Web Resources

Introduction

Researchers generally turn to the Internet in their quests for information, and those researching forensic topics are no exception. Some interesting and useful sources for forensic-related research can be found online. This article describes some of the best resources for forensic researchers.

General Forensic Resources

The Forensic Sciences webpages of the *National Institute of Justice (NIJ)* (http://www.ojp.usdoj.gov/nij/topics/forensics) provide the full text of many NIJ publications related to forensic sciences, as well as descriptions of related NIJ programs and funding sources. One of the essential works for forensics is the s *Handbook of Forensic Services* of the Federal Bureau of Investigation (FBI), the text of which can be found online at the FBI website (http://www.fbi.gov/hq/lab/handbook/forensis.pdf). It provides "guidance and procedures for safe and efficient methods of collecting, preserving, packaging, and shipping evidence" and describes "the forensic examinations performed by the FBI's Laboratory Division and Investigative Technology Division." This online version is divided into four major sections: Submitting Evidence, Evidence Examinations, Crime-Scene Search, and Crime-Scene Safety. The Handbook is also available in print from the FBI.

While the full text of the *Handbook of Forensic Services* is available online free of cost, it is generally true that the full texts of most books are not readily accessible online. However, a database of books from *FORENSICnetBASE/LawENFORCEMENTnet BASE* (http://www.forensicnetbase.com) provides the texts of dozens of forensic science and criminal justice books that are available in full text on the website. The database is available from CRCnetBASE for an annual subscription fee.

The *National Clearinghouse for Science, Technology and the Law* (*NCSTL*) at Stetson University College of Law provides a website (http://ncstl.org) that brings a wide variety of forensic-related information together in one place, freely available to public users. It offers a database of bibliographic information for a wide variety of resources, including books, scientific and legal journal articles, newspaper and magazine articles, seminars and conference sessions, dissertations and organizations. Books and journals that are indexed in the NCSTL database are available through interlibrary loan from the Stetson Law Library. The Related Links portion of the website provides a directory of hundreds of scientific and law-related links that are useful for forensic researchers. The section Education provides handouts created for professional development presentations by NCSTL staff, as well

as transcripts, podcasts, and webcasts of lectures on forensic science and technology.

There are several useful forensic-related websites providing either research pathfinders or a categorized directory of web links, or both. *Gelman Library's Forensic Sciences* pathfinder (http://libguides.gwu.edu/content.php?pid=9061) is an excellent resource for those performing research in topics related to forensic sciences. Provided by the George Washington University, this site provides information about forensic-related resources in both print and online formats. Additionally, *Science and Technology Resources on the Internet* (http://www.istl.org/03-spring/internet.html) by science and engineering librarian Cynthia Holt, is an online article from 2003 that describes and links to the best forensic resources on the Internet. Ms Holt updated this article in her 2005 book, *Guide to Information Sources in the Forensic Sciences*, published by Libraries Unlimited.

Zeno's Forensic Site (http://forensic.to/forensic.html) is a web directory of hundreds of forensic-related sites on the Internet, which are topically categorized. An interesting feature of the site is that users can rate the links, thus providing useful feedback to future researchers visiting the site. Zeno also provides a page of links to those sites he finds particularly interesting, and provides the opportunity to be informed by e-mail when new links are added. *Kruglick's Forensic Resource and Criminal Law Search Site* (http://www.bioforensics.com/kruglaw) is a web resource sponsored by Forensic Bioinformatics that lists links to over 1500 sites related to forensics or law. Informational articles on the site include Kim Kruglick's "A Beginner's Primer on the Investigation of Forensic Evidence," Donald E. Riley's "DNA Testing: An Introduction for Non-Scientists – An Illustrated Explanation," Frederic Douglas' "GC/MS Analysis," and Paul C. Giannelli's "Expert Qualifications & Testimony". Forensic researchers can also use *Reddy's Forensic Page* (http://www.forensicpage.com), which is a web directory of dozens of links related to forensic science and law. It is organized into topical categories.

The library at Johnson & Wales, a member of the Higher Education Library Information Network (HELIN), maintains several sites that describe and link to various – and often eclectic – research resources, including public and government criminal justice resources. The *Criminal Justice: General Resources* webpage (http://library.jwu.edu/research/websites/criminal.htm) points users to gateways, directories, academic sites, clearinghouses, think tanks, and professional associations/organizations. These resources do not include government agencies, which are located and described by the library's *Criminal Justice: Official Resources* webpage (http://library.jwu.edu/research/websites/criminal_official.htm). This site describes the Internet resources supporting the criminal justice activities of official federal, state, and local government agencies. Some of these links provide gateways to government agencies, international organizations, law enforcement directories, clearinghouses, and sites that maintain criminal justice statistical data.

Some unique forensic-related sites can be found online. These include *Crime and Clues: The Art and Science of Criminal Investigation* (http://www.crimeandclues.com), a site that pulls together articles about various aspects of criminal investigation. It not only provides links to new articles on its home page but also has archives of older articles categorized by topics, including different types of scientific evidence, crime-scene and death investigation, and testimony and ethics. The links are annotated, giving the reader the time-saving capability of narrowing down possibilities before following a link. Additionally, the Forensics and Investigations pages of *TruTV's Crime Library: Criminal Minds and Methods* (http://www.trutv.com/library/crime/criminal_mind/forensics/index.html) provide background reports on some of the most notorious or unusual investigations.

To make sure that forensic scientists from various countries are all on the same page, the Multilingua Group within the European Network of Forensic Science Institutes has created a *glossary of forensic-related terms* that are immediately translated into several European languages (http://www.enfsi.org/standingcommittees/eafs/multilingua).

Keeping up to date in any discipline can be a challenge. The *Crime Lab Project Forum* (http://crimelabproject.blogspot.com) is a blog that reports on the latest news about crime labs and other forensic-related stories. The webpage–based blog is frequently updated; however, readers who wish to follow the updates can opt to have them delivered to their e-mail addresses by means of a Yahoo! group. Additionally, *Daubert Tracker* (http://www.

dauberttracker.com) is a database that tracks US federal and state court decisions and supporting documents about "evidentiary gatekeeping". It provides a database of all reported decisions and many unreported decisions dating back to 1993. It also supplies information regarding the expert's name, discipline, area of expertise challenged, and results of the challenge. The Daubert Tracker is a fee-based service, with a subscription available.

Finding Forensic-Related Articles

Searching for an article related to forensic science is easy when using *Forensic Science Abstracts* (http://www.elsevier.com/wps/find/journaldescription. cws_home/506012/description#description). This index can be found in hard copies in many libraries, but is also part of the online collection of database called *EMBASE*, a biomedical and pharmacological database available from Elsevier (http://www.info.embase.com/embase_suite/about). *EMBASE* is available for a subscription fee.

The *FORS Forensic Bibliographic Database* (http://www.forensic.gov.uk/forensic_t/inside/products/fors/fors.htm) indexes articles about drugs and toxicology, forensic biology, DNA, forensic chemistry, forensic medicine and pathology, digital evidence, computer crime, documents and firearms examination, arson investigation, image processing, fingerprints, safety, quality and management aspects of running a forensic science service. It routinely covers 150 worldwide journals, but indexes from over 1900 sources. This index dates back to 1976. *FORS* is available for a subscription fee.

PubMed (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi) is a free service from the National Library of Medicine, National Institutes of Health. It includes over 16 million citations to life-science and biomedical articles dating back to the 1950s. It includes links to online sources of the full text of indexed articles, as well as a list of libraries that hold the individual article.

When searching for dissertations, researchers can use *ProQuest Digital Dissertations* (http://www.il.proquest.com/products_umi/dissertations), which indexesand abstracts dissertations and theses in all disciplines. Online guest users can search the most recent two years of the index and abstracts database. Full subscribers can access the entire database of over two million records.

The *National Criminal Justice Reference Service* (*NCJRS*) is sponsored by several offices within the United States Department of Justice and Executive Office of the President. *NCJRS* provides one of the largest criminal and juvenile justice libraries and databases in the world, the *NCJRS* Abstracts Database (http://www.ncjrs.gov/library.html). The *NCJRS* collection contains more than 185 000 publications, reports, articles, and audiovisual products from the US and around the world. These resources include statistics, research findings, program descriptions, congressional hearing transcripts, and training materials. The collection dates back to the 1970s. The Abstracts Database is available online for free, and links to the full text of documents whenever available.

The database at the *National Clearinghouse for Science, Technology and the Law* (http://ncstl.org) indexes forensic-related articles whose focus is science, technology, law or criminal justice, thus pulling together a variety of forensic-related disciplines into one place. The NCSTL database is available for free, and links to freely available full text of articles where they exist, or the webpage where such articles can be purchased if necessary.

Most journal articles can be accessed in full text online for a fee, but there are some journals that provide their full text for free. Some of these are American Journal of Pathology (from the American Society for Investigative Pathology) (http://ajp.amjpathol.org), Archives of Pathology & Laboratory Medicine (from the College of American Pathologists) (http://arpa.allenpress.com/arpaonline/?request=index-html), Crime Lab Minute (from the American Society of Crime Laboratory Directors (ASCLD)) (http://www.ascld.org), The Forensic Echo (http://echo.forensicpanel.com), Forensic Magazine (http://www.forensicmag.com), Forensic Nurse archives (http://www.forensicnursemag.com/archives.html), Forensic Science Communica tions (http://www.fbi.gov/hq/lab/fsc/current/index.htm), INTERfaces (newsletter of the Forensic Science Society) (http://www.forensic-science-society.org.uk/publications/interfaces.html), Journal of Clinical Pathology (http://jcp.bmj.com), Laboratory Investigation (from the United States and Canadian Academy of Pathology) (http://www.nature.com/labinvest/index.html), Microgram Journal (http://www.dea.gov/programs/forensicsci/microgram/journals_index.html), NIJ Journal (http://www.ojp.usdoj.gov/nij/journals/welcome.htm), Science and Technology

Review (http://www.llnl.gov/str/str.html), and TechBeat (http://www.nlectc.org/techbeat/justnet.html).

Forensic Associations

Most forensic associations and societies maintain an online presence through their websites, which generally provide information about the organizations and their membership. The *American Academy of Forensic Sciences* (*AAFS*) (http://www.aafs.org) is a professional society dedicated to the application of science to the law and is "committed to the promotion of education and the elevation of accuracy, precision, and specificity in the forensic sciences". AAFS was founded in 1948 and has nearly 6000 members, including physicians, attorneys, dentists, toxicologists, physical anthropologists, document examiners, psychiatrists, physicists, engineers, criminalists, educators, and others. AAFS publishes the *Journal of Forensic Sciences*. Current abstracts are available to the public at http://www.blackwellpublishing.com/journal.asp?ref=0022-1198, while abstracts for volumes 1–50 can be found online at http://journalsip.astm.org/JOURNALS/FORENSIC/jofs_home.html. The AAFS website also provides a wealth of information about education and career planning for forensic scientists, as well as links to other valuable forensic science resources.

The *ASCLD* (http://www.ascld.org) is a nonprofit professional society of crime laboratory directors and forensic science managers. It is "dedicated to providing excellence in forensic science through leadership and innovation." The website offers information about forensic science education and careers, as well as the society's official newsletter, the *Crime Lab Minute*, and links to other online articles focusing on forensic science in the news.

The *Canadian Society of Forensic Science* (*CSFS*) (http://www.csfs.ca) is a nonprofit organization of professionals with an active interest in forensic science. It is open to an international membership, and incorporated to maintain professional standards and to promote the study and enhance the stature of forensic science. Abstracts of the *CSFS Journal* are available on the CSFS website. Website visitors can link from the home page to the Population Studies Data Centre, which provides raw DNA data and frequency tables from the Royal Canadian Mounted Police and the Centre of Forensic Sciences in Toronto, Ontario. Visitors can also find information about forensic science education and careers.

Formed in 1915, the *International Association for Identification* (*IAI*) (http://www.theiai.org) is the oldest and largest forensic science/forensic identification organization in the world, with over 6400 members from many different forensic disciplines. Its website provides information about the association and its many divisions, publications, and professional opportunities.

Websites of other forensic-related associations or societies include: American Chemical Society (http://www.acs.org), American Society of Forensic Odontology (http://www.newasfo.org), American Society of Questioned Document Examiners (http://www.asqde.org), Armed Forces Institute of Pathology (http://www.afip.org), Association for Crime Scene Reconstruction (http://www.acsr.org), Association of Firearms and Tool Mark Examiners (http://www.afte.org), Association of Forensic DNA Analysts and Administrators (http://www.afdaa.org), Entomological Society of America (http://www.entsoc.org), Evidence Photographers International Council (http://www.epic-photo.org), Forensic Science Society (http://www.forensic-science-society.org.uk), International Association of Bloodstain Pattern Analysts (http://www.iabpa.org), International Association of Crime Analysts (http://www.iaca.net), International Association of Forensic Toxicologists (http://www.tiaft.org), International Institute of Forensic Engineering Sciences (http://www.iifes.org), Microscopy Society of America (http://www.microscopy.org), National Association of Medical Examiners (http://www.thename.org), and Society of Forensic Toxicologists (SOFT) (http://www.soft-tox.org).

Additionally, some scientific working groups or technical working groups related to forensics have their own websites. The Federal Bureau of Investigation has a general website for scientific working groups at http://www.fbi.gov/hq/lab/fsc/backissu/july2000/swgroups.htm. Other specific working group sites include SWGDAM – Scientific Working Group on DNA Analysis Methods (http://www.fbi.gov/hq/lab/fsc/backissu/april2003/swgdambylaws.htm), SWGDE – Scientific Working Group on Digital Evidence (http://www.swgde.org), SWGDOC – Scientific Working Group for Forensic Document Examination (http://www.fbi.gov/hq/lab/fsc/

backissu/april2000/swgdoc1.htm), SWGDOG – Scientific Working Group on Dog and Orthogonal Detector Guidelines (http://www.fiu.edu/~ifri/SWGDOG.htm), SWGDRUG – Scientific Working Group for the Analysis of Seized Drugs (http://www.swgdrug.org), SWGFACT – Scientific Working Group on Forensic Analysis of Chemical Terrorism (http://www.fbi.gov/hq/lab/fsc/backissu/april2004/standards/2004_02_standards01.htm), SWGFAST – Scientific Working Group on Friction Ridge Analysis, Study, and Technology (http://www.swgfast.org), SWGGUN – Scientific Working Group for Firearms and Toolmarks (http://www.swggun.org), SWGIT – Scientific Working Group on Imaging Technology (http://www.theiai.org/guidelines/swgit/index.php), SWGMAT – Scientific Working Group on Materials Analysis (http://www.fbi.gov/hq/lab/fsc/backissu/oct1999/trace.htm), SWGMGF – Scientific Working Group on Microbial Genetics and Forensics (http://www.fbi.gov/hq/lab/fsc/backissu/oct2003/2003_10_guide01.htm), SWGSTAIN – Scientific Working Group on Bloodstain Pattern Analysis (http://www.swgstain.org), SWGTREAD – Scientific Working Group on Shoeprint and Tire Tread Evidence (http://www.theiai.org/guidelines/swgtread/index.php), TWGFEX – Technical Working Group for Fire and Explosives (http://ncfs.ucf.edu/twgfex/home.html).

Web Resources Related to Specific Forensic Topics

Forensic Pathology

Visible Proofs: Forensic Views of the Body (http://www.nlm.nih.gov/visibleproofs) is an exhibition at the National Library of Medicine, National Institutes of Health, in Maryland. An online exhibit complements the exhibition in Maryland. The website features information about the history of forensic medicine, galleries of famous forensic cases throughout history, libraries of images and video files, and other educational resources.

Forensic Anthropology

Established in 1988 and reformed in 1992 as the *International Association for Craniofacial Identification* (*IACI*) (http://www.forensicartist.com/IACI/index.html), this association focusing on craniofacial identification is composed primarily of medical and scientific professionals from all over the world. Members also specialize in the related fields of forensic odontology, forensic anthropology, two and three-dimensional skull reconstruction techniques, computer-based skull reconstruction, facial aging for law enforcement, facial mapping and composite sketching. The organization offers educational opportunities that include classes ranging from "How to be a forensic artist" and "Understanding the human face" to both basic and advanced classes such as "Facial reconstruction sculpture". The site includes links to selected historical exhumation projects as well as nearly 30 related craniofacial identification sites and publications.

Wesley Neville, an IACI member and forensic artist for the Florence County Sheriff's Office in South Carolina, maintains *Forensic Art* (http://www.forensicartist.com), a website providing a brief description of the various facets of forensic art as well as links to several additional resources.

The *Central Identification Laboratory* (CIL) at the Joint POW/MIA Accounting Command (JPAC) in Hawaii is the world's largest forensic anthropology laboratory in operation. JPAC's mission is to account for missing Americans lost during military conflicts. The material evidence – personal effects, uniform items, weapons, tools, and more – collected throughout the world to identify the missing individuals is curated in the CIL and retained items are shared with museum partners. The website of the CIL (http://www.jpac.pacom.mil) provides an interesting presentation in describing how the staff of more than 30 civilian forensic anthropologists employs the related sciences of archaeology, odontology, odontosearch, and optosearch (which allows the investigator to determine the frequency of an eyeglass prescription in a population and therefore, determine the strength of match between an unknown and a known sample) to accomplish its identification missions. The site also includes general scientific information about forensic anthropology and related fields, news releases regarding its forensic work including its project to collect family reference samples (mtDNA) to aid in identification of missing service men and women, recounts of historical missions, and career opportunities.

The *American Association of Physical Anthropology* (*AAPA*) (http://www.physanth.org) is a

professional organization formed in 1930, whose members include more than 1700 professionals, scientists, and physical anthropology students who have demonstrated qualification for membership through publication or professional activities. The AAPA site provides information such as funding opportunities, the location of graduate programs, position statements by the organization, job postings, and access to its official journal, the *American Journal of Physical Anthropology*. Other pages include information about careers in physical anthropology, the organization's annual meeting, and links to other related scientific associations.

Biometrics

A trilogy of sites developed to foster greater collaboration and sharing of information related to biometrics among federal and state government agencies, commercial entities, and the public starts at *Biometrics.gov* (www.biometrics.gov). This site is the authoritative source for all biometrics-related activities within the federal government and is divided into three main areas. First, the Biometrics Reference Room tab leads to general information about biometrics technology, provides a synopsis of biometrics programs hosted by various federal agencies like the Department of Defense, Department of Commerce, Department of Homeland Security, or the Department of Justice, and offers *Privacy & Biometrics: Building a Conceptual Foundation*, a publication that presents the government's perspective on privacy with respect to biometrics. Next, the site provides information on the National Science and Technology Council's (NSTC) subcommittee on biometrics, including presentations, publications, and additional technical information. Third, a Media tab provides press releases, graphics and 'fast facts' geared for use among members of the press.

From *Biometrics.gov*, users can link to the two partner sites: *Biometrics Catalog* (http://www.biometricscatalog.org) and *The Biometric Consortium* (http://www.biometrics.org). Each of the partner sites contains general, introductory information on biometrics and the government's privacy policy, similar to the information found on Biometrics.gov, but each site also contains additional and distinct information.

Biometrics Catalog is a free, public repository of biometrics information sponsored by the federal government but kept current by its public users – both the biometrics scientific and commercial communities – who add information as it becomes available. Postings include news releases, related government documents as well as private and scholarly, peer-reviewed reports. The government documents and private-sector reports are maintained in a searchable database (through a combination of Boolean operators and/or date ranges). The database includes more than one hundred government reports published since 1991, as well as thousands of research papers, reports, conference presentations, publications and analysis across all fields of biometrics (fingerprints, voiceprints, facial recognition, ID cards, and iris recognition to name only a few topics). Examples of article titles include "Compound Stochastic Models for Fingerprint Individuality" and "3D Face Recognition Across Pose and Expression". While most of the publications referenced on the site are recent, it also includes the Bertillon Documents in French from 1881 and a 1919 study published by the Carnegie Institution of Washington on the basal metabolism in man.

Several additional databases maintained on the site allow users to search for information such as government-authored or government-sponsored evaluations of specific biometrics technologies (examples: *Independent Testing of Iris Recognition Technology* (May 2005); *Studies of Plain-to-Rolled Fingerprint Matching Using the NIST Algorithmic Test Bed (ATB)* (April 2004)). Federal solicitations for biometrics research and development or services and reports on biometrics-related legislative or governmental agency activity are also available. The remaining databases consist of postings listing conferences and educational courses, commercial products and their companies, and biometrics consultants and analysts.

The *Biometric Consortium* (http://biometrics.org) is described as a "focal point for research, development, testing, evaluation, and application of biometric-based personal identification/verification technology". The organization has hosted annual conferences since 1992 and provides links to conference material from the later conferences as well as information on upcoming conferences. Links to other meetings or educational events and materials are also available. Additionally, the site provides links to universities researching biometric technology and their specific programs, links to selected periodicals and research publications arranged by

year of publication from 1994 through 2003, and provides an extensive biometrics glossary.

Forensic Botany

The *Internet Directory of Botany*, compiled by Anthony R. Brach, Harvard University Herbarium, and others (http://www.botany.net/IDB), is an extensive, award-winning alphabetical index of links to online botanical information. Users may peruse the lists of links to specific databases, articles, and other resources or search terms or keywords found within the titles. For example, "A Mini-Course in Medical Botany Syllabus by James A. Duke" is listed under the "A" index and will not be found if searching the "M" index (for "mini-course" or "medical botany"). However, a keyword search for "medical botany" or "James Duke" will locate the article.

Toxicology

TIAFT.org (http://www.tiaft.org) is the official website of The International Association of Forensic Toxicologists and offers direct links to additional resources from both public and members-only areas. For example, the public section maintains a user-contributed collection of reference analytical data to assist in the identification of unknown toxic substances, an *Observatory* section composed of links and a large directory of all genres of toxicology-related websites arranged by forensic specialty, and a *Powersearch* area to search for and access various scientific and medical literature and technological information.

Membership to TIAFT is $40 annually and allows access to the organization's *therapeutic and toxic drug concentrations list* developed by toxicology and medical specialists and from information obtained through the field's literature, pharmaceutical industries and by comparison with established drug data lists. Members may also access the complete text of the organization's quarterly journal, retrieve previously published journal *case notes*, read *online reviews* of various articles focused on state-of-the-art forensic and analytical toxicology topics, and tap into a collection of nearly 1500 papers presented at TIAFT meetings spanning three decades.

Soft-tox.org is the website of the SOFT, an organization of practicing forensic toxicologists and others who share an interest in the discipline. The site provides brief and general information introducing forensic toxicology and allows public users to download guidelines for the practice of forensic toxicology in the two defined areas of postmortem forensic toxicology and human-performance forensic toxicology. Additional downloads include the organization's *Drug Facilitated Sexual Assault Survey* and *Drug Facilitated Sexual Assault Drug List and Cutoffs* and the new *American Academy of Forensic Sciences (AAFS) Toxicology Section Mass Spectrometry Database*, a comprehensive drug library of the spectra for more than several hundred substances including a mini library of the mass spectra of newer drugs, metabolites, and some breakdown products.

The *California Association of Toxicologists* maintains a website (http://www.cal-tox.org) that provides an online toxicology-related exchange and discussion forum ("CAT Forum") interlinked with the Southwestern Association of Toxicologists. In addition to its links page, the site also features an extensive searchable database of online literature reviews by Dr Rodger Foltz with particular emphasis on analysis of new drugs, newer chromatographic and mass spectrometric techniques, and, in general, their application for the identification and quantitation of drugs and metabolites in physiological specimens.

Fingerprints

The FBI maintains a webpage (http://www.fbi.gov/hq/cjisd/cjis.htm) from which users may access general and historical information on fingerprint identification (*All About Fingerprints* tab), become familiar with the federal integrated automated fingerprint identification system (IAFIS), learn the proper method for taking legible fingerprints, and discover training opportunities.

Latent Print Examination: Fingerprints, Palmprints and Footprints (http://www.onin.com/fp/index.htm) is possibly the best of several solid online resources dedicated to latent-print examination, and is useful for both the novice and the experienced. The website, maintained by Ed German, a recognized expert in latent-print examination, is an extensive repository of relevant latent fingerprint, handprint, and footprint technology, history, news, case law, and links. Topics from challenges to reliability of fingerprint evidence to the latest technologies for crime-scene processing and evidence collection are explored in depth; online discussion forums and opportunities

to "ask an expert" are also available to seek additional answers and information.

Ridges and Furrows at (http://www.ridgesandfurrows.homestead.com) focuses on the enhancement of latent prints and latent-print identification through the use of advances in digital technology. The history of latent-print examination and court challenges to fingerprint science, particularly as they relate to the advancements in digital technology, are offered, in addition to links to relevant the newsletter, discussion forums, and other online learning opportunities. The website is maintained by Mary Beeton, a Canadian fingerprint technician with the Durham Regional Police Service in Ontario.

DNA

MITOMAP: A Human Mitochondrial Genome Database (http://www.mitomap.org) offers a comprehensive and searchable "compendium of polymorphisms and mutations of the human mitochondrial DNA". The database can be searched by gene, disease, or enzymes, for example, and keywords may be abbreviated or truncated. Alternatively, the user may use subsections divided into areas including "mtDNA Polymorphisms" and "mtDNA Mutations with Reports of Disease-Associations" and organized by mtDNA location or phenotype. The website is supplemented by several illustrations and tables, a "Mitomap Quick Reference" section that includes an extensive bibliography of mitochondrial references, and links to additional databases and related sites.

DNA● VIEW (http://dna-view.com) is the website maintained by forensic DNA identification consultant Charles Brenner, and presents a comprehensive look at forensic DNA analysis particularly as it relates to mathematics. The site provides topical news, articles, archived discussions on DNA identification including recent identifications after mass disasters, information on DNA identification software, and data tables organized by subject (for example, "Allele Frequencies for US Populations" provided by Cellmark Diagnostics). Topic areas such as "Discussions" are enhanced with photos, PowerPoint® presentations, news, and other articles provided by the participants.

Firearms

FirearmsID (http://www.firearmsid.com), privately maintained by Jeffrey Scott Doyle, a firearm and tool-mark examiner with the Kentucky State Police Regional Forensic Lab, provides extensive educational and investigative information including firearm safety, topical articles, expert testimony, firearms testing, and an introduction to firearms and ballistics; it is arranged by categories like the *History of Firearm ID* and *Case Profiles*. The website offers a discussion area, *Forensic Forum*, and a *resource area* provides such subtopics as *Ballistics Picture Book* and *Virtual Comparison Microscope*, while another area leads the user to databases containing rifling data and a bullet and shotshell component search.

Odontology

Forensic odontologist Mike Bowers created the website *Issues in Human and Animal Bite mark (Bitemark) Analysis* (http://www.forensic.to/webhome/bitemarks), which provides an extensive overview of the subject, supplemented by hyperlinks to several case studies (for example, serial killer Ted Bundy and the crash of Alaska Airlines Flight 261 in 2000), photos of bitemark evidence, journal articles, and links to similar websites.

Forensic Dentistry Online (http://forensicdentistryonline.com) is the International Organisation for Forensic Odontostomotology (IOFOS)'s contribution to forensic dentistry. The website includes information on bitemarks and bitemark identification, including new resources using DNA from teeth and saliva, as well as information covering the legal aspects of bitemark evidence admissibility. Also included are sections posting news items and commentaries from users, book reviews on the subject, and links to continuing education courses and odontologists.

Questioned Documents

The website, maintained by certified questioned document examiner Emily J. Will (http://www.qdewill.com), offers an overview of the subject through areas exploring the theories and applications of questioned-document examination, the external tools used in examination and a presentation of famous cases. At *Identifont* (http://www.identifont.com/index.html), users can use the largest independent directory of online typefaces to search for

type and picture or symbol fonts by font appearance, name, or similarity. Users can also download a wide selection of fonts for free. *Identifont* was developed by Human–Computer Interface, an information design and documentation company specializing in high-tech products. Similarly, *Omniglot* (http://www.omniglot.com) provides details of alphabets and writing systems, both current and ancient. Each writing system is illustrated, with information provided about its origin, usage, notable features, and the languages written with it.

DIANA BOTLUK AND ELIZABETH FITTERMAN

Weisgram v. Marley

Facts, Issues, and Interest of Forensic Scientists Summarized

Plaintiff Weisgram, in this product liability case, brought suit for the wrongful death of his mother who was killed during a fire. The fatal conflagration was allegedly caused by a defective baseboard heater manufactured by the defendant Marley Company. Plaintiff sought to prove his case by the testimony of three witnesses, proffered as experts, in his attempt to causally connect the heater defect to the fire. Defendant objected to the testimony on the ground that it was unreliable under *Daubert*. The trial court overruled the objections and permitted the witnesses to testify. The trial resulted in a verdict for the plaintiff. Defendant requested judgment in his favor as a matter of law, urging that the court erred in admitting the expert opinion testimony. The trial court denied the request and entered judgment for plaintiff.

On appeal to the intermediate reviewing tribunal – the Court of Appeals for the Eighth Circuit – the court held that Marley's motion for judgment as a matter of law should have been granted. Weisgram's expert testimony was found to be too speculative and not shown to be scientifically sound. Without this testimony, Weisgram failed to prove his case. The reviewing court decided it was not required to reverse and remand for a new trial where the evidence was not even close.

The issue was one on which different Courts of Appeal had arrived at conflicting decisions. The Supreme Court agreed to resolve the conflict. In doing so, it affirmed the Eighth Circuit court's decision and approved its reasoning. The Court simply adhered to its pre-*Daubert* 1967 opinion in *Neely* [1] and held that pursuant to procedural rules courts have the power to enter a judgment as a matter of law against a party if there is no legally sufficient evidentiary basis which a reasonable jury could find for that party.

The case addresses the scenario where an expert witness's testimony is challenged at trial, but the court, applying *Daubert* criteria, allows the expert to testify. This ruling is the lastest of three decisions further explaining the meaning of Daubert v. Merrell Dow Pharamaceuticals. See endnotes below for more information on these rulings.[a.] On appeal, however, the reviewing court concludes that the basis for the expert's opinion was not validated and therefore not "reliable" under *Daubert*.

While *Weisgram* is technically applicable only to federal courts of the United States, it has applicability in many jurisdictions that have similar procedural statutes or rulings.

Forensic scientists should note the authority of courts in the United States to enter, under certain circumstances, a judgment as a matter of law when they rule expert evidence inadmissible. They should also pay heed to the Court's words in Weisgram:

> Since *Daubert,* . . ., parties relying on expert evidence have had notice of the exacting standards of reliability such evidence must meet. It is implausible to suggest, post-*Daubert*, that parties will initially present less than their best expert evidence in the expectation of a second chance should their first try fail [2].

Be aware of the admonition that "[l] itigatgors must prepare their cases with an understanding that heightened scrutiny of all but the most noncontroversial expert testimony is likely" [3] in jurisdictions that are steeped in the adversary system.

End Notes

[a.] For summaries and importance of the other two decisions, see the separate entries: *General Electric v. Joiner* (***General Electric v. Joiner***), *Kumho Tire*

v. Carmichael (***Kumho Tire v. Carmichael***) and *Daubert v. Merrell Dow Pharmaceuticals* (***Daubert v. Merrell Dow Pharmaceuticals***).

References

[1] Neely v. Martin K. Eby Construction, 386 U.S. 317, 87 S.Ct. 1072, 18 L.Ed.2d 75 (1967).
[2] Weisgram, 528 U.S. at 455, 120 S.Ct. at 1021 (2000).
[3] Ollanik, S.A. (1999). *Expert Testimony Defeating the Kumho Challenge*, Trial, Nov., at pp. 28, 30.

ANN C. SMITH

Whole Genome Amplification

Introduction

DNA typing has revolutionized forensic biology and has resulted in a dramatic increase in the ability to obtain probative information from crimes involving biological material [1–8]. The adoption of PCR-based methods resulted in a significant enhancement in the capacity for successful typing of biological crime scene samples [9–13]. However, there are still situations where it is extremely challenging to obtain a standard autosomal short tandem repeat (STR) DNA profile, although cellular material is present in the evidence. Some of these situations involve low copy number (LCN) samples that contain <100 pg of template DNA, equivalent to 15 diploid or 30 haploid cells [14]. The presence of such LCN samples could be due to several factors including damaged or degraded DNA, oligospermic or aspermic perpetrators, or from extended interval postcoital samples, where sperm have been lost over time due to the effects of drainage or host cell metabolism [15–17]. Other trace biological evidence also contains small quantities of cells, including fingerprints, particulate matter, and aerosols [18, 19]. Standard typing protocols permit the detection of as little as 100 pg of DNA and therefore may not be able to accommodate LCN samples frequently encountered in forensic casework.

In an attempt to increase the recovery of DNA profiles from LCN samples, the use of increased cycle number (ICN) has been suggested to permit the formation of sufficient amounts for analysis. A critical assumption of the ICN approach is that any expected stochastic effects can be dealt with by appropriate interpretation algorithms [20, 21]. However, this method may not always be efficient, with additional cycles at high temperatures leading to a decrease in efficiency of Taq DNA polymerase [22]. This results in less amplified product being produced per cycle, unless the process is halted for fresh enzyme to be added before the additional cycles. Similarly, nested polymerase chain reaction (PCR) using two sets of locus specific primers has been suggested for LCN samples, but it is difficult to envision its use with forensic specimens due to primer compatibility problems with multiple multiplex PCR systems [23, 24].

Another approach to the analysis of LCN samples utilizes whole genome amplification (WGA) strategies. WGA methods employ various random-sequence primers and low-stringency annealing conditions to amplify large tracts of the genome in an attempt to increase the effective number of starting genomic templates prior to any downstream analysis. Theoretically, by preamplifying the limited amount of genetic material in the sample, sufficient quantities of template can be produced to overcome stochastic effects resulting from LCN templates. Several WGA methods have been developed and reported in the forensic literature, including primer extension preamplification (PEP) [25], degenerate oligonucleotide primed PCR (DOP-PCR) [26–29], long fragments from low quantities DOP-PCR (LL-DOP-PCR) [30], improved primer extension preamplification PCR (IPEP-PCR) [31], and multiple displacement amplification (MDA) [32–38].

PCR-Based Primer Extension Methods

DOP-PCR is a WGA method first described in 1992 (Figure 1) [29]. This method involves the use of a primer with three distinct regions: a 3′ end with a specified sequence, a middle region with a stretch of six degenerate bases, and a 5′ end with a longer specified sequence. Using a small number of low-annealing temperature cycles, the short specified sequence on the 3′ end hybridizes to multiple sites of the template DNA and complementary sequences are generated. The low-annealing temperature used

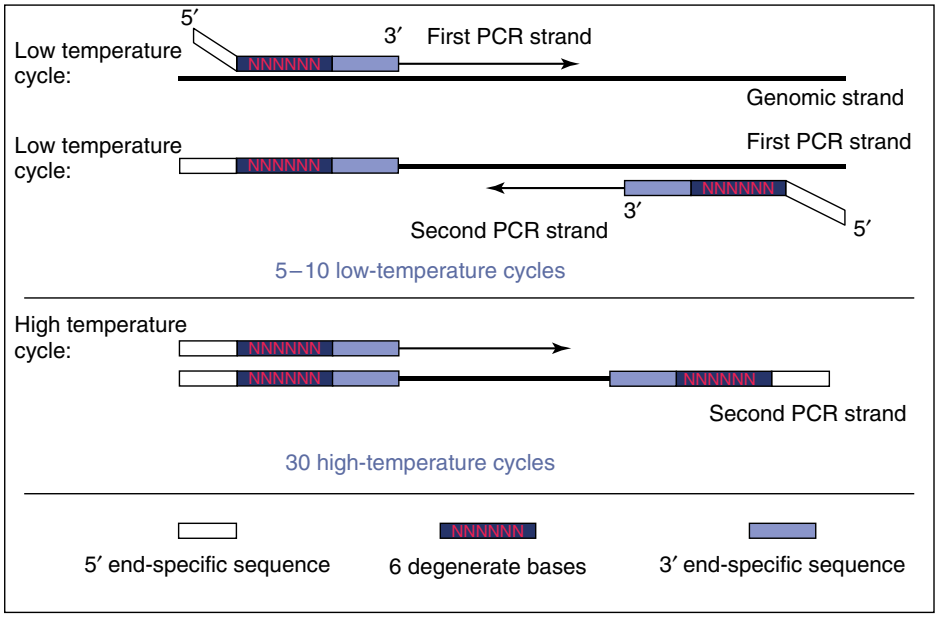


Figure 1 Schematic diagram of the degenerate oligonucleotide primed PCR process. Using a small number of low-annealing temperature cycles, the short specified sequence on the 3′ end of the primer hybridizes to multiple sites of the template DNA and complementary sequences are generated. Annealing of the 3′ end is then stabilized by the stretch of six degenerate bases adjacent to this region of the primer. The newly synthesized strands generated during the low-annealing temperature cycles serve as primer binding sites for subsequent cycles [Reproduced from Ref. 29. © Elsevier, 1992.]

in these initial cycles facilitates less stringent binding. Annealing of the 3′ end is then stabilized by the stretch of six degenerate bases adjacent to this region of the primer. The newly synthesized strands generated during the low-annealing temperature cycles serve as primer binding sites for subsequent cycles. These subsequent cycles utilize a higher annealing temperature to restrict nonspecific hybridization and allow for more specific binding based on the more stringent amplification conditions.

LL-DOP, or long products from low DNA quantities DOP, was developed to improve upon the original DOP method and provide better coverage [30]. The first modification was the use of a mixture of Taq DNA polymerase and Tgo polymerase, which was available commercially as the Expand High Fidelity Enzyme Mix (Roche Applied Sciences). It is reported that this mixture of enzymes has a threefold greater accuracy than Taq polymerase alone ($\sim 2 \times 10^{-4}$ misincorporation rate [22]). The 3–5′ exonuclease proofreading ability of the Tgo polymerase (later replaced with two polymerase) results in less sequence errors, which leads to a reduction in the number of truncated amplification products. It is reported that fragments up to 10 kb in length can be produced using this method. The LL-DOP method also incorporates increased annealing and extension times as well as increased annealing temperatures.

PEP was first described in 1992 and utilizes a mixture of random 15-base oligonucleotide primers to bind to various locations throughout the genome (Figure 2) [25]. Theoretically, 1×10^9 ($\sim 4^{15}$) primer sequences are possible using this length of primer, allowing for a plethora of sites within the genome to serve as primer binding sites. Extension from these primers then occurs with a large number of low-annealing temperature amplification cycles.

A modified version of the PEP method, IPEP, was described in 1999 and included two key changes [31]. The first modification was the addition of an elongation step prior to the denaturation step within each amplification cycle. Additionally, the mixture of Taq polymerase and a proofreading enzyme, the High Fidelity Enzyme mix, as described previously, was

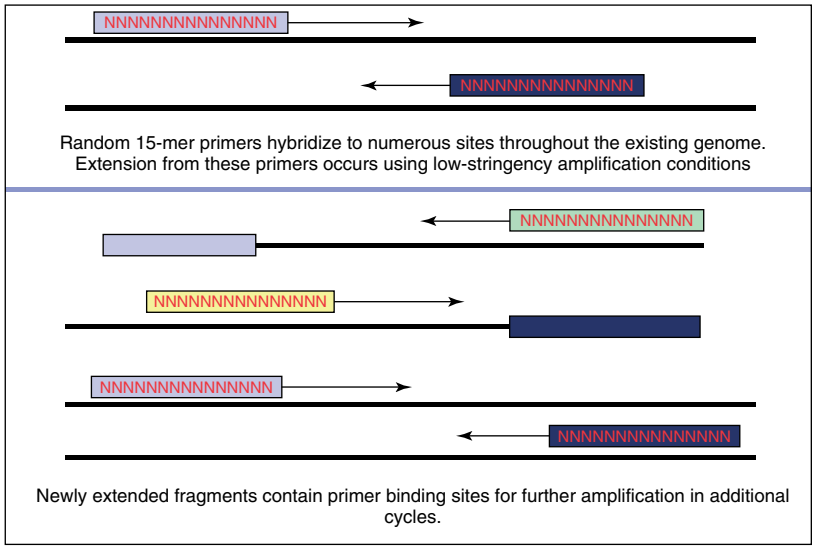


Figure 2 Schematic diagram of the primer extension preamplification process. Random 15-base oligonucleotide primers to bind to various locations throughout the genome. Extension from these primers then occurs with a large number of low-annealing temperature amplification cycles [Reproduced with permission from Ref. 25. © National Academy of Sciences, 1992.]

also utilized in the IPEP method. Further modifications to the IPEP method were made in 2005 [39]. A significant increase in the concentration of the Taq/Tgo enzyme mix and in the random 15-mer primer concentration resulted in the development of "modified improved primer extension preamplification (mIPEP)", which led to an improvement in the sensitivity of the IPEP method. The mIPEP method was also specifically designed for use with forensic casework samples.

Isothermal Amplification Reactions

An alternative WGA method is MDA, which uses the highly processive bacteriophage ϕ29 DNA polymerase [32]. Owing to the high processivity of this polymerase, primer extension from random hexamers allows products greater than 70 kb to be generated in an isothermal reaction. During this amplification process, ϕ29 DNA polymerase displaces the 5′ end of a newly synthesized DNA strand, thereby introducing a single-stranded product for further priming and extension (Figure 3). The MDA reaction is able to amplify DNA samples up to 100 000 fold and is reported to generate microgram quantities of DNA from picogram amounts of starting template.

Several amplification kits utilizing an MDA approach have been developed and are commercially available. The GenomiPhi™ v2 and the GenomiPhi™ HY Amplification kits are reported to result in the production of 4–7 μg or up to 50 μg of template DNA, respectively, in <2 h. The Repli-g Mini, Midi, and Ultrafast Mini Kit are available from Qiagen and are capable of producing 10 μg (Mini) to 40 μg (Midi). An isothermal-based reaction is also the basis for the GenomePlex® WGA products (Sigma-Aldrich). The GenomePlex® kits begin with a fragmentation of the input templates into a series of shorter, overlapping fragments. These fragments are then used to generate a library of DNA fragments, which is then amplified using a linear, isothermal amplification in the initial stages, followed by a limited round of geometric (PCR) amplifications using a universal primer to generate sufficient amounts of templates for downstream applications. Several GenomePlex® kits are available including the GenomePlex® Complete WGA kit (allows for amplification of a variety of starting materials), the GenomePlex® Single Cell WGA kit (reported to result in a millionfold amplification capable of producing microgram quantities of DNA from a single lysed

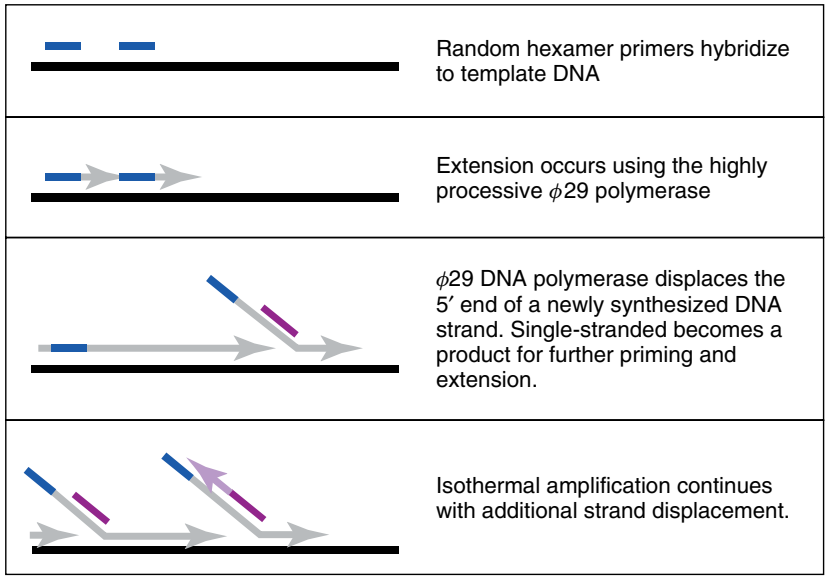


Figure 3 Schematic diagram of the multiple displacement amplification process. ϕ29 DNA polymerase displaces the 5′ end of a newly synthesized DNA strand, thereby introducing a single-stranded product for further priming and extension

cell), the GenomePlex® Tissue WGA kit (allows for amplification from fresh or frozen tissue), and also the GenomePlex® WGA Reamplification kit (allows for subsequent reamplification of the initial WGA product to produce large quantities of the product).

Genome Coverage

One of the main concerns regarding any WGA method is the successful amplification coverage of the genome present in the original sample with little amplification bias between or within commonly used genetic loci. The random nature of primers and the low stringency of amplification conditions utilized in most amplification strategies should theoretically allow for hybridization throughout the genome. Lower bias rates are typically reported for MDA (as low as sixfold between loci [32, 40]) compared to other PCR-based WGA methods (as high as $10^3 - 10^5$ [32, 41, 42]); however, individual reports vary widely, depending on the amount of input DNA used. Greater MDA bias is typically observed with smaller amounts of input DNA.

Many attempts have been made to use MDA to amplify low DNA quantity forensic samples. However, owing to the observed increase in amplification bias with smaller input amounts, attempts have been made to modify the MDA protocol to reduce this associated bias and increase the success of STR genotyping. It has been suggested that eliminating denaturation of the template DNA prior to MDA may result in a significant reduction in amplification bias [32, 43]. In a study conducted using the GenomiPhi™ amplification kit and subsequent autosomal STR typing, elimination of the template denaturation resulted in an increase in the allelic bias for those loci that were previously not showing a large amount of bias and resulted in a decrease in the allelic bias for those loci that were previously affected by amplification bias [43]. As a result, it was suggested that a pooling of template-denatured and template-nondenatured samples could be used in order to balance the effects at each of the included STR loci. In a separate study, again using the GenomiPhi™ amplification kit, the affects of molecular crowding (introducing additional sources of DNA into a single source DNA samples) on subsequent STR genotyping were investigated [44]. It was demonstrated that the incorporation of a larger amount of human DNA, animal DNA, or even glycol 400 as molecular crowding agents into low template samples resulted in an increase in the number of alleles detected.

Applications

WGA methods have aided in the development of LCN analysis of samples in the clinical research arena, although the impact of WGA in forensics has so far been relatively minor. However, recently, a few studies have been published involving the application of WGA methods to forensic samples, specifically the mIPEP and MDA methods.

PEP/IPEP/mIPEP

Early studies demonstrated the ability to obtain genetic profiles from several blood grouping loci (ABO, Lewis, and Pm) from PEP products [45, 46]. These studies were limited to relatively pristine samples using relatively large sample sizes (1–3 ng). More recent studies involving a broader range of forensic casework type samples evaluated the sensitivity and specificity of the modified PEP methods (IPEP and mIPEP). Autosomal STR and single nucleotide polymorphism (SNP) genotypes were obtained from genomic DNA samples preamplified using the IPEP method [47]. However, other studies have claimed that IPEP performed poorly with forensic samples including aged bloodstains, hair shafts, and bone samples [48]. mIPEP was successfully used to obtain autosomal and Y-chromosome STR profiles from forensically relevant samples, including buccal epithelial cells, vaginal epithelial cells, and sperm cells using only 5 pg of input DNA (~equivalent to 1 diploid or 2 haploid cells) [39]. STR profiles were also obtained from contact DNA samples (single dermal ridge fingerprints) that had been preamplified with the mIPEP method [39]. Additionally, partial Y- and autosomal STR profiles were obtained from mIPEP-treated DNA recovered from bloodstains exposed to the outside environment for one year, whereas non-mIPEP-treated samples did not produce profiles [39].

MDA

Significant effort has been extended in evaluating applications of the MDA method. With a reported ability to obtain microgram quantities of DNA from picogram amounts of input template DNA, the MDA method could have a significant impact on the analysis of forensic LCN samples. Additionally, several commercial kits are available that provide standardized protocols and optimized reagents. Successful MDA of various sample types including buccal cells, fresh and aged bloodstains, semen, hair shafts, and bone samples has been demonstrated using MDA [34, 36, 43, 47, 49]. Two commercially available MDA kits were used to determine if MDA could be used to improve the number of STR alleles recovered from degraded DNA samples [43]. While there was a difference between the two kits, both demonstrated the ability to recover additional alleles that were not obtained with non-MDA samples. The usefulness of this method in the analysis of forensic LCN samples has been demonstrated by the ability to amplify extracts from contact DNA samples such as fingerprints [34]. Recently, numerous studies have also reported the ability to analyze single cells after amplification with MDA [33, 50–53].

While reports of the ability to analyze single cells using the MDA method are encouraging for future forensic applications, there appears to be a large variation in the minimum amount of input template DNA required to obtain successful results in downstream analyses. There are numerous studies that provide evidence of problems with downstream analysis of MDA-amplified samples, even when greater input DNA amounts are used. Using 4–5 ng of input genomic DNA, a significant reduction in the ability to recover STR and SNP genotype data from MDA-amplified samples was obtained compared to the original DNA sample as a result of allelic loss and amplification bias [49]. Another study reported that the use of a minimum of 10 ng of template DNA is required to overcome the effects of the amplification bias associated with MDA [54]. Other studies have demonstrated that, even though the MDA method results in the production of microgram quantities of DNA (increased quantity compared to the sample samples when amplified with IPEP), there was no significant improvement in the downstream analysis of MDA-amplified forensic and artificially degraded samples, and oftentimes this resulted in the production of extraneous amplification products [48].

Selective Genomic Enrichment

Recently, a novel technology has been developed that allows for the selective amplification of human mitochondrial DNA (REPLI-g Mitochondrial DNA kit, http://www1.qiagen.com/Products/

GenomicDnaStabilizationPurification/replig/REPLIgMitochondrialDNAKit.aspx?ShowInfo=1) using an MDA-based approach. All of the previously developed WGA methods and commercially available kits are designed to amplify nuclear genomic DNA. However, this kit is reported to effect an enrichment of human mitochondrial DNA in total DNA samples with minimal contamination from nuclear DNA in a simple one-tube procedure. This eliminates the need for additional isolation or purification steps required for some downstream applications. This WGA-based selective enrichment technology has other promising applications with admixed forensic samples such as selective enrichment of Y-chromosome sequences in sexual assault cases.

Generation of Large Quantities of Starting DNA Template for Multiple Analyses

A potential application of WGA is to produce enough original starting template to permit multiple, including replicate, testing. The efficacy of such an approach assumes that the WGA treatment results in accurate replication of the starting templates. It is likely that the accuracy rates vary between different WGA methods and different starting template quantity and quality. Forensic samples are often limited in terms of quantity, and this limits the amount of DNA testing possible with these specimens. The fidelity of starting template replication caveat notwithstanding, more DNA tests could be carried out on the same specimen. For example, mixed samples that possess minor components showing stochastic variation in allelic signal intensity could be rerun multiple times after WGA pretreatment and a consensus profile of the minor contributor deduced. MDA was reported to provide enough starting DNA template to conduct 100 separate amplifications of distinct Alu elements in an attempt to predict the ethnoancestry of the DNA donor [55].

Summary

In an attempt to increase the recovery of profiles from LCN samples, WGA strategies have been developed. WGA methods attempt to increase the effective number of starting genomic templates prior to any downstream analysis. Although WGA methods have yet to demonstrate their utility for routine forensic analysis, selective genomic enrichment and the production of sufficient sample for multiple repeat analyses are possible novel applications.

References

[1] Brinkmann, B. (1998). Overview of PCR-based systems in identity testing, *Methods in Molecular Biology* **98**, 105–119.

[2] Carey, L. & Mitnik, L. (2002). Trends in DNA forensic analysis, *Electrophoresis* **23**(10), 1386–1397.

[3] Chimera, J.A. & Dyer, D. (1992). DNA analysis for human identification using the polymerase chain reaction, *Journal of Insurance Medicine* **24**(2), 109–113.

[4] Fukushima, H. (1999). Forensic DNA analysis – past and future, *Nihon Hoigaku Zasshi* **53**(3), 276–284.

[5] Gurvitz, A., Lai, L.Y. & Neilan, B.A. (1994). Exploiting biological materials in forensic science, *Australasian Biotechnology* **4**(2), 88–91.

[6] Hohoff, C. & Brinkmann, B. (1999). Human identity testing with PCR-based systems, *Molecular Biotechnology* **13**(2), 123–136.

[7] Jobling, M.A. & Gill, P. (2004). Encoded evidence: DNA in forensic analysis, *Nature Reviews. Genetics* **5**(10), 739–751.

[8] Nishimi, R.Y., O'Connor, K.W., Gwin, H.L. & Anderson, M.A. (1991). Genetic witness: forensic uses of DNA tests, *Journal International de Bioéthique* **2**(1), 29–32.

[9] Butler, J.M. (2006). Genetics and genomics of core short tandem repeat loci used in human identity testing, *Journal of Forensic Sciences* **51**(2), 253–265.

[10] Chakraborty, R., Stivers, D.N., Su, B., Zhong, Y. & Budowle, B. (1999). The utility of short tandem repeat loci beyond human identification: implications for development of new DNA typing systems, *Electrophoresis* **20**(8), 1682–1696.

[11] Edwards, A., Civitello, A., Hammond, H.A. & Caskey, C.T. (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats, *American Journal of Human Genetics* **49**(4), 746–756.

[12] Morling, N. (1998). Amplification of short tandem repeat loci using PCR, *Methods in Molecular Biology* **98**, 173–180.

[13] Tamaki, K. & Jeffreys, A.J. (2005). Human tandem repeat sequences in forensic DNA typing, *Legal Medicine* **7**(4), 244–250.

[14] Kirby, L. (1990). Specimens, *DNA Fingerprinting: An Introduction*, Stockton Press, New York.

[15] Hall, A. & Ballantyne, J. (2003). Novel Y-STR typing strategies reveal the genetic profile of the semen donor in extended interval post-coital cervicovaginal samples, *Forensic Science International* **136**(1–3), 58–72.

[16] Shewale, J.G., Sikka, S.C., Schneida, E. & Sinha, S.K. (2003). DNA profiling of azoospermic semen samples from vasectomized males by using Y-PLEX 6 amplification kit, *Journal of Forensic Sciences* **48**(1), 127–129.

[17] Sibille, I., Duverneuil, C., Lorin de la, G.G., Guerrouache, K., Teissiere, F., Durigon, M. & de Mazancourt,

P. (2002). Y-STR DNA amplification as biological evidence in sexually assaulted female victims with no cytological detection of spermatozoa, *Forensic Science International* **125**(2–3), 212–216.

[18] Schulz, M.M. & Reichert, W. (2002). Archived or directly swabbed latent fingerprints as a DNA source for STR typing, *Forensic Science International* **127**(1–2), 128–130.

[19] van Oorschot, R.A. & Jones, M.K. (1997). DNA fingerprints from fingerprints, *Nature* **387**(6635), 767.

[20] Gill, P., Whitaker, J., Flaxman, C., Brown, N. & Buckleton, J. (2000). An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA, *Forensic Science International* **112**(1), 17–40.

[21] Gill, P. (2001). Application of low copy number DNA profiling, *Croatian Medical Journal* **42**(3), 229–232.

[22] Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K. & Erlich, H. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase, *Science* **239**(4839), 487–491.

[23] Strom, C.M., Rechitsky, S., Wolf, G. & Verlinsky, Y. (1994). Reliability of polymerase chain reaction (PCR) analysis of single cells for preimplantation genetic diagnosis, *Journal of Assisted Reproduction and Genetics* **11**(2), 55–62.

[24] Strom, C.M. & Rechitsky, S. (1998). Use of nested PCR to identify charred human remains and minute amounts of blood, *Journal of Forensic Sciences* **43**(3), 696–700.

[25] Zhang, L., Cui, X., Schmitt, K., Hubert, R., Navidi, W. & Arnheim, N. (1992). Whole genome amplification from a single cell: implications for genetic analysis, *Proceedings of the National Academy of Sciences of the United States of America* **89**(13), 5847–5851.

[26] Barbaux, S., Poirier, O. & Cambien, F. (2001). Use of degenerate oligonucleotide primed PCR (DOP-PCR) for the genotyping of low-concentration DNA samples, *Journal of Molecular Medicine* **79**(5–6), 329–332.

[27] Buchanan, A.V., Risch, G.M., Robichaux, M., Sherry, S.T., Batzer, M.A. & Weiss, K.M. (2000). Long DOP-PCR of rare archival anthropological samples, *Human Biology* **72**(6), 911–925.

[28] Cheng, J., Waters, L.C., Fortina, P., Hvichia, G., Jacobson, S.C., Ramsey, J.M., Kricka, L. & Wilding, P. (1998). Degenerate oligonucleotide primed-polymerase chain reaction and capillary electrophoretic analysis of human DNA on microchip-based devices, *Analytical Biochemistry* **257**(2), 101–106.

[29] Telenius, H., Carter, N.P., Bebb, C.E., Nordenskjold, M., Ponder, B.A. & Tunnacliffe, A. (1992). Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer, *Genomics* **13**(3), 718–725.

[30] Kittler, R., Stoneking, M. & Kayser, M. (2002). A whole genome amplification method to generate long fragments from low quantities of genomic DNA, *Analytical Biochemistry* **300**(2), 237–244.

[31] Dietmaier, W., Hartmann, A., Wallinger, S., Heinmoller, E., Kerner, T., Endl, E., Jauch, K., Hofstadter, F. & Ruschoff, J. (1999). Multiple mutation analyses in single tumor cells with improved whole genome amplification, *American Journal of Pathology* **154**(1), 83–95.

[32] Dean, F.B., Hosono, S., Fang, L., Wu, X., Faruqi, A.F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J., Driscoll, M., Song, W., Kinsgmore, W., Kingsmore, S.F., Egholm, M. & Lasken, R.S. (2002). Comprehensive human genome amplification using multiple displacement amplification, *Proceedings of the National Academy of Sciences of the United States of America* **99**(8), 5261–5266.

[33] Jiang, Z., Zhang, X., Deka, R. & Jin, L. (2005). Genome amplification of single sperm using multiple displacement amplification, *Nucleic Acids Research* **33**(10), e91.

[34] Sorensen, K.J., Turteltaub, K., Vrankovich, G., Williams, J. & Christian, A.T. (2004). Whole-genome amplification of DNA from residual cells left by incidental contact, *Analytical Biochemistry* **324**(2), 312–314.

[35] Sussman, H.E. (2003). Whole-genome amplification easy as Phi: Amersham biosciences amplifies genomic DNA with a unique DNA polymerase, *The Scientist* **17**(16), 54.

[36] Thompson, M.D., Bowen, R.A., Wong, B.Y., Antal, J., Liu, Z., Yu, H., Siminovitch, K., Kreiger, N., Rohan, T.E., Cole, D.E. (2005). Whole genome amplification of buccal cell DNA: genotyping concordance before and after multiple displacement amplification, *Clinical Chemistry And Laboratory Medicine* **43**(2), 157–162.

[37] Tranah, G.J., Lescault, P.J., Hunter, D.J. & De Vivo, I. (2003). Multiple displacement amplification prior to single nucleotide polymorphism genotyping in epidemiologic studies, *Biotechnology Letters* **25**(13), 1031–1036.

[38] Watanabe, M.E. (2003). Advancing whole genome amplification: molecular staging makes inroads on both genomics and proteomics fronts, *The Scientist* **17**(1), 45.

[39] Hanson, E.K. & Ballantyne, J. (2005). Whole genome amplification strategy for forensic genetic analysis using single or few cell equivalents of genomic DNA, *Analytical Biochemistry* **346**(2), 246–257.

[40] Hosono, S., Faruqi, A.F., Dean, F.B., Du, Y., Sun, Z., Wu, X., Du, J., Kingsmore, S.F., Egholm, M. & Lasken, R.S. (2003). Unbiased whole-genome amplification directly from clinical samples, *Genome Research* **13**(5), 954–964.

[41] Jiao, Z.X., Zhuang, G.L., Zhou, C.Q., Shu, Y.M., Li, J., Liang, X.Y., Zhang, M.F. & Deng, M.F. (2003). Successful preimplantation genetic diagnosis for beta-thalassemia using primer extension preamplification, *Zhonghua Fu Chan Ke Za Zhi* **38**(3), 143–146.

[42] Xu, K., Tang, Y., Grifo, J.A., Rosenwaks, Z. & Cohen, J. (1993). Primer extension preamplification for detection of multiple genetic loci from single human blastomeres, *Human Reproduction* **8**(12), 2206–2210.

[43] Ballantyne, K.N., van Oorschot, R.A., Muharam, I., van, D.A. & John, M.R. (2007). Decreasing amplification bias associated with multiple displacement amplification and short tandem repeat genotyping, *Analytical Biochemistry* **368**, 222–229.
[44] Ballantyne, K.N., van Oorschot, R.A., Mitchell, R.J. & Koukoulas, I. (2006). Molecular crowding increases the amplification success of multiple displacement amplification and short tandem repeat genotyping, *Analytical Biochemistry* **355**(2), 298–303.
[45] Itoh, Y. & Kobayashi, R. (2000). Evaluation of ABO and Lewis genotypes using primer extension preamplification, *Forensic Science International* **113**(1–3), 139–141.
[46] Satoh, K., Takahashi, K., Itoh, Y. & Kobayashi, R. (1998). Typing of DNA using the primer extension preamplification (PEP) method – studies of reliability of typing and detection limits, *Nihon Hoigaku Zasshi* **52**(3), 184–190.
[47] Sun, G., Kaushal, R., Pal, P., Wolujewicz, M., Smelser, D., Cheng, H., Chakraborty, R., Jin, L. & Deka, R. (2005). Whole-genome amplification: relative efficiencies of the current methods, *Legal Medicine* **7**(5), 279–286.
[48] Barber, A.L. & Foran, D.R. (2006). The utility of whole genome amplification for typing compromised forensic samples, *Journal of Forensic Sciences* **51**(6), 1344–1349.
[49] Bergen, A.W., Haque, K.A., Qi, Y., Beerman, M.B., Garcia-Closas, M., Rothman, N. & Chanock, S.J. (2005). Comparison of yield and genotyping performance of multiple displacement amplification and OmniPlex whole genome amplified DNA generated from multiple DNA sources, *Human Mutation* **26**(3), 262–270.
[50] Handyside, A.H., Robinson, M.D., Simpson, R.J., Omar, M.B., Shaw, M.A., Grudzinskas, J.G. & Rutherford, A. (2004). Isothermal whole genome amplification from single and small numbers of cells: a new era for preimplantation genetic diagnosis of inherited disease, *Molecular Human Reproduction* **10**(10), 767–772.
[51] Hellani, A., Coskun, S., Benkhalifa, M., Tbakhi, A., Sakati, N., Al-Odaib, A. & Ozand, P. (2004). Multiple displacement amplification on single cell and possible PGD applications, *Molecular Human Reproduction* **10**(11), 847–852.
[52] Spits, C., Le, C.C., De, R.M., Van, H.L., Van, S.A., Liebaers, I. & Sermon, K. (2006). Optimization and evaluation of single-cell whole-genome multiple displacement amplification, *Human Mutation* **27**(5), 496–503.
[53] Spits, C., Le, C.C., De, R.M., Van, H.L., Van, S.A., Liebaers, I. & Sermon, K. (2006). Whole-genome multiple displacement amplification from single cells, *Nature Protocols* **1**(4), 1965–1970.
[54] Lasken, R.S. & Egholm, M. (2003). Whole genome amplification: abundant supplies of DNA from precious samples or clinical specimens, *Trends in Biotechnology* **21**(12), 531–535.
[55] Ray, D.A., Walker, J.A., Hall, A., Llewellyn, B., Ballantyne, J., Christian, A.T., Turteltaub, K. & Batzer, M.A. (2005). Inference of human geographic origins using Alu insertion polymorphisms, *Forensic Science International* **153**(2–3), 117–124.

JACK BALLANTYNE AND ERIN K. HANSON

Widmark Factor *see* Alcohol: Analysis

Wildlife

Prior to July 1, 1975, it was a rare event for a forensic scientist to apply the resources of a crime laboratory to wildlife- or animal-related evidence. When these applications did occur, it was mostly in an effort to determine the "family", "genus", or (ideally) "species" sources of bloodstains and loose hairs collected at human crime scenes. These identifications were generally based upon immunodiffusion tests using relatively nonspecific antisera, and microscopic comparisons utilizing small collections of known hairs from local museums or zoos. And the results were often more useful in eliminating a suspect (i.e., the blood on the suspect's shirt is not of human origin) than in trying to link suspect and victim to a specific crime scene.

During this time period, a few forensic scientists did apply their forensic protocols and tools to evidence seized by wildlife law enforcement officers; but their resources (lab space, instrumentation, databases, and "standards" collections) were extremely limited. And their efforts were mostly limited to the identification of blood, meat, and hair from locally hunted species, utilizing the same immunodiffusion and microscopic comparison techniques.

July 1, 1975 was a significant date for wildlife forensics in that 80 nations formally agreed to work together to enforce each other's endangered species laws via the establishment of CITES (The Convention

on International Trade in Endangered Species Fauna and Flora) (www.cites.org). This international agreement encouraged the member countries to enforce uniform import and export laws related to agreed-upon lists of endangered, threatened, and protected species. In doing so, the establishment of CITES brought forth an underlying forensic issue: the reality that illegal trafficking of regulated species would occur mostly in the form of parts and products, rather than whole animals; and that species-specific identifications of questioned items would be needed to enforce the CITES regulations in courts of law. Thus, the need for wildlife forensics on an international scale was born.

In 1986, in response to this need, and in an effort to support CITES, the United States established the National Fish and Wildlife Forensics Laboratory (NFWFL) in Ashland, Oregon (www.lab.fws.gov). The primary mission of this new laboratory was – and still is – to develop reliable wildlife forensic procedures and to provide forensic support to state, federal, and international wildlife law enforcement officers.

In 1993, at a meeting of the Environmental Crimes Group of Interpol in Lyon, France, the role of the NFWFL in assisting CITES and Interpol was documented in the form of signed letters of agreement. As a result of those treaties, the laboratory now works with scientists and law enforcement officers from the CITES and the Wildlife Subgroup of Interpol to continually develop and refine reliable methods of identifying wildlife parts and products and to link suspects and their animal victims to specific crime scenes.

Today, a wide range of protocols are available for the identification and comparison of wildlife related evidence. These protocols are typically divided into the following categories:

- morphology;
- molecular biology (Genetics);
- criminalistics;
- analytical chemistry;
- pathology; and
- digital evidence.

Morphology

Morphology is the study of structure and shape. In wildlife forensic science, morphological examinations of submitted evidence items are normally conducted by eye, and with the use of simple, compound, or scanning electron microscopes. These are often the simplest examinations performed in a wildlife crime lab; but at the same time, they address some of the most complex and difficult identification problems.

A protocol for a method of morphological identification would be difficult to write: the lack of standard definitions for individuals, and the fact that no two individuals (even genetic twins) are exactly alike. Two animals (e.g., two whitetail deer that are genetic twins) may start out looking very much alike, but the normal wear and tear that a young whitetail deer experiences literally from the moment of birth creates individual characteristics (a healed cut, a chipped hoof, or broken antler) that quickly separate those twins into distinct individuals.

Thus, the immediate problem for a wildlife forensic scientist is to come up with, in a manner similar to other forensic morphological identifications such as footmarks, "class characteristics" that distinguish and identify family, genus, and species of animals that are separate and distinct from population and individual characteristics. This is not a problem with whole animals; but it is very much a problem in the case of wildlife parts and products wherein the commonly occurring species-defining characteristics of the animal-source may not be present.

These identification characteristics typically fall within one of the following morphological categories:

- hair and fur;
- leather and hides;
- bones and skulls;
- teeth, claws, and beaks;
- hooves, horns, and antlers;
- feathers and down; and
- other miscellaneous parts.

Molecular Biology (Genetics)

Molecular biology involves the study of genetic information encoded in the DNA molecule, and the expression of that coding into proteins and related biological structures. Given the incredible diversity of biological structures present in the known plant and animal kingdoms, molecular biology offers the wildlife forensic scientist an extremely powerful tool to

1. determine family, genus, and species;
2. determine gender; and
3. individualize blood and tissue samples.

Family/Genus/Species ID

Prior to 1996, the forensic process of determining the species origin of an unknown tissue normally began with a series of screening (immunological) tests designed to narrow the possibilities down to the species comprising a single family (e.g., bears – family Ursidae or deer – Family Cervidae). Once the family source of the specimen is determined, the examiners went forward with either protein or DNA/PCR analysis (along with the necessary and comprehensive databases) to determine the actual genus and species involved.

In 1996, scientists at the National Fish and Wildlife Forensics Lab in Ashland, OR, utilized a Matrix-assisted laser desorption/ionization (MALDI) Mass Spectrograph to conduct research on a procedure whereby the species of a blood or meat sample is determined by the analysis of the isolated hemoglobin structure. The development and verification of this protocol made it possible to determine the species source of a blood or meat sample in a few minutes as opposed to hours or days. And the ability of the MALDI instrument to work with extremely small samples, along with the habit of hunters and poachers not to wash their hunting clothes during an entire hunting season (in the belief that the animals will smell – and be scared off by – the scent of soap) now enables a wildlife forensic scientist to sample perhaps a hundred blood spots at random from a set of hunting clothes, and shortly thereafter provide the game warden with a list of all the species the hunter "engaged with" during that hunting season; a very powerful investigative tool, indeed.

Gender ID

There are a number of nuclear DNA-based gender-determining tests available for blood and tissue samples from mammalian species. The tests generally use the polymerase chain reaction (PCR) to amplify specific sequences of the zinc finger-containing protein Y (ZFY) and the sex determining region Y (SRY) genes, both of which are located on the mammalian Y chromosome.

Individualization of Blood and Tissue

Early work on individualizing animal blood and tissue samples involved multilocus DNA probe hybridization techniques. However, new PCR methods for detecting single-locus short tandem repeat (STR) markers have been developed and applied in human forensic casework, demonstrating the technical feasibility of similar applications to animal species. Pending the arrival of new technologies, wildlife research and forensic laboratories focus a considerable amount of effort on the development of STR markers for determining the individual origins of wildlife evidence tissues.

Criminalistics

Wildlife forensic scientists frequently process evidence from an illegal hunt much in the same way that a police scientist works evidence from a homicide scene. In fact, the events associated with a typical illegal hunt often involve the following categories of "criminalistics" (or police forensic science) evidence:

- trace evidence;
- firearms;
- other weapons;
- impression marks;
- latent-prints; and
- questioned documents.

Trace Evidence

Trace evidence in an animal case can involve a wide range of materials. A classic example is a case in which a mountain lion was held captive for a period of time and then killed in an illegal "canned" hunt. As it turned out, the mountain lion tried to chew his way loose from the synthetic fiber ropes (two types were used by the suspects to secure the lion), and a forensic scientist was able to link the fibers from the lion's stomach back to the chewed ropes found at the crime scene.

Firearms

The typical circumstances in which an animal is killed with a firearm vary greatly from those of a homicide case. The most significant differences include the following:

- the distance from the suspect to the victim;
- the choice of firearm;
- the ability of the suspect to "clean up" the scene;
- the tendency of the suspect to take the victim from the scene; and
- the tendency of the suspect to frequently reuse the same firearm.

Unlike human crime situations in which the victim is most commonly killed with a pistol at short – "contact" to 25 yard – distances, the typical animal kill involves a high-powered (and large caliber) rifle or a shotgun at relatively long distances (50–300 yards).

Given the nature of the typical hunting area (brush, trees, and ground cover), it is often difficult for an illicit hunter to retrieve his expended casings; however, the long shooting distances and the fact that the shot could have come from any 360° vector point makes it extremely difficult for a crime scene investigator to locate the shooting point, much less the expended casings. However, all of these advantages (to the illicit hunter) tend to be negated by two simple facts:

1. The point of the illicit hunting is for the suspect to take the victim (as a trophy or meat) back to his residence. Thus, the bullet is likely to either be in the carcass of the animal, or in the "gut pile" left at the scene (which can be matched to the trophy head or meat with DNA techniques).
2. The typical illicit hunter spends a lot of money on his rifle or shotgun, and rarely willing to discard his weapon after a single illicit kill. Thus, it is very likely that a succession of illegal kills can be linked to a single poacher by matching the spent bullets or casings to his rifle or shotgun.

Other "Weapons"

Other hunting weapons typically associated with an illicit animal kill include the following:

- long bows and arrows;
- crossbows and "bolts";
- spears;
- spring traps;
- poison discharge devices; and
- nets.

Impression Marks

The fact that most illicit hunting situations occur in remote areas or "off-road" situations makes it extremely likely that the suspect leaves tire tracks and boot impressions in soil, mud, or snow. And the fact that these tires and boots are typically used in off-road situations makes it all the more likely that the tire or boot treads possess individualistic wear marks.

Latent Fingerprints

Latent fingerprints are the classical means of linking the suspected, the victim, and the crime scene through physical evidence. The following types of latent-bearing evidence are frequently submitted to a crime lab in wildlife cases:

- firearms;
- expended casings (shotgun and rifle);
- knives;
- "no trespassing" and "no hunting" signs;
- game tags; and
- import/export (CITES) permits.

Questioned Documents

Questioned documents are frequently encountered in wildlife investigations involving the import and/or export of wildlife parts and products. The question most frequently by the investigator is whether or not the seized documents (typically import/export permits) are valid. This can be an extremely difficult question to answer when the authorizing seals vary between countries, the names of individuals authorized to approve import/export documents change on a frequent basis, and the shipments must be "cleared" (the documents examined and approved) at the local port of entry.

Questioned documents typically associated with a wildlife case include the following:

- forged or altered hunting licenses;
- forged or altered game tags; and
- forged or altered import/export permits.

Analytical Chemistry

Chemical analysis techniques are most often used in a wildlife crime lab to provide toxicological information for a veterinary pathologist conducting a necropsy. As such, a scientist assigned to the chemistry section of a wildlife crime laboratory spends a great deal of time examining blood, urine, tissue,

and stomach/crop contents in a search for pesticides and poisons. Analytical chemistry procedures (that use chemical biomarkers) are also used to identify the species source of animal products such as bear bile and deer musk; and standard toxicological methods are routinely employed to identify chemical baits and poisons used to trap and kill wildlife.

Pathology

Veterinary pathologists are responsible for determining cause of death of an animal carcass submitted as evidence. This is accomplished thorough necropsy (autopsy) protocols involving a search for lethal wounds cause by bullets, arrows, spears, and traps; a comprehensive toxicological workup of blood, urine, tissue, and stomach/crop contents to eliminate or confirm a poison or contaminant cause of death; and a professional evaluation of the underlying health of the animal prior to death. In the process of conducting these examinations, the veterinary pathologist also searches for signs of disease that may indicate a natural cause of death.

Issues that often complicate a cause-of-death determination in an animal (but should not impact the results of a careful and professional necropsy examination), include the following:

- The possibility that the animal might have been struck by additional (nonlethal or crippling) bullets, pellets, or other projectiles on days, months, or years prior to the questioned incident.
- The possibility that an illicit bow-hunter has shot the animal with a firearm first (because of the difficulty in getting close to an alert animal), and then sticking an arrow into the bullet wound.
- The possibility that the animal was killed or fatally weakened by a modern pesticide or poison designed (as the result of environmental protection laws) to decompose rapidly after a few hours of exposure to air or sunlight.
- The likelihood that scavengers would have destroyed a considerable amount of useful blood, tissue or bone evidence if the carcass was not found and collected in a timely manner.

In conducting necropsies involving bullet wounds, it is often extremely important that the veterinary pathologist determine the trajectory of the bullet into or through the body. This information may resolve the question of where the accused hunter was properly defending himself against a charging animal, or illegally hunting a protected species. The information may also be used by investigators, during the interview process, in determining the veracity of suspects and witnesses.

Digital Evidence

Modern poachers add digital cameras to their lists of equipment, and take hundreds of digital photos to share (over the Internet) with their fellow violators. It is now a rare warrant search of a poacher's residence or business that does not include the seizure of at least one computer, or (often in the case of a business that the investigators are not allowed to shut down) the "mirroring" of the computer's hard drive data into the investigator's portable computer.

Further Reading

Aasen, E. & Medrano, J.F. (1990). Amplification of the ZFY and ZFX genes for sex determination in humans, cattle, sheep and goats, *Biotechnology* **8**, 1279–1281.

Adamczyk, M. & Gebler, J.C. (1997). Electropspray mass spectrometry of a and b chains of selected hemoglobins and their TNBA and TNB conjugates, *Bioconjugate Chemistry* **8**, 400–406.

Andrasko, J. & Rosen, B. (1994). Sensitive identification of hemoglobin in bloodstains from different species by high performance liquid chromatography with combined UV and fluorescence detection, *Journal of Forensic Sciences* **39**(6), 1018–1025.

Ashton, G.C. (1958). Beta-globulin polymorphism in cattle, sheep and goats, *Science* **182**, 945.

Brohn, A. & Korschgen, L.J. (1950). Precipitin test – a useful tool in game law enforcement, *Transactions of the 15th North American Wildlife Conference* 467–478.

Clarke, F.D. (1914). Forensic value of the precipitin test in the enforcement of game laws in California, *University of California Publications in Pathology* **2**, 131.

Cowan, I.McT. & Johnston, P.A. (1962). Blood serum protein variation at the species and subspecies level in deer of the genus Odocoileus, *Systematic Zoology* **II**(3), 131.

Cronin, M.A., Vyse, E.R. & Cameron, D.G. (1988). Genetic relationship between muledeer and white-tailed deer in Montana, *The Journal of Wildlife Management* **52**(2), 320.

Dilworth, T.G. & McKenzie, J.A. (1970). Attempts to identify meat of game animals by starch-gel electrophoresis, *The Journal of Wildlife Management* **34**(4), 917.

Espinoza, E.O., Baker, B.W. & Berry, C.A. (2007). The analysis of sea turtle and bovid keratin artifacts using DRIFT spectroscopy and discriminant analysis, *Archaeometry* **49**(4), 685–698.

Espinoza, E.O., Cech Lindley, N., Gordon, K.M., Ekhoff, J.A. & Kirms, M.A. (1999). Electrospray ionization mass spectrometric analysis of blood for differentiation of species, *Analytical Biochemistry* **268**, 252–261.

Espinoza, E.O., Kirms, M.A. & Filipek, M.S. (1996). Identification and quantitation of source from hemoglobin of blood and blood mixtures by high performance liquid chromatography, *Journal of Forensic Sciences* **41**(5), 804–811.

Fain, S.R. & LeMay, J.P. 1995). Gender identification of humans and mammalian wildlife species from PCR amplified sex-linked genes. *Proceedings of the 47th American Academy of Forensic Sciences*. Vol. 1, p. 34.

Gay, F.P. (1908). A contribution to the forensic value of the musculo-precipitin test, *The Journal of Medical Research* **19**(1), 219.

Grabar, P. & Burtin, P. (1964). *Immuno-Electrophoretic Analysis*, Elsevier Publishing, New York, p. 97.

Fenstermacher, R. & Pomeroy, B.S. (1938). Identification of suspected beaver blood stains, *Cornell Veterinarian* **28**(3), 257.

Irwin, D.M., Kocher, T.D. & Wilson, A.C. (1991). Evolution of the cytochrome b gene of mammals, *Journal of Molecular Evolution* **32**, 128–144.

Jackson, C.F. (1962). *Use of Paper Chromatography In Identifying Meat Of Game Animals*, New Hampshire Fish and Game Department, Tech. Circ. 19.

Karpas, A.B., Wyers, W.L. & Segre, D. (1970). Serologic identification of species of origin of sausage meats, *Journal of Food Science* **36**, 150.

Lawrence, B. (1951). *Post-Cranial Skeletal Characteristics Of Deer, Pronghorn And Sheep-Goat, With Notes On Bos And Bison, Part II Harvard*, Papers of the Peabody Museum of Archaeology and Ethnology, Vol. 35(3).

Moore, T.D., Spence, L.E. & Dugnolle, C.E. (1974). *Identification Of The Dorsal Guard Hairs or Some Mammals Of Wyoming*, Wyoming Game and Fish Department, Laramie, Bulletin No. 14.

Myers, M. (1900). Experiments upon the new specific test for blood. Preliminary note, *British Medical Journal* **1**, 1141.

Nuttall, G.H.F. (1904). *Blood Immunity And Blood Relationship*, Cambridge University Press.

Oates, D.W., Brown, C.W. & Weigel, D.L. (1974). *Blood and Tissue Identification Of Selected Birds And Mammals*, JPR Study XII, Project W-38-R, Nebraska Game and Parks Commission, Lincoln.

Ouchterlony, O. (1948). Antigen-antibody reaction in gels, *Acta Pathologica et Microbiologica Scandinavica* **25**, 186.

Reichert, D.T. & Brown, A.P. (1909). *The Differentiation And Specificity Of Corresponding Proteins And Other Vital Substances In Relation To Biological Classification And Organic Evolution: The Crystallography Of Hemoglobins*, Carnegie Institute of Washington.

Sibley, C.S. & Hendrickson, H.T. (1970). A comparative electrophoretic study of avian plasma proteins, *The Condor* **72**(1), 43.

Straughan, D.J., Burnham-Curtis, M.K. & Fain, S.R. (2002). Experimental search for forensically useful markers in the genus *Scaphirhynchus*, *Journal of Applied Ichthyology* **18**, 621–628.

Taylor, A.J., Linforth, R., Weir, O., Hutton, T. & Green, B. (1993). Potential of electrospray mass spectrometry for meat pigment identification, *Meat Science* **33**, 75–83.

VanTets, P. & Cowan, I.Mc.T. (1966). Some sources of variation in the blood sera of deer (Odocoileus) as revealed by starch gel electrophoresis, *Canadian Journal of Zoology* **44**, 631.

Wasserman, A. & Schutze, A. (1901). Ueber eine neue forensische Methode zur Unterscheid ung von Menschen und Thierblut, *Berliner klinische Wochenschrift* **38**, 187.

Wolfe, H.R. (1939). Serologic relationships among bovidae and cervidae, *Zoologica* **24**, 309.

Wolfe, H.R. (1933). Factors which may modify precipitin tests in their applications to zoology and medicine, *Physiological Zoology* **6**, 55.

KEN GODDARD

Witness *see* Eyewitness: Suggestibility of

Witness: Child *see* Children: as Witnesses

Witness: Elderly *see* Elderly in Court

Wood

Geographical Extent and Distribution of Forests

According to the Food and Agriculture Organization [1], 30.3% of the land area of the earth is covered by forest. This area has been further classified into 13 temperate and boreal forest types and 15 tropical

forest types. It is estimated that there are between 80 000 and 100 000 tree species worldwide. Each species has a set of identifying features of potential forensic value. These include

ecological/genetic factors,
microscopic structure,
dendrochronology, and
external appearance of timber.

The geographical distribution of forest species is only the first stage in the understanding of the importance of our use of wood. The timber- and wood-processing industries are substantial. In 2000, the worldwide production of timber was worth US $342 billion and represented 1.6% of world gross domestic product [2]. Wood and wood products can be found everywhere. This wide availability ensures that wood can be a valuable forensic tool.

Value of Wood

The forensic value of wood ranges from wood shavings found on the gloves of a suspect [3] to sections of ladders found at the crime scene in the Lindbergh kidnapping [4]. The analysis of wood is an important part of accident investigation, engineering reports, boundary disputes, archaeology as well as traditional crime cases where the suspect may be linked to a crime scene.

Ecological/Genetic Factors

The 28 forest types are classified according to ecological conditions. Trees are adapted to a variety of site factors. These include soil type, climate, altitude, and latitude. Acid soils, for example, will preclude trees that are adapted to alkaline soils. It is very unlikely that shoes covered in alkaline soil and owned by a suspect come from a coniferous plantation with acid soil. Similarly, it would be surprising to find tropical species in a temperate climate such as that as experienced in the United Kingdom or Norway. Such an unusual occurrence may suggest that the tropical sample was taken from a specialist hothouse and did not occur naturally.

Forest types are not usually monocultures. The Scottish Caledonian Pine Forest contains Scots pine, birch, rowan, aspen, juniper, oak as well as certain ferns, mosses, insects, and mammals. Any combination of these species might suggest that the forensic samples may be indicative of a Caledonian Pine Forest.

Origin is defined as the natural genetic location of a tree species. Sometimes seed is collected from a particular origin and exported to other sites around the world. These new trees then acquire a provenance. Thus, Sitka spruce seed from *Queen Charlotte Islands*, British Columbia, has been imported into the United Kingdom. The resulting progeny has a provenance referred to as *Queen Charlotte Islands*. A provenance may have slightly different growth characteristics than either the original stock or other provenances. This is thought to be as a result of phenotypic plasticity. Such a phenomenon may be of use in identification. Thus. a provenance of *Queen Charlotte Islands* may flush later, and is more frost hardy in Britain, than a provenance of *Washington*, which comes from a more southerly latitude. In practical terms, differences in provenance will usually be identified by differences in bark appearance, stem form, branchiness, and timber quality all within the same species.

The use of ecological factors in a forensic context is likely to be circumstantial. It is possible in some cases to rule out suspects at an early stage of an investigation, but it is unlikely that definitive proof of association can be obtained. It is likely that the testing of tree DNA will be used much more in the future to protect forest reserves. It is estimated that the loss of income to governments from illegal logging is between $10 and 15 billion worldwide [5]. The use of microsatellite DNA markers is already being used to identify trees and timber that have been stolen from protected reserves [6]. It is estimated that in Great Britain alone out of some 40 000 subsidence cases 75% are mainly due to woody material such as tree stumps invading adjacent property. In 1999, this amounted to over £300 million in claims.[a]

Forensic dendrologists are appointed to investigate such things as boundary disputes where trees are either the cause of damage or are themselves damaged. Cases of subsidence caused by trees and trees as nuisance value are also considered in this category. Road accidents where it has been alleged that trees are responsible or when road signs are obscured by trees are also topical. The condition, size, and health of the trees are usually assessed and taking measurements of height and estimations of root mass

are common. The engineering value of wood can be assessed by experts to see whether the wood is fit for purpose.

The use of DNA profiling is a useful but not foolproof technique for the identification of individual wood samples. In the case of poplar, willow, or walnut clones, for example, DNA profiling may identify the clone but not the individual. Samples taken from old furniture or heavily contaminated wood may have an incomplete DNA profile. Wood is often dead and may not contain any viable DNA. DNA databases are not complete and the analysis of some examples may be difficult to justify in court. Other techniques that can be used alone or in conjunction with DNA profiling to achieve accurate identification are available.

Microscopic Structure

The microscopic investigation of wood is perhaps the most useful alternative technique available to the forensic scientist. Not only can timbers be compared and identified but also engineered timber products such as cardboard, papers, and other processed timbers can be analyzed and identified. Furthermore, the analysis can be thorough and supported by strong statistical methods.

Wood (xylem) is a natural material with predictable structural variation. The variation is determined mainly by species though the location of the sample from the original log is also relevant. Logs consist of three sections: heartwood, sapwood, and bark. The heartwood is inactive but the sapwood is used for conduction. Wood consists of bundles of cellulose tubes (cells) glued together with lignin. In conifers (softwoods), these cells are known as *tracheids* and are typically 2–4 mm long and approximately 30 μm wide. They lie parallel to the stem and are used for support as well as conduction. They are characterized by having thin cell walls and open channels. Macerated tracheids can be seen in Figure 1.

In broadleaved species (hardwoods), these cells are differentiated into fibers and vessels. Fibers are rather thick walled and closed and are normally used only for support. They are typically shorter and thinner than tracheids, which are about 1–2 mm long and 15 μm wide. Vessels, which are between 0.2 to 1.2 mm long and less than 0.5 mm in diameter, are used for storage of water and minerals. Lignin is a resin found in the cell wall of plant cells that acts as glue, which holds together the tracheids, fibers, and vessels. It strengthens the overall structure of the wood. An example of oak vessels can be seen in Figure 2.



Figure 1 Macerated tracheids

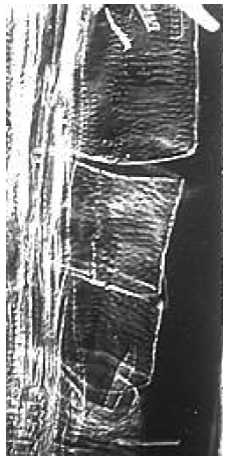


Figure 2 Oak *Quercus robur*. A decomposing vessel. The body of the long tube consists of short vessel elements that are stacked end to end

There are other structures within wood that can also be used for identification [7]. These include collenchyma and sclerenchyma, which are supporting tissues, and parenchyma, which is used mainly for storage. Medullary rays, pits, resin ducts, and pores can also be used for identification purposes. These traits vary with species.

The scientific analysis of wood has traditionally concentrated on the examination of fiber length. There is a correlation between the engineering performance of wood and the distribution of fiber length. The stiffness, hardness, durability and flexibility of timber are determined by the fiber structure. The importance of fiber length is even more evident in the wood-processing sector. The pulping process involves the extraction of the tracheids from the wood. These tracheids will later be reassembled as paper. In this context, softwood tracheids are often referred to as *fibers*. The wood is first chipped and then macerated so that the lignin can be removed and the fibers released. If the wood is rich in lignin, then the maceration process must be more active. This may include a bleaching process that is potentially hazardous to the environment. If the chosen species of tree has short fiber lengths, the resulting paper will tear easily. Long fiber lengths are associated with stronger papers. Manufacturers are concerned with producing an homogeneous product. This requires close monitoring of fiber length, which is one of the main quality procedures. Research into fiber length is therefore highly advanced. The examination of fiber length in papers, cardboard, liner paper, fiber board, and wood wool can be a useful form of identification.

Wood maceration consists of dissolving the lignin that cements wood fibers together. Various methods of wood maceration have been reviewed [8]. Various laboratory techniques have been used. They include Schulze's method [8], which consists of mixing nitric acid with potassium chlorate. This solution is often aggressive, especially when heated, so that some of the fibers are actually destroyed. An alternative milder treatment method known as *Jeffrey's method* [8] is used to mix equal portions of nitric acid and chromic acid together. Again this is heated to quicken the reaction. Franklin [8] proposed mixing together acetic acid and hydrogen peroxide. This method is effective but it can be slow. It is a particularly mild technique that does not damage the fibers appreciably. The Spearing and Isenburg [8] method uses a mixture of sodium chlorite and acetic acid. The Burkhart method [8] uses a mixture of boiling triethylene glycol mixed with a catalyst.

The choice of maceration process ultimately depends on the durability of the wood fibers, the time available and the size of the sample. Large samples of hard and dense timber, such as teak, may require a longer and harsher maceration process than say a soft pine. In the author's opinion it is probably best to start of with Franklin's method or Jeffrey's method as they are both predictable and controllable.

Industrial maceration techniques include Kraft pulping or the sulfate process [9]. The lignin from the wood is extracted by following a cooking process in digesters using caustic sodium hydroxide and sodium sulfide. This process has been designed to produce several useful by-products including the production of steam, electricity, soaps, and turpentine. The sulfite process is an alternative industrial method. Neither industrial method is appropriate for forensic investigations.

The results of the maceration process consist of an opaque liquid containing the separated fibers. This liquid can be washed, dried, and then mounted on a microscopic slide. Staining may sometimes improve clarity. The fibers are then physically separated and the lengths measured.

The objective of the analysis is the examination of fiber lengths. It might be expected that there would be a normal distribution of fiber lengths. Such an event would place great relevance to the mean value. In practice, however, some of the fibers are damaged and do not reflect the original length [10]. In these cases, the distribution is skewed or truncated because of the addition broken fibers. Almost all wood samples contain fibre damage. This could be due to saw damage, the chipping process, and natural fracturing. The maceration process itself could cause fiber damage. A reliance on mean fiber length could be misleading if the distribution is skewed due to damaged fibers. The examiner must take account of broken fibers either by excluding them from further examination or by assessing the actual length distribution for all of the fibers. In the latter case, the analysis is a form of pattern fitting. In effect, instead of comparing two normal distributions, the examiner would be comparing two truncated normal distributions or two skewed normal distributions [10]. The analysis becomes an estimate of fiber length distribution. A typical distribution of fibers and broken fibers can be seen in Figure 3.

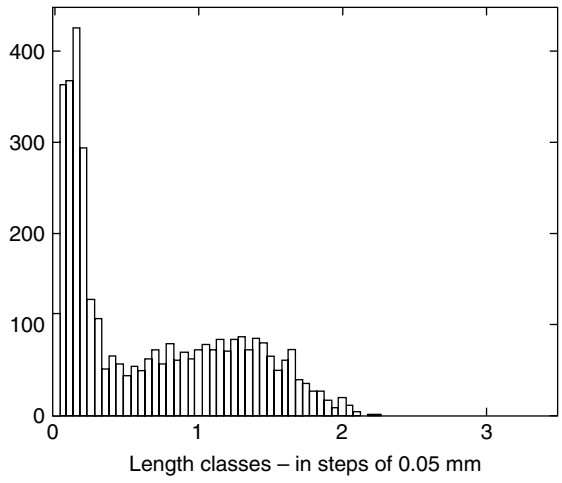


Figure 3 Histogram of fine and fiber length data from a sample from a 22-year-old Scot pine tree from Northern Sweden

It has been found that fiber length distributions have considerable intraspecies variation. Furthermore, variation exists within individuals. Fiber lengths will vary depending on where the sample is taken from the tree. Fibers closer to the core will differ from those nearer to the bark. This need not alarm the investigator because the comparison of fiber length will be between wood that has grown from the same part of the tree as the sample. A recent case in Texas illustrates this point [11]. A splintered pool cue had been used on someone's skull. Samples from the murder weapon were compared with splinters found in the suspect's car. Microscopic analysis confirmed that the splinters in the suspect's car came from a particular type of pool cue found at the crime scene. The microscopic structure of the splinters found in the suspect's car was identical to the structures found on the splintered part of the pool cue.

Good results can be achieved by measuring fiber lengths using a microscope. A less time-consuming method is by the use of an industrial fiber analyzer. Such instruments are designed for the pulp and paper industries. They automatically produce accurate morphological data on all aspects of fibers. They contain statistical analysis software including useful graphical data for distributions. Further research is required to investigate the full range of features that may be useful for forensic investigations.

Dendrochronology

Dendrochronology is the study of tree rings. Tree rings, often known as *annual rings*, mark the passage of growth of the tree throughout the year. They can be seen in the horizontal cross section cut through the tree trunk. Tree rings are common in temperate zones where the seasons are differentiated. They consist of spring wood (early wood) and summer wood (latewood). The spring wood grows fast and tends to be less dense than the summer wood. The comparative widths of both spring and summer wood are determined by the growing conditions existing at the time. Poor summers, for example, are reflected by narrow summer tree rings. All of the trees in a given locality have the same growth patterns. Given a sufficient period of growth, dendrochronologists can estimate variations in past weather by taking timber cores. The use of dendrochronology has expanded in recent years. Climatologists now use tree ring data to monitor changes in climate. Radiocarbon dating is often calibrated by using accurate dating of tree rings sometimes going back 10 000 years. The study of past events such as insect behaviour, wild fires, volcanology, storms, archaeology, and glaciology are all being enhanced by the use of dendrochronology. Tree ring data allows the researcher to cross reference past events accurately. This may have some value especially in the field of forensic archaeology.

External Appearance of Timber

Wood is the secondary xylem of trees. Xylem is used to transport water and nutrients throughout the tree. This is facilitated by transpiration and root pressure. The structure and appearance of xylem are largely determined by ecology and growing conditions. The proportion and density of spring wood to summer wood have already been discussed in relation to dendrochronology. Other aspects of structure such as the appearance of imperfections such as knots and grain effects can also be used for comparison and analysis.

Variations between species and within species pose a problem for the users of wood. Wood is

an heterogeneous material. The natural variation within wood must be accounted for by engineers and other users. Factors such as strength, hardness, modulus of elasticity, durability, tensile strength, and density are variable. Timbers are selected and graded to ensure that specific qualities are selected. For example, structural timbers are produced to meet international standards. Narrows variation down to within acceptable limits. This is usually based on minimum standards. Variation in standards aids identification and differentiation.

The external appearance of timber provides an obvious means of identifying, comparing, and matching. Timber is normally identified by looking at three dimensions or planes. These are illustrated in Figure 4.

The end-on view of a log is referred to as the *transverse section*. The main stem has been cut perpendicular to the grain and the spring and summer wood can be seen as a series of roughly concentric circles. The tangential plane is often found on the surface of veneered timbers. The wavy pattern of annual rings is chosen to enhance the attractiveness or figure of the wood. The radial plane accentuates the grain of the wood that lies perpendicular to the annual rings. The wood is usually very resistant to wear because the percentage of summerwood is maximized in the profile.

Samples of wood may be presented in one or more of these planes. Identification may be enhanced if similar samples that possess matching grain, frequency, and size of knots, as well as imperfections such as regular nail marks and colour are found. Associated paints, varnishes, and oils can be chemically analyzed as can the comparison of tool marks. Splinters of wood that do not conform to species guides because of the smallness of the sample can be identified using microscopic analysis.

The analysis of wood and wood products often requires techniques ranging from DNA profiling to good old-fashioned microscopic observation. Often these techniques must be used together to obtain the maximum amount of identifying information. It is likely that there will be significant growth in this area in the future, especially in civil cases, though one can never discount the need for such expertise in criminal cases.

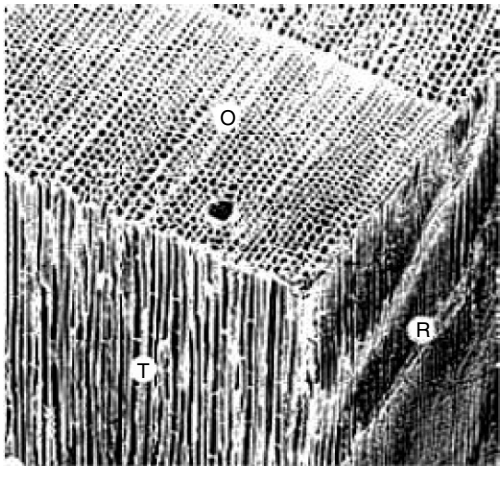


Figure 4 Spruce wood (a coniferous wood example) displaying the connection of cross-, radial, and tangential sections

End Notes

[a.] Association of British Insurers Report (1999).

References

[1] FAO (2005). *Global Forest Resources Assessment 2005*, FRA.

[2] Lebedys, A. (2004). Trends and current status of the contribution of the forestry sector to National economies 2004 FAO Working paper: FSFM/ACC/07.

[3] Comptroller and Auditor General (2003). *Improving Service Delivery*, The Forensic Science Service, HC 523 Session 2002–2003.

[4] Federal Bureau of Investigation. *Famous Cases. The Lindbergh Kidnapping*, FBI.gov; [updated 2007 July 16; cited 2007 July 16], http://www.fbi.gov/libref/historic/famcases/lindber/lindbernew.htm.

[5] The World Bank (2002). *Revised Forestry Strategy*, The World Bank, Washington, DC.

[6] Stacy, E.A., Dayanandan, S., Dancik, B.P. & Khasa, P.D. (2001). Microsatellite DNA markers for the Sri Lankan rainforest tree species, *Shorea cordifolia* (Dipterocarpaceae), and cross-species amplification in *S. megistophylla*, *Molecular Ecology Notes* **1**(1–2), 53–54.

[7] University of Hamburg (2007). *Botany Online. The Internet Hypertextbook*, University of Hamburg, http://www.biologie.uni- hamburg.de/b-online/e00/contents.htm.

[8] Han, J.S., Mianowski, T. & Lin, Yi-Yu. (1999). *Kenaf Properties, Processing and Products*, Mississippi State University, Ag & Bio Engineering, pp. 149–167.

[9] Gustafson, R.R., Sielcher, C.A., McKean, W.T. & Finlayson, B.A., (1983). Theoretical Model of the Kraft Pulping Process, *Industrial and Engineering Chemistry Process Design and Development* **22**, 87–96.
[10] Morling, T., Sjostedt-de Luna, Svensson, I., Fries, A., Ericsson, T. (2003). A method to estimate fibre length distribution in conifers based on wood samples from increment cores, *Holforschung* **57**, 248–254.
[11] Milius, S. (2002). The wood detective: the cases of the sunken pirate ship, the misunderstood antique, and the wicked pool cue, *Science News* **162**(12), 184–185.

Related Articles

Amplified Fragment Length Polymorphism

DNA

Environmental Science

Fibers

Microscopy: High Power

Microscopy: Low Power

Mitochondrial DNA: Profiling

Microscopy: Scanning Electron Microscopy

Microsatellites

Microscopy: FTIR

Phenotype

Palynology

ANTHONY M. COWELL

Workplace Violence *see* Threat Assessment: Workplace

Wounds: Gunshot *see* Gunshot Wounds

Wounds: Sharp Injury

Introduction

The term "wounds due to sharp force" comprises injuries that are caused by pointed and/or sharp-edged implements. The main categories are stab wounds, incised wounds (cuts, slashes), and chop wounds.

In stabs, the causative object penetrates along its longitudinal axis producing a wound track that is longer than the skin severance at the entrance site. Incised wounds are inflicted by a cutting edge that slices the affected tissue when the instrument is drawn over the body surface, exerting sufficient pressure onto it; the resulting injuries are longer than they are deep. Chop wounds are caused by heavy metallic implements with a cutting edge such as axes or machetes; their outer appearance resembles cuts, often accompanied by corresponding fractures of the underlying bones.

Epidemiology

Sharp-force trauma is a major topic both in clinical forensic medicine and forensic pathology. In non-fatal bodily injuries, the study material includes a great variety of constellations such as survived knife attacks, attempted suicides (e.g., by wrist cutting), other kinds of deliberate self-infliction (simulation of a criminal offense, self-mutilation for the purpose of insurance fraud, dermal artifacts in patients with personality disorders), and various kinds of accidents.

Fatalities by sharp force account for a considerable part of the forensic autopsy material with a clear predominance of homicidal deaths. In several regions of the world (for instance, in Britain and continental Europe), the use of sharp force, especially stabbing with a knife, is the leading method of killing [1–3]. In most of the cases, the offenses arise from domestic disputes between life partners, interpersonal conflicts, street violence, and brawls associated with alcohol [2, 4]. Apart from this common background, there are also sexually motivated homicides committed by sharp force. In most cases, the perpetrators and the victims know each other.

Cutting the wrists is often seen in attempted suicides, but plays a minor role in suicide acts with fatal outcome. Suicidal stabbing is less frequent than

cutting and is usually confined to the precordial region or the neck. Accidental deaths from sharp force occur sporadically, for instance, when falling into a glass pane.

Stab Wounds

A stab wound is caused by a pointed object that produces a slitlike severance of the penetrated tissues. In most knife stabs, there is a clean division of the skin with sharp, straight, and cutlike edges that usually do not show any abrasions or contusions. Owing to the elastic retraction of the margins, a stab wound will gape, more or less depending on its location and its orientation with respect to the local fiber pattern of the skin. The most gaping wounds are found if the blade hits the skin transversely to the so-called cleavage lines [5, 6]. In contrast to lacerations, no tissue bridges can be seen between the margins of a sharp-force injury. Wounds from single-edged weapons may show one sharply pointed end (corresponding to the cutting edge of the blade) and one rounded, squared off, blunt, or fishtail-like end on the side of the knife's back (Figure 1, [1, 5]). In some types of knives, the back edge is serrated and can be used as a saw.

If a knife is thrust into the body and twisted before withdrawal, the skin wound gets the appearance of a "V", "Y", or "L" (Figure 2, [4, 5]). The wound edges can be considerably longer than the maximum width of the blade, as the stabbing action is often combined with an additional cutting component (incised-stab wounds [5]). The wound may start as a cut, which terminates as a stab wound; on the other hand, stab wounds can turn into an incised wound when the knife is withdrawn at a shallow angle. Stabbing actions do not always result in actual stab wounds but can also produce a cut if the blade contacts a curved part of the body tangentially. In transfixion stab wounds, the weapon may reenter the body after having passed through a limb or a sagging female breast.

Sometimes, the margins show a patterned imprint abrasion from the knife's handle or crossguard if the weapon was vigorously pushed in up to the end of the blade. The details of such contusion marks may allow conclusions as to the respective construction part of the knife [1, 5, 7, 8]. If the plane of the blade enters the skin at an oblique angle, the wound margins are beveled (shelved) on one side and undermined (overhanging, undercut) on the other, thus indicating the direction of the thrust (cf. Figure 1). Knives with serrated cutting edges do not necessarily produce stab wounds with characteristic features. When a serrated blade is drawn over the body surface tangentially, the teeth of the serration scrape across the skin, leaving parallel, linear abrasions [6, 8, 9].

Similar to gunshot injuries, taking X rays of the affected body regions is also advisable in cases of



Figure 1 Stab wound with clean-cut margins and one pointed end (corresponding to the cutting edge) and one fishtail-like end (corresponding to the back of the knife). When the knife enters at an oblique angle, one side is beveled and the other is undermined



Figure 2 Homicide of a 39-year-old man by a stab that penetrated above the right clavicle and severed the brachiocephalic trunk. The V-shaped appearance indicates that the knife was twisted in the wound before withdrawal

stabbing, not only in surviving victims but also in corpses destined for autopsy. Roentgenographs are a valuable means for postmortem documentation of a pneumo- and/or hematothorax, pericardial tamponade, and air embolism [8]; last but not the least, X rays help in recovering broken-off knife blade tips (Figure 3, [5, 8, 10]). Clothing should be examined thoroughly (number, length, and location of perforations and their correspondence to the stab wounds on the body).

In homicide investigations, the question often arises whether the wound dimensions and the other characteristics permit conclusions regarding the size and type of the weapon used. When measuring the length of a gaping surface wound, the edges should be brought into apposition [1]. Even after restoration by approximating the margins, the length of the skin wound need not reach the full width of the penetrating blade (due to the elasticity of the dermis). Moreover, the length of a stab wound also depends on the taper from the tip to the hilt: If only the narrow front section of a tapered blade has penetrated, the size of the wound corresponds to the width at the respective level. As mentioned earlier, a combination of stabbing and cutting may cause a wound slit much longer than the maximum width of the blade.

The length of the entire wound track indicates the depth to which a knife has penetrated. If the whole blade has entered, the depth of the wound may give an estimate of the blade's length. A vigorous stab can indent the body surface (for instance the abdominal wall) and compress the soft tissues so that the depth of the wound track may exceed the length of the causative knife blade [1, 8, 10].



(a)



(b)

Figure 3 (a) Kitchen knife with blood-soiled blade. The broken-off tip got stuck in the victim's skin (b). Homicide of a 49-year-old man by multiple stabs to the chest

It is difficult to verbally describe the amount of force required to inflict a given stab wound. For this reason, experienced authors [1] suggest the use of an easily understandable quantification (e.g., "slight pressure", "moderate force", "considerable force", and "extreme force"). There are several factors that influence the force necessary for penetration. An essential point is the sharpness of a knife's tip and cutting edge [4, 11]. (From a scientific perspective, the actual parameter that should be addressed is pressure, which is the amount of force per unit area ($P = F/A$). In this way, it is easy to see how a small force exerted with a pointed weapon causes a penetrating injury, whereas the same force used with, for example, a plank of wood causes only bruising with no penetration of the skin.) A high velocity of the thrust facilitates penetration. Apart from the properties of the weapon and the dynamics of stabbing [12, 13], the resistance offered by clothing, skin, bone, cartilage, and even soft tissue is of major importance [14].

Pointed weapons other than knives are also capable of inflicting stab wounds. Injuries from scissors have a different appearance according to the distance between the two blades: A closed pair of scissors will produce a singular skin wound with abraded margins, the shape often resembling a flat "Z" [1]. Open scissors with separated blades may cause a pair of stabs with one wound above and somewhat oblique to the other [6].

Another stabbing instrument, which can often be identified from the wound appearance, is the blade of a screwdriver. This is especially true for the Phillips screwdriver with its cross-shaped tip resulting in skin wounds consisting of four equally spaced cuts, whereas standard screwdrivers leave slitlike wounds with squared ends and abraded margins [5, 8, 15, 16]. Stab wounds can also be inflicted with a broken bottle or a drinking glass and by falling through a glass door if a pointed sliver enters the body [4, 17, 18]. Other possible stabbing objects include forks, ice picks, pencils, files, screws, and similar instruments with a more or less pointed tip [5, 6].

Analysis of the evidence at scene, the clothes and the injured body, can help differentiate whether

a stab wound was homicidal or self-inflicted [8, 19, 20]. Accidental stabs with a fatal outcome are extremely rare; they may occur in persons working professionally with sharp knives such as butchers when deboning slaughtered animals.

The following signs may be indicators of suicidal self-stabbing [8, 18, 21] :

- localization in the precordial region (Figure 4), sometimes involving the adjacent areas of the upper abdomen and, in rare cases, also stabs to the neck [22] or to the *cubital fossa*, the forearm, or the groin;
- grouping of the stab wounds (in cases with a multitude of individual injuries, Figure 5);
- presence of superficial "tentative stabs" [23];
- additional presence of wrist cuts or hesitation cuts to the neck;
- absence of defense injuries; and
- findings compatible with a retained ability to act up to the last injury.

Many suicides bare the respective body region before stabbing, although this is not true in all cases [24, 25].

In homicides, the fatal injuries are predominantly localized on the chest (Figure 6), neck, and abdomen [3]. The number of wounds is no reliable criterion for differentiation according to the manner of death, as single and multiple stab wounds are seen both in homicides and suicides [19, 26]. Infliction by another person is obvious when the affected body region cannot be reached by the victim's own hand. The additional presence of typical defense injuries proves the homicidal character of stabs.

Incised Wounds

Incised wounds or slashes are characterized by a clean-cut division of the skin and underlying tissues, resulting in a cut that is longer than deep. The wound edges usually do not show any contusion or abrasion, there are no bridges of tissue between the margins, and the wound ends are acute, often forming a superficial "tail" or shallow scratch. Cuts tend to gape in a spindle-shaped way, if they have sufficient depth. When the cutting edge is moved to and fro within one and the same wound, the slash may terminate in a multitude of shallow ends due to the "sawing" effect



Figure 4 (a, b) Suicide of a 55-year-old butcher who used a so-called deboning knife for inflicting two penetrating stab wounds to the precordial region. The wound axes are parallel to each other. The lateral wound ends are squared off, reflecting the back of the blade

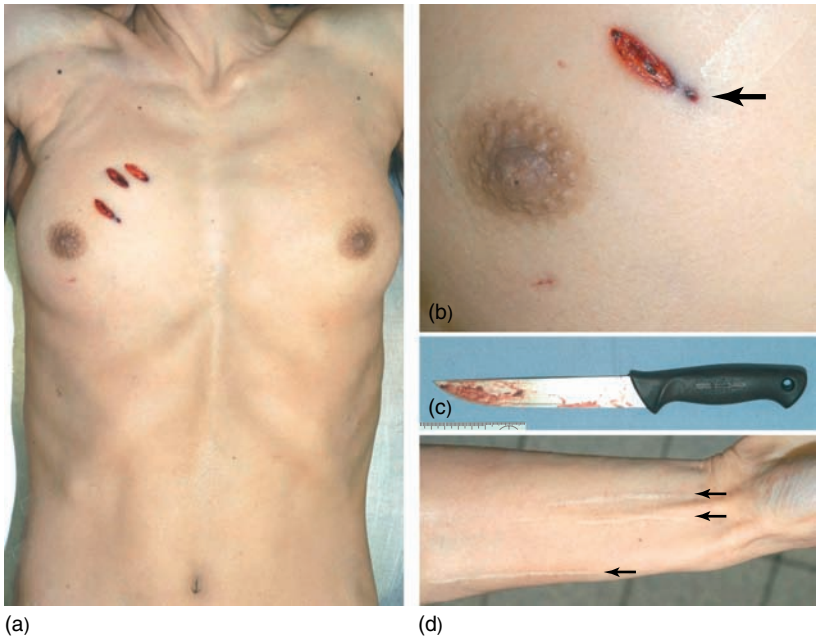


Figure 5 (a) Suicide of a 41-year-old woman by stabs to the chest. Atypical localization of the injuries on the right-hand side. The clothing had been pushed upward and therefore remained undamaged. Beside the wound next to the nipple, there was a punctiform "hesitation" stab (b), and the knife (c) was found close to the right hand. On the left forearm, there were whitish scars from a previous suicide attempt (d)



Figure 6 (a–c) Homicidal killing of a 37-year-old man by singular stab to the chest inflicted with a kitchen knife after a dispute in a state of intoxication. Death from exsanguination into the right pleural cavity [Reproduced from Ref. 21. © Elsevier, 2003.]

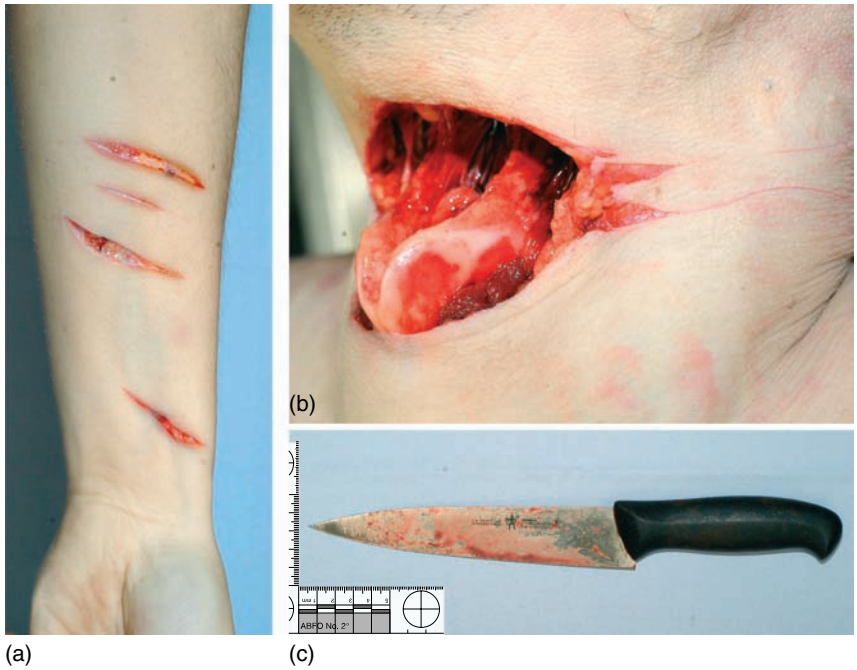


Figure 7 Suicide of a 26-year-old man who suffered from persecutory delusion. On the left forearm, there were four parallel, oblique, shallow cuts (a). Deep incised wound running transversely across the throat, exposing the laryngeal prominence with shallow ends on the left-hand side (b). The tool used was a kitchen knife (c) [Reproduced from Ref. 18. © Elsevier, 2004.]

(Figure 7b). If an edged weapon is drawn across the crests of loose skin or over bony ridges, a singular slash leaves an incised "wrinkle" wound with a linear, but interrupted, pattern of "skipping" cuts [5, 6].

Incised wounds are most often due to knives or other instruments with cutting edges such as razor blades, glass fragments, or pieces of sheet metal.

Cut wounds can occur in homicides, suicides, self-inflicted injuries without suicidal intention, and in accidents. Forensic differentiation between self-infliction and homicide is based on criteria similar to those described in connection with stab wounds [4].

Suicidal cuts are preferably located on the flexor sides of the distal forearms ("wrist cuts" (Figures 8 and 9b)) and on the neck ("cuts of throat" (Figures 7b and 9a)) [18]. In right-handed individuals, most of the wounds are expected to be on the left forearm and *vice versa* in left-handed persons. In most cases, there are multiple injuries, often parallel and arranged in groups, some of them being very superficial and having the character of "hesitation" or "tentative" marks resembling linear scratches (Figure 8b) [10]. Scars on the above-mentioned "sites of election" suggest previous suicide attempts (cf. Figure 5d). If the cuts do not lead to death or do not cause death rapidly enough, the fatal outcome may be achieved by applying another method such as hanging (complex suicide). The implement used to inflict suicidal cuts and/or stabs is typically left at the scene of death, even though the ability to act can be retained long enough to replace the weapon in its usual storage place or hide it [5]. In rare cases, the weapon may be still found in the hand of the deceased. As a rule, suicidal cuts are not accompanied by corresponding damage to the clothing. Unlike homicidal scenes, the place where a suicide is found normally does not show disturbance, especially regarding the distribution and spread of blood stains.

Fatal incised wounds from knife attacks mostly involve the neck [18]. Homicidal throat slashings can be single or multiple: the latter are sometimes in a "crisscross" arrangement with concomitant damage to the clothing. Characteristic hesitation cuts are missing. Additional defense wounds indicate the



Figure 8 (a, b) Suicide of a 54-year-old man by mostly shallow transverse cuts to the flexor side of both forearms. Death from exsanguination due to severance of the left radial and ulnar arteries

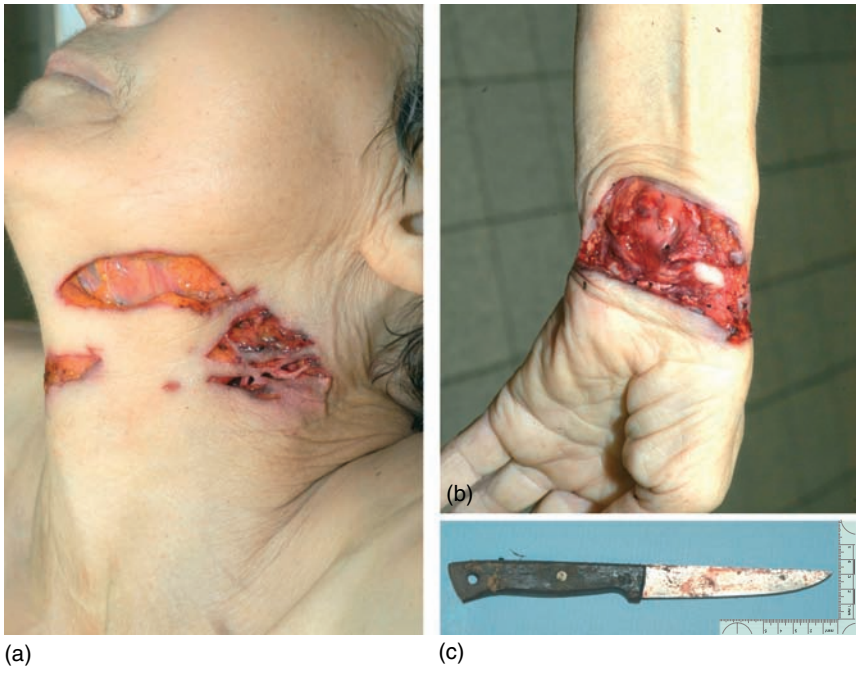


Figure 9 Suicide of a 66-year-old woman with a kitchen knife. Stabs and cuts arranged in groups in the anterior and lateral cervical region. Deep incised wound to the left wrist [Reproduced from Ref. 18. © Elsevier, 2004.]

homicidal manner of death, but they are by no means compulsory. For overall evaluation, it is essential to take into account the case history, the scene (including trace evidence, whereabouts of the weapon *etc*.), and the autopsy findings. Deaths from accidental cuts are occasionally seen after a fall through a sheet of glass or onto fragments of broken glass ware (Figure 10).

In clinical forensic medicine, the physical examination of living persons with stabs and cuts is part of the daily routine work. Apart from victims of assaults (mainly knife attacks), self-inflicted injuries play an important role [1, 4, 27]. A main category of self-induced bodily harm is the simulation of a criminal offense: The alleged victims falsely claim to have been raped or attacked for some fabricated reason (Figures 11 and 12). The primary intention of the informants is mostly to attract attention or affection. Usually, the self-inflicted injuries are caused with the help of pointed or cutting tools. The resulting lesions often consist of either shallow cuts or linear abrasions with the following typical features:

- equally shallow, nonpenetrating cuts;
- multitude of individual lesions;
- uniform shape, linear or slightly curved course of the lesions;
- grouped and/or parallel and/or crisscross arrangement;
- symmetry or preference of the nondominant side of the body (usually the left);
- location in easily reachable body regions;
- omission of especially sensitive body regions;
- no damage of the clothes or inconsistent damage;
- lack of defense injuries;
- sometimes additional presence of scars from former self-injurious behavior.

Some informants simulate a criminal offense by scratching letters, words, or political symbols such as swastikas into their skin [28].

Another category of self-inflicted cuts (including even mutilations) is seen in persons with emotional disturbances or mental diseases. Again, the lesions are typically multiple and located in easily reachable body regions. The coexistence of fresh wounds and



(a) (b)

Figure 10 (a) Accidental cuts suffered by a 61-year-old female (fall onto the fragments of a broken glass in a strongly intoxicated condition). The fatal loss of blood was accelerated by a coagulation disorder due to liver cirrhosis. (b) In the depth of some wound slits, splinters of glass could be detected



Figure 11 Right cheek of a 15-year-old girl with several scratchlike, linear, and equally shallow skin lesions arranged in a group. The girl had initially claimed to have been kidnapped by unknown perpetrators who raped her and caused the injuries on her cheek. Later, she admitted to have simulated the offense in order to justify her late arrival at home. The injuries had been self-inflicted with the tip of a pocket knife

scars of different ages points to repeated episodes of self-cutting (Figure 13).

Chop Wounds

Chop wounds are caused by the impact of rather heavy instruments with a more or less sharp edge, some of them having especially long blades (sword, saber, machete), whereas another group of implements comprises axes, hatchets, choppers, and meat cleavers. The severing of the tissue is effected by pressure and not by a tangential, drawing movement. Powerful blows can not only produce incisionlike wounds of the soft tissues but also notch or break the underlying bone [5, 8, 29]. If the hitting edge is dull, the wound margins show narrow abrasions.

In homicidal attacks with chopping tools, the perpetrators mostly aim at the head (Figure 14), and occasionally also at the neck [18]. Suicides are very rare, but may be committed, for instance, by multiple axe blows to the frontal, temporal, or parietal regions, resulting in parallel wound slits of varying depth. Accidental deaths are seen in persons who sustained injuries from a boat or airplane propeller [8].

In surviving victims, possible self-mutilation for the purpose of insurance fraud has to be considered. In such cases, an accident is simulated in a fraudulent attempt to obtain compensation from an insurance company [27, 30]. Voluntarily inflicted mutilations from axe blows are often claimed to have been caused by a misplaced stroke when chopping wood. Mostly, the thumb or the index finger is cut off close to the proximal joint, whereas in authentic accidents the amputation is usually a distal and incomplete one. A proximal and complete amputation of the index finger in a right angle to the axis is especially suspect of intentional self-mutilation.

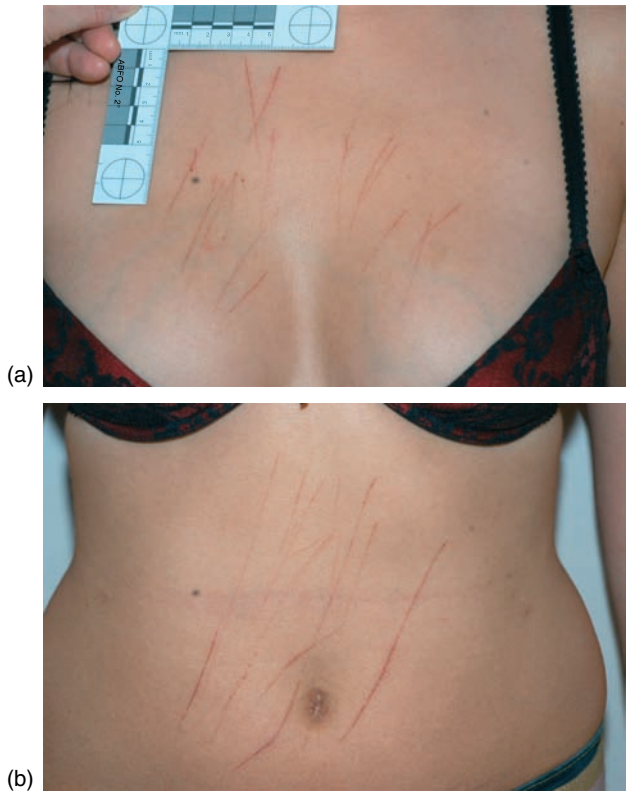


Figure 12 A 17-year-old girl who claimed to have been assaulted by an unknown perpetrator who allegedly injured her with a pointed instrument. The skin of the upper chest region (a) and of the abdominal region (b) showed a multitude of shallow scratchlike cuts, most of them running in parallel and without concomitant damage of clothes. Later, she admitted that she had invented the assault in order to attract attention and to win sympathy

Causes of Death and Ability to Act

A fatal outcome of sharp-force injuries is mostly the consequence of exsanguination caused by lesions to the heart, the great vessels, or the internal organs (*see* **Autopsy** [2, 4, 5, 21]). Penetrating wounds of the chest are often complicated by air in the pleural space (pneumothorax) with consecutive impairment of ventilation, especially when both halves of the thorax are affected. A cardiac tamponade (blood in the pericardial sac) can be caused by stabs to the heart if the blood cannot flow off into one of the pleural cavities. In stabs and cuts to the neck, a so-called air embolism may be the immediate cause of death: Air is sucked into a gaping venous vessel due to a negative intravasal pressure during inspiration. Another possible mechanism in sharp-force injuries of the neck is blood aspiration via the larynx or trachea. Late complications of stabs and cuts are local or generalized wound infections, for instance, peritonitis after injuries of the intestinal tract.

Even penetrating injuries of the heart, the aorta, or other major vessels do not necessarily mean that the ability to act is lost immediately [2, 21, 31, 32]. Incapacitation is to be expected only with the onset of unconsciousness. The duration and extent of the ability to act after stabs to internal organs or great vessels mainly depend on the size of the lesion and thus the amount and speed of blood loss [8].

Defense Injuries

Defense wounds indicate that an individual has been the victim of an assault. They also prove that the



Figure 13 Extensor side of the left forearm of a 16-year-old boy with repeated episodes of self-cutting due to a personality disorder



Figure 14 (a) Homicide of an 18-year-old man who was killed with a chopper. Gaping cutlike wounds with severance of the underlying bones and embedding of fragments broken off from the weapon (b). Defense injuries on the extensor side of the left hand with traumatic amputation of the index finger (c)



Figure 15 Defense injury of the flexor side of the left index finger (homicide of a 63-year-old woman being stabbed with a knife)



Figure 16 (a) Defense injuries on the left forearm of a 24-year-old woman, who was killed by several stabs to the chest. In the elbow region, there is a so-called incised-stab wound. (b) The tip of the bloodstained blade is broken off

attacked person was – at least initially – conscious and able to use his or her limbs for protection. When sustaining defense wounds, the victim also must have been aware of the assault. In knife attacks, the victim attempts to ward off the weapon either by seizing it or by raising the forearms and hands in order to protect other body regions such as the head, neck, and chest [1, 2, 4, 5, 8, 18, 33]. "Active" defense wounds occur when the threatened person grasps the knife with the hand so that the palmar surface is exposed to the blade (Figure 15). The "passive" injury type is seen when the raised (upper) limbs are held in front of the attacked body region so that the majority of the wounds is located on the extensor side (Figure 16).

The frequency of defense injuries indicated for homicide by sharp force ranges between 30 and 50% [34]. In victims primarily able to react, the incidence increases with the number of hits placed on the body. The absence of defense wounds in victims with a multitude of stabs/cuts may be an indication that the ability to act was impaired or lost (e.g., in persons who were held, tied, heavily intoxicated, or already unconscious before the knife attack). In rare cases, defense wounds are even localized on the legs: if the victim was in a lying position, the legs may be used to ward off the knife.

Some perpetrators of knife attacks show sharp-force injuries on their hands, which they unintentionally inflicted upon themselves in the dynamics of the offense. If the knife does not have an adequate handguard and the tip of the blade is abruptly decelerated by a solid resistance such as bone, the perpetrator's hand can slip off the hilt and slide onto the blade, causing an injury [35].

Dismemberment

Postmortem dismemberment of a homicide victim is a rare incident and cannot be dealt with extensively in this context. The instruments used comprise sharp-edged tools such as knives, saws, machetes, axes,



Figure 17 Severed left breast of a 25-year-old woman who was killed by blunt force to the head. Subsequently, the perpetrator dismembered the corpse in the bath tub. The upper part of the breast shows shallow parallel cuts

and meat cleavers [36]. The wounds do not show any vital reaction (Figure 17). The main questions to be answered by the medicolegal expert concern the identity of the victim(s) (one person or commingled remains?), the cause of death, and any clues as to the dissecting instrument and possible motives (mutilations for sexual reasons?). Competent disarticulation may be suggestive of a perpetrator with some knowledge of anatomy [21]. A specific search should be made for tool marks, particularly on the articular cartilages and bones [8].

References

[1] Saukko, P. & Knight, B. (2004). *Knight's Forensic Pathology*, 3rd Edition, London, Arnold, pp. 153–166, 240–243.

[2] Bohnert, M., Hüttemann, H. & Schmidt, U. (2006). Homicides by sharp force, in *Forensic Pathology Reviews*, M. Tsokos, ed, Humana Press, Totowa, NJ, Vol. 4, pp. 65–89.

[3] Hunt, A.C. & Cowling, R.J. (1991). Murder by stabbing. *Forensic Science International* **52**(1), 107–112.

[4] Vanezis, P. (2003). Sharp force trauma, in *Forensic Medicine: Clinical and Pathological Aspects*, J. Payne-James, A. Busuttil & W. Smock, eds, Greenwich Medical Media, London, pp. 307–319.

[5] Di Maio, D.J. & Di Maio, V.J.M. (1989). *Forensic Pathology*, Elsevier, New York, pp. 171–206.

[6] Spitz, W.U. (2006). Sharp force injury, in *Spitz and Fisher's Medicolegal Investigation of Death*, 4th edition, W.U. Spitz, ed. Thomas, Springfield, IL, pp. 252–309.

[7] Pollak, S. & La Harpe, R. (1992). Defined contusion marks caused by the knife handle. *Archiv fur Kriminologie* **190**(1–2), 1–8.

[8] Lew, E. & Matshes, E. (2005). Sharp force injuries, in *Forensic Pathology: Principles and Practice*, D. Dolinak, E. Matshes & E. Lew, eds, Elsevier Academic Press, Burlington, MA, pp. 143–162.

[9] Pollak, S. (1989). Pattern of findings in injuries caused by "survival knives". *Archiv fur Kriminologie* **183**(1–2), 11–20.

[10] Pounder, D.J. (2000). Sharp injury, in *Encyclopedia of Forensic Sciences*, J.A. Siegel, P.J. Saukko & G.C. Knupfer, eds, Academic Press, London, Vol. 1, pp. 340–345.

[11] Payne-James, J., Vanezis, P. (2005). Sharp and cutting-edge wounds, in J. Payne-James, R.W. Byard, T.S. Corey & C. Henderson, eds, *Encyclopedia of Forensic and Legal Medicine*, Elsevier Academic Press, Oxford, Vol. 3, pp. 119–129.

[12] Chadwick, E.K., Nicol, A.C., Lane, J.V. & Gray, T.G. (1999). Biomechanics of knife stab attacks. *Forensic Science International* **105**(1), 35–44.

[13] Kaatsch, H.J., Mehrens, C. & Nietert, M. (1994). The biomechanics of perforating stab wounds: experimental studies. *Rechtsmedizin* **4**(3), 91–98.

[14] O'Callaghan, P.T., Jones, M.D., James, D.S., Leadbeatter, S., Holt, C.A. & Nokes, L.D. (1999). Dynamics of stab wounds: force required for penetration of various cadaveric human tissues. *Forensic Science International* **104**(2–3), 173–178.

[15] Nadjem, H. & Pollak, S. (1993). Manifestation of screwdriver injuries. *Archiv fur Kriminologie* **192**(1–2), 27–36.

[16] Faller-Marquardt, M. & Pollak, S. (1996). Homicide with a screwdriver and simulation of a similar offense by self-infliction of injuries. *Journal of Clinical Forensic Medicine* **3**(3), 141–147.

[17] Ambach, E., Tributsch, W. & Rabl, W. (1991). Fatal injuries caused by glass fragments. Case reports and autopsy findings. *Archiv fur Kriminologie* **187**(1–2), 39–46.

[18] Pollak, S. & Saukko, P. (2003). *Atlas of forensic medicine. CD-ROM*, Elsevier, Amsterdam, chapter 6 (sharp trauma).

[19] Karlsson, T. (1998). Homicidal and suicidal sharp force fatalities in Stockholm, Sweden. Orientation of entrance wounds in stabs gives information in the classification. *Forensic Science International* **93**, 21–32.

[20] Karlsson, T. (1998). Multivariate analysis ('forensiometrics') – a new tool in forensic medicine. Differentiation between sharp force homicide and suicide. *Forensic Science International* **94**(3), 183–200.

[21] Eisenmenger, W. (2004). Spitze, scharfe und halbscharfe Gewalt, in *Handbuch gerichtliche Medizin*, B. Brinkmann & B. Madea, eds, Springer, Berlin, Vol. 1, pp. 571–591.

[22] Pollak, S. & Ropohl, D. (1992). Morphologic patterns in suicidal stab injuries of the neck. *Archiv fur Kriminologie* **190**(3–4), 72–81.

[23] Vanezis, P. & West, I.E. (1983). Tentative injuries in self stabbing. *Forensic Science International* **21**(1), 65–70.

[24] Start, R.D., Milroy, C.M. & Green, M.A. (1992). Suicide by self-stabbing. *Forensic Science International* **56**(1), 89–94.

[25] Bohnert, M., Ropohl, D. & Pollak, S. (1997). Suicidal stab wounds through clothing. *Archiv fur Kriminologie* **200**(1–2), 31–38.

[26] Ormstad, K., Karlsson, T., Enkler, L., Law, B. & Rajs, J. (1986). Pattern in sharp force fatalities – a comprehensive forensic medical study. *Journal of Forensic Medicine* **31**(2), 529–542.

[27] Pollak, S. & Saukko, P.J. (2000). Self-inflicted injury, in *Encyclopedia of Forensic Sciences*, J.A. Siegel, P.J. Saukko & G.C. Knupfer, eds, Academic Press, London, Vol. 1, pp. 391–397.

[28] Faller-Marquardt, M. & Pollak, S. (2006). Self-inflicted injuries with negative political overtones. *Forensic Science International* **159**(2–3), 226–229.

[29] Nadjem, H., Bohnert, M. & Pollak, S. (1999). Appearance of injuries caused by machetes and unusually large knives. *Archiv fur Kriminologie* **204**(5–6), 163–174.
[30] Bonte, W. (1983). Self-mutilation and private accident insurance. *Journal of Forensic Science* **28**(1), 70–82.
[31] Karger, B., Niemeyer, J. & Brinkmann, B. (1999). Physical activity following fatal injury from sharp pointed weapons. *International Journal of Legal Medicine* **112**(3), 188–191.
[32] Große Perdekamp, M., Riede, U.N. & Pollak, S. (2000). Penetrating stab injury to the heart with unusually long survival. *Archiv fur Kriminologie* **206**(3–4), 102–109.
[33] Pollak, S. & Saukko, P.J. (2000). Defense wounds, in *Encyclopedia of Forensic Sciences*, J.A. Siegel, P.J. Saukko & G.C. Knupfer, eds, Academic Press, London, Vol 1, pp. 374–378.
[34] Schmidt, U. & Pollak, S. (2006). Sharp force injuries in clinical forensic medicine – findings in victims and perpetrators. *Forensic Science International* **159**(2–3), 113–118.
[35] Schmidt, U., Faller-Marquardt, M., Tatschner, T., Walter, K. & Pollak, S. (2004). Cuts to the offender's own hand – unintentional self-infliction in the course of knife attacks. *International Journal of Legal Medicine* **118**(6), 348–354.
[36] Püschel, K. & Koops, E. (1987). Dismemberment and mutilation. *Archiv fur Kriminologie* **180**(1–2), 28–40; **180**(3–4), 88–100.

Related Articles

Autopsy

STEFAN POLLAK AND PEKKA J. SAUKKO

Wounds: Inflicted by Firearms

see Gunshot Wounds

Writing Instruments and Printing Devices

Introduction

Determining the general class of writing instrument or printing device that produced an entry is fundamental to many forensic document examinations. It may be that making the determination is all that is required in a case, or it may be the first stage of a more in-depth examination process.

Some examples of examinations where determining the class of writing instrument may assist are as follows:

- quickly identifying added or altered entries if different types of pen have been used;
- indicating the best method for subsequent chemical analysis, as different classes of pen employ inks with quite different compositions each of which may be more responsive to differing analytical techniques; and
- determining direction of stroke in handwriting examinations.

The detection of counterfeits or fraudulently produced documents is an obvious examination type where determining the class of printing device used is very important. However, this determination is also a basic part of many other examinations as it can allow you, among other things, to

- identify areas of alteration in a document;
- determine whether a document is an original or a copy;
- guide investigators as to what types of device to take samples from; and
- identify the device that produced a document.

Once the general class has been identified, further analysis can be undertaken to assist in identifying possible brands and/or models of writing instrument or printing device. For example, with a typewritten document this analysis may involve consulting a typestyle database to narrow down the possible makes and/or models of typewriter. With a handwritten or inkjet printed document it may involve a chemical analysis of the ink and comparison to an ink manufacturer database.

There are a wide and ever increasing variety of analytical techniques that can be employed in the examination of the output from writing instruments and printing devices. These techniques can be used to distinguish between output from items of the same class, such as thin layer chromatography (TLC) to differentiate between ballpoint pen inks or Fourier transform infrared spectroscopy (FTIR) to distinguish between brands of toner.

In combination with ink libraries and spectroscopic databases, these techniques can be used to

obtain brand and model information. This may allow an investigation to be focused and provide a guide as to what devices to take samples from.

Ultimately, however, the goal of the examination is often to identify the source of a questioned document, using individuating faults to link to a suspect device. Determining the class of printing device allows the examiner to assess the significance of marks and characteristics in terms of their value in linking to a particular device.

This article focuses on the visual examination of the characteristics that allow the different classes of writing instruments and printing devices to be distinguished.

Writing Instruments

Writing instruments constitute a very broad category, which can really include anything capable of visibly recording handwriting on a surface. Obviously there have been, and still are, a large number of devices used for this purpose. This article discusses only pens that are most commonly encountered in document examination cases; that is, nib or fountain pens, ballpoint, porous tip, rollerball, and gel pens. The scope of the article is restricted to the general physical characteristics of the writing line. Ink analysis is covered in **Ink Analysis**.

Nib Pens

Nib pens, such as the steel dip pen or the fountain pen, use capillary action to flow ink down a split in the middle of the nib. Because of the split in the nib, these pens can create double furrows within the ink line (Figure 1). The degree to which this is noticeable depends on the stiffness of the nib and the writing pressure employed.

Dip pens need to be regularly refilled, usually as their name suggests, through dipping in an ink well. Because of this their ink line gradually reduces in intensity between refillings.

Fountain pens on the other hand have an inbuilt ink reservoir. Because of this they generally produce a continuous and even ink line. Fountain pen inks need to be free flowing and are likely to be readily absorbed by paper and bleed into the fibers, causing feathering of the ink line.

The direction of stroke in a nib pen entry can often be determined through the phenomenon of flowback [1], where additional ink flows back down the ink line at the end of a stroke, darkening that portion of the ink line.

Ballpoint

After many early attempts, reliable ballpoint pens became available from around 1950 [2]. The basic writing mechanism of the ballpoint pen is that it has a small steel ball, which is the writing tip, sitting in a housing in which it can freely rotate. As the ball rotates it picks up ink from the reservoir behind it and rolls it onto the page. As the ball keeps rolling it continues to pick up ink from the reservoir, producing a continuous ink line. Because of the pressure applied during writing, the ball generally produces a groove along the middle of the ink line. Ballpoint pen strokes are distinctive owing to the highly viscous, paste-like, ink that is used in them. This ink does not flow, so when writing with light to medium pressure writing the ink often only records on the top surface of the paper fibers, leaving numerous white voids in the ink line.

Stroke direction can be determined in ballpoint pen writing in a number of ways. Two techniques arising from faults are discussed. They also help to characterize the visual appearance of a ballpoint ink line.



Figure 1 A fountain pen entry showing double furrows from the split nib and feathering of the ink line



Figure 2 A ballpoint pen entry showing fine uninked striations and ink "goops"

Firstly, fine uninked striations running through the stroke [3] (Figure 2). These are caused by imperfections, or a build up of debris, between the ball and its housing. As the ball rotates they wipe off the ink, producing fine, uninked lines on the ball, which then transfer to the page. Owing to the rotation of the ball in the housing, as these striations travel along the ink line, they appear to move from the inside to the outside of a curved stroke. This phenomenon very clearly and quickly shows the direction in which the stroke was made. Occasionally these striations occur in sufficiently distinctive patterns to enable the writing from an individual pen to be uniquely identified [4].

Secondly, because of its viscous nature, the ink is prone to accumulating on the lip of the ball housing. These accumulations wipe off onto the ink line appearing as small dark spots and are generally referred to as "*goops*". They tend to wipe off immediately after a change in direction, showing the direction of stroke.

Porous Tip

This class of writing instrument is sometimes also referred to as *fiber tip* as they were the first type of pens in this class, first introduced in the early 1960s [2]. However, the general class of porous tip pens also includes felt tip pens and models with a perforated plastic tip. The two main common features of porous tip pens are that their ink is stored in a saturated fibrous column and that their porous tip works like a wick, allowing the liquid ink to flow onto the writing surface. Because of the very fluid nature of their ink, porous tip pens tend to produce saturated, continuous ink lines, often with feathered edges if writing on paper (Figure 3). Because capillary action is constantly feeding ink to the nib, if the nib is paused during the stroke, ink keeps running onto the paper. This results in a dark dot and/or local thickening of the width of the ink line.

Rollerball

The rollerball pen was introduced in the late 1960s. Like a ball point pen, it uses a rolling ball ink delivery system. However, it has a very liquid ink like a fountain or porous tip pen. Generally, its writing strokes exhibit the groove caused by the rolling ball and seen in ballpoint pen strokes (Figure 4).

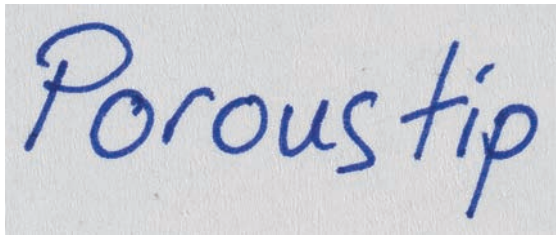


Figure 3 A porous tip pen entry showing feathered ink line and dark dots associated with pauses

However, it can be distinguished from ballpoint pens by the very liquid appearance of the ink line. This has the feathered edge appearance associated with porous tip and fountain pens. Additionally, it sometimes also exhibits the darkening of the ink line at the end of a stroke, reminiscent of the flowback associated with fountain pens.



Figure 4 A rollerball pen entry showing feathered ink line and groove produced by the ball

Gel Pen

The first gel pens were created by Sakura Color Products Corporation of Japan in 1984 [5]. Gel pens utilize a similar rolling ball ink delivery system to rollerballs and ballpoint pens; where they differ is their ink. Gel pen ink is water based and designed to be environmentally friendly with limited volatile components [6]. The ink has a high viscosity, although this is designed to lower when the pen is used. The

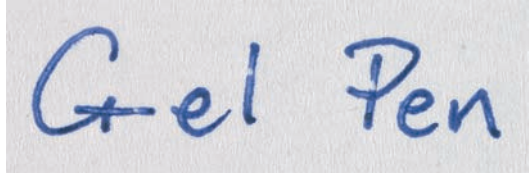


Figure 5 A gel pen entry showing tramline effect and "wiping" of overlapping strokes

appearance of the written line is similar to that from a rollerball. However, the gel-like nature of the ink can become apparent through the tramline effect that can occur along the edges of the stroke [4] (Figure 5). The ink is sometimes forced outward by the ball to the edge of the stroke, creating a dark pair of parallel lines. This, and a lower tendency toward feathering of the stroke edges, enables gel pen writing to be distinguished from that of a rollerball.

Gel pen inks are available in a wide variety of vibrant colors. Those containing reflective particles giving the ink a metallic sheen seem to be particularly popular.

Like rollerballs, gel pens tend to leave a dark dot of extra ink at the end of a stroke. Where two strokes intersect, the tramline edges of the second stroke wipe through and remove those of the first stroke. These characteristics can help determine direction and sequence of stroke in gel pen writing.

Printing Devices

The variety of printing devices and their associated printing technology is something that is constantly evolving. In this article, only the more common printing technologies are covered: typewriting, dot matrix, electrophotographic, ink jet, and thermal printing. Some of the specific issues of photocopier and facsimile machine examination are also discussed.

A discussion of the many different analytical techniques employed to examine the class characteristics of the different inks and toners associated with different printing technologies is beyond the scope of this article. The latter instead concentrates on the likely sources of individuating faults associated with the printing technologies and devices. That is, faults or characteristics that may allow a particular questioned document to be matched back to its source through a comparison with a suspect device.

As rollers or paper grabbers are used to some extent in all printing devices, they are a valuable potential source of evidence. The indented marks they leave on a page can often be visualized with a device such as the electrostatic detection apparatus (ESDA) [7]. While many of the impressions they make may be common to other printers of the same model or make, some rollers may develop individuating characteristics that may link them to their output.

Printing Technologies

The purpose of a forensic examination of a printed document is usually twofold. The first step is to identify the printing process(es) used to produce the document. In some cases, such as general questions of authenticity or counterfeiting, this level of examination is all that is required. Often though, the goal of the examination is to determine whether the printed document contains evidence that positively links it to a particular printing device. That is, the examiner is looking for some form of individual characteristic in the print that is also present on known samples taken from a suspect printing device.

Typewriting. Typewriters differ from the computer driven printing devices more commonly in use now, in that each character or symbol that is going to be typed exists as an examinable physical object. Computer-based printers generate characters and images by creating them from a series of dots; the appropriate placement of which is determined by the computer and printer. However, on a typewriter the individual characters are already there. Pressing a key on the typewriter keyboard causes that character to physically strike the page, generally through a ribbon, forming an image of that character on the page. The distinct advantage this gives an examiner is that if that individual character contains some form of defect or damage that is generally recorded in its image on the page. This is one way for typewritten documents to be identified back to their source.

The two main classes of traditional typewriter technology are the older typebar machines and the electronic single-element typewriters.

Typebar machines are the traditional “basket” typewriters with a row of metal typebars, each of which had a head like a metal stamp with two characters on it. Pressing a key caused an individual typebar to fly forward striking the page through an inked ribbon. Holding down the shift key would adjust the positioning of the typebar so that the other character on the head would strike the page when the key was pressed. Each key press would also move the carriage holding the paper and backing platen one space to the left so that the typed impressions were correctly spaced. In manual typewriters (nonelectric) the pressure with which the typebar hit the page was directly related to the force with which the key was pressed. Accordingly, one of the characteristics of

typing from a manual typebar machine is that overall the typing impression tends to be heavy and the impression of individual characters within a word or line is often uneven.

In addition to developing various alignment and movement faults, this type of machine was particularly prone to developing faults in their type characters. Often when typing it was common to simultaneously hit more than one key, causing two typebars to activate. This often resulted in one head hitting the back of another, causing physical damage to the character. The type characters in a typebar machine are physically part of the typewriter. This means that an individuating fault in a typewritten document positively identifies the machine that typed it. Also, fixing damaged characters requires major repair work; hence, character damage once formed is likely to remain.

This level of certainty was undermined by the development of the single-element typewriters. In these machines the type characters are contained on a removable single element. These elements take the form of either a *typeball* (these machines are often known as *golf ball typewriters*, Figure 6) or a *typewheel* (also commonly referred to as a *daisy wheel*, Figure 7).

Single-element typewriters do not have the direct relationship between typing pressure and the impression on the page that manual typebar machines do.



Figure 6 A typeball or "golf ball" from single-element typewriter

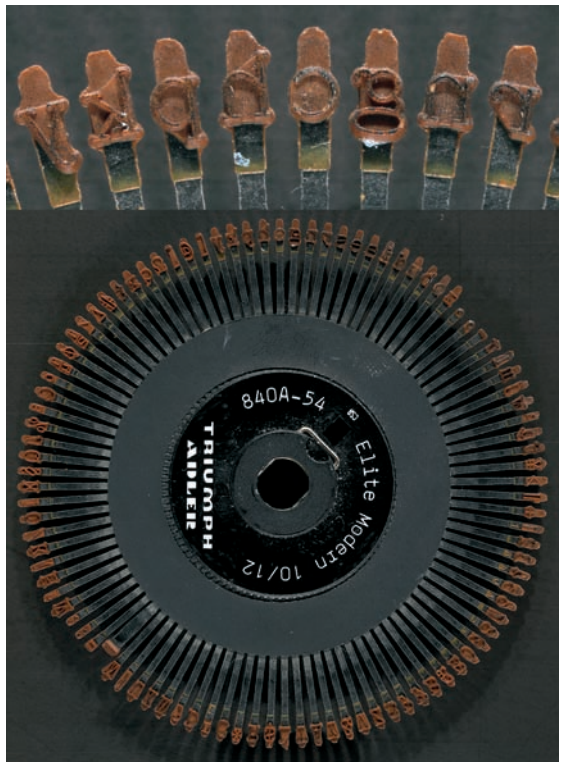


Figure 7 A typewheel or "daisy wheel" from single-element typewriter. The top of the figure shows an enlargement of the characters

The movement of the type element is electronically controlled. Assuming no fault develops in this mechanism, it produces a more even pressure between typed characters. Another major difference is that in single-element machines there is no moving carriage. The platen with the paper is fixed and the type element moves across the page.

The first single-element typewriter was the IBM Selectric, introduced in 1961 [8]. It used a typeball element. The typeball is a sphere (truncated top and bottom) with the characters positioned in rows and columns on its surface. When a key is pressed the typeball rotates and tilts as necessary to move the required character into position. It then moves the entire head forward, impressing the character onto the page through a ribbon.

A typewheel element consists of a central core encircled by radiating flexible spokes, each with a head containing a single character. When a key is pressed the typewheel is rotated to move the spoke

with the required character into position. A hammer then presses the back of the character, pushing it against the ribbon and page to form the typed impression.

Because the elements are removable, a single-element typewriter can produce typing in many different fonts and styles just by changing elements. Changing the elements can also be done without having to remove the page being typed; hence, a page with more than one typestyle can be easily produced. It also means that a character fault can only be matched back to the element, not the typewriter itself.

Individual faults can still develop within the characters on the elements. However, faults found in the characters can sometimes be class faults resulting from the manufacture process. An examination of the actual element, rather than just other typed specimens, can be necessary to more accurately determine the nature of the fault [9]. It has also been noted that once character damage occurs in a typewheel element, the character may continue to break down with use, and the damage becomes significantly worse in a relatively short time [10]. This increases the likelihood of an operator noticing the damage and eliminating it by replacing the type element.

While developments in typewriter mechanics made some aspects of typewriting examinations more difficult, developments in typewriter ribbon technology provided new ways to link questioned typing to its source.

Unlike their fabric predecessors, single strike carbon ribbons were designed to be used only once and then replaced. They consist of a polymer film backing with a coating that is transferred to the paper to form the image when a character strikes the ribbon. This leaves a hole in the coating corresponding to the typed character. With a properly working ribbon mechanism, the next typed character appears one space further on down the ribbon and not overlapping the previous character. A subsequent examination can therefore effectively read back everything typed on the ribbon. Additionally, the pressure of the type element forcing the ribbon against the page causes an impression of the paper fibers at that point of the page to record onto the polymer film backing [11]. This means that questioned typing on a page can be physically matched to the carbon film ribbon that produced it (Figure 8).

Another source of evidence for the document examiner came with the introduction of liftoff correction ribbons by IBM in the Correcting Selectric typewriter in 1973 [12]. This introduced a correction ribbon consisting of an adhesive tape that was used to lift the carbon film impression of an incorrectly typed character off the page. The result is a correction ribbon with a sequential record of the carbon film characters that have been corrected on that typewriter. Unless a significant amount of correction was being done many carbon film cassettes would be used and discarded before the correction ribbon would need replacing. Accordingly, even if the questioned text could not be found on the carbon ribbon currently in the typewriter when located, a fiber impression could be present and a comparison may be possible between any corrections in the questioned text and the images recorded on the correction ribbon [13].

Dot Matrix. The printhead of a dot matrix printer consists of a vertical stack of pins. A variety of printheads with different pin configurations have been used, but the most common are the 9 pin and the 24 pin (two columns of 12 pins) [14]. The name of the process comes from the fact that it uses a set of dots to form characters defined in a matrix form. The vertical columns of the matrix correspond to the printhead's stack of pins. The horizontal rows of the matrix are formed by iterations of the vertical pin stack as the printhead moves across the page.

A standard single pass of a dot matrix printhead produces what is known as "*draft" print quality*. Using a second, slightly offset, pass or a double column printhead produces a more complete image with the individual dots having less gap between them and blending more to look like a solid character. This is often described as "near letter quality" (NLQ) print [15].

With an impact printing dot matrix printhead, the printing controller moves the individual pins in and out, forming the image by striking the paper through an inked ribbon. Because of the impact process, these printers lend themselves to being used in conjunction with self-carbonating forms and this is often where they are used nowadays.

Thermal dot matrix printers also exist, which use individually heated pins to form an image. This can either be directly to darken heat sensitive thermal paper or indirectly using a heat sensitive thermal ribbon to print onto plain paper. Applying heat from

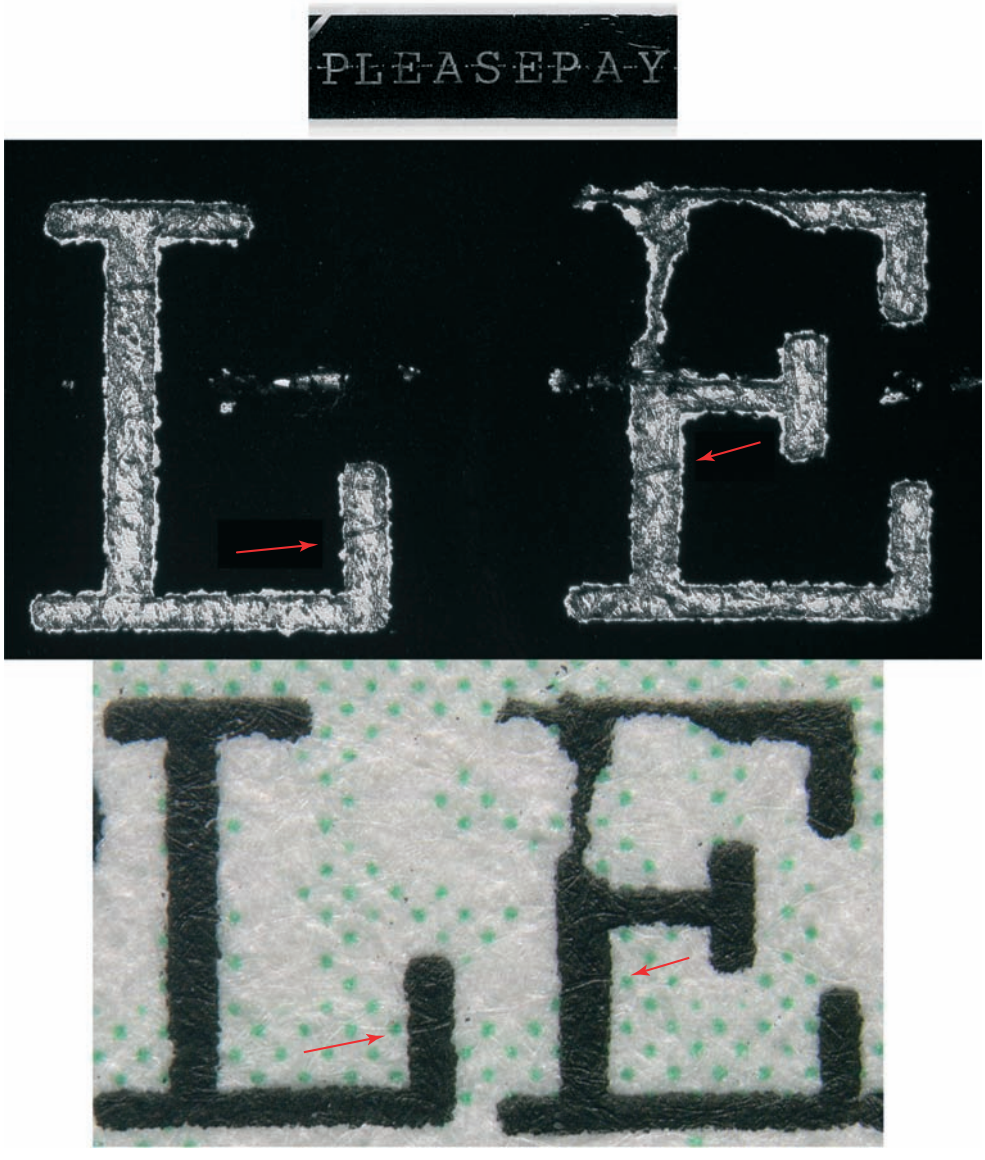


Figure 8 The top of the figure shows a used piece of single strike carbon ribbon. The middle image is an enlarged portion of the ribbon, showing paper fiber impressions that have been formed on the ribbon. The lower image shows the questioned document. The fiber that formed the prominent impressions on the ribbon is arrowed

the pins to the back of a thermal ribbon causes the ribbon to release ink or carbon onto the printing substrate. The process involved in printing with a thermal ribbon is that described in the thermal fusion process in this article.

There are several different types of fault that can develop in the printing mechanism of a dot matrix printer. A combination of some of these in a document may enable the printer that produced them to be identified. Some of the main faults are as follows:

1. Bent pins

A bent pin or misalignment in the firing mechanism that moves it in and out can cause the pin to be out

of alignment with the other pins in the printhead. Generally this is misalignment in the horizontal direction as movement in the vertical direction may result in neighbouring pins colliding, which causes a different fault.

2. Pin wear

Constant impact can lead to wear in the tips of the pins. This could take the form of broadening or "mushrooming" of the pin tip resulting in a larger dot, or uneven wear resulting in a noncircular dot. Significant, sudden damage of this type could occur through an impact with a foreign body, such as a bit of grit in the paper.

3. Weak impression

A pin may record faintly due to a fault in the firing mechanism or a build up of ink and dust that retards its movement. This type of fault can sometimes be intermittent.

4. Print-stack collapse

This is where misalignment in the vertical direction reaches the point of pins collapsing down onto their neighbors. Once this happens to a significant degree it is unlikely that the printer will continue to be used owing to the quality of its output.

When considering pin wear and tear faults it is worth bearing in mind that some pins in the head do considerably more work than others and so could be expected to wear out faster. For instance, in a nine pin head the lower two pins are generally only used (not considering images and vertically enlarged text) for the portions of the image that are printed below the line. Accordingly they get much less use. While the top two pins get more use (capitals, numbers, ascenders) it is the middle five pins that are used constantly. It would be more likely then for a wear associated fault to develop in pins 3–7.

Electrophotographic. Laser printers use the electrophotographic printing process, which is also sometimes referred to as *electrostatic printing* or *xerography*. This is also the process used in most photocopiers and many facsimile machines.

The basic mechanism for this printing process is that light is used to create an electrostatically charged latent image on the surface of a photoconductive drum or belt. Oppositely charged toner is applied to the drum, adhering only to the latent image. The toner image is then transferred to a sheet of paper and fused to it using heat and/or pressure (Figure 9).

There are many variations on this fundamental design.

In traditional photocopiers, the imaging drum is given an initial uniform charge. Light is shone onto the document to be copied and is reflected from the white, nonimage areas of the page onto the photosensitive drum. This reflected light dissipates the charge in these areas of the drum. Accordingly, only the areas on the drum that correspond to the dark, light absorbing image details on the page to be copied are left with charge.

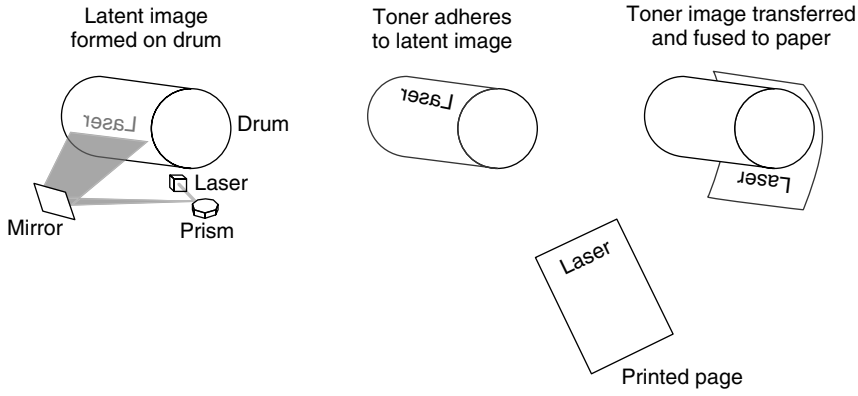


Figure 9 The basic electrophotographic printing process used in laser printers. In this illustration a pulsed laser beam and scanning prism/mirror arrangement are used to from an electrostatically charged latent image on the photoconductive drum. Toner is then applied and adheres to the latent image. This toner image is then transferred to the page and fused to produce the printed document

In a laser printer there is no document for light to be reflected from. Instead, the source of the image is a digital signal sent from a computer. Traditionally, the image is formed through pulsed laser light being scanned across the photosensitive drum. Depending on the printer design, the image can be formed by one of two ways. Either, as with the photocopier design above, the drum is precharged and the laser light is used to dissipate the charge in the nonimage areas that are to be white in the completed print. This is often described as print white. The alternative method is described as print black, where the laser is directed onto the image areas that are going to be black in the completed print. This is the method more commonly used in laser printers today.

A number of laser printers now no longer use the scanning laser beam to create the image on the drum. Instead they use a fixed array of light-emitting diodes (LEDs) positioned in front of the drum. These are switched on and off by the printer controller to selectively expose the imaging drum.

Another alternative technology is the use of liquid crystal shutters. With an electronic signal these shutters can be set to either black (closed) or transparent (open). In this method a constant light source is used, with an array of liquid crystal shutters positioned across the drum. The printer controller opens and closes the shutters to selectively allow the light to the drum to form the image.

There are a number of faults or defects, which can develop in a laser printer that allow documents to be positively matched back to that particular device. However, given the scope of this article, I shall only discuss the two most common faults, those associated with the imaging drum [16] and the fuser roller.

There are many ways damage can occur to the photosensitive surface of the imaging drum, including in the manufacturing process, during installation, from paper jams, foreign objects pulled in through the paper feed, and natural wear. Depending on the nature of the damage these faults can either always attract toner and print black on the page or reject toner and print white (Figure 10). As most documents usually consist mostly of white page with some black text, the faults that print white are often difficult to locate. Accordingly, it is more often the black printing faults that are used for comparison.

A fault on a laser printer imaging drum prints with each revolution of the drum. The circumference of most laser printer imaging drums are smaller than a standard page. Accordingly, a fault on a drum prints repeatedly down the page as a dot in the same horizontal position across the page. A common circumference for a laser printer drum is 94 mm. A fault on a drum of this size prints three or sometimes four times down the page. This makes the marks produced by a drum fault easily separable from the random background scatter that sometimes occurs with laser printers.

When comparing marks on a questioned document back to known samples, the horizontal position of the mark on the page is its initial identifying characteristic. However, given that there are only so many positions that a mark could form in across the page



Figure 10 The top half of the figure shows text from a threatening letter. Faults in the laser printer's photoconductive drum have rejected toner and recorded as white dots. The bottom half of the figure shows the same faults appearing in a sample print from the suspect's laser printer

the possibility that another laser printer could also develop a fault in the same horizontal position is very real and needs to be considered. A repeating pattern of several marks increases the likelihood of being able to positively identify a printer significantly. So too do faults of distinctive size and shape (Figure 11).

The heat and pressure fuser roller mechanism used in laser printers and photocopiers is particularly susceptible to damage [17]. Partly this is through the natural wear that occurs given the twin stresses of heat and pressure. But additionally the fuser is an area often prone to paper jams. This increases the chance of damage when the jam is being cleared. Faults associated with fuser rollers can be extraneous toner marks or smears and heavily embossed patterning of the paper (Figure 12).

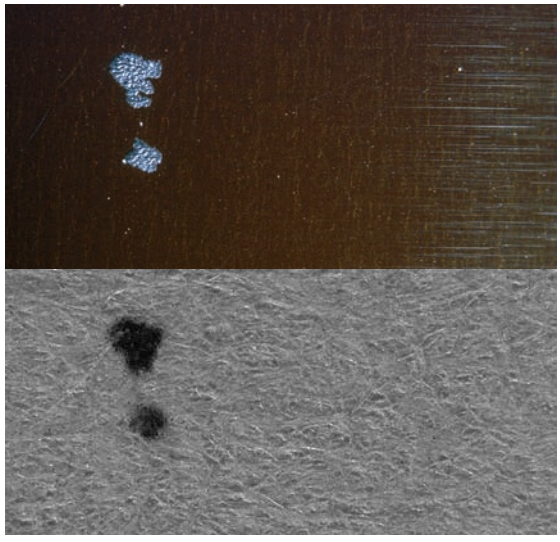


Figure 11 The top half of the figure shows distinctively shaped damage on the surface of a laser printer's photoconductive drum. The bottom half of the figure shows how this damage reproduces as a toner mark on documents printed with this drum. The photograph of the drum has been mirrored to allow for a direct comparison

Inkjet. The output of an inkjet printer is formed from droplets of liquid ink. These are fired at the printing surface through a printhead made up of an array of fine nozzles. The formation of the image is determined by a digital signal.

There are two broad categories of inkjet printing. These are "continuous stream" and "drop-on-demand".

In continuous stream inkjet printing, the printhead produces a constant flow of ink droplets. The droplets that are actually going to be required to form the image are given an electrical charge. By using an electrostatically charged deflection plate, droplets are then deflected, either toward the printing surface if they are to form part of the image or to a waste collection gutter if they are not required. Continuous

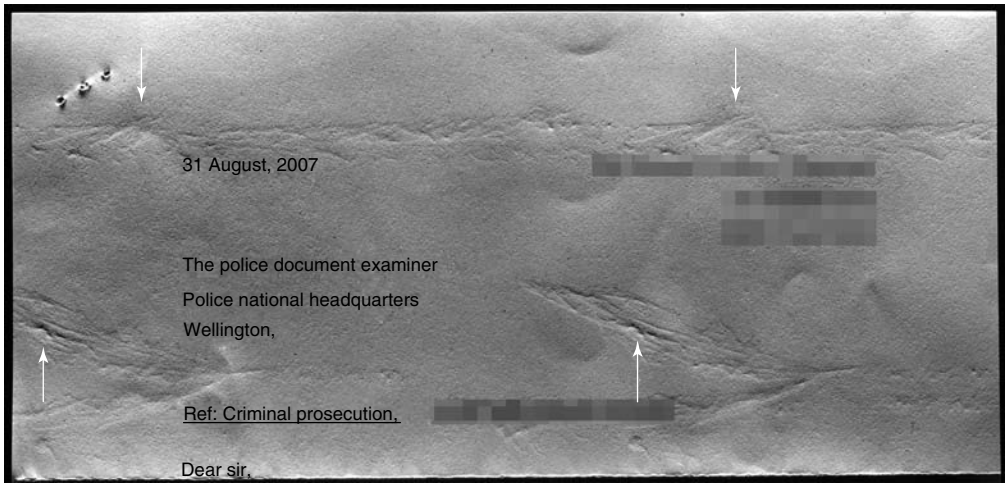


Figure 12 This photograph shows the heavily embossed pattern caused by damage to a photocopier's fuser roller. This pattern is embossed onto all the documents it prints. The arrows highlight repeating marks, demonstrating the circumference of the roller (the pixelization is for privacy purposes)

stream inkjet tends to be utilized in high volume commercial applications such as product dating and barcodes.

As the name suggests, drop-on-demand inkjet printers only produce a droplet when it is required to form the printed image. Drop-on-demand inkjet printing is generally performed using one of three technologies. These are thermal inkjet, piezoelectric, and solid inkjet.

Piezoelectric. The basic principle behind this method is that the application of voltage to a piezoelectric crystal causes it to deform. By alternating the voltage a piezoelectric element can be caused to rapidly flex. This vibration of a piezoelectric element can be used to produce a pressure pulse in a printhead ink chamber, forcing a droplet of liquid ink out of the printing nozzle. Epson is probably the largest user of piezoelectric inkjet printing technology.

Thermal Inkjet. This process uses a heater element in the printhead ink chamber to create a pressure difference to force ink droplets out of the printing nozzle. Rapid heating of the element causes the liquid ink next to it to vaporize, creating a bubble. This rapidly expanding bubble forces a droplet of ink out of the printhead nozzle, firing it toward the page. As the bubble collapses more ink is drawn into the chamber. This process is often referred to as *bubble-jet printing* and tends to dominate the desktop printer market.

Solid Inkjet. Like other inkjet devices, solid inkjet printers have a print head consisting of a series of nozzles through which the ink is propelled to form the printed image. The main difference is the ink itself. This is solid and waxy, rather like crayon. Typically the four process colours cyan, magenta, yellow, and black (CMYK) are used. In the printing process, the ends of the solid ink blocks are melted and the resultant liquid ink is sprayed through the printhead nozzles onto a drum, where the image is formed. The paper is then pressed against the drum, transferring the image to the paper. The ink then cools and hardens into its original solid state. The output from solid inkjet is easily identified by its raised, waxy appearance.

Some printers sprayed the ink directly onto the printed page. However, the interim step of forming the image on a drum and transferring it to the page was introduced as it allowed for faster printing and improved image quality.

Another difference from the majority of inkjet printers is that the solid inkjet printhead is page wide, i.e., the printing nozzles are in fixed positions and the print head does not scan back and forth across the page.

Solid inkjet was originally popular with graphic artists for its saturated colour and as an alternative to thermal wax transfer printers for printing colour proofs. However, Xerox has been developing solid inkjet printing technology and is beginning to market it for general office use. This has included incorporating solid inkjet technology in multifunction devices (all-in-one devices that photocopy, scan, print, and fax).

Positively linking a questioned document to a specific inkjet printer is notoriously difficult. While printers can develop blocked nozzles, this type of fault is often highly transitory as most printers perform regular cleaning cycles to try and eliminate this fault. With colour image printing, a blockage of several nozzles of the same colour often results in a noticeable colour cast, which may well prompt the user to initiate a nozzle cleaning cycle. Stripes through the image that look like blocked nozzles can also be caused through faults with the printhead's electrical contacts. This can sometimes appear quite dramatic with many nozzles not firing. In cases where this is encountered, investigations would need to be made to determine whether that was a common failure in that model of printer.

A not uncommon fault with inkjet printers is paper misfeeds, or paper too small for the image, leading to the printer printing onto the backing platen. Subsequent pages may pick up this ink on their rear providing a possible link to the printer.

With solid inkjet printers the waxy nature of the ink increases the chance that blockages are more persistent. Also, the full page-width printhead means that any nozzle faults occur in the same position on each page, making them much easier to be identified (see Figure 13).

Dye Sublimation. The chemical process of sublimation is where the application of heat transforms a substance directly from a solid state to a gaseous state without an intervening liquid phase. The dye sublimation printing method is also known as *dye diffusion*



Figure 13 This is a print from a solid inkjet printer. The banding seen slightly right of center is caused by blocked cyan ink nozzles

thermal transfer (D2T2), *dye diffusion*, or *thermal dye* [18]. It uses a full page-width print head consisting of a line of heating elements. These elements are heated to transfer sublimable dyes from a ribbon onto the printing substrate. They heat the dye until it sublimates to its gaseous phase and then transfers to the printing substrate. The substrate requires a special coating, which absorbs the gaseous dye, returning it to a solid state and so producing the image. Dye sublimation cannot print directly onto plain paper.

The amount of dye that transfers can be varied by altering the degree of heating of the printhead elements. This means that dye sublimation can print shades of color, producing continuous tone images of near photographic quality.

The ribbon is a clear plastic film coated in dyes. The dyes are in repeating sets of full page sized panels, generally of yellow (Y), magenta (M), and cyan (C). Some ribbons may also include a black (K) panel and/or a protective overcoat (O) panel. The final print is made up through a series of separate passes with each color being printed separately. For example, using a YMC ribbon the yellow component of the image would be printed, then the magenta, then the cyan. At the end of the yellow and magenta passes the printing substrate would be drawn back into the printer to allow for the addition of the next color. All the panels in a set are used to print a single page and cannot (in normal usage) be reused.

The best way of identifying printed output back to a specific printer is through the ribbon. Each panel on a used ribbon retains a negative image (corresponding to the lost dye) of what was printed (Figure 14). A microscopic examination of the ribbon and print often locates areas where a piece of dust or



Figure 14 The image on the left is a print from a dye sublimation printer. The image on the right is from the magenta panel of the printer ribbon, showing a negative image corresponding to the lost magenta dye that has gone to make up the print

other obstruction has affected the transfer of dye and recorded on both the ribbon and print. This allows the ribbon to be "physically matched" back to the print, rather than just on its content.

Failure of heating elements in the printhead results in vertical lines on the printed page and could also be used to link a printer to its output.

Thermal Wax Transfer. This printing process is also known as *thermal mass transfer, hot wax transfer,* or *direct thermal transfer* (D1T2) [18]. The general mechanics of the printing process are sufficiently similar to dye sublimation that many machines are actually sold as dual technology, capable of printing both thermal wax and dye sublimation. The significant difference between the processes is the composition of the ribbon used. Thermal wax transfer uses ribbons coated with a wax layer containing pigments. The heated printhead elements melt the wax and it transfers onto the printing substrate.

Thermal wax transfer can print directly onto plain paper, although it transfers most easily onto a smooth or glossy substrate, working well on transparencies. To improve its performance on rougher papers some printers include a special precoating panel on the ribbon, which lays down a clear wax base onto the substrate before the color printing. Like dye sublimation the thermal wax transfer ribbon is composed of repeating sets of full page sized panels. Generally it consists of yellow (Y), magenta (M), and cyan (C) and sometimes with a black (K) panel and/or a protective overcoat (O).

Because the pigmented wax either transfers from the ribbon or it does not, thermal wax transfer cannot produce the same range of tones and colors that dye sublimation can. The visual appearance of the output from the two processes is very different. Dye sublimation has a distinct grid like appearance, but with blurred individual pixels giving the smooth gradual changes in tone. The output from thermal wax transfer is made up of dots of solid color with distinct edges. Printed text particularly has the characteristic stepped edge associated with low resolution digital printing.

One of the products that dye sublimation and thermal wax transfer are often used to produce is plastic identity cards. These often incorporate dual technology ribbons, where dye sublimation panels are used to print images, such as the holder's photograph, and then text is printed with a thermal wax transfer ribbon. This has the advantage of continuous tone photographic quality images with solid crisp text, all printed in the one machine.

The process of matching thermal wax transfer output back to a specific printer is the same as for dye sublimation. That is, the identification of negative images on a used ribbon and the detection of heating defects with the thermal printhead elements.

Thermal Fusion. Thermal wax transfer is typically associated with color printing using semitransparent pigmented wax. There are other thermal ribbon based printing processes, sometimes grouped under the term thermal fusion, which use carbon or ink rather than wax. They also generally produce only black and white output. The printing mechanics are the same, with a printhead consisting of heated elements being brought into contact with a thermally sensitive ribbon. The heating of individual elements causes the carbon or ink to be released from the ribbon and transfer onto plain paper. The type of devices that use this basic technology can range from small heads and ribbons, such as thermal dot matrix or the IBM Quietwriter, through to page-width printheads and ribbons such as used in some facsimile machines. This technology is also found in specialist devices like labeling machines such as the Brother P-Touch range.

Again, a record of the output from all these devices remains on the ribbons they use and provides a possible avenue for the identification of the printer. Heating element failure, or disruption caused by an accumulation of debris, are the most likely causes of identifiable image faults.

Photocopiers

There have been, and still are, many different print technologies used to reproduce documents in photocopiers [19–22]. The process that became the most successful and widespread is dry toner electrophotographic, described in a previous section. This is still the process used in the majority of photocopiers today, although nowadays the analog photocopiers, which reflected the copied document's image onto the imaging drum, have been largely superseded by digital copiers, which perform the separate action of scanning the document and then sending this information to the printing unit which operates like a laser printer. Many of these devices are designed to also be used as networked printers, scanners, and facsimile machines.

Smaller versions of these devices, often using inkjet rather than electrophotographic technology, are being produced as low cost multifunction devices for the small and home office (SOHO) market.

The faults associated with the print technology used in most photocopiers have been discussed in the appropriate sections earlier. The main device specific characteristic of the photocopier that is of interest to an examiner is the glass platen. Marks on the platen, which may come from such things as scratches, dirt, ink residue, and opaquing fluid can be reproduced on every copy that machine subsequently makes. When they appear on a document they are often referred to as "*trash marks*". If the marks on the platen are noticed, they may be removed by cleaning the glass. But if the glass is not regularly cleaned and the marks are not obtrusive they can remain a feature of the output of that copier indefinitely.

Platen marks occur in the same position on each copied page. Often there are several different marks on the glass platen, which all reproduce onto the copied page as trash marks. Their fixed nature means the spatial relationship between the trash marks remains constant. Some authors liken these patterns of trash marks to constellations of stars. The fixed nature of platen marks helps to differentiate them from imaging drum marks in electrophotographic copiers, as the drum marks repeat along the page with every revolution of the imaging drum. This repeating pattern occurs if the circumference of the imaging drum is less than the length of the page, as is often the case. However, some high volume photocopiers use an electrophotographic imaging belt. The length of some of these belts means that there will be several pages gap between repetitions of a fault on the belt.

The gradual appearance and evolution of platen and drum marks can assist in dating documents. This requires a body of documents of known dates copied on the machine in question.

Duplication of platen marks provide evidence that a document has been copied twice on the same machine. This is sometimes seen in documents created using a "cut and paste" method.

Marks on the underside of the platen's cover can also be reproduced onto copied documents like platen marks. However, their occurrence tends to be unpredictable as it requires that they are not covered by the document being copied.

Another difference between platen and drum marks that is worth noting is what happens when a document is enlarged while being copied. Enlargement works by scanning a smaller than normal portion of the platen and then reproducing that smaller area across the entire normal page size. Hence, platen marks that are outside of the scanned area do not appear on the copy and those that are within the area being reproduced are enlarged. There is no change to the drum marks.

When taking samples from the photocopier you need to remember what you are capturing. Copies taken with the lid down and no paper on the platen capture marks from the underside of the lid, the platen, and the drum. Copies taken with a piece of paper fully covering the platen capture platen and drum marks. This piece of paper should be retained so that it can be determined that none of the copied marks came from a defect in the paper. Copies taken with the lid open and nothing on the platen produce black pages that show up any white faults on the imaging drum. If the copier is also set up as a printer, printing a largely white page captures any drum faults.

The potential for color copiers to be used for counterfeiting purposes prompted some manufacturers to introduce specific anticounterfeiting measures in their color copiers. These included software designed to recognize the patterns from major currencies and prevent the copier from reproducing those patterns. They also created what is in effect a built in drum fault. The copier deliberately prints a pattern of dots, unique to each machine, onto each copy it makes. The machine can then be identified by comparing the dot pattern in the questioned document to that seen on a sample print out from the machine. Through close collaboration with manufacturers, central banks and counterfeiting investigators can also use these proprietary codes to identify a machine from its printed output [23, 24].

Facsimile Machines

A facsimile machine, like a digital copier, consists of a scanner and a printing engine. The print technology employed in them is usually either direct thermal paper, thermal fusion, laser (electrophotographic), or inkjet. Direct thermal paper was the process often associated with facsimile machines with its rolls of shiny thermal paper that quickly faded. After a hesitant start, thermal fusion (using a thermal ribbon to print onto plain paper) took over from thermal paper

because it produced a permanent image and was cheaper than a laser print engine. While the laser print engine facsimiles still seem to dominate the business market, the rise of multifunction inkjet devices means inkjet printed faxes are possibly appearing more frequently.

One of the main differences between faxes and other documents is usually their resolution. While a multifunction device may have a high resolution scanner and printer, the standard protocols that enable different models of facsimile machines to talk to each other mean that transmissions usually take place at around 200 dpi or less. This is what causes the highly pixelated images we associate with facsimile transmissions.

The other distinctive part of the facsimile document is the Transmit Terminal Identifier (TTI) strip along the top of the document. This strip is supposed to carry information regarding the transmission such as the date/time, who from/to, telephone, and page numbers. However, the vast majority of this information can be edited and so cannot be relied on as accurate. Because manufacturers employ different fonts and layouts in these strips, an examination of these aspects may assist in determining the likely make or model of the sending machine. An important point to remember about the TTI strip is that it is generated by the sending machine electronically. That is, it has not been scanned. Accordingly, a TTI on an original facsimile document should be better defined than the body of the fax. If the TTI contains evidence of scanning (not just its normal pixelated appearance), then this suggests that the document has either been copied on a facsimile machine, or is not a first generation fax.

The types of individuating faults generally associated with facsimile machines are vertical black or white lines running through the page. If the line does not extend into the TTI strip, this shows the fault lies with a scanning element in the sending machine. If a scanning element has become obscured (from something picked up off a document), then it reproduces a black line. If there is a fault in the scanning element, it may produce a white or black line, depending on whether it is defaulting to off or on.

If a vertical line extends into the TTI strip, this indicates a fault with the receiving machine.

These faults, along with any print technology faults, can be used to link a document back to a particular facsimile machine. For those machines using a thermal ribbon, a negative image of anything printed on the machine (received faxes, direct copies, and transmission/machine status reports) is present on the ribbon.

References

[1] Hilton, O. (1982). *Scientific Examination of Questioned Documents*, Elsevier, New York.

[2] Brunelle, R.L. & Reed, R.W. (1984). *Forensic Examination of Ink and Paper*, Charles C Thomas, Springfield.

[3] Snape, K.W. (1980). Determination of the direction of ball-point pen motion from the orientations of burr striations in curved pen strokes, *Journal of Forensic Sciences* **25**(2), 386–389.

[4] Lindblom, B.S. (2006). Pens and pencils, in *Scientific Examination of Questioned Documents*, 2nd Edition, J.S. Kelly & B.S. Lindblom, eds, CRC Press, Boca Raton, pp. 147–158.

[5] Wilson, J.D., LaPorte, G.M. & Cantu, A.A. (2004). Differentiation of black gel inks using optical and chemical techniques, *Journal of Forensic Sciences* **49**(2), 364–370.

[6] Gernandt, M.N. & Urlaub, J.J. (1996). An introduction to the gel pen, *Journal of Forensic Sciences* **41**(3), 503–504.

[7] LaPorte, G.M. (2004). The use of an electrostatic detection device to identify individual and class characteristics on documents produced by printers and copiers – a preliminary study, *Journal of Forensic Sciences* **49**(3), 610–620.

[8] Hilton, O. (1962). Identification of the work from an IBM Selectric typewriter, *Journal of Forensic Sciences* **7**(3), 286–302.

[9] Behrendt, J.E. (1988). Class defects in printwheel typescript, *Journal of Forensic Sciences* **33**(2), 328–335.

[10] Allen, M.J. & Hardcastle, R.A. (1990). The distribution of damage defects among characters of printwheel typing elements, *Forensic Science International* **47**, 249–259.

[11] Hahn, G.H. (1974). Paper fiber impressions on carbon tape ribbons, *Journal of Forensic Sciences* **19**(1), 136–141.

[12] Casey, M.A. & Purtell, D.J. (1976). IBM correcting selectric typewriter: an analysis of the use of correctable film ribbon in altering typewritten documents, *Journal of Forensic Sciences* **21**(1), 208–212.

[13] Gerhart, F.J. (1989). Methods of associating typewriter ribbons and correcting tapes with a questioned text, *Journal of Forensic Sciences* **34**(5), 1183–1195.

[14] Blanco, J.A. (1993). Identifying documents printed by dot matrix computer printers, *Forensic Science International* **59**, 35–47.

[15] Allen, M.J. (1987). Dot-matrix printers, *Forensic Science International* **35**, 283–295.

[16] Arbouine, M.W. & Day, S.P. (1994). The use of drum defects to link laser printed documents to individual laser printers, *Journal Forensic Science Society* **34**, 99–104.
[17] Gerhart, F.J. (1992). Identification of photocopiers from fusing roller defects, *Journal of Forensic Sciences* **37**(1), 130–139.
[18] LaPorte, G.M., Wilson, J.D., Mancke, S.A., Payne, J.A., Ramotowski, R.S. & Fortunato, S.L. (2003). The forensic analysis of thermal transfer printing, *Journal of Forensic Sciences* **48**(5), 1163–1171.
[19] Holland, N.W. (1984). Photocopy classification and identification, *Journal Forensic Science Society* **24**, 23–41.
[20] Crown, D.A. (1989). Dates of introduction of specific photocopy machines, *Journal of Forensic Sciences* **34**(1), 110–141.
[21] James E.L. (1987). Classification of photocopy machines by physical characteristics, *Journal of Forensic Sciences* **32**(5), 1293–1304.
[22] Shiver, F.C. & Nelson, L.K. (1991). Nondestructive differentiation of full-color photocopies, *Journal of Forensic Sciences* **36**(1), 145–152.
[23] Li, C.K. & Leung, S.C. (1998). The identification of colour photocopiers: a case study, *Journal of American Society of Questioned Document Examiners* **1**(1), 8–11.
[24] Tweedy, J.S. (2001). Class characteristics of counterfeit protection system codes of color laser copiers, *Journal of American Society of Questioned Document Examiners* **4**(2), 53–66.
[25] Lindblom, B.S. (2006). Facsimile machines, in *Scientific Examination of Questioned Documents*, 2nd Edition, J.S. Kelly & B.S. Lindblom, eds, CRC Press, Boca Raton, pp. 227–234.

GORDON A.I. SHARFE



Y-Chromosome Short Tandem Repeats

Introduction

The physical map of the Y chromosome consists of three distinct regions: the euchromatin, the heterochromatin and pseudoautosomal regions (PARs). The PARs are located at the telomeric ends of the chromosome and constitute ~5% of the Y chromosome sequence. PARs are the only regions of the chromosome to undergo recombination with the X chromosome during male meiotic events. The remaining 95% of the Y chromosome does not undergo recombination referred to as the *nonrecombining region of the Y chromosomes* (*NRY*) and comprises euchromatin sequences (that contain all of the chromosome's known genes) and the functionally inert heterochromatin. Thus, the entire NRY is inherited in a patrilinear manner, with a haplotype of physically and genetically linked genetic markers passed essentially unchanged, barring any rare mutations, from father to son [1].

The unique biology of the genetic markers present on the NRY region has resulted in their widespread use in determining patrilineal relationships within and between populations to aid in understanding of human migration and evolution [2, 3]. Y-chromosome microsatellites or short tandem repeats (STRs) are one class of these genetic markers that have been incorporated into a variety of multiplex PCR assays for potential forensic casework applications [4–12]. Their intended use is not to replace autosomal STR loci but to employ them in casework situations in which the traditional autosomal loci would not be expected to yield sufficient probative information. For example, Y-chromosome short tandem repeats (Y-STRs) are particularly useful for cases involving admixed male and female DNA for which a separation of the male- and female-derived cells is unsuccessful or not feasible [13–15].

Commonly Used Y-STR Markers

A major international multicenter study of 13 candidate Y-STR markers resulted in recommendations for the use of nine core loci for standard forensic haplotyping [3]. These loci are referred to as the *minimal haplotype loci* (*MHL*), and include DYS19, DYS385 (a) and (b), DYS389 (I and II), DYS390, DYS391, DYS392, and DYS393 (*see* **Mini-STRs** for a description of the naming conventions for different loci and alleles). Despite the initial utility of this set of markers in forensic casework, there was a need for additional Y-STR loci to be used in conjunction with these markers to improve the discriminatory capacity of Y-STR testing. In 2003, the use of two additional loci, DYS438 and DYS439, was recommended by the Scientific Working Group on DNA Analysis Methods (SWGDAM) [16]. The MHL loci plus the two additional loci are referred to as the "SWGDAM core loci."

Various combinations of the common Y-STR loci have been incorporated into commercially available multiplex systems (Table 1) [6, 7, 10–12]. The

Table 1 Commercial Y-STR multiplex amplification kits

Commercial kit	Manufacturer	Incorporated loci
Commonly used commercial kits		
PowerPlex® Y	Promega Corporation, Madison, WI, USA	DYS19, DYS385a/b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439
AmpF*l*STR® Yfiler™	Applied Biosystems, Foster City, CA, USA	DYS19, DYS385a/b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS456, DYS458, DYS448, DYS635(C4), H4
Previously available commercial kits		
Y-Plex™ 5	Reliagene Technologies Inc., New Orleans, LA, USA	DYS19, DYS385a/b, DYS389II, DYS390, DYS391, DYS393
Y-Plex™ 6	Reliagene	DYS389I, DYS389II, DYS392, DYS393, DYS438, DYS439
Y-Plex™12	Reliagene	DYS19, DYS385a/b, DYS389I, DYS389II; DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439

Promega PowerPlex® Y kit, which contains the SWDGAM core loci plus DYS437, and the Applied Biosystems AmpF*l*STR® Yfiler™ kit, which also contains the SWGDAM core loci plus several additional markers (DYS437, DYS448, DYS456, DYS458, C4 (DYS635), and H4) are two of the most commonly used multiplex kits in the US [6, 7]. Figure 1 gives an example of a 17-locus Y-STR profile from an individual obtained using the Yfiler™ kit. Three additional Y-STR systems were developed by Reliagene Technologies (Y-Plex™5, Y-Plex™ 6, and Y-Plex™ 12) but are no longer commercially available [10–12]. All of these kits have been fully validated for forensic use as required by US national DNA standards and in accordance with SWGDAM validation guidelines [6, 7, 10–12]. They have been optimized to exhibit a high degree of sensitivity and specificity for male DNA, even in the presence of a vast excess of female DNA.

Interpretation of a Y-STR Match

Since the loci in the NRY region of the Y chromosome are inherited as a physical block of physically linked haplotypes, independent assortment does not occur and the loci are in linkage disequilibrium with each other. As a result the product rule cannot be used to determine multilocus Y-STR frequencies. A counting method is commonly employed in order to estimate the frequency of occurrence of a particular Y-STR haplotype. With the counting method, the frequency of the haplotype is the number of times the haplotype is observed divided by the number of samples in the database. Alternative methods for Y-STR haplotype frequency estimates have been suggested including one based upon a mismatch distribution approach that allows the evaluation of the proportion of pairs of Y-STR haplotypes that are prone to become identical by state (IBS), in one generation, by recurrent mutation [17].

To correct for possible sample effects when using databases of limited size, the use of confidence intervals around the haplotype frequency estimate has been recommended. Reporting a haplotype frequency without associated confidence intervals may still be acceptable as a factual statement regarding observations in a particular database. If a confidence interval is applied and the haplotype has been observed in the database, the upper 95% confidence limit would be calculated using the following formula

$$p + 1.96\sqrt{\frac{(p)(1-p)}{n}} \qquad (1)$$

where p is x/n (n is the database size, x is the number of times observed in the database). If

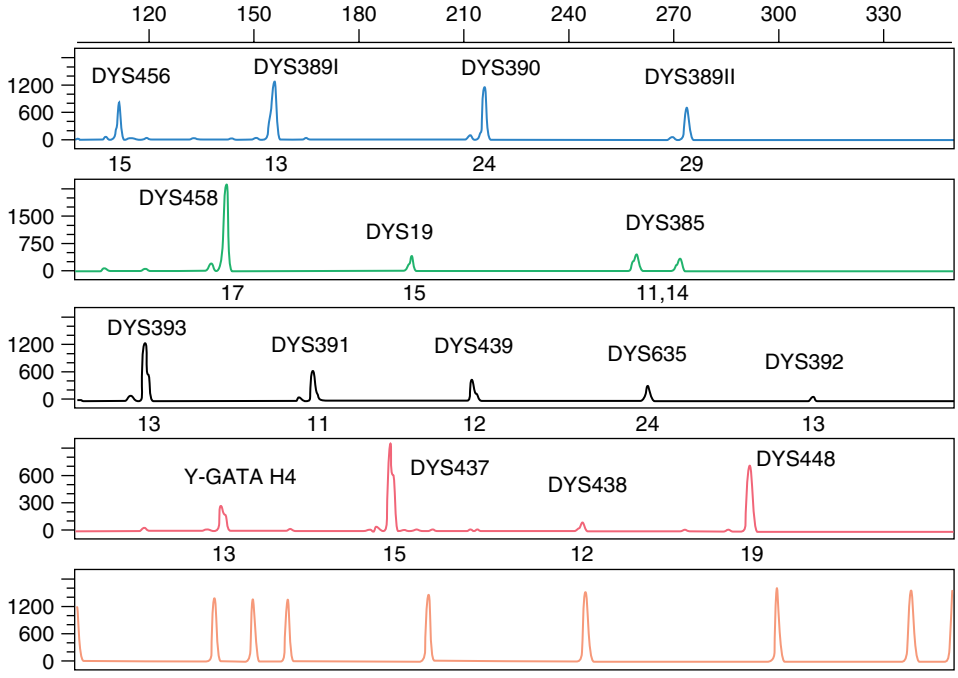


Figure 1 Y-STR profile using the Applied Biosystems AmpF*l*STR® Yfiler™ Amplification Kit. 17 Y-STR loci are coamplified in a single reaction, separated by capillary electrophoresis, and displayed as an electropherogram. The *x* axis represents fragment size (base pairs) and the *y* axis represents signal intensity (relative fluorescence intensity (RFU)). Each locus is labeled with a fluorescent dye: top channel – 6-FAM (blue); second channel – VIC (green); third channel – NED (yellow); fourth channel – PET (red); bottom channel – LIZ (orange) – internal size standard

the haplotype has not been previously observed in the database, the formula used to calculate the upper 95% confidence limit would be the following

$$1 - (0.05)^{1/n} \qquad (2)$$

where n is the size of the database. A simplified alternative formulation can also be used in this instance; $3/n$ (n = size of the database). This value is close to the earlier formula and for 95% of the time the real frequency is less than that estimate. Owing to population differentiation between subpopulations within the broader ethnic groups an appropriate θ (or F_{ST}) correction, akin to the situation with autosomal STRs, may be applicable [18, 19].

Empirical studies have confirmed the nonlinkage of Y-STR haplotypes and autosomal STR profiles [18]. Therefore, in cases with both autosomal and Y-STR analysis performed, it is possible to multiply the autosomal STR profile frequency with the upper confidence limit bounded Y-STR haplotype frequency to produce an overall frequency estimate for the combined profiles.

Casework Applications

Y-STRs can be particularly useful when trying to determine the genetic profile of the male donor in a male/female DNA admixture when the female DNA component is present in vast excess (e.g., ≥100-fold) and when traditional autosomal STR analysis fails or is not well suited for the particular analysis [13–15]. Samples containing a mixture of body fluids other than semen, such as in saliva/saliva mixtures, saliva/vaginal secretion mixtures in cases of oral sodomy, or fingernail scrapings with cells from the perpetrator may not be suitable for autosomal STR analysis. In these types of samples, unlike sperm containing samples, a differential lysis approach to separate the male and female cells is not possible. The minor male component in nondifferentially extracted

samples would often not be detectable with the PCR-based autosomal STR systems routinely used. This is due to the kinetics of the PCR process itself, which does not permit minor components to be detected at low levels (i.e., $\leq 1/20$) because of titration of critical reagents by the major DNA component [20]. Autosomal STR analysis may also fail with some semen containing samples in which few sperms are present, or are present in an extremely fragile state, such as in extended interval (i.e.,>48 h) postcoital samples. Differential extraction of these particular samples may not result in male profile being obtained due to a combination of premature lysis of the sperm's cellular constituents into the nonsperm fraction and to sperm loss during the physical manipulations required of the DNA isolation process. Therefore the use of Y-STR systems can eliminate the need for differential extractions (separating the male and female fractions), thus reducing the potential loss of the small amount of male DNA that may be present in such samples.

Y-STRs are hemizygous in nature, with only one allele being found at each single-copy locus. Therefore, Y-STRs can be useful in determining the number of male contributors in an admixed DNA sample [8, 21–23]. Mixture analysis using autosomal analysis becomes quite complex when the number of donors exceeds two. However, the presence of multiple alleles at each Y-STR locus in a sample can give a more precise indication of the number of male contributors.

Y-chromosome polymorphisms may also be useful in criminal paternity cases, missing persons cases, and in the identification of victims involved in mass disasters where a reference sample from the (male) victim may not be available [24–30]. Y-chromosome markers are inherited as a physical block unchanged from one generation to the next (barring any random mutation events). Therefore, a Y-STR profile obtained by typing a male relative in the same lineage should be the same as that of the victim. Additionally, Y-STR analysis may provide additional discrimination in situations where a complete autosomal profile is not obtained, such as in mixtures or partially degraded samples [20, 31–35].

Y-STR analysis may aid in familial searches, which refers to the use of low stringency autosomal STR matching of a crime scene profile with an offender DNA database to identify possible biological relatives of the donor of the crime scene sample [36, 37]. With the large number of samples present in offender databases, several individuals from different candidate families can arise as potential relatives to the donor of the crime scene profile. Subsequent Y-STR analysis of the crime scene profile and the offenders in the database identified as potential relatives should eliminate most of the adventitious candidates. This not only facilitates the efficiency of the investigative process but also precludes unnecessary invasion of the privacy of noninvolved families. However, coincidental matches between unrelated individuals still occur with the limited number of Y-STR loci typically used in forensic casework. Therefore, it may be necessary to use additional highly discriminating "noncore" Y-STR loci in order to resolve coincidental matches and provide a more accurate indication of true relatedness.

References

[1] Tilford, C.A., Kuroda-Kawaguchi, T., Skaletsky, H., Rozen, S., Brown, L.G., Rosenberg, M. McPherson, J.D., Wylie, K., Sekhon, M., Kucaba, T.A., Waterston, R.H. & Page, D.C. (2001). A physical map of the human Y chromosome, *Nature* **409**(6822), 943–945.

[2] De, K.P., Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G, Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, M., Krawczak, M., Leim, K., Meuser, S., Meyer, E., Oesterreich, W., Pandya, A., Parson, W., Penacino, G., Perez-Lezaun, A., Piccinini, A., Prinz, M. & Roewer, L. (1997). Chromosome Y microsatellites: population genetic and evolutionary aspects, *International Journal of Legal Medicine* **110**(3), 134–149.

[3] Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, M., Krawczak, M., Leim, K., Meuser, S., Meyer, E., Oesterreich, W., Pandya, A., Parson, W., Penacino, G., Perez-Lezaun, A., Piccinini, A., Prinz, M., Schmitt, C. & Roewer, L. (1997). Evaluation of Y-chromosomal STRs: a multicenter study, *International Journal of Legal Medicine* **110**(3), 125–129.

[4] Corach, D., Filgueira, R.L., Marino, M., Penacino, G. & Sala, A. (2001). Routine Y-STR typing in forensic casework, *Forensic Science International* **118**(2–3), 131–135.

[5] Kayser, M., Brauer, S., Willuweit, S., Schadlich, H., Batzer, M.A., Zawacki, J., Prinz, M., Roewer, L. & Stoneking, M. (2002). Online Y-chromosomal short tandem repeat haplotype reference database (YHRD) for U.S. populations, *Forensic Science International* **47**(3), 513–519.

[6] Krenke, B.E., Viculis, L., Richard, M.L., Prinz, M., Milne, S.C., Ladd, C., Gross, A.M., Gornall, T., Frappier, J.R., Eisenberg, A.J., Barna, C., Aranda, X.G., Adamowicz, M.S. & Budowle, B. (2005). Validation of a male-specific, 12-locus fluorescent short tandem repeat (STR) multiplex, *Forensic Science International* **148**(1), 1–14.

[7] Mulero, J.J., Chang, C.W., Calandro, L.M., Green, R.L., Li, Y., Johnson, C.L. & Hennessy, L.K. (2006). Development and validation of the AmpFlSTR Yfiler PCR amplification kit: a male specific, single amplification 17 Y-STR multiplex system, *Forensic Science International* **51**(1), 64–75.

[8] Prinz, M. & Sansone, M. (2001). Y chromosome-specific short tandem repeats in forensic casework, *Croatian Medical Journal* **42**(3), 288–291.

[9] Roewer, L., Krawczak, M., Willuweit, S., Nagy, M., Alves, C., Amorim, A., Anslinger, K., Augustin, C., Betz, A., Bosch, E., Caglia, A., Carracedo, A., Corach, D., Dekairelle, A.F., Dobosz, T., Dupuy, B.M., Furedi, S., Gehrig, C., Gusmao, L., Henke, J., Henke, L., Hidding, M., Hohoff, C., Hoste, B., Jobling, M.A., Kargel, H.J., De, K.P., Lessig, R., Liebeherr, E., Lorente, M., Martinez-Jarreta, B., Nievas, P., Nowak, M., Parson, W, Pascali, V.L., Penacino, G., Ploski, R., Rolf, B., Sala, A., Schmidt, U., Schmitt, C., Schneider, P.M., Szibor, R., Teifel-Greding, J. & Kayser, M. (2001). Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes, *Forensic Science International* **118**(2–3), 106–113.

[10] Shewale, J.G., Nasir, H., Schneida, E., Gross, A.M., Budowle, B. & Sinha, S.K. (2004). Y-chromosome STR system, Y-PLEX 12, for forensic casework: development and validation, *Forensic Science International* **49**(6), 1278–1290.

[11] Sinha, S.K., Nasir, H., Gross, A.M., Budowle, B. & Shewale, J.G. (2003). Development and validation of the Y-PLEX 5, a Y-chromosome STR genotyping system, for forensic casework, *Forensic Science International* **48**(5), 985–1000.

[12] Sinha, S.K., Budowle, B., Arcot, S.S., Richey, S.L., Chakrabor, R., Jones, M.D., Wojtkiewicz, P.W., Schoenbauer, D.A., Gross, A.M., Sinha, S.K., Shewale, J.G. (2003). Development and validation of a multiplexed Y-chromosome STR genotyping system, Y-PLEX 6, for forensic casework, *Forensic Science International* **48**(1), 93–103.

[13] Betz, A., Bassler, G., Dietl, G., Steil, X., Weyermann, G. & Pflug, W. (2001). DYS STR analysis with epithelial cells in a rape case, *Forensic Science International* **118**(2–3), 126–130.

[14] Dekairelle, A.F. & Hoste, B. (2001). Application of a Y-STR-pentaplex PCR (DYS19, DYS389I and II, DYS390 and DYS393) to sexual assault cases, *Forensic Science International* **118**(2–3), 122–125.

[15] Martin, P., Albarran, C., Garcia, O., Garcia, P., Sancho, M. & Alonso, A. (2000). Application of Y-STR analysis to rape cases that cannot be solved by autosomal STR analysis, *Progress in Forensic Genetics* **8**, 526–528.

[16] SWGDAM Y-STR Subcommittee (2007). Report on the current activities of the scientific working group on DNA analysis methods Y-STR subcommittee, *Forensic Science Communications* **6**(3), 1–2.

[17] Pereira, L., Prata, M.J. & Amorim, A. (2002). Mismatch distribution analysis of Y-STR haplotypes as a tool for the evaluation of identity-by-state proportions and significance of matches–the European picture, *Forensic Science International* **130**(2–3), 147–155.

[18] Budowle, B., Adamowicz, M., Aranda, X.G., Barna, C., Chakraborty, R., Cheswick, D., Dafoe, B., Eisenberg, A., Frappier, R., Gross, A.M., Ladd, C., Lee, H.S., Milne, S.C., Meyers, C., Prinz, M., Richard, M.L., Saldanha, G., Tierney, A.A., Viculis, L. & Krenke, B.E. (2005). Twelve short tandem repeat loci Y chromosome haplotypes: genetic analysis on populations residing in North America, *Forensic Science International* **150**(1), 1–15.

[19] National Research Council (1996). *The Evaluation of Forensic DNA Evidence*, National Academy Press, Washington, DC.

[20] Prinz, M., Boll, K., Baum, H. & Shaler, B. (1997). Multiplexing of Y chromosome specific STRs and performance for mixed samples, *Forensic Science International* **85**(3), 209–218.

[21] Daniels, D.L., Hall, A.M. & Ballantyne, J. (2004). SWGDAM developmental validation of a 19-locus Y-STR system for forensic casework, *Forensic Science International* **49**(4), 668–683.

[22] Hanson, E.K. & Ballantyne, J. (2004). A highly discriminating 21 locus Y-STR "megaplex" system designed to augment the minimal haplotype loci for forensic casework, *Journal of Forensic Sciences* **49**(1), 40–51.

[23] Hanson, E.K., Berdos, P.N. & Ballantyne, J. (2006). Testing and evaluation of 43 "noncore" Y chromosome markers for forensic casework applications, *Journal of Forensic Sciences* **51**(6), 1298–1314.

[24] Alshamali, F., Qader Alkhayat, A., Budowle, B. & Watson, N. (2004). Y chromosome in forensic casework and paternity testing, *Progress in Forensic Genetics* **10**, 353–356.

[25] Andelinovic, S., Sutlovic, D., Erceg, I., Skaro, V., Ivkosic, A., Paic, F., Rezic, B., Finis-Gojanovic, M. & Primorac, D. (2005). Twelve-year experience in identification of skeletal remains from mass graves, *Croatian Medical Journal* **46**(4), 530–539.

[26] Anjos, M.J., Carvalho, M., Andrade, L., Lopes, V., Serra, A. & Batista, L., Oliveira, C., Tavares, C., Balsa, F., Corte-Real, F., Vieira, D.N. & Vide, M.C. (2004). Individual genetic identification of biological samples: a case of an aircraft accident, *Forensic Science International* **146**, S115–S117.

[27] Jobling, M.A., Pandya, A. & Tyler-Smith, C. (1997). The Y chromosome in forensic analysis and paternity

testing, *International Journal of Legal Medicine* **110**(3), 118–124.
[28] Kayser, M., Kruger, C., Nagy, M., Geserick, G., de Knijff, P. & Roewer, L. (1998). Y-chromosomal DNA-analysis in paternity testing: experiences and recommendations, *Progress in Forensic Genetics* **7**, 494–496.
[29] Li, J. (2004). Chromosome STR genetic markers in paternity identification, *Zhong Nan Da Xue Xue Bao Yi Xue Ban* **29**(4), 432–434.
[30] Rolf, B., Keil, W., Brinkmann, B., Roewer, L. & Fimmers, R. (2001). Paternity testing using Y-STR haplotypes: assigning a probability for paternity in cases of mutations, *International Journal of Legal Medicine* **115**(1), 12–15.
[31] Berger, B., Niederstätter, H., Köchl, S., Steinlechner, M. & Parson, W. (2003). Male/female DNA mixtures: a challenge for Y-STR analysis, *Progress in Forensic Genetics* **9**, 295–299.
[32] Cerri, N., Ricci, U., Sani, I., Verzeletti, A. & De, F.F. (2003). Mixed stains from sexual assault cases: autosomal or Y-chromosome short tandem repeats? *Croatian Medical Journal* **44**(3), 289–292.
[33] Parson, W., Niederstatter, H., Brandstatter, A. & Berger, B. (2003). Improved specificity of Y-STR typing in DNA mixture samples, *International Journal of Legal Medicine* **117**(2), 109–114.
[34] Tsuji, A., Ishiko, A., Ikeda, N. & Yamaguchi, H. (2001). Personal identification using Y-chromosomal short tandem repeats from bodily fluids mixed with semen, *American Journal of Forensic Medicine and Pathology* **22**(3), 288–291.
[35] Yoshida, Y., Fujita, Y. & Kubo, S. (2004). Forensic casework of personal identification using a mixture of body fluids from more than one person by Y-STRs analysis, *Journal of Medical Investigation* **51**(3–4), 238–242.
[36] Bieber, F.R., Brenner, C.H. & Lazer, D. (2006). Human genetics. Finding criminals through DNA of their relatives, *Science* **312**(5778), 1315–1316.
[37] Greely, H.T., Riordan, D.P., Garrison, N.A. & Mountain, J.L. (2006). Family ties: the use of DNA offender databases to catch offenders' kin, *Journal of Law Medicine and Ethics* **34**(2), 248–262.

JACK BALLANTYNE AND ERIN K. HANSON

Y-STR *see* Y-Chromosome Short Tandem Repeats

Young Offenders *see* Death Penalty and Age

Glossary

AAFS	American Academy of Forensic Sciences, representing over 5000 scientists including physicians, attorneys, dentists, toxicologists, psychiatrists, engineers, and educators, in the application of science to the law.
AAS	Atomic Absorption Spectroscopy.
Accompanying Drop	A small droplet that forms between a larger drop and its blood source when the larger drop breaks away and falls free.
Accuracy	The degree of closeness between a measured or calculated quantity and the actual value, cf Precision.
ACE-V	Analysis, Comparison, Evaluation, Verification; a method used by fingerprint examiners (especially in the United States) to standardize evidence examination steps.
Admissibility	Admissibility of expert evidence in legal proceedings is a matter to be determined by the judge (the finder of law) rather than the jury (the finder of fact). Expert evidence that is deemed inadmissible is excluded from the proceedings.
Admission	(i) The acceptance of a fact in a legal procedure; (ii) A confession of wrongdoing or of guilt; (iii) The acknowledgment of the truth of a fact.
Agonal Respiration	Abnormal breathing, characterized by irregular, gasping, shallow, and slow breaths interrupted by irregular pauses.
Algorithm	Mathematical routine used in computer processing of data to sort or classify information.
Aliquot	A portion of a total amount of solution (e.g., the 10-ml solution was added to the beaker in 2-ml aliquots).
Alkaline Phosphatase	An enzyme produced in the liver, bone, and placenta.

Allele An alternative form of a gene (one member of a pair), that occupies a given locus (location) on a chromosome.

Allelic Frequency The frequency with which a particular allele appears at a specific DNA locus (location) compared with all the possible alleles in a defined population. It is generally expressed as a percentage or proportion of the specified population.

Amplicon Pieces of DNA created by natural amplification processes (such as natural gene duplication) or artificial means (such as a polymerase chain reaction or ligase chain reaction).

Amplification Multiplication of the number of DNA molecules from an initial "template".

Analyte The substance that is being identified or measured in an analytical procedure (e.g., a titration).

Anastomosis An opening between two normally distinct spaces or organs, caused by surgery, pathology, or trauma.

Angle of Impact The internal angle at which blood strikes a target surface relative to the horizontal plane of that target surface.

ANSI American National Standards Institute. This organization promotes and facilitates voluntary consensus standards and conformity assessment systems across a range of areas, and oversees their integrity.

Aptamer Oligonucleic acid or peptide macromolecules whose shape assists in binding to a specific target molecule.

Arterial Spurting (or Gushing) Characteristic bloodstain patterns on a target surface resulting from blood exiting under pressure from a breached artery. Characterized by their size and shape.

ASCII American Standard Code for Information Interchange. A coding system that converts font characters to standard numeric values.

Assay A technique in which an analyte's properties or concentration are measured. Includes immunoassays (e.g., see CEDIA, ELISA), bioassays, stem cell assays, and microbiological assays.

ASTM American Society for Testing and Materials, now ASTM International. A voluntary standards development organization producing technical standards for services, systems, products, and materials across a range of industries.

Atomic Number The number of protons in the nucleus of an atom; generally equal to the number of electrons revolving around the nucleus.

Atomic Weight The relative average weight of an atom of an element, compared to a value of exactly 12 for the principal stable isotope of carbon (carbon-12).

Autosome Any chromosome other than the sex chromosomes (X and Y). Humans have 22 pairs of autosomes.

Back Spatter Blood that is directed back toward its source of energy; often associated with gunshot wounds of entrance.

Bacteria Microscopic organisms in which single cells do not have a membrane-bounded nucleus or any other membrane-bounded organelles like mitochondria and chloroplasts.

Blind Verification The independent application of an accepted methodology to the analysis reached by another examiner-researcher who has no expectation or knowledge of the conclusion of the original examiner. A verification is double-blind when the reviewing examiner is not aware that any analysis has been effected earlier by another person.

Blood The "circulating tissue" of the body; the fluid and its suspended formed elements that are circulated through the heart, arteries, capillaries, and veins.

Blood Clot A blood clot is formed as a result of a complex mechanism involving the plasma protein fibrinogen, platelets, and other clotting factors. It is observed visually as an insoluble network of fibrous material (fibrin and red blood cells).

Bloodstain The resulting transfer when liquid blood comes into contact with a surface, or when a moist or wet surface comes into contact with dried blood.

Bubble Rings Rings in blood that result when blood containing air bubbles dries retaining the circular configuration of the bubbles.

Buccal Swab Device similar to a cotton bud used to obtain a DNA sample from a person by wiping the inside of the cheek.

Buffer Solution that resists change to its pH (acidity/basicity) when small amounts of acid or base are added to it, or when it is diluted.

Burden of Proof The duty to present evidence to support the elements of a cause of action or an asserted legal defense. In criminal cases, the primary burden of proof of guilt of the accused is on the prosecution; in civil cases, the primary burden of establishing the existence of the elements of a cause of action is on the plaintiff. The burden of proof may shift during legal proceedings from one party to another, in accordance with specific rules of procedure. See also Quantum of Proof.

Cast-Off Pattern Blood that has been projected onto a surface as a result of being cast off or flung from an object in motion.

CEDIA Cloned Enzyme Donor Immunoassay. A homogeneous assay used in biochemistry that allows highly sensitive detection of low-molecular-mass analytes without separation steps. The bacterial enzyme b-galactosidase is genetically engineered into two inactive fragments (an enzyme donor and enzyme receptor). The enzyme donor fragment is able to covalently bind to both analytes and ligands, and this binding does not affect the ability of the enzyme donor to then reassociate with the enzyme receptor to form an active enzyme. The amount of enzyme formed is directly proportional to the concentration of analyte in the sample, and is detectable by spectrophotometric methods.

Chimerism Extremely rare disorder in which a single organism has genetically different chromosomes. It can manifest as two sets of DNA, organs that do not match the DNA of the rest of the organism, the presence of both male and female sex organs, or small patches of DNA throughout the body that are different from the DNA of the rest of the body.

Chromatin A combination of DNA and proteins found in the cell nucleus that together make up chromosomes. The name is derived from "colored material" because chromatin can be easily stained and thus visualized.

Chromatography A process of separation of compound materials by percolation through a selectively absorbing medium (paper, liquid, gas-liquid, gas-solid, etc.).

Chromosome A very long DNA molecule with associated proteins that is located in an organism's cells and carries portions of the hereditary information of an organism. Humans have 23 paired chromosomes.

Chronotrope That which affects the heart rate; positive chronotropes increase heart rate and negative chronotropes decrease the heart rate.

Circumstantial Evidence — Indirect evidence. Evidence of one fact from which a different but related fact may be inferred.

CODIS Database — Combined DNA Index System database, funded by the Federal Bureau of Investigation.

Compression Transfer Pattern — A contact bloodstain pattern created when a wet, bloody surface contacts a second surface with minimal lateral motion. A recognizable mirror image, or at least a portion of geometry of the original surface, is transferred to the second surface.

Confession — An admission of guilt or of an incriminating fact following the commission of a crime. In most jurisdictions, confessions that are the product of police questioning can be used as evidence only if it is established that the confession was voluntarily made and not the product of coercive methods, torture, or inhumane treatment. In most countries, there exists a right to remain silent during police interrogation and a corresponding right not to be forced to incriminate oneself.

Confidence Interval — A statistical estimate of an interval that is thought likely to contain the true value of whatever is being estimated or measured. The more confidence sought in the reliability of the results, the larger the interval is likely to be.

Contamination — Error in which unrelated material is introduced into a sample, thus tainting it (e.g., sneezing over a DNA sample to be analyzed in a forensic laboratory, thus contaminating the sample with saliva).

Controls — A set of positive or negative experimental conditions that ensure that a scientific experiment is working correctly. The negative control should always produce a negative result and the positive control should always produce a positive result.

CRFP — Council for the Registration of Forensic Practitioners (UK).

Defendant — The person or entity against whom a legal action is brought by another.

Density — The mass-per-unit volume of a substance, usually expressed in grams per cubic centimeter, under special or standard conditions of pressure and temperature.

Direct Evidence — Evidence that, if believed, establishes a fact without resort to any further inferences of other facts. Example : an eyewitness testimony as to the identity of a perpetrator of an act.

Directionality — Relating to or indicating the direction that blood was traveling in space from its point of origin.

Directionality Angle The angle between the long axis of a bloodstain and a predetermined line on the plane of the target surface taken as 0°.

DNA Advisory Board (US) Appointed by Director of the FBI to set and oversee quality assurance standards in US DNA crime laboratories and for the CODIS database.

Draw-Back Effect Blood in the barrel of a firearm, which has been drawn back into the muzzle of a firearm that was discharged when its muzzle was in contact or near contact with a victim.

Drip Pattern A bloodstain pattern that results from blood dripping into blood.

Drop A volume of blood of sufficient weight to overcome its surface tension and fall free from the mass of blood from which it was formed; typically a volume of 0.05 ml.

Droplet A blood drop volume less than the typical volume of 0.05 ml that was produced by energy exceeding that of gravitational attraction alone.

EAP Extensible Authentication Protocol. Universal authentication framework used commonly in wireless networks and point-to-point network connections.

ECD Electron Capture Detector, used in gas chromatography.

EIA Enzyme Immunoassay. This terms covers a range of biochemical assays in which antigen and antibodies can be quantifiably analyzed. Antigen-antibody reactions are indicated by a color change produced when enzymes react with a substrate (or inhibition occurs). Examples of EIA include EMIT and ELISA.

Electron Microprobe Analysis A research instrument for analyzing any element heavier than sodium in the periodic table.

Electrophoresis Technique to separate charged molecules by migration on a support medium exposed to an electric potential.

ELISA Enzyme-Linked Immunosorbent Assay. A technique used to detect the presence of antigens or antibodies in a sample.

EMIT Enzyme Multiplied Immunoassay Technique.

ENFSI European Network of Forensic Science Institutes. Organizes proficiency tests, seminars, collaborative studies, best practice manuals and glossaries of forensic terms for forensic science organizations throughout Europe.

Enzyme Protein catalysts that speed up the rate of a reaction.

ESDA Electrostatic Detection Apparatus.

Euchromatin A relatively low-density chromatin (i.e., DNA arranged with proteins to make chromosomes), which also has a relatively high concentration of genes.

Expirated Blood Blood that is blown out of the nose, mouth, or a wound as a result of air pressure and/or air flow, which is the propelling force.

Fact Finder In a legal proceeding, the person or persons who determine all disputed factual issues after considering all the evidence. In jury trials, the jury is the fact finder; in bench trials (nonjury) the trial judge is the fact finder.

Familial Search/Kinship Testing Search of a DNA database to identify other people who share a number of alleles or a particular allele (e.g., rare alleles). The search can identify related people or random matches.

Felony Murder A form of common law murder where the killing occurs accidentally while the accused is engaged in the commission of a common law felony. Proof of the intent necessary for the triggering felony substitutes for proof of malice, so that the unintended death can be prosecuted as murder. Even in countries that have the common law as the basis of their criminal justice system, most criminal offenses are today defined by statute.

FEPAC Forensic Educational Program Accreditation Commission. Now a standing committee of the American Academy of Forensic Sciences. FEPAC accreditation is voluntary but is established and standardized. FEPAC deals with accreditation, including external validation, helping students select institutions, and serving as a means for forensic scientists and employers to be able to judge the credentials of graduates. It also can help improve program quality, and it provides a certain demonstrable level of competency to the graduates.

FID Flame Ionization Detector. A type of detector used in gas chromatography, which is particularly useful for the detection and analysis of hydrocarbons and organic compounds with a high proportion of carbon.

Flight Path The path of blood as it moves through space from its origin to the surface it strikes.

Forensic Belonging to, used in, or suitable to courts of law or to public discussion or debate.

Forensic Science The use of scientifically based principles as these principles intersect with and provide evidence for legal proceedings.

Forward Spatter Blood that travels in the same direction as the source of energy or force causing the spatter. Most often associated with gunshot wounds of exit.

FPIA Fluorescence Polarization Immunoassay.

Gas Chromatography Laboratory method and instrument for the separation of complex mixtures to produce a chromatogram, which, by comparison with ASTM chromatogram standards, enables tentative identification of a sample examined.

GC/MS See Gas Chromatography and Mass Spectrometry.

Genotype The specific alleles that make up the genetic code for an individual. Different from phenotype, which refers to the outward appearance of the individual and which is affected by the genotype as well as the environment.

Grain Equivalent to 0.00648 g in weight. Also, Standard American unit of weight for bullet weight and weight of powder charge; 7000 grains per pound.

Gram Metric unit of weight; one gram is the equivalent of 15.4324 grains.

Haplotype A combination of alleles at multiple loci that are transmitted together on the same chromosome, or a set of single nucleotide polymorphisms (SNPs) on a single chromatid that are statistically associated. Can also refer to an individual collection of short tandem repeat (STR) allele mutations within a specific genetic segment.

Harmless Error A technical violation of some procedural rule that does not substantially affect the outcome of litigation and thus does not require a new trial. See also, Prejudicial Error.

Heat of Passion Term used in criminal law and in behavioral science to distinguish intentional killings from murder to manslaughter when the killing occurs in the heat of passion from lawful provocations on the part of the victim.

Heterogeneous Having a diverse, nonuniform composition. See Homogeneous.

Heterozygote Individual with different alleles at a specific locus on homologous chromosomes, one inherited from each parent.

High-Velocity Impact Spatter	A bloodstain pattern caused by a high-velocity impact/force to a blood source of approximately 30 meters per second or greater. (Examples of such impact are gunshot, high speed machinery, or fast moving objects like an airplane propeller.)
Homogeneous	Having a uniform structure or composition (or components of the same or similar kind). See Heterogeneous.
Homozygote	Individual with two identical alleles at a specific locus on homologous chromosomes, one inherited from each parent.
HPLC	High Performance (or High Pressure) Liquid Chromatography.
Hypothesis; Hypothesis Testing	A hypothesis is the postulating of an educated guess that is then tested and compared to experimentally collected data to determine whether the hypothesis can be either proved correct or false.
IAFIS	Integrated Automated Fingerprint Identification System.
Iatrogenic	A disease, disorder, complication, or effect caused by medical treatment.
Impact Pattern	A bloodstain pattern that is created when a blood source receives an impact that results in the random dispersion of smaller drops of blood.
Infrared Spectroscopy	Use of absorption of infrared radiation to obtain the composition of a substance.
Inotropic	That which affects the force of muscle contraction.
Involuntary Manslaughter	A common law felonious but unintended killing.
Isotope	Different versions of the same chemical element, in which the atoms have the same number of protons (atomic number), but different numbers of neutrons (and therefore a different atomic mass).
LASER	Light Amplification by Stimulated Emission of Radiation.
Ligand	An ion, molecule, or molecular group that binds to another chemical entity to form a larger complex. (Latin ligare, to bind).
Liquid Chromatography-Mass Spectrometry	(LCMS) Analysis using a liquid chromatograph instead of a gas chromatograph.

Lividity A sign of death in which the blood in the body causes a purplish red discoloration of the skin in whatever parts of the body are lowest (but not those in contact with the ground or another surface).

Locus A specific position on a chromosome. Plural loci.

Low-Velocity Impact Spatter A bloodstain pattern caused by a low-velocity impact/force of approximately 1.5 meters per second or less to a blood source.

Lumen The SI unit of luminous flux, a measure of the perceived power of light. Also; the central part of a tube.

Lysis The death of a cell by breaking of the cellular membrane. The solution containing the contents of lysed cells is a "lysate".

Major Profile In DNA analysis where there are more than two alleles at a locus (i.e., a mixture of DNA from at least two people), the ratio of peak areas may indicate a distinctive major contributor (larger peaks) and a distinctive minor contributor (smaller peaks).

Malice A mental state in criminal law and in behavioral science that was required in common law jurisdictions for unlawful killings regarded as murder and for the crime of arson. Malice can be either express or implied. Express malice is evidenced by proof of an intent to kill. Implied malice is evidenced by proof that the accused killed another person, either by intending to do great bodily harm to the victim, or by acting with wanton and willful disregard for the safety of the victim or others.

Mark Term commonly used In the United Kingdom and some Commonwealth countries to designate any item of trace evidence found during a criminal investigation and believed to be associated with the crime that can be studied and compared. If the mark is invisible or nearly invisible until revealed by an enhancement method, in the United States and Canada it is more commonly referred to as a latent impression.

Mass Spectrometry (MS) Analytical method usually employed with gas chromatography (GC/MS) to identify materials based upon production and detection of characteristic molecular fragments.

Medium-Velocity Impact Spatter A bloodstain pattern caused by a medium velocity impact/force of approximately 7.6 meters per second to a blood source.

Microtome Instrument used to cut biological specimens into thin transparent sections suitable for microscopic examination.

Minor Profile In DNA analysis, where there are more than two alleles at a locus (i.e., a mixture of DNA from at least two people), the ratio of peak areas may indicate a distinctive major contributor (larger peaks) and a distinctive minor contributor (smaller peaks).

Misting Blood that has been reduced to a fine spray as a result of applied energy.

Modus Operandi (Latin) The manner of operation or the means of accomplishing an act often revealing characteristics and methods peculiar to certain criminals.

Mole Avogadro's number of constituent entities (approximately 6.02214 1023). The 'constituent entities' may be atoms, ions, atomic particles etc.

Multiplex System of testing multiple DNA loci at the same time; for example, SGM+ tests 10 loci, CODIS tests 15.

Murder Under common law, the killing of a human being by another human being with malice aforethought.

Mutation An inherited or acquired permanent change in the DNA sequence of a gene.

Mydriasis Prolonged or excessive dilation of the pupil due to disease, trauma, or drugs and/or failure of the pupil to constrict when exposed to light.

NIST National Institute of Standards and Technology, a U S Department of Commerce division that sets, approves, and maintains measurements standards in the United States. Formerly called the Bureau of Standards.

Nosocomial Any disease contracted by a patient while under medical care, secondary to the patient's original condition. A category of iatrogenic effects.

Offer or Proof The offering of evidence, the admissibility of which is disputed, to permit a judicial officer to render a decision as to its admissibility. Offers of proof may require the presentation of testimonial evidence, or the presentation of documents or physical evidence for the inspection of the judicial officer.

Parent Drop A drop of blood from which a wave cast off or satellite spatter originates.

Passive Drop A drop that is created or formed solely as a result of gravity.

Pattern A form, shape, or outline that is recognized. If comprised of small spots, there must be an adequate number to be classified as a pattern.

PCR Polymerase Chain Reaction. A technique for making multiple copies of DNA from an initial sample.

Peak Height Imbalance In DNA profiling, a peak height imbalance occurs when two peaks that were expected to be the same are more than 30-40 different in terms of their height.

Perimeter Stain A bloodstain that is comprised only of its peripheral outline. (This stain forms when the central area has been removed by wiping after the blood has partially dried, or by its center flaking off after it has completely dried.)

PGM Phosphoglucomutase. An enzyme that facilitates the interconversion of glucose 1-phosphate and glucose 6-phosphate. Variations can be identified from blood tests and used as genetic markers.

Phenotype The outward appearance of the individual, affected by the genotype and the environment. Different from genotype, which refers to the specific alleles that make up the genetic code for an individual.

Plaintiff The person or entity who initiates a legal action against another.

Point (Area) of Convergence A point or area, on a two dimensional surface, from which a bloodstain pattern originated as determined by tracing the long axis of well-defined bloodstains back to a common point or area of convergence.

Point (Area) of Origin The three-dimensional point or area from which a bloodstain pattern originated as determined by projecting angles of impact of individual, well-defined bloodstains back to a common point or area of convergence.

Polarized Light Microscope (PLM) A microscope that is equipped with two polarizing elements that are placed in the optical path of a microscope.

Polymerase An enzyme that catalyzes the polymerization of deoxyribonucleotides into a DNA strand. In DNA replication, polymerase uses an existing DNA strand as a template, and adds nucleotides to create a new strand.

Precision Reproducibility or repeatability, that is, the degree to which further measurements or calculations will show the same or similar results. Different from accuracy.

Prejudicial Error See Reversible Error.

Presumption In law, a device established either by custom, common law, statute, or rule, which requires that the existence of the presumed fact be taken as established when certain other basic facts are established, admitted, or uncontested. The existence of a legally recognized presumption shifts the burden of disproving the existence of the presumed fact to the party who denies or contests its existence. Most legally recognized presumptions merely create inferences, which, in criminal cases, the fact finder is permitted but not required to entertain. "Conclusive presumptions" are not really presumptions, but are rules of substantive law. Example: Asserting that everyone is presumed to know the law merely precludes, as a matter of law, use of "ignorance of the law" as a defense.

Presumptive A presumptive test identifies when something can be excluded (e.g., the sample is not blood) and identifies what it might be (e.g., the sample might be blood).

Prima Facie Case A party who has the burden of proof (plaintiff) has made a prima facie case when he has presented sufficient evidence to allow the tribunal to reasonably determine the issue in his favor, unless or until the opposing party presents evidence to contradict the plaintiff's evidence.

Primer A strand of nucleic acid that is required for the initiation of DNA synthesis (e.g., in PCR). A primer is chosen to be complementary to the DNA region targeted for amplification under specific thermal cycling conditions.

Projected Blood Pattern A pattern created when blood is projected or released as a result of force.

Pull Up Also termed "bleed-through". Occurs when DNA analysis software fails to discriminate between different dye colors, thus creating false peaks. Can be identified by analysis of the position of peaks across the color spectrum. Can be caused by sample overload.

Pyrolysis The chemical decomposition of matter into new compounds through the action of heat or burning.

Qualitative Analysis Determination of the type of material(s) in a sample.

Quantitative Analysis Determination of the amount of material(s) in a sample.

Quantum of Proof — In satisfying the burden of proof in a legal matter, the sufficiency of the evidence depends on the type of legal matter that is pursued. In criminal cases, proof of guilt must be beyond a reasonable doubt. In civil cases, depending on the jurisdiction, the evidence must be proved by a preponderance of the evidence, by clear and convincing evidence, or proved on the balance of probabilities.

Racemization — A racemic mixture, or racemate, has equal amounts of left- and right-handed enantiomers of a chiral molecule (i.e., mirror-image molecules that can not be superimposed, because of their geometry). Racemization is the transformation of one half of the molecules of an optically active compound into molecules having exactly the opposite configuration (thus the compound is now optically inactive overall).

RADAR — Radio Detection and Ranging. Used extensively in speed detection devices that depend on use of the Doppler Effect.

Refractive Index — The ratio of the speed of light in a vacuum to the speed of light through a transparent medium (such as glass).

Relative Fluorescence Unit — A unit of measurement for electrophoreses, which uses a fluorescence detector.

Relevant Evidence — Evidence that tends to make the existence of a fact of consequence to the determination of a triable issue in court more probable or less probable than it would be without the evidence.

Reversible Error — Sometimes also called Prejudicial Error; the erroneous admission or exclusion of evidence or an erroneous procedural that substantially affects the legal or constitutional rights of a party, entitling that party to a new trial or dismissal of prior proceedings. See also: Harmless Error.

RFLP — Restriction Fragment Length Polymorphism.

Ricochet or Secondary Splash — The deflection of large volumes of blood after impact with a target surface that results in staining of a second surface. (Ricochet does not occur when small drops of blood strike a surface.)

Satellite Spatter — Small droplets of blood that are projected around or beside a drop of blood upon impact with a surface. (A wave cast off is also considered a form of satellite spatter.)

Scallop Pattern — A single drop bloodstain pattern that is characterized by a wavelike, scalloped edge. (When scallops appear only on one side they identify the blood drops directionality prior to impact with the surface.)

Scanning Electron Microscopy (SEM) — A microscope that allows for the viewing of samples at much greater magnification (up to 100 000 times) than is possible with a light microscope.

Scientific Method — While there is no standard definition of the scientific method, it is a process of inquiry to explain observational collected data that are critically compared to known information and systematically analyzed so as to lead to a hypothesis, which is then tested, documented, and refined in a manner that minimizes bias and subjectivity. The result is a conclusion that is accepted as reliable by scientists in the same and other fields. The scientific method is marked by strict adherence to observation, objectivity, rationality, testing, and revising (if revision is indicated).

Sensitivity — Measures the proportion of positive results that are correctly identified as such. In other words, it is a measure of the probability of correctly diagnosing a condition/positive result (cf specificity, which measures the proportion of negatives that are correctly identified, i.e., specificity is a measure of the probability of correctly identifying a nondiseased person/negative result.)

Serum Stain — A clear, yellowish stain, usually having a shiny surface, normally associated with a bloodstain after some red cells have retracted following the clotting mechanism.

SGM+ — Second Generation Multiplex Plus (SGM Plus), a DNA profiling package manufactured by ABI (Applied Biosystems), in which 10 STR loci are analyzed.

SLP/MLP — Single Locus Probe / Multi Locus Probe. SLPs are DNA or RNA sequences that hybridize (i.e., form a DNA-DNA or DNA-RNA duplex) with DNA from a specific restriction fragment. MLPs are DNA or RNA sequences that hybridize with DNA at a number of different sites in the genome of an organism.

Smear — A relatively large volume of blood, at least 0.5 ml, that has been distorted to such a degree that further classification is not possible. Similar to a smudge, but a smear is a stain produced by a larger volume of blood. See also Swipe.

Smudge — A small bloodstain that has been distorted to such a degree that further classification is not possible.

SNP Single Nucleotide Polymorphism. A variation in a sequence of DNA that occurs when a single nucleotide (A, T, C, or G) differs between paired chromosomes in an individual.

Spatter Blood that is dispersed as a result of impact.

SPE Solid Phase Extraction (toxicology/drugs).

Spine The pointed, often elongated, linear streaks that radiate away from the center of a bloodstain.

Spoliation The intentional alteration or destruction of documents or evidence.

Spot A spot of blood is not a pattern. It may exhibit directionality, various degrees of size and spatter, and other recognizable characteristics, but a single spot of blood does not constitute a pattern.

Stenotic Narrowed or constricted, as in a stenotic artery.

Stutter Spurious "shadow" bands produced during PCR amplification of STR loci, which manifest as peaks in an electrophoretogram.

Surface Tension Two fluids in contact exhibit this phenomenon due to molecular attractions that appear to arise from a tension in the surface of separation. It is measured in dynes, the force that produces acceleration of 1 cm s?2 in a mass of 1 gram.

SWGDAM Scientific Working Group on DNA Analysis Methods (United States). Falls under auspices of the Federal Bureau of Investigation (FBI) and proposes and recommends revisions to the National Quality Assurance Standards (QAS) and liaises with the forensic DNA community.

Swipe Pattern A bloodstain pattern that results from a lateral transfer of blood onto a surface. (A feathered edge indicates direction of travel.)

Tachycardia Rapid heartbeat.

Tachypnea Rapid breathing.

Target A surface that has become bloodstained.

Tattooing Punctate abrasions, also termed stippling and gunpowder tattooing, are caused by unburned or partially burned pieces of gunpowder impacting the skin upon a weapon being fired in close proximity to the skin.

Teardrop The typical shape of a bloodstain that is produced when a drop or droplet of blood strikes a surface at an acute angle.

Telomere A region of repetitive DNA (TTAGGG) at the end of chromosomes, which stabilizes the chromosomes and protects the ends from attaching to one another. Many cells lose parts of the telomeres during cell division.

Terminal Velocity The maximum speed to which a free-falling drop of blood can accelerate in air. This is approximately 7.7 meters per second for a typical 0.05-ml drop.

TLC Thin Layer Chromatography.

Transfer Pattern A contact bloodstain created as a result of compression and/or lateral movement of a wet, bloody surface against a second surface.

TWGED Technical Working Group on Education and Training in Forensic Science (USA), which addresses qualifications for a career in forensic science, undergraduate programs, graduate programs, and continuing professional training. TWGED guidelines are voluntary and consensus derived.

UKAS United Kingdom Accreditation Service. Sole national accreditation body recognized by the UK government to assess, against internationally agreed standards, organizations that provide certification, testing, inspection, and calibration services. Accreditation by UKAS is intended to demonstrate the competence, impartiality, and performance capability of these evaluators.

Validation Process of checking that a test/process measures what it purports to measure, and performs effectively and reproducibly.

VASCAR Visual Average Speed Computer. VASCAR is used by some law enforcement agencies to measure the speed of vehicles.

Venous Insufficiency Syndrome Small bloodstains can be produced by ruptured varicose veins. These can resemble medium-velocity impact spatter.

Viscosity The internal friction within a fluid, which is described as the resistance to changing the form of the fluid. It is measured in centipoises, which is the tangential force/unit area required to maintain unit difference of 1 cm s?1 between two parallel plains separated by 1 cm of fluid.

Void or Shadow Pattern	Absence of bloodstain in an otherwise continuous bloodstain pattern.
Wave Cast Off	A small blood droplet that originates from a parent drop of blood due to the wavelike action of the liquid in conjunction with striking a surface at an angle less than 90°.
Wipe Pattern	A bloodstain pattern that is created when an object moves through blood that has not completely dried, thereby altering its appearance.
Y Chromosome	The chromosome that determine sex in mammals, including humans. In humans, it consists of approximately 60 million base pairs (G-C and A-T).

Author Index

Abbondante, Serena F. **2**:500–507; **4**:1883–1887
Abdel-Monem, Tarik **2**:760–764
Abou-Khalil, Bassel **5**:2298–2305
Abram, Karen M. **4**:1877–1883
Adams, Holly A. **1**:51–58
Aitken, Colin G. G. **3**:1579–1586; **5**:2281–2291
Alberink, Ivo **3**:1624–1632
Aleksander, Adam K. **3**:1483–1495
Allen, Rebecca S. **2**:717–720
Almog, Joseph **3**:1211–1216
Anderson, Robert N. **4**:1680–1688
Andrews, Paul **4**:1989–2001, 2173–2186
Anetzberger, Georgia J. **2**:912–916
Angelis, Danilo De **3**:1511–1518; **5**:2611–2617
Aumeer-Donovan, Shaheen **3**:1292–1317

Baden, Michael M. **1**:243–248
Bader, Scott **3**:1458–1459; **4**:2007–2008, 2075–2075
Baker, David W. **1**:141–149
Balding, David **5**:2365–2376
Baldwin, David **3**:1240–1243, 1248–1252
Ballantyne, Jack **5**:2628–2635, 2677–2682
Barnes, Sean **3**:1500–1505
Barni, Filippo **3**:1645–1656
Benbow, M. Eric **2**:934–945
Berkowitz, Shari R. **2**:1075–1079
Bernet, William **2**:529–537; **3**:1335–1342; **4**:1981–1984; **5**:2602–2611
Beyer, Jochen **1**:134–140; **4**:2057–2068
Bicknell, Danna E. **3**:1255–1276
Black, Sue **1**:152–179; **2**:764–772
Block, Stephanie **2**:549–552, 553–556
Bohnert, Michael **3**:1529–1541
Botluk, Diana **5**:2619–2627
Bottoms, Bette L. **2**:549–552
Bowman-Fowler, Nicci **2**:1075–1079
Braun, Michelle **2**:444–450
Bresler, Scott A. **3**:1459–1465
Brick, John **1**:99–108, 108–120, 120–125
Bright, Jo-Anne **4**:1749–1750; **5**:2354–2365
Bryant, Vaughn M. **4**:1954–1968
Buchholz, Bruce A. **1**:418–422, **5**:2231–2233
Buckleton, John S. **2**:831–839; **3**:1566–1575; **4**:1838–1842
Bulling, Denise **2**:760–764
Burgess, Ann W. **5**:2311–2324

Cantu, Antonio A. **3**:1541–1546
Carpenter, Douglas J. **3**:1175–1189
Cattaneo, Cristina **1**:179–187, 188–191, **1**:191–198; **3**:1086–1092, 1495–1499, 1511–1518; **4**:1889–1895, 2089–2093; **5**:2328–2332, 2393–2397, 2557–2565, 2611–2617
Champod, Christophe **2**:968–976, **3**:1277–1282, **3**:1508–1511
Cheng, Wing-Chi **4**:2119–2136
Choi, Hyeyoung **2**:844–851
Choi, Sangkil **2**:844–851
Christensen, Thomas C. **5**:2250–2257
Chung, Heesun **2**:844–851
Clegg, Carl **1**:270–272; **5**:2241–2242
Coble, Michael D. **4**:1804–1810
Connor, Melissa **4**:1674–1679
Cornell, Dewey G. **5**:2454–2460
Costanzo, Mark **3**:1586–1590
Costello, Jan **4**:1984–1989
Court, Denise S. **1**:338–359
Cowell, Anthony M. **5**:2640–2646
Cross, Douglas W. **2**:946–953
Curran, James M. **3**:1351–1360, **5**:2590–2593

Davis, Malcolm **4**:1762–1791
Day, Stephen P. **3**:1451–1458
De Angelis, Danilo **4**:1889–1895
De La Torre, Rafael **5**:2420–2431

den Dunnen, M. **2**:584–588
Dickson, Stuart **2**:431–437
Dietz, Park **5**:2460–2465
Douglas, Kevin S. **2**:667–677; **5**:2272–2279
Drummer, Olaf H. **1**:134–140, 290–293, 293–298; **2**:595–601; **4**:2115–2119
Dutton, Gerard **3**:1204–1211, 1216–1219
Du Preez, Charl **5**:2480–2484
Duvinage, Nicolas **3**:1360–1371

Eangelis, Danilo De **3**:1511–1518
Easteal, Patricia **1**:272–276; **4**:2149–2156
Edelman, Gerda **3**:1624–1632
Edelstein, Barry A. **2**:450–455
Edwards, Carl N. **1**:281–289; **2**:728–740; **3**:1474–1480; **4**:2156–2161; **5**:2271–2271, 2431–2441
Elkington, Kate S. **2**:649–655
Erickson, Steven K. **5**:2576–2580, **5**:2580–2584, **5**:2584–2587
Eriksson, Anders F. **5**:2541–2545
Ervin, Thomas **1**:141–149
Esseiva, Pierre **2**:851–859
Evett, Ian Webber **2**:968–976

Fagan, Jeffrey **3**:1612–1618
Felgate, Peter **5**:2509–2533
Fiddian, Susan **1**:422–429
Finch, Indra A. **2**:456–463
Fineschi, Vittorio **2**:468–482
Finkenbine, Ryan **1**:253–256
Fitterman, Elizabeth **5**:2619–2627
Fitzpatrick, Robert W. **5**:2377–2388
Flanagan, Robert J. **4**:2219–2229
Found, Bryan **3**:1436–1451
Fowler, Nicci B. **4**:1712–1716
Frazier, Leeanne **2**:579–583
Fremouw, William J. **1**:270–272; **3**:1480–1482; **5**:2241–2242, 2418–2419
Friedman, Susan Hatters **2**:556–561; **4**:2136–2141; **5**:2449–2454
Frudakis, Tony **4**:2021–2035
Frumkin, I. Bruce **2**:463–467; **3**:1590–1594

Gallo, Frank J. **4**:2068–2071, 2071–2075
Ganis, Giorgio **2**:724–728
Gardner, Ross M. **2**:625–643
Geller, Jeffrey L. **3**:1225–1229
Geradts, Zeno **2**:584–588, **3**:1520–1527
Geraerts, Elke **5**:2243–2250
Gerostamoulos, Dimitri **1**:134–140; **4**:1895–1903, 2115–2119
Gert, De Boeck **4**:1903–1925
Gibelli, Daniele **4**:2089–2093
Giblin, Mary **5**:2291–2295
Gilder, Jason R. **2**:816–820
Gillman, Victoria C. **2**:500–507, 507–528
Gitlow, Stuart **1**:18–22
Goddard, Ken **5**:2635–2640
Goodman, Gail S. **2**:549–552, 553–556
Goodwin, Kerri A. **2**:1065–1072
Gould, Christine E. **2**:450–455
Graham, Eleanor A. M. **2**:800–816
Greaves, Caroline **1**:36–51
Greenberg, Martin S. **2**:643–648
Greene, Edie **3**:1602–1607

Hackman, Lucina **2**:764–772
Hamilton, Warren D. **2**:860–868
Hammer, Lesley **3**:1244–1248, 1252–1255
Hanson, Erin K. **5**:2628–2635, 2677–2682
Han, Eunyoung **2**:844–851
Harbison, Sally-Ann **2**:821–830, 1060–1065
Hart, Stephen D. **4**:2193–2197, 2197–2201
Hasel, Lisa E. **2**:1072–1075
Hayne, Harlene **2**:656–662
Hazelwood, Robert R. **4**:2161–2173; **5**:2311–2324
Henderson, Carol **2**:1013–1019; **4**:1663–1668
Hicks, Tacha N. **3**:1351–1360
Hill, Cheryl A. **1**:253–256
Honts, Charles R. **2**:720–724
Hopen, Thomas J. **4**:1762–1791
Hunter, John **1**:199–207

Ikegaya, Hiroshi **3**:1342–1348
Isenschmid, Daniel S. **2**:562–569

Jackson, Graham **2**:483–497
Jackson, Michael **3**:1200–1204
Jamieson, Allan **1**:17–17, **5**:2555–2557
Jones, Alan W. **1**:58–81, **1**:81–99
Jones, Graham R. **5**:2495–2503
Jones, Philip J. **2**:483–497
Just, Rebecca S. **4**:1804–1810

Kassin, Saul **2**:588–595
Katsumata, Yoshinao **3**:1342–1348
Katterwe, Horst **5**:2485–2495

Kaye, David H. **3**:1561–1566
Kaye, Neil S. **4**:2149–2156
Keaton, Ralph **1**:1–10
Keereweer, Isaäc **3**:1230–1239
Kehn, Andre **5**:2306–2311
Kenan, Joseph **2**:542–549; **4**:1981–1984, 1984–1989; **5**:2602–2611
Kennedy, Robert **3**:1244–1248
Kernbach-Wighton, Gerhard **4**:2076–2088
Khanmy-Vital, Aita **4**:1793–1804
Kim, Eunmi **2**:844–851
Kintz, Pascal **3**:1427–1432
Kirkbride, K. Paul **4**:1750–1758
Knoll IV, James L. **1**:243–248; **2**:977–985; **4**:2161–2173; **5**:2311–2324, 2597–2602
Koehler, Jonathan J. **5**:2401–2409
Koperski, George J. **2**:500–507; **4**:1883–1887
Kosslyn, Stephen M. **2**:724–728
Krane, Dan E. **1**:126–128, **3**:1639–1645

Lamendola, Gretchen M. **2**:463–467; **3**:1590–1594
Lancaster, Sarah L. **2**:1028–1060
Langenburg, Glenn **3**:1282–1292
Laporte, Gerald M. **3**:1255–1276
Laux, Dale L. **1**:314–319
Lebeau, Marc A. **2**:868–876
Lee, Juseon **2**:844–851
Lee, Li-Wen G. **2**:437–444
Lee, Sooyeun **2**:844–851
Lennard, Chris **3**:1292–1317
Lentini, John J. **1**:207–224; **3**:1103–1112, 1112–1122, 1122–1136, 1136–1137, 1137–1171, 1171–1174
Lenz, Kurt W. **2**:1013–1019; **4**:1663–1668
Leong, Gregory B. **2**:741–745; **3**:1432–1436
Leo, Richard A. **3**:1586–1590
Lewis, Simon W. **3**:1645–1656
Lim, Miae **2**:844–851
Liptai, Laura L. **5**:2250–2257, 2565–2576
Loftus, Elizabeth F. **2**:1075–1079; **4**:1709–1712, 1712–1716
Lovell, Robert W. **4**:2210–2218
Lovelock, Tina J. **3**:1348–1351
Luong, Susan **4**:1668–1674
Lynn, Steven Jay **3**:1500–1505

Maat, G. J. R. **2**:891–897
Macdonell, Herbert L. **1**:359–396
Maceo, Alice V. **3**:1322–1331
Macvaugh III, Gilbert S. **4**:1730–1737
Madea, Burkhard **2**:697–716; **4**:1689–1709; **5**:2466–2479
Maiden, Nicholas R. **3**:1219–1225; **5**:2324–2327
Mallett, Xanthé **1**:152–179
Margot, Pierre **2**:851–859
Marleen, Laloup **4**:1903–1925
Marshall, Maurice **2**:1028–1060
Martell, Daniel A. **3**:1459–1465
Massonnet, Geneviève **4**:1943–1954
Mastruko, Vojin **5**:2257–2267
Matthews, Abigail **3**:1500–1505
Mazzella, W. D. **3**:1557–1561
McCoy, Katrina **5**:2241–2242, 2418–2419
McCullough, John **4**:1931–1943
McDermott, Sean D. **5**:2534–2540
McKenna, Louise **1**:320–321
McNally, Richard J. **5**:2243–2250
Meijerman, L. **2**:891–897
Meloy, J. Reid **5**:2397–2401
Melson, Kenneth E. **1**:1–10
Melton, Terry **4**:1833–1837
Merritt, Richard W. **2**:934–945
Meuwly, Didier **1**:249–253; **5**:2389–2392
Micheals, Anastasia D. **4**:1680–1688
Miles, Samuel I. **2**:728–740, **4**:2201–2210
Moenssens, Andre A. **1**:23–24, 150–151, 298–299; **2**:498–500, 561–562, 662–665, 745–747, 758–760, 775–778, 778–781, 840–841, 955–957, 963–963, 998–1000, 1001–1002, 1012–1013; **3**:1093–1094, 1095–1095, 1276–1277, 1331–1332, 1333–1333, 1465–1468, 1505–1505, 1528–1529, 1599–1600, 1601–1602, 1607–1608, 1623–1624; **4**:2009–2011; **5**:2296–2297, 2589–2590,
Mohammed, Linton A. **1**:128–134
Monnard, Florence **4**:1943–1954
Moretti, Marlene M. **1**:36–51
Morrish, Bronwyn C. **1**:299–313; **2**:500–507
Mueller-Johnson, Katrin **2**:916–920
Murphy, John P. **4**:1968–1980

Nehse, Kornelia **2**:985–998
Nele, Samyn **4**:1903–1925
Nelson, Kally J. **2**:1075–1079; **4**:1712–1716
Nerenberg, Lisa **2**:902–912
Neumann, Cedric **3**:1546–1552

Neuner, John K. **1**:1–10
Neuschatz, Jeffrey S. **2**:1065–1072
Nicholls, Tonia L. **1**:36–51
Nikolova, Natalia L. **5**:2272–2279
Norman, Keith W. **2**:500–507, 507–528
Nunez, Narina L. **5**:2306–2311

Ojanperä, Ilkka **5**:2503–2509
Olley, J. Gregory **4**:1724–1730, 1730–1737
Öström, Mats G. **5**:2541–2545
Oxley, Jimmie C. **2**:1028–1060

Park, Yonghoon **2**:844–851
Payne-James, Jason **1**:234–242
Perry, Sylvia **2**:549–552
Peterson, Tiamoyo **4**:1709–1712
Pichini, Simona **5**:2420–2431
Pinals, Debra A. **3**:1552–1557
Piper, August **2**:728–740
Pollak, Stefan **1**:256–262, 396–411; **3**:1380–1401, 1468–1474; **5**:2646–2660
Pollanen, Michael S. **1**:224–234
Porta, Davide **3**:1086–1092, 1511–1518; **5**:2328–2332, 2393–2397, 2557–2565
Porter, Glenn **4**:2036–2057
Poulsen, Helen **2**:431–437

Quinn, Mary J. **3**:1371–1380

Raes, Elke **2**:877–885
Raff, Adam N. **2**:576–579
Randolph-Quinney, Patrick **1**:152–179
Raymond, Jennifer J. **3**:1318–1322
Ray, Neelanjan **2**:529–537
Reed, Tom **4**:2149–2156
Resnick, Phillip J. **2**:977–985
Resor, Michelle R. **5**:2413–2417
Ressler, Robert K. **5**:2311–2324
Riezzo, Irene **2**:468–482
Robertson, James **1**:422–429; **3**:1095–1103; **3**:1415–1427, **4**:1758–1761; 1791–1793
Roffey, Paul E. **1**:299–313; **2**:500–507
Romero, Erin G. **2**:649–655
Roux, Claude **3**:1095–1103, 1292–1317, 1318–1322; **4**:1668–1674; **5**:2480–2484
Royds, David **1**:411–418
Ruifrok, Arnout C. C. **3**:1081–1086

Salekin, Karen L. **2**:717–720, **4**:1724–1730, 1730–1737, 1818–1823
Salsarola, Dominic **3**:1495–1499
Saukko, Pekka J. **1**:256–262; 396–411, **3**:1342–1348, 1380–1401, 1468–1474; **5**:2646–2660
Sauvagnat, François **4**:2186–2193
Schiffer, Beatrice **1**:10–16; **5**:2545–2555
Schneck, William M. **4**:1743–1749, 2001–2007
Schneider, Richard D. **4**:1717–1724
Scott, Allan Mathieson **2**:602–614, 614–619, 619–625, 963–968; **4**:1927–1931
Scott, Charles L. **3**:1552–1557
Segovia, Daisy A. **2**:553–556
Sharfe, Gordon A. I. **5**:2660–2675
Shefchick, Thomas P. **2**:920–934
Shiver, Farrell C. **3**:1594–1599
Siegel, Jay A. **2**:897–902
Silva, Arturo J. **1**:75–79
Singh, Rajvinder **2**:748–757
Skopp, Gisela **4**:2093–2114
Smarty, Sylvester **1**:263–269; **2**:537–542
Smith, Ann C. **2**:692–696; **3**:1334–1335, 1619–1621, **5**:2627–2628
Smith, Delaney M. **4**:1737–1742; **5**:2409–2413
Smyth, Larry D. **4**:2141–2148
Sorrentino, Renee **5**:2332–2338
Spiegel, David **2**:784–792
Squier, Waney **5**:2339–2350
Stankowski, Joy E. **2**:885–890; **5**:2443–2448
Stauffer, Eric **1**:10–16; **3**:1632–1639; **5**:2545–2555
Steinberg, Laurence **3**:1608–1612
Strub, Diane S. **5**:2272–2279
Studebaker, Christina **3**:1602–1607

Takatori, Takehiko **3**:1342–1348
Taroni, Franco **1**:276–281
Teplin, Linda A. **2**:649–655; **4**:1877–1883
Thakar, Mukesh Kumar **2**:748–757
Thean, A. **2**:891–897
Thomas, Tracy A. **1**:270–272; **3**:1480–1482; **5**:2418–2419
Thompson, Christopher **2**:542–549; **4**:1984–1989
Thompson, William C. **3**:1575–1579
Thurman, James T. **2**:1019–1027
Toglia, Michael P. **2**:1065–1072

Tramontana, Michael G. **4**:1869–1877
Tridico, Silvana R. **3**:1403–1415
Tully, Gillian **4**:1823–1832
Turillazzi, Emanuela **4**:1843–1861
Turner, Barry **2**:1003–1004; 1004–1007, 1007–1011

van Waalwijk van Doorn, K. **2**:584–588
Verstraete, Alain G. **2**:877–885
Viner, Mark D. **5**:2233–2240
Vitacco, Michael J. **4**:1657–1663; **5**:2576–2580, 2580–2584, 2584–2587
Vittorio, Fineschi **4**:1843–1861
Vuori, Erkki **5**:2503–2509

Walker, James S. **4**:1862–1869, **5**:2298–2305
Walsh, Simon J. **1**:140–140, 262–262, **2**:497–498, 569–570, 677–684, 757–757, 792–799, 831–839, 842–843, 954–955; **3**:1318–1322, 1518–1520; **4**:1688–1689, 2149–2149, **4**:2219–2219; **5**:2595–2596
Washburn, Jason J. **4**:1877–1883
Weinstock, Robert **2**:741–745; **3**:1432–1436
Weir, Bruce S. **4**:1810–1818
Wells, Gary L. **2**:1072–1075
Wenger, Eric **2**:500–507
West, Sara G. **2**:556–561, **5**:2449–2454
Wetton, Jon **4**:1823–1832
Weyermann, Céline **2**:684–692
Wheate, Rhonda M. **2**:772–774, 774–775, 781–784; **5**:2268–2270
Whelpton, Robin **4**:2219–2229
Wilson, Catherine M. **2**:667–677
Wong, Steven H. Y. **4**:2012–2021

Yang, Suzanne **1**:24–36; **2**:571–575; **3**:1225–1229; **4**:1737–1742, **4**:2186–2193
York, Catherine **2**:549–552

Zajac, Rachel **2**:656–662
Zeichner, Arie **3**:1189–1200, **5**:2351–2354
Zoun, Rikkert **1**:321–338

Subject Index

1 : 1 taping method **2**:986, 993, 994
2-D–2-D identification techniques **5**:2614
2p rule (low template DNA) **3**:1570
3-D–2-D identification techniques **5**:2614, 2615
3D modeling **3**:1628
3D reconstructions **5**:2257–67

A4 paper **4**:1972, 1974
AA (angular apertures) **4**:1767
AAFS *see* American Academy of Forensic Sciences
AAIDD (American Association on Intellectual and Developmental Disabilities) **4**:1724–5
AAMR (American Association on Mental Retardation) **4**:1732
abandonment of elders **2**:904
abbreviated disclosure **2**:452
ABC *see* American Board of Criminalistics
abdominal injuries **1**:410–1
abduction of elders **2**:904
aberrations, objective lenses **4**:1768
ABFM (American Board of Forensic Medicine) **2**:1015
ABI *see* Applied Biosystems
ability to act **3**:1385; **5**:2654–5, 2658
ABO blood group system **1**:339, 341–9
abrasion collar/ring/margin/rim **3**:1386–7
abrasion traces, paint **4**:1950
abrasions (skin) **1**:396–7
absorbent substrates **3**:1650
absorption of drugs **1**:60; **2**:565; **4**:1900, 2115
  see also administration routes (drugs)
absorption-elution technique **1**:342, 343, 347
absorption-inhibition technique **1**:343
absorption mode
  fingermarks in blood **3**:1310
  optical enhancement lighting **4**:2042–6, 2047
abuse
  allegations **2**:531, 535, 539; **5**:2607
  denials **2**:534–5
  elders **2**:902–12, 912–5; **3**:1373
  radiological diagnosis **5**:2236
  see also alcohol abuse; child abuse; childhood sexual abuse; domestic abuse; drug abuse; physical abuse; sexual abuse; substance abuse
'abuse of discretion' **3**:1334
'abuse excuse' (battered woman syndrome) **1**:270
'abuse therapy' business **2**:734–5
accelerants (fire) **3**:1133; **4**:1929–30; **5**:2537
accelerated fire myth **1**:210, 213, 216, 219–20
  see also arson
acceleration of head **5**:2572–4
accelerator mass spectrometry (AMS) **1**:419–20; **5**:2231–2
accelerometers (air bag systems) **1**:52–3, 55, 57
acceptor, iron traces **3**:1215
accidental characteristics, shoemarks **3**:1232–3
  see also occasional features
accidents
  analysis (product use) **3**:1489
  deaths **5**:2235, 2541, 2646, 2653
  explosions **2**:1024, 1030; **3**:1136
  fires **1**:210; **3**:1123, 1136
  injuries **1**:100–4, 266; **5**:2237
  reconstructions **2**:580, 581–2; **5**:2250–7, 2483
  see also road traffic accidents
accommodation to abuse (child sexual abuse) **2**:537–41
accreditation **1**:11–2
  advantages/disadvantages **1**:15
  crime laboratories **1**:1–8
  definition **1**:11; **3**:1173
  fire investigation **3**:1171–4
  forensic science education programs **2**:898, 899–901
  laboratory accreditation **1**:1–8, 10–6; **4**:2220–1
  organisational **1**:10–6
  process **1**:12–3
accumulated degree-days (ADD) **2**:942
accumulated degree-hours (ADH) **2**:942
accuracy
  confidence relationship **2**:1074
  cross-examination **2**:657–8, 660
  eyewitness identification **2**:1074
  traumatic memory recollection **5**:2243–8
'accusatorial' model **2**:782
ACE-V *see* analysis, comparison, evaluation and verification
acetaldehyde **1**:114
acetaminophen **1**:110
acetylcodeine **4**:1917
6-acetylmorphine (6-AM) **2**:865
acetylsalicylic acid (ASA) **1**:115
ACFE (American College of Forensic Examiners) **2**:1014–5
acid back-extraction **4**:2131
acid digestion extraction method **2**:753, 754
acid etching serial number restoration **5**:2326–7
acid phosphatase (AP) **1**:17, 315; **2**:822–3, 954
acid phosphatase 1 (ACP1) **1**:351–2
acid sized papers **4**:1974
acidic drug sample extraction **5**:2517
acidification of specimens **4**:2104

acidosis, lactic **4**:2080
aconite **4**:2058, 2059–60
Aconitum napellus **4**:2059–60
acoustic–phonetic speech analysis **5**:2390
ACP1 (acid phosphatase 1) **1**:351–2
acquiescence (suggestibility) **3**:1593
acquiescence response bias **2**:591
acquired immune deficiency syndrome (AIDS) **4**:1880, 1881
  see also human immunodeficiency virus
acquired standards, inks **3**:1542
acquisition stage memory **2**:1076; **4**:1709–10
acrylamide gels **1**:349
ACS *see* activated carbon strips; acute chest syndrome; acute coronary syndrome
activated carbon strips (ACS) **3**:1139–40, 1142
active defense wounds **5**:2658
activities of daily living (ADLs) **2**:445
activity issues (fingerprint interpretation) **3**:1279
activity-level evidence, CAI **2**:493–4
activity-level hypothesis, paint evidence **4**:1951, 1952
activity-level propositions **2**:973; **3**:1356–7, 1580
acts of commission **2**:955
acts of omission **2**:955
actuarial predictions **2**:672, 673; **5**:2274, 2275–6
actuarial risk assessment instruments (ARIs) **5**:2598, 2599
actus reus **1**:254; **3**:1339
acute chest syndrome (ACS) **4**:1849–50
acute coronary syndrome (ACS) **2**:475
acute ethanol intoxication **1**:227
acute factors, risk assessments **5**:2599
acute hepatic insufficiency **4**:2081
acute homosexual panic *see* homosexual panic
acute interstitial pneumonitis (AIP) **4**:1845–6
acute myocardial infarction (AMI) **2**:469
acute renal failure **4**:2082
acute respiratory distress syndrome (ARDS) **4**:1845
acute stress disorder (ASD) **4**:2141
acute tolerance **1**:75, 123
acutely psychotic filicide **4**:2139
ADA (adenosine deaminase) **1**:349–50
adaptational model of malingering **4**:1658
adaptive cruise control **1**:57
adaptive deficits (malingering) **4**:1734–5
adaptive functioning **4**:1726–7, 1733
ADD (accumulated degree-days) **2**:942
addictions **1**:18–21
  behavior **2**:571, 572–3
  definition **1**:18, 123, 287
  disease course **1**:19
  epidemiology **1**:19–20
  incidence **1**:20
  prevalence **1**:19
  relapse prevention **1**:20–1
  treatment **1**:20
  see also alcohol abuse; dependence (drugs); drug abuse; substance abuse
additive alterations, documents **1**:129–31
additive color synthesis model **4**:2043–5
additive effects (alcohol–drug interactions) **1**:109
additive tape lifting **2**:986
additives (paint) **4**:1932
adenosine deaminase (ADA) **1**:349–50
adenosine triphosphate (ATP) **2**:705
adenylate kinase (AK) **1**:350–1
ADH *see* accumulated degree-hours; alcohol dehydrogenase
ADHD *see* attention deficit hyperactivity disorder
adhesive lifters **3**:1319
adhesive surface fingermark detection **3**:1305–8
adipocere formation, mass graves **4**:1674, 1675
adjudicative competence/culpability **2**:543–6
ADLs (activities of daily living) **2**:445
ADM (alcohol, drug and mental) disorders **4**:1877–82
administration guidelines, Hare Psychopathy Checklists **4**:2198
administration routes (drugs) **4**:2116
  amphetamines **1**:137
  cannabis **2**:435
  cocaine **2**:565–6
  opioids **4**:1902
  see also absorption of drugs; intravenous drug use
administrator influence, lineups **2**:1074
admissibility
  computer animation evidence **2**:581, 582
  expert opinion evidence **2**:500, 692–6, 760, 1001–2, 1003–4, 1004–7, 1007–11; **3**:1331–2, 1333, 1334–5, 1528, 1619–21; **5**:2627–8
  hair comparison evidence **3**:1424
  hypnosis testimony **3**:1502
  neuropsychological evidence **4**:1862–3
  simulation evidence **2**:581, 582
  truth serum evidence **2**:735–6
  see also foundation evidence; laying the foundation
admixture
  AIMs **4**:2025–8
  definition **4**:2023
  mapping process **4**:2032
  multidimensional continuum **4**:2024–5
  see also mixtures (DNA)
Adobe Photoshop *see* Photoshop
adolescent limited (AL) conduct disorder **1**:45
adolescents
  abstract thinking **2**:544
  age determination of remains **1**:179
  aggression **1**:45
  anti-social behavior **2**:544, 547–8
  arrest process **2**:546–7
  asphyxial games **1**:246
  culpability **2**:543–6, 547
  custody/visitation evaluation **5**:2605
  death penalty abolition **2**:717–8
  development **2**:543–4; **3**:1608–12
  developmental immaturity **2**:543–4
  'evil' identification **2**:982
  false confession vulnerabilities **2**:591–2
  identity **2**:544
  intelligence **2**:546
  interrogation **2**:546–7
  mental disorders **2**:544–6, 547
  perceived autonomy **2**:544
  right to treatment **5**:2581, 2582–3

sadism **5**:2320
serial murder **5**:2314
sexual homicide **5**:2314
sexually motivated assault **1**:238–9
substance abuse **5**:2414
substance use disorders **2**:546
transfer to adult justice **3**:1612, 1613, 1615
see also children; juvenile. . .; pediatric. . .; subadult age category
adrenaline **1**:134
adult criminal courts, juveniles in **2**:547–8
adult protective services (APS) **2**:908
adulteration
drug manufacture **4**:2095
urine test obfuscation **2**:863–5
adults
age determination
living individuals **1**:189
remains **1**:180–5
cross-examination **2**:660
guardianships **3**:1371–80
offenders' rights **5**:2583
advance directives **2**:439; **5**:2452
adventitious DNA matches **2**:832–4, 837, 838
adversarial justice system **1**:23–4; **2**:1004–6, **5**:2297
CAI application **2**:495
cross-examination **2**:663, 664
Daubert case impact **2**:695
demonstrative evidence **2**:745
discovery **2**:775
expert evidence admissibility **2**:1007, 1008
expert witness definition **2**:1012
jury instructions **3**:1607
learned treatise evidence **3**:1623–4
see also common law systems
adverse events (AEs) **4**:1690
definitions **4**:1690
epidemiology **4**:1691, 1692
hospitalized patients **4**:1690, 1692, 1703
malpractice **4**:1703, 1704–5
medication errors **4**:1704
advisory role of CSI **2**:616
advocacy logic **5**:2439
advocates **2**:663; **5**:2269
AEA (autoerotic asphyxiation) **1**:244–7
AEDs (antiepileptic drugs) **5**:2303–4
AEME (anhydroecgonine methylester) **4**:1913
aerial photography searches **3**:1497
AEs *see* adverse events
AES (Auger electron spectroscopy) **3**:1195; **4**:1800
AFE (amniotic fluid embolism) **4**:1847–8
affect, neuropsychology **3**:1463
affective violence **5**:2398
affirmative defense **1**:254
AFIS (Automated Fingerprint Identification Systems) **1**:249–53
AFM (atomic force microscopy) **3**:1598
African
admixture **4**:2023, 2024–5, 2026–8
ancestry **1**:163
DNA Database **2**:679
African-Americans **5**:2306
AFTE (Association of Firearm and Toolmark Examiners) **3**:1218
age determination
documents **2**:686–8
human remains **1**:156–61, 179–85; **4**:1890; **5**:2237, 2557–60
inks **3**:1542, 1543–4, 1544–5
living individuals **1**:188–90; **4**:1890–1
skeletal lesions **5**:2557–60
age factors
capacity to waive Miranda rights **2**:466
cognitive ability changes **2**:450, 453
death penalty for prisoners **2**:718–9
eyewitness memory **2**:1068
interrogative suggestibility **3**:1592
memory changes **2**:916–7
mental retardation onset **4**:1734
psychopathy **4**:2194
sentencing defendants **5**:2307–8
see also children; older adults
agency guardianships **3**:1372
agent role, power of attorney **3**:1374
agglutination **1**:339
aggravating circumstances, definition **4**:1818
aggression **1**:24–32
assessment **1**:30
behavior **1**:26–7; **2**:572
biological studies **1**:29
causes **1**:28–30
consequences **1**:27
definition **1**:24
impulsive behavior **2**:572
male aggression rates **1**:38
prevalence data **1**:28
prevention **1**:31
seizure disorder **5**:2304–5
sex differences/similarities **1**:38–42
sex-specific variables **1**:42–4
substance abuse **5**:2415
treatment **1**:30
types **1**:36
see also female aggression; stalking
aggressor identification syndrome **5**:2411
aging *see* age determination; age factors
agony/agonal period **2**:697; **4**:2098
agreement, definition problems **2**:892
agriculture **1**:421
see also plants
Agurs v. United States 427 U.S. 97 (1976) **2**:779, 781
ahaptoglobinemia **1**:355
AHG test (antihuman globulin test) **1**:346
Aide memoir document **2**:613–4
AIDS *see* acquired immune deficiency syndrome
AIMs *see* ancestry informative markers
AIP (acute interstitial pneumonitis) **4**:1845–6
air bag systems **1**:51–7; **5**:2543, 2572
air-drying method, swabs **1**:321
air filters, vehicles **4**:1965
aircraft operation **1**:103
AK (adenylate kinase) **1**:350–1
AKA (alcoholic ketoacidosis) **4**:1852–3
akathisia **4**:2204
AL (adolescent limited) conduct disorder **1**:45

Alaska Psychiatric Institute case **5**:2585–6
albino phenotype **4**:2030, 2031
albumin **1**:154; **5**:2396
alcohol **1**:120–4
  absorption **1**:60
  addiction to **1**:18
  analytical methods **1**:81–99; **5**:2505
  behavioral toxicology **1**:290, 291, 292, 293
  biochemical markers **1**:72–3
  dependence **1**:21, 122–4
  distribution **1**:60–1
  drug–drug interactions **1**:108–17; **4**:2095
  drug-facilitated crime **2**:868, 869
  excretion **1**:62–3
  interpretation of BAC **1**:58–80
  intoxication **1**:99–105; **5**:2415
  lethal dose **1**:77–8
  metabolism **1**:62–3
  multiple drug use **4**:2119
  pharmacogenomics **4**:2017
  postmortem toxicology **4**:2095, 2100–1, 2122, 2127–8; **5**:2498, 2499
  punishable concentration limits **1**:82–3
  sexually motivated assault **1**:236, 239
  temporary insanity **5**:2446
  tolerance to **1**:74–6
  total amount ingested **1**:67
  toxicity **1**:76–7
  truth serum use **2**:728, 729, 730
  use (definition) **1**:122–3
  volume percentages **1**:59–60
  see also alcoholism; ethanol
alcohol abuse **1**:122–3; **3**:1430
  see also alcoholism; drug abuse; substance abuse
alcohol consumption *see* alcohol
alcohol dehydrogenase (ADH) **1**:62, 63, 84, 121
alcohol–drug interactions **1**:108–17; **4**:2095
alcohol, drug and mental (ADM) disorders **4**:1877–82
alcohol impaired driving **2**:877, 878, 879, 882; **5**:2542–3
  see also drug impaired driving
  Alcohol, Tobacco, Firearms and Explosives (ATF), US Bureau of **3**:1132
  alcohol use *see* alcohol
  alcoholic ketoacidosis (AKA) **4**:1852–3
  alcoholism **1**:19, 21; **4**:2017, 2080
  see also alcohol abuse
aldehyde dehydrogenase (ALDH) **1**:62, 63
ALDH *see* aldehyde dehydrogenase
ALFPs *see* amplified fragment length polymorphisms
algor mortis **2**:708–14; **5**:2467–72
  see also postmortem changes
alienation, parental **4**:1981–4
aligned parent (parental alienation) **4**:1981, 1982
aliphatics **3**:1149
alkaline sized papers **4**:1974
alkalizing agents, urine tests **2**:863–4
alkaloids **2**:854
alkanes **3**:1148, 1157, 1162
alleged father hypotheses **4**:1814–5
alleles **1**:339, 340
  identical by descent **4**:1810, 1811, 1812
allelic attribution **1**:126–8
allelic designation (STRs) **5**:2356
allelic drop-in **3**:1568, 1640, 1643
allelic drop-out **3**:1567, 1640, 1643
  allelic frequency
  evolutionary relatedness **4**:1816
  Hardy–Weinberg equilibrium **3**:1458
  parentage/missing persons calculations **4**:1811–3
population-specific estimates **2**:497
allelic ladders **2**:807
Allewalt case **5**:2436, 2437
alligatoring effect (fire myths) **1**:209, 210–1
alloantibodies **1**:339, 340
alloimmunization **1**:340
Allport, Gordon **4**:2179
Alltrix Healthcare **4**:1916
alone elders, measures for **3**:1375
α-amylase **2**:954
alterations to documents **1**:128–9, 129–31
alternate light sources **1**:314, 315, 317; **3**:1318
alternative specimens
  BAC interpretation **1**:71–2
  hair **3**:1427–32
  oral fluid **4**:1903–20
  sweat **5**:2420–9
altruistic filicide **4**:2139
aluminum **2**:952; **5**:2326–7
aluminum alloys **5**:2326–7
aluminum flake 'effect' paints **4**:1935
Alzheimer's disease **2**:718–9, 948–9
  see also dementia
6-AM *see* 6-acetylmorphine
AM *see* antemortem...
Amanita phalloides **4**:2060
amanitin **4**:2060, 2066
ambient light **3**:1222
ambient temperature
  algor mortis **2**:708–14
  larval development **2**:937, 941
  nomogram method **5**:2470–1
amelogenin **5**:2354–5
America *see* South America; United States
American Academy of Forensic Sciences (AAFS) **5**:2622
  education program accreditation **2**:899, 900–1
  engineering section **2**:922
  ethics **2**:958
American Association on Intellectual and Developmental Disabilities (AAIDD) **4**:1724–5
American Association on Mental Retardation (AAMR) **4**:1732
American Board of Criminalistics (ABC) certification **3**:1174
  code of ethics **2**:959
  quality triangle **3**:1172
  American Board of Forensic Medicine (ABFM) **2**:1015
American College of Forensic Examiners (ACFE) **2**:1014–5
American Law Institute (ALI) **3**:1553; **5**:2444
American Psychiatric Association (APA) **4**:1725
American Psychological Association (APA) **2**:959
American Society of Crime Laboratory Directors (ASCLD) **3**:1173

American Society of Crime Laboratory Directors/Laboratory Accreditation Board (ASCLD/LAB) **1**:2–3
International program **1**:2, 3–8
Legacy program **1**:2, 12, 15, 16
American Society of Metals (ASM) **4**:1688
American Society for Testing and Materials (ASTM) **3**:1138–9, 1172; **4**:2004
'Amerithrax' attacks **2**:501
Ames, D. **3**:1594
AMI (acute myocardial infarction) **2**:469
amido black **3**:1320
amines **3**:1653
amino acids **1**:184; **3**:1296
ammonium microchemical tests **4**:1746
ammonium nitrate (AN) **2**:1037, 1038, 1046
ammonium perchlorate (AP) **2**:1046
ammunition **3**:1380–6
chemistry **3**:1189–90
identification **3**:1204–11
packaging **4**:1929
shooting scene investigation **3**:1219–20, 1223
shotguns **3**:1382
see also bullets; projectiles
amnesia **2**:784, 785–8; **4**:1716; **5**:2244–6
amniotic fluid embolism (AFE) **4**:1847–8
amorphous medulla **3**:1404
AMP FLPs *see* amplified fragment length polymorphisms
AmpF/STR® systems **2**:570, 802; **3**:1518
amphetamine-type stimulants (ATS) **2**:851–2, 852, 855, 857
amphetamines **1**:134–9; **4**:2207
abuse of **1**:136
acid back-extraction **4**:2131
ADHD treatment **4**:2214
alcohol interactions **1**:116–7
behavioral toxicology **1**:291, 292
driver impairment **2**:877–8
drug-facilitated crime **2**:872
illicit drug analysis **2**:847
oral fluid **4**:1911
pharmacology/disposition **1**:136–8
repeated use **4**:2118
types **1**:134–5
urine testing **2**:862
see also phenethylamines
ampholytes **1**:339
amplification of DNA **5**:2630–1
see also polymerase chain reaction
amplified fragment length polymorphisms (AMP FLPs) **1**:140; **4**:1749
amputations **1**:194
AMS *see* accelerator mass spectrometry
amylase **1**:317; **2**:954
amytal interviews **2**:729
AN *see* ammonium nitrate
anabolic steroids **1**:291–2
anagen hairs **2**:825–6; **3**:1417, 1420
anagraphical age *see* chronological age
anal fissures, definition **1**:239
anal injuries **1**:239; **2**:534
anal intercourse **1**:239
analgesic effects of drugs **2**:872; **4**:1896, 1899
analysis, comparison, evaluation and verification (ACE-V) **3**:1277–8, 1280, 1285–7, 1289; **4**:2011
analyte changes after death **4**:2098–101
analytical error **2**:955
analytical methods **4**:2220
alcohol **1**:81–99
amphetamines **1**:138–9
artifacts **4**:2106–7
automotive trauma causation **5**:2565–75
benzodiazepines **1**:296–7
breath-alcohol analysis **1**:94–7
cannabis **2**:434
chemical warfare agents **2**:508–23
cocaine use **2**:567–8
DFC investigations **2**:874–6
diatoms **2**:753
driver impairment evidence **2**:881–2, 883
drug profiling **2**:852
ethanol determination **1**:83, 88
explosion debris **2**:1028–60
failure of materials **4**:1680–8
fiber analysis **2**:988–92
fire debris **3**:1137–70
glass evidence **3**:1349–50
network analysis **1**:141–8
opioids **4**:1902–3
oral fluid **4**:1908
paint examination **4**:1936–9, 1942
poisonous plants/fungi **4**:2066–7
postmortem toxicology **4**:2119–34
printing devices **5**:2660–1
soil analysis **5**:2377–87
speed of vehicles **5**:2251–4
sweat testing **5**:2426–7
toxicology **2**:595–600
wildlife forensics **5**:2638–9
writing instruments **5**:2660–1
see also individual methods; methodologies
analytical profiles, inks **3**:1543
anaphylaxis **4**:1854–5
anatomical dissections **1**:256, 259
anatomical pathology technology (APT) **2**:769–70
anatomically detailed dolls **2**:554
anatomy
DNA **2**:792
friction ridge skin **3**:1322–6
ancestry
aging the living **1**:188
human remains **1**:163–6, 191–4; **4**:1890; **5**:2238
see also coancestry coefficient; race
ancestry informative markers (AIMs) **4**:2025–8, 2033
ancient cultures **3**:1282
'ancient document' rule **2**:840
ancient human migration **3**:1342
ancillary autopsy investigations **1**:260
anemophilous plants **4**:1955, 1956
anesthetics **2**:563; **4**:1896
Angel's Trumpet plant **4**:2062
anger problems **5**:2273
angiosperms **4**:1976
angle of impact
air bag systems **1**:53
bloodstain patterns **1**:366–8

angle of minutiae (fingerprints) **3**:1283
angle of V-pattern (fire myths) **1**:209, 220–1
angled collision speed analysis **5**:2251–2
Anglo–Saxon jurisprudence *see* adversarial justice system
angular apertures (AA) **4**:1767
anhedral crystal faces **4**:2004
anhydroecgonine methylester (AEME) **4**:1913
animal-pollinated plants **4**:1956
animals
  bones **1**:153–4
  cruelty to **5**:2318
  DNA profiling **5**:2363
  evidence related to **5**:2635–9
  hair **3**:1403–15, 1418, 1419, 1420; **4**:1965
  remains **5**:2393–4, 2394–5
animations **2**:579–82; **5**:2259, 2264–6
  see also 3D reconstructions
anisotropic particles **4**:1770–1, 1773–5, 1779
anisotropic properties of paper **4**:1971–2
ANN *see* artificial neural network
Annual Accreditation Audit Report, ASCLD/LAB program **1**:5
annual rings, trees **5**:2644, 2645
anodes **4**:1795
anodic reaction, materials **4**:1685
anogen hairs *see* anagen hairs
anogenital injuries **1**:238
  see also anal injuries
anomalies in human remains **1**:194–6
anomalous coronary artery **2**:479
anonymizer services **1**:143
another match (P) error **5**:2404
antagonism **1**:110
antemortem (AM) factors
  disaster victim information **2**:766–7, 770
  odontological identification data **4**:1891
  positive identification material **3**:1512, 1515
  postmortem artifacts **4**:2094–6
  skeletal remains **5**:2557–9
  toxicology specimens **5**:2501–2
  trauma **5**:2557–9
antequem non temporal limit **4**:2091
anterior neck compression **1**:228–32
anthrax attacks **1**:300, 301–2, 311; **2**:501
anthropogenic soils **5**:2379, 2383
anthropology **1**:152–79
  disaster victim identification **2**:770
  firesetting **3**:1225
  living individuals
    aging **1**:188
    identification **5**:2611–5
  positive identification of remains **3**:1512–6
  search scenarios **3**:1495
  skeletal remains trauma analysis **5**:2557–63
  web resources **5**:2623–4
anthropometric approach, sex determination **5**:2330
antiandrogens **5**:2334
antibiotics **1**:110
anticlimb paint **4**:1933
anticoagulant medications **1**:111
anticontamination measures **4**:1927
  see also contamination
anticonvulsants **1**:111
antidepressants **1**:111–2; **4**:2205–6, 2215
antidiabetic medications **1**:112
antiepileptic drugs (AEDs) **5**:2303–4
antigens **1**:339, 340, 341, 342
antiglobulin test **1**:340
antihistamines (H_1-antagonists) **1**:112; **2**:872–3
antihuman globulin (AHG) test **1**:346
antipsychotic medication **1**:112; **4**:2017, 2203–5, 2215–6
antisera **1**:347
antisocial behavior **2**:544, 547–8; **3**:1336; **4**:1658
antisocial personality disorder (APD) **4**:1881, 2145
anvil marks, cartridge cases **3**:1211
anxiety **2**:531; **3**:1463
anxiolytics **4**:2206
aortic dissection **2**:475
aortic ruptures **1**:409
AP *see* acid phosphatase; ammonium perchlorate
APA *see* American Psychiatric Association; American Psychological Association
APCI (atmospheric pressure chemical ionization) **5**:2529–30
APD *see* antisocial personality disorder
apertures (microscopy) **4**:1767
  see also diaphragms
aphasia **4**:1866
apochromatic lenses **4**:1758
apocrine glands **5**:2420, 2422
apparent sheet density (paper) **4**:1973
apparent thickness (paper) **4**:1968, 1973
appeals, expert testimony in **2**:998–1000
Appelbaum, P.S. **2**:453, 454
appellate proceedings **2**:998–1000
Apple computers **2**:585–6, 587
application documents, laboratory accreditation **1**:11, 14
application methods, paint **4**:1932
Applied Biosystems (ABI)
  AmpFISTR® SGM Plus™ system **2**:802
  low copy number DNA test kits **3**:1640, 1643
applied knowledge, fire investigation **3**:1176
APS (adult protective services) **2**:908
APT (anatomical pathology technology) **2**:769–70
aquaeroticum **1**:246
aqueous solutions (explosion debris) **2**:1037
arbitration committees **4**:1692–9, 1705
arc-discharge light **3**:1634
archaeology **1**:199–207; **3**:1495–9; **4**:2090
ARDS (acute respiratory distress syndrome) **4**:1845
area of expertise **5**:2269
  see also expert...
Arendt, H. **1**:31
ARIs *see* actuarial risk assessment instruments
Arizona State DNA database **2**:833
arm avulsion case **3**:1492
Armed Forces DNA Identification Laboratory **4**:1833
aromatics **3**:1111, 1148, 1155, 1161
arrests
  competence/culpability **2**:546–7
  severe mental illness **2**:650–1
arrhythmogenic right ventricular disease (ARVD) **2**:477–8
arson **1**:207–23; **3**:1169, 1225–8

art objects **5**:2232
arterial bloodstain patterns **1**:378
arthropods **2**:934–45
artifacts
  disaster victim identification **2**:770
  DNA matrix **4**:1689
  DNA profiling **1**:127
  handwriting **3**:1438
  pixel defects **3**:1523
  postcollection **4**:2118
  postmortem toxicology **4**:2093–108, 2118
  signatures **3**:1444
artificial neural network (ANN) method **2**:856, 858
ARVD *see* arrhythmogenic right ventricular disease
ASA (acetylsalicylic acid) **1**:115
ASCLD *see* American Society of Crime Laboratory Directors
ASCLD/LAB *see* American Society of Crime Laboratory Directors/Laboratory Accreditation Board
ASD (acute stress disorder) **4**:2141
ASM (American Society of Metals) **4**:1688
aspermic semen **1**:316–7
asphalt **3**:1111, 1158, 1160
asphyxia **1**:78, 224–33; **5**:2345, 2562–3
asphyxial games **1**:246
asphyxiant gases **1**:232
asphyxiophilia **1**:243
aspirations, ethics **2**:959
aspirin **1**:115
assays **4**:2134
  see also genetic assays; immunoassays
assessments
  ASCLD/LAB program **1**:5
  crime scenes **2**:615
  elder abuse cases **2**:909
  environmental science **2**:946
  evidential strength (mtDNA) **4**:1824–7
  facial comparison **3**:1085
  genotype probabilities (LTDNA) **3**:1572
  independent living capacity **2**:444–9
  maturity of young offenders **3**:1611
  medical decision-making capacity **2**:452–4
  neuropsychological **4**:1862–9
  replicate probabilities (LTDNA) **3**:1572–3
  scene footwear impressions **3**:1250
  sexually motivated assault **1**:234–5, 236–7
  see also case assessment and interpretation; evaluation; risk assessment; threat assessment
assignment of contributors, DNA mixtures **4**:1839
assisted employment **4**:1729
Association of Firearm and Toolmark Examiners (AFTE) **3**:1218
Association for the Treatment of Sexual Abusers (ATSA) **5**:2336
assortment process, DNA **2**:796
ASTM *see* American Society for Testing and Materials
asylum model, mental health laws **5**:2581
ATF *see* Alcohol, Tobacco, Firearms and Explosives, US Bureau of
Atkins v. Virginia, 536 U.S. 304 (2002) **4**:1730, 1735
atmospheric pressure chemical ionization (APCI) **5**:2529–30
atomic absorption spectrophotometry (AAS) **3**:1192
atomic force microscopy (AFM) **3**:1598
atomoxetine **4**:2214
ATP (adenosine triphosphate) **2**:705
ATR *see* attenuated total reflectance spectroscopy
Atropa belladonna **4**:2060–1
atropine **4**:2061
ATS *see* amphetamine-type stimulants
ATSA (Association for the Treatment of Sexual Abusers) **5**:2336
attachment pathology **5**:2398
attention deficit hyperactivity disorder (ADHD) **2**:547; **4**:1866, 1873, 2214–5
attention, neuropsychology **3**:1462; **4**:1866
attenuated total reflectance (ATR) spectroscopy **2**:1037; **4**:1756
attenuated total reflectance-Fourier transform infrared (ATR-FTIR) **3**:1598
attorney custody/visitation evaluation **5**:2606
attribution errors, memory **4**:1711
atypical antipsychotic drugs **4**:2204, 2205
atypical autoerotic asphyxiation (AEA) **1**:245–6
atypical hanging **1**:231
audio recordings **5**:2389–91
  see also recordings
auditory hallucinations **2**:458; **3**:1432, 1433, 1434, 1435
auditory–perceptual speech analysis **5**:2390
audits
  environmental **2**:951
  laboratory accreditation **4**:2221
Auger electron spectroscopy (AES) **3**:1195; **4**:1800
Auger electrons **4**:1798, 1800
auricular surface method, age determination **1**:182
Australia
  battered woman syndrome **1**:272–4, 275
  doctor-patient confidentiality **2**:888–9
  expert evidence admissibility **2**:1003–4
  laboratory accreditation **1**:12
  National DNA Database **2**:678–9
  soil database **5**:2383
  spousal homicides **1**:275
  Standards Australia **4**:1916
Australia Group alliance **1**:301
Austrian decree (1855) **1**:257
authentic standard, mass spectrometry **2**:596
authentication of evidence **2**:840–1
  see also laying the foundation
'authoritative' texts **2**:664
autoerotic asphyxiation (AEA) **1**:244–7
autoerotic deaths **1**:231, 243–8; **4**:2170
autoerotic hanging **1**:231
autofluorescence **4**:2047
autolysis *see* degradation
Automated Fingerprint Identification Systems (AFIS) **1**:249–53
automatic approaches, speech analysis **5**:2390–1, 2391
automatic matching, earprints **2**:892–3
automation developments, AFIS system **1**:252
automatism defense **1**:253–6

automobile accident reconstruction *see* road traffic accidents; vehicles
automotive paints **4**:1932, 1939, 1943–7
automotive trauma **5**:2565–75
  see also road traffic accidents
autonomy preservation, older adults **2**:445
autopedestrian incidents **5**:2254–5
  see also road traffic accidents
autopsies **1**:256–61
  alcohol analysis specimens **1**:89–90
  ancillary investigations **1**:260
  BAC interpretation samples **1**:70
  choking diagnosis **1**:227
  drug concentrations **4**:2208
  fire victims **3**:1535–6
  histology **3**:1468–73
  history **1**:256–9
  medical malpractice claims **4**:1699
  objectives **1**:259–61
  reports **1**:261
  techniques **1**:260
  see also postmortem...
autoradiography **1**:262
autosomal chromosomes **2**:795, 821
autosomal short tandem repeats analysis **5**:2679
autosomes **2**:795, 821
autosuggestion **2**:1066, 1069
availability, mtDNA databases **4**:1828–9
avenger-type sexual homicide **5**:2315
aversive racism **5**:2306
Avuncular Index **4**:1815
awareness, CAI **2**:493
axial illumination **4**:2050

BAC *see* blood–alcohol concentration
bachelor's degree programs **2**:898
Bacillus anthracis see anthrax attacks
back-calculation **1**:58
back-tracking **1**:58
backdoor systems **1**:143
background colors, optical enhancement **4**:2045
background noise **1**:126–8
background parameters, paint **4**:1950, 1952
background probabilities, glass evidence **3**:1357
background sampling, explosions **2**:1031, 1044
background subtraction, ILR **3**:1153
backscatter **3**:1348
backscattered electrons (BSEs) **4**:1687, 1799, 1802
backspatter, bloodstains **1**:372
backward extrapolation, BAC **1**:69–70
backward fragmentation **3**:1348
Bacon, Francis **3**:1576
BACs *see* blood–alcohol concentration
bacteria **1**:301; **2**:501, 505–6; **3**:1346; **4**:2099–100
bacterial pneumonia **4**:1844
'bagging and tagging' evidence **2**:1026
baking soda **2**:863–4
balance of harms **2**:577
balance of probability **2**:947
balancing test, police use of force **4**:2069
Balding and Donnelly DNA databases **2**:835, 837
Bali bombings **1**:413, 414, 415, 417; **2**:1028
ballistics
  3D reconstructions **5**:2263, 2264
  bloodstain pattern interpretation **1**:362
  definition **3**:1216
  external **3**:1202
  intermediate **3**:1201
  internal **3**:1200–1
  terminal **3**:1202–4
  wound ballistics **3**:1202–3, 1382
  see also firearms...
ballpoint pens **2**:686; **5**:2661–2
Balthazard, V. **3**:1288
banding, animal hair **3**:1405
bandpass (BP) filters **4**:1786
BAPP (beta amyloid precursor protein) **5**:2340
barbiturates **1**:116; **2**:729, 730, 736
barefoot impressions **3**:1244, 1245–7
  see also foot, impressions
bark **5**:2642
barn values, neutron activation **1**:151
barrel distortion **4**:2038
barrels (firearms) **3**:1198
barrier filters **4**:1786, 1787, 2047
Bartlett, Sir Frederic C. **2**:1066
basal ganglia **4**:2203
base fractures, skull **1**:402
base meth *see* methamphetamine
basement membrane **3**:1325
basic drug sample extraction **5**:2517
basis weight *see* grammage of paper
bast fibers **4**:1978
batch analyses **4**:2224–5
batch variations
  automotive paints **4**:1946
  household paints **4**:1949
  illicit drugs **2**:857, 858
  ink **3**:1547, 1549
  paint **4**:1946, 1949
battered child syndrome **1**:263–9; **2**:540; **4**:1990
battered spouse syndrome *see* battered woman syndrome
battered woman syndrome (BWS) **1**:270–2, 272–4; **5**:2433, 2447–8
  see also domestic violence
battered woman's reality (BWR) **1**:272–5
battery theory of liability **2**:451
Bayer filters **4**:2041, 2049
Bayes' rule
  parentage testing **3**:1563
  probabilistic statements **3**:1562
Bayes theorem **2**:968, 970
  CAI **2**:485–7, 489
  full form **2**:487
  handwriting examination **3**:1440
  identification/individualization **3**:1508, 1510
  likelihood ratio **3**:1583
  odds form **2**:486
  paint evidence **4**:1951–2
  retrocausal probability **5**:2493
Bayesian approach
  document dating **2**:689
  fibers evidence **3**:1100
  glass evidence interpretation **3**:1351–8
  quantity estimation **5**:2288–9
  sample size **5**:2284, 2285–6, 2287

Bayesian inference **3**:1350, 1580
Bayesian networks (BNs) **1**:276–80; **3**:1584; **4**:1841
1,4-BD (1,4-butanediol) **2**:869–70
BE *see* benzoylecgonine
beach marks **4**:1684
beam condensers **4**:1751
beam splitters **4**:2050
beatings, bloodstain patterns **1**:370–3, 374, 375, 378–81, 395
Becke line technique **4**:1770
Becton, Judge Charles **4**:2157
bedding packaging **4**:1927–8
bedsores **2**:905
beetles **2**:934, 936, 940
beginning phase of livor mortis **2**:704
behavior–brain relationship **4**:1862–9
behavior–free will debate **2**:574
behavioral artifacts **3**:1438, 1444
behavioral effects
  alcohol **1**:99–105
  benzodiazepines **1**:295–6
behavioral genomics **3**:1335–41
behavioral profiles **4**:2156–60
behavioral science **4**:1877–82
  evidence **1**:281–8; **5**:2431–2, 2438
  experiments **5**:2405–6
  issues/controversies **1**:287–8
  syndromes **5**:2431–2, 2433, 2438
behavioral tolerance **1**:76
behavioral toxicology **1**:290–3
behavioral treatment, sex offenders **5**:2335
beliefs, suggestibility **2**:733
believed to be (BTB) identification **1**:172
belting systems (vehicles) **1**:52, 53, 57; **5**:2543, 2566
belts, fabric marks **4**:1672
bench trials **2**:1005
beneficiaries, definition **3**:1376
benefits/costs analysis, victim decision making **2**:644, 645–6, 647
bent pin fault, printers **5**:2666–7
benzodiazepines **1**:293–7; **4**:2206
  alcohol interactions **1**:116
  amnesia **4**:2202
  behavioral toxicology **1**:291, 292
  degradation mechanisms **4**:2105
  drug-facilitated crime **2**:870–1
  multiple drug use **4**:2119
  opioid interactions **4**:1901
  oral fluid **4**:1914–6
  postmortem toxicology **4**:2131
  truth serum use **2**:730
  types **1**:293–4
benzoflavone **3**:1304, 1305, 1319
benzoylecgonine (BE) **2**:565, 566; **4**:1912, 1913, 1917
Bertillon, Alphonse **2**:746–7
best-case scenarios, eyewitness testimony **4**:2022
best evidence rule **1**:298–9
best interest standard **2**:558
best practices
  network analysis **1**:148
  paint examination **4**:1940
beta amyloid precursor protein (BAPP) **5**:2340
beta-binomial distribution **5**:2287
betrayal trauma **2**:785
beyond reasonable doubt cases **2**:947
Beyond the Pleasure Principle (Freud) **2**:571
bias
  acquiescence response bias **2**:591
  confirmation bias **3**:1575–8
  fingerprint interpretation **3**:1280
  ingroup/outgroup bias **5**:2306
  juries **3**:1604–5, 1605; **5**:2308
  mtDNA databases **4**:1830
  see also observer effects
biaxial crystals **4**:1770, 1777
bi-dimensional marks (tires) **5**:2480
Bichat, Marie-François Xavier **3**:1468
bicycling injuries **1**:102
bile **4**:2066, 2103, 2122; **5**:2499
bilirubin **4**:2082
binary approaches **1**:147–8; **4**:1840
binary errors **2**:955
  see also Type I error; Type II error
binders
  ammunition primers **3**:1190
  printing inks **3**:1541, 1544
  see also resin
binding in autoerotic asphyxiation **1**:246
binding thread security features **3**:1273
Binet, Alfred **2**:1066; **4**:2177
binge drinking **1**:103
ß-binomial distribution **5**:2287
bioaccidents **1**:300
bioagents *see* biological agents
bioavailability of drugs **4**:2116
biochemical examinations postmortem **4**:2076–87
biochemical markers, alcohol **1**:72–3
biocrime **1**:300–1, 307–9
biological age **1**:159, 188
biological agents **1**:299–313; **2**:501, 505–6
  see also chemical, biological, radiological and nuclear agents
biological material
  cannabis **2**:434
  explosive residues **2**:1033
  extraction techniques **2**:598
  fingerprint detection interaction **3**:1318–21
  sample packaging **4**:1930–1
  stain identification **1**:314–9
  terminal ballistics **3**:1202–3
  traces transfer **5**:2537
  see also individual materials
biological models, LTDNA profiles **3**:1569–70
biological profile **4**:1889–91
  age determination **1**:179–85
  anomalies **1**:194–6
  facial reconstruction **3**:1086
  identikit provision **3**:1511
  pathology **1**:194–6
  race/ancestry determination **1**:191–4
  sex determination **5**:2328–31
  stature determination **1**:194
biological specimens *see* biological material
biological swabs **1**:320–2

Biological and Toxins Weapons Convention (BWC) **2**:500–1
biological treatment, sex offenders **5**:2333–5
biology
  DNA **2**:798–9
  hair **3**:1415–8
  soil analysis methods **5**:2386
  wildlife forensics **5**:2636–7
  see also biological...
bioluminescence **4**:2046
biomechanics **5**:2346–8
biomedical engineering **5**:2565–75
biometric devices **1**:322–38
biometrics
  definition **1**:322
  documents **3**:1275
  facial comparison **3**:1085
  web resources **5**:2624–5
biomolecular methods **1**:181, **2**:755–6; **5**:2394, 2395
bioterrorism **1**:300–1, 307–9
biowarfare **1**:300, 301–2
bipeds **3**:1628
bipolar disorder **4**:2206, 2215; **5**:2415
  see also postpartum psychosis
birefringence **4**:1781
birth-related subdural hemorrhage **5**:2341, 2345
bitemarks **1**:399; **4**:1894–5
bites (insects) **2**:945
bizarre delusions **2**:742
BK virus (BKV) **3**:1343, 1345
black electrical tape **3**:1307
'black eyes' **1**:399
black powder fingerprint enhancement **3**:1318
black sheep effect (BSE) **5**:2306
Black & White adjustment, Photoshop **4**:2054–5
black widow killers **5**:2314
bladder injuries **1**:410
blameworthiness **3**:1608, 1609
  see also criminal responsibility
blanched bloodstain patterns **1**:362
Bland–White–Garland syndrome **2**:479
blank-cartridge guns **3**:1396
blastomycosis **1**:169
bleach, urine tests **2**:864
bleached hardwood **4**:1979
bleed-throughs *see* pull-ups
bleeding *see* bloodstains/bloodstain patterns; hemorrhages
blind procedures, confirmation bias **3**:1576, 1577
blind proficiency tests **2**:956
'blind stick' method **5**:2496
blind tests **2**:956; **3**:1090
blind verification **4**:2010
blister agents **2**:501
blobs, DNA profiling **1**:127
blocking tests, blood groups **1**:346
blood
  alcohol elimination rates **1**:68–9
  animal individualization **5**:2637
  benzodiazepine concentrations **1**:296
  characteristics **1**:362, 365, 369
  chemiluminescence reaction **3**:1648
  cocaine concentrations **2**:566–7
  DFC investigations **2**:873–4
  DNA source **2**:821
  driver impairment evidence **2**:881, 883
  drops
    characteristics **1**:362
    shape of **1**:365
    size considerations **1**:369
    surface texture influence **1**:364, 365
  drug concentrations **2**:566–7; **4**:2117–8
  extraction techniques **2**:598
  fingermark enhancement **3**:1309–10, 1320
  glucose metabolism **4**:2078
  microchemical tests **4**:1747
  oral fluid drug transfer **4**:1906
  photoluminescence optical enhancement **4**:2047
  sampling **1**:88–9; **4**:2102
  toxicology **4**:2097, 2098–9, 2102, 2107, 2121, 2123, 2128; **5**:2496–7, 2504
  transfer **5**:2537
  see also bloodstains/bloodstain patterns
blood agents **2**:501
blood–alcohol concentration (BAC) **1**:291; **2**:877, 879
  analytical methods **1**:82–3, 92, 93–4, 97–8
  interpretation **1**:58–80
  intoxication **1**:100
  lethal dose **1**:121
  postmortem toxicology **4**:2100, 2128
blood-alcohol curve **1**:63–4
blood ethanol concentration *see* blood–alcohol concentration
blood group polymorphisms **1**:339, 340–9
blood pattern analysis **2**:821
  see also bloodstains/bloodstain patterns
blood pressure deception test **3**:1331–2
bloodspatter **1**:362, 363–5, 365–6, 369–77
bloodstains/bloodstain patterns **1**:360–96
  absorption-elution technique **1**:342
  expirated **1**:381–3
  haptoglobin detection **1**:355
  identification **1**:314–5
  impact angle **1**:366–8
  lifting **1**:386–8
  luminol tests **3**:1645, 1648–54
  principles **1**:362
  projected blood **1**:378–83
  recovering/documenting **1**:386–8
  shape of stain **1**:365–6
  size of stain **1**:368–9
  transfer patterns **1**:384–5, 390
  within body **1**:362
blowflies **2**:936, 940
blue ribbon juries **3**:1605
bluebottles **2**:936, 940
Bluestar Forensic formulation **3**:1654
blunt force injuries **1**:396–411; **5**:2560–1
BMI *see* body mass index
BNs *see* Bayesian networks
Bobbitt, Lorena **5**:2447–8
body damages, malpractice cases **4**:1692
  see also human bodies; human remains
body effects, diatoms **2**:751
body fluids **5**:2555
  see also fluids; *individual fluids*

body height estimation *see* height
body mass index (BMI) **1**:61, **4**:1848, 1849
body modifications **1**:171
body position at scene **1**:227
body recovery, disaster scenarios **2**:768–9
  see also recovery methods
body removal team **4**:1676
body temperature *see* algor mortis; temperature...
bodyweights
  algor mortis **2**:712
  BAC correspondence **1**:61
bomb-pulse dating **1**:418–21
Bombay blood group **1**:341
bombings/bombs
  Bali **1**:413, 414, 415, 417; **2**:1028
  Madrid train **3**:1290, 1577
  mechanisms **2**:823, 1034, 1035, 1044–5
  scene management **1**:412–8
  World Trade Center 1993 **2**:1031; **3**:1577
bond paper **4**:1972
bond strength, polymers **3**:1110
bones
  age determination **1**:179–80, 180, 183, 189
  anomalies/pathology **1**:194
  burn injuries **3**:1534
  DNA extraction **2**:1063
  DNA source **2**:826
  fractures **1**:170, 265, 266, 403
  impressions **1**:403
  lesions **1**:168, 170; **5**:2557–63
  mtDNA analysis **4**:1834
  PMI determination **4**:2090
  positive identification of remains **3**:1512, 1513–6
  race/ancestry determination **1**:192, 194
  radiocarbon dating **1**:420; **5**:2232
  sex determination **5**:2328
  skeletal trauma analysis **5**:2557–63
  soft tissue relationship **3**:1087, 1090
  species determination **1**:153–4; **5**:2393–5
  stature determination **1**:194
  see also cranium; human remains; skeletal...
BootCamp, Macintosh computers **2**:585, 586
borderline personality disorders **2**:572
Boston Process neuropsychological assessment **4**:1874
BOT networks **1**:144
botany **1**:423–30; **4**:2089, 2091; **5**:2625
  see also plants
bovine albumin **1**:346
bow effect, bullet impacts **3**:1220
boxer's attitude **3**:1532
BP (bandpass) filters **4**:1786
BrAC (breath-alcohol analysis) **1**:81–99
Bracton, Henry de **1**:282, 283
brain
  adolescent development **3**:1611
  behavior relationship **4**:1862–9
  burn injuries **3**:1534–5
  fingerprinting **2**:725
  imaging **2**:724–7
  postmortem toxicology **4**:2122; **5**:2500
  swelling **1**:407
  tissue sampling **4**:2103
brain-damage/injury **1**:268; **4**:1863
brainwashing *see* persuasion
braking distance (automobiles) **5**:2255–6
Brandeis, Justice **5**:2576
brands, writing instruments **3**:1547, 1548
breaking glass **3**:1348–9
breath-alcohol analysis (BrAC) **1**:81–99
breath control play **1**:246
breech face, cartridge cases **3**:1210
breeder documents **3**:1256
Brennan, M. and R.E. **2**:657–8
brightfield microscopy **2**:989; **4**:1759, 1763–73
British Infanticide Act (1922) **4**:2139
Brittain, Robert **5**:2312
brittle behavior, materials **4**:1682
brittle fractures, light globes **3**:1637
'broad brush' courses **5**:2545
bronchopneumonia **4**:1844
Brown Simpson, Nicole **1**:391–5
Brugada syndrome **2**:471
bruising **1**:398–400, 406, 408; **2**:905
brush abrasions **1**:396
brushes/brushing
  fingerprint powders **3**:1294–5
  paint application **4**:1932
  powder brush DNA transfer **3**:1320
  trace evidence sample recovery **5**:2294
BSE (black sheep effect) **5**:2306
BSEs *see* backscattered electrons
BSS *see* battered woman syndrome
BTB (believed to be) identification **1**:172
bubble-jet printing **5**:2670
bulk analysis tools **3**:1191–2; **4**:1886
bulk explosives **2**:1034, 1036
bulk instrumental analytical methods **3**:1191–2
bulk sampling, trace evidence **5**:2293
bulking thickness of paper **4**:1973
bullet lesions *see* gunshot wounds
bullet wipe-off **3**:1387
bullets **3**:1189
  fragment analysis **3**:1197
  holes in glass **3**:1220–1
  identification **3**:1207–9
  impact damage **3**:1220
  ricochet **3**:1221
  shooting scene investigation **3**:1219–20
  trajectory approximation **3**:1220
  see also ammunition; ballistics; projectiles
Burch v. Zinermon, 494 U.S. 113 (1990) **5**:2586
burden of proof **1**:282
burglary cases **2**:616–7, 824, 826–7; **4**:1947
burials **1**:200, 202; **3**:1496–9
  see also graves
burking (positional asphyxia) **1**:227
burns **3**:1529–37
  alcohol-related injuries **1**:102
  battered children **1**:266
  chemical injuries **3**:1538–9
  classification systems **3**:1532
  external findings **3**:1529–32
  internal findings **3**:1533–5
  see also scalds
burnt remains **3**:1495
1,4-butanediol (1,4-BD) **2**:869–70

BWC (Biological and Toxins Weapons Convention) **2**:500–1
BWR (battered woman's reality) **1**:272–5
BWS *see* battered woman syndrome
by-products of illicit drugs **2**:853, 854
Byron Mitchell v. United States, CA-No. 96–407 (1999) **3**:1289
bystander injuries **4**:2074

C_3 alkyl benzenes **3**:1150
C_4 alkyl benzenes **3**:1153
^{14}C *see* Carbon-14
CABs (conformity assessment bodies) **4**:2220
CAC (California Association of Criminalistics) **5**:2548
CACs (child advocacy centers) **2**:550–1
cacti **4**:2063
CAD *see* coronary artery disease
cadaver dogs **3**:1497
see also dogs
cadavers **3**:1087, 1305; **5**:2504
see also human remains; postmortem...
Caddell v. State, 287 N.C.266: 215 S.E.2d 348 (1975) **1**:254
Caesar, Julius **3**:1576
CAF *see* cyanoacrylate fuming
CAGE *see* Computerised Assistance for Glass Evidence
CAI *see* case assessment and interpretation
calcar area, feet **3**:1323
calcination of bones **3**:1534
calcium, water imbalances **4**:2085
calcium carbonate **4**:1974
caliber of firearms **3**:1381
calibration
breath-alcohol analytical instruments **1**:97
cameras **3**:1625
graphs **4**:2224
preparation/storage solutions **4**:2223–4
calibration laboratories **1**:3–8, 11–2, 13, 14–5, 16
California Association of Criminalistics (CAC) **5**:2548
caliper of paper **4**:1968, 1973
callus formation, skeletal remains **5**:2558–9
Camelford poisoning case **2**:948–9, 949, 952
camera match moving **5**:2261
cameras
calibration **3**:1625
control **2**:635–41
matching (height estimations) **3**:1628
nightshot **1**:337
see also digital cameras; images, processing; photography; surveillance images
Cameron v. Rouse, 373 F.2d 451 (D.C. 1966) **5**:2583
CAN (cardiovascular autonomic neuropathy) **4**:1851
Canada
adult female offenses **1**:38
dangerous offender/long term offender legislation **2**:670–1
duty to warn of dangerousness **2**:670, 888
expert evidence admissibility **2**:1004
mental health courts **4**:1719, 1720
National DNA Database **2**:678
Canadian Society of Forensic Science (CSFS) **5**:2622
cannabinoids **1**:113; **2**:432–3; **4**:1908–11
cannabis **2**:431–6; **5**:2363
abuse **2**:433–4
alcohol interactions **1**:113
botanic identification **1**:424, 425–6, 429
DNA evidence **2**:827
driver impairment **2**:877
drug-facilitated crime **2**:871
illicit drug analysis **2**:845–7
immunoassays **5**:2512
medical use **2**:433
oral fluid **4**:1906, 1908–11, 1918
pollen evidence **4**:1966
postmortem toxicology **4**:2124
profiling **2**:851–2, 852, 855
psychoses **1**:292
sources **2**:432–3
sweat **5**:2425–6
tolerance **1**:292
urine testing **2**:866, 871–2
see also Δ^9-tetrahydrocannabinol
Cannan, P.N. **2**:660
Canter, David **4**:2157, 2158
CAP *see* common approach path
capacitive sensors **1**:326
capacity
assessment **2**:437–43
civil law **2**:438–40
criminal law **2**:440–3
behavioral science evidence **1**:287
consent
elder abuse **2**:907–8
to medical treatment **2**:439–40, 450–4
to research **2**:440
contractual **2**:439
decision-making
about capacity **2**:448–9
adolescents **3**:1609
definition **2**:450
diminished capacity defense **3**:1339, 1435; **4**:2150–1, 2152; **5**:2446–7
independent living assessment **2**:444–9
insanity defense waiver **2**:442
legal counsel waiver **2**:442
medical treatment consent/refusal **2**:439–40
Miranda rights waiver **2**:463–7
privilege waiver **2**:440
research consent **2**:440
sentencing **2**:442
soil databases **5**:2382–3
testamentary **2**:438–9
trial process **2**:440–2, 456
see also civil commitment
capillary action, bloodstain patterns **1**:389
capillary electrophoresis (CE) **2**:845
Capital Jury Project (CJP) **4**:1819, 1822
capital litigation **4**:1818–23
capital punishment *see* death penalty
capture
devices **1**:324, 334; **4**:2048–9
files **1**:146

images in AFIS system **1**:249
motion in 3D reconstructions **5**:2262
carbohydrate deficient transferrin (CDT) **1**:73
Carbon-14 (^{14}C) **1**:418–21
see also radiocarbon...
carbon dioxide (CO_2) **1**:418–9, 419–20; **5**:2231, 2232
carbon disulfide **3**:1141
carbon monoxide (CO)
embalmed specimens **4**:2102
poisoning **1**:233; **4**:2133
stability during storage **4**:2105
carbon monoxide hemoglobin (CO-Hb) **3**:1535–6
carbon ribbons, typewriters **5**:2665
carbon strip ILR isolation methods **3**:1139–40, 1142
carbonization of remains **3**:1513
carboxy-THC **2**:866
carboxyhemoglobin **1**:233; **5**:2505
cardboard surfaces **3**:1319
see also porous surfaces
cardiac...
causes of sudden death **2**:468–80
injury histology **3**:1472–3
see also heart...
cardiomyopathies **2**:472, 473–4
cardiovascular autonomic neuropathy (CAN) **4**:1851
cardiovascular medications **1**:113–4
Cardozo, Justice Benjamin **2**:439
care issues **2**:906, 909
facilities for elders **2**:904, 905, 914
mistakes (malpractice) **4**:1702
see also community care; health care; perpetrator risk factors
Carmichael v. Kumho Tire see Kumho Tire v. Carmichael
carrion beetles **2**:940
cars *see* automotive. . .; vehicles
cartilage **1**:420
cartridge cases **3**:1189
DNA evidence **3**:1222
estimation of time since discharge **3**:1198
fingermark detection **3**:1308–9
identification **3**:1209–11
mass grave investigation **4**:1678
shooting scene investigation **3**:1219–20
see also cartridges
cartridge discharge residues (CDR) **3**:1189, 1198
cartridges **3**:1200, 1203–4, 1204, 1381–2
see also ammunition; cartridge cases
case assessment and interpretation (CAI) **2**:483–96, 972–4
case-by-case basis, hypnotic testimony **3**:1502–3
case congregation analysis **1**:285, 288
case histories, paint evidence **4**:1940–2
case law, expert reports **5**:2269, 2270
case manager's bias **3**:1577
case reviews, pharmacogenomics **4**:2018
case-to-case comparison, illicit drugs **2**:857
Caspi, Avshalom **3**:1338
cast-off bloodstain patterns **1**:365, 378–81, 390
casting techniques **2**:963–7; **3**:1087; **5**:2325, 2489–92
castor oil plant **4**:2064–5
castration treatment **5**:2334, 2336
casts *see* casting techniques
Casualty Bureaus **2**:765–6
CAT *see* computed axial tomography
cat hairs **3**:1404, 1409–10, 1411
catagen hairs **2**:825–6; **3**:1417, 1420
categorical opinions **3**:1564–5
Category Test (Halstead Impairment Index) **4**:1865
Catha edulis **4**:2061
cathine **4**:2061
cathinone **4**:2061
cathodes **4**:1795
cathodoluminescence (CL) microscopy **4**:1788, 1798, 1800
Cattell, James M. **4**:2177
Caucasians
characteristics **1**:191–2
sentencing **5**:2306
causal component (capacity assessment) **2**:453
causality
probability concept of **5**:2492, 2493
retrocausal probability **5**:2492–3
causation
automotive trauma analysis **5**:2565–75
behavioral genomics **3**:1341
elder abuse **2**:913
fire scenes **3**:1123
tort of malpractice **4**:1664
see also risk factors
cause of death
fire victims **3**:1537
functional **4**:2076
histology **3**:1468–73
natural causes **4**:1843–61
noncardiac **4**:1843–61
postmortem toxicology **4**:2119, 2120
sharp force injuries **5**:2654–5
skeletal remains trauma analysis **5**:2557
sudden death **2**:468–80
cause and effect (Bayesian networks) **1**:278
CBRN *see* chemical, biological, radiological and nuclear agents
CCD *see* charge coupled device
CCTV systems *see* closed-circuit television
CD *see* conduct disorder; cross direction
CDR *see* cartridge discharge residues; crash data recorders
CDs, Macintosh computers **2**:585
CDT (carbohydrate deficient transferrin) **1**:73
CE *see* capillary electrophoresis; tetryl
CEDIA (cloned enzyme donor immunoassay) **5**:2514–5
ceiling jets **3**:1178
ceiling principle **2**:497–8
celebrity stalking **5**:2399
cell phones *see* mobile phones
cells
computational fluid dynamics **3**:1183
DNA biology **2**:792–4
cellular phones *see* mobile phones
Center for Disease Control and Prevention (CDC) **3**:1459
center punching serial number obliteration **5**:2325
central blood **5**:2496

central nervous system (CNS)
depressants **1**:108, 117, 291, 293; **2**:868–9, 871
opioid effects **4**:1896
stimulants **1**:291; **2**:563
centric diatoms **2**:749
centric GSM **3**:1360–71
ceramics **4**:1687
cerebral contusions **1**:406
cerebral edema **1**:407
cerebral injuries **1**:406–7
cerebrospinal fluid (CSF) **1**:71; **4**:2103; **5**:2500
certification
accreditation distinction **1**:11
criminalistics **5**:2545–54
definition **3**:1173
fire investigation **3**:1171–4
forensic electrical engineering **2**:921–2
laboratory personnel **1**:7
process **5**:2549–50
programs **5**:2548–9, 2550–4, 2554
CF *see* corrective factors
CFA (confirmatory factor analysis) **4**:2199
CFAST zone model **3**:1183
CFC (ChloroFluoroCarbon) **3**:1296
CFD *see* computational fluid dynamics
CGS (Crow–Glassman scale) **3**:1532
chain of custody/possession of tangible evidence **2**:498–500, 606–7, 759; **4**:2228
chain rule (Bayesian networks) **1**:277
chamber marks, cartridge cases **3**:1210
chance matches, STR profiles **5**:2371–2
chancery court **1**:282, 284
changes after death *see* postmortem changes
char
alligatoring effect myth **1**:209, 210–1
combustion of solids **3**:1112
depth myth **1**:209, 213
interpretation of **1**:211
location myth **1**:213
see also burns
character modeling **5**:2261
characterization of soil **5**:2380, 2384–5, 2385–6
charge coupled device (CCD) **2**:990; **3**:1523, 1524, 1525; **4**:2049
Charge Switch™ DNA extraction **2**:1062
Charpy impact test **4**:1685–6
charring *see* char
charts, quality control **4**:2225–6
Chatterjee, Salil **3**:1284
Chelex® DNA extraction **2**:804, 1061–2
chemical analysis
explosion debris **2**:1032, 1035–42
microbiological investigations **1**:308
wildlife forensics **5**:2638–9
see also chemical imaging
chemical anoxia **1**:233
chemical, biological, radiological and nuclear (CBRN) agents **2**:500–6; **4**:1883
see also chemical warfare agents; nuclear forensics
chemical erasures, documents **1**:132
chemical ignitions **3**:1114
chemical imaging **4**:1757
chemical injuries **3**:1538–9
chemical ionization mass spectrometry (CI-MS) **2**:595, 597
chemical microscopy *see* microchemistry
chemical oxidation, alcohol **1**:84
chemical pulps **4**:1970, 1978
chemical reactions, document security **3**:1273
chemical reagents, fingerprint enhancement **3**:1319–20
'chemical restraint' **4**:2216
chemical treatment, evidence enhancement **4**:2040–1
chemical warfare agents (CWAs) **2**:501, 507–28
see also chemical, biological, radiological and nuclear agents
Chemical Weapons Convention (CWC) **2**:500–1, 508
chemicophysical age determination methods **1**:184
chemiluminescence **3**:1646, 1648; **4**:2046
see also luminol
chemiluminescence immunoassay **5**:2515
chemiresistor detection **2**:513–9
chemistry
fire **3**:1103–12
ink features **3**:1543
'kitchen sink' ecology **2**:951–2
luminol properties **3**:1645–6
paper **4**:1974
see also chemical...
chert **4**:2003
chest compression asphyxia **1**:227
chest injuries **1**:407–9
Chestnut Lodge case *see Osheroff v Chestnut Lodge*
child abuse **4**:1989–93
battered child syndrome **1**:263–9
causation theories **4**:1991–2
chemical injuries **3**:1538
contextual factors **4**:1992
forensic interviews **2**:550
functional amnesia **2**:786
implanted memories **2**:733, 734–5
parental factors **4**:1992–3
physical effects **4**:1990
protective factors **4**:1991
psychological effects **4**:1990–1
scalding injuries **3**:1538
see also child neglect; childhood sexual abuse
child advocacy centers (CACs) **2**:550–1
child custody **1**:287; **5**:2602
see also custody evaluation
child forensic interviews **2**:532–4, 550, 553, 554
child neglect **4**:1991
see also child abuse
child protection **4**:1987, 1989–98; **5**:2543
see also child custody; parental rights
child sexual abuse *see* childhood sexual abuse
child sexual abuse accommodation syndrome (CSAAS) **2**:537–41
child witnesses **2**:549–52
cross-examination **2**:656–60, 660
interviews **2**:550
jurors' reactions **2**:551–2
research **2**:549–50, 552
sexual abuse cases **2**:549–50, 551, 552

testimony **2**:549–50, 550–1, 552
childhood amnesia **4**:1716; **5**:2245
childhood fire discovery **3**:1226
childhood sexual abuse (CSA) **1**:239; **2**:529–36; **4**:1990
child witnesses **2**:549–50, 551, 552
clinical features of victims **2**:531
close associate patterns **2**:530–1
cross-examination **2**:657
definitions **2**:529
epidemiology **2**:529–30
evaluation process **2**:531–2
female aggression **1**:39
interviews **2**:532–4
intrafamilial patterns **2**:530
laboratory examination **2**:534
male perpetrators **4**:1992
neurobiological consequences **2**:535
parental alienation **4**:1981
patterns **2**:530–1
perpetrators **2**:530–1; **4**:1992
physical examination **2**:534
prognosis after abuse **2**:535–6
protective factors **4**:1991
psychosocial sequellae **2**:535–6
recovered memory **4**:1712, 1713–5; **5**:2246–7, 2248
serial murderer history **5**:2318
special tests **2**:534
stepwise interview **2**:533
stranger abuse **2**:531
traumatic memory **5**:2245–7, 2248
validity determination **2**:534–5
see also child abuse; child sexual abuse accommodation syndrome
childrearing evaluation **5**:2606
children
abstract thinking **2**:544
age determination of remains **1**:179
aggression **1**:45
arrest process **2**:546–7
asphyxial games **1**:246
civil commitment **2**:559
conflict-escaping behavior **4**:1982
culpability **2**:543–6, 547
custody/visitation evaluation **5**:2607
delinquency **4**:1985–6
developmental immaturity **2**:543–4
direct examination **2**:658
disability **4**:1876
earnings **4**:1984
epilepsy **4**:1872
fire discovery **3**:1226
functional amnesia **2**:786
genetic disorders **4**:1871
homicide–suicide **5**:2418–9
inhibition skills **2**:555
intelligence **2**:546; **4**:1870
interrogation **2**:546–7; **3**:1592
learning disabilities **4**:1872, 1873
lying child syndrome **5**:2440
meetings with **5**:2604–5
mental disorders **2**:544–6, 547
mental illness **4**:1872
neurological disorders **4**:1871–2
neuropsychological assessment **4**:1869–76
parental alienation **4**:1981–4
parental punishment **4**:1985
perceived autonomy **2**:544
poisonous plant ingestion **4**:2058
respiratory deaths **4**:1843
sadism **5**:2320
serial murderers **5**:2314
substance use disorders **2**:546
suggestibility **2**:549, 551, 553–5; **3**:1592
systemic illness **4**:1872
traffic fatality prevention **5**:2543
traumatic brain injury **4**:1872
see also adolescents; child. . .; infants; juvenile. . .; minors; pediatric. . .; subadult age category; young adults; young offenders
Children's charter **1**:264
China
forensic hospitals **2**:560
National DNA Database **2**:678
Chinese medicine **4**:2058, 2060
chiral analysis **2**:599
chivalry thesis **5**:2309
chlorates **2**:1044
chloride **4**:2085
ChloroFluoroCarbon (CFC) **3**:1296
choking **1**:227; **5**:2345
choking agents (chemical warfare) **2**:501
chop wounds **5**:2646, 2654
Christmas Tree stain **1**:316
chromatic beam splitter **4**:1786
chromatic modification, digital images **4**:2054–5
chromatograms
definition **1**:85
examples **1**:88
explosion debris **2**:1042
ILR **3**:1142, 1143–4, 1146, 1153, 1167
see also thin layer chromatography
chromatography tests **2**:595, 597–8
analytical artifacts **4**:2106
cocaine analysis **2**:567–8
plant analysis **4**:2066
postmortem toxicology **4**:2124, 2125–7
sweat **5**:2427
urine **2**:862
see also gas chromatography; liquid chromatography
chromosomal circumstances, definition **3**:1336
chromosomes **2**:795
chronic renal failure **4**:2082, 2087
chronic tolerance **1**:75, 123
chronological age **1**:159, 188
chronology approach, document dating **2**:688
CI (cognitive interviews) **2**:554
CI-MS *see* chemical ionization mass spectrometry
CIDI (Composite International Diagnostic Interview) **4**:1880
cigarette ends **2**:824; **4**:1930
circle hypothesis, serial murder **5**:2316
circle search strategy **2**:623
circuit breakers **2**:927, 929
circumstances, framework of **2**:971, 974
circumstantial evidence, ecological cases **2**:948–9

CITES *see* Convention on International Trade in Endangered Species Fauna and Flora
citizen–police encounters **2**:650–1
CITs (concealed information tests) **2**:721, 725, 726
civil cases *see* civil law
civil commitment **2**:556–60
  abuses **2**:560
  affecting factors **2**:558–9
  children **2**:559
  civil rights **5**:2586
  dangerousness **2**:668–70; **5**:2578
  forced medications **2**:559
  incompetence **2**:558–9
  inmates **2**:559
  inpatients **2**:556–8
  insanity acquittees **2**:559
  legal foundations **5**:2577
  medical treatment consent/refusal **2**:440
  minors **2**:559
  outpatients **2**:558
  personality disorders **5**:2578
  severe mental illness **2**:649
  sex offenders **2**:559–60, 668–9; **5**:2578–9
  special cases **2**:559–60
  therapeutic jurisprudence **5**:2451–2
  see also capacity, assessment; compulsion
Civil Contingencies Act (UK) **2**:765
civil engineering technology **2**:950
civil law **2**:561–2, 1004–6
  capacity assessment **2**:438–40
  consent to research capacity **2**:440
  contractual capacity **2**:439
  cross-examination **2**:663
  dangerousness **2**:668–70
  Daubert case impact **2**:695
  discovery **2**:782
  expert evidence admissibility **2**:1003, 1004, 1007–11
  medical treatment consent/refusal capacity **2**:439–40
  privilege waiver capacity **2**:440
  psychological autopsy **4**:2163
  testamentary capacity **2**:438–9
  see also civil litigation
civil liberties **5**:2576, 2578, 2579
  see also civil rights
civil litigation **2**:775–8, 782, 783
  see also civil law
Civil Procedure Rules (CPR) **2**:774, 775, 783; **5**:2268, 2269
civil rights **2**:560; **5**:2584–7
  see also civil liberties
civil settings, malingering **4**:1660
Civilization and its Discontents (Freud) **4**:2187
CJA *see* Criminal Justice Act
CJP *see* Capital Jury Project
CL microscopy *see* cathodoluminescence microscopy
clades **1**:423
claims
  management by ecologists **2**:946
  medical malpractice cases **4**:1698–9, 1699, 1705
clamshell marks **4**:1684
clandestine laboratories **2**:503
class characteristics
  firearms identification **3**:1205–6, 1207, 1208
  printed documents **3**:1558
  shoemark examination **3**:1232–3
classical identification markers **1**:339–56
classification systems
  burns **3**:1532
  diatoms **2**:749
  entrance wounds **3**:1390–3
  fingerprints **1**:249
  plants **1**:423–5
  textile fibers **2**:986–8
clean man/dirty man approach **2**:504
cleaning attempts, hemoglobin **3**:1652
clearance (drug elimination) **4**:2116
clearcoat *see* top coat paints
Clifford v. Michigan **3**:1123
climate data **2**:939, 943
clinical assessment **4**:2173–5, 2182
  see also clinical risk assessments
clinical autopsies **1**:257–8, 259, 260, 261
  see also autopsies
clinical biochemistry **4**:2076
clinical confusion **5**:2244–6
clinical factors
  base fractures **1**:402
  crash tests **5**:2570
  independent living assessment **2**:447–8
  pharmacogenomics application **4**:2015–7, 2020
  truth serum application **2**:736
clinical forensic toxicology **5**:2506–7
clinical neuropsychologists **4**:1862
  see also neuropsychological assessment
clinical prediction
  risk assessment **5**:2275
  risk factors **5**:2274
  violence **5**:2274
clinical psychologists *see* mental health professionals
clinical psychopharmacology *see* psychopharmacology
clinical risk assessments **5**:2598, 2599
clinical supervision **2**:762
clinical toxicology **5**:2503, 2506–7
clinical violence risk assessment **5**:2599
clinician role, syndromes **5**:2436
clinicide **5**:2315
cloned enzyme donor immunoassay (CEDIA) **5**:2514–5
close associate patterns, child sexual abuse **2**:530–1
close-range gunshots **3**:1392–3
close-up photography **1**:386; **2**:630, 632, 633–5
closed-circuit television (CCTV) **2**:551, 552, 587; **3**:1520–2, 1526
  see also surveillance images
closed questioning **2**:656
cloth gloves **4**:1670
clothing
  autopsy external examination **1**:260
  fabric marks **4**:1669
  firearm discharge residue patterns **5**:2351–2
  packaging **4**:1927–8, 1930
  paint found at random on **4**:1950
  pollen evidence **4**:1964

residues, shooting scene investigation **3**:1223
transport **4**:1930
video imaging **2**:995–6
see also textiles
clusters, Macintosh computers **2**:586
CMO (common main objective) stereomicroscope **4**:1762–3
CMOS (complementary metal oxide semiconductor) **4**:2049
CMR (comprehension of Miranda rights) **2**:465
CMR-R (comprehension of Miranda rights-recognition) **2**:465
CMS (consecutively matching striae) **3**:1209
CMV (comprehension of Miranda vocabulary) **2**:465
CMY color model **4**:2043–5
CN *see* cyanide
CNS *see* central nervous system
CO *see* carbon monoxide
CO_2 *see* carbon dioxide
CO-Hb *see* carbon monoxide hemoglobin
coalescent model, mtDNA **4**:1827, 1828
coancestry coefficient **4**:1811, 1817; **5**:2367
see also population genetics parameter
cocaethylene **1**:117
cocaine **2**:562–8
abuse **2**:562–3
alcohol interactions **1**:117
behavioral toxicology **1**:290, 291, 292
driver impairment **2**:878
history of **2**:562
illicit drug analysis **2**:848–9
metabolism **4**:1912
oral fluid **4**:1912–3, 1917
perpetrators as victims **5**:2434
pollen evidence **4**:1966
profiling **2**:851–2, 852–3, 855
stability **4**:2105
sweat **5**:2422, 2423, 2425–6, 2426
codeine **2**:865; **4**:1896, 1899, 1913, 1914, 2017; **5**:2422
codes of ethics **2**:957–62
coding regions, genes **2**:795
coding systems
footwear impressions **3**:1250
tread patterns **3**:1241–2
CODIS *see* combined offender DNA index system
codon, nucleotides **1**:340
coerced-compliant false confessions **3**:1591, 1593
coerced-internalized false confessions **3**:1591
coerced-reactive false confessions **3**:1591
coercion **2**:571, 573, 732; **3**:1591, 1593, 1611
see also compulsion; peer pressure; persuasion
coextraction, DNA/RNA **2**:1064
cofactors, enzymes **2**:954
'coffin-flies' **2**:943
COfiler™ systems **2**:569–70
cognitive abilities
adolescents **3**:1609
age-related changes **2**:450, 453
genomic information **3**:1337
cognitive behavior therapy (CBT) **4**:2146
cognitive-behavioral treatment **3**:1226; **4**:2146; **5**:2335
cognitive deficits **2**:918; **4**:1734
see also cognitive impairment
cognitive dissonance **4**:1982
cognitive effects, benzodiazepines **1**:295
cognitive functioning, independent living capacity **2**:445–7
cognitive impairment *see* cognitive deficits
cognitive interviews (CI) **2**:554, 918
cognitive malingering **4**:1660, 1868
cognitive reevaluation **2**:646–7
cognitive testing **2**:448; **3**:1553; **4**:2175
see also neuropsychological assessment
colchicine **4**:2061
Colchicum autumnale **4**:2061
cold case reviews **2**:677, 682–3
cold field emission **4**:1796
Colello, G. **2**:887
collagen extraction **5**:2232
collateral interviews **4**:2165
collected handwriting exemplars **3**:1440
collection control, network data **1**:142
collection methods/devices
bloodstain luminol tests **3**:1649
casting evidence **2**:963–7
CBRN evidence **2**:504
CWA samples **2**:523
DNA **2**:821, 822
ecological evidence **2**:948–9, 952
entomological evidence **2**:943–4
explosion debris **2**:1029
explosion scene evidence **2**:1023, 1026
oral fluid **4**:1906–7, 1909, 1911
pollen samples **4**:1960–6
postmortem artifacts **4**:2101–2
soil **5**:2381
sweat **5**:2422–6
urine for drug tests **2**:860–1
see also extraction methods; recovery methods
collision-phase prevention measures **5**:2543
collisions *see* road traffic accidents
color deconvolution **4**:1688
color film spectral sensitivity **4**:2049
color frequency distributions, paint **4**:1945–6
color of law standards **4**:2068–9
color of particles
brightfield microscopy **4**:1772
polarized light microscopy **4**:1778, 1779, 1780
color perception **3**:1293
color photocopiers **5**:2673
color synthesis models **4**:2043–5
color tests **1**:314; **3**:1191
see also spot tests
color wheels **4**:2045
Colorado v. Connelly, 479 U.S. 157 (1986) **2**:464
colorants, ink analysis **3**:1541
colorimetric detectors **2**:513
colorimetric tests **2**:1038
column supports, drug screening **5**:2522–4, 2527–9
columnar-pattern fire plumes **3**:1120
coma cases **4**:2076–7, 2081, 2086
combined offender DNA index system (CODIS) **2**:678; **4**:2021, 2032, 2033
COfiler systems **2**:569–70
loci (mini-STRs) **4**:1804–5, 1806, 1807

combined offender DNA index system (CODIS) *(cont)*
short tandem repeats **5**:2357–9
combined soil analysis methods **5**:2386
combustion **3**:1107–12, 1175–6
see also fire
command hallucinations **3**:1433, 1435; **5**:2274
commercial market place development **2**:494
see also market-based forensic science
commercial products
short tandem repeats assays **5**:2357
urine test adulteration **2**:864–5
see also individual products
commingled remains **3**:1495; **5**:2237
see also mass graves
commission acts **2**:955
commitment to psychiatric hospital *see* civil commitment
committees, electrical engineering **2**:922
common approach path (CAP) **2**:620, 622
common authorship, handwriting **3**:1438
common knowledge, matters of **3**:1601
common law systems
depositions **2**:772
discovery **2**:782
ethical rules **2**:958
expert malpractice **4**:1663
expert reports **5**:2269
hypothetical question evidence **3**:1505
judicial notice **3**:1601
trial and appeals procedure **2**:998–9
see also adversarial justice system
common main objective (CMO) stereomicroscope **4**:1762–3
common sense concept, product use **3**:1489
Commonwealth v. Louraine, 453 N.E. 2D, 437 (Mass. 1983) **5**:2586
Commonwealth v. Serge 586 Pa. 674 (Pa. 2006) **2**:581
communication
CAI **2**:492, 493
firesetting as **3**:1227–8
risk communication **2**:761
statistical evidence **5**:2401–8
workplace threats **5**:2460, 2462–4
communicative arson **3**:1227, 1228
community-based treatment **5**:2577
community care **2**:649, 669–70
community epidemiology **2**:952
community psychology **2**:760
comorbidity
ADM disorders **4**:1881
psychopathy **4**:2194
compactor rollover case **3**:1493–5
comparator samples *see* control samples
comparison
barefoot impressions **3**:1244, 1245
definition **3**:1081, 1252
facial comparison **3**:1081–6
fingerprints **1**:251–2; **3**:1277–8, 1282–90, 1602; **4**:2011
foot impressions **3**:1244, 1245, 1252–3
glass evidence **3**:1349–50
hair evidence **3**:1424
handwriting **3**:1438, 1439–40, 1443, 1451, 1452, 1454–6
identification techniques (radiology) **5**:2238
illicit drug profiling **2**:857
ink evidence **3**:1542, 1543, 1546, 1548–50
process **3**:1083–5
shoe to foot **3**:1247–8
shoemark examination guidelines **3**:1230–4
signatures **3**:1456–7
soil samples **5**:2383–6
striation marks **5**:2487, 2489
see also matching process
comparison light microscopy **5**:2487–9
comparison microscopy **3**:1422; **4**:1760, 1788–9
animal hairs **3**:1410
design **4**:1789
firearms identification **3**:1206–7, 1207, 1208, 1209
forgeries/counterfeits **3**:1257
paint **4**:1935
toolmark examination **5**:2487–9
comparison question test (CQT) **2**:722–3
comparison scanning electron microscopy **5**:2489
compartment fires **3**:1118–20, 1178
compensating eyepieces **4**:1759
compensation cases, suicide **4**:2164
competency
civil commitment **2**:558–9
criminalistics certification **5**:2548–9, 2550
definition **2**:437, 450
experts **2**:960
genotyping for criminal trials **3**:1340
informed consent to medical treatment **4**:2211
juvenile defendants **2**:543–6
laboratory personnel **1**:8
to confess **2**:464
see also Miranda rights
to stand trial **2**:440–2, 456–63
court recommendations **2**:461–2
definition **2**:456
delusions **2**:744
evaluation **2**:457–9, 459–61
hallucinations **3**:1435
legal process **2**:456–7
mental status examination **4**:1738
psychopathy **4**:2195
psychopharmacology **4**:2202
report writing **2**:461
testimony **2**:461
traumatic brain injury **3**:1460
see also capacity
competent ignition sources **3**:1109
competitive market place development **2**:485
see also market-based forensic science
complementary colors **4**:2043
complementary metal oxide semiconductor (CMOS) **4**:2049
complete relocation, livor mortis **2**:705
complex microsatellites **2**:796
complex PTSD **2**:536
complex repeats (STRs) **5**:2356
complexity
ecological **2**:950
evidence **3**:1605

handwriting examination **3**:1443
language in cross-examination **2**:656–7, 660
compliance
analysis in ecology **2**:946
drug testing in urine **2**:860
false confessions **2**:589; **3**:1591, 1593
complications
rib fractures **1**:408
surgical **4**:1702, 1703
Composite International Diagnostic Interview (CIDI) **4**:1880
composite materials analysis **4**:1687
composite samples **5**:2281, 2384–5, 2385–6
compositing **5**:2261
composition (photographic) **2**:626–35
compound heterozygotes **4**:2031
compound microsatellites **2**:796
compound microscopes **4**:1758–61, 1791
see also microscopy
compound questions **2**:758
compound repeats, STRs **5**:2356
comprehension of Miranda rights (CMR) **2**:465
comprehension of Miranda rights-recognition (CMR-R) **2**:465
comprehension of Miranda vocabulary (CMV) **2**:465
compression systems, image processing **3**:1522
compression transfer bloodstain patterns **1**:385
compromised systems **1**:143–5
compulsion **2**:571–5; **4**:2191
behavior **2**:572–3; **5**:2317
defenses **2**:576–9
definition **2**:571
involuntariness **2**:571–2
legal applications **2**:574, 576–9
subjective experience **2**:573
see also civil commitment; insanity defense
computational fluid dynamics (CFD) **3**:1183, 1184
computed axial tomography (CAT) **5**:2234, 2235, 2236, 2238
computed tomography (CT) **4**:1871
computer-aided facial reconstructions **3**:1090
computer animation evidence **2**:579–82
computer forensics **2**:584–8
computer simulations **2**:579–82; **5**:2250–1
see also simulations
computer systems **3**:1368; **5**:2382–3, 2574–5
see also hardware; software
Computerised Assistance for Glass Evidence (CAGE) **3**:1358; **5**:2590, 2591
computerization *see* computer systems
Conan Doyle, Sir Arthur **5**:2380
concealed information tests (CITs) **2**:721, 725, 726
concealed object detection/retrieval methods **5**:2239
concentration (cognitive ability) **3**:1462; **4**:1866
concentration of soil materials **5**:2381
conceptual model, capacity assessment **2**:452–3
conchoidal glass fractures **4**:2005
conclusion scale, toolmarks **5**:2494
conclusion section, statements **2**:974
concurrent validity **4**:2200
concussion **1**:406; **3**:1460
condensed-phase explosions **2**:1029
condensed ring aromatics **3**:1148
condenser lenses **4**:1796
condensers **4**:1751, 1758, 1759, 1766
conditional probability **2**:970, 975; **4**:1812; **5**:2405–6
see also prosecutor's fallacy
conditioned tolerance **1**:124
conditioning contributors, DNA mixtures **4**:1839
conditioning paper for testing **4**:1971
condoms **2**:822; **4**:1931
conduct (legal provisions) **1**:282
conduct disorder (CD) **4**:1881
conduction **3**:1117–8
conductors, fire investigation **2**:929
cone-pattern fire plumes **3**:1120
cone photoreceptors **4**:2041
conference attendance, criminalistics training **5**:2547
confession evidence **2**:463–4, 464, 588–93; **3**:1590–3
see also false confessions
confidence
accuracy relationship **2**:1074
eyewitness identification **2**:1074; **4**:1711
malleability **2**:1074
confidence intervals **4**:1824–5; **5**:2288, 2678–9
confidentiality **5**:2336
custody/visitation evaluation **5**:2603
mental health patients **2**:885–6, 887–8, 889
mental status examination **4**:1739
patient confidentiality **2**:670, 885–6, 887–8, 889
rules of **2**:960
sexual assault examination **1**:236
confirmation bias **3**:1575–8
see also context effects; observer effects
confirmation testing
amphetamines **1**:138
benzodiazepines **1**:297
cocaine analysis **2**:567–8
opioids **4**:1902–3
stain identification **1**:314, 316, 317
systematic toxicological analysis **4**:2127
toxicology **2**:595–600; **4**:2127
urine drug tests **2**:862
confirmatory factor analysis (CFA) **4**:2199
confirmatory feedback, eyewitness testimony **4**:1711
confirmed identity, definition **1**:172
conflict-escaping behavior **4**:1982
conflicted disclosure of abuse **2**:539
confluence, livor mortis **2**:704
conformity assessment bodies (CABs) **4**:2220
congenital diseases **1**:194, **2**:479–80
coniferous trees **4**:1976
see also softwoods
Conium maculatum **4**:2061–2
Conjectures and Refutations (Popper) **3**:1093
conjoint therapy **4**:1983
Connelly v. Colorado, 479 U.S. 157 (1986) **2**:464
Connor v. Graham, 490 U.S. 386 (1989) **4**:2069–70
conscious intent **2**:577–9
consciousness research **2**:577–9
consecutively matching striae (CMS) **3**:1209
consensual sexual activity **1**:237, 238, 239

consensual sexual intercourse (CSI) **1**:238
consensus profiles **3**:1569–70, 1643
consent
  capacity
    elder abuse **2**:907–8
    medical treatment **2**:450–4
    research **2**:440
  civil commitment **5**:2452
  definition **2**:450
  family law/healthcare **3**:1375
  independent living assessment **2**:447–8
  sex offender treatment **5**:2336
  sexual assault cases **1**:236; **2**:660
  to medical treatment **2**:439–40, 450–4
  to research capacity **2**:440
  treatment refusal **5**:2584–7
  see also informed consent
conservatorships **2**:558; **3**:1371
  see also guardianships
consistency of handwriting **3**:1442
constitutional requirements, discovery **2**:779, 780
constitutional rights **4**:2071
construct validity **4**:2200
constructional tests **3**:1462
constructive errors, memory **4**:1710
Consumer Product Safety Commission (CPSC) **3**:1486
consumers
  expectation test **3**:1495
  misuse of products **3**:1490
  use of products **3**:1488–92
contact details, crime scene documents **2**:611
contact lenses **2**:827
contact range, shooting distance **5**:2351
contact shots **3**:1390–1
contact sites, sexual assault **1**:237–9
contact traces **2**:992–4; **4**:1951; **5**:2378–83
containers
  fire scene samples **3**:1129
  storage of specimens **4**:2104
contaminated fingers, blood **3**:1309
contamination
  CBRN agents **2**:503–5
  DNA extraction **2**:1064
  explosion debris **2**:1044, 1045
  'Keumdong No 5' oil spill **2**:948
  low copy number DNA tests **3**:1643
  low template DNA **3**:1568
  minimization strategies (bomb scenes) **1**:414–5
  mtDNA analysis samples **4**:1834
  see also anticontamination measures
context effects **3**:1575–8
  see also confirmation bias; observer effects
context–person interactions **2**:453
contextual risk factors **2**:914; **5**:2273
continuing training, criminalistics **5**:2546–8
continuity measures, packaging/transport **4**:1927
continuity of purpose, stalking **5**:2400
continuous performance testing **4**:1866
continuous stream inkjet printing **5**:2669–70
contractual capacity **2**:439
contrary opinions, experts **2**:664–5
contrast
  adjusting **4**:2039–40, 2054
  CCTV images **3**:1521
  crime scene photography **2**:627
  digital imaging **4**:2054
  enhancing **3**:1521
  evidence photography **4**:2039–40, 2054
filters **4**:2046
  light microscopy **4**:1768, 1770
contrecoup contusions **1**:406
contributors, DNA mixtures **4**:1838, 1839
control factor, elder abuse **2**:906
control region, mtDNA **4**:1833, 1835
control samples (*K*)
  diatoms **2**:754–5
  explosive residues **1**:415
  glass **4**:1929
  paint **4**:1935
  pollen **4**:1962–3, 1965
  soil **5**:2383
  see also exemplars; reference samples
controlled use treatment (addictions) **1**:19
controls (errors/error rates) **2**:955
contusions **1**:398–400, 406, 408; **2**:905
convection **3**:1117, 1119, 1178
convenience samples **3**:1581
Convention on International Trade in Endangered Species Fauna and Flora (CITES) **5**:2635–6
converging connections, Bayesian networks **1**:278
cooling conditions, algor mortis **2**:708, 709, 711, 714
coomassie blue **3**:1320
Coombs test **1**:346
coordination of forensic specialists **1**:414
coordinator log, crime scenes **2**:611–3
copper concentration cases **2**:952
copper tests **3**:1220
copper wiring **2**:925
copybook system of writing **3**:1437
copycat behavior, multiple murders **3**:1478
coronary artery anomalies **2**:479
coronary artery disease (CAD) **2**:469–70, **4**:1852
coronary artery dissection **2**:479
coronary artery tunneling **2**:477
coronary atherosclerotic plaque **2**:470
coronary bridging **2**:477
Coroner's inquest, Leskie case **2**:818–9
correction fluid obliterations **1**:133
correction ribbons, typewriters **5**:2665
correctional facilities, juveniles **3**:1613–4, 1615
  see also incarceration; jail populations; prisoners
corrective factors (CF), nomograms **2**:711, 712
correlations
  behavioral genomics **3**:1341
  fire models **3**:1181
corroboration, witness testimony **4**:1678
corrosion of materials **4**:1685
corruption, ecological cases **2**:952
cortical cells **3**:1418
Corymbia species **4**:1979
cost-effectiveness of CAI **2**:483–4, 485
costs/benefits analysis, victim decision making **2**:644, 645, 647
cotton, papermaking **4**:1978
cotton cloth **5**:2353
Cotton, Ronald **2**:1076

cough reflex suppression **4**:1899
counseling, elder abuse cases **2**:909
counterfeit documents **3**:1255–76; **5**:2673
countermeasures, deception detection **2**:726
countertransference, malingering **4**:1658
counting method, mtDNA **4**:1824, 1826, 1828
coup contusions **1**:406
The Courage to Heal (Bass and Davis) **4**:1713
courses *see* training
Court of Appeals
  General Electric v. Joiner **3**:1334
  Weisgram v. Marley **5**:2627
Court of Criminal Appeal, UK **2**:1000
court orders, parental alienation **4**:1983–4
court processing
  guardianships **3**:1377–9
  severe mental illness **2**:651
  youths as adults **4**:1879
  see also courts
court reports
  expert opinions **5**:2268–70
  identification evidence **3**:1509–10
  courts
  chancery **1**:282, 284
  communication of statistical evidence **5**:2401–8
  competency to stand trial recommendations **2**:461–2
  confession evidence **2**:592–3
  context (human factors experts) **3**:1491–2
  external compulsion treatment **2**:571
  guardianship alternatives **3**:1375, 1376
  illicit drug evidence **2**:857–8
  mtDNA information **4**:1836
  questions (explosion debris) **2**:1044–5
  specialized courts **5**:2450–1
  truth serum admissibility **2**:736
  see also court processing; judiciary
covert security features, documents **3**:1265
cowitness suggestion **2**:1067
CPA (cyproterone acetate) **5**:2334
CPIA *see* Criminal Procedure and Investigations Act
CPK (creatine phosphokinase) **4**:2205
CPR *see* Civil Procedure Rules
CPSC (Consumer Product Safety Commission) **3**:1486
CQT (comparison question test) **2**:722–3
crack cocaine **2**:563, 848; **4**:1912; **5**:2434
cradle-to-grave DNA databases **2**:681
craniofacial superimposition **3**:1515–6
cranium
  age determination **1**:179, 181
  blunt trauma **5**:2561
  cranial cavity injuries **3**:1534–5
  facial reconstruction **3**:1087–8
  race/ancestry determination **1**:166
  sex determination **1**:156; **5**:2329–30
crash data **1**:53–4, 55–6, 57; **5**:2572
  see also road traffic accidents
crash data recorders (CDR) **5**:2251
crash tests **5**:2569–74
craters, explosion scenes **2**:1029–30
crazing of glass myth **1**:208, 209, 211–2
creases
  embryonic formation **3**:1328, 1330
  hands/feet **3**:1323
creatine phosphokinase (CPK) **4**:2205
creatinine **1**:318; **2**:863, 864, 866; **4**:2082–3, 2084
credentials of experts **2**:1013–8; **4**:1862
  see also qualifications
credibility
  cross-examination **2**:656, 657, 660
  older witnesses **2**:918–9
creep fracture, plastics **4**:1687
Crime in the United States (FBI) **1**:28
crime investigation role **2**:495
  see also crime scene investigator
crime laboratories **1**:1–8
  see also laboratories
crime propositions **3**:1581
crime range, DNA databases **2**:681
crime scene coordinator (CSC) **2**:621
crime scene investigator (CSI) **2**:615
  burglary scenes **2**:616–7
  casting evidence **2**:963–7
  documentation **2**:602–14
  major crime scenes **2**:621, 622–3
  outside scenes **2**:619
  quality assurance measures **2**:625
  vehicle crime scenes **2**:617–9
  volume crime scenes **2**:615–6
  see also crime scenes; scene investigation
crime scene management/manager (CSM) **2**:611–3, 619–25
crime scenes
  3D modeling **3**:1628; **5**:2258–60, 2261
  botany **1**:425
  CBRN scenes **2**:503–5
  DNA phenotype inference **4**:2021–34
  documentation of evidence **2**:602–14
  entomology **2**:939, 943–4
  explosions **2**:1030
  fingermark detection **3**:1294–5
  footwear impressions **3**:1249
investigation **2**:614–9; **4**:1674–**5**:2480–1
  see also crime scene investigator
  modeling **5**:2258–60, 2261
  movements (bomb scenes) **1**:413
  odontology **4**:1889
  photography **2**:625–43; **4**:1669
  pollen evidence **4**:1963, 1967
  profiles for Y-chromosomal STR analysis **5**:2680
  reconstructions (3D) **5**:2258–60, 2261
  reports **2**:607
  shooting scene examination **3**:1219–25
  stain typing **4**:1834
  trace evidence sampling **5**:2291
  see also scene examination; scene investigation
crime victims
  decision to call police **2**:643–7
  defining crime **2**:644–5
  inaction **2**:647
  see also victim...
crimes against humanity **4**:1676, 1677
Criminal Appeal Act, 1907 (UK) **2**:1000
criminal behavior
  diversity **4**:1879
  genetics **3**:1335–6

criminal defenses, compulsion **2**:576–9
criminal injury diagnosis **5**:2237
Criminal Interrogations and Confessions **2**:590
Criminal Justice Act (CJA), UK **2**:782, 1000
criminal law
  capacity assessment **2**:440–3
  capacity to proceed with trial **2**:440–2
  dangerousness **2**:670–1
  execution capacity **2**:442–3
  insanity defense waiver capacity **2**:442
  legal counsel waiver capacity **2**:442
  mentally ill accused **4**:1717–24
  sentencing capacity **2**:442
  see also criminal proceedings; criminal trials
Criminal Lunatics Act (1800) **3**:1553
criminal mass graves **4**:1674, 1676–8, 1678
  see also mass graves
Criminal Procedure and Investigations Act (CPIA), UK **2**:782
Criminal Procedure Rule 24 **2**:1003
criminal proceedings
  discovery
    motions **2**:774–5
    of expert finding **2**:782–3
    United States **2**:778–81
  see also criminal law
criminal profiling **5**:2455
  see also behavioral profiles; psychological profiling
criminal responsibility
  assessment **3**:1554–5
  automatism defense **1**:253–6
  compulsion defenses **2**:577–9
  defendant's account of offense **3**:1554–5
  dissociative disorders **2**:788
  forensic interviews **3**:1554
  insanity defense **5**:2443
  neuroscientific research **2**:577–9
  sanity opinion formation **3**:1555–6
  see also culpability
criminal trials **2**:969–71; **3**:1338–40
  see also criminal law; criminal proceedings; trials
criminalistics
  gunshot wounds **3**:1386
  organisational accreditation **1**:10–6
  toolmark examination **5**:2485–94
  training/certification **5**:2545–54
  wildlife forensics **5**:2637–8
criminalization
  definition **2**:649
  mental illness **2**:649–53; **4**:1718
  substance abuse **5**:2414
criminological model of malingering **4**:1658
critical difference, storage stability **4**:2104
critical illumination **4**:1758
critical incident teams **4**:2071–5
critical radiant heat flux **3**:1105
crocodiling effect *see* alligatoring effect
cross-contamination, eyewitness reports **2**:1067
crossdirection (CD), paper **4**:1971–2
cross-examination
  accuracy **2**:657–8, 660
  adults **2**:660
  children **2**:656–60
  definition **2**:656
  expert malpractice **4**:1664
  expert opinion evidence **2**:662–5; **3**:1505
  leading questions **2**:758
  learned treatise evidence **3**:1624
  qualifications of experts **2**:663–4, 759
  testimony impact **2**:656–62
cross-linking acrylamide gels **1**:349
cross-racial identification **2**:1077
cross-reactivity, immunoassays **2**:862; **4**:2125; **5**:2512
crossreference documents **2**:613
cross-sectional hair examination **3**:1409
cross-tolerance, alcohol **1**:124
cross-trenching mass graves **4**:1676
crossed polar particle characterization **4**:1778–83
crossing-over (DNA) **2**:796
crossover, thermoionic gun SEM **4**:1795
Crow–Glassman scale (CGS) **3**:1532
Crowe, Michael **2**:589
Crown Prosecution Service **2**:1006
crude oil **3**:1149
cruise control systems **1**:57
crush asphyxia **1**:227
crushing abrasions **1**:397
cryptomnesia **4**:1711
crystal meth *see* methamphetamine
crystal violet *see* gentian violet
crystals
  compounds **4**:1744–5, 1749
  fusion methods **4**:1783
  interference figures **4**:1783
  particle forms **4**:2002
  perpendicular plane polarization **4**:1779
  refractive index **4**:1770, 1775–7
CSA *see* childhood sexual abuse
CSAAS *see* child sexual abuse accommodation syndrome
CSC *see* crime scene coordinator
CSF *see* cerebrospinal fluid
CSFS (Canadian Society of Forensic Science) **5**:2622
CSI *see* consensual sexual intercourse; crime scene investigator
CSM *see* crime scene management/manager
CT scanning *see* computed axial tomography
cued recall **2**:554, 917
culpability **2**:543–6, 547; **3**:1608, 1609
  see also blameworthiness
cultural context, elder abuse **2**:914
cultural modifications **1**:170–1
cultural violence **1**:25
cumulative drug detection **5**:2426
Curran, J.M. **5**:2591
curriculum vitae (CV) **2**:922, 923
curves function, Photoshop **4**:2054
curvilinear distortion **4**:2038
custodial chain *see* chain of custody/possession of tangible evidence
custodial sources, footwear **3**:1241
custody evaluation **5**:2602–11
customer requirements, CAI **2**:489–90
customized crash tests **5**:2571–4

cut marks, skeletal remains **5**:2561–2
'cut and paste' document alterations **1**:131
cut wounds **5**:2649–54
  see also incised wounds
cuticle of hair **3**:1404, 1406–7, 1418
cutoffs
  applications **2**:598
  drug testing **2**:861; **3**:1428; **4**:1908; **5**:2506, 2507
  immunoassays **4**:2125
cutting agents, illicit drugs **2**:853–4
cutting back treatment (addiction) **1**:19
CV *see* curriculum vitae
CWAs *see* chemical warfare agents
CWC *see* Chemical Weapons Convention
cyanide (CN)
  fire victims **3**:1535
  poisoning **1**:233; **4**:2101, 2133
  stability during storage **4**:2105
cyanoacrylate fuming (CAF) **3**:1300–2, 1305, 1309, 1319
cyclic stress, materials **4**:1683
cycloalkanes **3**:1148, 1157, 1162
CYP genes
  clinical applications **4**:2015, 2017
  cytochrome P450 monooxygenases **1**:63
  forensic applications **4**:2018
  race **4**:2014
  tests **4**:2014
cyproterone acetate (CPA) **5**:2334
cytisine **4**:2063
cytochrome P450 monooxygenases **1**:63
  see also CYP genes

d-connection, Bayesian networks **1**:278
D-loop, mtDNA **4**:1833
d-separation, Bayesian networks **1**:277, 278
DA (dopamine) **2**:563
DAB (diaminobenzidine) **3**:1320
dactyloscopic points *see* minutiae
DAD *see* diffuse alveolar damage; diode-array detectors; drowning-associated diatoms; photodiode array
DADP *see* diacetone diperoxide
DAGs (directed acyclic graphs) **1**:276
Dahmer, Jeffrey **5**:2318
DAI (diffuse axonal injury) **1**:407
daisy wheel typewriters **5**:2664–5
Dallagher, Mark **2**:891
Dallas criteria, myocarditis **2**:475
damage to fiber evidence **2**:995
damages (malpractice cases) **4**:1664, 1692
Damasio, Antonio **2**:578–9
damp items, packaging **4**:1928, 1929
Dangerous Offender (DO)
  legislation (Canada) **2**:670–1
  sexual offender distinction **2**:674
dangerous severe personality disorder (DSPD) legislation **2**:669
dangerousness **2**:670, 886–9
  civil commitment **2**:668–70; **5**:2578
  civil law **2**:668–70
  criminal law **2**:670–1
  death penalty **2**:671
  definition **2**:668
  delusions **2**:744
  genotyping for criminal trials **3**:1340
  insanity defense acquittees **5**:2579
  mandated treatment **5**:2577–9, 2585
  risk assessment **2**:667–77
  sexual offenders **2**:668–9, 671
  violence by suspects **4**:2070
  violence prediction **2**:671–4
  see also duty to warn
dark field microscopy **4**:1760
data analysis
  crime victim decision making **2**:644
  legal/scientific method comparison **5**:2296–7
  pollen **4**:1956–9
data collector, total station device **2**:580
data interpretation *see* data analysis
data retrieval, Macintosh computers **2**:585–6
data sources
  air bag systems **1**:53–4, 55
  biomedical engineering **5**:2572
  DNA mixture interpretation **4**:1839–40
  medical malpractice **4**:1692–704
  mobile phones **3**:1365–8
  network analysis **1**:141–2, 147, 148
  odontological identification **4**:1891–2
  vehicle speed **5**:2250–1
database match probability, definition **2**:835
  see also match probabilities
databases **2**:677–84
  automotive paints **4**:1939, 1946
  barefoot impressions **3**:1246–7
  footwear impressions **3**:1249–51
  household paints **4**:1949–50
  mtDNA evidence **4**:1827, 1828–30
  paint evidence **4**:1939, 1946, 1949–50, 1952
  size issues **4**:1830
  soil **5**:2382–3
  sole patterns **3**:1240–3, 1254
  STRs **5**:2357–9, 2372–3
  toxicology research **5**:2508–9
  see also DNA databases
dating techniques **1**:418–21; **2**:684–90; **3**:1542; **5**:2231–3
see also age determination; ink analysis; line intersection (documents); paper, analysis
  Datura stramonium **4**:2062
  Daubert standard
  child sexual abuse accommodation syndrome **2**:540
  homosexual panic defense **3**:1481–2
  premenstrual syndrome **4**:2151
  psychological autopsy **4**:2166
Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993) **2**:692–6, 1001–2, 1004; **3**:1289, 1290, 1331, 1334, **5**:2627, 2628
  error rates **2**:956, 957
  expert credentials **2**:1015, 1017
  falsifiability theory **3**:1093, 1094
  FRE 702 **3**:1095
  general acceptance test **3**:1333
  ipse dixit evidence **3**:1600
  Kumho Tire ruling **3**:1619, 1620
  neuropsychological evidence **4**:1862

Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993) *(cont)*
peer reviews **4**:2011
scientific evidence admissibility **1**:271
scientific expert testimony **1**:285
see also Daubert standard
DB (defense-biased) jurors **5**:2308
'dead-in-bed syndrome' **4**:1851
deadly force **4**:2070–1, 2073
deadly nightshade plant **4**:2060–1
death
agonal period **2**:697
changes occurring after **4**:2096–102
classification **4**:2162
opioid use **4**:1900–2
processes **2**:697–716
radiology of deceased **5**:2235–6
scenes (autoerotic deaths) **1**:246–7
time of **2**:697–716; **5**:2466–79
see also cause of death; fatalities; lethal cases; manner of death; postmortem...
death cap mushroom **4**:2060
death penalty **2**:442–3
abolition **2**:717–8, 719
aging inmates **2**:718–9
dangerous offenders **2**:671
'evil' inmate cases **2**:982
fingerprint comparison cases **3**:1289
inmate age **2**:717–9
litigation **4**:1818–23
mental retardation **2**:718; **4**:1730–5
weight of expert evidence **2**:1016
young offenders **3**:1608
debossing documents **3**:1273
debris *see* explosion debris; fire, debris; residues
decay, neutron activation **1**:150
see also decomposition of remains; putrefaction
deceased checklist document **2**:612–3
deceit, experts **2**:960
deception detection **2**:720–3, 724–7, 728–37; **3**:1331–2
see also lie detection
decision-making
adolescents **3**:1609, 1610
Bayesian networks **1**:279
capacity
assessment **2**:448–9
competency to stand trial **2**:458–9
elder abuse **2**:907–8
independent living assessment **2**:445
medical treatment **2**:452–4
crime victims notifying police **2**:643–7
financial **3**:1375–6
juries **3**:1602–6; **4**:1819
declarants, hearsay evidence **3**:1465, 1467
declared proficiency tests **2**:956
declining health-type spousal homicide–suicide **5**:2418
decompensated retention, kidney **4**:2082
decomposition of remains
entomology **2**:936, 940, 943–4
fingerprints **3**:1512–3
mass graves **4**:1674, 1675
PMI determination **4**:2089
time of death determination **5**:2474–8
see also putrefaction
decontamination methods, CBRN agents **2**:505
DECTs (digital enhanced cordless telecommunications) **3**:1360
decubitus ulcers **2**:905
deductive identification **5**:2237–8
deep venous/vein thrombosis (DVT) **4**:1846
deer hair **3**:1407
Deese-Roediger-McDermot (DRM) procedure **2**:1069
defeating urine drug tests **2**:862–5
defects
materials **4**:1683
products **3**:1484, 1487–8, 1495
see also flaws
defendants
account of offense **3**:1554–5
physical appearance **2**:980
see also behavioral profiles; offender...; psychological profiling
defense
diminished capacity **3**:1339
discovery **2**:782–3
motions **2**:774, 775
United States **2**:779, 780
environmental prosecutions **2**:946
profile testimony **4**:2159
proposition **2**:970, 971–2
'ultimate issue' opinion testimony **5**:2590
see also insanity defense; not guilty by reason of insanity
defense (attorney) fallacy **3**:1584; **5**:2405
defense-biased (DB) jurors **5**:2308
defense mechanisms **5**:2410–1
defense wounds **5**:2651, 2655–8
deflagrations, definition **2**:1023
deflection of projectiles **3**:1384
deformation of materials **4**:1681–2
degenerate oligonucleotide primed (DOP) DNA quantities **5**:2629
degenerative diseases **1**:196
degradation
chemical warfare agents **2**:526–7
DNA samples **2**:816–9
human remains **4**:2089, 2091
pollen **4**:1957, 1958, 1959
postmortem toxicology **4**:2105–6, 2107
see also postmortem changes; postmortem redistribution
degree-days/degree-hours (fly development) **2**:942
degree of bloodspatter **1**:363–5
degree of similarity *see* comparison
degree programs **2**:898–9, 901; **5**:2545–6
dehumanization **1**:30
dehydration **4**:2084
deinstitutionalization **2**:649
see also hospitalization
Delaney case **5**:2434
delay discounting (reward/satisfaction) **2**:572
delayed cross-examination **2**:657, 658–9
delayed disclosure of abuse **2**:539
delayed discovery doctrine **4**:1713

delayed effects
Camelford poisoning case **2**:948–9, 949
gunshot wounds **3**:1386
deliberations of juries **3**:1607, 1608
delinquency **4**:1878, 1985–6
see also Northwestern Juvenile Project
deltas, friction ridge skin **3**:1323
delusions **2**:459, 733, 741–4; **5**:2274
see also insanity defense
demarcation lines, fire myths **1**:209, 214–6
dementia **2**:445–6, 718–9, 915; **4**:2189
see also Alzheimer's disease
demethylation process, inks **3**:1545
demographic factors
firesetting **3**:1227
guardianships **3**:1373
juvenile rearrests **3**:1613
multiple murders **3**:1478–9
sentencing **5**:2306–11
demonstrative evidence **2**:745–7
dendrochronology **5**:2644
see also growth ring analysis
dendrology **5**:2641–2
see also wood
denial syndrome **5**:2410–1
Denny-Shaffer case **5**:2435
density of paper **4**:1973
dental age **1**:188, 189
dental identification techniques **5**:2238
see also dental remains
dental records **2**:767, 770
dental remains
age determination **1**:179, 180, 182
anomalies/pathology **1**:194
race/ancestry determination **1**:192–3
sex determination **5**:2328
see also odontology; teeth
dental stone **2**:963, 964–5
dental superimposition **4**:1893
deoxyribonucleic acid *see* DNA...
dependence (drugs)
alcohol **1**:122–4
benzodiazepines **1**:296
opioids **4**:1899
prevalence **1**:19
see also addictions
dependency **2**:906–7, 915; **4**:1987
depersonalization disorder **2**:784
depolymerization **3**:1110
depositions **2**:772–4
depressants **1**:108, 117, 291, 293; **2**:868–9, 871
see also alcohol
depressed skull fractures **1**:403
depression **1**:111; **2**:531, 547; **4**:2215
depth of char myth **1**:209, 213
depth of field (photography) **2**:639–41
dermal...
nitrate test **3**:1191
papillae **3**:1324
see also skin
dermestid beetles **2**:940
dermis, friction ridge skin **3**:1323, 1325, 1326
description of body (autopsies) **1**:260
descriptive methods *see* morphology
desensitization **1**:30
design defects, products **3**:1484, 1487
desired combustion **3**:1175
Deskovic, Jeffrey **2**:591–2
desmosomes **3**:1325
DESNOS (disorder of extreme stress not otherwise specified) **2**:536
destruction of CBRN agents **2**:505
destructive restoration techniques **5**:2326–7
detailed characterization, soil **5**:2385–6
detection methods
biological agents **1**:303–6
chemical warfare agents **2**:511–22
concealed/embedded objects **5**:2239
DNA/fingerprint interaction **3**:1318–21
drug impaired driving **2**:881–2
drug screening **5**:2524–5
elder abuse **2**:904–6
fingermarks **3**:1292–311, 1320
foot impressions **3**:1254
malingered deficits **4**:1868–9
see also deception detection; limit of detection
detection period, cannabis testing **2**:866
determining charges, mass graves **4**:1676–8, 1678
deterministic fire models **3**:1177–8, 1181–4, 1186
detonations, definition **2**:1023
detonator examinations **2**:1035
deuterated triglycine sulfate (DTGS) detector **4**:1937
deuterium **4**:2222
developmental aspects
aging **1**:160
juveniles **3**:1608–12, 1613–4
mitigation **4**:1819–20
developmental disorders **2**:547
developmental immaturity **2**:543–4
developmental noise **3**:1330
developmental risk factors **5**:2273
deviant behavior **4**:2187
deviant fantasies **5**:2315, 2321
dexamphetamines *see* amphetamines
DFC *see* drug-facilitated crime
DFO *see* diazafluorenone
DFSA *see* drug-facilitated sexual assault
DHCP (Dynamic Host Configuration Protocol) **1**:146
diabetes mellitus (DM) **1**:112, **4**:1851–2, 2076–81, 2086
diabetic coma **4**:2076, 2077, 2086
diacetone diperoxide (DADP) **2**:1037, 1041, 1047–8
diagnosis
G x E interactions **3**:1337–8
genotype designation **3**:1337
kidney function disturbances **4**:2084
per exclusionem **4**:2086, 2087
radiology use **5**:2235–7
strangulation **1**:232
subdural hemorrhage **5**:2345
vitality (fire victims) **3**:1535–6
water/electrolyte imbalances **4**:2085
Diagnostic and Statistical Manual of Mental Disorders (DSM) **2**:651
NJP study **4**:1880

Diagnostic and Statistical Manual of Mental Disorders (DSM) (cont)
PTSD definition **5**:2436, 2437
Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-R) **4**:1657–8
Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) **4**:2187–8; **5**:2443
Diagnostic and Statistical Manual of Mental Disorders (DSM-V) **4**:2190–1
Diagnostic Interview Schedule (DISC) **4**:1880
diagnostic reasoning, Bayesian networks **1**:278
diagnostic trouble codes (DTCs) **1**:55
diaminobenzidine (DAB) **3**:1320
Diamond, Shari **3**:1604
diaphanoscopy **1**:400
diaphragms **4**:1751, 1752
see also apertures
diastatic skull fractures **1**:402
diatomic gases **3**:1107
diatomological maps **2**:755
diatoms **1**:428–9, **2**:748–56
classification **2**:749
definition **2**:748
extraction methods **2**:753–4
fate inside body **2**:751
identification techniques **2**:754–5
particle forms **4**:2002
skeletal remains analysis **5**:2563
structure **2**:749
tests **2**:751–2
diazafluorenone (DFO) **3**:1297–8, 1310, 1319
diazepam **1**:294, 295, 297
dichroic mirror (DM) **4**:1786
dichroism **4**:1778
DID *see* dissociative identity disorder
diesel fuel identification **3**:1157–8, 1160
diethyl ether **3**:1141
differences *see* differentiation
differential diagnosis, subdural hemorrhage **5**:2345
differential extraction
DNA **2**:1061
semen **2**:757
differential scanning calorimeter (DSC) **2**:1036
differential shedding, fibers **3**:1097, 1098
differential stains **4**:1979
differentiation
cat vs dog hairs **3**:1409–10
facial comparison **3**:1083, 1084
diffraction **4**:1752, 1767
diffuse alveolar damage (DAD) **4**:1846
diffuse axonal injury (DAI) **1**:407
diffuse fuel explosions **2**:1023
diffuse reflectance infrared Fourier transform (DRIFT) **5**:2385
diffuse reflection
fingermarks in blood **3**:1310
photographic enhancement methods **4**:2050–1
diffusion flames **3**:1116
digital cameras **2**:633, 636, 639–41; **4**:2049
see also digital imaging; digital photography
digital compression, images **3**:1522
digital data
biomedical engineering **5**:2572
vehicle speed **5**:2250–1
digital enhanced cordless telecommunications (DECTs) **3**:1360
digital evidence **2**:584–8; **5**:2639
digital filters **4**:2041, 2049
digital fingerprints **2**:587, 588
digital imaging **4**:2036
color synthesis models **4**:2044
evidence enhancement techniques **4**:2053–5
fingerprint spoofing **1**:330, 332
hand vein pattern spoofing **1**:337
height estimations **3**:1624–32
lens distortion correction **4**:2038
see also digital cameras; images, processing
digital photography **4**:2036
see also digital cameras
digital sensors **4**:2049
digital video systems **2**:584, 587
digital watermarks **3**:1275
dihydrocodeine **4**:1914
dilated cardiomyopathy **2**:473
dimensional integrity (photography) **4**:2037–8
dimethylnaphthalenes **3**:1154
diminished capacity defense **3**:1339, 1435; **4**:2150–1, 2152; **5**:2446–7
diminished responsibility **1**:283–4
dimorphism **1**:155–6
diode-array detectors (DAD) **2**:990, **4**:2126, 2128, 2130
dip pens **5**:2661
diplotypes **4**:2029
dipping application, paint **4**:1932
direct causality *see* causality
direct examination **2**:658, 662, 758–60; **3**:1505
direct methods
ink volatile components **3**:1544
phenotype inference **4**:2022, 2029–33
direct reflection **4**:2050
direct sensing fingerprint recognition **1**:325
direct thermal transfer printing **5**:2672, 2673
directed acyclic graphs (DAGs) **1**:276
directional sensitivity, air bag systems **1**:53
directionality, bloodstains **1**:363, 366
Dirichlet sampling formula **4**:1812, 1813, 1815, 1816
dirt samples, pollen evidence **4**:1963, 1966
'dirty bombs' **4**:1884
dirty man/clean man approach **2**:504
disability **2**:743–4; **3**:1434; **4**:1876; **5**:2303
see also mental retardation
disaster mental health **2**:760–3
disaster victim identification (DVI) **2**:764–72
anthropology **1**:153
bomb scene management **1**:414
DNA sources **2**:826
Identifiler™ system **3**:1518
mini-STRs **4**:1806
odontology **4**:1893
United Kingdom **2**:765–7, 771
DISC (Diagnostic Interview Schedule) **4**:1880
disciple/ideological murders **3**:1476
disciplinary actions, school violence **5**:2456
disciplining children **4**:1985
disclosure **2**:452, 539, 552
see also discovery

discoloration of ink **3**:1545
discontinuous memories **2**:786
discovery
  draft reports **5**:2270
  expert findings **2**:772–4, 774–5, 775–8, 778–81
  motions **2**:774–5
  see also disclosure
discrete data **5**:2284–7
discriminant function analysis **5**:2330
discriminating power (DP) **3**:1581; **4**:1943
discrimination
  epilepsy **5**:2303
  racism **5**:2306
disease course (addictions) **1**:19
diseases
  drug concentration effects **4**:2094
  drug detection interpretation **4**:2117, 2119
  markers **1**:168–70, 194–6
  opioid use **4**:1902
  traffic fatality prevention **5**:2542
  see also individual diseases; medical disorders
disguised handwriting **3**:1444, 1456
disguised signatures **3**:1445, 1447
Disk Arbitration, Macintosh computers **2**:586
dismemberment **5**:2658–9
disorder of extreme stress not otherwise specified (DESNOS) **2**:536
disorders, use of term **5**:2431
  see also individual disorders; syndromes
dispersed-phase explosions **2**:1030
dispersion of light **4**:1765
dispersion staining **4**:1771–2
disposable gloves **2**:615
disposal of CBRN agents **2**:505
dispositional component, capacity assessment **2**:453
dispositional risk factors **5**:2273
dispositional tolerance **1**:74, 123
disputes
  opinion evidence **2**:760
  workplace threats **5**:2464
dissecting microscopes *see* stereomicroscopy
dissection of body (autopsies) **1**:256, 259
dissocial personality disorder *see* psychopathy
dissociation **2**:531; **5**:2243, 2245, 2247
  see also dissociative disorders
dissociative amnesia **2**:784, 785–7, 788
dissociative disorders **2**:784–9
  see also dissociation
dissociative fugue **2**:784
dissociative identity disorder (DID) **2**:784–6, 788–9
dissonance, cognitive **4**:1982
distal region, hands/feet **3**:1322
distance determination *see* range determination
distant-range gunshots **3**:1393; **5**:2351
distillates
  analysis **3**:1143
  identification criteria **3**:1150, 1155–61
  source identity **3**:1163–5
distortion
  friction ridges **3**:1283
  photographic images **4**:2037–8
distributed air bag systems **1**:52, 55
distribution of drugs *see* drug distribution
disulfiram **1**:114
dithiothreitol (DTT) **2**:804
diverging connections, Bayesian networks **1**:277
diversion programs **4**:1721–3
diversionary devices, police use **4**:2073
diversity
  criminal behavior **4**:1879
  naturally occurring soils **5**:2379
  racial/ethnic **4**:1879
divorce, child custody **5**:2602–3
DM *see* diabetes mellitus; dichroic mirror
DNA/DNA evidence **2**:792–9
  autoradiography **1**:262
  Bayesian networks **1**:279
  botany and **1**:424, 425, 429
  ceiling principle **2**:497–8
  cellular biology **2**:792–4
  confirmation bias **3**:1576, 1577
  crime scene phenotype inference **4**:2021–34
  degraded samples **2**:816–9
  disaster victim identification **2**:767–71
  documentation **2**:602, 610
  earprint interpretation **2**:891
  evaluation **5**:2591–2
  extraction **2**:757, 803–5, 1060–4; **4**:2219
  fingermarks in blood **3**:1309
  fingerprints **3**:1290, 1318–21; **5**:2555, 2556
  footwear analysis **3**:1247
  genetics **2**:794–8
  Green River murders **4**:1940
  hair **3**:1420, 1424
  handwriting **3**:1453, 1454
  Hardy–Weinberg equilibrium **3**:1458
  HLA-DQα system **2**:842–3
  Identifiler™ system **3**:1518
  increased cycle number samples **5**:2628
  individualization **3**:1509
  IQ™ extraction **2**:1062
  lineups **2**:1072
  low copy number **5**:2268, 2628
  low template DNA **3**:1566–73
  matrix/matrix files **4**:1688–9
  microsatellites **4**:1749–50
  mini-STRs **4**:1804–9
  mitochondrial DNA interpretation **4**:1823–32
  mixtures **4**:1838–41, 2008; **5**:2680
  molecular biology **2**:798–9
  paraffin-embedded tissue **3**:1469
  peak heights **3**:1567, 1571; **4**:2007–8; **5**:2591
  pharmacogenomic principles **4**:2013
  plant analysis **1**:429
  polymorphism **4**:2075
  powder brush transfer **3**:1320
  PowerPlex systems **4**:2149
  primary transfer **5**:2556
  primer extension preamplification of samples **5**:2629–30
  profile matching **5**:2555, 2556, 2628–33
  QiaAmp® extraction **4**:2219
  quantification **2**:805–6
  reference samples **1**:320
  sampling **1**:320; **2**:816–9
  secondary transfer **5**:2555, 2556
  sex determination **5**:2328

DNA/DNA evidence *(cont)*
shooting scene investigation **3**:1222
sources **2**:821–8
species determination **1**:154; **5**:2395, 2396
stain identification **1**:314–5, 317, 318
starting templates **5**:2633
STR profiles **5**:2365–76, 2628
submission forms **2**:610
trace DNA **2**:823
transfer **5**:2537, 2538, 2555–6
ultrasensitive tests **3**:1643
variable number tandem repeats **5**:2595–6
vehicles **2**:618, 619
violence predisposition **3**:1336
virus histology **3**:1470
web resources **5**:2626
whole genome amplification **5**:2628–33
wildlife forensics **5**:2636–7
see also DNA profiling; genomics; mitochondrial DNA; nuclear DNA; pharmacogenetics
DNA databases **2**:677–84, 831–8; **5**:2361, 2680
cold case review **2**:677, 682–3
evidence associated with **2**:831–8
intelligence use **2**:681–2
investigative use **2**:681–2
match statistics **2**:832–7, 837, 838
mitochondrial DNA **4**:1828–30
operational impact **2**:680–1
postconviction testing **2**:682–3
socio-legal issues **2**:683–4
DNA evidence *see* DNA/DNA evidence
DNA fingerprinting **2**:800
see also DNA profiling
DNA markers (alternative) **2**:802–3
DNA polymerases **2**:954
DNA profiling **2**:792, 799
allelic attribution **1**:126–8
animal hair **3**:1403, 1411
history **2**:800–1
interpretation **2**:806–8
low amounts **3**:1639–45
low template analysis **2**:808–10, **3**:1567, 1569–72
matching evidence **5**:2555
missing persons calculations **4**:1810–7
mitochondrial DNA **4**:1833–7
paradigms **3**:1643
parentage calculations **4**:1810–7
polymorphisms comparison **1**:339–40, 356
probability approach **2**:972, 973, 975
quality **3**:1567
STR interpretation **5**:2365–76
wood **5**:2641, 2642, 2645
see also DNA/DNA evidence; DNA fingerprinting
DNA sequence **2**:798–9
DNA testing *see* DNA/DNA evidence
DNA typing, overview **2**:800–10
DNAMIX program **4**:1840; **5**:2592
DNAPrint genomics laboratory **4**:2030, 2033
DNAWITNESS™ **4**:2033
DNS (domain name system) **1**:146
DO *see* Dangerous Offender
doctorate programs **2**:899, 901
doctors
patient confidentiality **2**:670, 885–6, 887–8, 889
reporting battered child syndrome **1**:268
serial murderers **5**:2315
see also mental health professionals; physicians
documentary evidence
authentication **2**:840–1
best evidence rule **1**:298–9
dating **2**:684–90; **5**:2232
examination **1**:128–33; **4**:2036–56
forgeries/counterfeits **3**:1255–76
foundation testimony **3**:1277
handwriting/signatures **3**:1436–48, 1451–7
interpretation **3**:1557–61
line intersection **3**:1594–8
photographic/optical examination **4**:2036–56
printing devices **5**:2660–1, 2663–74
radiocarbon dating **5**:2232
security features **3**:1265–73
wildlife forensics **5**:2638
writing instruments **5**:2660–3
see also ink. . .; paper...
documentation procedures
bloodstain patterns **1**:373–4, 386–8
crime scene management **2**:622–3
of evidence **2**:602–14
explosion scenes **2**:1024
fire scenes **3**:1124–6, 1126, 1136
gunshot wounds **3**:1397–8
health care providers **5**:2571–2
image processing **3**:1526
laboratory accreditation **1**:11, 12, 13, 14
mass graves **4**:1676
mental health practitioners **5**:2598, 2600–1
photography **1**:373–4, 386; **2**:603, 608, 625–43
post scene **2**:608–10
principles **5**:2601
radiocarbon dating **1**:421
sexual assault cases **1**:237
see also reports/reporting
dog hairs **3**:1404, 1407, 1409–10, 1410
dogs **1**:204, **3**:1497
Doheny and Adams v. R [1996] EWCA Crim 728 **2**:975
doll aids in child interviews **2**:554
domain name system (DNS) spoofing **1**:146
domestic abuse **5**:2411
domestic violence **1**:272–5; **5**:2415
see also battered woman syndrome; spousal...
Donaldson v. O'Connor, 422 U.S. 563 (1975) **5**:2581
Donnelly and Balding DNA databases **2**:835, 837
door hinges, fire scenes **3**:1133
DOP *see* degenerate oligonucleotide primed DNA quantities
dopamine (DA) **2**:563
doping **3**:1430; **5**:2507
dose dependency, poisons **4**:2120
doses, drug pharmacokinetics **4**:2116–7
dossiers **2**:562, 663, 1006, 1007
dot matrix printers **5**:2665–7
double-blind lineups **2**:1074, 1077
double-blind verification **4**:2010
double dissociation **5**:2247

'double file' bloodstain pattern **1**:381
doughnut patterns, fire **1**:216
Down's syndrome **3**:1337, 1341
Downs v. United States, 522 F.2d 990 (6th Cir., 1975) **4**:2074
downward comparisons **2**:646–7
DP *see* discriminating power
DQA1 locus (human leukocyte antigen) **2**:842
draft print quality **5**:2665
draft reports **5**:2270
DRIFT *see* diffuse reflectance infrared Fourier transform
drilling (serial number obliteration) **5**:2325
dripped blood spatter patterns **1**:369–70
driver impairment **2**:877–83
  alcohol intoxication **1**:100
  alcohol–drug interactions **1**:113
  benzodiazepines **1**:296
  blood alcohol concentration **1**:58, 65, 67
  breath-alcohol analysis **1**:96, 97
  cannabis **2**:434, 436
  oral fluid **4**:1918–9
  pharmacogenomics **4**:2019
  punishable alcohol limits **1**:82–3
  sweat testing **5**:2428
  toxicology **1**:291, 292; **5**:2505–6
  traffic fatality prevention **5**:2542–3
  see also driving under the influence of drugs
driving force, materials corrosion **4**:1685
driving license regranting **3**:1430
driving restrictions, epilepsy **5**:2303
driving under the influence of drugs (DUID) **2**:878–82, 882; **5**:2428, 2511
  see also drug impaired driving
drop-in *see* allelic drop-in
drop-on-demand inkjet printers **5**:2670
drop-out *see* allelic drop-out
drowning **1**:225, 429, **2**:751, 755; **5**:2563
drowning-associated diatoms (DAD) **2**:752–3
  see also diatoms
drug abuse **4**:2059; **5**:2434
  see also alcohol abuse; alcohol, drug and mental disorders; substance abuse
drug–alcohol interactions **1**:108–17
drug analysis *see* illicit drugs, analysis
drug classes
  amphetamines **1**:134–9
  benzodiazepines **1**:293–7
  cannabis **2**:431–6
  cocaine **2**:562–8
  opioids **4**:1895–902
drug concentrations
  alterations postmortem **4**:2099–101
  blood **4**:2117–8, 2121
  gastric contents **4**:2122
  postmortems **4**:2093–108; **5**:2495–502
  urine **4**:2121
  see also toxicology
drug consignment sampling **5**:2282–9
drug consumption *see* drug use
drug courier profiles **4**:2157
drug courts **4**:1719; **5**:2416, 2450
drug dependence **1**:19, 122–4, 296; **4**:1899
drug detection
  interpretation of results **4**:2115–9
  microchemical tests **4**:1747
  postmortem toxicology **4**:2127, 2128–34
  see also drug screening; drug testing
drug distribution **1**:60–1; **4**:2115; **5**:2420–2
drug–drug interactions **1**:108–17; **4**:1901–2, 2095
  see also multiple drug use
drug-facilitated crime (DFC) **1**:295–6; **2**:868–76; **3**:1430–1
drug-facilitated sexual assault (DFSA) **1**:236; **2**:868–76
drug impaired driving **2**:877–83
  alcohol–drug interactions **1**:113
  alcohol intoxication **1**:100
  BAC interpretation **1**:58, 65, 67
  breath-alcohol analysis **1**:96, 97
  cannabis **2**:434, 436
  oral fluid **4**:1918–9
  pharmacogenomics **4**:2019
  punishable alcohol limits **1**:82–3
  sweat testing **5**:2428
  traffic fatality prevention **5**:2542–3
  see also driving under the influence of drugs
drug-induced hypertrophy **2**:472–3
drug–nutrient interactions **4**:2095
drug paraphernalia **4**:2123; **5**:2501
drug profiling **2**:851–9
drug reactions, malpractice cases **4**:1703, 1704–5
drug-resistant bacteria **3**:1346
drug safety research **5**:2508–9
drug screening **5**:2511–6, 2522–30
drug testing **4**:2207–8
  cutoffs **2**:598
  driver impairment **2**:881–2
  EQA schemes **4**:2226
  hair analysis **3**:1427–32
  oral fluid **4**:1903–20
  roadside **4**:1918–9
  sports **5**:2507
  sweat **5**:2426–8
  traffic medicine **5**:2506
  urine **2**:860–7
  workplace **2**:860, 866; **3**:1428; **4**:1916–7; **5**:2507
  see also drug detection; drug screening
drug trafficking **2**:851–9
drug transfer mechanism (oral fluid) **4**:1906
drug use
  alcohol interactions **1**:109
  behavioral toxicology **1**:290–3
  history of **3**:1429; **5**:2501
  multiple **1**:138; **4**:2095, 2119
  repeated **4**:2118–9
  see also drug abuse; drug testing
drugs
  antidepressants **4**:2205–6
  antipsychotic medication **4**:2203–5
  anxiolytics **4**:2206
  hallucinations **3**:1432
  interactions **4**:2207
    see also drug–alcohol interactions; drug–drug interactions
  metabolization **4**:2202
  plants used as **1**:425–6
  psychiatry **4**:2201–8

drugs *(cont)*
stimulants **4**:2207
therapeutic classes **4**:2203
truth serum use **2**:729–30, 731, 732, 736
see also individual drugs; medications; pharmacology
drugs-of-abuse **2**:869
see also alcohol; drug. . .; illicit drugs
Drugwipe immunoassay **5**:2426, 2429
drum marks **5**:2668–9, 2673
drunken driving *see* alcohol impaired driving
dry miscellaneous traces **4**:1929
dry swabbing technique **1**:320, 321
drying methods, explosion debris **2**:1033
DSC (differential scanning calorimeter) **2**:1036
DSM see Diagnostic and Statistical Manual of Mental Disorders
DSM-IV-R (Diagnostic and Statistical Manual of Mental Disorders) **4**:1657–8
DSM-IV-TR (Diagnostic and Statistical Manual of Mental Disorders) **4**:2187–8; **5**:2443
DSM-V (Diagnostic and Statistical Manual of Mental Disorders) **4**:2190–1
DSPD (dangerous severe personality disorder) legislation **2**:669
DTCs (diagnostic trouble codes) **1**:55
DTGS (deuterated triglycine sulfate) detector **4**:1937
DTT (dithiothreitol) **2**:804
dual detector systems, drugs **5**:2525–6
dual remote diaphragms **4**:1752
dual role dilemma **2**:558
ductile behavior, materials **4**:1682, 1683
due process **2**:957, 958, 960; **3**:1378
DUID *see* driving under the influence of drugs
Dunkle v. State 139 P.3d 228 (Okla. Crim. App. 2007) **2**:581
dura mater **1**:402, 404
durable powers of attorney **3**:1374, 1375
dural hemorrhage **5**:2343
duration of action, amphetamines **1**:136–7
duress defense **2**:576–7, 577
Durham Standard **1**:283; **5**:2444
Durham v. United States, 214 F.2d at 862 **5**:2444
Dusky v. United States, 362 U.S. 402, 80 S. Ct.788 (1960) **2**:441, 456
dust samples, pollen evidence **4**:1963
duty to warn **2**:670, 885–9
DVI *see* disaster victim identification
DVT (deep venous/vein thrombosis) **4**:1846
dyadic death *see* homicide–suicide
dye decomposition, ink aging **3**:1545
dye sublimation printing **5**:2670–2, 2672
dynamic approach, document dating **2**:686–8
dynamic headspace concentration **3**:1139
Dynamic Host Configuration Protocol (DHCP) **1**:146
dynamic risk factors **5**:2274, 2598, 2599
dysfunction (PTSD) **4**:2143
dystonia **4**:2204

e-discovery **2**:777–8
EA (enzyme acceptor) **5**:2515
EAP (erythrocyte acid phosphatase) **1**:351–2
early warning biodetection systems **1**:306
earprint interpretation **2**:891–6
Earth population paradigm **3**:1278, 1279, 1508, 1509, 1510
Ebanks v. People, 117 Cal. 652, 665, 49 p. 1049, 1053 (1897) **3**:1500
Ebbinghaus memory experiments **4**:1710
eccrine glands **5**:2420
ECD *see* electron capture detector
ecgonine methylester (EME) **2**:566; **4**:1912, 1913
ECHR (European Court of Human Rights) **2**:774
ecological succession, definition **2**:939
ecology **2**:946–53; **5**:2641–2
ecosystem sampling **2**:947–8
ecstasy (drug) **2**:849, 878
see also 3,4-methylenedioxy-methamphetamine
ED (enzyme donor), immunoassays **5**:2515
EDDs *see* electrostatic detection devices
edema, cerebral **1**:407
EDM (etching and electrical discharge machining) **5**:2324
EDMI (electronic distance measuring instrument) **2**:580
education **2**:897–901
criminalistics certification **5**:2549, 2554
expert witnesses **2**:1014
fire safety **3**:1226
forensic electrical engineering **2**:921–2
public view of sex offenders **5**:2337
training distinction **5**:2554
see also qualifications; training
EDX *see* energy dispersive X-ray
EEG *see* electroencephalogram
EEPROM storage, air bag systems **1**:57
EFA (exploratory factor analysis) **4**:2199
effect, cause and (Bayesian networks) **1**:278
'effect' paints **4**:1935
effective magnification (high power) **4**:1758
eggs (flies) **2**:944
see also oviposition
ego depletion **2**:573
Ehrenburg, Prof **5**:2380
EI-MS *see* electron impact positive ion mass spectrometry
EIA *see* enzyme immunoassays
EIC *see* extracted ion chromatogram
Eichel, Dr. Steve K.D. **2**:1015
EIP *see* extracted ion profiles
ejector marks, cartridge cases **3**:1210
elastic deformation, materials **4**:1681, 1682
elasticity of bone **5**:2559
elder abuse **2**:902–12
causation **2**:913
consequences **2**:907
definitions **2**:903
guardianships **3**:1373
risk factors **2**:912–5
see also older adults
electric field sensors **1**:326
electrical energy **3**:1104
electrical engineering **2**:920–33
electrical excitability of muscles **2**:698–701; **5**:2468, 2469–72

electrical pathway, materials **4**:1685
electrical systems, fire scenes **3**:1131
electrical tape **3**:1307
electricians, definition **2**:920
electro-convulsive therapy (ECT) **4**:2138
electro dip coating, paints **4**:1943
electro-optical fingerprint recognition **1**:325
electrochemical detection, chemical warfare agents **2**:513–9
electrocution cases **1**:245; **2**:923
electroencephalogram (EEG) **2**:724–5; **5**:2300–1
electrolyte imbalances **4**:2084–5
electromagnetic spectrum **3**:1293
electron beam irradiation **3**:1320
electron capture detector (ECD) **2**:1040, 1041; **4**:2126; **5**:2524
electron column, SEM **4**:1796–7
electron impact positive ion mass spectrometry (EI-MS) **2**:595
electron interaction, SEM **4**:1798–800
electron microscopy **1**:309; **4**:1686–7, 1793–804; **5**:2385
electron source, SEM **4**:1793–5
electronic data, ILRs **3**:1169
electronic devices **3**:1491; **4**:2089
  see also electronic media
Electronic Discovery Amendments, Rules of Civil Procedure **2**:777–8
electronic distance measuring instrument (EDMI) **2**:580
electronic media **2**:584; **3**:1273–5
  see also electronic devices
electronic nose device **4**:2089
electronically stored information (ESI) **2**:584, 777–8
electropherogram (epg) **1**:127; **5**:2591
electrophoresis **1**:339, 349, 350, 353
electrophotographic printers **5**:2667–9, 2672
  see also laser printers
electrophotography **3**:1558
electrospray ionization (ESI) **4**:2126–7; **5**:2529–31
electrostatic detection apparatus (ESDA) **2**:688; **3**:1597
electrostatic detection devices (EDDs) **1**:129, 130; **3**:1257
electrostatic lifting apparatus (ESLA) **2**:620
electrostatic printing *see* electrophotographic printers
elemental analysis
  glass **3**:1349–50, 1354, 1356
  paint **4**:1937
  soil **5**:2386
elemental gases **3**:1107
ELEMENTAL (software package) **3**:1358
eliciting expert testimony **2**:759–60
elimination actions, crime scenes **2**:607
elimination of drugs **1**:68–9; **2**:435, 870; **4**:2116
ELISA *see* enzyme-linked immunosorbent assays
elliptical bloodstain patterns **1**:366
elliptical forms, microscopy **4**:1777
elongation sign, fibers **4**:1782–3
elution technique **1**:342, 343, 347
embalmment **4**:2101–2
embedded objects **5**:2239
embolization, projectiles **3**:1385–6
embossed paper **3**:1273, 1597
embryology, friction ridge skin **3**:1327–30
EME *see* ecgonine methylester
emergency funds, elder abuse **2**:909
emergency guardianships **3**:1378–9
emergency medical treatment **2**:439; **4**:2095
emergency response, fires/explosions **3**:1136
emergency triage **5**:2461
EMIT *see* enzyme multiplied immunoassay technique
emitter filters *see* barrier filters
emotions
  crime victims **2**:644, 645–6
  neuroscientific research **2**:579
emphysema **1**:408
empirical approaches **3**:1278–9; **4**:2022, 2023
empirical standard *see* minutiae (fingerprints)
EMPOP database **4**:1829, 1830
EN 17025 standard **5**:2493–4
EnCase tools **2**:585, 586–7, 587
encephalopathy **5**:2340
enclosing barriers, suffocation **1**:226
encoding, incomplete **5**:2245
encryption **2**:587
endocarditis **2**:475–6
endogenous hepatic coma **4**:2081
endogenous hyperinsulinism **4**:2080–1
endurance level of materials **4**:1684
energy dispersive X-ray (EDX) **2**:989–90, 1036; **4**:1800
energy of bloodspatter *see* velocity of blood
energy of fire **3**:1103–5
ENFSI *see* European Network of Forensic Science Institutes
Engel, R.S. **2**:650–1, 651
engineering
  biomedical trauma causation **5**:2565–75
  ecological techniques **2**:950
  electrical **2**:920–33
  human factors **3**:1483–95
  materials engineering **4**:1680–8
  stress/strain **4**:1681
England *see* United Kingdom
engravings **3**:1273; **5**:2325
enhanced product ion (EPI) **4**:2131
enhancement techniques
  fingermarks **3**:1292–311, 1320
  fingerprints/DNA interaction **3**:1318–20
  photographic/optical **4**:2039–55
  spectral information **2**:596–7
enrollment mode, biometric devices **1**:324
enterohepatic recirculation **4**:2115
entomological succession **2**:937, 939–40
entomology **2**:934–45; **4**:2089, 2091
entomophilous plants **4**:1956
entrance hole (ammunition) **3**:1386
entrance wounds **3**:1196–8, 1386–7, 1390–3; **5**:2562
entrapment **2**:538–9; **4**:2159
enumerated population probabilities **3**:1562–3, 1563, 1564
envelopes, saliva DNA **2**:824
environmental audits **2**:951

environmental hypoxia **1**:232
environmental risk factors, elder abuse **2**:915
Environmental Science Research (ESR) Institute **5**:2592
environmental sciences **1**:199
  see also environmental toxicology
environmental SEM (ESEM) **4**:1793, 1801–3
environmental systems **2**:946–53
environmental testing, biometric devices **1**:328, 329
environmental toxicology **5**:2508
  see also environmental sciences
enzyme acceptor (EA), immunoassays **5**:2515
enzyme donor (ED), immunoassays **5**:2515
enzyme immunoassays (EIA) **4**:1908; **5**:2513–4
enzyme-linked immunosorbent assays (ELISA) **5**:2396, 2427, 2428, 2512, 2513, 2514
enzyme multiplied immunoassay technique (EMIT) **5**:2428, 2513–4
enzymes **2**:954
  CWA detection techniques **2**:521–2
  diatom extraction method **2**:753–4
  histochemical methods **3**:1471
  oxidation in alcohol analysis **1**:84–5
  Rhesus blood group system **1**:346
  see also acid phosphatase; enzyme. . .; protein...
epg *see* electropherogram
Ephedra species **1**:134
EPI (enhanced product ion) **4**:2131
epidemiology
  communities **2**:952
  drug impaired driving **2**:878–9
  forensic toxicology **5**:2508–9
  medical malpractice **4**:1690–2
  sharp force injuries **5**:2646–7
  SMI in jail studies **2**:652
epidermis **3**:1323–6
epidural hematoma **3**:1534
epidural hemorrhages **1**:404
epilepsy **5**:2298–305
  aggression and the law **5**:2304–5
  automatism defense **1**:254, 255
  children **4**:1872
  classification **5**:2299
  consequences **5**:2302–3
  definition **5**:2298
  diagnostic methods **5**:2300–1
  differential diagnosis **5**:2301–2
  disability **5**:2303
  driving restrictions **5**:2303
  epidemiology **5**:2299–300
  incidence/prevalence **5**:2298
  seizures classification **5**:2298–9
  syndromes **5**:2299
  treatment **5**:2303–4
epinephrine **1**:134
episodic dyscontrol syndrome **5**:2301
episodic memory **2**:916
epitheca valve, diatoms **2**:749
epithelial material **2**:757, 822, 1061
EQA *see* external quality assessment
'equality of arms' theory **2**:775
  see also 'equivalence of arms' theory
equilibrium **3**:1458
equipment documents **1**:128–9
'equivalence of arms' theory **2**:783
  see also 'equality of arms' theory
equivocal deaths **4**:2162, 2170
erasures, documents **1**:128–9, 131–2
ergonomics **3**:1483–95
ergot **2**:728
erotomanic delusions **2**:742, 743
ERPs (event-related potentials) **2**:724–5
error rates **2**:694, 955–7; **4**:1698; **5**:2402–3, 2406, 2408
errors **2**:955–7
  3D reconstructions **5**:2264–6
  of execution **4**:1690
  fingerprint interpretation **3**:1280
  hair examination **3**:1423
  human factors context **3**:1490
  interpretation of results **3**:1583–4
  match statistics **5**:2403–4, 2408
  measurement **2**:956
  medical malpractice **4**:1690, 1698, 1702, 1703, 1704, 1705
  mtDNA evidence **4**:1825, 1829
  of planning **4**:1690
  reconstructive memory **4**:1710, 1711, 1712
  STR profiles **5**:2371–2
  see also error rates
erythroblastosis foetalis **1**:345
erythrocyte acid phosphatase (EAP) **1**:351–2
Erythroxylum coca v. coca see cocaine
escalation of aggression **1**:27
Escherichia coli **1**:343
ESD (esterase D) **1**:352
ESDA (electrostatic detection apparatus) **2**:688; **3**:1597
ESEM *see* environmental SEM
ESI *see* electronically stored information; electrospray ionization
ESLA (electrostatic lifting apparatus) **2**:620
ESR (Environmental Science Research) Institute **5**:2592
establishment of guilt, civil law **2**:1005
Estelle v. Gamble, 429 U.S. 97 (1976) **5**:2583
esterase D (ESD) **1**:352
estimation
  age of single ink entry **3**:1542, 1544
  biochemical **4**:2076
  DNA match statistics **2**:834–7
  height **3**:1624–32; **5**:2615
  post-mortem interval **2**:936–9, 942–3
  quantities **5**:2281–90
  time of death **5**:2466–79
  of time since discharge (firearms) **3**:1198
  upper bound frequency estimation **4**:1824–5
estimator variables **2**:1072, 1077–8, 1078
etching and electrical discharge machining (EDM) **5**:2324
etching techniques **5**:2324, 2326–7
ethanol **1**:58–9, 81, 121
  analytical methods **1**:83, 88
  intoxication effects **1**:227
  lethal dose **1**:77–8
  physiochemical properties **1**:58
  postmortem alterations **4**:2100–1
  tolerance to **1**:74–6

toxicity **1**:76–7
see also alcohol
ethics **2**:443, 957–62; **4**:2032–3, 2165; **5**:2336
ethnicity **1**:163–6; **4**:1835, 1879
see also race
ethyl alcohol *see* alcohol
ethylglucuronide **3**:1430
etiology of psychopathy **4**:2194–5
Eucalyptus species **4**:1979
eugenics movement **3**:1336
euhedral crystal faces **4**:2004
euphoric hypoxia **1**:246
Europe
admixture **4**:2024–8
doctor-patient confidentiality **2**:888
inpatient civil commitment **2**:557–8
see also individual countries; United Kingdom
European Court of Human Rights (ECHR) **2**:774
European Network of Forensic Science Institutes (ENFSI) **1**:10, 15; **5**:2494
evaluation
AFIS technology **1**:253
CAI reporting **2**:494
capacity to waive Miranda rights **2**:463–7
crime investigation phase **2**:495
DNA evidence **5**:2591–2
elder abuse **2**:904–6
ignitable liquid residue samples **3**:1138
independent living assessment **2**:448
laboratory accreditation **1**:13
numerical height estimations **3**:1630
scientific evidence **1**:276–80
soil sample comparisons **5**:2383
sole database use **3**:1243
statistical (glass evidence) **5**:2590–1
see also assessments
evaluative opinion (Bayes theorem) **2**:487, 488
evaluator role **3**:1454; **4**:1981
evanescent waves **4**:1756
event-related potentials (ERPs) **2**:724–5
everyday forgetfulness/amnesia distinction **5**:2244
evidence
analysis (drug impaired driving) **2**:881–2, 883
animal-related **5**:2635–9
authentication of documents **2**:840–1
automobile accident reconstruction **5**:2250
barefoot impressions **3**:1244
behavioral science **1**:281–8; **5**:2431–2, 2438
best evidence rule **1**:298–9
bloodstain patterns **1**:360–96
bomb scene management **1**:413, 415–7
burglary scenes **2**:616–7
case assessment and interpretation **2**:493–4
casting **2**:963–7
categorical opinions **3**:1564–5
chain of possession **2**:498–500, 759
collection methods **2**:504–5, 952, 963–7
computer animations/simulations **2**:579–82
computer forensics **2**:584–8
confessions **2**:588–93
control criteria **1**:6
demonstrative **2**:745–7
depositions **2**:772–4
disaster victim identification **2**:770
DNA databases **2**:831–8
documentation **2**:602–14
drowning cases **2**:751
ecological **2**:948–9, 952
enhancement techniques **4**:2039–55
entomological **2**:943–4
evaluation (Bayesian networks) **1**:276–80
explosion scenes **2**:1023, 1025–6
fibers **3**:1095–101
fire scenes **3**:1128–9
first officer attending scene **2**:620
footwear impression quality **3**:1250
glass **3**:1348–50; **5**:2590–1
hair comparisons **3**:1424
handwriting **3**:1438
hearsay **2**:664; **3**:1465–7, 1624
height estimations **3**:1631
histological findings **3**:1471–2
identification reporting **3**:1509–10
illicit drug profiling **2**:857–8
ink evidence interpretation **3**:1546–52
ipse dixit **3**:1599–600, 1620
judicial notice **3**:1601
learned treatises **2**:664; **3**:1623–4
legal interpretation **3**:1561–5
logical approach to interpretation **2**:968–76
mitochondrial DNA **4**:1823–32
neuropsychological **4**:1862–3
optical examination techniques **4**:2036–56
outside scenes **2**:619
packaging **2**:1029; **4**:1927–31
paint **4**:1933
case histories **4**:1940–2
interpretation of **4**:1943–53
investigations **4**:1931, 1939
recovery methods **4**:1934–5
significance/reporting **4**:1939–40
photography **2**:625–43, 746–7; **4**:2036–56
casting procedure **2**:963
documentation **2**:603, 608
examination **4**:2036–56
plants **2**:827–8
pollen **4**:1954–67
postmortem toxicology **4**:2123
preservation **2**:963–7
probabilistic statements **3**:1561–4
rules of **1**:23; **2**:963
standards **5**:2438
statistical (court communications) **5**:2401–8
storage (CBRN scenes) **2**:505
strength assessment (mtDNA) **4**:1824–7
tampering with STR profiles **5**:2371–2
transport **4**:1927–31
vehicle crime scenes **2**:617–9
weight of expert credentials **2**:1016–7
wood (physical fit) **1**:427
see also DNA/DNA evidence; exhibits; expert opinion evidence; *individual evidence types*; sampling; specimens; trace evidence
Evidence Act 1995 (Australia) **2**:1003
evidence close-up photographs **2**:633–5
evidence establishing photographs **2**:629–33
evidence vouchers **2**:607

evil
courtroom applications **2**:980–1
dangers of concept **2**:981–2
forensic interest in concept **2**:979
forensic mental health **2**:977–85
human face **2**:980
mental illness and **2**:981–2
redundant concept **2**:981
subjective, moral concept **2**:977–8
witchcraft hysteria **2**:978–9
evolution **3**:1342
evolutionary relatedness **4**:1812, 1815, 1816, 1817
EWG (Expert Working Group) **5**:2494
examination of evidence
animal hairs **3**:1409, 1413
documents **3**:1257–75, 1436–48
expert opinion **2**:662–5, 758–60
explosion debris **2**:1032–5
fibers/textiles **2**:985–96
firearms **3**:1200–4, 1217, 1218
gunshot wounds **3**:1397–8
hair **3**:1420–3, 1427
human remains **5**:2474–8
light globes/filaments **3**:1632–8
luminol tests **3**:1653
paint **4**:1935, 1939, 1940
photographic/optical techniques **4**:2036–56
shoemark guidelines **3**:1230–1
toolmarks **5**:2485–94
see also individual evidence types; microscopy
examination of scenes **2**:616; **3**:1636
see also scene...
examination of witnesses **3**:1505
see also cross-examination; direct examination
examination strategy, CAI **2**:492
examinations
biochemical **4**:2076–87
criminalistics certification **5**:2549, 2550
postmortem (traffic fatalities) **5**:2544
radiological **5**:2236–7
sexual assault cases **1**:236, 238
see also autopsies; examination of...
examiner bias *see* observer effects
examining magistrates **2**:561–2, 1005, 1006, 1007
excavations **1**:199, 204–6, 206; **3**:1497–9
exceptions to hearsay evidence rule **3**:1466–7
excessive force concept **4**:2072–4
excitability
electrical **5**:2468, 2469–72
facial muscles **5**:2468, 2469–72
iris **2**:701–2; **5**:2468
mechanical **5**:2468
pharmacological **5**:2468
skeletal muscle **2**:698–703; **5**:2468
excitation filters **4**:1786
excitation ranges, photoluminescence **4**:2047
exclusion
fingerprint comparison **3**:1277, 1285, 1288
handwriting examination **3**:1441, 1443
mtDNA matching **4**:1824
personal identification **1**:172
probability (match statistics) **5**:2402
excretion of drugs **1**:62–3, 137; **2**:566; **4**:1900
excuse defense **2**:577; **4**:2150–1
execution elements, handwriting **3**:1442
execution killings **3**:1477
executions **2**:442–3; **4**:1731
see also death penalty
executive functioning **3**:1462; **4**:1867
exemplars
glass **4**:1929
reasoning by **5**:2406
writings **3**:1438, 1439–40, 1444
see also control samples
exhibits
handling at crime scenes **2**:624
labels **2**:606–7
list **2**:607–8, 770
packaging **4**:1927–31
store register **2**:608
transport **4**:1927–31
see also evidence; specimens
exit wounds **3**:1386, 1388; **5**:2562
Exner personality assessment **4**:2184
exogenous hyperinsulinism **4**:2080–1
exons **1**:340
see also coding regions
exothermic reactions **2**:965; **3**:1114
exotic species fiber analysis **4**:1979
expectation component, suggestibility **3**:1591
experience of expert witnesses **2**:1014
experimentation **2**:950, 951–2; **5**:2487–9, 2565–75
expert-based approaches, speech analysis **5**:2390, 2391–2
expert behavior, ethics **2**:957–62
expert evidence *see* expert opinion evidence; expert testimony
expert opinion evidence
admissibility **2**:500, 692–6, 760, 1001–2, 1003–4, 1004–7, 1007–11; **3**:1331–1333, 1334–5, 1528, 1619–21; **5**:2627–8
chain of possession **2**:499–500
cross-examination **2**:662–5
direct examination **2**:662, 758–60
FRE 702 **3**:1095
hypothetical question **3**:1505
impeachment **2**:662, 663
ipse dixit **3**:1599–600
jury instructions **3**:1607–8
legal vs scientific method **5**:2297
paint interpretation **4**:1947
parental alienation **4**:1981
quality of **2**:483
report writing **5**:2268–70
rules of **2**:963
toolmarks **5**:2492–3
'ultimate issue' **5**:2589–90
expert testimony
appeal use **2**:998–1000
battered woman syndrome **1**:272–3
battered woman's reality **1**:273–5
behavioral evidence **1**:285, 286
child sexual abuse accommodation syndrome **2**:540–1
confession evidence **2**:593
defendant's personality profile **4**:2159–60
ecological cases **2**:949
Federal Rules of Evidence 702 **2**:540

firearms investigation reports **3**:1218
Frye test **2**:540
interrogations **3**:1589
mental retardation **4**:1731
mitigation **4**:1821
syndromes **5**:2438, 2439–40
trial use **2**:998–1000
see also expert opinion evidence
expert witnesses
certification **5**:2550
credentials **2**:1013–8; **4**:1862
definition **2**:1012–3
direct examination **2**:758–60
discovery of findings **2**:772–4, 774–5, 775–8, 778–81
foundation testimony **3**:1276
hair comparison evidence **3**:1424
human factors **3**:1491–2
malpractice **4**:1663–7
neuropsychological **4**:1862
qualifications **2**:663–4, 759, 1012–3, 1014
selection **2**:1013–8
see also expert opinion evidence; expert testimony; party-appointed experts
Expert Working Group (EWG) **5**:2494
expirated bloodstain patterns **1**:381–3
exploratory factor analysis (EFA) **4**:2199
explosion debris **2**:1028–60
explosions
definition **1**:412
investigations **2**:1019–27; **3**:1136–7, 1137, 1173, 1175
scene investigation **2**:1019–27
see also bomb. . .; fire...
explosive residues **1**:415–7; **2**:1026, **4**:1746
explosives **3**:1399
see also explosion...
exposure-based cognitive behavior therapy **4**:2146
extended suicide *see* homicide–suicide
external ballistics **3**:1202
external compulsion **2**:571, 573
see also compulsion
external examination
autopsies **1**:260
burns (findings) **3**:1529–32
scalds (findings) **3**:1537
time of death determination **5**:2474–5
external force *see* external compulsion
external quality assessment (EQA) **4**:2224, 2226–7
external support, workplace **5**:2464–5
extinction, particles **4**:1782
extracted ion chromatogram (EIC) **3**:1143–4, 1153, 1167
extracted ion profiles (EIP) **3**:1143, 1153
extraction methods
analytical artifacts **4**:2106–7
collagen **5**:2232
confirmation testing **2**:598–9
diatoms **2**:753–4
DNA **2**:757, 803–5, 1060–4; **4**:2219
features (biometric devices) **1**:324
friction ridge features **1**:249–51
ink aging **3**:1544
mtDNA **4**:1834
pollen samples **4**:1960–6
postmortem toxicology **4**:2125, 2130, 2131
see also collection methods/devices
extractor marks, cartridge cases **3**:1210
extradural hemorrhages **1**:404
extrafamilial homicide–suicide **5**:2419
extralegal information **3**:1603
extrapolation methods, BAC **1**:58, 69–70
extremity injuries **1**:411
eye
facial reconstruction **3**:1088
periorbital hematoma **1**:399
photoreceptors **4**:2041
vitreous humor **1**:71; **4**:2103, 2122; **5**:2472–3, 2498–9
see also iris...
eyelet document security features **3**:1272
eyepieces, microscopy **4**:1758, 1759, 1766, 1791
eyewitness identification **2**:1072–5
see also eyewitness testimony
eyewitness memory **2**:549, 553, 918, 1065–70
see also hypnosis; repressed memory; suggestibility
eyewitness reports **2**:1067
eyewitness testimony **2**:1075–9
children **2**:549–51, 659
crime scenes **4**:2021
estimator variables **2**:1077–8
juror knowledge **2**:1078
juror trust **2**:1078
reconstructive memory **4**:1711
system variables **2**:1077–8
truth serum use **2**:731
see also eyewitness identification; testimony
eyewitnesses
credibility **2**:918–9
older adults **2**:916–9
see also eyewitness...
Eysenck, Hans **4**:2179
f-stop settings **2**:639–41

fabric bloodstain patterns **1**:385
fabric marks **4**:1669, 1670–4
see also glove marks
fabrication of documents **4**:1968
face aging techniques **1**:190
face masks **2**:615
facial approximation *see* facial reconstruction
facial characters/indices **5**:2614
facial comparison **3**:1081–6
facial compression **1**:226
facial fiducial points **5**:2614
facial injuries, gagging **1**:226
facial muscles **5**:2468, 2469–72
facial recognition **1**:190; **5**:2611, 2614–5
facial reconstruction **3**:1086–91; **5**:2238
facility-related elder abuse **2**:904, 905, 914
facsimile machines **5**:2673–4
fact settings, ethics **2**:959–60
fact witnesses **2**:758–9; **5**:2268, 2270
see also facts; lay witnesses
factitious disorders **4**:1660
see also malingering; Munchausen syndrome

factor analysis **4**:2199–200
facts
  adversarial system **1**:23
  behavioral science **1**:285–6, 286–7
  finder of **2**:735–6
  judicial notice **3**:1601–2
  reporting (CAI) **2**:494
  trier of **1**:286–7
  see also fact witnesses
FAEE (fatty acid ethyl esters) **3**:1430
failure analysis (materials) **4**:1680–8
failure mechanisms, light globes **3**:1634
failure modes (materials) **4**:1683–5
faintings imitating seizures **5**:2302
fair trial concept **2**:774
faking *see* malingering
fall-down injuries **1**:100
fall-off-the-cliff effect **3**:1582
fallacy
  defense (attorney) **3**:1584; **5**:2405
  fingerprint examiner's **3**:1509
  prosecutor's **2**:975; **5**:2403
  of the transposed conditional **3**:1562, 1583
false accept rate (FAR) **1**:329
false accusations/allegations **2**:531, 535; **4**:1713
false confessions **2**:588–9, 591; **3**:1588, 1589
  classification **2**:588
  definition **3**:1590
  false evidence tactic **2**:592
  interrogative suggestibility **3**:1590–1, 1591, 1592–3
  suspect vulnerabilities **2**:591–2
false confidence, eyewitnesses **2**:1074
false denials, child sexual abuse **2**:534–5, 539
false evidence tactic **2**:592; **3**:1587
false identification of suspects **2**:917
false information, child interviews **2**:553
false matches, STR profiles **5**:2371–2
false memories **2**:589, 916, 1069; **4**:1710–1, 1712, 1714, 1716
  children's suggestibility **2**:555
  hypnosis **3**:1501, 1502
  syndrome **4**:1714
false negatives **2**:862, 955; **3**:1651; **4**:2028
  see also Type II error
false positives **2**:862, 955; **3**:1651, 1653; **4**:1658, 2028, 2125
  see also Type I error
false reject rate (FRR) **1**:329
falsifiability
  definition **3**:1093
  theory of **2**:694; **3**:1093–4
falsified logs, network analysis **1**:142
FAME (fatty acid methyl esters) **3**:1168
familial relatedness *see* relatedness
familial searching **2**:682; **5**:2680
familicide–suicide **5**:2419
family consent laws (healthcare) **3**:1375
family courts, parental alienation **4**:1983–4
family identification (wildlife) **5**:2637
family liaison officers (FLOs) **2**:766, 767, 771
Family Relations Test **5**:2606
family violence *see* domestic violence
fantasies **5**:2315, 2321, 2399
FAR (false accept rate) **1**:329
fast-developing fire myth **1**:216
fast fourier transform (FFT) **3**:1525, 1526
Fast Patch devices **5**:2423
FASTs (financial abuse specialist teams) **2**:909
fatal maltreatment filicide **4**:2139
fatal toxicity index **5**:2508–9
fatalities
  human factors **3**:1492–5
  traffic fatalities **5**:2541–4
  see also death; lethal cases
fatigue failure, materials **4**:1683–4
fatty acid ethyl esters (FAEE) **3**:1430
fatty acid methyl esters (FAME) **3**:1168
Faulds, Henry **3**:1282
fault codes **1**:55
faults
  dot matrix printers **5**:2666–7
  facsimile machines **5**:2674
  inkjet printers **5**:2670
  laser printers **5**:2668–9
  photocopiers **5**:2673
  typewriters **5**:2665
  see also error...
Faust, David **4**:2182
fax machines **5**:2673–4
FBI *see* Federal Bureau of Investigation
FDEs *see* forensic document examiners
FDR *see* firearm discharge residues
FDS *see* Fragment Data System
FE gun *see* field emission gun
FearID project **2**:892–3, 894
feature comparison, friction ridge skin **1**:251–2
feature extraction, friction ridge skin **1**:249–51
feature lists, facial comparison **3**:1083, 1084
fecal material **1**:318–9, 427; **2**:826–7, 1063
Federal Bureau of Investigation (FBI)
  CODIS system **2**:678
  DNA databases **4**:1835, 1836
  Handbook of Forensic Services **5**:2619
  IED group **2**:1045
  NAA use **1**:151
  source attribution approach **2**:832
  threat assessment **5**:2454, 2455
federal death penalty *see* death penalty
federal requirements, discovery **2**:779–80
Federal Rules of Criminal Procedure
  16 **2**:779–80
  17 **2**:780
Federal Rules of Evidence (FRE)
  402 **2**:694
  702 **2**:693, 694, 1001, 1002; **3**:1095, 1624
  703 **3**:1505
  704 **5**:2589–90
  720 **2**:540
  801 **3**:1466, 1467
  803 **3**:1466, 1467
  901 **2**:841
  902 **2**:841
  best evidence rule **1**:299
  document authentication **2**:841
  General Electric v. Joiner **3**:1334
  hearsay **3**:1466, 1467
  see also Rules of Civil Procedure

feet *see* foot
Feigl, F. **4**:1745
feigned delusions **2**:743
feigned illness *see* malingering
feigned mental retardation **4**:1734–5
feigned psychosis **3**:1434
felonious violent crimes **2**:934
  see also medicolegal investigations
felt marks **4**:1973–4
female aggression **1**:36–51
female criminality **1**:40; **4**:2151; **5**:2314, 2399
  see also gender; women
female remains *see* sex determination
female sample, Northwestern Juvenile Project **4**:1879
female serial murderers **5**:2314
female stalkers **1**:40; **5**:2399
femoral blood **4**:2102
femurs **5**:2330, 2394
FEPAC *see* Forensic Science Education Programs Accreditation Commission
Ferroin test **4**:2122
ferrozin **3**:1212–5
fetal development **3**:1327–30
fetal remains **1**:179
fetishism **5**:2317
FFT *see* fast fourier transform
FHEs (forensic handwriting examiners) **3**:1436–48
FIA phenomenon **5**:2246, 2247
fiber analysis **2**:988–92; **4**:1968, 1970, 1975
  furnish analysis **4**:1968–9, 1979
  length analysis **5**:2643–4
  microchemical techniques **2**:992
  microscopy **2**:988–90; **4**:1775, 1781, 1782–3, 1975–80
  spectroscopy **2**:990–1
  staining **4**:1979
  weighting **4**:1975
  see also fibers; paper, analysis
fiber evidence *see* fibers
fiber tapings **4**:1929
fiber tip pens **5**:2662
fibers **3**:1095–101
  Bayesian approach **3**:1100
  classification **2**:986–8
  composites failure analysis **4**:1687
  document security features **3**:1267
  examination of **2**:985–96
  frequency **3**:1098–100
  intelligence work **2**:992–6
  interpretation **3**:1097–101
  particle forms **4**:2006
  persistence **3**:1098; **5**:2535
  probabilistic model **3**:1100–1
  recovery methods **2**:985–6, 993, 994
  statistical treatment **3**:1100–1
  target fiber studies **3**:1099
  transfer **3**:1098; **5**:2534–5, 2538
  value **3**:1096–7
  wood **5**:2642, 2643–4
  see also fiber analysis
fibular side, feet **3**:1322, 1323
FID *see* flame ionization detectors
field emission (FE) gun **4**:1793, 1795
field instruments, explosion debris **2**:1028
field investigations, explosions **2**:1020, 1027
field models, fire **3**:1183–4, 1185
field-of-view index, microscopy **4**:1766
field walking **3**:1497
Fifth Amendment discovery requirements **2**:779
filaments **3**:1632–8; **4**:1795
filicide **4**:2138–9
filicide–suicide **5**:2418
filing (serial number obliteration) **5**:2325
filing guardianship petitions **3**:1378
fill-flash photography **2**:637–8
fillers
  automotive paints **4**:1943
  lineups **2**:1072, 1073
  paper analysis **4**:1968, 1974
film emulsions **4**:2049
film photography **4**:2048–9
  see also photograph...
film reference cards **2**:626
filter cubes **4**:1786
filters
  digital photography **4**:2041, 2049
  light examinations **3**:1258–9
  optical evidence enhancement **4**:2041–2, 2046, 2047
final surveys, explosion scenes **2**:1026
financial abuse of elders **2**:903, 905, 906, 907
financial abuse specialist teams (FASTs) **2**:909
financial decision-making **3**:1375–6
finder of fact, truth serum **2**:735–6
  see also trier of fact prerogatives
Finger Tapping Test **4**:1865
fingermarks **3**:1282
  definition **1**:249; **3**:1282
  detection/enhancement **3**:1292–311, 1320
  diffuse reflection lighting **4**:2051
  digital imaging **4**:2054, 2055
  glove mark comparison **4**:1669
  see also fingerprints
fingernail scratches **1**:397
fingernails, DNA **2**:825
fingerprint donor category **3**:1215
fingerprint examiner's fallacy **3**:1509
fingerprint powders **3**:1294–5
fingerprint tape **1**:387
fingerprints **1**:322
  AFIS system **1**:249–53
  burglary scenes **2**:617
  classification systems **1**:249
  comparison **1**:251–2; **3**:1277–8, 1282–90, 1602; **4**:2011
  decomposed remains **3**:1512–3
  definition **1**:249; **3**:1282
  disaster victim identification **2**:768, 769
  DNA
    biological material interaction **3**:1318–21
    evidence **5**:2555, 2556
    source **2**:825
  documentation **2**:604–6, 609
  evidence close-up photographs **2**:635
  historical evidence **3**:1282–3
  identification **3**:1509, 1510, 1577, **5**:2611
  image processing **3**:1525–6

fingerprints *(cont)*
individualization **1**:168; **3**:1509, 1510
infrared microspectral imaging **4**:1757
interpretation **3**:1277–80
judicial notice **3**:1565
living individual identification **5**:2611
recognition devices **1**:325–7, 330–4
shooting scene evidence **3**:1222
spoofing **1**:330–4
submission forms **2**:609
web resources **5**:2625–6
wildlife forensics **5**:2638
see also DNA fingerprinting; fingermarks; friction ridge skin
finishes *see* top coat paints
finite microscopes **4**:1768
finite population correction (fpc) **5**:2284
Finland **2**:880; **5**:2507
fire
accelerants **3**:1133; **4**:1929–30; **5**:2537
accidental injuries **1**:102
chemical processes **3**:1103–12
consumption by **3**:1530–2
debris **3**:1111, 1137–70
dynamics **3**:1112–21, 1176
effects **3**:1112, 1120
energy **3**:1103–5
patterns **3**:1112–21
safety education programs **3**:1226
see also arson
fire investigation **3**:1136–7, 1137
applied knowledge use **3**:1176
burns **3**:1529–37
electrical engineering **2**:922, 924–5, 929
fire modeling applications **3**:1175–88
misconceptions/myths **1**:207–23
standardization/accreditation/certification **3**:1171–4
see also fire scenes
fire load myth **1**:209, 218–9
fire modeling **3**:1175–88
see also fire investigation
fire officer's documentation **2**:613
fire plumes **3**:1120
fire scenes **3**:1122–36
accelerant sample packaging **4**:1930
accidental cause **3**:1123
avoiding spoliation **3**:1127–8
documentation **3**:1124–6, 1126, 1136
evidence collection/preservation **3**:1128–9
fragile item packaging **4**:1930
hypothesis developing/testing **3**:1133–5
initial survey **3**:1124
inventory **3**:1127
need for investigation **3**:1122–3
origin determination **3**:1129–32, 1133–5, 1136
planning investigation **3**:1123–4
reconstruction **3**:1126
reporting procedure **3**:1135, 1136
see also fire investigation; fire modeling
firearm discharge residues (FDR) **3**:1189–98; **5**:2351–3
see also cartridge discharge residues
firearms
barrel lead deposits **3**:1198
DNA evidence **3**:1222
estimation of time since discharge **3**:1198
examination **3**:1200–4
fingermark detection **3**:1308–9
gunshot entry association **3**:1196–8
identification **3**:1204–11, 1211–5
investigation **3**:1216–8, 1219; **4**:1678
mass grave investigation **4**:1678
packaging **4**:1929
serial number restoration **5**:2324–7
shooting scene investigation **3**:1219
shooting-distance estimation **5**:2351–3
web resources **5**:2626
wildlife forensics **5**:2637–8
wounds **3**:1380–401
see also ammunition; firearm discharge residues; shooting scene investigation
firesetting **1**:207–23; **3**:1169, 1225–8
FireWire Target Disk Mode **2**:585–86
firing-distance estimation **5**:2351–3
firing pin impressions **3**:1210
firmware, mobile phone handsets **3**:1364
first-degree burns **3**:1530
First Officer Attending (FOA) **2**:615, 619–20
first-pass metabolism **4**:1900, 1902
first report to police **2**:619
Fisher–Race hypothesis **1**:346
fissures, anal intercourse **1**:239
fixing techniques **3**:1304, 1310, 1470
flail chest **1**:408
flame ionization detectors (FID) **4**:2126; **5**:2524
flame photometric detection (FPD) **2**:520–1
flame-spread index **1**:218–9
flame test, explosion debris **2**:1036
flames **1**:218–9; **3**:1115–6
see also fire
flammability **3**:1116
Flash memory removable cards **3**:1362, 1365, 1366
flash photography **2**:633–4, 637–9
flash point, vapour combustion **3**:1109
flash search strategy **2**:622, 623
flashover, definition **3**:1180
flashover point (compartment fires) **3**:1118–20
flaws, materials **4**:1683, 1686
see also defects
flesh flies **2**:940
flexography **3**:1264
flies **2**:934, 936
accumulated degree-days/-hours **2**:942
collecting for evidence **2**:943–4
crime scene value **2**:944–5
DNA source **2**:827
life cycle/development **2**:940–1
PMI estimates **2**:937, 942–3
floor effect tests **4**:1868
floor holes fire myth **1**:209, 219–20
floor plans, fire scenes **3**:1124
floppy MVPS **2**:476–7
FLOs *see* family liaison officers
flow of ridges, fingerprints **3**:1283
flowering plants **4**:1976
FLSs *see* forensic light sources

FLTs (forklift trucks) **3**:1489
fluids
  alcohol analysis **1**:81–99
  BAC interpretation **1**:71–2
  body fluids **5**:2555
  postmortem toxicology **5**:2500–1
  stain identification **1**:314–9
  storage containers **4**:2104
  swab removal **1**:320–1
  see also blood; cerebrospinal fluid; oral fluid; seminal fluid; urine...
flunitrazepam **4**:1915
fluorescein **3**:1320
fluorescence **2**:989; **4**:1760, 1784–8, 1958–9, 2046, 2047
  see also fluorescent...; photoluminescence
fluorescence polarization immunoassay (FPIA) **5**:2515
fluorescent dyes **4**:1688–9
fluorescent powders **3**:1295
fluorescent whitening agents (FWAs) **4**:1968, 1974
fluoride preservation **4**:2104
fluoroscopy **2**:769; **5**:2234
flux *see* velocity of blood
Flynn effect **4**:1729, 1733
fMRI (functional magnetic resonance imaging) **2**:725–6
FOA *see* First Officer Attending
focal concerns theory **5**:2309
focal plane array **4**:1757
focus (photography) **2**:639–41
focus of attention (trauma) **2**:787, 1077
foil stamp document security **3**:1273
folie à deux/trois **5**:2446
folkways **1**:281, 282
follicle-stimulating hormone (FSH) **5**:2333
follicles of hair **3**:1417
follow up assessments **2**:448
Folstein Mini-Mental State Examination **4**:1864
food residues **1**:427–8
foot
  anatomical regions **3**:1322–3
  comparison of shoes to **3**:1247–8
  embryonic development **3**:1327–8
  impressions **3**:1244–8, 1252–5
  structure of **3**:1244
  see also footwear...
footer information, pixel defects **3**:1524
footwear
  bloodstain patterns **1**:388–9
  comparison of foot to **3**:1247–8
  comparisons between **3**:1252
  DNA analysis of insides **3**:1247
  documentation **2**:602–3, 610
  linking foot to **3**:1244–8, 1253
  outsole databases **3**:1254
  overview **3**:1252–5
  packaging **4**:1927–8
  submission forms **2**:610
footwear impressions **3**:1253–4
  burglary scenes **2**:616, 617
  casting **2**:966, 967
  databases **3**:1249–51
  first officer attending scene **2**:620
  intelligence sources **3**:1248–52
  packaging **4**:1929
  sizing **3**:1243
  sole databases **3**:1240–3
  see also shoemark examination guidelines
force, police use of **4**:2068–71, 2072–4
force of attack, stab wounds **5**:2648
  see also sharp force injuries
forced choice tests **4**:1868
forced confessions **3**:1586
forced entry, burglary scenes **2**:616, 617
forced medications **2**:559
Ford v. Wainwright, 106 S. Ct. 2595 (1986) **2**:442–3
Ford v. Wainwright, 477 U.S. (1986) **2**:719
Fordisc program **1**:194
forensic 3D reconstruction **5**:2257–67
forensic anthropology **1**:152–79; **5**:2557–63, 2611–5, 2623–4
  see also anthropology
forensic archaeology **1**:199–207; **3**:1495–9
  see also archaeology
forensic assessment perspectives **4**:2182
forensic associations **5**:2548, 2622–3
  see also forensic institutes; professional associations
forensic ballistics *see* forensic firearms investigation
forensic biomedical engineering **5**:2565–75
  see also biomedical engineering
forensic botany **1**:423–30; **5**:2625
  see also botany
forensic diatomology **2**:749–56
  see also diatoms
forensic DNA databases **2**:677, 680–2, 683–4
  see also DNA databases
forensic document examiners (FDEs) **3**:1257–9, 1275
  see also documentary evidence
forensic ecology *see* ecology
forensic electrical engineering **2**:920–33
forensic engineering **2**:920–33; **3**:1483–95, 1493; **5**:2565–75
  see also engineering
forensic entomology **2**:934–5, 936, 939, 943; **4**:2089
  see also entomology
forensic evaluations, mental status **4**:1738–9
forensic evidence admissibility **2**:1005
  see also evidence
forensic firearms investigation **3**:1216–8
  see also firearms
forensic geology **5**:2378
forensic handwriting examiners (FHEs) **3**:1436–48
forensic histology **3**:1468–73
  see also histology
forensic hypnosis **3**:1500–4
see also hypnosis
  forensic image processing **3**:1520–6
  forensic institutes **3**:1241; **4**:1699–704
see also forensic associations
forensic interviews **3**:1554
  see also interviews
forensic light sources (FLSs) **3**:1293, 1294

forensic neuropsychology **4**:1862
see also neuropsychological assessment
forensic odontology **4**:1889–95
see also odontology
forensic pathology **5**:2623
see also pathology
forensic pollen *see* pollen
forensic psychology **4**:2181–3
see also psychological...
Forensic Quality Services-International (FQS-I) **1**:2, 4
forensic radiologist/radiographer role **5**:2234
forensic radiology **5**:2233–9
see also radiology
forensic remains **4**:2090
see also human remains
forensic science education programs **2**:897–901
see also education
Forensic Science Education Programs Accreditation Commission (FEPAC) **2**:899–901, 1017
Forensic Science Service (FSS) **2**:488, 972; **5**:2591, 2592
Forensic Science Society (FSS) **2**:959; **5**:2548
forensic soil science **5**:2377–8, 2383
see also soil analysis
forensic specialists' coordination **1**:414
Forensic Specialties Accreditation Board (FSAB) **3**:1174
forensic speech analysis **5**:2389–90
see also speech analysis
forensic toxicology **2**:595–600; **4**:2012–20
see also toxicology
forensics, term derivation **1**:284
foreseeability, product use **3**:1488–92
forests **5**:2640–1
see also wood
forgeries (documents) **3**:1255–76
forgetting
amnesia distinction **5**:2244
eyewitness memory **2**:1068
forgot-it-all-along (FIA) phenomenon **5**:2246, 2247
forklift trucks (FLTs) **3**:1489
form of particles **4**:2001–7
formal police–citizen encounters **2**:650
formal testing **4**:2173–5, 2175–7
formaldehyde-fixed tissue **2**:1063
formaldehyde storage **4**:2101–2
formalin-fixed tissue **2**:826
formalism (Bayesian networks) **1**:276–80
formation mechanism, friction ridges **3**:1283
formic acid **1**:115, 121
Formica **5**:2352
forms *see* documentation procedures
forward ABO blood group **1**:341
forward commander role **1**:412, 414
forward extrapolation, BAC **1**:69–70
foster parent evaluation **5**:2608
Foucha v. Louisiana, 499 U.S. 946 (1994) **5**:2583
foundation evidence (documents) **2**:840
foundation testimony **3**:1276–7
see also laying the foundation
founder effects, Rhesus blood group **1**:347
fountain pens **5**:2661
Fourier transform infrared (FTIR)
microscopy **4**:1750–7
spectroscopy **2**:990–1; **4**:1936–7, 1942; **5**:2482
see also diffuse reflectance infrared Fourier transform
Fourier transformation **4**:1751
fourth rib method, age determination **1**:181
Foveon sensors **4**:2049
fpc (finite population correction) **5**:2284
FPD (flame photometric detection) **2**:520–1
FPIA (fluorescence polarization immunoassay) **5**:2515
FQS-I *see* Forensic Quality Services-International
fracture matching
paint evidence **4**:1933
toolmarks **5**:2487, 2489
see also physical fit evidence
fracture mechanisms, materials **4**:1683, 1687
fracture surfaces, light globes **3**:1634, 1637–8
fractures
alcohol-induced **1**:104
individualization markers **1**:170
ribs **1**:407–8
skull **1**:402–4
see also lesions of bone
fragile item packaging **4**:1930
fragile watermarks **3**:1275
Fragment Data System (FDS) **3**:1183–4, 1358; **5**:2591
fragmentary bones **1**:154
see also bones
frame shifts, polymorphisms **1**:340
framework of circumstances **2**:971, 974
framing, photography **2**:627
France **2**:561, 783, 1007–8
Franklin, George **2**:787; **4**:1713
Franklin rods **2**:920
Franklin's wood maceration method **5**:2643
fraud **2**:960; **3**:1369; **4**:1968
see also hacking; spoofing
FRE *see* Federal Rules of Evidence
free recall **2**:554, 917, 918
free will **2**:574, 577–9
freebase *see* crack cocaine
freehand simulated signatures **3**:1445
freezer registers **2**:608
freezing method, swabs **1**:321
French Napoleonic Code **2**:561
frequency
distributions (paint) **4**:1944, 1945–6
estimation (mtDNA) **4**:1824–5
fiber evidence **3**:1098–100
parameters (paint) **4**:1943, 1944, 1945–6, 1952
sample matches **3**:1561–2
see also allelic frequency
frequentist approach **5**:2283–4, 2284, 2286–7, 2288
Freud, Sigmund **1**:28; **2**:571, 733; **4**:1712, 2179; **5**:2436
FRI (function of rights in interrogation) **2**:465–6
friction materials, ammunition primers **3**:1190
friction ridge skin
AFIS system **1**:249–53
anatomy **3**:1322–6
embryology **3**:1327–30

features **3**:1283–5
fingerprint comparison **3**:1282–7
fingerprints/DNA interaction **3**:1318–21
interpretation **3**:1277–80
morphogenesis **3**:1322–30
overview **3**:1322–30
persistence **3**:1327
physiology **3**:1326–7
uniqueness **3**:1329–30
see also fingerprints
frontal sinus shape identification **3**:1514
FRR (false reject rate) **1**:329
Fruit Fly Eradication Project, South America **2**:951
frustrated total internal reflection (FTIR) **1**:325
frustules **1**:428; **4**:2002; **5**:2563
Frye test, expert testimony **2**:540
Frye v. United States, 193 F. 1013 (D.C. Cir. 1923) **2**:693, 694, 695, 1001–2; **3**:1095, 1331–1333, 1334
Fry's reagent **5**:2326, 2327
FSAB (Forensic Specialties Accreditation Board) **3**:1174
FSH (follicle-stimulating hormone) **5**:2333
FSS *see* Forensic Science Service; Forensic Science Society
F_{ST} *see* population genetics parameter
FTA cards **2**:1063
FTA paper **1**:320
FTIR *see* Fourier transform infrared; frustrated total internal reflection
FTP servers **1**:145
fuel cell instruments **1**:95
fuel load fire myth **1**:209, 218–9
fuels
ammunition primers **3**:1190
explosions **2**:1023
Fugate, Caril Ann **3**:1475, 1478
full disclosure **2**:774–5
fully method of stature determination **1**:163
fuming techniques **3**:1300–2, 1304, 1305, 1309, 1319
function of rights in interrogation (FRI) **2**:465–6
functional amnesia **2**:786
functional assessments **2**:452–3; **4**:1863, 1865–7
see also functional testing
functional capability
delusions **2**:743–4
hallucinations **3**:1434
functional causes of death **4**:2076
functional lineup size **2**:1074
functional magnetic resonance imaging (fMRI) **2**:725–6, 726
functional testing **2**:448
see also functional assessments
functional tolerance **1**:76
funding graduate programs **2**:901
fungi **1**:292, 301, 423; **2**:828; **4**:2057–67
see also magic mushrooms
funnel approach, child interviews **2**:554
furniture polish debris **3**:1158
furniture spring sagging fire myth **1**:209, 216–7
fuse boxes, vehicles **3**:1636
fuse compartments, switches **2**:924
fuser roller faults, printers **5**:2669
fusion marks, textiles **2**:994
fusion methods, microscopy **4**:1783
FWAs *see* fluorescent whitening agents

G × E interactions **3**:1337–8, 1341
GABA receptors **2**:729, 730
see also gamma aminobutyric acid
gagging **1**:226
gait recognition systems **5**:2612
Galen's theory **3**:1087
Galileo's Revenge: Junk Science in the Courtroom (Huber) **3**:1576
Galton, Sir Francis **3**:1282; **4**:2177
Galton points **1**:251
see also minutiae
galvanized steel **5**:2352
Gamble v. Estelle, 429 U.S. 97 (1976) **5**:2583
gambling **1**:18
gametes **2**:793
gamma aminobutyric acid (GABA) **1**:116
see also GABA receptors
gamma butyrolactone (GBL) **2**:869–70
gamma hydroxybutyrate (GHB) **2**:869–70; **4**:2101
Gardner, R. **4**:1981
Garner v. Tennessee, 471 U.S. 1 (1985) **4**:2070
gas chromatography (GC)
alcohol analysis **1**:85–8, 98
amphetamine analysis **1**:138–9
analytical artifacts **4**:2107
cannabis analysis **2**:434
confirmation testing **1**:297; **2**:595, 597, 598, 599
driver impairment evidence **2**:881
drug profiling **2**:844
drug screening **5**:2522–30
explosion debris **2**:1039, 1040–1, 1042
fiber analysis **2**:991
hair analysis **3**:1427, 1431
paint analysis **4**:1938
plant analysis **4**:2066
postmortem toxicology **4**:2125–6, 2130, 2133
sweat testing **5**:2428
see also gas chromatography-mass spectrometry
gas chromatography-electron capture detector (GC-ECD) **2**:1041
gas chromatography-mass spectrometry (GC-MS)
drug profiling **2**:844–5
drug screening **5**:2526
fire debris analysis **3**:1168
see also gas chromatography
gas wash, explosion debris **2**:1034
gaseous organic bases, inks **3**:1545
gases **3**:1107
see also gas. . .; *individual gases*
gasoline **3**:1149–55, 1165–7, 1168
gastric contents
botanical analysis **1**:427
postmortem redistribution **4**:2097
sampling **4**:2103
time of death determination **5**:2473–4
toxicology **4**:2122; **5**:2499–500
upper airway obstruction **1**:226
gastrointestinal tract **1**:410; **3**:1533; **4**:1899
gauge length, tensile testing **4**:1685

Gaussian mixture models (GMM) **5**:2391
gay panic **3**:1480–2; **5**:2447
GBL *see* gamma butyrolactone
GC *see* gas chromatography; group specific component
GC-ECD *see* gas chromatography-electron capture detector
GC-MS *see* gas chromatography-mass spectrometry
gel pens **5**:2662–3
gelatin fingerprints **1**:330–1
gender
  aggression differences **1**:36–51
  alcohol distribution differences **1**:61
  determination **5**:2354, 2637
  psychopathy **4**:2194
  sentencing defendants **5**:2308–9
  see also sex determination
general acceptance test **2**:694, 695, 1001; **3**:1331, 1333
General Electric v. Joiner, 522 U.S. 136 (1997) **3**:1334–5, 1600, 1621
General Neuropsychological Deficit Scale (GNDS) **4**:1865
general pattern, friction ridge skin **1**:251
general power of attorney **3**:1375
general rifling characteristics (GRC) file **3**:1207
general unknown screening (GUS) **4**:2126, 2128–31
generalized seizures **5**:2298–9, 2299
Generation X **2**:1017
Generation Y **2**:1017
generic code names, tread patterns **3**:1242
genes, definition **2**:795
genetic assays **1**:305–6
genetic diseases **5**:2359
genetic disorders **4**:1871
genetic fitness mechanism **2**:795
genetic markers, Y-chromosomal STRs **5**:2677
genetic matches, Y-chromosomal STRs **5**:2678–9
genetic typing (psychopathology) **4**:2189
genetics **2**:794–6
  behavior and **3**:1335–6
  definition **3**:1336
  population parameter **5**:2367
  studies of **4**:1836
  wildlife forensics **5**:2636–7
  wood **5**:2641–2
  see also DNA. . .; genetic. . .; genomics; genotyping; molecular biology; pharmacogenetics
Geneva conventions **2**:764
genipin **3**:1299
genital examinations **2**:534
genitoanal injuries **1**:238
  see also anal injuries
genocide **4**:1676–7, 1677
genomes, definition **2**:794
genomics **3**:1335–41
genotyping **1**:339; **3**:1337, 1338–40, 1341, 1572
gentian violet **3**:1302, 1307, 1319
genuine signatures **3**:1444, 1447
genus
  plants **1**:423
  wildlife **5**:2637
geoforensics **5**:2378
geographic genotyping **3**:1342–7
geographical distribution, forests **5**:2640–1
geographical origin of drugs **2**:854–5
geographical positioning, mobile phones **3**:1369–70
geology **5**:2378
geophysical survey methods **1**:199, 203
George III, King **3**:1553
Germany **2**:782, 783, 1008–9; **4**:1692–1705
germline mutations **2**:794
GHB *see* gamma hydroxybutyrate
'ghost pills' **5**:2500
Gilligan, James **1**:25
girdle bands, diatoms **2**:749
girls' aggression **1**:37, 43, 44
  see also female aggression
gist memory traces **2**:1067, 1068, 1069
GKTs (guilty knowledge tests) **2**:725
glands
  epidermis **3**:1324
  salivary **4**:1904–5
  sweat **5**:2420, 2421, 2422
Glasgow Coma Scale (GCS) **3**:1459
glass/glass evidence **3**:1348–50
  activity level propositions **3**:1356–7
  analysis/comparison **3**:1349–50
  background/transfer probabilities **3**:1357
  breaking glass **3**:1348–9
  bullet holes **3**:1220–1
  computer software packages **3**:1358
  continuous approach **3**:1355–7
  crazing myth **1**:208, 209, 211–2
  elemental analysis measurements **3**:1354, 1356
  exemplars **4**:1929
  failure analysis procedures **4**:1687
  fingerprint enhancement **3**:1319–20
  firearm discharge residue patterns **5**:2352
  interpretation **3**:1350, 1351–8
  knowledge-based systems **5**:2590–1
  packaging **4**:1929
  particle forms **4**:2004–6
  persistence **5**:2536
  preassessment **3**:1352–3
  refractive index measurements **3**:1353–4, 1356
  source level propositions **3**:1353–6
  statistical evaluation **5**:2590–1
  storage **4**:1929
  technology **3**:1348
  transfer **3**:1357; **5**:2535–6, 2538
  transition temperature **3**:1111
  transport **4**:1929
  two-stage approach to **3**:1353–5
  working propositions **3**:1352–3
  see also light globes/filaments; nonporous surfaces
glia **4**:2202
GLO-I (glyoxalase I) **1**:352–3
global economy issues **3**:1491
Global System for Mobile (GSM) **3**:1360–71
globes, light **3**:1632–8
glove marks **2**:617; **4**:1668, 1669–70
  see also fabric marks
gloves, disposable **2**:615
glucose concentrations, vitreous humor **5**:2498

glucose metabolism **4**:2076–81
glucuronidation **4**:1900
glutaraldehyde **2**:864
glycophorins **1**:344
glyoxalase I (GLO-I) **1**:352–3
GMM (Gaussian mixture models) **5**:2391
GNDS (General Neuropsychological Deficit Scale) **4**:1865
GnRH *see* gonadotropin-releasing hormone
go-cart case **3**:1492–3
gold chloride **4**:1747
golf ball typewriters **5**:2664–5
gonadotropin-releasing hormone (GnRH) **5**:2334–5
Good Samaritan laws **2**:763
'goops', ballpoint pens **5**:2662
government role, product liability **3**:1486
GPR *see* ground-penetrating radar
graduate forensic science programs **2**:898–9, 901
graduation hypothesis, sadism **5**:2318
graffiti cases **4**:1941–2, 1947
Graham v. Connor, 490 U.S. 386 (1989) **4**:2069–70
graminaceous fibers **4**:1978
grammage of paper **4**:1968, 1972–3
grandiose delusions **2**:741–2
grandparent evaluation **5**:2608
graphology **3**:1451
grasses **2**:827; **4**:1978–80
graves **1**:200–2, 203, 204, 206; **3**:1516; **4**:1808, 1966
see also mass graves
gray zone, air bags **1**:55
graze shots **3**:1384
GRC (general rifling characteristics) **3**:1207
'grease ring' **3**:1387
Green River murders **4**:1940–1
Greenberg–Ruback model **2**:644–7
greenbottles *see* blowflies
Greenough stereomicroscope **4**:1762, 1763
grid searches **2**:623, 1026
Griess reaction **5**:2351
grinding (serial number obliteration) **5**:2325
Grisso, T. **2**:452–3, 454
Grisso tests **2**:465–6
grommets **3**:1272
groove impressions **3**:1207
grooves (bullets) **3**:1207
ground disturbances (graves) **1**:203
ground-penetrating radar (GPR) **1**:199, 203
'ground truthing' **1**:204
groundwood pulp **4**:1978
group norms, violence **5**:2273
group specific component (Gc) **1**:354
growth ring analysis **1**:427
see also dendrochronology
GSM (Global System for Mobile) **3**:1360–71
GSRs *see* gunshot residues
GSS *see* Gudjonsson Suggestibility Scales
guanidinium thiocyanate (GuSCN) **2**:804–5
guard hairs **3**:1406
guardian ad litem **4**:1988
guardianships **2**:558; **3**:1371–80
Gudjonsson, Gisli **3**:1591, 1592, 1593
Gudjonsson Suggestibility Scales (GSS) **2**:466–7, 918, 1068; **3**:1591–2, 1593
'guesswork' technique **2**:950
guidance role of CSI **2**:616
guilloche patterns **3**:1267
guilty knowledge tests (GKTs) **2**:721, 725
gun blue **3**:1309
gunpowder particle forms **4**:2007
gunpowder residues **3**:1195–6, 1197
see also gunshot residues
gunshot entries/lesions *see* entrance wounds; gunshot wounds
gunshot residues (GSRs) **3**:1201
analysis **3**:1191–5, 1196–7, 1224
classification **3**:1193–4
colour tests **3**:1191
composition **3**:1193–4
detection **3**:1191–5
interpretation **3**:1193–4
persistence **3**:1192–3; **5**:2537
range determination **3**:1203
sampling **3**:1192–3, 1194
scanning electron microscopy **3**:1192, 1195; **4**:1803–4
SEM/EDX analysis **3**:1192, 1195
transfer **5**:2536–7
see also gunpowder residues
gunshot wounds **3**:1380–401
examination/documentation **3**:1397–8
firearm discharge residue patterns **5**:2353
internal findings **3**:1394–5
interpretation of results **3**:1217–8
lesions in skeletal remains **5**:2562
manner of death **3**:1398
see also entrance wounds; firearms
gunsmoke, definition **3**:1390
GUS *see* general unknown screening
GuSCN (guanidinium thiocyanate) **2**:804–5
Gustafson–Johanson test **1**:183
gustatory hallucinations **3**:1432
gymnosperms **4**:1976
gypsum compounds **2**:964–5
see also dental stone

H_0 (null hypothesis) **3**:1582
H_1-antagonists *see* antihistamines
H_2-antagonists *see* histamines
H_2S (hydrogen sulfide) **4**:2133
habeas corpus petitions **2**:999
hacking, mobile phones **3**:1368–9
hackle marks, glass **4**:2005
Hadfield, James **3**:1553
hair/hair analysis **3**:1415–27, 1427–32
applications of analysis **3**:1429–31
biology **3**:1415–8, 1427
cocaine use **2**:567
color (phenotype inference) **4**:2031
DFC investigations **2**:874
DNA extraction **2**:1063
DNA source **2**:825–6
drug impaired driving **2**:882
forensic processes **3**:1418–24
heat effects **3**:1530

hair/hair analysis *(cont)*
individualization **3**:1423–4
microscopic structure **3**:1418, 1420–2, 1424
morphology **3**:1418
mtDNA analysis **4**:1833, 1834
particle forms **4**:2006
persistence **3**:1418; **5**:2535
physiology **3**:1415–8
pollen evidence **4**:1965
radiocarbon dating **1**:420
sampling hair **4**:2104
toxicology **4**:2122; **5**:2501
transfer **3**:1418; **5**:2535, 2538
see also animals, hair
hair cycle **3**:1417
hair swipe bloodstain patterns **1**:384–5, 385, 390
half-lives (drugs) **2**:870; **4**:2116
hallucinations **3**:1432–5; **5**:2274
see also delusions
hallucinogens **1**:292; **4**:2059, 2064
halogen-type light bulbs **3**:1633
Halstead Impairment Index **4**:1865
Halstead-Reitan Neuropsychological Battery (HRNB) **4**:1873
Halstead–Reitan Neuropsychological Test Battery (HRNTB) **4**:1864–5
hammer beating bloodstain patterns **1**:377
hand... *see* hands
'hand' picking trace evidence **5**:2294
Handbook of Forensic Services (FBI) **5**:2619
handling
crime scene exhibits **2**:624
firearms **3**:1211–5
handling errors, STR profiles **5**:2371–2
'handover' (cellular phones) **3**:1360
hands
anatomical regions **3**:1322–3
embryonic development **3**:1327–8
geometry recognition **1**:327, 335–6
vein pattern recognition **1**:327–8, 336–7
handsets (mobile phones) **3**:1362–7, 1368–9
handwriting **3**:1451–7
characteristics **3**:1441–4
comparison **2**:747; **3**:1438, 1439–40, 1443, 1451, 1452, 1454–6
definition **3**:1437
DNA interpretation **3**:1453
examination **3**:1436–48
identification **3**:1437–9, 1441, 1443
handwritten document alterations **1**:130
hanging **1**:228, 229–32; **5**:2563
hanging drop method (microchemistry) **4**:1746
haplotypes **1**:346; **4**:1824, 1827, 2029
see also mitochondrial DNA
haptenization **5**:2512–3
haptoglobin (Hp) **1**:355
harassment, stalking **5**:2397–400
hard disks/drives **2**:584–5, 587
hard surfaces **1**:365, 388
hard tissue individualization **1**:167, 168
see also bones
hardiness **4**:2142
hardness of materials **4**:1682
hardware, mobile phones **3**:1362–3
hardwoods **1**:426; **4**:1976, 1978, 1979, **5**:2642
Hardy–Weinberg Equilibrium (HWE) **3**:1458; **5**:2368
Hare Psychopathy Checklists **4**:2197–201
harmful dysfunction **4**:2187
harmful use **5**:2413
harmonized conclusion scale, toolmarks **5**:2494
harms reduction, necessity defense **2**:577
Harper v. Washington, 494 U.S. 210, 110 S.Ct. 1028 (1990) **5**:2586
hash function, definition **2**:588
hash value (computers) **2**:587, 588
Hathaway, **4**:2179
hatred/political murders **3**:1476
Haversian canals **5**:2395
Haversian systems, bone **1**:154
Hawkins, Robert A. **3**:1475, 1476
Hayne, H. **2**:658, 659
Hb see heterozygote balance
HCI *see* hydrochloride
HCM *see* hypertrophic cardiomyopathy
HCN poisoning **4**:2133
head acceleration, biomedical engineering **5**:2572–4
head injuries **1**:104, 401–6
see also head trauma; traumatic brain injury
head pattern matching **5**:2614
head-rests (vehicles) **5**:2543
head trauma **1**:263, 266; **5**:2300
see also head injuries
header information pixel defects **3**:1524
headspace
analysis **5**:2524
concentration, ILRs **3**:1139–40
packaging evidence **4**:1930
healing subdural membranes **5**:2343–5
health and safety risks **2**:615
health care **3**:1374–5; **5**:2571–2
see also care issues; medical treatment
Healthcare Directives **3**:1374
hearings
guardianships **3**:1378
in limine motions **3**:1528
hearsay evidence **3**:1465–7
children's testimony **2**:551
cross-examining experts **2**:664
definition **3**:1467
document authentication **2**:841
exceptions to rule **3**:1466–7
foundation testimony **3**:1277
learned treatises **3**:1624
Hearst, Patricia (Patty) **5**:2410, 2411, 2439
heart
amphetamine effects **1**:136, 138
blunt force injuries **1**:409
cannabis effects **2**:434
disease **3**:1472; **4**:2117, 2119
see also cardiac...
heart blood **4**:2097, 2102, 2107
heartwood **5**:2642
heat
chemical processes **3**:1105–12
of combustion **3**:1107
effects on human body **3**:1529–38
energy **3**:1104, 1107

flow **3**:1104
flux **3**:1105
serial number restoration process **5**:2327
transfer **3**:1117–8, 1178
heat release rate (HRR) **1**:219; **3**:1104
heavy drinkers **1**:122, 123
heavy machinery damage **4**:1676
heavy metal analysis **4**:2134
heavy petroleum distillates (HPDs) **3**:1143, 1155–7, 1157, 1158, 1160
height
estimations **3**:1624–32; **5**:2615
human remains **1**:163, 194
see also stature
helmets **5**:2543
helplessness **2**:538
see also learned helplessness
hematogenous osteomyelitis **1**:169
hematological disorders **4**:1849–51
hematomas **1**:399, 404–5; **3**:1534
see also hemorrhages
hemicellulose **4**:1970
hemlock plant **4**:2061–2
hemoglobin **3**:1648, 1652
see also blood
hemoglobin A_{1c} **4**:2078–9
hemoglobinopathies **4**:1849–50
hemolysis **4**:2098
hemolytic disease of the newborn **1**:345
hemorrhages
burns **3**:1532
intracranial **1**:404–6
petechiae **1**:225
pulmonary **1**:227
shaken baby syndrome **5**:2339–40, 2341–5
subarachnoid bleeding **1**:406
see also hematomas
hemorrhagic shock **1**:104
hemothorax **1**:408
hemp fiber **2**:432–3
Hendricks v. Kansas, 521 U.S. 346 (1997) **2**:669; **5**:2578
hepatic...
cirrhosis **4**:2081
coma **4**:2081
failure **4**:2081
insufficiency **4**:2081
see also liver
hepatitis (alcohol-induced) **1**:104–5
herbal medications **4**:2058, 2060
HERG (human ether-a-go-go-related gene) **2**:471
heritability *see* genetics
heroin **4**:1895–6
analytical methods **2**:847–8; **4**:1903
driver impairment **2**:878
metabolism **4**:1900
multiple drug use **1**:109
oral fluid **4**:1913
paraphernalia **5**:2501
postmortem toxicology **4**:2124
profiling **2**:851–2, 852–3, 854, 855
tolerance/dependence **4**:1899
urine testing **2**:865
Herschel, Sir William **3**:1282
heterogeneous immunoassays **5**:2512
heterogeneous line intersections **3**:1594, 1595
heteroplasmy **4**:1834
heterozygosity
alleles **1**:340
chromosomes **2**:795
peak heights **4**:2008
STRs **4**:2075; **5**:2361
heterozygote balance (*Hb*) **2**:807–8, **3**:1567
hexamethylene triperoxide diamine (HMTD) **2**:1037, 1039, 1041, 1044, 1048–9
hexogen (RDX) **2**:1054
HFs *see* human factors
HHV (human herpes virus) **3**:1343
HID (high-density discharge) light **3**:1634
'hidden suicides' **5**:2541
hierarchy of propositions **2**:487–8, 493–4, 973; **4**:1951–2
Higgins v. People, 5 N.Y.2d 607 (NY. 1959) **1**:254
high-density discharge (HID) light **3**:1634
high excitation **4**:2086
high performance liquid chromatography (HPLC) **2**:844; **5**:2526–9
see also liquid chromatography
high performance liquid chromatography with a pendant mercury drop electrode detector (HPLC/PMDE) **3**:1196
high power microscopy **4**:1758–61
high-velocity impact, bloodspatter **1**:376–7
higher limit of quantitation (HLoQ) **4**:2224
higher order mixtures **4**:1839
Hinckley, John **1**:283; **5**:2444, 2445
hindsight test, police use of force **4**:2070
hinge skull fractures **1**:402
Hinkley, Peter **3**:1176–7
histamines (H_2-antagonists) **1**:114
histology **1**:260; **3**:1468–73, 1536; **5**:2394, 2395
histopathology **3**:1468
see also microscopy
historical risk factors **5**:2273, 2599
history documentation
crime scenes **2**:611–2
sexual assault cases **1**:237
history of forensics
autopsies **1**:256–9
bloodstain pattern interpretation **1**:360–1
case assessment and interpretation **2**:484–5
CBRN agent malicious use **2**:500–1
consent to treatment doctrine **2**:450–1
crime laboratory accreditation **1**:1–3
diatomology **2**:749–51
DNA profiling **2**:800–1
electrical engineering **2**:920–1
entomology **2**:934–5
ethanol determination procedures **1**:88
fire modeling **3**:1176–7
genetics–crime relationship **3**:1335–6
guardianships **3**:1373
handwriting comparison **3**:1452
Hare Psychopathy Checklists **4**:2197
histopathology **3**:1468
image processing **3**:1520
interpretation of results **3**:1579–84
microscopy **3**:1468

history of forensics *(cont)*
psychological testing **4**:2177–8
serial murder **5**:2312
shaken baby syndrome **5**:2339
soil analysis **5**:2380
sudden cardiac death **2**:468–9
threat assessment **5**:2460–1
hit/no hit system, DNA databases **2**:679
Hitler Diaries case **3**:1452
HIV *see* human immunodeficiency virus
HLA (human leukocyte antigen) **2**:842–3
HLoQ (higher limit of quantitation) **4**:2224
HMTD *see* hexamethylene triperoxide diamine
HMX *see* octogen
hoaxes (bioterrorism/biocrime) **1**:301
Hochstedler, E. **2**:651
HoG *see* hypoglycemia
hogtying in prone position **1**:228
hole fractures, skull **1**:403
holes in floor fire myth **1**:209, 219–20
holistic approaches **3**:1279, 1287, 1288
home visits, custody evaluation **5**:2605
home washer case **3**:1492
homemade bombs *see* improvised explosive devices
homeschooling **4**:1984
homicide–suicide **2**:903, 905; **5**:2418–9
homicides **1**:26
disguised as suicides **4**:2165, 2167
drug-related **1**:290
elders **2**:903, 905
fire victims **3**:1531, 1537
hanging **1**:232
juvenile perpetrators **4**:1985
mass murder **4**:1674
multiple **3**:1474–9
paint evidence cases **4**:1940–1
poisonous plants **1**:428
radiological diagnosis **5**:2236
school shootings **5**:2455, 2456
sharp force injuries **5**:2646, 2648, 2649, 2651–3, 2654, 2658–9
stalking **5**:2398
traffic fatalities **5**:2541
see also homicide–suicide; serial murder
homogeneous immunoassays **4**:2125; **5**:2511–2
homogeneous line intersections **3**:1594
homoplasmy **4**:1835
homosexual panic **3**:1480–2; **5**:2447
homosexuality, custody evaluation **5**:2609
homozygosity
alleles **1**:340
chromosomes **2**:795
peak heights **3**:(1567; **4**:2008
STRs **4**:2075; **5**:2360
hormonal treatment **5**:2334–5
hospital autopsies *see* clinical autopsies
hospitalization
adverse events (patients) **4**:1690, 1692, 1703
battered child syndrome **1**:267
severe mental illness **2**:649
see also deinstitutionalization
host networks **1**:143
hostage identification syndrome *see* Stockholm syndrome
hostage injuries **4**:2074
hostage taking **5**:2410
hot wax transfer printing **5**:2672, 2673
hotspots, mutations **4**:1827, 1831
House, Robert **2**:728
household paints **4**:1947–50
household substances **2**:864; **4**:2124
see also individual substances
households, custody evaluation **5**:2607
Hp (haptoglobin) **1**:355
HPDs *see* heavy petroleum distillates
HPLC *see* high performance liquid chromatography
HPLC/PMDE (high performance liquid chromatography with a pendant mercury drop electrode detector) **3**:1196
HRNB (Halstead-Reitan Neuropsychological Battery) **4**:1873
HRNTB (Halstead–Reitan Neuropsychological Test Battery) **4**:1864–5
HRR *see* heat release rate
HTTP *see* Hypertext Transfer Protocol
Huber, Peter **3**:1576
human bodies
fire scenes **3**:1131
firearm discharge residue patterns **5**:2353
see also body. . .; human remains
human characteristics **1**:324
see also physical characteristics
human consciousness **2**:577–9
human ether-a-go-go-related gene (HERG) **2**:471
human factors (HFs) **3**:1483–95, 1493
human figure drawings **2**:554
human hair vs animal hair **3**:1404–5
see also hair/hair analysis
human herpes virus type 1 (HHV-1) **3**:1343
human identification *see* identification
human immunodeficiency virus (HIV) **1**:240; **4**:1878, 1880, 1881
human leukocyte antigen (HLA) **2**:842–3
human migration **3**:1342
human observation, facial comparison **3**:1081–5
human remains
age determination **1**:156–61, 179–85; **4**:1890; **5**:2237
anomalies/pathology **1**:194–6
burn injuries **3**:1529–37
entomology **2**:936–40, 940, 942–4, 944
identification **1**:152–79
mass graves **4**:1674–9
PMI determination **4**:2089–92
positive identification **3**:1511–6
race/ancestry determination **1**:163–6, 191–4; **4**:1890; **5**:2238
radiological identification **5**:2237–8
recovery of **1**:199, 200, 204–6; **4**:1889
scalding injuries **3**:1537–8
search scenarios **3**:1495–9
sex determination **1**:155–6; **4**:1890; **5**:2237, 2328–31
species determination **1**:153–4; **4**:1889–90; **5**:2237, 2393–6
stature determination **1**:163, 194; **5**:2237
stratigraphy **1**:201–2
time of death determination **5**:2466–79

trauma analysis **5**:2557–63
see also body. . .; cadavers; graves; human bodies; postmortem. . .; skeletal remains
human resources, search strategies **2**:624
human rights violations **2**:904
human skin *see* skin
human viruses *see* viruses
'humane killers' **3**:1396
Hume-Rothery solution **5**:2327
Hungarian red **3**:1320
hunting weapons **5**:2637–8
Huntington's disease **3**:1337
Hurd guidelines, hypnotic testimony **3**:1503
HV regions, mitochondrial DNA **4**:1828, 1830
HWE *see* Hardy–Weinberg Equilibrium
hydraulic modeling **2**:950
hydraulic theory of social control **2**:649, 652
hydrocarbons **3**:1162, 1163–5; **5**:2524, 2537
hydrochloride (HCI) salt **4**:1912
see also crack cocaine
hydrogen isotopes **1**:309
hydrogen peroxide **2**:864, 865
hydrogen sulfide (H_2S) poisoning **4**:2133
hydrolysis **2**:527, 567, 874
hydropower, Mekong Basin **2**:951
hydrops fetalis **1**:345
hydroxy–ferryl–porphyrin radical (OH $-Fe^{4+}-$P) **3**:1648
hygiene, bomb scenes **1**:415
hymenal injuries **1**:238, 239
hyoid bone fracture **1**:229, 231, 232
hyperbolic discounting **2**:572
hyperglycemia **5**:2498
hypersomnolence **1**:136, 138
hyperspectral imaging **4**:1757
hypertensive left ventricular hypertrophy **2**:471
Hypertext Transfer Protocol (HTTP) **1**:143, 145
hypertonic dehydration **4**:2084
hypertrophic cardiomyopathy (HCM) **2**:472, 473–4
hypertrophy **2**:469, 471–3, 480
hypnosis
admissibility of testimony **2**:736; **3**:1500–3
dissociative disorders **2**:787
interviews **2**:731, 732, 733, 736
legal decisions/opinions **3**:1500–4
memory recall confidence **3**:1501, 1502
trauma **3**:1502
trier of fact prerogatives **1**:286–7
see also suggestibility
hypocalcemia **4**:2085
hypochlorite interference, luminol tests **3**:1652, 1653–4
hypodermis **3**:1323, 1326
hypoglycemia (HoG) **4**:1853, 2080–1; **5**:2302
hypokalemia **4**:2085
hypostasis **2**:703–5; **5**:2468
hypothenar side
feet **3**:1323
hands **3**:1323
hypothermia **4**:2086
hypothesis development, fire scenes **3**:1133–5
hypothetical question opinion evidence **3**:1505
hypotonic dehydration **4**:2084
hysteria **2**:531

IAAC (Inter-American Accreditation Cooperation) **1**:3
IAAI *see* International Association of Arson Investigation
IABPA *see* International Association of Bloodstain Pattern Analysts
IADLs (instrumental activities of daily living) **2**:445
IAI *see* International Association for Identification
ibd *see* identical by descent
IBS (identical by state) **5**:2678
ICCI (intraclass correlation coefficient) **4**:2199
ICCID (Integrated Circuit Card Identifier) **3**:1365
ICD-9 (International Classification of Diseases-9) **5**:2443
ICD-10 (International Classification of Diseases-10) **4**:1725–6, 2188
ice *see* methamphetamine
ICN (increased cycle number) **5**:2628
ICP-MS *see* inductively coupled plasma-mass spectrometry
ictal aggression **5**:2304
IDA (information-dependent acquisition) **4**:2131
IDEIA (Individuals with Disabilities Education Improvement Act 2004) **4**:1728
identical by descent (ibd)
alleles **4**:1810, 1811, 1812
STR profiles **5**:2370
identical by state (IBS), Y-chromosomal STRs **5**:2678
identification **3**:1508–10
animal hairs **3**:1405–9, 1409–10
biological stains **1**:314–9
biometric devices **1**:322, 324
bullets **3**:1207–9
cartridge cases **3**:1209–11
diatoms **2**:754–5
disaster victims **2**:764–72
earprints **2**:891–2
evidence reporting **3**:1509–10
eyewitness lineups **2**:1072–5
fingerprints **3**:1277, 1278–9
firearms identification **3**:1204–11
handwriting **3**:1437–9, 1441, 1443, 1451–2
human remains **1**:152–79; **3**:1511–6
ILR criteria **3**:1146–62
individualization distinction **3**:1580
ink evidence **3**:1546, 1548
living individuals **5**:2611–5
odontological methods **4**:1891–3
plants **1**:424, 425–6, 429
polymorphisms **1**:339–56
radiological methods **5**:2237–8
soil minerals/organic matter **5**:2384–5
wildlife forensics **5**:2637
wood **1**:426–7; **4**:1979
see also DNA. . .; fingerprints; individualization
'identification algorithm' **3**:1514
Identification Commission **2**:771
identification keys (plants) **1**:425
identification pictures *see* mugshots
Identifiler™ system **3**:1518; **5**:2354, 2362
identikits **3**:1511

identity **3**:1508–9
logical approach **2**:975
measures (allelic frequency) **4**:1811, 1812
nonprimary indicators **2**:768
right to possess **2**:764
testing **5**:2357
see also identification
ideological/disciple murders **3**:1476
idiolect **5**:2391
idiomuscular pads **2**:698
idiopathic epilepsies **5**:2299
idiopathic left ventricular hypertrophy **2**:469, 472–3
IEDs *see* improvised explosive devices
IEF (isoelectric focusing) **1**:339
ignitable liquid residues (ILR) **3**:1128, 1137–70
ignition **3**:1109, 1112–5
IHL (International Humanitarian Law) **4**:1677
Ikarian Reefer case **5**:2269
illicit drugs
analysis **2**:844–50
plants used as **1**:425–6
profiling **2**:851–9
see also drug...
illumination sources, microscopy **4**:1758, 1759, 1766–7, 1785
see also light sources
ILR *see* ignitable liquid residues
images
capture (AFIS system) **1**:249
distortion **4**:2037–8
facial comparison **3**:1082–3, 1085
integrity **3**:1522–3; **4**:2036–9
mechanisms (SEM) **4**:1793–800
processing **3**:1520–6
quality **4**:2039, 2053
see also digital imaging; imaging techniques; surveillance images
imaging drum faults **5**:2668–9
imaging signals, SEM **4**:1798–9
imaging submission forms **2**:609
imaging techniques **4**:1756–7, 1886; **5**:2234
imbricate scale pattern (hair) **3**:1404
IMEI *see* International Mobile Equipment Identity
imipramine **1**:111
immature bone lesions **5**:2558
immersed body time of death **5**:2474–8
immune system, opioid effects **4**:1900
immunities
disaster mental health **2**:763
expert malpractice **4**:1664–5
immunoassays
amphetamines **1**:138
analytical artifacts **4**:2106
biological agent detection **1**:304
cocaine analysis **2**:567
drug screening **5**:2511–6
oral fluid **4**:1908
plant analysis **4**:2066
postmortem toxicology **4**:2124–5, 2127, 2128
species determination **1**:154
sweat testing **5**:2426–7, 2428, 2429
urine drug tests **2**:861–2, 866–7
immunochromatographic membranes **1**:314, 316, 317
immunohistochemical methods **3**:1471
impact angle, bloodstains **1**:366–8
impact site, definitions **1**:396
impact testing **4**:1685–6; **5**:2572
see also crash tests
impact velocity, bloodspatter **1**:369–77
impaired control, alcohol dependence **1**:123
impairment of drivers *see* driver impairment
impairment legislation **2**:879, 882
impeachment of experts **2**:662, 663
implanted memories **2**:733–5
impressions
casting **2**:963–7
firearms identification **3**:1205, 1207, 1209–10
foot impressions **3**:1244–8, 1252–5
manufactured items **4**:1668–74
photographic/optical examination techniques **4**:2036–56
water-covered **2**:965
wildlife forensics **5**:2638
see also fingerprints; footwear impressions; tire marks; toolmarks
improved primer extension preamplification (IPEP) **5**:2629–30, 2632
improvised explosive devices (IEDs) **2**:1019–20, 1029, 1045
improvised laboratories **1**:417–8
improvised nuclear devices (INDs) **4**:1884, 1885
impulse control
adolescents **3**:1610
disorders **2**:572; **3**:1227
impulsive behavior **2**:572; **5**:2273, 2398
IMS *see* ion mobility spectrometry
in limine motions **2**:664; **3**:1528
in situ documentation, crime scenes **2**:625, 626
INAA *see* instrumental nuclear activation analysis
inadmissibility
evidence **3**:1603
hypnosis testimony **3**:1500–2
see also admissibility
inbreeding measures **4**:1810–1, 1812
incandescent light **3**:1633–4, 1637
incapacity, definition **2**:445
see also capacity
incarceration, gender disparity **1**:37
see also prisoners
incest **2**:530, 531
incidence, addictions **1**:20
incident light **3**:1310; **4**:2051
incident locations **2**:947–8
see also scene...
incipient ridges, fingerprints **3**:1285
incised wounds **5**:2646, 2649–54
inclusion
mtDNA evidence **4**:1824–7
probabilities (match statistics) **5**:2402
incompetency **2**:440–2, 450, 558–9
see also competency
incomplete encoding **5**:2245
incomplete shifting, livor mortis **2**:705
incomplete suspension hanging **1**:231
inconclusive results
fingerprint comparison **3**:1277, 1278, 1285
handwriting identification **3**:1441

mtDNA matching **4**:1824
incorrect selection of samples **4**:2102–4
increased cycle number (ICN) **5**:2628
incriminating statements **2**:463–4
IND *see* 1, 2-Indanedione
IND-Zn *see* 1, 2-Indanedione-Zinc
1, 2-Indanedione (IND) **3**:1298–9, 1319
1, 2-Indanedione-Zinc (IND-Zn) **3**:1298–9
indans **3**:1153
independence
interpretation of results **3**:1581
STR profiles **5**:2369
independent living capacity assessment **2**:444–9
indicatrix, refractive index **4**:1776–7, 1777
indigenous help, disasters **2**:762
indigenous species, fiber analysis **4**:1979
indirect assessment, workplace threats **5**:2462
indirect method, phenotype inference **4**:2023–8
individual differences, children's testimony **2**:659
individual experts vs institutional laboratories **2**:949
individual rights *see* civil rights
individualization **3**:1508–10
animal blood/tissue **5**:2637
animal hair **3**:1411
DNA evidence **3**:1509
exaggeration error **5**:2403
facial features **3**:1084
fingerprint comparison **3**:1277, 1278, 1285, 1288, 1289
firearms identification **3**:1206
hair **3**:1423–4
handwriting **3**:1451–2
human remains **3**:1514
identification distinction **3**:1580
markers of **1**:166–71
odontological methods **4**:1891–3
soil **5**:2379
speech analysis **5**:2391
subjectivity in **2**:892
textiles **2**:996
tool working surfaces **5**:2486–7
see also identification; living individuals; personal identification
Individuals with Disabilities Education Improvement Act 2004 (IDEIA) **4**:1728
INDs *see* improvised nuclear devices
induced delusions **2**:742
induction training programs **5**:2546
inductively coupled plasma-mass spectrometry (ICP-MS) **4**:1939, 2134
industrial human factors **3**:1483–95
inertia-based switches **1**:52
infanticide **4**:2138–9, 2139
infants **1**:179, 227; **5**:2339–50
see also children; neonatal. . .; perinatal. . .; preschool children; subadult age category; sudden infant death syndrome
infections, sudden death **4**:1855
infectious diseases **1**:196
inference **1**:278; **2**:968, 969–71; **4**:1710, 2021–34
inferior vena cava (IVC) **5**:2496
infinity microscopes **4**:1768
informal police–citizen encounters **2**:650
informant interview, independent living assessment **2**:448
information-dependent acquisition (IDA) **4**:2131
information processing theory **4**:1992
information references, materials **4**:1688
information standard, consent to treatment **2**:451–2
informed consent
civil commitment **5**:2452
history **2**:450–1
independent living assessment **2**:447–8
pediatric psychopharmacology medication **4**:2211–2
sex offender treatment **5**:2336
treatment refusal **5**:2584, 2586, 2587
see also consent
infrared chemical imaging **4**:1757
infrared examination, document line **3**:1596
infrared luminescence (IRL) properties **3**:1259
infrared microscopy/microspectrometry **4**:1751–6
infrared microspectral imaging **4**:1756–7
infrared peaks ratio **3**:1545
infrared reflectance microspectrometry **4**:1755–6
infrared (IR) spectroscopy **4**:1750–7
drug profiling **2**:845
explosion debris **2**:1037, 1046–60
paint analysis **4**:1936–7, 1942
soil analysis **5**:2385
see also Fourier transform infrared
infrared spectrum, specimens **4**:1750
infrared transmission microspectrometry **4**:1751–5
ingestion
alcohol (total amount) **1**:67
alkali (urine tests) **2**:863–4
poisonous plants **4**:2058–9
Ingram, Paul **4**:1712
ingroup/outgroup bias **5**:2306
inhalants **1**:245
inhibition (behavior) **2**:555, 572
inhibition techniques **1**:343; **2**:817, 818
initial characterization, soil **5**:2384–5
initial conference, custody evaluation **5**:2603
initial corners of the kerf **5**:2562
initial observations, crime scenes **2**:612
initial testing **1**:296–7; **2**:861, 866; **4**:1902; **5**:2504
see also screening
initiators (ammunition primers) **3**:1190
injection sites **5**:2500
injuries
alcohol intoxication **1**:100–4, 104–5
blunt force **1**:396–411; **5**:2560–1
burns **3**:1529–37
chemical injuries **3**:1538–9
electrical engineering **2**:923, 927
explosives **3**:1399
radiological diagnosis **5**:2237
scalds **3**:1537–8
sexually motivated assault **1**:237–9
shaken baby syndrome **5**:2339–45
sharp force injuries **5**:2646–60
shotguns **3**:1393–4
see also wounds
injury triad **5**:2339–45

ink analysis **1**:129, 131; **2**:686–9; **3**:1541–5, 1546–52
see also writing instruments
ink-based security measures **3**:1269
ink line microscopic examination **3**:1594–8
inkjet printers **5**:2669–70, 2674
inline collisions **5**:2252–4
inmates
civil commitment **2**:559
mental illness **2**:456
prisonization **2**:573
see also jail populations; prisoners
inorganic gunshot residues *see* gunshot residues
inpatient aggression **1**:41
inpatient civil commitment **2**:556–8
inquisitorial system *see* civil law
insane automatism **1**:255
'insane' suicide **4**:2163–4
insanity
definition **5**:2443
tests **3**:1552–3, 1556
see also mental...
insanity defense **1**:254–5, 283, 284; **3**:1552–6
acquittees **2**:559; **5**:2579
civil commitment **2**:559
competency to stand trial **2**:459
delusions **2**:744
dissociative identity disorder **2**:788
expert testimony **1**:286
genotyping applications **3**:1338
hallucinations **3**:1435
mandated treatment **5**:2579
maternal filicide **4**:2139
multiple murders **3**:1476
not guilty by reason of insanity **5**:2581, 2583
plea history **5**:2443–4
temporary insanity **5**:2443–8
waiver capacity **2**:442
Insanity Defense Reform Act **2**:441; **5**:2444
insects **2**:827, 934–45
insight (patient's) **3**:1463
inspections, laboratories **1**:13, 16
instantiated variables **1**:280
instantiation **3**:1584
instars **2**:940, 942
see also larval development stages
institutes *see* forensic institutes
institutional laboratories **2**:949
institutional settings, elder abuse **2**:914
see also long-term care facilities
institutionalization *see* hospitalization; prisonization
instruction failures, eyewitnesses **2**:1073
instructions to jury, expert opinion **3**:1607–8
instrument marks *see* toolmarks
instrumental activities of daily living (IADLs) **2**:445
instrumental nuclear activation analysis (INAA) **1**:150, 151
instrumented immunoassays **5**:2426–7
insulin treatment **4**:1853
insurance
claim management (ecology) **2**:946
medical malpractice cases **4**:1692
see also claims
intaglio printing **3**:1262
integrated air bag systems **1**:53, 55
Integrated Circuit Card Identifier (ICCID) **3**:1365
integrative functions, neuropsychology **4**:1867
integrity of images **3**:1522–3; **4**:2036–9
integument injuries **1**:396–400
intellectual disability *see* mental retardation
intelligence
adjudicative competence/culpability **2**:546
capacity to waive Miranda rights **2**:466
interrogative suggestibility **3**:1592
mental retardation diagnosis **4**:1732–3
method intelligence **2**:600
intelligence collection/analysis *see* intelligence work
intelligence quotient (IQ) **4**:1729, 2177–8, 2181
mental retardation **4**:1726, 1732–3, 1734
tests **4**:2177–8, 2181
intelligence sources *see* intelligence work
intelligence tests **4**:2177–8, 2181
children **4**:1870
mental retardation measurement **4**:1732–3, 1734
Miranda rights waiver **2**:465
neuropsychological assessment **4**:1864, 1870
see also intelligence quotient
intelligence work
CBRN forensic analysis **2**:505–6
DNA databases **2**:681–2
fibers/textiles **2**:992–6
footwear impressions **3**:1248–52
workplace threats **5**:2461
intentional ingestion, poisonous plants **4**:2058–9
intentional traffic fatalities **5**:2541
interactive component, capacity assessment **2**:453
interactive virtual reality **5**:2263–4
Inter-American Accreditation Cooperation (IAAC) **1**:3
interatrial septum **2**:480
inter-batch variation, illicit drugs **2**:857
see also batch variations
Intercept® collector **4**:1907
interference colors **4**:1779, 1780
interference figures **4**:1783
interfering substances, luminol tests **3**:1651–2, 1653–4
intergenerational transmission model **4**:1992
interictal abnormalities **5**:2300
interictal epileptiform discharges **5**:2300
interictal violence **5**:2305
interim ceiling principle **2**:497
interindividual variations, alcohol elimination **1**:68–9
interior environment, vehicles **5**:2571
interlock devices **1**:97
interlocutory orders **2**:1010
intermediate ballistics **3**:1201
intermediate-range gunshots **3**:1392–3; **5**:2351
intermediate targets (projectiles) **3**:1384
intermittent explosive disorder **5**:2301
internal ballistics **3**:1200–1
internal compulsion **2**:571, 572, 573
see also compulsion

internal findings **1**:260; **5**:2478
  burns **3**:1533–5
  gunshot wounds **3**:1394–5
  scalds **3**:1537
internal quality control (IQC) solutions **4**:2223, –4, 2224–5
internal standards (IS) **2**:600, 865; **3**:1141; **4**:2221
internal workplace threat assessment **5**:2464
internalized false confessions **2**:589; **3**:1591
international accreditation program (ASCLD/LAB) **1**:2, 3–8
International Association of Arson Investigation (IAAI) **2**:922; **3**:1172–3, 1174
International Association of Bloodstain Pattern Analysts (IABPA) **1**:395–6
International Association for Identification (IAI) **2**:959; **3**:1288; **5**:2622
International Classification of Diseases-9 (ICD-9) **5**:2443
International Classification of Diseases-10 (ICD-10) **4**:1725–6, 2188
international collaboration, nuclear forensics **4**:1887
International Humanitarian Law (IHL) **4**:1677
International Ink Library **3**:1543
International Mobile Equipment Identity (IMEI) **3**:1366, 1369
International Organization for Standardization (ISO) **4**:1971
  see also ISO standards
international policies, illicit drugs **2**:854–5
International Society of Forensic Genetics (ISFG) **5**:2356
Internet **1**:141–8, 247; **4**:2010; **5**:2619–26
Internet Protocol (IP) **1**:141, 142, 146
Internet Relay Chat (IRC) servers **1**:144
Internet Service Providers (ISPs) **1**:142
interoperability, AFIS system **1**:252
interpersonal conflict **2**:906
interpersonal trust **3**:1591, 1592
interpersonal violence **1**:26, 30
Interpol **2**:679; **5**:2357–9
interpolation, crash tests **5**:2569–71
interpretation
  amphetamine toxicology **1**:138
  benzodiazepine concentrations **1**:296
  blood-alcohol concentration **1**:58–80
  bloodstain patterns **1**:360–96
  cocaine use **2**:566–7
  DNA profiling **2**:806–8
  documents **3**:1453, 1456
  earprints **2**:891–6
  facial comparison **3**:1084
  facial reconstruction **3**:1088
  fibers **3**:1097–101
  fingerprints **3**:1277–80
  glass evidence **3**:1350
  gunshot wounds **3**:1217–8
  history of **2**:484–5; **3**:1579–84
  ink evidence **3**:1546–52
  light globes/filaments **3**:1637–8
  logical approach **2**:968–76
  luminol tests **3**:1653
  mtDNA **4**:1823–32
  observer effects **3**:1575
  opioid data **4**:1902
  oral fluid **4**:1908–16
  paint **4**:1943–53
  pollen data **4**:1956–9
  principles of **2**:971–2
  printed documents **3**:1557–61
  of results **2**:492, 494
  statement structure **2**:974
  statistical **3**:1566–73
  STR profiles **5**:2365–76
  toxicological **1**:138; **4**:2115–9
  urine drug tests **2**:865–6
  see also case assessment and interpretation
interrogations **3**:1586–90
  adjudicative competence/culpability **2**:546–7
  confession evidence **2**:590–1, 592
  expert testimony **3**:1589
  Miranda rights waiver **2**:463–7
  reforms **3**:1588
  truth serum **2**:736
  videotaping **3**:1588
  see also confession evidence; interviews
interrogative suggestibility **2**:466–7, 591, 917–8; **3**:1590–3
intersecting lines *see* line intersection (documents)
interventions
  disaster mental health **2**:761
  elder abuse cases **2**:908–9
  radiology **5**:2234
interviews
  child sexual abuse **2**:532–4
  child witnesses **2**:550
  children's suggestibility **2**:553–4
  contrasted with interrogations **3**:1586
  independent living assessment **2**:448
  psychological autopsy **4**:2165
  sensitivity **4**:2165
  suggestibility **2**:553–4, 1068
  see also interrogations
interwriter variation, handwriting **3**:1437, 1441–2
intestinal injuries **1**:410
intimate partner threats **5**:2463–4
intine of pollen **4**:1955, 1958
intoxication **1**:99–105; **5**:2415, 2447
intra-batch variation, illicit drugs **2**:857, 858
intraclass correlation coefficient (ICCI) **4**:2199
intracranial hemorrhages **1**:404–6
intradermal bruises **1**:398, 400
intrafamilial child sexual abuse **2**:530
intraindividual earprint variations **2**:894–5
intravenous drug use **1**:137; **2**:565; **4**:1900
  see also administration routes (drugs)
intricate printing, document security **3**:1267
introns **1**:340
inventories, fire scenes **3**:1127
investigations
  bioterrorism/biocrime **1**:307–9
  CBRN agents **2**:500–6
  crime scenes **2**:614–9; **4**:1674–**5**:2480–1
  DNA database use **2**:681–2
  drowning cases **2**:755
  ecological processes **2**:948
  entomological **2**:934
  explosion debris **2**:1028–31, 1045

investigations *(cont)*
explosion scenes **2**:1019–27
explosions **2**:1019–27; **3**:1136–7, 1137, 1173, 1175
fire myths **1**:207–23
firearms **3**:1216–8, 1219; **4**:1678
handwriting comparison **3**:1454–6
histological **3**:1468–73
illicit drug profiling **2**:855–7
mass graves **4**:1674–9
paint **4**:1931, 1939
phase of crime investigation **2**:495
shaken baby syndrome **5**:2347–8
sole database use **3**:1242–3
see also autopsies; crime scene investigator; fire investigation; scene investigation
investigative opinion **2**:487, 488, 492
see also investigations
investigative system of jurisprudence *see* civil law
invisibility of soil **5**:2381
invisible watermarks **3**:1275
involuntariness (compulsion) **2**:571
involuntary hospitalization *see* civil commitment
involuntary medication **2**:457
iodine **3**:1304, 1305, 1319
ion chromatograms **3**:1143–4, 1153, 1167
ion enhancement, oral fluid **4**:1908
ion mobility spectrometry (IMS) **2**:519–20, 1028; **3**:1196
ion suppression, oral fluid **4**:1908
ion-trap detectors **5**:2525
ionic pathway, materials **4**:1685
ionization, mass spectrometry **2**:595, 596–7; **4**:2126
IP *see* Internet Protocol
IPEP *see* improved primer extension preamplification
iPod forensics **2**:586–7, 587
ipse dixit evidence **3**:1599–600, 1620, 1621
IQ *see* intelligence quotient
IQC *see* internal quality control
IR spectroscopy *see* infrared spectroscopy
IRA *see* Irish Republican Army
Iraq **2**:501, 1045; **4**:1678
IRC (Internet Relay Chat) **1**:144
iris
color (phenotype inference) **4**:2026, 2029–30, 2031
pharmacological excitability **2**:701–2; **5**:2468
recognition **1**:327, 334–5
iris printing **3**:1261, 1267
Irish Republican Army (IRA) **4**:1941
IRL (infrared luminescence) **3**:1259
IRMS *see* isotope ratio mass spectrometry
iron, serial number restoration **5**:2326
iron acceptor **3**:1215
iron dissolution **3**:1215
iron traces (firearms) **3**:1212–5
irrebuttable presumption **2**:1007
irresistible impulse standard **1**:283, **3**:1553; **5**:2444
IS *see* internal standards
ischemic heart disease **3**:1472
ISFG (International Society of Forensic Genetics) **5**:2356
island problem **3**:1583
ISO standards **4**:2219–20
ISO 9001 **1**:12, 15
ISO/IEC 17011 **1**:3, 4
ISO/IEC 17020 **1**:12, 15, 16
ISO/IEC 17025 **1**:3–4, 5, 11–2, 13, 14, –16
see also International Organization for Standardization
isoelectric focusing (IEF) **1**:339
isoelectric point, protein polymorphisms **1**:339
isoenzymes **1**:17, 349
see also acid phosphatase
isolated network analysis **1**:148
isolation methods
confirmation testing **2**:598–9
ignitable liquid residues **3**:1138–45
isoparaffinic hydrocarbons **3**:1162
isopropanol **1**:121, **4**:2095; **5**:2425
isotonic dehydration **4**:2084
isotope ratio mass spectrometry (IRMS) **2**:522–3; **3**:1424
isotopes **1**:309; **2**:855
isotrope particles **4**:1770, 1777
ISPs (Internet Service Providers) **1**:142
Italy **2**:782, 1009
item response theory (ITR) **4**:2199
IVC (inferior vena cava) **5**:2496

J.-LJ v. R [2000] 2 SCR 600 [10] **2**:1004
Jackson v. State, 553 So. 2d 719 (Fla. 4th DCA., 1989) **4**:2163
Jacksonian seizure **5**:2298
Jaffe test **1**:318
jail populations **2**:652
see also inmates; prisoners
Janssen, W. **3**:1471–2
Japan **5**:2507–8
JCV (polyomavirus JC) **3**:1342–6
jealousy **2**:742, 743; **5**:2399, 2418
Jeffrey's wood maceration method **5**:2643
Jewell, Richard **4**:2158
Johnson & Wales library resources **5**:2620
Joiner v. General Electric see General Electric v. Joiner
joint legal custody **5**:2609
joint ownership of property **3**:1376
joint physical custody **5**:2609
Jones v. Smith, 1 S.C.R. 455 (1999) **2**:670
Jones v. Vitek, 445 U.S. 480 (1980) **5**:2583
Jonestown murders **3**:1476
Joule, James P. **3**:1104
Joule's effect **3**:1634
journal resources **5**:2621–2
see also peer reviews; professional literature
judges *see* judiciary
judgment
capacity assessment **2**:453
eyewitness identification **2**:1073
neuropsychological assessment **3**:1462
sampling **5**:2293
values **4**:2187
judicial context **1**:278–9
see also judiciary; juries; legal...
judicial experts **2**:1010, 1011

judicial notice **2**:665; **3**:1565, 1601–2, 1624
judiciary
  behavioral evidence **1**:284–5
  instructions to juries **3**:1603–4
  see also courts; magistrates
juge d'instruction see examining magistrates
junk DNA **2**:795
juries
  biases **3**:1604, 1605
  child witnesses, reactions to **2**:551–2
  complex evidence evaluation **3**:1605
  composition diversity **3**:1604–5
  comprehension assistance **3**:1605–6
  confession evidence **2**:593
  decision making **3**:1602–6; **4**:1819
  defendant appearance **2**:980
  deliberations **3**:1604–5
  extralegal information **3**:1603
  eyewitness testimony knowledge **2**:1078
  instructions on expert opinions **3**:1607–8
  match statistics use **5**:2404–6
  memory assistance **3**:1605–6
  notetaking during trial **3**:1606
  nullification **1**:285; **5**:2435
  performance improvement **3**:1605
  prejudice reduction **3**:1605
  questionnaires **3**:1605
  racial composition **3**:1604–5
  racial prejudice **3**:1604
  subtle biases **3**:1604
  trial evidence evaluations **3**:1604
  weight of expert evidence **2**:1016–7
  see also jury trials
jurisprudence systems *see* adversarial justice system; civil law; criminal law
jurors *see* juries
jury trials **2**:1005
  adversarial system **1**:23
  common law systems **2**:998, 999
  demographic factors in sentencing **5**:2306–11
  evidence comprehension **3**:1605–6
  evidence evaluations **3**:1604
  learned treatises **3**:1624
  see also juries
justice
  personalized **4**:2020
  restorative **4**:1719
  see also adversarial justice system; civil law; criminal law; juvenile justice; legal context
justification defense **2**:577
juvenile courts **2**:542–43 **3**:1340–1
  see also juvenile justice
juvenile justice **2**:547–8; **3**:1608–12, 1612–5; **4**:1985–7
  see also juvenile courts
juveniles
  adjudicative competence **2**:543–6
  ADM disorder studies **4**:1877–82
  in adult criminal courts **2**:547–8; **3**:1612–5
  attention deficit hyperactivity disorder **2**:547
  capacity to waive Miranda rights **2**:466
  culpability **2**:543–6, 547
  death penalty abolition **2**:717–8
  as defendants **2**:542–9
  depression **2**:547
  developmental disorders **2**:547
  false confession vulnerabilities **2**:591–2
  offender's right to treatment **5**:2582–3
  pornography involving **1**:190
  pretrial proceedings **2**:547
  sentencing **5**:2307–8
  social anxiety disorder **2**:547
  trial proceedings **2**:547
  see also adolescents; children; young adults

K see control samples; reference samples
k-coefficients **4**:1811
Kansas v. Hendricks, 521 U.S. 346 (1997) **2**:669; **5**:2578
Katyn Forest mass graves **4**:1675, 1678
kava plant **4**:2064
KBSs (knowledge-based systems) **5**:2590–3
Kempf's Disease *see* homosexual panic
Kendra's Law **2**:558
keratin **3**:1417
keratinization **3**:1326
keratinocytes **3**:1325, 1326
kerf floor injury **5**:2562
Kerley–Ubelaker method **1**:183
kerosene **3**:1157–8, 1160
ketamine **1**:292; **2**:872; **4**:2131
ketone bodies **2**:864; **4**:2079
'Keumdong No 5' oil spill **2**:948
keypoint matching, earprints **2**:893
keys *see* identification keys
keystoning **4**:2037
khat plant **4**:2061
kidnapping, parental **5**:2609
kidney(s)
  blunt force injuries **1**:410
  diseases **4**:2094, 2117
  drug testing in urine **2**:860
  function disturbances **4**:2082–4, 2087
  postmortem toxicology **4**:2122; **5**:2500
  sampling **4**:2103
Kidwell, D.A. **5**:2423
kinetic interaction of microparticles in solution (KIMS) **5**:2516
Kingston, C.R. **3**:1580, 1581
kinship cases, Identifiler™ **3**:1518
kinship coefficient **5**:2367
  see also population genetics parameter
Kirchhofrosen **2**:703
Kirk, P.L. **1**:360, 361; **3**:1580, 1581
'kitchen sink' experimentation **2**:951–2, 952
Klein v. Rennie, 462 F. Supp. 1131 (1978) **5**:2585
Klindt, Joyce **3**:1562
knives
  packaging **4**:1929
  wounds **5**:2647–9, 2649–54
  see also sharp force injuries
knocking (gasoline identification) **3**:1155
knowing waiver of rights, definition **2**:464
knowledge-based systems (KBSs) **5**:2590–3
Knowledge of Eyewitness Behavior Questionnaire (KEBQ) **2**:1078
knowledge, skills and abilities (KSA) **5**:2548, –9

known writing *see* exemplars, writings
Köhler illumination **4**:1758, 1766, 1769
Koss case **5**:2433
Kraft Ebing, Richard von **5**:2312
kraft pulp process **4**:1978
Kramers–Kronig transformation **4**:1756
Kromekote paper technique **3**:1595–6
KSA *see* knowledge, skills and abilities
Kumho Tire v. Carmichael, 526 U.S. 137 (1999) **2**:1002, 1015; **3**:1093, 1600, 1619–21

L. C. v. Olmstead (98–536) 527 U.S. 581 (1999) **5**:2577
L/*D* ratio, aminoacids **1**:184
LAB *see* Laboratory Accreditation Board
labels
  entomological evidence **2**:944
  textiles **2**:995
laboratories
  accreditation **1**:1–8, 10–6; **4**:2220
  analyst's confirmation bias **3**:1577
  bomb scene management **1**:417–8
  CBRN scenes **2**:503
  chain of possession **2**:499
  chemical warfare agents analysis **2**:522–3
  child sexual abuse examinations **2**:534
  drug-facilitated crime examinations **5**:2506–7
  ecological evidence **2**:949
  errors (STR profiles) **5**:2371–2
  explosion debris analysis **2**:1028–60
  fire debris analysis **3**:1137–70
  firearms **3**:1216–8
  footwear casework **3**:1241
  light globe examinations **3**:1637
  management/operations criteria **1**:6–7
  mtDNA analysis **4**:1836–7
  paint examinations **4**:1935
  quality management **5**:2493–4
  quality triangle **3**:1171–2
  sampling trace evidence **5**:2291–2
  submission forms **2**:610
  toolmark examinations **5**:2494
  urine drug tests **2**:861–2
Laboratory Accreditation Board (LAB) **3**:1173
  see also American Society of Crime Laboratory Directors
Laburnum anagyroides **4**:2062–3
Lacan, J. **1**:30
lacerations **1**:400
  abdominal **1**:410
  anal intercourse **1**:239
  lung **1**:408
lactic acid (lactate) **4**:2078
lactic acidosis **4**:2080
LaFarge, Marie **4**:2120; **5**:2504
Lambertian surface reflection **4**:2051
Lamendin test **1**:182–3, 183
laminate-based document security **3**:1271
laminated plastic **5**:2352
land impressions **3**:1207
landform, soil analysis **5**:2386
lands (bullets) **3**:1207
landscape analysis **1**:199
Langerhans cells **3**:1326
language deficits **4**:1866
language skills **3**:1462
large consignment sample size **5**:2284–6
larval development stages **2**:936, 937, 940–1, 942, 943, 944–5
laryngeal cartilage fracture **1**:229, 231, 232
laser-engraved document security **3**:1273
laser-induced breakdown spectroscopy (LIBS) **4**:1939
laser printers **3**:1560; **5**:2667–9, 2674
laser profilometry **3**:1598
lasers
  bullet trajectory **3**:1220
  fingerprint enhancement **3**:1318
latent fingermarks **3**:1292, 1294, 1296, 1309, 1310; **4**:1669
  see also fingermarks
latent fingerprints **1**:249; **2**:604–6; **5**:2638
  see also fingerprints
latent interval *see* lucid interval
lateral expansion, handwriting **3**:1442
law *see* legal...
law enforcement agencies **2**:503, 614–5, 855–7; **3**:1241, 1590–3; **5**:2460–1
  see also police
Law Enforcement Assistance Administration (LEAA) **1**:1, 2, 207–8, 209–10
laws of probability **2**:969
lay witnesses **2**:758–9, 1003, 1006
layer structure, paint **4**:1933, 1935, 1943–4, 1944
laying the foundation **2**:498–9, 746, 840, **3**:1276–7, 1624
LC *see* liquid chromatography
LC-MS *see* liquid chromatography-mass spectrometry
LC-MS/MS (liquid chromatography tandem mass spectrometry) **5**:2529–30
LCN *see* low copy number DNA
LCP (life-course persistent) conduct disorder **1**:45
LCV (leucocrystal violet) **3**:1320
LEAA *see* Law Enforcement Assistance Administration
lead tests **3**:1220
leading questions **2**:656, 657, 758
learned helplessness **1**:271; **5**:2411, 2432, 2439
learned treatises as evidence **2**:664; **3**:1467, 1623–4
learning disabilities **4**:1872, 1873
  see also mental retardation
learning function, neuropsychology **4**:1866
learning research, alcohol dependence **1**:124
leather gloves **4**:1670
LEDs *see* light-emitting diodes
Leeuwenhoek, Antony van **4**:1791
left ventricular hypertrophy (LVH) **2**:469, 471–3
Legacy program, ASCLD/LAB **1**:2, 12, 15, 16
legal applications *see* legal context
legal assistance, elders **2**:909
legal cases **2**:922–9
  see also individual cases
legal challenges, fingerprint comparison **3**:1289
legal concepts **2**:460
legal context
  applied psychology **4**:2181–3

behavioral science **1**:282–8
capacity assessment **2**:447
child sexual abuse accommodation syndrome **2**:540–1
compulsion **2**:574, 576–9
disaster mental health **2**:762–3
DNA databases **2**:683–4
elder abuse cases **2**:910
expert reports **5**:2269–70
insanity **1**:283
product liability **3**:1486–7
rape trauma syndrome **5**:2241–2
seizures and aggression **5**:2304–5
severe mental illness **2**:649
sex offender treatment **5**:2336
Stockholm syndrome **5**:2412
syndromes **5**:2412, 2431–40
therapeutic jurisprudence **5**:2449–54, 2580–3
truth serum **2**:735–6
young offenders **3**:1611
see also adversarial justice system; civil law; court. . .; criminal law; juries; medicolegal autopsies
legal counsel waiver **2**:442, 462
legal method-scientific method comparison **5**:2296–7
legal professionals **2**:442, 462; **5**:2606
see also individual types
legal-psychological perspectives **4**:2182–3
legal system *see* legal context
legislation, drug impaired driving **2**:879–81 882
see also individual legislation; legal context
Leitspur Principle **2**:993
length heteroplasmy, mtDNA **4**:1835
length measurement *see* height
leniency bias **3**:1604
lenses
distortion **3**:1625; **4**:2038
electron microscopy **4**:1796
evidence photography **2**:627–8, 632, 633; **4**:2038
high power microscopy **4**:1758–9
light microscopy **4**:1765, 1768
lesions of bone **1**:168, 170; **5**:2557–60, 2560–3
see also fractures
Leskie, Jaidyn **2**:818–9
Lessard v. Smith, 349 F. Supp. 1078 (E.D. Wis.1972) **5**:2577
lethal cases
dosage of ethanol **1**:77–8, 121
gunshot injuries **3**:1384–5
medical malpractice **4**:1692, 1699, 1701–3, 1705
see also death; fatalities
letterpress **3**:1264
leucocrystal violet (LCV) **3**:1320
leukemia **4**:1850–1
Lewis antigens **1**:342
LH (luteinizing hormone) **5**:2333
liability **2**:451, 762–3; **4**:1692
see also product liability
Libet, Benjamin **2**:578
libriform fibers **4**:1978
LIBS (laser-induced breakdown spectroscopy) **4**:1939
lice, DNA source **2**:827
lie detection **1**:286–7; **2**:590, 592–3
see also deception detection; polygraph testing; truth serum
Lie Detection and Criminal Interrogation (Inbau) **3**:1586
life-course persistent (LCP) conduct disorder **1**:45
life cycle, flies **2**:940–1
lifetime, product **3**:1485
lifting evidence
bloodstain patterns **1**:386–8
document line intersection **3**:1595–6
fibers **2**:986, 993, 994
fingerprints **3**:1319
footwear impressions **2**:620
ligature strangulation **1**:229, 232; **5**:2563
light, properties **3**:1293–4, 1310; **4**:1764–5
see also light sources; lighting
light bulbs **3**:1633–4
see also light globes/filaments
light drinkers *see* social drinkers
light-emitting diodes (LEDs) **3**:1634; **4**:1766; **5**:2668
light examination, forgeries **3**:1258–9
see also light microscopy; light sources; lighting
light globes/filaments **3**:1632–8
light microscopy (LM) **4**:1743, 1762–89
design concepts **4**:1764
fiber analysis **2**:988–9, 989; **4**:1975–80
Fourier transform infrared **4**:1750–7
toolmark examination **5**:2487–9
see also brightfield microscopy; comparison microscopy; fluorescence; microscopy; polarized light microscopy; stereomicroscopy
light petroleum distillates (LPDs) **3**:1155
light signals, SEM **4**:1798, 1800
light sources
fingerprint enhancement **3**:1318
shooting scene investigation **3**:1222
specimen spectral response **4**:2041–2
spectral distribution **4**:2041
stain identification **1**:314, 315, 317
see also illumination sources; light globes/filaments
lighting
crime scene photography **2**:627, 633–4, 637–9, 642–3
document examination **1**:129, 132
optical enhancement modes **4**:2042–8
photographic definition **4**:2050
specialized photographic techniques **4**:2050–3
see also light, properties; light sources
lightning experiments **2**:920
lignin **4**:1970; **5**:2643
likelihood ratios (LRs) **2**:487, 970, 971, 972; **3**:1579, 1582–3, 1631
CAI **2**:490, –1, 492
DNA mixture interpretation **4**:1838, 1839, 1840
document interpretation **3**:1558, 1559, 1560, –1
hierarchy of propositions **2**:973
identification/individualization **3**:1508–9, 1510
ink evidence interpretation **3**:1550
match statistics **5**:2402
missing persons **4**:1815
mtDNA evidence **4**:1824, 1825–7

likelihood ratios (LRs) *(cont)*
paint evidence **4**:1951–2
paternity **4**:1814; **5**:2374–5
random man not excluded **5**:2373, 2374
sample matches **3**:1562, 1565
significance probabilities relationship **3**:1583
single-source identification **5**:2366–73
statement structure **2**:974
toolmark examination **5**:2493
likelihoods **2**:487
limbic system **4**:2203
limit of detection (LOD) **1**:126–7; **4**:2224
limit of quantitation (LOQ) **1**:83, 126–7
limited power of attorney **3**:1375
Lindbergh kidnapping case **1**:427
line intersection (documents) **3**:1594–8
see also alterations; erasures; ink analysis; obliterations; paper, analysis
line search strategy **2**:623
linear fractures, skull **1**:402
linear prediction cepstrum coefficients (LPCC) **5**:2391
linear predictive coding (LPC) **5**:2390
linear scales, evidence photography **4**:2038–9
lines of demarcation fire myth **1**:209, 214–6
lineups **2**:1072–5, 1077–8
see also eyewitness testimony
linguistic complexity, cross-examination **2**:656–7, 660
'linked' scenes, footwear intelligence **3**:1251
lipid-reducing medications **1**:114
lipomatous hypertrophy **2**:480
liquefied petroleum (LP) gas **3**:1107
liquid accelerant fire myth **1**:219–20
liquid blood, characteristics **1**:362
see also blood
liquid chromatography (LC) **2**:597–8
explosion debris **2**:1039, 1041
postmortem toxicology **4**:2125, 2126–7, 2128, 2130, 2131
liquid chromatography tandem mass spectrometry (LC-MS/MS) **5**:2529–30
liquid chromatography-mass spectrometry (LC-MS) **2**:595, 597; **4**:2066
drug classes **1**:138–9, 297; **2**:434
drug impaired driving **2**:881
drug profiling **2**:845
drug screening **5**:2529–30
hair analysis **3**:1431
oral fluid **4**:1908
postmortem toxicology **4**:2107, 2126–7, 2131
liquid crystal shutters (laser printers) **5**:2668
liquid–liquid extraction (LLE) **4**:2125; **5**:2516–8
liquids
blood characteristics **1**:362
combustion **3**:1108
packaging/storing samples **4**:1929
transporting samples **4**:1930
see also ignitable liquid residues; *individual types*
listening efforts, earprints **2**:895
literature review, SMI criminalization **2**:650–2
lithium **4**:2206, 2207
lithography **3**:1261
live ammunition **3**:1219–20
see also ammunition
live investigations **2**:587–8
see also investigations
liveness detection, biometric devices **1**:328, 336
liver
alcohol-induced injury **1**:104–5
blunt force injuries **1**:410
diseases **4**:2094, 2117
function alterations **4**:2081–2
postmortem toxicology **4**:2094, 2097, 2103, 2121; **5**:2499
sampling **4**:2103
livestock stunners **3**:1396
lividity **2**:703–5
living individuals
age determination **1**:188–90; **4**:1890–1
blood alcohol sampling **1**:88–9
concealed/embedded object detection **5**:2239
identification techniques **5**:2611–5
radiological diagnosis **5**:2236–7
living wills **2**:439; **3**:1374
livor mortis **1**:362; **2**:703–5
see also postmortem changes
LL-DOP *see* long products from low DNA quantities DOP
LLE *see* liquid–liquid extraction
LLOQ *see* lower limit of quantitation
LM *see* light microscopy
LNNB-CR *see* Luria-Nebraska Neuropsychological Battery-Children's Revision
local anesthetics **2**:563
Locard, Edmond **3**:1284
Locard exchange principle **5**:2378, 2480
Locard's Law **2**:614
location of char fire myth **1**:213
loci designation, STRs **5**:2355–6
LOD *see* limit of detection
Loftus, Elizabeth **2**:1066
log documents **2**:608, 611–3; **4**:1676
logging mechanisms, network analysis **1**:142, 143, 146
logical approach
evidence interpretation **2**:968–76
identity and **2**:975
long-acting GnRH agonists **5**:2334–5
long products from low DNA quantities DOP (LL-DOP) **5**:2629
long QT syndromes (LQTS) **2**:470–1, **4**:1851–2, 1853
long-term care facilities **2**:904, 905, 914
Long-Term Offender (LTO) legislation (Canada) **2**:670–1
longdrop (judicial) hanging **1**:231
longitudinal skull fractures **1**:402
longitudinal studies **4**:1879
longpass (LP) filters **4**:1787
longwave UV fingerprint enhancement **3**:1318
Lophophora williamsii **4**:2063
LOQ *see* limit of quantitation
lorazepam **4**:1915
lost-in-the-mall technique **4**:1711
Louraine v. Commonwealth, 453 N.E. 2D, 437 (Mass. 1983) **5**:2586

low burning fire myth **1**:209, 219–20
low copy number (LCN) DNA **3**:1566
  degraded samples **2**:818
  hair **2**:826
  profiling **2**:809; **3**:1639–45
  saliva **2**:824
  samples **2**:818; **5**:2628
  techniques **2**:823–4
  testing **3**:1643; **5**:2268
low-level falls **5**:2341
low-power microscopy **2**:988–9; **4**:1791–3
low template DNA (LTDNA) **2**:808–10, **3**:1566–73
low-velocity impact, bloodspatter **1**:369–70
lower limit of quantitation (LLOQ) **2**:599; **4**:2224
LP (longpass) filters **4**:1787
LPC (linear predictive coding) **5**:2390
LPCC (linear prediction cepstrum coefficients) **5**:2391
LPDs (light petroleum distillates) **3**:1155
LQTS *see* long QT syndromes
LRs *see* likelihood ratios
LSD (lysergic acid diethylamide) **1**:292; **2**:849
LTDNA *see* low template DNA
LTO (Long-Term Offender) legislation **2**:670–1
lucid interval, intracranial hemorrhages **1**:404
luminescence **4**:1784–5, 2046
luminescent stains **3**:1302
luminol **2**:821; **3**:1645–56
  chemical/physical properties **3**:1645–6
  factors influencing use **3**:1649–52
  fingerprint enhancement **3**:1320
  formulation improvements **3**:1653–4
  interpretation of test results **3**:1653
  operational use **3**:1648–9
  reaction **3**:1646–54
lung
  disease **4**:2119
  injuries **1**:408
  postmortem toxicology **4**:2097, 2122
  tissue toxicology **4**:2122
  see also pulmonary...
Luria-Nebraska Neuropsychological Battery-Children's Revision (LNNB-CR) **4**:1873
luteinizing hormone (LH) **5**:2333
LVH *see* left ventricular hypertrophy
lying child syndrome **5**:2440
  see also deception detection; lie detection
lysergic acid diethylamide (LSD) **1**:292; **2**:849
lysis **2**:757, 1061
  see also extraction methods

M-FAST (Miller Forensic Assessment of Symptoms Test) **4**:1659
M8 detection paper **2**:513
M9 detection paper **2**:513
MacArthur Competence and Assessment Tool-Treatment (MacCAT-T) **2**:453, 454
MacArthur Structured Assessment of the Competencies of Criminal Defendants (MacSAC-CD) **2**:441
MacCAT-T *see* MacArthur Competence and Assessment Tool-Treatment
maceration process, wood **5**:2642, 2643
McGarry Scale **2**:441
machine direction (MD), paper analysis **4**:1971–2
machine-generated signatures **3**:1446–7
McHugh v. People 124 Misc.2d 559 (N.Y. Supp. 1984) **2**:580–1
Macintosh computers **2**:585–6
McKinley, - **4**:2179
McKinney, Aaron **5**:2447
McMahon, Thomas **4**:1941
McNaughten Rule **1**:255, 283; **3**:1553, **4**:2152; **5**:2444
macrolenses **2**:633; **4**:2038
macromorphological soil descriptors **5**:2384
macroscopic age determination methods **1**:181–4
MacSAC-CD *see* MacArthur Structured Assessment of the Competencies of Criminal Defendants
McVeigh, Timothy **2**:1044; **3**:1475, 1476, 1477
'madness', suggestibility **2**:733
Madrid train bombing **3**:1290, 1577
MagAttract® DNA extraction **2**:1062
magazine (firearms) packaging **4**:1929
maggots **2**:827, 937, 940, 942
  see also larval development stages
magic mushrooms **2**:828, 849–50; **4**:2059, 2064
  see also Psilocybe species
magistrates **2**:561–2, 1005, 1006, 1007
  see also judiciary
magnetic lenses **4**:1796
magnetic particle restoration (serial numbers) **5**:2326
magnetic powder **3**:1318, 1320
magnetic resonance imaging (MRI) **4**:1871; **5**:2234, 2236, 2301
magnetic storage media **2**:584
magnetic susceptibility methods **5**:2385
magnetometry **1**:203
magnification **4**:1763, 1765
  high power **4**:1758
  low power **4**:1791
  SEM **4**:1797
  stereomicroscopy **4**:1763
  see also microscopy
mail servers **1**:145
mainlining *see* intravenous drug use
maintenance faults, circuit breakers **2**:927
major contributor, DNA mixtures **4**:1838
major crime scenes **2**:620–3
  see also volume crime scenes
major incident forms **2**:610–4
major mental disorders **4**:1881
MALDI/TOF (matrix-assisted laser desorption/ionization time-of-flight) **4**:1805
males
  aggression rates **1**:38
  child sexual abuse victims **2**:539
  DNA admixtures **5**:2679
  sentencing **5**:2308–9
  stalking **5**:2398–9
  see also gender; sex determination
malicious use
  CBRN agents **2**:500–1
  explosions **2**:1023–4
malignant narcissism **5**:2317

malingering
behavioral science evidence **1**:287
countertransference **4**:1658
delusions **2**:743
detection **4**:1659, 1660–1, 1868–9
evaluation **4**:1657–61
genotyping for criminal trials **3**:1340
hallucinations **3**:1434
mental illness **3**:1556
mental retardation **4**:1734–5
neurocognitive dysfunction **4**:1660–1
neuropsychological assessment **3**:1463; **4**:1867–9
PTSD **4**:2141, 2143, 2144–6
seizure imitation **5**:2302
malpractice
expert behavior **4**:1663–7
medical **4**:1689–709
malware **1**:147–8
MAM (Mechanical, Analytical and Medical) analysis **5**:2566
mammal remains **5**:2393–4, 2394–5
see also human remains
management systems **1**:6–7, 12, 13, 15, 16; **2**:611–3, 619–25
see also postassault management; quality management; risk management plan; scene management
mandated treatment **5**:2451–2, 2576–9
mandible **4**:1892
manmade items
fibers **2**:986, 987–8
particles **4**:2001
see also manufactured items
Mann, Dale **3**:1165
manner of death
gunshot wounds **3**:1398
'undetermined' **5**:2542
Mant, A.K. **4**:1674, 1675
manual strangulation **1**:232; **5**:2563
manufactured items
defects **3**:1487
fiber particle forms **4**:2006
impressions **4**:1668–74
technical surfaces **5**:2486–7
see also manmade items
manufacturer information **3**:1206, 1240
MAOA *see* monoamine oxidase A gene
MAOIs *see* monoamine oxidase inhibitors
mapping
diatoms **2**:755
saliva stains **1**:317
soil **5**:2383, 2386
marijuana *see* cannabis
market-based forensic sciences **2**:485, 494, 680–1
marketing defects (products) **3**:1484, 1488
Markov, Georgi **5**:2508
marks, photographic **4**:2036–56
see also fingermarks; impressions; shoemark examination; tire marks; toolmarks
Marley v. Weisgram, 528 U.S. (2000) **5**:2627–8
Marsh test **4**:2120
Marshall and Hoare's formula **2**:708, 709, 714
Martin K. Eby Construction v. Neely, 386 U.S. (1967) **5**:2627
masking tape photographic method **1**:374
masks **2**:615
mass disasters **3**:1516; **4**:1893
see also disaster victim identification
mass graves
anthropologist role **3**:1516
definition **4**:1674
forensic archaeology **1**:200, 206
investigation **4**:1674–9
mini-STRs **4**:1808
pollen evidence **4**:1966
taphonomy **4**:1674–5
mass invariance, inks **3**:1544
mass migration, mtDNA **4**:1829
mass murder **3**:1474–5, 1478, 1479, **4**:1674
see also multiple murders
mass spectrometry (MS) **2**:595–7
amphetamine analysis **1**:138–9
analytical artifacts **4**:2107
cannabis analysis **2**:434
cocaine analysis **2**:568
confirmation testing **1**:297
CWA analysis **2**:522–3
driver impairment evidence **2**:881
explosion debris **2**:1039, 1040, 1041
fiber analysis **2**:991
hair analysis **3**:1424, 1427, 1431
ignitable liquid residues **3**:1143, 1162, 1168
MALDI/TOF **4**:1805
matching criteria **2**:596
paint analysis **4**:1938, 1939
plant analysis **4**:2066
postmortem toxicology **4**:2107, 2126, –7, 2130–1, 2131, 2133, 2134
radiocarbon samples **1**:419–20; **5**:2231–2
sweat testing **5**:2427, 2428
urine drug tests **2**:862
see also gas chromatography-mass spectrometry; liquid chromatography-mass spectrometry
mass spectrum detectors (MSD) **5**:2525
mass storage (MMS), mobile phones **3**:1362
mass/volume magnetic susceptibility, soil **5**:2385
master's degree programs **2**:898–9
match probabilities (MPs) **5**:2366, 2367–8
DNA databases **2**:832, 833–4, 834–5, 835, 836–7
errors **5**:2371–2
fingerprint comparison **3**:1278
Hardy–Weinberg equilibrium **3**:1458
identification/individualization **3**:1509
mtDNA evidence **4**:1827–8
multiple loci **5**:2369–70
single-locus **5**:2368–9
Y chromosome STR profiles **5**:2373
match statistics
characterization of **5**:2401–4
DNA databases **2**:832–7, 837, 838
mtDNA evidence **4**:1824, 1825–7
see also match probabilities
matching centers, disaster victims **2**:770–1
matching process
earprints **2**:892–3
hair examination **3**:1422, 1423
mass spectral criteria **2**:596

odontological identification **4**:1892–3
see also comparison
materials
characterization **4**:1685–7
engineering (failure analysis) **4**:1680–8
firearm discharge residue patterns **5**:2352–3
maternal aggression **1**:41
maternal filicide **4**:2138–9
maternity testing *see* parentage testing
mathematical definition, surfaces **5**:2486
mathematical modeling **2**:856; **3**:1177; **5**:2565–75
see also deterministic fire models
The Matrix (movie) **5**:2448
matrix/matrix files **4**:1688–9
matrix-assisted laser desorption/ionization time-of-flight (MALDI/TOF) **4**:1805
matrix matching solutions **4**:2223
mature adult category **1**:161
maturity assessments, young offenders **3**:1611
maximizing strategy, crime victims **2**:645
maximum phase, livor mortis **2**:704
Mayfield Incident see Madrid train bombing
MB (myocardial bridging) **2**:477
MCMC (Monte Carlo Markov Chain) **4**:1827
MCMI (Millon Clinical Multiaxial Inventory) **2**:695
MCT (mercury cadmium telluride) detector **4**:1937
MD (machine direction), paper **4**:1971–2
MDA *see* 3,4-methylenedioxyamphetamine; multiple displacement amplification
MDMA *see* 3,4-methylenedioxy-methamphetamine
MDTs (multidisciplinary teams) **2**:909
mean values, time of death **5**:2468
measurement
bloodstains **1**:368
errors **2**:956
height estimation digital images **3**:1624–32
ink volatile components **3**:1544
instruments **3**:1083; **4**:1878, 1880
radiocarbon samples **1**:419–20; **5**:2231–2
uncertainty of **2**:600
validation **3**:1628–30, 1631
Mechanical, Analytical and Medical (MAM) analysis **5**:2566
mechanical asphyxia **1**:225; **5**:2562–3
mechanical erasures, documents **1**:131–2
mechanical excitability of muscles **2**:698; **5**:2468
mechanical factors
crash tests **5**:2570, 2571
material properties **4**:1680, 1685–6
mechanical fit evidence **2**:618
mechanical stimulus, oral fluid collection **4**:1907
mechanisms of action
amphetamines **1**:136
behavioral toxicology **1**:290–1
benzodiazepines **1**:294–5
cannabis **2**:434
cocaine **2**:563
oral fluid drug transfer **4**:1906
traumatic memory **5**:2243–4
media
desensitization to violence **1**:30
multiple murder coverage **3**:1478
Stockholm syndrome **5**:2409–10
mediastinal emphysema **1**:408
Medicaid programs **2**:649; **4**:1729
medical disorders **2**:469–80
see also individual disorders
medical emergencies **2**:439; **4**:2095
medical estimation, time of death **5**:2467
medical examinations
battered child syndrome **1**:267
shooting scene investigation **3**:1224
see also examinations
medical factors
alcohol effects **1**:99–105
cannabis use **2**:433
crash tests **5**:2570, 2571–2
see also biomedical engineering
medical malpractice **4**:1689–709, 2164
medical models, disaster mental health **2**:760–1
medical negligence **4**:1690, 1701–2, 1703
medical radiology **5**:2233–4, 2234, 2235–9
see also radiology
medical treatment
capacity to consent **2**:439–40, 450–4
consent/refusal capacity **2**:439–40
decision-making capacity **2**:452–4
see also medications; treatment
Medicare programs **2**:649
medications
alcohol interactions **1**:108–17
classes **4**:2203
compliance (civil commitment) **5**:2452
definition **1**:108
drug-facilitated crime **2**:869
errors (malpractice) **4**:1703, 1704, 1705
informed consent **4**:2211–2
pediatric psychopharmacology **4**:2214–5
postmortem toxicology **4**:2124
refusal (civil commitment) **2**:559
see also drug. . .; *individual medications*; treatment
medicolegal autopsies **1**:256–7, 258, 259, 260, 261; **3**:1468, 1469, 1473
see also autopsies
medicolegal investigations **2**:934
meditation treatment **5**:2416
medium petroleum distillates (MPDs) **3**:1150, 1155–7, 1158
medium-range gunshots **3**:1392
see also intermediate-range gunshots
medium-velocity impact, bloodspatter **1**:370–5
medroxyprogesterone (MPA) **5**:2334
medulla of hair **3**:1404, 1407, 1418
medullary index (MI) **3**:1409, 1418
meiosis **2**:793
Mekong Basin hydropower **2**:951
mel frequency cepstral coefficients (MFCC) **5**:2391
melanin pigments, hair **3**:1418
melanocytes **3**:1326, 1418
Mellanby effect **1**:75
melted fractures, light globes **3**:1634, 1637
membership requirements, professional associations **2**:1014–5
memory
adult suggestibility **2**:1065–70
aging effects **2**:916–7
formation processes **2**:1076

memory *(cont)*
interrogative suggestibility **3**:1592
loss (benzodiazepine use) **1**:295
malingered deficits **4**:1868–9
neuropsychological assessment **3**:1462; **4**:1866, 1868–9
recall confidence **3**:1501, 1502
reconstructive memory **4**:1709–12
retrieval **2**:786–8
stages of **4**:1709–11
traumatic memory recollection **5**:2243–8
truth and **2**:728–9, 730–5
see also amnesia; false memories; hypnosis; implanted memories; recovered memory; repressed memory; suggestibility
memory cards, mobile phones **3**:1363
memory distrust syndrome **2**:589
Memory and Encoding Related Multifaceted Electroencephalographic Response (MERMER) **2**:725
'memory impairment' **2**:1067
men *see* gender. . .; male...
Mendel, Gregor **4**:2012, 2020
mens rea **1**:282; **3**:1339
automatism defense **1**:254
dissociative identity disorder **2**:788
insanity defense **3**:1552
neuropsychological impairment **3**:1460
premenstrual syndrome **4**:2152; **5**:2447
shared psychotic disorder **5**:2446
unconscious motivation **2**:574
menstrual blood **2**:821
mental asylums **2**:556
mental competency *see* competency
mental disorders *see* mental illness
mental health
assessments (elder abuse) **2**:909
disaster situations **2**:760–3
juvenile correctional settings **3**:1614, 1615
services **4**:1717–8
therapeutic jurisprudence laws **5**:2449–54, 2580–3
see also mental health professionals; mental illness
Mental Health Act (UK) **2**:669
mental health courts **2**:462; **4**:1717–24; **5**:2450–1
see also therapeutic jurisprudence
mental health professionals
clinical assessment **4**:2173–5
competency to stand trial evaluation **2**:457–9
defendant competency assessment **2**:461
doctor-patient confidentiality **2**:885–6, 887–8, 889
ethical issues **2**:443
Miranda rights waiver **2**:464, 465
risk assessment **5**:2272
threat assessment **5**:2597–601
violence prediction **2**:671–4
see also mental health
mental illness
capacity to waive Miranda rights **2**:466
children **4**:1872
civil commitment **2**:556–60
court-mandated treatment **2**:573
criminalization **2**:649–53; **4**:1718
death penalty abolition **2**:719
definitions **2**:651
diversion programs for accused **4**:1721–3
'evil' association **2**:982
female/male aggression **1**:43
folkways **1**:281
inmates **2**:456, 719
insanity defense relationship **1**:283
juvenile competence/culpability **2**:544–6, 547–8
legal relevance **1**:287
malingering **3**:1556
mandated treatment **5**:2576–9
multiple murder causes **3**:1475–6
Northwestern Juvenile Project **4**:1880, 1881
rights to treatment **5**:2580–3
substance abuse **5**:2415
symptom severity **5**:2445
temporary insanity **5**:2443–8
treatment refusal **5**:2584–7
violence **5**:2274
see also alcohol, drug and mental disorders; insanity defense; postpartum psychosis; psychiatric disorders; severe mental illness
mental impairment defense **1**:253–6
mental retardation **4**:1724–9
assessment **4**:1726–7
community life **4**:1727–8
death penalty **2**:718; **4**:1730–5
definitions **4**:1724–6, 1732
diagnosis in death penalty cases **4**:1732–4
etiology **4**:1726
false confession vulnerabilities **2**:591
malingering **4**:1734–5
services/support **4**:1728–9
mental state
at time of offense **3**:1435
expert testimony **1**:286
legal responsibility **1**:284
mental status examination (MSE) **4**:1737–42
mercaptoethanol solution **1**:353
mercury cadmium telluride (MCT) detector **4**:1937
mercury lamps **4**:1785
Merkel cells **3**:1326
MERMER (Memory and Encoding Related Multifaceted Electroencephalographic Response) **2**:725
Merrell Dow Pharmaceuticals v. Daubert see Daubert v. Merrell Dow Pharmaceuticals
mescaline **1**:134; **4**:2063
message digest *see* hash value
metabolic disease **1**:196
metabolic disorders **4**:1851–4
metabolic tolerance **1**:75, 123–4
metabolism
alcohol **1**:62–3
amphetamines **1**:137
benzodiazepines **1**:295; **2**:870–1
cocaine **2**:565; **4**:1912
drug testing in urine **2**:860
glucose **4**:2076–81
hydrogen sulfide **4**:2133
opioids **4**:1900
metabolite stability **4**:2105, 2106

metadata **2**:778
metal salt treatment, fingermarks **3**:1297
metal surface explosion debris **2**:1030
metallic iron dissolution **3**:1215
metallic paints **4**:1935
metallic powder **3**:1318
metallography **4**:1686
metals
  failure analysis procedures **4**:1687
  luminol test results **3**:1653
  particle transfer **5**:2537
  pretreated automotive paints **4**:1943
  see also individual metals; materials. . .; metal...
metameric inks **3**:1269
metatarsal ridge **3**:1244
methadone **1**:115; **4**:1899, 1918, 2018, 2019, 2117, 2118
methamphetamine **1**:135, 136, 138
  analytical methods **1**:138; **2**:847
  driver impairment **2**:878
  excretion **1**:137
  oral fluid **4**:1911
  repeated use **4**:2118
methane **3**:1107
methanol **1**:115, 121; **3**:1162
  see also alcohol
methodologies
  behavioral science **1**:285
  forensic archaeology **1**:200
  height estimations **3**:1626–8
  implementation **4**:2221–5
  intelligence **2**:600
  network analysis **1**:145–8
  pharmacogenomics **4**:2014–5
  search strategies **2**:623
  standardization (fire investigation) **3**:1172–3
  validation **1**:90–2; **4**:2009, 2010, 2221–5
  see also analytical methods; *individual methods*; scientific method
4-methoxyamphetamine (PMA) **1**:135, 137, 138
methyl alcohol *see* alcohol
methylbenzoylecgonine *see* cocaine
3,4-methylenedioxy-methamphetamine (MDMA) **1**:135, 136
  drug-facilitated crime **2**:872
  excretion **1**:137
  oral fluid **4**:1906, 1911, 1919
  sweat **5**:2422, 2426, 2427
  toxicity **1**:138
  see also ecstasy
3,4-methylenedioxyamphetamine (MDA) **1**:135, 137; **4**:1911
methylnaphthalenes **3**:1154
methysticum species **4**:2064
metric approaches **1**:166, 192, 194; **5**:2328, 2330–1
Meyer, Adolf **4**:1737, 1738
MFCC (mel frequency cepstral coefficients) **5**:2391
MGT (modified Griess test) **5**:2351–2
MHL (minimal haplotype loci) **5**:2677
MI *see* medullary index; motivational interviewing
Michel–Levy chart **4**:1781
Michigan v. Clifford **3**:1123
Michigan v. Tyler **3**:1123
micro-X-ray fluorescence **3**:1195; **5**:2353
microalgae **1**:428–9
microanalysis, nuclear forensics **4**:1886
microarrays **1**:306
microbial forensics **2**:505–6
microbiological investigations **1**:307–9
microcantilevers **1**:306
microchemistry **2**:992; **4**:1743–9, 1784
microcrystal tests **4**:1744–5, 1749
microorganism histology **3**:1470
microsatellites **4**:1749–50, 2021
  see also short tandem repeats
microscopy
  age determination **1**:181, 183–4
  animal hair **3**:1403, 1404–10
  biological agents **1**:309
  bone structure **5**:2394
  brightfield **2**:989
  diatoms **2**:754
  explosion debris **2**:1034, 1036
  fecal stains **1**:318
  fiber analysis **2**:988–90; **4**:1975–80
  firearms **3**:1206–7, 1207, 1208, 1209
  forgeries/counterfeits **3**:1257
  Fourier transform infrared **4**:1750–7
  hair structure **3**:1418, 1420–2, 1424
  high power **4**:1758–61
  history **3**:1468
  ink aging **3**:1544
  light microscopy **4**:1762–89
  low-power **2**:988–9; **4**:1791–3
  materials characterization **4**:1686–7
  paint examination **4**:1934, 1935, 1937, 1942
  perimortem trauma analysis **5**:2559
  pollen **4**:1954, 1958–9
  scanning electron microscopy **4**:1793–804
  shooting-distance estimation **5**:2351–3
  soil analysis **5**:2385
  toolmark examination **5**:2487–9
  wood **5**:2642–4, 2645
  see also microchemistry
microspectral imaging **4**:1756–7
microspectrometry **4**:1751–6
microspectrophotometry (MSP) **2**:990, **4**:1938
microstructural materials characterization **4**:1686–7
microvariants, STRs **5**:2356
mid-range photographs **2**:630–2
midazolam **4**:1915
middle adult category **1**:161
migraine **5**:2302
migration effects, mtDNA databases **4**:1829
mild steel **5**:2326
miles per hour (mph), air bag deployment **1**:55
military biological agents *see* biowarfare
military explosives **2**:1044
Millennials jury pool **2**:1017
Miller Forensic Assessment of Symptoms Test (M-FAST) **4**:1659
milling serial number obliteration **5**:2325
Millon Clinical Multiaxial Inventory (MCMI) **2**:695
minerals **4**:2003, 2006; **5**:2384–5
  see also individual minerals
minimal haplotype loci (MHL) **5**:2677
Mini-Mental State Examination (MMSE) **4**:1739, 1864

minimization of contamination **1**:414–5
minimization (interrogation tactic) **2**:592
minimum number of individuals (MNI) present **1**:155
minisatellites **1**:140; **4**:1749; **5**:2595–6
  see also variable number tandem repeats
mini-short tandem repeats **2**:817; **4**:1804–9; **5**:2362
mini-tapes **1**:321
Minnesota Multiphasic Personality Inventory (MMPI) **3**:1593; **4**:1869, 2179–80, 2183
Minnesota Multiphasic Personality Inventory-2 (MMPI-2) **4**:1659, 2189; **5**:2605
minor contributor, DNA mixtures **4**:1838
minority syndromes **5**:2439
minority tolling doctrine **4**:1713
minors' civil commitment **2**:559
  see also children; pediatric psychopharmacology
minutiae (fingerprints) **1**:251; **3**:1283, 1287, 1288
  see also Galton points
mIPEP (modified primer extension preamplification) **5**:2632
Miranda rights
  capacity to waive **2**:463–7
  interrogative suggestibility **3**:1590
  warnings **3**:1586
Miranda v. Arizona, 384 U.S. 436 (1966) **2**:463–4
misattributing information source **2**:916
misconceptions, fire investigation **1**:207–23
misdemeanant defendants **2**:651
misidentification
  delusions **2**:742, 743
  in lineups **2**:1073–4, 1074
misinformation effects **2**:1066–7; **4**:1710
misrepresentation, experts **2**:959–60
missing persons
  calculations **4**:1810–7
  Casualty Bureau system **2**:765–6
  DNA database program **4**:1836
  DNA profiling **4**:1836
  face aging techniques **1**:190
  odontological identification **4**:1893
  search scenarios **1**:202–4
  see also disaster victim identification
misuse of products **3**:1490
Mitchell v. United States, 365 F.3d 215 (3rd Cir. 2004) **4**:2011
mitigation
  factors (juvenile justice) **3**:1609
  genotyping for criminal trials **3**:1339
  testimony **4**:1818–23
mitochondrial DNA (mtDNA) **2**:795
  amplification method **5**:2632–3
  analysis **4**:1834
  databases **4**:1828–30
  hair **3**:1424
  insect sources **2**:827
  interpretation **4**:1823–32, 1835–6
  mini-STRs **4**:1805
  nonforensic analyses **4**:1836
  profiling **4**:1833–7
  sample interpretation **4**:1835–6
  typing analyses **4**:1833
mitosis **2**:793
mitral valve **2**:478–9
mitral valve prolapse (MVPS) **2**:476–7
'mixed is burned' fire model **3**:1184
mixed sexual homicide **5**:2313
mixtures (DNA)
  analysis (Y-chromosomal STRs) **5**:2680
  peak heights **4**:2008
  statistical interpretation **4**:1838–41
  see also admixture
MLE (most likely estimate) **4**:2026
MLP *see* multilocus probing/profiling
MMC (Multi Media Card system) **3**:1363
MMD *see* multimetal deposition
MMDII *see* multimetal deposition II
MMPI *see* Minnesota Multiphasic Personality Inventory
MMPI-2 *see* Minnesota Multiphasic Personality Inventory-2
MMS (mass storage) **3**:1362
MMSE *see* Mini-Mental State Examination
M'Naghten standard *see* McNaughten Rule
MND (malingering, neurocognitive dysfunction) **4**:1660–1
MNI (minimum number of individuals) present **1**:155
MNS blood group system **1**:344
mobile laboratories **1**:417–8
mobile phases, drug screening **5**:2527–9
mobile phones **2**:584, 586–7
  centric **3**:1360–71
  data sources **3**:1365–8
  handsets
    architecture **3**:1362–5
    data sources **3**:1365–7
    hacking **3**:1368–9
    user computer **3**:1368
mock jury study **5**:2404–5
mock witness procedure **2**:1074
modafinil **4**:2207
modeling
  Bayesian networks **1**:278–9
  bloodstain pattern reconstructions **1**:390–5
  case assessment and interpretation **2**:488–9
  characters **5**:2261
  crime scenes **5**:2258–60
  definition **3**:1177
  fingerprint interpretation **3**:1280
  fire modeling **3**:1175–88
  height estimations **3**:1628
  mathematical **2**:856; **3**:1177; **5**:2565–75
  probability theory (toolmarks) **5**:2487
  sex offender treatment **5**:2335
  see also 3D reconstructions
moderate drinkers **1**:122
modern racism **5**:2306
modified Griess test (MGT) **5**:2351–2
modified physical developer (MPD) **3**:1303
modified primer extension preamplification (mIPEP) **5**:2632
modulus of elasticity **4**:1682
moisture absorbency, paper **4**:1971
moisture availability, mass graves **4**:1675
molds, fingerprints **1**:330, 332
molecular analysis, polymorphisms **1**:340, 347
molecular autopsies **4**:2018

molecular biology **2**:798–9; **5**:2636–7
see also genetics
Mongolid characteristics **1**:191–3
monkshood plant **4**:2059–60
monoamine oxidase A (MAOA) gene **3**:1338, 1339, 1341
monoamine oxidase inhibitors (MAOIs) **1**:112; **4**:2205, 2206
monochromatic light sources **4**:2041–3
monoclonal antibodies **1**:314, 316, 318, 344
monomer reversion, polymers **3**:1110
monozygotic twins **3**:1330
Monte Carlo Markov Chain (MCMC) **4**:1827
mood assessment **3**:1463
mood disorders **1**:292, 293; **4**:2215
see also bipolar disorder; depression
mood stabilizers **4**:2216
Moore v. Parker, 425 F.3d 250 (6th Cir., 2005) **4**:1821–2
moral development, multiple murderers **3**:1476
morbid jealousy **5**:2399
morbid obesity **4**:1848–9
morphine **2**:847–8
bile content **5**:2499
gastrointestinal effects **4**:1899
metabolism **4**:1900
oral fluid **4**:1913, 1914
sources **4**:1896
urine testing **2**:865
morphogenesis, friction ridge skin **3**:1322–30
morphology
foot **3**:1244
hair **3**:1418
particles **4**:1769
positive identifications **3**:1512–6
race/ancestry determination **1**:166
sex determination **1**:156; **5**:2328, 2331
soil **5**:2380, 2383, 2384
wildlife forensics **5**:2636
see also anthropology; odontology
mortality rates, juveniles **4**:1882
mortuaries, bomb scenes **1**:415
mosquitoes **2**:827
most likely estimate (MLE), phenotypes **4**:2026
motion capture systems **5**:2262
motion picture evidence **2**:746
see also video/videotape
motions
discovery **2**:774–5
in limine **2**:664; **3**:1528
pretrial depositions **2**:773
motivational interviewing (MI) **5**:2416
motives/motivations
elder abuse **2**:906
expert testimony **1**:286
reduction of **2**:646
serial murder **5**:2315–6
motor seizures **5**:2298
motor vehicles *see* vehicles
Mountbatten, Lord **4**:1941
mounting media for microscopy **4**:1769
mouth reconstruction **3**:1088
movements, bomb scenes **1**:413
moving images *see* computer animation evidence; motion. . .; video/videotape
Moye, J. **2**:454
MPA (medroxyprogesterone) **5**:2334
MPD *see* modified physical developer; multiple personality disorder
MPDs *see* medium petroleum distillates
mph *see* miles per hour
MPs *see* match probabilities
MRI *see* magnetic resonance imaging
MRM (multiple reaction monitoring) **4**:2131
mRNA pharmacogenomics **4**:2013
MS *see* mass spectrometry
MS/MS *see* tandem mass spectrometry
MSD (mass spectrum detectors) **5**:2525
MSE (mental status examination) **4**:1737–42
MSP *see* microspectrophotometry
mtDNA *see* mitochondrial DNA
mugging (strangulation) **1**:232
mugshots **2**:746–7; **3**:1501, 1502
Mulica case **5**:2432
multi-channel spectroscopy **2**:990
multicomponent analysis **4**:1688
multidimensional continuum, admixture **4**:2024
multidisciplinary teams (MDTs) **2**:909
multilocus probing/profiling (MLP) **2**:800–1; **5**:2595
see also multiple loci match probabilities
multimedia **2**:579–82, 584; **3**:1363
Multi Media Card (MMC) system **3**:1363
multimetal deposition (MMD) **3**:1303–4, 1305, 1308
multimetal deposition II (MMDII) **3**:1304
multipage document alterations **1**:130, 131
multiple bullet lesions **5**:2562
see also multiple projectile cartridges
multiple data sources, network analysis **1**:141–2, 148
multiple displacement amplification (MDA) **2**:808–9; **5**:2630–1, 2632–3
multiple doses of drugs **4**:2116–7
multiple drug use **1**:109, 138; **4**:2119
see also drug–alcohol interactions; drug–drug interactions
multiple interviews **2**:550, 553
multiple loci match probabilities **5**:2369–70
see also multilocus probing/profiling
multiple murders **3**:1474–9
multiple personality disorder (MPD) **5**:2434–5
see also dissociative identity disorder
multiple projectile cartridges **3**:1203–4
see also multiple bullet lesions
multiple reaction monitoring (MRM) **4**:2131
multiplex short tandem repeats **5**:2354, 2357, 2677
multivariate analysis **2**:858
mummification of remains **3**:1513
Munchausen syndrome **5**:2302
Münsterburg, Hugo **2**:1066; **4**:2181
murder *see* homicide. . .; multiple murders
Murray, Henry **4**:2178, 2179
muscles
excitability **2**:698–701; **5**:2468, 2469–72
facial reconstruction **3**:1088
specimen sampling **4**:2103

mushrooms *see* fungi; magic mushrooms
mutations **2**:794, 795, 796–8
  hotspots **4**:1827, 1831
  polymorphism **4**:2075
  short tandem repeats **5**:2359–61, 2375
muzzle flash **3**:1201, 1204
muzzle-to-target range **3**:1390–3
MVPS (mitral valve prolapse) **2**:476–7
Mycoplasma pneumoniae **4**:1845
Myers v. Alaska Psychiatric Institute, 138 P.3d 238 (2006) **5**:2585–6
myocardial bridging (MB) **2**:477
myocarditis **2**:474–5; **3**:1472
myoclonic seizures **5**:2298
myths
  fire investigation **1**:207–23
  'pure evil' **2**:978

NA *see* numerical aperture
NAA *see* neutron activation analysis
NAD^+ (nicotinamide adenine dinucleotide) **1**:62
NAEs *see* negligent adverse events
NAFE (National Academy of Forensic Engineers) **2**:922
NAFI (National Association of Fire Investigators) **3**:1174
NAHI *see* nonaccidental head injury
nails **5**:2501
  see also fingernail...
nanoparticle detection techniques **3**:1304
nanowires **1**:306
naphthenic–paraffinic products **3**:1162
Napoleonic Code **2**:561
narcissism **1**:30; **5**:2399
narcolepsy **5**:2302
nascent ridges **3**:1285
NATA supplementary requirements **1**:13–4
National Academy of Forensic Engineers (NAFE) **2**:922
National Association of Fire Investigators (NAFI) **3**:1174
National Bureau of Standards (NBS) **1**:208, 209
National Clearinghouse for Science, Technology and the Law (NCSTL) **5**:2619–20, 2621
national context
  expert malpractice **4**:1663–4
  illicit drugs policies **2**:854–5
  see also nationhood
National Crime Victimization Survey (NCVS) **1**:28; **2**:643, 645, 647
National Criminal Justice Reference Service (NCJRS) **5**:2621
National DNA Database Programs **2**:677–80; **4**:1836
National Fire Protection Association (NFPA) **1**:208; **3**:1172, 1173
National Forensic Science Technology Center (NFSTC) **1**:2
National Institute of Child Health and Human Development (NICHD) **2**:554
National Institute of Justice (NIJ) **2**:899, 900, 901; **3**:1288, 1290; **5**:2619
National Institute of Standards and Technology (NIST) **1**:151
National Missing Person DNA Database program **4**:1836
National Research Council (NRC) **2**:497–8, 834–5; **3**:1290
nationhood **1**:281–2
  see also national context
native gel electrophoresis (Native PAGE) **1**:349
natural causes of death **4**:1843–61; **5**:2541–2
natural diseases *see* diseases
natural fibers **2**:986, 987
natural gas **3**:1107
natural polymers **3**:1109
natural sciences training **5**:2545
natural surfaces, definition **5**:2486
natural variation, handwriting **3**:1437, 1442
naturally occurring soils **5**:2379, 2383
nausea **4**:1899
Naxos disease **2**:477
NBS *see* National Bureau of Standards
NC *see* nitrocellulose
NCJRS (National Criminal Justice Reference Service) **5**:2621
NCSI *see* nonconsensual sexual intercourse
NCSTL *see* National Clearinghouse for Science, Technology and the Law
NCVS *see* National Crime Victimization Survey
nDNA testing **3**:1420, 1424
near-axial illumination **4**:2050, 2053
near contact shooting range **5**:2351
near letter quality (NLQ) printers **5**:2665
'near-table' methods **4**:2076
necessity defense **2**:577
neck
  compression **1**:228–32
  fractures **3**:1492–3, 1493
  internal examination **1**:232
  restraint **1**:232
needs (product) **3**:1485–6
Neely v. Martin K. Eby Construction, 386 U.S. (1967) **5**:2627
negative corpus, fire scenes **3**:1134
negative ion MS **2**:595, 597
negative stereotypes **2**:554
neglect
  children **1**:265; **4**:1991
  elders **2**:904, 905, 907
  entomological analysis **2**:945
negligence
  disaster mental health practice **2**:762–3
  liability **2**:762–3; **3**:1487
  medical **4**:1690, 1701–2, 1703
  see also duty to warn
negligent adverse events (NAEs) **4**:1690, 1691, 1692, 1703
Negroid characteristics **1**:191–3
NEO-Personality Inventory-Revised (NEO-PI-R) **4**:2184
neonatal line **4**:1891
neonatal remains **1**:179
neonaticide **4**:2138
neoplastic diseases **1**:196, **4**:1850–1
Nerium oleander **4**:2063–4

nerve agents **2**:501, 513, 521
nervous system drug effects **4**:2202–3
nested PCR strategy **2**:809
Netherlands **2**:436, 1010
networks
  analysis **1**:141–8
  Bayesian **1**:276–80; **3**:1584; **4**:1841
  cellular phones **3**:1360–2, 1367–8
neurally mediated syncope **5**:2302
neurobiological effects, child sexual abuse **2**:535
neurocognitive malingering **4**:1660
neurodegenerative conditions **2**:948–9, 949
neuroimaging psychopathology **4**:2189
neuroleptic malignant syndrome **4**:2205
neurological disorders **4**:1871–2
neurons **4**:2203
neuropsychological assessment **4**:1862–9, 1869–76
neuropsychological testing **4**:2175, 2180
neuroscientific research **2**:577–9
neurosurgery, sex offenders **5**:2334
neurotransmitters **1**:290, 291; **4**:2203
neutron activation analysis (NAA) **1**:150–1; **3**:1191
New York City region **3**:1612–3, 1614
New York Innocence Project **2**:682–3
New York state **2**:558, 682–3
New Zealand DNA Database **2**:679
Newton's colors **4**:1779
Newton's law of cooling **2**:708, 709
NFPA *see* National Fire Protection Association
NFSTC (National Forensic Science Technology Center) **1**:2
NG *see* nitroglycerine
NGRI *see* not guilty by reason of insanity
nib pens **5**:2661
NICHD (National Institute of Child Health and Human Development) **2**:554
nichrome wire **2**:924
Nickell, Rachel **4**:2157
nicotinamide adenine dinucleotide (NAD^+) **1**:62
nicotine–cytisine comparison **4**:2063
Niemann, Albert **2**:562
nightshot cameras **1**:337
nihilistic delusions **2**:742
NIJ *see* National Institute of Justice
ninhydrin **3**:1296, 1299, 1310, 1319
Ninth Circuit Court of Appeals, US **2**:693
NIST *see* National Institute of Standards and Technology
nitrates **2**:1031, 1037, 1038, 1046
nitrite, urine tests **2**:864–5
nitrocellulose (NC) **2**:1051
nitrogen, urea **5**:2473
nitrogen phosphorus detector (NPD) **2**:595, 598; **4**:2126; **5**:2524
nitroglycerine (NG) **2**:1052
NJP (Northwestern Juvenile Project) **4**:1877–82
NLQ (near letter quality) printers **5**:2665
NMR (nuclear magnetic resonance) spectroscopy **2**:845
'no correction' approach, DNA databases **2**:837–8
nodes, Bayesian networks **1**:276, 278, 280
noise thresholds (DNA) **4**:2008
  see also background noise; developmental noise
nomograms **2**:710–4; **5**:2467–72
nonabsorbent substrates **3**:1650, 1651
nonaccidental head injury (NAHI) **5**:2339
  see also shaken baby syndrome
nonautosomes **2**:795
nonbiological matter, ballistics **3**:1202
nonbizarre delusions **2**:742
noncardiac causes, sudden death **4**:1843–61
noncolored components, inks **3**:1544
nonconformities
  ASCLD/LAB program **1**:5
  laboratory accreditation **1**:13
  proficiency tests **2**:956
nonconsensual sexual intercourse (NCSI) **1**:238
  see also sexually motivated assault
nondestructive techniques
  neutron activation analysis **1**:150
  serial number restoration **5**:2325–6
nondisclosure **2**:452; **5**:2245
nongenuine questioned signatures **3**:1447
nonhuman analysis, STRs **5**:2363
nonhuman/human bone differentiation **1**:153–4
  see also animals
noninvasive optical imaging **2**:726
nonjudgmental role, clinicians **5**:2436
nonmatching entities, fingerprints **3**:1278
non-narcotic pain relievers **1**:115
nonnucleotide linker molecules **4**:1807
nonnumerically based sampling plans **5**:2293
nonpermanent contact traces, textiles **2**:993–4
nonporous surfaces **1**:369; **3**:1292, 1300–3, 1319–20
nonprimary identity indicators **2**:768
nonrecombining region of Y chromosomes (NRY) **5**:2677
nonrepinephrine *see* noradrenaline
nonsecretors, ABO blood group **1**:342
nonstate acquisition/production/use, CBRN agents **2**:502
nonstatistically based sampling plans **5**:2293
nonsteroidal anti-inflammatory drugs (NSAIDs) **1**:115
nontort solutions, malpractice **4**:1663–4
nonvolatile components, inks **3**:1544
nonwood fibers, paper **4**:1978–80
11-nor-9-carboxy-Δ^9-tetrahydrocannabinol *see* THC-COOH
noradrenaline **1**:134, 136
nordiazepam **1**:295
normal heart cases, sudden death **2**:470–1
normality **4**:2186–7
norms **4**:2186–7, **5**:2297
norpseudoephedrine **4**:2061
Northern Ireland **2**:557
Northwestern Juvenile Project (NJP) **4**:1877–82
nose reconstruction **3**:1088
not guilty by reason of insanity (NGRI) **5**:2581, 2583
  see also insanity defense
not identified *see* inconclusive results
notetaking during trial (juries) **3**:1606
novel expert evidence **3**:1333
  see also expert opinion evidence
novel scientific methods **4**:2009–11
novel sequences, mtDNA **4**:1835

novel theories **1**:288
nozzle faults, printers **5**:2670
NP *see* neuropsychological...
Np approach, DNA databases **2**:835, 836, 837, 838
NPD *see* nitrogen phosphorus detector
NRC *see* National Research Council
NRY (nonrecombining region of Y chromosomes) **5**:2677
NSAIDs (nonsteroidal anti-inflammatory drugs) **1**:115
nuclear attribution **4**:1884, 1886–7
nuclear DNA **2**:794; **4**:1833, 1834, 1836
nuclear forensics **2**:506; **4**:1883–7
  see also chemical, biological, radiological and nuclear agents
nuclear magnetic resonance (NMR) spectroscopy **2**:845
nuclear medicine **5**:2234
nucleases **2**:954
nucleotides **1**:340; **2**:798; **4**:1833, 1834, 1835; **5**:2354, 2359
nuDNA *see* nuclear DNA
null alleles **5**:2360–1
null hypothesis (H_0) **3**:1582
numerical aperture (NA) **4**:1758–9, 1767
numerical conversion error **3**:1584; **5**:2404
numerical evaluation, height estimations **3**:1630
numerical specification, Bayesian networks **1**:279
numerical standard *see* minutiae (fingerprints)
numerically based sampling plans **5**:2292–3
nursing homes **2**:905, 914
  see also long-term care facilities
Nysten's rule **2**:707

oath-taking **3**:1467
OBAs *see* optical brightening agents
obesity **4**:1848–9
obfuscation
  source addresses **1**:142–3
  urine drug tests **2**:862–5
object-oriented Bayesian networks (OOBNs) **1**:279
objections to leading questions **2**:758
objective component, compound microscopes **4**:1758, 1791
objective lens system **4**:1765, 1768, 1796
objective reasonableness **4**:2073
objective testing **4**:2069, 2071, 2178; **5**:2494
oblique lighting **3**:1595; **4**:2052–3
obliterations
  documents **1**:128–9, 132–3
  serial numbers **5**:2325
observations
  explosion scenes **2**:1028
  facial comparison **3**:1081–5
observer effects **3**:1575–8
  see also confirmation bias; context effects
obsessional behavior **5**:2317, 2399
obsessive-compulsive disorder (OCD) **2**:572; **4**:2137
OCA2 gene *see* Oculocutaneous Albinism 2 gene
occasional features, friction ridges **3**:1285
occupant position sensing **1**:57
occupant restraint **5**:2566, 2574
OCD (obsessive-compulsive disorder) **2**:572
O'Connor v. Donaldson, 422 U.S. 563 (1975) **5**:2581
octogen (HMX) **2**:1039, 1041, 1050
ocular petechiae **1**:225
Oculocutaneous Albinism 2 (OCA2) gene **4**:2030, 2031
odds
  concept of **2**:969
  ratios **3**:1613
  see also posterior odds; prior odds
odds form (Bayes theorem) **2**:486
odontology **2**:770; **3**:1090, 1512; **4**:1889–95; **5**:2626
  see also dental...; teeth
odor
  gastric contents **4**:2103, 2122
  ignitable liquid residues **3**:1138
  PMI determination **4**:2089
OEM paints *see* original equipment manufacturer paints
offender databases **2**:677, 680, 832–3; **5**:2680
  see also databases
offender profiles **4**:2156–60
  see also criminal profiling; perpetrator profiles
offense level hypothesis, paint evidence **4**:1951, 1952
offense level propositions **2**:973
offenses (cannabis use) **2**:435–6
  see also drug-facilitated crime
official document authentication **2**:840
offset lithography **3**:1261
OH –Fe^{4+}–P (hydroxy–ferryl–porphyrin radical) **3**:1648
Ohio v. Terry, 392 U.S. 1 (1968) **4**:2157
oil circuit breakers **2**:927
oil contamination **2**:948
Oil Red O (ORO) **3**:1299–300
old age determination **1**:181, 189
  see also age determination
older adults
  behavioral science evidence **1**:287
  cognitive changes **2**:450, 453
  cognitive interviews **2**:918
  death penalty **2**:718–9
  eyewitness memory **2**:917–8
  gagging fatalities **1**:226
  guardianships **3**:1371–80
  independent living capacity **2**:444–9
  inmates **2**:718–9
  offender sentencing **5**:2307
  person identification performance **2**:917
  prison inmate death penalty **2**:719
  sentencing **5**:2307
  suggestibility **2**:917–8
  traffic fatalities **5**:2543
  witnesses in court **2**:916–9
  see also elder abuse
oleander plant **4**:2063–4
olfactory hallucinations **3**:1432
Olmstead v. L. C. (98–536) 527 U.S. 581 (1999) **5**:2577
omission acts **2**:955

on-site testing
  sweat **5**:2426–7
  urine drug tests **2**:866–7
on-the-job training **5**:2546
ongoing guardianships **3**:1379
OOBNs (object-oriented Bayesian networks) **1**:279
OPCW *see* Organization for the Prohibition of Chemical Weapons
open access points, Internet **1**:142
open-ended questions **2**:554
operating system (OS), mobile phones **3**:1364
operational factors
  DNA databases **2**:680–1
  luminol use **3**:1648–9
OPGs (orthopantomographs) **4**:1891
opiates
  alcohol interactions **1**:115
  drug-facilitated crime **2**:872
  as illicit drugs **2**:847–8
  oral fluid **4**:1913–4, 1917
  repeated use **4**:2118
  urine testing **2**:865
  see also opioids
opinion evidence
  Bayes theorem **2**:487–8
  CAI **2**:492
  identification **3**:1438
  importance in disputed issues **2**:760
  lay versus expert witnesses **2**:758–9
  report writing **5**:2268, 2270
  statistics vs **2**:972
  'ultimate issue' **5**:2589–90
  see also expert opinion evidence
opinion identification evidence, definition **3**:1438
opioids **4**:1895–902, 2017
  see also opiates
opium **2**:854
  see also opiates
oppositional child parental alienation **4**:1982, 1983
optical bridge (microscopes) **4**:1771
optical brightening agents (OBAs) **4**:1968, 1974
optical cube (microscopes) **4**:1786–7
optical enhancement *see* optical techniques
optical features
  high power microscopy **4**:1760
  ink profiles **3**:1543
optical microscopy **4**:1686
optical prism (total station device) **2**:580
optical sensors **1**:325, 332, 334
optical storage media **2**:584
optical techniques
  evidence examination **4**:2036–56
  fingermark detection **3**:1293–4
  lighting mode enhancement **4**:2042–8
  photographic enhancement of evidence **4**:2040–9
  see also optical microscopy
optical variable devices (OVDs) **3**:1271
oral fluid **4**:1903–20
  applications **4**:1916–9
  cannabis tests **2**:436
  cocaine tests **2**:567
  collection techniques **4**:1906–7, 1909, 1911
  interpretation of results **4**:1908–16
  pharmacokinetics **4**:1908–16
  physiology **4**:1904–5
  sweat testing alternative **5**:2427–8
  see also saliva
oral route (drug use) **1**:137; **2**:435
  see also administration routes (drugs)
orbicularis oculi muscle **2**:700
Orfila, M. **4**:2120; **5**:2503–4
organ... *see* organs
organic amnesia **5**:2244
organic bases, inks **3**:1545
organic DNA extraction **2**:1061
organic drugs/poisons **4**:2128–31
  see also drug...; poisons
organic explosive analysis **2**:1038–9
organic gunshot residues *see* propellants, ammunition
organic matter, soil **5**:2384–5
Organization for the Prohibition of Chemical Weapons (OPCW) **2**:508
organizational accreditation **1**:10–6
  see also accreditation
organizational support, CAI **2**:493
organized-disorganized serial murder type **5**:2312–3
organs
  'puppet' organs **3**:1533
  removal **1**:260
  see also individual body organs
orientation
  minutiae of fingerprints **3**:1283
  neuropsychological assessment **3**:1461
origin of fire **3**:1129–32, 1133–5, 1136
origin of wood, definition **5**:2641
original equipment manufacturer (OEM) paints **4**:1943, 1945, 1946, 1947
original evidence **1**:298–9
Orkin v. Rogers, 478 F. Supp. 1342 (D. Mass. 1979) **5**:2585
ORO (Oil Red O) **3**:1299–300
O'Rourke, M. **4**:1958
orthographic hallucinations **3**:1433
orthopantomographs (OPGs) **4**:1891
orthoscopic lenses **4**:2038
OS (operating system), mobile phones **3**:1364
O'Shaughnessy, **2**:431
Osheroff v Chestnut Lodge, 490 A.2d 720 MD (1985) **5**:2582
osmium tetroxide **3**:1304–5
osteons **1**:183; **5**:2394–5
osteophytes **1**:196
other evidence ratio **5**:2366
out-of-court statements **3**:1466
outdoor scenes *see* outside scenes
outer morphology, friction ridge skin **3**:1322
outpatient civil commitment **2**:558, 571
outside scenes **1**:425, **2**:619
  see also scene...
outsole databases **3**:1254
  see also sole databases
outsourced workplace threat assessment **5**:2464–5
OVDs (optical variable devices) **3**:1271
over-stamping serial number obliteration **5**:2325
over-the-counter pharmaceuticals **2**:869
  see also medications
overall photographs **1**:386; **2**:627–8

overdoses **4**:1900, 2059, 2064
overexposure of photographs **2**:637–9
overhairs **3**:1405
overlay tracing impressions **3**:1244
overt document security features **3**:1265
overweight, definition **4**:1848
  see also obesity
oviposition (flies) **2**:936, 937, 940, 942
Oxford, Edward **3**:1553
oxidation
  alcohol analysis **1**:84–5
  light globes/filaments **3**:1638
  materials corrosion **4**:1685
oxidizers (ammunition primers) **3**:1190
oxycodone **4**:2018, 2019
oxygen isotopes **1**:309
oxytocin **2**:730

P (another match) error **5**:2404
P2P (Peer-to-Peer technologies) **1**:146
P300 peak deception detection **2**:725, 726
PAC (plasma alcohol concentration) **1**:83
PACE (Police and Evidence Act) **3**:1249
packaging evidence **2**:1029; **4**:1927–31
paclitaxel **4**:2065
Paddy's Bar *see* Bali bombings
PAEs *see* preventable adverse events
PAI (Personality Assessment Inventory) **4**:2183
pain relief **1**:115; **4**:1896
paint **4**:1931–42
  analytical methods **4**:1936–9, 1942
  application methods **4**:1932
  case histories **4**:1940–2
  composition **4**:1932
  databases **4**:1939, 1946, 1949–50, 1952
  definition **4**:1931
  forensic aspects **4**:1931–2
  found at random on clothing **4**:1950
  interpretation of evidence **4**:1943–53
  layer structure **4**:1933, 1935
  persistence **5**:2536
  pigment particle forms **4**:2003
  transfer **5**:2536, 2538
paint-with-light photography **2**:639
palatal rugae **1**:168; **4**:1893
palmar side, hands **3**:1323
palynology **1**:426; **4**:1954–67, 2002
  see also pollen
palynomorphs **4**:1954–67
Panetti v. Quarterman, 551 U.S. (2007) **2**:719
panic attacks **5**:2301
panning function, video **2**:642
PAP *see* prostatic acid phosphatase
Papaver somniferum **1**:426
  see also opium
paper
  analysis **1**:129, 132; **2**:686; **4**:1968–80
  basic properties **4**:1971–4
  fingermark detection **3**:1300, 1310
  fingerprint enhancement **3**:1319
  making **4**:1969–70, 1976–8
  physical properties **4**:1970–1
  spoof methods **1**:334–5, 335–6
  testing **4**:1969–70, 1971
  see also porous surfaces
paper fold (line intersection) **3**:1596
paper grabbers, printing devices **5**:2663
paper misfeeds, printers **5**:2670
papillary crests **3**:1513
papillary ridges **3**:1292
Paracelsus **5**:2509
paradigm shifts **3**:1280, 1448
paradoxical symptoms, benzodiazepines **1**:296
paraffin-embedded tissue **2**:826, 1063; **3**:1469
paraffin gunshot residue test **3**:1191
parallel history, psychopathology **4**:2189
parameter learning **1**:280
paranoid schizophrenia **2**:741
paraphilia **5**:2317, 2332–3, 2334, 2335
  see also sex offenders
parasitic genotypes **3**:1342–3, 1343–5, 1346
parcentration **4**:1759
parens patriae doctrine **2**:542; **3**:1373; **4**:1985, 1994; **5**:2576, 2577
parent-child relationship **4**:1984–5, 1985, 1987, 1996–7
  see also parental...; parenting
parent meetings **5**:2604
parentage testing **4**:1813–5; **5**:2357
  calculations **3**:1563; **4**:1810–7
  probability calculations **3**:1563
  Rhesus blood group system **1**:348
  short tandem repeats **5**:2362, 2680
parental aggression **1**:43
parental alienation **4**:1981–4
parental attachment evaluation **5**:2607
parental kidnapping evaluation **5**:2609
parental pressure, child interviews **2**:553–4
parental rights **4**:1984–9
parenting
  adequacy/inadequacy **4**:1995
  capacity assessment **4**:1989–98
  common elements **4**:1995
  custody/visitation evaluation **5**:2604, 2606–7, 2607–8, 2609, 2610
  observing parent and child **4**:1996–7
  see also child custody; parent-child relationship; parental...
Parenting Stress Index **5**:2606
Parenting Stress Inventory (PSI) **4**:1996
parenting time evaluation **5**:2610
parfocality **4**:1759
Parker v. Moore, 425 F.3d 250 (6th Cir., 2005) **4**:1821–2, 1822
Parkinson's syndrome **4**:2204
parricide **1**:263, 268
PARs (pseudoautosomal regions) **5**:2677
partial malingering **4**:1660
partial profiles, STRs **5**:2372
partial remains log **4**:1676
partial seizures **5**:2298, 2299, 2300
The Particle Atlas **4**:2001
particles
  analysis (gunshot residues) **3**:1192, 1193–4
  characterization (microscopy) **4**:1769–73, 1773–83
  forms **4**:2001–7

metal transfer **5**:2537
size (brightfield microscopy) **4**:1769
partitions, Macintosh computers **2**:585, 586
partner violence **1**:40, 41
see also domestic violence
party-appointed experts **2**:1010–1
passenger air bag systems **1**:51–7; **5**:2543, 2572
passive defense wounds **5**:2658
passive diffusion **2**:860; **5**:2421
passive headspace concentration **3**:1139–40
patch technology **5**:2422–3, 2425, 2426, 2427–8
patent fingerprints **2**:604–6
paternalism **5**:2450
paternity index likelihood ratios **3**:1562
paternity testing **4**:1813–5; **5**:2374
see also parentage testing
path animation of camera **5**:2259
path-counting coancestry method **4**:1811
path of ridges, fingerprints **3**:1283
pathogenic malingering model **4**:1658–9
pathogens *see* biological agents
pathological narcissism **5**:2399
pathologist role **1**:414; **2**:769–70
pathology
elder abuse perpetrators **2**:914–5
human remains **1**:194–6
individualization markers **1**:168–70
web resources **5**:2623
wildlife forensics **5**:2639
see also diseases
patient consent/refusal of treatment **5**:2584, 2586, 2587
patient waivers **2**:439–40
pattern of bloodstains *see* bloodstains/bloodstain patterns
pattern development, fire **3**:1112–21
pattern recognition, speech **5**:2391
PB (prosecution-biased) jurors **5**:2308
PC-Crash program **2**:581–2
PCBs *see* polychlorinated biphenyls
PCC (pyridinium chlorochromate) **2**:865
PCL : SV *see* Screening Version of the Hare Psychopathy Checklist-Revised
PCL-R (Psychopathy Checklist-Revised) **4**:2197–201
PCP (phencyclidine) **1**:292
PCPs (person-centered plans) **4**:1728, 1729
PCR *see* polymerase chain reaction
PCT (procalcitonin concentration) **4**:1856
PD *see* physical developer
PDAs *see* personal digital assistants
PDD (psychophysiological detection of deception) **2**:619–20
PDM (Psychodynamic Diagnostic Manual) **4**:2190
PDT trace metal detection **3**:1212–5
PE (pulmonary embolism) **4**:1846–7
peak heights (DNA) **3**:(1567; **4**:1840–1, 2007–8; **5**:2591
pedestrians
automobile accidents **5**:2254–5
biomedical analysis **5**:2566–7
injuries **1**:100
pediatric neurological exam (PNE) **4**:1870
pediatric psychopharmacology **4**:2210–6
pedo-pornography **1**:190
pedology, definition **5**:2378
pedophiles *see* childhood sexual abuse
peer pressure **2**:553–4; **3**:1610–1
peer reviews **2**:693, 694; **4**:2009–11
Peer-to-Peer (P2P) technologies **1**:146
pellet impacts (shotguns) **3**:1221
pelvis **1**:156; **5**:2328–9, 2330
pencil erasures, documents **1**:131
see also writing instruments
Pendleton v. R [2002] 1 W.L.R. 72 **2**:1000
penile plethysmography **5**:2333
penis injuries **2**:534
pennate diatoms **2**:749
pens *see* writing instruments
pentaerythritol tetranitrate (PETN) **2**:1038, 1041, 1053
pentane **3**:1141
People v. Ebanks, 117 Cal. 652, 665, 49 p. 1049, 1053 (1897) **3**:1500
People v. Higgins, 5 N.Y.2d 607 (NY. 1959) **1**:254
People v. McHugh 124 Misc.2d 559 (N.Y. Supp. 1984) **2**:580–1
PEP *see* primer extension preamplification
PEP-PCR *see* primer extension preamplification polymerase chain reaction
per exclusionem diagnosis **4**:2086, 2087
per se legislation **2**:880–1, 881, 882; **3**:1500–2, 1502
see also zero tolerance approach
percent elongation, materials **4**:1682
perception–response time, automobiles **5**:2255–6
perceptions
observer effects **3**:1575
of risk **5**:2455–6
perceptual speech analysis approach **5**:2390
perchlorates **2**:1044, 1046
perforating shot wound track **3**:1384
perforations, documents **3**:1272
performance, influence of drugs on **2**:877–8; **5**:2507
performance-enhancing drugs **5**:2507
performance testing biometric devices **1**:329
perimortem injuries **1**:226
perimortem specimens **5**:2501–2
perimortem trauma **5**:2559
perinatal disorders **4**:2136–40
periorbital hematoma **1**:399
periostitis **5**:2558
peripheral blood **5**:2496
perjury prosecutions **2**:1015
permanent fixed contact traces **2**:994
permanent vs temporary insanity **5**:2445
peroxidase **2**:865; **3**:1648
peroxide explosives *see* diacetone diperoxide; hexamethylene triperoxide diamine; triacetone triperoxide
perpendicular planes **4**:1779, 1780
perpetrator-absent lineups **2**:1073
perpetrator characteristics *see* perpetrator profiles
perpetrator dysfunction, elder abuse **2**:906
perpetrator images **3**:1625–6, 1628
see also facial recognition; mugshots
perpetrator models **1**:390–5

perpetrator profiles **1**:265–6; **2**:530–1; **3**:1477
see also offender profiles
perpetrator risk factors, elder abuse **2**:914–5, 915
see also care issues
perpetrator substitution **2**:535
perpetrators as victims **5**:2432–5
persecutory delusions **2**:741, 743
persistence of evidence **5**:2534–9
animal hair **3**:1403–4
fibers **3**:1098; **5**:2535
friction ridge skin **3**:1327
hair **3**:1403–4, 1418; **5**:2535
low copy number DNA **3**:1643
paint **4**:1933, 1951; **5**:2536
probability **2**:971
see also transfer of evidence
person-centered plans (PCPs) **4**:1728, 1729
person-context information, formal testing **4**:2175
person–context interactions **2**:453
person identification *see* personal identification
personal digital assistants (PDAs) **2**:584, 586–7
personal effects, disaster victims **2**:769–70, 770
personal identification **1**:153, 166–71, 171–2; **2**:917; **3**:1514
see also identification; individualization
personal nature of propositions **2**:969
personal protective equipment (PPE) **2**:503, 504
personal risk factors, violence **5**:2273
Personality Assessment Inventory (PAI) **4**:2183; **5**:2605
personality disorders **2**:572; **3**:1476; **5**:2274, 2434–5, 2578
see also psychopathy
personality testing **2**:465, 466; **4**:2175, 2179–80, 2183–4
see also dangerousness; deception detection; Minnesota Multiphasic Personality Inventory
personalized justice **4**:2020
personnel qualifications, laboratories **1**:7–8
perspective distortion, photography **4**:2037
persuasion **1**:287; **2**:732
perversion **3**:1476; **5**:2314
pests *see* biological agents
PET (positron emission tomography) **5**:2301
petechial hemorrhages **1**:225; **3**:1532
petition filing, guardianships **3**:1378
PETN *see* pentaerythritol tetranitrate
petrography **5**:2384
petroleum
distillates **3**:1143, 1150, 1155–61, 1163–5
gas **3**:1107
persistence **5**:2537
transfer **5**:2537
peyote cactus **4**:2063
PFA (psychological first aid) **2**:761
PGC *see* Pyrolysis–Gas Chromatography
PGM (phosphoglucomutase) **1**:353–4
pH manipulation, urine tests **2**:863–4
phalanges (fingers) **3**:1323
pharmaceutical safety research **5**:2508–9
see also pharmacology
pharmacodynamics **1**:109–10; **2**:563–6; **4**:2202, 2207, 2213
pharmacogenetics **4**:2012, 2013–4
pharmacogenomics **4**:2012–20
see also pharmacogenetics
pharmacokinetics **4**:2115–6, 2202, 2207
alcohol–drug interactions **1**:109
benzodiazepines **1**:294–6
cocaine **2**:565–6
oral fluid **4**:1908–16
pediatric psychopharmacology **4**:2213
pharmacological excitability, iris **2**:701–2; **5**:2468
pharmacology
alcohol–drug interactions **1**:109–10
amphetamines **1**:136–8
benzodiazepines **1**:294–6
cocaine **2**:563–6
opioids **4**:1896–900
sex offender treatment **5**:2334–5
THC **2**:434
see also drug…; medications
pharmacotherapy **1**:20; **4**:2215; **5**:2505
PharmChek patch **5**:2422–3
phase contrast microscopy **4**:1760
PhD programs **2**:899, 901
phencyclidine (PCP) **1**:292
phenethylamines **2**:849
see also amphetamine-type stimulants; amphetamines; ecstasy
phenol–chloroform DNA extraction **2**:803–4
phenothiazine drugs **4**:2204
phenotypes **1**:339
active loci/variants **4**:2022
crime-scene DNA **4**:2021–34
direct inference method **4**:2022, 2029–33
indirect inference method **4**:2023–8
philosophical basis, CAI **2**:485–8
phone forensics **2**:584, 586–7; **3**:1360–71
phonetic speech analysis approach **5**:2390
phonotactics **5**:2391
Phoridae ('coffin-flies') **2**:943
phosphatase **2**:954
see also acid phosphatase
phosphoglucomutase (PGM) **1**:353–4
phosphorescence **4**:1785, 2046
phosphoric acid tests **4**:1747
photocopied documents **1**:131, 132, 133; **3**:1559–60
photocopiers **5**:2667, 2672–3
photocopy paper **4**:1972
photodiode array (DAD) **2**:597–8
photogrammetry **3**:1626
photographic log **2**:608
photographic paper **3**:1310
photography
aerial **3**:1497
authentication **2**:840
best evidence rule **1**:298
crime scenes **2**:625–43
disaster victim postmortem **2**:769
documentation **1**:373–4, 386; **2**:603, 608, 625–43
evidence **2**:625–43, 746–7; **4**:2036–56
casting procedure **2**:963
documentation **2**:603, 608
examination **4**:2036–56
fire scenes **3**:1124

see also images, processing
photoionization detectors (PIDs) **2**:521
photoluminescence **3**:1294; **4**:2042, 2046–8
photometric inaccuracy **4**:1752–5
photoreceptors **4**:2041
Photoshop **4**:2054–5
PHT *see* pulmonary hypertension
phylogeny **1**:423; **4**:1824, 1827–8, 1829
physical abuse
  battered child syndrome **1**:263–9
  elders **2**:903, 904–5, 905, 907, 914
  radiological diagnosis **5**:2236
  see also abuse
physical barriers, upper airway **1**:225–6
physical characteristics **1**:324
  see also human characteristics
physical developer (PD) **3**:1299–300, 1303, 1319
physical evidence
  authentication **2**:840–1
  automobile accident reconstruction **5**:2250
  barefoot impressions **3**:1244
  best evidence rule **1**:298
  bloodstain patterns **1**:360–96
  documentation **2**:602
  foundation testimony **3**:1277
  photographic/optical examination **4**:2036–56
  see also evidence
physical examination
  battered child syndrome **1**:267
  child sexual abuse **2**:534
  explosion debris **2**:1032–5
  see also examination...
physical exercise **5**:2426
physical fit evidence **1**:427; **4**:1933
  see also fracture matching
physical handicaps **1**:281
  see also disability
physical illnesses **3**:1475
  see also diseases
physical modeling **3**:1177
physical requirements, legal responsibility **1**:284
physicians
  patient confidentiality **2**:885–6, 887–8, 889
  reporting battered child syndrome **1**:268
  serial murderers **5**:2315
  see also doctors; mental health professionals
physiochemical properties, ethanol **1**:58
physiology
  cannabis effects **2**:433
  drug distribution in sweat **5**:2420–2
  friction ridge skin **3**:1326–7
  hair **3**:1415–8
  oral fluid **4**:1904–5
phytane **3**:1157
picking-off fiber recovery method **2**:986
PIDs (photoionization detectors) **2**:521
piezoelectric inkjet printers **5**:2670
piezoelectric sensors **1**:327
piezoelectric transducers **3**:1201, 1204
pigments (paint) **4**:1932
pilocarpine stimulation **5**:2422
piloted ignition temperature **3**:1109
pin wear faults, printers **5**:2667
pinch point, bullet impacts **3**:1220
pincushion distortion **4**:2038
Piotrowski, Dr Eduard **1**:361
pipe bombs **2**:823
Piper methysticum **4**:2064
pixels
  defects **3**:1523
  infrared imaging **4**:1756–7
pj values **5**:2368, 2371
placards (photography) **2**:626, 628, 632–3, 634–5
placement continuity evaluation **5**:2606
planchets **3**:1267
plane polarized light **4**:1776, 1778–9, 1779
plankton **1**:428, 429
  see also centric diatoms
planning interpretation activities **3**:1581
plant criteria, laboratories **1**:8
plants/plant material **1**:423–5
  classification systems **1**:423–5
  DNA
    analysis **1**:429
    extraction **2**:1064
    profiling **5**:2363
    sources **2**:827–8
  fiber sources **4**:1976–80
  food residues **1**:427–8
  identification **1**:424, 425–6, 429
  as illicit drugs **1**:425–6
  poisonous plants **4**:2057–67
  radiocarbon dating **1**:421
  toxicology **1**:428
  see also botany; pollen
plasma
  blood ratios **4**:2098
  cocaine concentrations **2**:563, 565–6
  postmortem toxicology **4**:2128
  saliva-to-plasma ratios **4**:1906
plasma alcohol concentration (PAC) **1**:83
plastics
  deformation **4**:1681, 1682
  failure analysis procedures **4**:1687
  fingerprint enhancement **3**:1319–20
  firearm discharge residue patterns **5**:2352
  see also nonporous surfaces
platen marks **5**:2673
plausible presentation, malingering **4**:1659
pleochroism **4**:1778
pleura sign **3**:1537
plexiform pattern, bones **1**:154
PLM *see* polarized light microscopy
'plug-in' values, population genetics **5**:2371
plywood **5**:2352
PM *see* postmortem...; probability of a match
PMA *see* 4-methoxyamphetamine
PMDD *see* premenstrual dysphoric disorder
PMI *see* postmortem interval
PMR *see* postmortem redistribution
PMS *see* premenstrual syndrome
PNES (psychogenic nonepileptic seizures) **5**:2301
pneumonia **4**:1843, 1844–6
pneumothorax **1**:408
POCT (point-of-care testing) **4**:2228
poikilothermic organisms **2**:941
point-of-care testing (POCT) **4**:2228
point-to-point search strategy **2**:623

points *see* minutiae (fingerprints)
poisons
  antemortem factors **4**:2094
  Camelford case **2**:948–9, 949, 952
  definition **4**:2120
  environmental toxicology **5**:2508
  epidemiology **5**:2508
  plants **1**:428; **4**:2057–67
  postmortem alterations **4**:2101
  postmortem toxicology **4**:2120, 2122, 2127, 2128–31, 2133–4; **5**:2504–5, 2505
  small molecules **4**:2133–4
  see also toxicology
Poisson distribution **4**:1946
polarization colors **4**:1780, 1783
polarized light microscopy (PLM) **2**:989; **3**:1544; **4**:1743, 1773–84
polarizing microscopy **4**:1760
police
  citizen encounters **2**:650–1
  crime scene investigation **2**:614–5
  disaster victim identification **2**:765–6
  explosion scene role **2**:1021
  firearms laboratories **3**:1216
  first report to **2**:619–25
  induced confessions **2**:589, 590; **3**:1586–8, 1589
  interrogations **2**:590–1, 592; **3**:1586–90
  notifying (crime victims) **2**:643–7
  preinterrogation interviews **2**:589–90
  training **4**:2072
  use of force **4**:2068–71, 2072–4
  see also law enforcement agencies; SWAT teams
Police and Evidence Act (PACE) **3**:1249
'policeman at the elbow' test **3**:1555
policy implications
  adolescent development studies **3**:1611
  juvenile transfer to adult justice **3**:1614–5
political boundaries **2**:950
political dimension, illicit drugs **2**:854–5
political/hatred murders **3**:1476
pollen **1**:421, 426; **4**:1954–67, 2002; **5**:2537
pollen rain **4**:1954, 1965
pollution toxicology **5**:2508
polyads **4**:1959
polyampholytes **1**:339
polychlorinated biphenyls (PCBs) **3**:1334
polyclonal antisera **1**:347
polycyanoacrylate **3**:1300
polyethylene **3**:1110, 1160–1
polygraph testing **2**:721–3
polymarker loci **2**:842
polymer... *see* polymers
polymerase chain reaction (PCR) **1**:126; **2**:798, 800, 801; **3**:1639; **4**:1749–50, 1834, 2015
  AMP FLPs **1**:140
  biological agent detection **1**:305
  bone/teeth DNA **2**:826
  Chelex® extraction **2**:1061, 1062
  degraded DNA samples **2**:817
  DNA quantification **2**:806
  DNA samples **2**:817; **5**:2628
  fecal DNA **2**:827
  HLA-DQα **2**:842
  low template DNA **2**:808, 809–10; **3**:1566, 1567
  mini-STRs **4**:1805, 1806
  paradigms **3**:1643
  plant DNA evidence **2**:827–8
  PowerPlex **4**:2149
  primer extension methods **5**:2628–30
  sensitivity increase **3**:1642
  spermatozoa DNA **2**:822
  stochastic effects **3**:1640
  STRs **2**:801; **4**:1805, 1806; **5**:2356, 2357
  VNTR analysis **5**:2595
  whole genome amplification **2**:808
polymers **3**:1109–11
  see also resin
  polymorphisms **4**:2075; **5**:2595
  AIMs **4**:2025–8
  DNA genetics **2**:795–6
  HLA-DQα system **2**:842
  identification markers **1**:339–56
  Y chromosomes **4**:2023
  see also individual types
polyomavirus JC (JCV) **3**:1342–6
polyphenolic derivatives **3**:1651
polyvinyl chloride (PVC) **3**:1110, 1111
Popp, George **5**:2380
Popper, Karl **2**:694; **3**:1093–4
poppy plant **1**:426
  see also opium
poppy seeds **2**:865; **4**:1914
population determination, trace evidence **5**:2292
population genetics parameter (θ or F_{ST}) **5**:2367, 2368, 2370, 2371
population sampling **5**:2281, 2284
population-specific estimates, ceiling principle **2**:497–8
population statistics **4**:1810; **5**:2408
population studies **4**:1836, 1944–5, 1948, 1950
pores of ridges (fingerprints) **3**:1284
pornography, juvenile **1**:190
porous surfaces
  bloodspatter degree **1**:365
  explosion debris **2**:1030
  fingermark detection **3**:1292, 1296–300
  fingerprint enhancement **3**:1319
  lifting bloodstain evidence **1**:388
  wettability **1**:369
porous tip pens **5**:2662
Portuguese mtDNA database **4**:1830
posing concept **5**:2316
positional asphyxia **1**:227–8
positioning facial comparisons **3**:1082
positive identification
  definition **1**:172
  human remains **3**:1511–6
  odontology **4**:1891–3
  see also identification
positron emission tomography (PET) **5**:2301
possession of tangible evidence *see* chain of custody
possible identification, definition **1**:172
post scene documentation **2**:608–10
postanalytical laboratory operations **4**:2220
postassault management **1**:240
postblast investigation **2**:1019–20, 1023, 1027
  see also explosions

postcollection artifacts **4**:2118
postcollision prevention measures **5**:2543
postconcussional disorder **3**:1460
postconviction actions **2**:999
postconviction testing **2**:682–3
posterior odds **2**:970; **5**:2493
posterior probabilities **2**:487; **3**:1562
postmortem (PM)
  biochemical examinations **4**:2076–87
  data (odontological) **4**:1891–2
  diagnosing neck compression **1**:228
  disaster victim identification **2**:769–70
  histology **1**:260; **3**:1468–73
  lesions **5**:2560
  material for positive identification **3**:1512
  samples (diatoms) **2**:753–4
  shooting scene investigation **3**:1223
  suicide risk assessment **4**:2165–6
  traffic fatality examination **5**:2544
  see also autopsies; human remains; postmortem toxicology
postmortem changes **2**:697–716; **3**:1529; **4**:2089, 2091, 2096–102; **5**:2467
  see also decomposition of remains; putrefaction
postmortem interval (PMI) **2**:936–9, 942–3; **4**:2089–92
postmortem redistribution (PMR) **2**:567; **4**:2096–7, 2121
postmortem toxicology **4**:2208; **5**:2504–5
  analysis **4**:2119–34
  artifacts **4**:2093–108, 2118
  BAC interpretation **1**:70
  benzodiazepines **1**:296
  databases **5**:2508
  opioid data **4**:1902
  plant analysis **4**:2066
  specimens **1**:89–90; **4**:2101–6; **5**:2495–502
postnatal depression **4**:2137–8
postnatal psychosis **4**:2136–40
postpartum depression (PPD) **4**:2137–8
postpartum psychosis (PPP) **4**:2136–40
postquem non temporal limit, PMI **4**:2091
postsynaptic receptors **4**:2203
posttraumatic growth **4**:2142
posttraumatic stress disorder (PTSD) **2**:785; **4**:2141–7; **5**:2243
  Allewalt case **5**:2436, 2437
  child abuse **4**:1990
  child sexual abuse **2**:536
  definition **1**:288; **4**:2141; **5**:2432, 2436, 2437
  diagnosis **4**:2142–6
  DSM definition **5**:2436, 2437
  hallucinations **3**:1433
  incomplete encoding **5**:2245
  juvenile correctional settings **3**:1614
  malingering **4**:2141, 2143, 2144–6
  perpetrators **1**:27; **5**:2434
  rape trauma syndrome **5**:2241–2
  remission **4**:2142
  risk factors **4**:2141
  symptoms **4**:2143, 2146
  traumatic brain injury **3**:1464
  treatment **4**:2146
  victims of violence **1**:27
  see also traumatic memory
postural asphyxia **1**:227–8
posture, facial comparisons **3**:1082
potassium **4**:2085; **5**:2472–3
potency of amphetamines **1**:136–7
potential source population, definition **5**:2408
potentiated drug effects **1**:110
pour patterns, fire **1**:214
powder brush DNA transfer **3**:1320
powder tattooing (shootings) **3**:1223
powdering
  fingermark detection **3**:1294–5, 1302, 1307
  fingerprint enhancement **3**:1318–9, 1320
power (energy), definition **3**:1104
power factor, elder abuse **2**:906
PowerPlex systems **4**:2149; **5**:2362
powers of attorney **3**:1374, 1375
PPD (postpartum depression) **4**:2137–8
PPE *see* personal protective equipment
PPP (postpartum psychosis) **4**:2136–40
pragmatic inference **4**:1710
preanalytical laboratory operations **4**:2220
preassessment, CAI **2**:974
precipitation techniques **2**:599
precollision prevention measures **5**:2542–3
preconcentration step, ILRs **3**:1141
preconceptions, confirmation bias **3**:1575–7
precrash data, air bag systems **1**:56
predatory violence **5**:2398
predictions, drug concentrations **4**:2117–8
predictive reasoning **1**:278
predictive validity **4**:2200
pregnancy **1**:411, **4**:1847–8; **5**:2331
preinterrogation interviews **2**:589–90
prejudice, juries **3**:1604, 1605
prejudicial effect (evidence) **2**:581, 746
prelineup instructions, eyewitnesses **2**:1073
premenstrual dysphoric disorder (PMDD) **4**:2150, 2152
premenstrual syndrome (PMS) **4**:2149–56; **5**:2446–7
premorbid function **4**:1863
premortal changes, time of death **5**:2467
preparation
  calibration solutions **4**:2223–4
  child cross-examination **2**:659–60
prerogatives, trier of fact **1**:286–7
preschool children **5**:2604
  see also infants
prescription medications **2**:869
  see also medications
presentation, expert testimony **2**:998
preservation of evidence **1**:321; **2**:944, 963–7; **4**:2104
President's DNA Initiative **4**:1837
pressure abrasions **1**:397
pressure ulcers **2**:905
presumption of sanity **5**:2431
presumptions of law, expert evidence **2**:1007
presumptive analysis, chemical warfare agents **2**:511–22
presumptive identification, definition **1**:172
presumptive tests **1**:314, 417; **3**:1645, 1648–54
pretrial defendant treatment refusal **5**:2586

pretrial motions **2**:773
pretrial proceedings, juveniles **2**:547
prevalence
  addictions **1**:19
  drug impaired driving **2**:878–9
  epilepsy **5**:2298
  female aggression **1**:36–8
  psychopathy **4**:2194
  violence **1**:28
preventable adverse events (PAEs) **4**:1690, 1691, 1692, 1703
preventative law **5**:2452
prevention measures, traffic fatalities **5**:2542–3
pRIA (protein radioimmunoassays) **1**:154
prima facie cases **2**:1006
primary alcohols **1**:81
primary delusions **2**:741
primary fluorescence **4**:1785
primary legal custody **5**:2609
primary physical custody **5**:2609
primary scenes **1**:415, **2**:621
primary transfers
  animal hair **3**:1404
  definition **5**:2378
  DNA evidence **5**:2556
primer binding site mutations **5**:2359–61
primer design, mini-STRs **4**:1807
primer extension preamplification (PEP) **5**:2629–30, 2632
primer extension preamplification polymerase chain reaction (PEP-PCR) **2**:808
primer pairs, mtDNA **4**:1834
primer residues *see* gunshot residues
primers (firearms) **3**:1190, 1200, 1204
print-stack collapse fault **5**:2667
printed documents **3**:1557–61
  see also documentary evidence
printing devices **1**:130, **5**:2660–1, 2663–74
printing inks **3**:1541, 1544
see also ink...
printing processes **3**:1261–5, 1267–9
prints *see* fingerprints
prior contact, LTDNA **3**:1569
prior function **4**:1864
prior odds **2**:970, **5**:2493
prior probabilities **2**:487, **3**:1562–3, 1564
prior ratio **5**:2366
prisoners
  civil commitment **2**:559
  right to treatment **5**:2581, 2582–3
  treatment refusal **5**:2586
  see also incarceration; inmates; jail populations
prisonization **2**:573
pristane **3**:1157
private actions, crime victims **2**:646
private professional guardians **3**:1372
privilege
  sex offender treatment **5**:2336
  vs confidentiality **2**:886
pro forma documents **3**:1440
pro se (self-representation) defense **2**:462
probability **2**:968–76; **3**:1579
  allelic frequency **4**:1812
  basic concepts **2**:969
  Bayesian networks **1**:276–80
  causality concept **5**:2492, 2493
  of discrimination **3**:1581; **4**:1943
  DNA databases **2**:832, 833–4, 834–5, 835, 836–7
  earprint matching **2**:892–3
  examples **2**:968–9
  fingerprint interpretation **3**:1278, 1280
  handwriting examination **3**:1440
  Hardy–Weinberg equilibrium **3**:1458
  ibd alleles **4**:1810, 1811, 1812
  intervals **5**:2288
  laws of **2**:969
  LTDNA profiles **3**:1572–3
  match statistics **3**:1581; **5**:2402, 2404, 2405–6
  models **3**:1100–1, 1177, 1283, 1287–9; **5**:2487
  mtDNA evidence **4**:1827–8
  reverse causality **5**:2492–3
  sampling trace evidence **5**:2293
  significance probabilities **3**:1581–2
  single-source identification **5**:2366, 2367–70, 2371–2, 2373
  soil transfer/retention **5**:2380
  see also Bayes theorem
probability of a match (PM) **3**:1581
probative effect, demonstrative evidence **2**:746
probing methods, archaeology **1**:204
procalcitonin concentration (PCT) **4**:1856
procedural issues, phenotype inference **4**:2032–3
procurement of forensic services **2**:494
procureur see state prosecutors
product analysis **5**:2568
product defects **3**:1484, 1487–8, 1495
product liability **3**:1483–95; **4**:2164
product lifetime **3**:1485
product need **3**:1485–6
'product rule' **5**:2369–70
product test of insanity **1**:283; **3**:1553; **5**:2444
productions
  disaster victim identification **2**:770
  documentation **2**:607
  traces of (illicit drugs) **2**:853
professional associations **2**:663–4, 957–62, 1014–5
  see also forensic associations
professional literature **2**:664–5, 693, 694; **3**:1623–4
  see also journal resources; learned treatises; scientific literature
professional societies *see* professional associations
proficiency testing **1**:6, 7, 8; **2**:956; **4**:2225
Profiler Plus™ system **2**:570
profiling *see* behavioral profiles; criminal profiling; DNA profiling; illicit drugs, profiling; psychological profiling; soil profiling
prognathism **1**:192
prohibited conduct, definition **2**:958
projected bloodstain patterns **1**:378–83
projectiles **3**:1189
  analysis **3**:1197–8
  deflection **3**:1384
  embolization **3**:1385–6
  external ballistics **3**:1202
  impact damage **3**:1220
  intermediate ballistics **3**:1201
  internal ballistics **3**:1200–1

stopping power **3**:1385
terminal ballistics **3**:1202–4
see also ammunition; bullets
projective tests **4**:2178, 2184
Prometheus myth **3**:1225
promulgation of myths **1**:208–10
proof
burden of **1**:282
of cause **5**:2436–8
facility **3**:1201, 1204
of guilt **2**:1005
of possession **2**:498–500
propaganda **2**:732
see also persuasion
propane combustion **3**:1108
propellants, ammunition **3**:1190
property ownership **3**:1376
propositions **2**:969
defense **2**:970, 971–2
hierarchy of **2**:487–8, 493–4, 973; **4**:1951–2
interpretation principles **2**:971–2
personal nature of **2**:969
prosecution **2**:970, 971–2; **3**:1580–1
refinement (handwriting comparison) **3**:1454–6
proprietary software **2**:587
prosecution
discovery **2**:775, 779, 780, 783
environmental cases **2**:946
profile testimony **4**:2159–60
propositions **2**:970, 971–2; **3**:1580–1
trial and appeals procedure **2**:998–9
prosecution-biased (PB) jurors **5**:2308
prosecutor's fallacy **2**:975; **3**:1562, 1583; **5**:2403
see also transposed conditional
prosody **5**:2391
prostate-specific antigen (PSA) **1**:316
prostatic acid phosphatase (PAP) **1**:17
prosthetic devices **1**:194
protection patterns (fire) **1**:215; **3**:1119
protective devices, child traffic safety **5**:2543
protective factors
juvenile projects **4**:1880
mitigation testimony **4**:1819, 1820
threat assessment **5**:2457
protein polymorphisms **1**:339, 340, 349–55
protein radioimmunoassays (pRIA) **1**:154
proteinases **2**:954, 1061
proteins **1**:184; **5**:2396
proteolytic enzymes **1**:346
protocols, network analysis **1**:141, 142, 143, 145, 146, 148
protrusion of tongue, burns **3**:1532
provenance of wood **5**:2641
provision/interpretation of results distinction **2**:494
proximal region, hands/feet **3**:1322
proximate cause (disaster mental health) **2**:763
proxy servers **1**:143, 144
Prussian edict (1875) **1**:257
PSA (prostate-specific antigen) **1**:316
pseudoalkaloids **4**:2065
pseudoautosomal regions (PARs) **5**:2677
pseudobruises **1**:232
pseudocount methods **5**:2371
pseudo kerosene **3**:1160
pseudo-peroxidase activity **3**:1648
pseudoseizures **5**:2301
PSI (Parenting Stress Inventory) **4**:1996
psilocin **2**:849–50
Psilocybe species **4**:2064, 2066
see also magic mushrooms
psilocybin **2**:849–50
psychiatric advance directives **2**:571
psychiatric disorders **3**:1337–8, 1373; **4**:2017
see also mental illness; psychiatric illness; psychological syndromes
psychiatric drugs **4**:2131, 2134
see also medications
psychiatric hospitalizations **2**:556, 649
psychiatric illness **2**:556–60
see also mental illness; psychiatric disorders
psychiatric malpractice cases **4**:2164
psychiatric testing **4**:2174
see also psychological testing
psychiatric treatment **2**:439–40
see also psychiatric drugs; psychological treatment; psychopharmacology
psychiatrists
ethical issues **2**:443
training **4**:2174
see also mental health professionals
psychoactive drugs **4**:2202–3
psychoactive plants **4**:2059, 2064
Psychodynamic Diagnostic Manual (PDM) **4**:2190
psychogenic amnesia **5**:2244
psychogenic nonepileptic seizures (PNES) **5**:2301
psychological abuse (elders) **2**:904, 905
psychological autopsy **1**:246–7; **4**:2161–73
psychological first aid (PFA) **2**:761
psychological-legal perspectives **4**:2182–3
psychological profiling **4**:2156–60
'psychological self-defense' **1**:271
psychological syndromes **5**:2431–40
see also psychiatric disorders
psychological testing **4**:2173–84; **5**:2605–6
see also deception detection; neuropsychological assessment; risk assessment
psychological treatment **2**:760; **5**:2335, 2337
see also psychiatric treatment
psychologists
expert testimony **2**:593
training **4**:2174
see also mental health professionals
psychometric properties, Hare Checklists **4**:2199
psychometric validation **4**:2189
psychopathic personality disorder *see* psychopathy
psychopathology **3**:1226, 1227; **4**:1659, 2186–91
psychopathy **4**:2193–7, 2197–201; **5**:2273
see also antisocial personality disorder; personality disorder
Psychopathy Checklist-Revised (PCL-R) **4**:2197–201
psychopharmacology **4**:2201–8
see also addictions; pediatric psychopharmacology; substance abuse
psychophysiological detection of deception (PDD) **2**:619–20
psychosis
capacity to waive Miranda rights **2**:466

psychosis *(cont)*
competency to stand trial **2**:458, 459
drugs and **1**:292, 293
hallucinations **3**:1433
medication **4**:2203–5
multiple murder causes **3**:1476
tort law **1**:282
see also postpartum psychosis; psychotic disorders
psychosocial dysfunction **4**:2143
psychosocial sequellae, child sexual abuse **2**:535–6
psychotherapeutic medications **4**:2211–2
see also medications
psychotherapists **5**:2605
see also mental health professionals
psychotic disorders **2**:434, 563–5
see also psychosis
psychotropic medications **2**:457
PTE (pulmonary thromboembolism) **4**:1847
PTSD *see* posttraumatic stress disorder
puberty stages **1**:190
pubic bones **5**:2328
pubic symphysis method **1**:181, 183
public authority expert opinions **2**:1009
public education, sex offenders **5**:2337
public policy, syndromes **5**:2438–40
public records' authentication **2**:840
public safety, mandated treatment **5**:2579
puerperal psychosis **4**:2136–40
pugilistic attitude **3**:1532
pull-ups (DNA) **1**:127; **2**:805, 808; **4**:1689
pulmonary chemical warfare agents **2**:501
pulmonary embolism (PE) **4**:1846–7
pulmonary hemorrhage **1**:227
pulmonary hypertension (PHT) **4**:1849–50, 1850
pulmonary thromboembolism (PTE) **4**:1847
pulmonary tissue burns **3**:1535, 1537
pulmonary..., *see also* lung
pulp **4**:1968, 1970, 1975, 1978
see also fiber...
punching serial numbers **5**:2325
punishable concentration limits, alcohol **1**:82–3
puparium of fly **2**:941, 944
'puppet' organs **3**:1533
pure gases **3**:1107
purines **2**:798
purpose section of statements **2**:974
putrefaction **4**:2089, 2098, 2099, 2100, 2101, 2123; **5**:2474–8
see also postmortem changes
PVC (polyvinyl chloride) **3**:1110, 1111
pyridinium chlorochromate (PCC) **2**:865
pyrimidines **2**:798
pyrolysis **3**:1107, 1111
Pyrolysis–Gas Chromatography (PGC) **2**:991; **4**:1938
Pyrolysis–Mass Spectrometry **2**:991
pyromania **3**:1226–7, 1227

Q see questioned samples
QC *see* quality control
QiaAmp® DNA extraction **2**:805, 1062–3; **4**:2219
QM *see* quality management
QPNs (qualitative probabilistic networks) **1**:279
QT interval
left ventricular hypertrophy **2**:471
long QT syndromes **2**:470–1, **4**:1851–2, 1853
obesity **4**:1849
qualifications
experts **2**:663–4, 759, 1012–3, 1014
Hare Checklist users **4**:2198
laboratory personnel **1**:7–8
see also certification; credentials of experts; education
qualified exclusion, handwriting **3**:1441
qualified identification, handwriting **3**:1441
qualifying process (experts) **2**:759
see also qualifications
qualitative data, DNA **4**:1839–40
qualitative descriptions **5**:2401
qualitative differences, traumatic memory **5**:2243
qualitative opinions **2**:972
see also opinion evidence
qualitative probabilistic networks (QPNs) **1**:279
qualitative testing **4**:2134
quality
assurance measures (crime scenes) **2**:624–5
CAI approach **2**:483, 493
DNA profiles **3**:1567
evidence photography **4**:2039, 2053
expert opinion evidence **2**:483
facial comparison images **3**:1082, 1085
handwriting exemplars **3**:1439
images **3**:1082, 1085; **4**:2039, 2053
mtDNA databases **4**:1829–30
scene footwear impressions **3**:1250
systems (postmortem toxicology) **4**:2134
see also quality management
quality control (QC) **4**:2224, 2225–7
quality management (QM) **1**:6, 7, 12, 13, 15, 16; **4**:2219–29; **5**:2493–4, 2554
quality triangle **3**:1171–2
quantification
DNA **2**:805–6
handwriting exemplars **3**:1439
human remains **1**:155
see also quantity estimation
quantitation *see* limit of quantitation
quantitative descriptions **5**:2401–2
quantitative methods **2**:599–600
head acceleration analysis **5**:2572–4
living individual identification **5**:2614
systematic toxicological analysis **4**:2127
validation **4**:2134, 2221
quantity estimation **5**:2281–90
see also quantification
Quarterman v. Panetti, 551 U.S. (2007) **2**:719
Queen Charlotte Islands' tree species **5**:2641
questioned documents **5**:2626, 2638
see also documentary evidence
questioned hairs **3**:1419, 1421
questioned persons, digital images **3**:1625–6
questioned samples (*Q*)
mitochondrial DNA **4**:1823, 1824, 1825–7
paint **4**:1935
soil **5**:2383
questioned writing

definition **3**:1438
identification/exclusion approach **3**:1443
signatures **3**:1447
timing **3**:1440
questioning rules (cross-examination) **2**:656
see also leading questions

R&D *see* research and development
R v. Doheny and Adams [1996] EWCA Crim 728 **2**:975
R v. J.-LJ [2000] 2 SCR 600 [10] **2**:1004
R v. Pendleton [2002] 1 W.L.R. 72 **2**:1000
R v. Silverlock [1894] 2 QB 766 [3] **2**:1003
R v. Sullivan, 2 A111 E.R. 673, 675–676 (1983) **1**:255
R v. Ward (1993), 1 WLR 619 **2**:782; **5**:2270
rabbit experiments **1**:361
rabbit hair **3**:1407
race
aging the living **1**:188
defendant sentencing **5**:2306
determination of remains **1**:163–6, 180, 191–4; **4**:1890; **5**:2238
diversity (Northwestern Juvenile Project) **4**:1879
genetic variations **4**:2014
jury prejudice **3**:1604
Race–Fisher hypothesis **1**:346
'racial profiling' **4**:2032
racism **5**:2306
radar technology **1**:199, 203
radial side, hands **3**:1322, 1323
radiation **3**:1117, 1119
radio communication networks **3**:1360–1
radio waves **2**:921
radioactive materials **2**:506; **4**:1884–5
see also chemical, biological, radiological and nuclear agents; nuclear forensics
radiocarbon dating **1**:418–21; **5**:2231–3
radiochemical tests **4**:2091
radiographer role **5**:2234
radiography **2**:769; **5**:2234
see also radiology; X-rays
radioimmunoassay (RIA) **5**:2427, 2514
radioisotopes **1**:262; **2**:502
radiolarian ooze **4**:2003
radiologic technologist role **5**:2234
radiological dispersion devices (RDDs) **4**:1884, 1886
radiological emission devices (REDs) **4**:1884, 1886
radiologist role **5**:2234
radiology **2**:769; **5**:2233–9, 2570
radionuclide imaging **5**:2234
rage attacks **5**:2301
RAID (redundant array of independent disks) systems **2**:585
rainbow printing **3**:1261, 1267
RAM storage, air bag systems **1**:57
Raman spectroscopy **2**:991, 1037, 1046–60; **4**:1938
random man not excluded (RMNE) **4**:1838–9; **5**:2373–4
random match probabilities (RMP) **5**:2402
allelic frequencies **3**:1458
DNA databases **2**:832, 833, 834–5, 836
sample matches **3**:1561–2
random mating rule **3**:1458
random occurrence ratio **3**:1453
random paint studies **4**:1950
random parameters, technical surfaces **5**:2486–7
random samples **5**:2281, 2282–3, 2293
random scission **3**:1110
randomness **3**:1581
see also random...
range determination
multiple projectile cartridges **3**:1203–4
muzzle to target (entrance wounds) **3**:1390–3
single projectile cartridges **3**:1203
terminal ballistics **3**:1203–4
rape
serial murder **5**:2313
victim reporting **2**:644
rape trauma syndrome (RTS) **5**:2241–2
see also posttraumatic stress disorder
raphe (diatoms) **2**:749
rapid anoxial death *see* asphyxia
rare alleles, STRs **5**:2359–61
rarity statistic, DNA **2**:835
ratio of infrared peaks, inks **3**:1545
rational choice perspective **2**:644–7
RCMP *see* Royal Canadian Mounted Police
rCRS *see revised Cambridge Reference Sequence*
RDDs *see* radiological dispersion devices
RDX *see* hexogen
readiness potential (RP), consciousness **2**:578
Reagan, Ronald **5**:2444, 2445
real-time biodetection systems **1**:306
real-time PCR (RT-PCR) **1**:305; **2**:806
ream wrappers (paper) **4**:1969
rear-impact inline collisions **5**:2252–4
rearrest rates **3**:1613
see also recidivism
reasoning
Bayesian networks **1**:278
by exemplars **5**:2406
deficits **4**:1868–9
recall, definition **3**:1081
receptors
benzodiazepines **1**:294–5
cannabis **2**:434
postsynaptic **4**:2203
see also GABA receptors; photoreceptors; ryanodyine receptor
recertification, criminalistics **5**:2549–50
recidivism
rates for sex offenders **5**:2333, 2335
risk assessment **5**:2276
see also rearrest rates
recognition
definition **3**:1081
tasks (aging effects) **2**:917
tests (facial reconstruction) **3**:1090
recollective accuracy, traumatic memory **5**:2243–8
recombination of DNA **2**:796
recommendations to the court *see* courts
RECON (relationship and context) typology **5**:2398
reconstructions
3D **5**:2257–67
automobile accidents **5**:2250–7

reconstructions *(cont)*
bloodstain patterns **1**:390–5
facial **3**:1083, 1086–91; **5**:2238
fire scenes **3**:1126
height estimations **3**:1630
shooting incidents **3**:1217; **5**:2259, 2261–4
textiles **2**:995
see also accidents, reconstructions
reconstructive memory **4**:1709–12
record keeping
fire debris analysis **3**:1169
quality management **4**:2227–8
see also reports/reporting
recordings **1**:298, 299; **2**:840; **5**:2389–91
recovered memory **2**:787; **4**:1712; **5**:2246–7, 2248
childhood sexual abuse **4**:1712, 1713–5
delayed discovery doctrine **4**:1713
dissociative amnesia **2**:786–8
experience diversity **5**:2246–7
false accusations **4**:1713
false memories **4**:1714, 1716
hypnotic testimony **3**:1500–4
media contribution **4**:1712–3
movement for **2**:734–5
research **4**:1714–5
retractors **4**:1713–4
statute of limitations **4**:1713
true memories **4**:1715
see also hypnosis; posttraumatic stress disorder; repressed memory
recovery methods
bloodstain pattern evidence **1**:386–8
bomb mechanisms **2**:1034
casting evidence **2**:966–7
fibers **2**:985–6, 993, 994
foot impressions **3**:1254
hair **3**:1419, 1423
human remains **1**:199, 200, 204–6; **2**:768–9; **4**:1889
paint evidence **4**:1934–5
trace evidence sampling **5**:2294–5
see also collection methods
recreation animation **2**:582
see also animations
recreational drugs **2**:869
see also illicit drugs
re-cross-examination of experts **2**:665
rectal injuries **2**:534
rectal temperature **2**:708–9, 712; **5**:2469–70
recurrence models, toolmarks **5**:2487
recycled fiber **4**:1980
recycling process, pollen **4**:1957–8, 1959
red hair color (RHC) phenotypes **4**:2031
redirect examination, experts **2**:665
REDs *see* radiological emission devices
reduced-scale modeling **3**:1177
redundant array of independent disks (RAID) systems **2**:585
refereeing *see* peer reviews
reference bias **4**:1830
reference collections, paints **4**:1946, 1949–50
reference compounds **4**:2223
reference databases **3**:1240–3, 1249–50
see also databases
reference DNA, disaster victims **2**:767
reference numbers, crime scene documents **2**:611
reference population
ibd alleles **4**:1810
match statistics **5**:2408
reference samples (*K*) **1**:320, **2**:753, 754; **4**:1823, 1824, 1825–7
see also control samples
reference sources, materials failure **4**:1688
reference standards **2**:1040; **3**:1542–4
referral questions **2**:447
refining propositions, handwriting **3**:1454–6
reflectance (surface vs. subsurface) **3**:1544
reflectance microspectrometry **4**:1755–6
reflected light stereomicroscopy **2**:988–9
reflection, enhancement methods **4**:2042–6, 2050–1
reflex cardiac arrest **1**:232
refractive index (RI) **4**:1764–5
brightfield microscopy **4**:1770–1
dispersion staining **4**:1771–2
glass evidence **3**:1349, 1350, 1353–4, 1356
high power microscopy **4**:1760
polarized light microscopy **4**:1775–7
refusal of treatment **5**:2336, 2452, 2584–7
Regents of University of California v. Tarasoff, 551 P.P2d 553 (Cal. 1974) **2**:886–9
regression analyses **4**:2026–8
regulation
cannabis use **2**:435–6
forensic science **2**:494–5
illicit drug profiling **2**:854–5
rehabilitating the witness **2**:665
Reid Technique (interrogation) **2**:590; **3**:1586, 1590
reincarceration rates, juveniles **3**:1613
relapse prevention, addicts **1**:20–1
relatedness
alleged father to true father **4**:1814–5
evolutionary **4**:1812, 1815, 1816, 1817
measures of **4**:1810–1, 1813, 1817
offender databases **2**:832–3
STR profiles **5**:2370
relational aggression **1**:37, 39, 40, 45
relative age of ink entries **3**:1542, 1544
relative fluorescent units (rfu) **4**:2007, 2008
relative frequencies **3**:1580–1
relative judgments **2**:1073
relative retention time (RRT) database **5**:2525
relative standard deviations (RSDs) **4**:2224
relatives *see* relatedness
relevance (mtDNA databases) **4**:1828–9
'relevant' evidence definition **2**:694
relevant-irrelevant deception test **2**:721–2
reliability
biometric devices **1**:322, 323, 328–30
errors **2**:956
expert opinion evidence **2**:693, 1001–2, 1002; **3**:1600
relief printing **3**:1264–5
religion evaluation **5**:2608
relocation disputes, custody evaluation **5**:2608
REM behavior disorder **5**:2302
remote data storage **1**:147
renal diseases **4**:2094
renal failure **4**:2082, 2087

see also kidney
Rennie v. Klein, 462 F. Supp. 1131 (1978) **5**:2585
rental equipment, human error **3**:1490
repeat offenses, juveniles **3**:1612–3
repeated drug use **4**:2118–9
repetitive stress **5**:2567–8
replicate probabilities **3**:1572–3
reports/reporting
ASCLD/LAB program **1**:5, 7
autopsies **1**:261
bomb scene management **1**:418
CAI **2**:494
competency to stand trial **2**:461
crime scene management **2**:607
crimes **2**:643–7
expert reports **2**:775, 1007–8, 1009, 1010, 1011; **5**:2268–70
fire debris analysis **3**:1168–9
fire scenes **3**:1135, 1136
firearms investigation **3**:1218
first report to police **2**:619–25
identification evidence **3**:1509–10
paint evidence **4**:1939–40
quality management **4**:2227–8
toolmark examination **5**:2485–6
see also record keeping
representation of scale (photography) **4**:2038–9
representative payee appointment **3**:1376
repressed memory **4**:1712–6; **5**:2243
see also hypnosis; posttraumatic stress disorder; recovered memory; suggestibility
reproducibility of evidence **2**:693
requested exemplars, handwriting **3**:1439–40
requested standards, inks **3**:1542
research
consent capacity **2**:440
criminal responsibility **2**:577–9
funding graduate programs **2**:901
requirements (bloodstain patterns) **1**:388–90
research and development (R&D) **5**:2392
research diagnostic criteria (RDC) **4**:2187
Research Network on Adolescent Development **3**:1609, 1611
resemblance ratings, facial reconstruction **3**:1090
residential care facilities *see* long-term care facilities
residential substance abuse treatment **5**:2416
residues
clothing **3**:1223
explosives **1**:415–7; **2**:1026; **4**:1746
see also explosion debris; fire, debris; gunshot residues; ignitable liquid residues
resiliency **4**:1820
resin **4**:1932
resisitivity **1**:203
Resnick, Philip **2**:743; **3**:1434
resolving power/resolution, microscopy **4**:1758, 1763, 1767, 1796
resource availability, searches **2**:624
respiratory causes, sudden death **4**:1843–6
respiratory depression **4**:1896
respiratory tract injuries **3**:1533, 1535–6
response mechanisms, CBRN scenes **2**:503–5
response/stimulus theory **5**:2436
response time, automobiles **5**:2255–6
Ressam, Ahmad **2**:1044
restoration of serial numbers **5**:2324–7
restorative justice **4**:1719
restraint asphyxia **1**:228
restraint in prone position **1**:228
restriction fragment length polymorphism (RFLP) **1**:262; **4**:1749; **5**:2595
resumes *see* curriculum vitae
resuscitation period **2**:698
retaliation, females **1**:40
retardation *see* mental retardation
retardation colors **4**:1780, 1783
retardation of waves **4**:1778, 1779
retention interval **2**:1068
retention probability, soil **5**:2380
retention stage, memory **2**:1076; **4**:1709, 1710
retention time (RT), gas chromatography **1**:85, 86
retinal hemorrhages (RHs) **5**:2339–40
Retinome™ **4**:2030, 2031
retraction of abuse allegations **2**:539
retractors, recovered memory **4**:1713–4
retrieval deficit, brain injury **3**:1462
retrieval methods, concealed objects **5**:2239
retrieval stage, memory **2**:1076; **4**:1709, 1710
retrieval tools, child interviews **2**:554
retrocausal probability **5**:2492–3
retrograde extrapolation **1**:58
reunification therapy **4**:1983
reuptake, neurotransmitters **1**:290
revenge (arson) **3**:1228
revenge killings **3**:1477
reverse ABO blood group **1**:342
reverse projection photogrammetry **3**:1626
revised Cambridge Reference Sequence (rCRS) **4**:1830, 1834
revised cognitive interview **2**:554
revocable living trusts **3**:1376
reward-related learning **2**:572
reward-satisfaction model **2**:572
RFLP *see* restriction fragment length polymorphism
rfu *see* relative fluorescent units
RGB color model **4**:2043–5
rhabdomyolysis **1**:138
RHC (red hair color) phenotypes **4**:2031
Rhesus blood group system **1**:345–9
Rhesus boxes **1**:347
rheumatic mitral stenosis **2**:478–9
RHs (retinal hemorrhages) **5**:2339–40
Rhythm Test, Halstead Impairment Index **4**:1865
RI *see* refractive index
RIA (radioimmunoassay) **5**:2427, 2514
rib fractures **1**:407–8
rib marks, materials **4**:1687
ribbon technology **5**:2665, 2671, 2672
ribs **1**:181, 407–8
rich false memories **4**:1710–1, 1712
ricin **4**:2064–5
rickettsiae **1**:301
ricochet (bullets) **3**:1221, 1384
ridgeology **3**:1279
see also friction ridge skin; papillary ridges
Ridgway, Gary **4**:1940–1
Rielly, Mary **5**:2445–6

rifled-barrel firearms **3**:1380
rifling, projectile movement **3**:1200, 1204
Righarts, S. **2**:659
right ventricular dysplasia **2**:477
rights
  mental health law orientation **5**:2450
  'right to die' **5**:2585
  to treatment **5**:2580–3
  see also parental rights
rigor mortis **2**:705–8; **5**:2468, 2469–70
  see also postmortem changes
rima oris **3**:1088
  see also mouth
ring binder photography method **1**:373
ring fractures, skull **1**:403
'ring of dirt' **3**:1387
risk assessment **5**:2271, 2272–7
  behavioral science evidence **1**:287
  clinical violence risk **5**:2599
  criminal profiling **5**:2455
  dangerousness **2**:667–77
  disaster mental health **2**:761
  documentation **5**:2600
  history **5**:2272
  models **5**:2274–6
  postmortem suicide risk assessment **4**:2165–6
  principles of **5**:2598
  schools **5**:2454
  search strategies **2**:623, 624
  violence **1**:29–30; **4**:2195
  workplace **5**:2461–2
  see also duty to warn; risk factors; threat assessment
risk behaviors, juveniles **4**:1878, 1880, 1881
risk factors **5**:2273–4
  adolescent impulse control **3**:1610
  categories **5**:2598
  childhood firesetting **3**:1226
  clinical violence risk **5**:2599
  crime scenes **2**:615
  driving under the influence of drugs **2**:879, 882
  elder abuse **2**:906–7, 912–5
  mitigation testimony **4**:1819–20
  suicide **4**:2165–6
  see also duty to warn; risk assessment
risk management plan **5**:2599–600, 2600
  see also risk assessment
risk perceptions, school threats **5**:2455–6
RMNE *see* random man not excluded
RMP *see* random match probabilities
RNA extraction **2**:1064
road traffic accidents
  alcohol intoxication **1**:100
  biomedical engineering **5**:2565–75
  cannabis use **2**:434, 436
  drug impaired driving **2**:877, 879, 882
  fatalities **5**:2541–4
  light globe/filament examination **3**:1632–8
  prevention systems **1**:57
  reconstructions **5**:2250–7
  toxicology **5**:2505–6
  trauma causation analysis **5**:2565–75
  see also crash data; vehicles
roadside drug testing **2**:881; **4**:1918–9
'roaming', cellular phones **3**:1360
robotic application, paint **4**:1932
rock cocaine *see* crack cocaine
rock fragment particles **4**:2003
rockets, paint types **4**:1933
rod photoreceptors **4**:2041
Rogers v. Orkin, 478 F. Supp. 1342 (D. Mass. 1979) **5**:2585
Rokitansky, Carl von **1**:257–8
rollerball pens **5**:2662
rollers (printing devices) **5**:2663, 2669
Roman civil law system **2**:561
Romero v. Youngberg, 457 U.S. 307 (1982) **5**:2583
room fires *see* compartment fires
root shapes, hair **3**:1404, 1407, 1409–10
Roper v. Simmons, 543 U.S. 551 (2005) **2**:717–8
Rorschach, Hermann **4**:2178
Rorschach inkblots **4**:2178, 2181, 2184
Rouse v. Cameron, 373 F.P2d 451 (D.C. 1966) **5**:2583
route attribution, nuclear forensics **4**:1887
routes of administration *see* administration routes (drugs)
routine procedures, alcohol analysis **1**:87–8
Royal Canadian Mounted Police (RCMP) **3**:1246, 1287
royal dispensations **1**:282
RP (readiness potential), consciousness **2**:578
RRT (relative retention time) database **5**:2525
RSDs (relative standard deviations) **4**:2224
RT *see* retention time
RT-PCR *see* real-time PCR
RTS *see* rape trauma syndrome
RTX (ruthenium tetroxide) **3**:1304–5
Ruback, R.B. **2**:644–7
rubber stamp fingerprint spoofing **1**:330, 332–3
rugae, palatal **1**:168
Ruhemann's purple **3**:1296, 1297
rule of thirds (photography) **2**:627
Rules of Civil Procedure (US) **2**:776–7
  26 **2**:776–7, 777–8
  Electronic Discovery Amendments 2006 **2**:777–8
rules of evidence **1**:23; **2**:963
  see also expert opinion evidence; Federal Rules of Evidence
run-time malware analysis **1**:148
RUNG program **5**:2590, 2591
rupture pattern, light globes **3**:1634, 1635
ruptures, aortic **1**:409
  see also lacerations
rural areas, mental illness **2**:652
Russian facial reconstruction method **3**:1087
ruthenium tetroxide (RTX) **3**:1304–5
ryanodyine receptor (Ry) **2**:471

S–N curve, materials **4**:1684
S/P ratios (saliva-to-plasma ratios) **4**:1906
SAC (serum alcohol concentration) **1**:83
sacrum **5**:2329
sadism **5**:2312, 2313, 2317, 2318, 2320
safes/safe packing materials **1**:428
safety
  CBRN scenes **2**:504

child traffic safety **5**:2543
crime scenes **2**:615
drug safety research **5**:2508–9
explosion scene assessments **2**:1021–3
fire **3**:1226
human factors interaction **3**:1484–5, 1486, 1489
public safety mandated treatment **5**:2579
urine drug tests **2**:860
safety belts *see* seat belts
safety engineering **3**:1484–5
sagged furniture spring fire myth **1**:209, 216–7
Sain v. Townsend, 372 U.S. 293 (1963) **2**:736
St. Jean v. United States (U.S. Ct. of Appeals for Armed Forces, 1996) **4**:2163
salbutamol **3**:1430
Salgo case **2**:451
saliva **4**:1904–5
DNA source **2**:824
stain identification **1**:317–8
testing for cocaine **2**:567
transfer **4**:1906; **5**:2537
see also oral fluid
saliva-to-plasma (S/P) ratios **4**:1906
salivary glands **4**:1904–5
salt form, cocaine **2**:562–3
SAMHSA *see* Substance Abuse and Mental Health Service Administration
sample chamber, SEM **4**:1797
sample collection *see* sampling
sample containers **3**:1129
sample evidence *see* trace evidence
sample extraction *see* sampling
sample matches **3**:1561–2, 1565
see also trace evidence
sample size
choice for discrete data **5**:2284–7
determination criteria **5**:2283–4
Northwestern Juvenile Project **4**:1878, 1879
sampling **5**:2281–90
biological swabs **1**:320–2
blood **1**:70, 88–9
brightfield microscopy **4**:1769
CBRN scenes **2**:504
contamination of samples **4**:1834
convenience samples **3**:1581
CWA samples **2**:523–5
definition **5**:2281, 2291
diatoms **2**:753–4, 754–5
differential extraction of semen **2**:757
Dirichlet formula **4**:1812, 1813, 1815, 1816
DNA **2**:816–9, 1060, 1063–4
ecosystems **2**:947–8
error (mtDNA) **4**:1825
explosion debris **1**:415; **2**:1031, 1044
gunshot residues **3**:1192–3, 1194
hair **3**:1419–20, 1421
handwriting **3**:1438
ignitable liquid residues **3**:1138–41, 1169
interpretation of samples **4**:1835–6
mtDNA **4**:1823, 1824, 1825–7, 1834, 1835–6
Northwestern Juvenile Project **4**:1878, 1879
packaging samples **4**:1929, 1930–1
paint **4**:1935
plans **5**:2291, 2292–3
pollen **4**:1960–6
postmortem artifacts **4**:2102–4, 2107
preparation of samples **2**:523–5; **4**:1801
questions **5**:2293–4
radiocarbon samples **1**:419–20; **5**:2231–2
sexual assault cases **1**:236–7
single-source DNA identification **5**:2367
soil analysis **5**:2380, 2383–6
strategy **5**:2291
tissues **3**:1469–70
toolmarks **5**:2485
toxicological **5**:2510–1, 2516–9
trace evidence **5**:2291–5
transporting samples **4**:1930
types **5**:2281
urine for drug tests **2**:860–1
wood **1**:427; **5**:2645
see also evidence; exhibits; specimens
sampling fraction **5**:2283
sampling frame **5**:2282
sanctions (ethics) **2**:957, 958, 960
sandbox networks **1**:148
sane automatism **1**:255
'sane' suicide **4**:2163, 2164
sanity
opinion (criminal responsibility) **3**:1555–6
presumption of **5**:2431
sapwood **5**:2642
sarcoidosis **2**:475
sarin hydrolysis **2**:527
SBS *see* shaken baby syndrome
SCA (sickle cell anemia) **1**:169
scalability, AFIS system **1**:252
scalds **3**:1537–8
scale
distance (explosion scenes) **2**:1031
ecosystem incidents **2**:947–8
evidence photography **2**:634–5; **4**:2038–9
of reference **2**:634–5
scale patterns (hair) **3**:1404, 1406–7
scanning devices **1**:330, 334; **4**:1797
scanning electron microscopy (SEM) **3**:1595; **4**:1793–804; **5**:2395
explosion debris **2**:1034, 1036
fiber analysis **2**:989–90
firearms identification **3**:1207
forgeries/counterfeits **3**:1257
materials characterization **4**:1686–7
paint analysis **4**:1937
soil analysis **5**:2385
toolmark examination **5**:2489
scanning electron microscopy/energy dispersive X-ray (SEM/EDX) **2**:989–90; **3**:1192, 1195
scanning electron microscopy/wavelength dispersive X-ray spectroscopy (SEM/WDX) **3**:1195
scanning transmission microscope (STEM) **4**:1793
SCD *see* sickle-cell disease; sudden cardiac death
'scenarios' (interrogations) **3**:1587
scene
definition **2**:615
processing **2**:616
see also crime scene. . .; incident locations
scene assessment **2**:615
scene attendance **2**:613, 615

scene databases **3**:1250–1
scene entry log **2**:610–1
scene examination **3**:1636
  see also crime scene...
scene hygiene **1**:415
scene investigation **2**:614–9
  drowning cases **2**:755
  explosions **2**:1019–27, 1028–31, 1045
  see also crime scene...; shooting scene investigation
scene management **1**:412–8; **2**:611–3, 619–25
scene records **2**:613
scene-to-scene linking **3**:1251
scene-to-suspect linking **3**:1251
scenes of crime officer (SOCO) **2**:615
  see also crime scene investigator
Schallamach pattern, shoemarks **3**:1234
schizophrenia **2**:461, 741; **3**:1433; **4**:2189, 2204; **5**:2415
Schneidman, Edwin **4**:2161, 2162
school age children evaluation **5**:2604–5
schools
  child sexual abuse **2**:530–1
  parental rights **4**:1984
  threat assessment **5**:2454–8
Schottky emitters **4**:1796
Schulze's maceration method **5**:2643
Schwartz, Adina **3**:1094
science
  human factors **3**:1484
  legal definitions **3**:1619–20
  legal role **1**:285–6
  see also scientific...
scientific evidence **1**:276–80; **2**:971
  see also evidence
scientific knowledge **3**:1621
scientific literature **5**:2546
  see also professional literature
scientific method **2**:1072; **3**:1175; **4**:1680–8, 2009–11, **5**:2296–7
  see also methodologies
scientific principles, judicial notice **2**:665; **3**:1601–2
Scientific Support Department (SSD) **2**:621
scientific support manager (SSM) **2**:621
scientific validity **2**:693, 695, 1002
Scientific Working Group on DNA Analysis Methods (SWGDAM) **4**:1835
Scientific Working Group for Friction Ridge Analysis, Study and Technology (SWGFAST) **3**:1285, 1290
scientific working groups **3**:1285, 1290; **4**:1835; **5**:2622–3
scissor stab wounds **5**:2648
scopolamine **2**:728, 729, 730
Scotland **2**:557
  see also United Kingdom
scraping sample recovery method **5**:2294–5
screening
  explosion scene evidence **2**:1026
  immunoassay screening **4**:2128
  soil samples **5**:2384–5
  toxicology **4**:2127; **5**:2504
  see also initial testing
Screening Version of the Hare Psychopathy Checklist-Revised (PCL : SV) **4**:2197–201
screwdriver stab wounds **5**:2648
scrutiny of experts **2**:1015
SDH (subdural hemorrhage) **5**:2341–5
SDS *see* sodium dodecyl sulfate; standard deviations
SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis) **1**:349
SEA (strategic environmental assessment) **2**:946
sea water **2**:754
search strategies
  bomb scene evidence **1**:415–7
  crime scene management **2**:622, 623–5
  database searches **2**:834–7; **5**:2372–3
  explosion scene evidence **2**:1026
  forensic archaeology **1**:199, 202–4; **3**:1495–9
  see also familial searching
searching microscope **4**:1934
seat belts **1**:52, 53, 57; **5**:2543, 2566
seat of explosion investigation **2**:1029–30
sebaceous glands **5**:2421, 2422
second-degree burns **3**:1530
second generation multiplex (SGM) systems **2**:801–2
secondary alcohols **1**:81
secondary delusions **2**:741
secondary electrons (SEs) **4**:1687, 1798, 1799, 1802
secondary fluorescence **4**:1785, 2047
secondary fractures, skull **1**:404
secondary metal salt treatment **3**:1297
secondary scenes **1**:415, **2**:621
secondary sexual features **1**:155
secondary transfers
  animal hair **3**:1404
  definition **5**:2379
  DNA evidence **5**:2555–6, 2556
  fibers **3**:1098
secrecy, child sexual abuse **2**:538
Secret Service threat assessment **5**:2454, 2455
secretors, ABO blood group **1**:339, 341, 342
sectioning *see* civil commitment
securing computer data **2**:585–6
security features, documents **3**:1265–73
sedative hypnotics **1**:116
seeds **1**:421, 425
  see also poppy seeds
SEIR *see* surface-enhanced irregular reflection
seized drugs *see* illicit drugs
seizures **5**:2298–305
selected ion monitoring (SIM) **2**:596, 862; **5**:2525
selection of experts **2**:1013–8
selective serotonin reuptake inhibitors (SSRIs) **1**:111; **4**:2205, 2206; **5**:2335
selegiline **1**:137, 139
self-authenticating documents **2**:840, 841
self-blame **2**:646
self-control **2**:573
self-defense **1**:40, 263, 268, 270, 271, 272, 274–5
self-heating ignition **3**:1114, 1115
self-incrimination, truth serum **2**:736
self-inflicted wounds **5**:2653–4, 2654
  see also suicide
self-report inventory **4**:2178, 2183

self-representation defense **2**:462
Sell v. United States, 539 U.S. (2003) **2**:457
SEM *see* scanning electron microscopy
SEM/EDX *see* scanning electron microscopy/energy dispersive X-ray
SEM/WDX (scanning electron microscopy/wavelength dispersive X-ray spectroscopy) **3**:1195
semen **1**:17, 315–7; **2**:757, 822–3; **5**:2537
semicircular-pattern fire plumes **3**:1120
semiconductors **4**:2049
seminal fluid, definition **1**:315
  see also semen
semiporous surfaces **3**:1292, 1305
semisynthetic drugs **4**:1896
senior citizens *see* older adults
senior identification manager (SIM) **2**:765, 766, 767
senior investigating officer (SIO) **2**:613, 620, 621, 765
sensitivity
  analysis **1**:279; **3**:1353
  interviews **4**:2165
  low copy number DNA tests **3**:1642, 1643
  photographic capture devices **4**:2048–9
sensitizers, ammunition **3**:1190
sensorimotor ability assessment **4**:1865
sensors
  air bag systems **1**:53–4, 57
  digital photography **4**:2049
  fingerprint recognition **1**:325–7
sentencing
  capacity of defendant **2**:442
  demographic factors **5**:2306–11
  see also death penalty
separation methods
  ignitable liquid residues **3**:1138
  soil **5**:2381
sequence heteroplasmy **4**:1835
sequence of DNA molecule **2**:798–9
sequence polymorphism **2**:842
sequential data, fire scenes **3**:1126
Serge v. Commonwealth 586 Pa. 674 (Pa. 2006) **2**:581
serial connections, Bayesian networks **1**:277
serial murder **3**:1474, 1476; **5**:2311–21
serial numbers
  obliteration **5**:2325
  restoration **5**:2324–7
seriousness of crime determination **2**:645, 646
serotonin reuptake *see* selective serotonin reuptake inhibitors
serrated knife stab wounds **5**:2647
serum alcohol concentration (SAC) **1**:83
serum method, haptoglobin **1**:355
serum tryptase level, anaphylaxis **4**:1855
servers, network analysis **1**:143, 144–5, 145
service use
  elder abuse cases **2**:908–9
  juvenile projects **4**:1878
SEs *see* secondary electrons
setigerum poppy subspecies **1**:426
severe mental illness (SMI) **2**:649–53
  see also mental illness
severe personality disorder **2**:669–70
sex determination
  human remains **1**:155–6; **4**:1890; **5**:2237, 2328–31
  wildlife forensics **5**:2637
sex offenders
  civil commitment **2**:559–60, 668–9; **5**:2578
  civil liberties **5**:2578
  dangerousness **2**:671; **5**:2578–9
  definition **5**:2332–3
  treatment of **5**:2332–7
  see also sexual abuse; sexual assaults; sexual homicide
sexism **4**:2151
sexual abuse
  cross-examination methods **2**:657, 660
  elders **2**:903, 906
  female aggression **1**:39
  implanted memories **2**:733, 734–5
  recovered memory **4**:1712, 1713–5
  see also childhood sexual abuse; sex offenders; sexual assaults
sexual arousal control **5**:2335
sexual assaults **1**:234–42; **2**:644, 660, 757, 822, 868–76
  see also sexual abuse
sexual behavior, child abuse **2**:531
sexual dimorphism **1**:155–6
sexual fatalities *see* autoerotic deaths; sexual homicide
sexual hair **3**:1417
sexual homicide **3**:1476; **5**:2312, 2313
  adolescent perpetrators **5**:2314
  child murder **5**:2315
  childhood sexual abuse **5**:2318
  motivations **5**:2315–6
  offender interviewing strategies **5**:2319–20
  offense behavior **5**:2316
  prevention **5**:2320
  psychiatric findings **5**:2317–8
sexual maturation **1**:190
sexual offenders *see* sex offenders
sexual sadism **5**:2313, 2317
sexual violence **2**:668–9; **5**:2415
  see also sexual abuse; sexual assaults; sexual homicide
sexually motivated assault **1**:234–42
  see also sexual assaults
sexually transmitted diseases (STDs) **2**:534
  see also human immunodeficiency virus
sexually transmitted infections (STIs) **1**:240; **4**:1878, 1880
sexually violent predator (SVP) legislation **2**:668–9
SGM (second generation multiplex) systems **2**:801–2
shaken baby syndrome (SBS) **1**:263; **5**:2339–50
shape characteristics
  bloodstains **1**:365–6
  shoemarks **3**:1232
shared delusions **2**:742
shared psychotic disorder **5**:2446
sharp force injuries **5**:2561–2, 2646–60
  see also chop wounds; cut wounds; incised wounds; stab wounds

shedding fibers **3**:1097, 1098
see also transfer of evidence
sheet thickness, paper **4**:1973
shell casing packaging **4**:1929
shellfish contamination **2**:948
shelters, elders **2**:908
Shepherd, Matthew **5**:2447
Shift scores, suggestibility measurement **2**:1068; **3**:1592
Shipman, Dr Harold **5**:2505
shoemark examination guidelines **3**:1230–4
see also footwear impressions
shoes *see* footwear
shooter position at time of firing **3**:1220
shooting-distance estimation **5**:2351–3
shooting scene investigation **3**:1219–25
3D reconstructions **5**:2259, 2261–4
ammunition **3**:1219–20, 1221, 1223
bloodstain patterns **1**:374, 376–7
cartridge cases **3**:1219–20
clothing residues **3**:1223
DNA evidence **3**:1222
fingerprint evidence **3**:1222
gunshot residue analysis **3**:1224
holes in glass **3**:1220–1
interpretation **3**:1222
lighting **3**:1222
mass graves **4**:1678
medical examinations **3**:1224
postmortems **3**:1223
powder tattooing **3**:1223
reconstructions **3**:1217, 1222; **5**:2259, 2261–4
schools **5**:2454–8
suicides **3**:1221–2
wound position/trajectory **3**:1223
short-sighted decision-making **3**:1610
short tandem repeats (STRs) **2**:796, 801–3, 821; **4**:1750, 2021; **5**:2354–65
allele designation **5**:2356
analysis **5**:2356–9
bone/teeth DNA **2**:826
cannabis DNA analysis **1**:429
COfiler™ systems **2**:569–70
core loci **5**:2357–9
databases **5**:2357–9
forensic applications **5**:2354–5
genetic diseases **5**:2359
Identifiler™ system **3**:1518
limit of detection **1**:126
loci designation **5**:2355–6
microvariants **5**:2356
mini-STRs **2**:817; **4**:1804–9
mutations **2**:796; **5**:2354, 2359–61
nomenclature **5**:2355–6
nonhuman forensic analysis **5**:2363
nonstandard **5**:2362
plant DNA evidence **2**:827
PowerPlex systems **4**:2149
primer binding site mutations **5**:2354, 2359–61
profiling DNA **2**:806–7; **3**:1639; **5**:2365–76, 2628, 2632
rare alleles **5**:2359–61
semen DNA **2**:822
triallelic patterns **5**:2361
visualization **5**:2356–7
X-chromosomes **5**:2362
Y-chromosomes **2**:817; **5**:2362, 2677–80
see also microsatellites
short-term assessment of risk and treatability (START) **5**:2276
shortpass (SP) filters **4**:1786
shortwave UV fingerprint enhancement **3**:1318
shotguns
ammunition **3**:1382
injuries **3**:1393–4
patterns **3**:1221
see also firearms
shrinkage of tissue, burns **3**:1532
shrubby plant fiber sources **4**:1978–80
sickle cell anemia (SCA) **1**:169
sickle-cell disease (SCD) **4**:1849–50
Sickles, Daniel **5**:2445
side-by-side visual comparison method **3**:1244
side chain scission, polymers **3**:1111
side lighting **3**:1595
see also oblique lighting
SIDS *see* sudden infant death syndrome
sifting explosion evidence **2**:1026
sign of elongation, fibers **4**:1782–3
signatures **3**:1451–7
significance assessment of evidence **4**:1939–40
significance levels **3**:1582
significance probabilities **3**:1581–2, 1583
significance tests **3**:1544
silibinin **4**:2060
silicon dioxide (silica) (SiO_2) **3**:1348; **4**:2219
silicone molds **1**:330, 332
Silver, E. **2**:650–1, 651
Silverlock v. R [1894] 2 QB 766 [3] **2**:1003
SIM *see* selected ion monitoring; senior identification manager
SIM cards **3**:1363, 1365–6
similarity, degree of *see* comparison
Simmons v. Roper, 543 U.S. 551 (2005) **2**:717–8
Simon, Theodore **4**:2177
simple correlations, fire models **3**:1181
Simple Mail Transfer Protocol (SMTP) **1**:145
simple microsatellites **2**:796
simple random samples **5**:2281, 2282–3
simple tandem repeats (STRs) **4**:2075; **5**:2356
simplicity (photographic) **2**:627
Simpson, O.J. **1**:391–5
SIMS (Structured Inventory of Malingered Symptomatology) **4**:1659
simulations **2**:579–82
handwriting **3**:1444
shooting incidents **5**:2261–4
signatures **3**:1444–7
vehicle speed **5**:2250–1
see also reconstructions
single bullet lesions **5**:2562
single cell techniques **2**:823–4
single court transactions **3**:1375, 1376
single dose of drugs **4**:2116–7
single-element typewriters **5**:2664–5
single-locus match probabilities **5**:2368–9
single locus probing/profiling (SLP) **2**:800–1; **5**:2595

single-metal deposition (SMD) **3**:1304
single nucleotide polymorphisms (SNPs) **2**:795, 796, 802–3
  ABO blood group system **1**:343
  ancestry informative markers **4**:2025
  iris color **4**:2030, 2031
  protein polymorphisms **1**:340, 352, 354, 355
single polar particle characterization **4**:1773–8
single projectile cartridges **3**:1203
single-source identification **5**:2366–73
single-stain mixtures **4**:1838
single thickness of paper **4**:1973
single view metrology **3**:1626–7
Singleton haplotypes **4**:1827
sinking speeds, pollen **4**:1956
SIO *see* senior investigating officer
SiO_2 *see* silicon dioxide (silica)
sipes, tires **5**:2481
Sipin v. State 130 Wash.App. 403 (Wash.App.Div. 1, 2005) **2**:581
SIRS (Structured Interview of Reported Symptoms) **4**:1659
site-dependent differences, PMR **4**:2097
site heteroplasmy, mtDNA **4**:1835
site investigation *see* crime scene. . .; scene investigation
situational factors
  firesetting **3**:1227
  violence risk **5**:2273
size
  bloodstain characteristics **1**:368–9
  databases **4**:1830
  footwear impressions **3**:1243
  sorting explosion debris by **2**:1034
  see also sample size
Sizemore, Chris **5**:2434–5
SK sequences, mtDNA **4**:1823, 1826–7
skeletal age **1**:159, 161, 188
skeletal fractures (alcohol-induced) **1**:104
skeletal lesions **1**:169; **5**:2557–60
skeletal muscle excitability **2**:698–701; **5**:2468
skeletal remains
  age determination **1**:159, 161, 179, 188; **5**:2557–60
  anthropological methods **3**:1495
  archaeology **1**:199
  identification **1**:152–79; **3**:1514
  mass graves **4**:1676
  mtDNA analysis **4**:1833
  PMI determination **4**:2090
  sex determination **5**:2328
  species determination **5**:2393–6
  trauma analysis **5**:2557–63
  see also bones; human remains
sketching scenes **2**:607, 613; **3**:1124–6
skid marks **5**:2250–1, 2482–3
skin
  abrasions **1**:396–7
  aging of lesions **5**:2557, 2559–60
  burns **3**:1530, 1532, 1538
  chemical injuries **3**:1538
  color (phenotype inference) **4**:2032
  fingermark detection **3**:1305
  injuries **1**:229; **3**:1538
  specimen sampling **4**:2103
  splitting of **3**:1532
  temperature (scalds) **3**:1537
  see also friction ridge skin
skull fractures **1**:402–4
skullcap burn injuries **3**:1534, 1535
slashes *see* incised wounds
'slaughterer's guns' **3**:1396
SLC6A4 gene **3**:1338, 1341
sleep deprivation **2**:878, **3**:1592
Slick criteria, malingering **4**:1660
slide preparation (microscopy) **2**:754
SLP *see* single locus probing/profiling
small consignment sample size **5**:2286–7
small molecule poisoning **4**:2133–4
Small Particle Reagent (SPR) **3**:1295, 1319
SMD (single-metal deposition) **3**:1304
smeared blood **1**:388–9
SMI *see* severe mental illness
Smith, Sandie **4**:2150
Smith v. Jones, 1 S.C.R. 455 (1999) **2**:670
Smith v. Lessard, 349 F. Supp. 1078 (E.D. Wis. 1972) **5**:2577
Smith v. Wiggins, 539 U.S. 510 (2003) **4**:1818
SMM (stepwise mutation model) **5**:2375
smoke horizons **3**:1131
smoke residues **3**:1111
smoking drugs **1**:137; **2**:434, 435
smooth-barrel firearms **3**:1380
smooth, hard surfaces **1**:365, 388
smothering **1**:226–7
SMS spoofing **3**:1369
SMTP (Simple Mail Transfer Protocol) **1**:145
smuggling objects **5**:2239
Snell's law **4**:1764
snow, casting in **2**:967
SNPs *see* single nucleotide polymorphisms
social aggression **1**:37
social anxiety disorder **2**:547
social control theory **2**:649, 652
social drinkers **1**:122
social influence **2**:644, 645–6; **3**:1478
social learning theory **5**:2316
social science expert testimony **3**:1588
  see also sociological approach
socialization, female aggression **1**:44
societal norms **4**:2186
societies (forensic) **2**:873, 874, 922; **3**:1428, 1520
  see also forensic associations; *individual societies*
Society of Forensic Toxicologists (SOFT) **2**:873, 874
Society of Hair Testing (SoHT) **3**:1428
sociobiological model, child abuse **4**:1991–2
sociocultural factors
  elder abuse **2**:914
  psychopathy **4**:2194
sociodemographic factors, firesetting **3**:1227
socio-legal issues, DNA databases **2**:683–4
sociological approach **2**:760–1
  see also social...
sociopathy *see* psychopathy
sock marks **4**:1669, 1670–1
SOCO *see* scenes of crime officer

soda pulp **4**:1978
sodium **4**:2085
sodium amytal **2**:729
sodium bicarbonate **2**:863–4
sodium dodecyl sulfate (SDS) **2**:804
sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) **1**:349
sodium pentothal **2**:729
SOFT *see* Society of Forensic Toxicologists
soft porous surfaces **1**:365, 369
soft tissue
  bone relationship **3**:1087, 1090
  individualization markers **1**:167–8
  radiocarbon dating **1**:420
  see also tissues
software
  minutiae thresholds **3**:1288
  mobile phone handsets **3**:1364–5
softwoods **1**:426; **4**:1976, 1978, 1979, 2005–6; **5**:2642, 2643
SoHT (Society of Hair Testing) **3**:1428
soil analysis **5**:2377–87
soil layers **1**:200
  see also stratigraphy
soil mapping **5**:2383, 2386
soil morphology **5**:2380, 2383, 2384
soil profiling **5**:2384
soil science **5**:2377–8
  see also soil analysis
Sokolow v. United States, 490 U.S. 1 (1989) **4**:2157
sole databases **3**:1240–3, 1254
  see also footwear impressions
sole legal custody evaluation **5**:2609
sole physical custody evaluation **5**:2609
solicitor-client privilege **2**:670
solid inkjet printers **5**:2670
solid-phase extraction (SPE) **2**:598–9, 1062–3; **4**:2107, 2125, **5**:2427, 2518–22
solid-phase microextraction (SPME) **2**:598, 599; **3**:1139, 1142; **5**:2519–22
solid-state sensors **1**:326–7
solids, combustion **3**:1109–12
solubility tests, fibers **2**:992
'solution-focused' evaluation **4**:1981
solvents
  evaporation (ink aging) **3**:1544–5
  extraction (drug samples) **5**:2516–8
  paint composition **4**:1932
  selection for ILR isolation **3**:1141
somatic complaints, child abuse **2**:531
somatic delusions **2**:741, 742
somatic mutations **2**:794
Sommers, Samuel **3**:1604–5
somnambulism **1**:254
somniferum poppy subspecies **1**:426
soot deposits **3**:1535–6
SOPs *see* standard operating procedures
sound recordings **2**:840
  see also recordings
source address obfuscation **1**:142–3
source attribution
  DNA databases **2**:832
  errors of memory **4**:1711
  nuclear forensics **4**:1886
source identification
  ILRs **3**:1163–8
  ink evidence **3**:1546–8
  match statistics **5**:2407
source-level evidence **2**:493–4
source-level propositions **2**:973, **3**:1353–6, 1580; **4**:1951–2, 1952
source memory **2**:916
source monitoring
  errors **2**:553; **4**:1711
  truth serum effects **2**:732, 733
source population (potential) **5**:2408
source probability error **5**:2404
South America **2**:951
Soviet Union **2**:560; **3**:1087
SP (shortpass) filters **4**:1786
Spain **2**:1010–1
spalling
  concrete combustion **3**:1112
  fire myths **1**:209, 217–8
spatial information, soil **5**:2383
spatial memory **2**:916
spattering blood *see* bloodspatter
SPE *see* solid-phase extraction
speaker recognition **5**:2389–91
special dispensations **1**:282
special needs trusts **3**:1376
special weapons and tactics (SWAT) teams **4**:2071–5
specialist applications
  behavioral toxicology **1**:290–3
  drug-facilitated crime **2**:868–76
  drug impaired driving **2**:877–83
  medical imaging **5**:2234
specialized courts **5**:2450–1
specialized training
  criminalistics **5**:2546
  sex offender treatment **5**:2337
specialty board ethical rules **2**:958
species, definition **1**:423
species determination
  remains **1**:153–4; **4**:1889–90; **5**:2237, 2393–6
  wildlife forensics **5**:2637
specific heat capacity **3**:1104
specific power of attorney **3**:1375
specificity, immunoassays **2**:862; **5**:2512
specimen writings *see* exemplars, writings
specimens
  acquisition of **4**:2102–4
  biochemical examinations **4**:2076
  cannabis analysis **2**:434
  infrared spectrum **4**:1750
  postmortem toxicology **1**:89–90; **4**:2101–6; **5**:2495–502
  preservation of **2**:944
  spectral response **4**:2041–8
  storage of **4**:1927, 1929, 1930, 2101–2, 2104–6
  toxicology **1**:89–90; **4**:2101–6, 2120–3; **5**:2495–502
  see also alternative specimens; collection methods/devices; evidence; exhibits; sampling
spectral distribution, light **4**:2041
spectral information *see* mass spectrometry
spectral response, specimens **4**:2041–8

spectral sensitivity (photography) **4**:2048–9
spectrophotometry **4**:1938
spectroscopy
  fiber analysis **2**:990–1
  infrared **2**:845, 1037, 1046–60; **4**:1750–7, 1936–1942; **5**:2385
  paint analysis **4**:1936–7, 1937, 1938, 1939, 1942
  SEM signals **4**:1800
  soil analysis **5**:2385
  see also individual types
specular reflection **4**:2050
speech analysis **5**:2389–91
Speech Sounds Perception Test **4**:1865
'speed' *see* methamphetamine
speed of vehicle
  angled collisions **5**:2251–2
  braking distance **5**:2255–6
  calculation methods **5**:2250–1
  inline collisions **5**:2252–4
  perception–response time **5**:2255–6
spermatozoa **1**:316; **2**:757, 822, 1061
  see also semen
spermatozoon, definition **1**:315
  see also semen
spider's web fracture pattern **1**:402
spikes, DNA profiling **1**:127
spines, bloodstains **1**:365, 369
spiral search strategy **2**:623
spitting method, oral fluid collection **4**:1907, 1909
SPJ *see* Structured Professional Judgment
splashed bloodspatter patterns **1**:369–70
spleen **1**:410; **5**:2500
split custody evaluation **5**:2609
split fountain printing **3**:1261, 1267
split image focus **2**:641
split/splitless injectors **5**:2524
splitting of skin, burns **3**:1532
SPME *see* solid-phase microextraction
spoliation, fire scenes **3**:1127–8
spontaneous false memories **2**:1069, 1070
spontaneous ignition **3**:1114–5
spoofing **1**:142, 146, 329–37; **3**:1369
spores **4**:1954–67
sporopollenin **4**:1958
sports
  injuries **1**:102–3
  performance-enhancing drugs **5**:2507
spot tests **4**:1745–6
spousal abuse *see* domestic abuse; domestic violence
spousal homicide **5**:2398
spousal homicide–suicide **5**:2418
spouse revenge filicide **4**:2139
SPR *see* Small Particle Reagent
spray paints **4**:1932, 1942, 1947, 1948, 1949, 1951
spray pattern, bloodstains **1**:376, 377
spree murders **3**:1474, 1475
spring sagging, furniture **1**:209, 216–7
spurious malingering presentation **4**:1659
spurious signatures **3**:1445
SQ sequences, mtDNA **4**:1823, 1826–7
SSD (Scientific Support Department) **2**:621
SSM (scientific support manager) **2**:621
SSRIs *see* selective serotonin reuptake inhibitors
STA (systematic toxicological analysis) **4**:2127
stab wounds **5**:2646, 2647–9
stabbing incident bloodspatter **1**:370, 374
stability
  camera control **2**:636–7
  storage of specimens **4**:2104–6
staff training **4**:2227
  see also training
StAG program *see* Statistical Analysis of Glass program
staged reporting, CAI **2**:494
Stagg, Colin **4**:2158
staging **4**:2170; **5**:2316
stain identification **1**:314–9; **2**:757; **4**:1838
  see also bloodstains
staining techniques
  brightfield microscopy **4**:1771–2
  fiber analysis **4**:1979
  fingermark detection **3**:1302
  fingerprint enhancement **3**:1319
  histology **3**:1470–1
  pollen data **4**:1958–9
stainless steel **5**:2326
stalking **1**:40; **5**:2397–400
stamps
  fingerprints **1**:330, 332–3
  saliva DNA **2**:824
  serial numbers **5**:2325
standard of care, experts **4**:1664
standard deviations (SDs) **4**:2224
standard dimensions, paper **4**:1972
standard operating procedures (SOPs) **1**:13; **4**:2220
standard writings *see* exemplars, writings
standardization
  behavioral evidence **5**:2438
  CBRN agent analysis **2**:505
  crash tests **5**:2569–71
  diatom samples **2**:754–5
  DNA mixture interpretation **4**:1841
  electrical engineering **2**:922
  explosion debris analysis **2**:1040
  fingerprint identification **3**:1278–9
  fire investigation **3**:1171–4
  ignitable liquid residues **3**:1141, 1169
  soil analytical methods **5**:2380, 2383
  timber **5**:2644–5
  see also reference standards
Standards Australia **4**:1916
starch gels **1**:349, 350, 353
stare decisis **2**:782
Starkweather, Charles Raymond **3**:1475, 1477, 1478
START (short-term assessment of risk and treatability) **5**:2276
Starzepyzel judgment **3**:1453
state-dependent learning **1**:124
state prosecutors **2**:1007
state provisions, discovery **2**:780
State v. Allewalt, 517 A.P2d 741 (Md. 1986) **5**:2436, 2437
State v. Caddell, 287 N.C.266: 215 S.E.2d 348 (1975) **1**:254
State v. Koss, 551 N.E.2d 970 (Ohio, 1990) **5**:2433
State v. Sipin 130 Wash.App. 403 (Wash.App.Div. 1, 2005) **2**:581

statements, structure of **2**:974
states of mind, automatism **1**:253–6
static analysis **1**:145, 147
static approach, document dating **2**:686
static headspace sampling **3**:1139
static risk factors **5**:2598
stationary phase drug screening **5**:2522–4
Statistical Analysis of Glass (StAG) program **3**:1358; **5**:2591
statistics
  court communication of evidence **5**:2401–8
  DNA database evidence **2**:831–8
  evaluating **5**:2590–1
  fiber evidence treatment **3**:1100–1
  glass evidence **3**:1358; **5**:2590–1
  inference distinction **2**:968
  ink analysis methods **3**:1544
  interpretation **3**:1566–73, 1579–84
  missing persons calculations **4**:1810–7
  modeling **3**:1280, 1570–2
  mtDNA evidence **4**:1823–32
  parentage calculations **4**:1810–7
  performance testing devices **1**:329
  qualitative opinions vs **2**:972
  sampling plans based on **5**:2292
  sampling/quantity estimation **5**:2281–90
stature determination **1**:163, 194; **5**:2237
  see also height
status offenses **4**:1986–7
statute of limitations **4**:1713
statutory procedures
discovery **2**:779–80
  environmental audits **2**:951
  STDs *see* sexually transmitted diseases
steady state *see* equilibrium
steam distillation **3**:1142
steam generating tanks **2**:924
steam inhalation **3**:1537
steel **5**:2326, 2352
STEM (scanning transmission microscope) **4**:1793
step-parent evaluation **5**:2608
stepwise interviews **2**:533
stepwise mutation model (SMM) **5**:2375
stereomicroscopy **2**:988–9; **3**:1420–1; **4**:1762–3, 1791, 1792, 2002
stereotype induction, child interviews **2**:554
steroids **1**:291–2
Stickney v. Wyatt, 344 F.Supp. 387 (M.D. Ala. 1972) **5**:2581
sticky-side powder **3**:1307, 1318
stigmatization **4**:2152, 2187
stimulants **1**:116–7, 291; **4**:2207
  see also amphetamines; cocaine
stimulus/response theory **5**:2436
stings (insects) **2**:945
stippling, gunshot wounds **3**:1393
STIs *see* sexually transmitted infections
stochastic effects **3**:1640–1, 1643
Stockholm syndrome **5**:2409–12, 2439
stomach **1**:410; **5**:2499–500
  see also gastric contents
stopping power, projectiles **3**:1385
storage
  calibration solutions **4**:2223–4
  CBRN scene evidence **2**:505
  cigarette ends **4**:1930
  computer forensics **2**:584
  containers/preservation **4**:2104
  documentation **2**:608
  fiber tapings **4**:1929
  formaldehyde **4**:2101–2
  glass **4**:1929
  liquid samples **4**:1929
  specimens **4**:1927, 1929, 1930, 2101–2, 2104–6
  stability during **4**:2104–6
  wet miscellaneous traces **4**:1929
  see also packaging evidence
StPO (StrafprozeBordnung) **2**:1008
STR *see* short tandem repeats
StrafprozeBordnung (StPO) **2**:1008
strain **4**:1681
  see also stress–strain curve
strangers
  child sexual abuse **2**:531
  stalking **5**:2398
strangulation **1**:228, 232; **5**:2563
strategic environmental assessment (SEA) **2**:946
stratified samples **5**:2281
stratigraphy **1**:201–2
  see also soil layers
stray light, microspectrometry **4**:1752–5
street explosion scenes **2**:1031
Street Survival Skills Questionnaire (SSSQ) **4**:1733
strength of evidence calculation **4**:1824–7
stress
  compared with trauma **4**:2142
  crime victim decision making **2**:645–6
  hallucinations **3**:1432
  traumatic brain injury **3**:1463, 1464
stress cracking, plastics **4**:1687
stress markers on remains **1**:194–6
stress–strain curve **4**:1680–2, 1685
stretch injuries **1**:239
striae/striations
  consecutively matching **3**:1209
  definition **3**:1205
  firearms **3**:1205, 1209
  tools **5**:2485, 2487, 2489
strict liability, human factors **3**:1487
stringlines **3**:1220
strings, malware **1**:147
strip search strategy **2**:623
stroke direction, pens **5**:2661, 2663
STRs *see* short tandem repeats
structural approaches, psychopathology **4**:2190
structural framing, fire scenes **3**:1131
structural violence **1**:25
structure fires *see* compartment fires
structure, friction ridges **3**:1283
Structured Interview of Reported Symptoms (SIRS) **4**:1659
Structured Inventory of Malingered Symptomatology (SIMS) **4**:1659
Structured Professional Judgment (SPJ) **2**:673–4; **5**:2274, 2276
structured risk assessment **5**:2274
stubborn child parental alienation **4**:1982
stud guns **3**:1396

student threat assessment **5**:2454–8
stunning device wounds **3**:1396
stutter
  DNA profiling **1**:127; **2**:807
  DNA quantification **2**:805
  low copy number DNA **3**:1641
  low template DNA **3**:1567
  size **3**:1567
stutter bands **2**:807
style elements, handwriting **3**:1442
subadult age category **1**:179–80
subarachnoid bleeding **1**:406
subaverage intellectual functioning *see* mental retardation
subcapsular lacerations **1**:410
subclass characteristics, firearms **3**:1206
subcutaneous bruises **1**:399, 400
subcutaneous emphysema **1**:408
subdural hematomas **1**:265, 404–5
subdural hemorrhage (SDH) **5**:2341–5
subjectivity **2**:892; **3**:1437; **4**:2070
sublevel 1 propositions **2**:973; **3**:1569
submerged plant pollen **4**:1956
submission forms **2**:609, 610
subsampling **4**:1879; **5**:2291–2
subset analysis **2**:878–9
subsource level propositions **2**:973; **3**:1569
substance abuse **5**:2413–6
  adolescents **5**:2414
  aggression **5**:2415
  court-mandated treatment **2**:573
  criminality **5**:2414
  disorders **4**:2145
  mental illness **5**:2415
  monitoring **4**:1917–8
  pediatric psychopharmacology **4**:2214
  treatment **2**:573; **5**:2415–6, 2427–8
  violence **5**:2274
  see also addictions; alcohol abuse; drug abuse; psychopharmacology; substance use disorders
Substance Abuse and Mental Health Service Administration (SAMHSA) **4**:1908, 1911, 1916–7; **5**:2414
substance dependence **5**:2413
  see also alcohol, dependence; substance abuse; substance use disorders
substance exposure, toxicology **5**:2504, 2505, 2507
  see also alcohol; drug...
substance use disorders **2**:546; **4**:1881
  see also substance abuse; substance dependence
substantive threats **5**:2457, 2458
  see also threat assessment
substitution, urine tests **2**:862–3
substrates
  fingermark detection **3**:1292, 1302
  luminol tests **3**:1650–2
  security features in **3**:1265–7
subsurface reflectance, ink aging **3**:1544
subsystem experiments **5**:2571–4
subtractive color synthesis model **4**:2043–5
succession-driven entomological PMI **2**:943
  see also entomological succession
Suchey–Brookes age determination method **1**:181, 183
SUD (sudden unexpected death) **4**:1843–61
sudden cardiac death (SCD) **2**:468–80; **3**:1472–3
sudden death
  definitions **2**:468
  natural causes **2**:468–80
  see also sudden unexpected...
sudden infant death syndrome (SIDS) **1**:226, **2**:468, **4**:1844, 1851, 1855
  see also sudden unexpected infant death syndrome
sudden unexpected death (SUD) **4**:1843–61
sudden unexpected death due to neoplastic disease in infancy and childhood (SUDNIC) **4**:1851
sudden unexpected death in epilepsy (SUDEP) **2**:468; **5**:2303
sudden unexpected infant death syndrome (SUIDS) **2**:468
SUDEP *see* sudden unexpected death in epilepsy
SUDNIC *see* sudden unexpected death due to neoplastic disease in infancy and childhood
sufficiency
  threshold of **3**:1288
  victim decision making **2**:645
suffocation **1**:226–7
suggestibility **2**:732–3
  adult eyewitness memory **2**:1065–70
  children **2**:549, 551, 553–5; **3**:1592
  interrogative **2**:466–7, 591, 917–8; **3**:1590–3
  older adults **2**:917–8
  tests **2**:1068
  see also hypnosis; interrogations; repressed memory
suicide
  alcohol intoxication **1**:103
  commonalities **4**:2162, 2163
  fire consumption **3**:1531
  gunshot wounds **3**:1396, 1398
  homicide–suicide **2**:903, 905; **5**:2418–9
  notes **4**:2164–5
  pediatric psychopharmacology **4**:2215
  poisonous plants **1**:428; **4**:2058–9
  postmortem risk assessment **4**:2161–73
  postmortem toxicology **5**:2505
  risk factors **4**:2164, 2165–6
  sharp force injuries **5**:2646, 2649, 2651, 2654
  shooting scene investigation **3**:1221–2
  traffic fatalities **5**:2541
SUIDS (sudden unexpected infant death syndrome) **2**:468
sulfate pulp process **4**:1978
sulfur mustard hydrolysis **2**:527
sulfur tests **4**:1746
Sullivan v. R, 2 A111 E.R. 673, 675–676 (1983) **1**:255
sum value method, glucose metabolism **4**:2078
Summit, Ronald C. **2**:537, 539–40
'super max' prisons **5**:2583
super-soft plastic fingerprint spoofing **1**:330, 331–2
superimposition techniques **3**:1515–6; **4**:1893
supervision orders, juveniles **4**:1986
supplementary accreditation requirements **1**:11, 13–4
support services, elders **2**:909
supported employment **4**:1729

supravitality **2**:697–703
Supreme Court **2**:692–6, 1009; **3**:1334–5, 1619–21; **5**:2627–8
see also United States Supreme Court
surface-enhanced irregular reflection (SEIR) **1**:325, 332, 334
surface features, particles **4**:1785
surface marks *see* toolmarks
surface persistence, gunshot residues **3**:1192–3
surface reflectance, ink aging **3**:1544
surface tension, blood **1**:362, 365, 369
surfaces
characteristics **5**:2485, 2486–7
definition **5**:2486
texture influence on bloodspatter **1**:363–5
surgical treatment **1**:170; **2**:439–40; **4**:1702, 1703; **5**:2334
surveillance images **3**:1085, 1625; **5**:2611, 2614
see also closed-circuit television
surveillance visits, ASCLD/LAB program **1**:5
survival identification *see* Stockholm syndrome
susceptibility to misinformation *see* suggestibility
suspects
databases (footwear) **3**:1250, 1251
false confession vulnerabilities **2**:591–2
physical abuse **3**:1586
population probabilities **3**:1562–3
shooting scene investigation **3**:1223–4
sustained attention **4**:1866
sustained release opioids **4**:1900
Suzanski, T.W. **3**:1410
SVP (sexually violent predator) legislation **2**:668–9
swabs/swabbing **1**:320–2
DNA collection **2**:821, 822
explosion debris analysis **2**:1039
explosion scene evidence **2**:1026
gunpowder residue analysis **3**:1196
oral fluid collection **4**:1911
packaging **4**:1930
types **1**:321
SWAT teams **4**:2071–5
sweat **5**:2420–9
collection devices **5**:2422–6
drug distribution **5**:2420–2
glands **3**:1324
testing **5**:2426–8
Swedish '*per se*' legislation **2**:880–1
sweep sensors **1**:327
swelling, brain **1**:407
SWGDAM *see* Scientific Working Group on DNA Analysis Methods
SWGFAST *see* Scientific Working Group for Friction Ridge Analysis, Study and Technology
swipe transfer bloodstain patterns **1**:384–5, 385, 390
switches, electrical **2**:923
symphyseal surface, pubic bones **1**:181, 183
symptomatic epilepsies **5**:2299
synapses **4**:2203
synchronization, mobile phones **3**:1366–7
syncope, seizure imitation **5**:2302
syndromes **5**:2431–40
see also individual syndromes
synergistic alcohol–drug interactions **1**:110
syntactic complexity, cross-examination **2**:657
synthesis, chemical warfare agents **2**:525–6
synthetic fiber frequencies **3**:1099
synthetic gunshot residue particles **3**:1195
synthetic polymers **3**:1109–11
system variables
eyewitness testimony **2**:1077–8
juror knowledge **2**:1078
lineups **2**:1072
systematic toxicological analysis (STA) **4**:2127
systematics **1**:423
systolic blood pressure deception test **3**:1331–2

tactile hallucinations **3**:1432
Tactual Performance Test **4**:1865
'tagging' evidence **2**:1026
Takayama test **4**:1747
tandem mass spectrometry (MS/MS) **1**:138–9; **2**:434, 597, 881; **4**:1908, 2131
tangential abrasions **1**:396
tangible evidence **2**:498–500, 759
see also physical evidence
Tanner criteria, puberty **1**:190
tape methods
bloodstain pattern evidence **1**:387
lifting fibers **2**:986, 993, 994
trace evidence sample recovery **5**:2294
taphonomy **1**:203, 204; **4**:1674–5, 2089; **5**:2560
tapings, fiber **4**:1929
see also tape methods
TAPPI *see* Technical Association of the Pulp and Paper Industry
Tarasoff v. Regents of University of California, 551 P.P2d 553 (Cal. 1974) **2**:886–9
tardive dyskinesia **4**:2204
Target Disk Mode (TDM) **2**:585–6, 586
target fiber studies **3**:1099
targeted evidence collection **2**:948–9
TAT (thematic apperception test) **4**:2178
TATP *see* triacetone triperoxide
tattooing (gunshot wounds) **3**:1393
taxane-type pseudoalkaloids **4**:2065
taxonomy **1**:423
Taxus baccata **4**:2065
TBI *see* traumatic brain injury
TBW *see* total body water
TCP *see* Transmission Control Protocol
TDM *see* Target Disk Mode; therapeutic drug monitoring
TdP (*torsades des pointes*) **2**:471
TEA *see* thermal energy analyzer
teachers, custody evaluation **5**:2605
tear injuries **1**:239
see also lacerations
tear seam, air bags **1**:52
technical artifacts, DNA **1**:127
Technical Association of the Pulp and Paper Industry (TAPPI) **4**:1970, 1971
technical/scientific knowledge distinction **3**:1621
Technical Support Working Group (TSWG) **3**:1288, 1290
technical surface characteristics **5**:2486–7

Technical Working Group on Education and Training in Forensic Science (TWGED) **5**:2546
techniques *see* methodologies
technology
  AFIS **1**:249–53
  cellular phone networks **3**:1361–2
  glass **3**:1348
  multimedia **2**:579–82
  printing devices **5**:2663–72
  see also biometric devices; *individual technologies*
teenagers *see* juvenile...
teeth
  age determination **1**:161, 188, 189; **4**:1890–1
  biological profile reconstruction **4**:1889
  bitemark analysis **4**:1894–5
  DNA extraction **2**:1063
  DNA source **2**:826
  positive identification **4**:1891, 1892, 1893
  race determination **4**:1890
  radiocarbon dating **1**:420–1
  species determination **4**:1889–90; **5**:2394, 2395
  see also dental remains; odontology
telecommunications **3**:1360–71
telogen hairs **2**:825–6; **3**:1417, 1420, 1424
TEM *see* transmission electron microscopy
temazepam **2**:730
temperature
  accumulated degree-days/-hours **2**:942
  crime scenes **2**:939, 943
  enzyme activity **2**:954
  fire myths **1**:210, 222–3
  larval development stages **2**:937, 941, 942
  sweat collection **5**:2426
  time of death determination **2**:708–14; **5**:2467–72
  urine drug tests **2**:863
temporal ordering **2**:916
temporary guardianships **3**:1378–9
temporary insanity defense **5**:2443–8
  see also insanity defense
Tennessee v. Garner, 471 U.S. 1 (1985) **4**:2070
tensile strength, materials **4**:1682
tensile testing **4**:1681, 1682, 1685
Teplin, L.A. **2**:650–1
Terman, Lewis **4**:2177
terminal ballistics **3**:1202–4
terminal renal failure **4**:2082
terminus ante quem, soil **1**:200
terminus post quem, soil **1**:201
terrorism **2**:501, 502–3, 736–7; **5**:2507–8
  see also bioterrorism; bombings
Terry v. Ohio, 392 U.S. 1 (1968) **4**:2157
Test of Memory Malingering (TOMM) **4**:1661, 1868
testability *see* falsifiability; tests
testamentary capacity **2**:438–9, 908
testamentary trusts **3**:1376
testator coercion **2**:438
testimony
  children **2**:549–50, 550–1, 552
  competency to stand trial **2**:461
  corroboration (mass graves) **4**:1678
  cross-examination impact **2**:656–62
  hypnotic **2**:736; **3**:1500–4
  older adults **2**:918
  trace evidence consistency **3**:1564–5
  see also expert testimony; eyewitness testimony; foundation testimony; witnesses
testing laboratories
  accreditation **1**:3–8, 11–2, 13, 14–5, 16
  quality management **5**:2493–4
testosterone **3**:1430; **5**:2333, 2334
tests
  norms **4**:2176–7
  pharmacogenomics **4**:2014–5
  police use of force **4**:2069–70
  results communication **2**:492
  results provision/interpretation **2**:494
  user qualifications **4**:2198
  see also individual tests
Δ^9−tetrahydrocannabinol (THC) **2**:432–3, 434
  administration route **2**:435
  alcohol interactions **1**:113
  analysis **2**:434
  drug-facilitated crime **2**:871
  mechanism of action **2**:434
  oral fluid **4**:1906, 1907, 1908–11, 1916, 1919
  pharmacological effects **2**:434
  regulatory aspects **2**:436
  sweat **5**:2425
  urine tests **2**:866
  see also cannabis
tetrazepam **4**:1915
tetryl (CE) **2**:1041
textiles
  contact traces **2**:992–4
  examination of **2**:985–96
  fibers **2**:986–8
  individualization **2**:996
  intelligence work **2**:992–6
  reconstruction of **2**:995
texture of surface influence **1**:363–5
thalamo-cortical system **4**:2203
THC *see* Δ^9-tetrahydrocannabinol
THC-COOH **2**:434, 435, 871–2; **4**:1911; **5**:2425
thematic apperception test (TAT) **4**:2178
'themes' (interrogations) **3**:1587
thenar side
  feet **3**:1323
  hands **3**:1323
theodolite device **2**:580
theoretical information, capacity assessment **2**:447
therapeutic communities **5**:2416
therapeutic drug monitoring (TDM) **4**:2015, 2019, 2226
therapeutic jurisprudence (TJ) **4**:1718–9; **5**:2449–54, 2580–3, 2586–7
therapeutic medications *see* medications
therapeutic privilege **2**:452
therapists, custody evaluation **5**:2605
  see also mental health professionals
thermal cycler, DNA **3**:1639
thermal decomposition **3**:1107, 1111
thermal dot matrix printers **5**:2665–6
thermal dye printing **5**:2670–2, 2672
thermal energy analyzer (TEA) **2**:1040, 1042
thermal fusion printing **5**:2672, 2673–4

thermal inkjet printers **5**:2670
thermal mass transfer **5**:2672, 2673
thermal paper **3**:1300; **5**:2673
thermal sensors **1**:327
thermal wax transfer printing **5**:2672, 2673
thermoionic gun **4**:1793, 1795
thermomechanical pulps (TMP) **4**:1978
θ *see* population genetics parameter
thickness of paper **4**:1968, 1973
thin film interference colors **4**:1779
thin layer chromatography (TLC) **2**:597, 844, 992, 1039; **3**:1543
thiobarbiturates **2**:729
thiodiglycol **2**:527
third-degree burns **3**:1530
Thomas, Philip **3**:1176–7
Thompson, Jennifer **2**:1075
Thompson v. Oklahoma, 487 U.S. 815 (1988) **2**:717
thought suppression **5**:2247
thread security (documents) **3**:1267, 1271, 1273
threat assessment
  behavioral science evidence **1**:287
  history **5**:2460–1
  limitations **5**:2458
  meaning variations **5**:2461–2
  mental health practitioners **5**:2597–601
  schools **5**:2454–8
  search strategies **2**:624
  workplace **5**:2460–5
  see also risk assessment
threat location analysis **1**:329
threatening communications *see* threat assessment
three-dimensional (3D)
  identification techniques **5**:2614, 2615
  modeling **3**:1628
  reconstructions **5**:2257–67
thrill killings **3**:1476
thromboembolism **4**:1846–7
thumb pressure criteria, livor mortis **2**:703–4, 705
thyroiditis **4**:1853–4
TIAFT.org **5**:2625
tibial side, feet **3**:1322, 1323
TIC *see* total ion chromatograms
'ticking bomb' scenario **2**:736–7
TICs *see* toxic industrial chemicals
timber **5**:2644–5
  see also wood
time of death determination **2**:697–716; **5**:2466–79
time-dependent differences
  postmortem redistribution **4**:2097
  stability during storage **4**:2105
time-of-flight (TOF) mass spectrometry **2**:597; **4**:1805, 2131
time-of-flight secondary ion mass spectrometry (TOF-SIMS) **3**:1195, 1598
time-oriented assessments **3**:1461
time since discharge (firearms) **3**:1198
time/temperature fire myth **1**:210, 222–3
timing, exemplar writings **3**:1440
TIMs *see* toxic industrial materials
tire marks **2**:602–3, 967; **5**:2480–4
tire tracks **5**:2482
tissues
  animal individualization **5**:2637
  bone relationship **3**:1087, 1090
  burn injuries **3**:1532, 1535, 1537
  chemical injuries **3**:1539
  depths (cranial points) **3**:1088
  DNA extraction **2**:1063
  DNA sources **2**:826
  forensic histology **3**:1469–70
  postmortem toxicology **5**:2500–1
  radiocarbon dating **1**:419, 420
  sampling **4**:2103
  shrinkage of **3**:1532
  storage containers **4**:2104
TJ *see* therapeutic jurisprudence
TLC *see* thin layer chromatography
TMDT (trace metal detection technique) **3**:1212, 1215
TMJ (total metal jacket) bullets **3**:1190
TMP (thermomechanical pulps) **4**:1978
TNAZ *see* 1,3,3-trinitroazetidine
TNT *see* 2, 4, 6-trinitrotoluene
toes *see* foot
TOF *see* time-of-flight
TOF-SIMS *see* time-of-flight secondary ion mass spectrometry
tolerance to drugs **1**:292–3
  alcohol **1**:123–4
  benzodiazepines **1**:296
  cocaine **2**:565, 566
  ethanol **1**:74–6
  opioids **4**:1899
  postmortem artifacts **4**:2095
tolerances (ACE-V process) **3**:1277
tolling doctrine **4**:1713
toluene **3**:1149, 1162, 1169
TOMM *see* Test of Memory Malingering
tonal modification of images **4**:2054–5
toner powder (photocopiers) **3**:1596–7
tongue **1**:230; **3**:1532
toolmarks
  casting **2**:967; **5**:2489–92
  documentation **2**:602–3
  examination of **5**:2485–94
  expert opinion evidence **5**:2492–3
  firearms identification **3**:1205–6, 1209
  harmonized conclusion scale **5**:2494
  individual characteristics **5**:2486–7
tools
  definition **3**:1205
  packaging **4**:1928–9
  working surface characteristics **5**:2486–7
tools (techniques) *see* methodologies
top coat paints **4**:1933, 1943, 1945–6
topography, search strategy **2**:624
Toronto mental health courts **4**:1719, 1720
torsades des pointes (TdP) **2**:471
tort law **1**:282; **4**:1664–5, 2210
torture **2**:736; **3**:1586
total body water (TBW) **1**:58, 61, 68
total ion chromatograms (TIC) **3**:1143, 1146; **5**:2526
total metal jacket (TMJ) bullets **3**:1190
total station device **2**:580
totality of circumstances
  confessions **3**:1590

Miranda rights waiver **2**:465
police use of force test **4**:2070
touchless fingerprint recognition **1**:325
toughness of materials **4**:1682
Tourette's disorder **5**:2431
towel fabric marks **4**:1669
Townsend v. Sain, 372 U.S. 293 (1963) **2**:736
toxic industrial chemicals (TICs) **2**:507–8, 519, 520, 522–3
see also chemical warfare agents
toxic industrial materials (TIMs) **2**:507, 519, 520, 522–3
see also toxic industrial chemicals
toxicity
amphetamines **1**:137–8, 138
benzodiazepines **1**:296
cocaine **2**:563–5, 566
ethanol **1**:76–7
opioids **4**:1900
toxicology
analytical methods **2**:595–600, 874–6; **4**:2119–34
applications **1**:290–3; **5**:2503–9
artifacts **4**:2093–108, 2118
autopsy ancillary investigations **1**:260
BAC interpretation **1**:70
burn injuries **3**:1535–6
external quality assurance **4**:2226–7
forensic/clinical distinction **5**:2503
immunoassays **5**:2511–6
initial screening **5**:2509–33
insect analysis **2**:944
interpretation of results **4**:2115–9
'kitchen sink' ecology **2**:951–2
opioid data **4**:1902
oral fluid **4**:1903–20
pharmacogenomics **4**:2012–20
plants **1**:428; **4**:2066
sample extraction **5**:2516–9
sample selection **5**:2510–1
specimens **1**:89–90; **2**:873–4; **4**:2101–6; **5**:2495–502
sweat **5**:2420–9
web resources **5**:2625
see also drug...; postmortem toxicology
toxins **1**:301
see also toxic...
trace, speech as a **5**:2389
trace DNA **2**:823; **4**:1838
trace evidence
categorical opinions **3**:1564–5
'consistent-with' testimony **3**:1564–5
diatoms **1**:428–9
documentation **2**:602
explosion debris **2**:1033, 1044
fibers/textiles **2**:985–96
ideal properties **5**:2379
legal interpretation **3**:1561–5
light microscopy **4**:1762–89
packaging **4**:1929
paint **4**:1950–1, 1952
particle forms **4**:2001–7
persistence **5**:2534–9
probabilistic statements **3**:1561–4
sampling **5**:2291–5
soil **5**:2378–83
transfer **5**:2534–9
wildlife forensics **5**:2637
see also hair/hair analysis; sample matches
trace metal detection (firearms) **3**:1211–5
traced simulated signatures **3**:1445–6
traces of production, illicit drugs **2**:853
tracheids **4**:1976, 1979, 2005; **5**:2642, 2643
traffic accidents *see* road traffic accidents
traffic fatalities **5**:2541–4
trafficking drugs **2**:851–9
see also illicit drugs
trail patterns, fire **1**:215, 217
train bombings, Madrid **3**:1290, 1577
training
criminalistics **5**:2545–54
education distinction **5**:2554
electrical engineering **2**:921–2
expert witnesses **2**:1014
handwriting examiners **3**:1447–8
quality management **4**:2227
sex offender treatment **5**:2337
trajectory determination, projectiles **3**:1202, 1204
transcapsular lacerations **1**:410
transfer of evidence **5**:2534–9
animal hair **3**:1403–4
biological traces **5**:2537
bloodstain patterns **1**:384–5, 390
CAI example **2**:489–92
DNA **3**:1643; **5**:2555–6
factors affecting **5**:2534
fibers **3**:1098; **5**:2534–5, 2538
glass **3**:1357; **5**:2535–6, 2538
gunshot residue **5**:2536–7
hair **3**:1418; **5**:2535, 2538
hydrocarbon fire accelerants **5**:2537
mechanical processes **5**:2534
metal particles **5**:2537
paint **4**:1933, 1940, 1951, 1952; **5**:2536, 2538
petrol **5**:2537
pollen **5**:2537
probability **2**:971; **3**:1357
problems **5**:2537–8
soil materials **5**:2378–9, 2380
transfer of juveniles to adult justice **3**:1612–5
transfer mechanisms, drugs **4**:1906
transfer patterns, bloodstains **1**:384–5, 390
TRANSFER software package **3**:1358; **5**:2591
transflection **4**:1756
transformative events, multiple murders **3**:1477
transient global amnesia **5**:2302
transient ischemic attacks (TIAs) **5**:2302
transient threats **5**:2457, 2458
transillumination **1**:400
translucency of particles **4**:1785
Transmission Control Protocol (TCP) **1**:141, 142
transmission electron microscopy (TEM) **4**:1793; **5**:2385
transmission microspectrometry **4**:1751–5
transmission mode, optical enhancement **4**:2042, 2046
Transmit Terminal Identifier (TTI) **5**:2674
transmitted light microscopy **3**:1421

transparency of particles **4**:1785
transporting evidence **4**:1927–31
transposed conditional **2**:975
  see also prosecutor's fallacy
transverse section, timbers **5**:2645
trash marks (photocopiers) **5**:2673
trauma
  automotive impact analysis **5**:2565–75
  battered child syndrome **1**:263–9
  causation (biomedical engineering) **5**:2565–75
  children's suggestibility **2**:555
  dissociative disorders **2**:784, 785–7, 788
  focus of attention **2**:787, 1077
  functional amnesia **2**:786
  hypnosis **3**:1502
  individualization markers **1**:169–70
  rape trauma syndrome **5**:2241–2
  skeletal remains analysis **5**:2557–63
  Stockholm syndrome **5**:2409–12
  stress comparison **4**:2142
  see also posttraumatic stress disorder; traumatic...
traumatic amnesia **4**:1715; **5**:2244–6
traumatic aortic ruptures **1**:409
traumatic asphyxia **1**:227
traumatic brain injury (TBI) **3**:1459–64; **4**:1872
traumatic memory **5**:2243–8
traumatic subarachnoid bleeding **1**:406
traumatic subdural hemorrhage **5**:2341
tread patterns **3**:1240–3, 1249; **5**:2481
  see also sole databases
treatment
  addictions **1**:20
  disaster mental health relationship **2**:762–3
  errors (malpractice) **4**:1702, 1703
  mental health **2**:762–3; **4**:1720; **5**:2580
  parental alienation **4**:1983–4
  psychopathy **4**:2195
  refusal of **5**:2336, 2452, 2584–7
  right to **5**:2580–3
  sex offenders **5**:2332–7
  substance abuse **5**:2415–6
  see also mandated treatment; medical treatment
trees **2**:828
  see also forests; plants; wood
tri-dimensional marks (tires) **5**:2480
triacetone triperoxide (TATP) **2**:1030, 1037, 1041, 1044, 1055–6
triad of injuries **5**:2339–45
triage **2**:769
triallelic patterns, STRs **5**:2361
trials
  evidence comprehension **3**:1605–6
  evidence evaluations **3**:1604
  expert testimony use **2**:998–1000
  juvenile defendants **2**:547
  see also criminal trials; jury trials
tribal groups **1**:281–2
trichromacy **4**:2043
tricyclic antidepressants **4**:2205
trier of fact prerogatives **1**:286–7
  see also finder of fact
Trikha v. Wenden, 116 A.R. 81 (Q.B.1991) **2**:670
1,3,3-trinitroazetidine (TNAZ) **2**:1060
2,4,6-trinitrotoluene (TNT) **2**:1033, 1044, 1060
Trojans **1**:143
truancy **4**:1984
true father hypotheses **4**:1814–5
true flies *see* flies
true memories **4**:1715
trust component, interrogative suggestibility **3**:1591, 1592
trusts (guardianships) **3**:1376
truth
  memory and **2**:728–9, 730–5
  type elicited by truth serum **2**:735
  see also true...
truth detection **2**:590, 592–3
  see also lie detection; truth serum
truth serum **1**:286–7; **2**:728–37
truthstate, Bayesian networks **1**:278
TSWG *see* Technical Support Working Group
TTI (Transmit Terminal Identifier) **5**:2674
tungsten bulbs **3**:1633
TWGED *see* Technical Working Group on Education and Training in Forensic Science
twins **3**:1330
two-dimensional (2D) techniques **5**:2614
two-vehicle collisions **5**:2251–4
Tylenol tampering incident **5**:2508
Tyler v. Michigan **3**:1123
Type I error **2**:955; **3**:1423
  see also false positives
Type II error **2**:955; **3**:1423
  see also false negatives
typeball technology **5**:2664–5
typebar machines **5**:2663–4
typewheel technology **5**:2664–5
typewriting **1**:130; **3**:1558, 1559, 1597; **5**:2663–5
tyres *see* tire...

U-pattern fire plumes **3**:1120
UAC (urine–alcohol concentration) **1**:77
UK *see* United Kingdom
Ulex europaeus **1**:343
ulnar side, hands **3**:1322, 1323
'ultimate issue' opinion testimony **5**:2589–90
ultrarapid metabolizers (UMs) **4**:2014, 2018
ultrasensitive DNA tests **3**:1643
ultrasonic sensors **1**:327
ultrasound **5**:2234
ultraviolet (UV) light
  drug profiling **2**:845
  fingermark detection **3**:1293, 1310
  fingerprint enhancement **3**:1318
  liquid chromatography **2**:597–8
  postmortem toxicology **4**:2126, 2128–30
  stain identification **1**:314, 315, 317
ultraviolet/visible (UV/Vis) spectrometry **2**:845, 990, 991; **4**:2128–30
UMs *see* ultrarapid metabolizers
Unallocated clusters, Mac computers **2**:586
unbefriended elders **3**:1375
uncertainty
  Bayes theorem **2**:485–7
  interpretation of results **3**:1581
  interrogative suggestibility **3**:1591
  of measurement **2**:600

unconscious motivation **2**:574
unconsciousness defense **1**:253–6
unconvincing disclosure of abuse **2**:539
undergraduate programs **2**:898
underhairs **3**:1406
undesired combustion **3**:1175–6
'undetermined' manner of death **5**:2542
undue influence **1**:287; **2**:438, 906; **3**:1377
unethical conduct *see* ethics
'unexplained' multiple murders **3**:1477
unforeseeability, product use **3**:1488–92
uniaxial crystals **4**:1770, 1775–6
Uniform Health-Care Decisions Act 1993 **2**:451
uniform prior probabilities **3**:1562–3
unintentional poison ingestion **4**:2058
unique reference numbers (URNs) **2**:766, 768–9
uniqueness **3**:1508, 1509
  friction ridge skin **3**:1329–30
  probabilities **3**:1565
  source attribution approach **2**:832
  STR profiles **5**:2373
United Kingdom (UK)
  adolescent female offenses **1**:37
  civil commitment **2**:557, 558, 669–70
  disaster victim identification **2**:765–7, 771
  discovery **2**:774–5, 782–3, 783
  expert evidence admissibility **2**:1003, 1004
  inpatient commitment **2**:557
  laboratory accreditation **1**:12
  major crime scene management **2**:620–1
  National DNA Database **2**:678
  trial and appeals system **2**:999–1000
  see also Civil Procedure Rules
United States (US)
  adolescent female offenses **1**:37
  alcohol use **1**:122
  Bureau of Alcohol, Tobacco, Firearms and Explosives **3**:1132
  cannabis regulation **2**:435, 436
  criminalistics training **5**:2546
  discovery **2**:774, 775–8, 778–81
  evidential role of mtDNA **4**:1836
  expert malpractice **4**:1663
  expert opinion evidence **2**:692–6, 1001–2, 1004, 1006; **3**:1331–1334–5, 1619–21; **5**:2627–8
  female incarceration **1**:38
  fingerprint comparison **3**:1288–9, 1290
  guardianships **3**:1373
  inpatient civil commitment **2**:557
  jail population mental illness **2**:652
  laboratory accreditation **1**:1–3, 12, 15, 16
  mental health courts **2**:462
  National DNA Database **2**:678
  outpatient civil commitment **2**:558
  peer reviews **4**:2010–1
  postconviction DNA testing **2**:682–3
  school threat assessment **5**:2454–8
  Tarasoff statutes **2**:888
  terrorism **5**:2508
  trial and appeals system **2**:998–9
  workplace drug testing **5**:2507
  see also Federal Rules of Evidence
United States Supreme Court (USSC) **2**:717–8, 719; **4**:1730–1, 1818
  see also Supreme Court
United States v. Agurs 427 U.S. 97 (1976) **2**:779, 781
United States v. Byron Mitchell, CA-No. 96–407 (1999) **3**:1289
United States v. Denny-Shaffer, 2 F.P3d 999 (10th Cir., 1993) **5**:2435
United States v. Mitchell, 365 F.P3d 215 (3rd Cir. 2004) **4**:2011
United States v. St. Jean (U.S. Ct. of Appeals for Armed Forces, 1996) **4**:2163
United States v. Sokolow, 490 U.S. 1 (1989) **4**:2157
universal gas constant **3**:1107
university-associated institutes **1**:151
university degrees **5**:2545–6
unpolarized light **4**:1769–73
unstructured risk assessment **5**:2274, 2275
  see also clinical prediction
unwanted child filicide **4**:2139
upper airway obstruction **1**:225–7
upper bound frequency estimation **4**:1824–5
urea **2**:1032, 1038; **4**:2083–4; **5**:2473
urine
  amphetamine effects **1**:137
  blood-alcohol concentrations **1**:71
  diabetes mellitus **4**:2079
  postmortem toxicology **4**:2121; **5**:2497–8
  sampling **4**:2103
  stain identification **1**:318
urine–alcohol concentration (UAC) **1**:77
urine tests **2**:860–7
  cannabis **2**:435
  cocaine **2**:567
  defeating **2**:862–5
  DFC investigations **2**:871–2, 873–4
  driver impairment **2**:881
  EQA schemes **4**:2227
  hair analysis comparison **3**:1430
  oral fluid concordance **4**:1917
  sweat testing alternative **5**:2427–8
  workplace drug testing **5**:2507
URNs *see* unique reference numbers
urobilinogen **1**:319
US *see* United States
user support, quality management **4**:2227
USSC *see* United States Supreme Court
utilitarianism **5**:2307
UV light *see* ultraviolet light
UV/Vis *see* ultraviolet/visible

V-pattern fire plumes **1**:209, 220–1; **3**:1120
VA *see* Veterans Administration
vacuum cyanoacrylate (VCA) fuming **3**:1301
vacuum metal deposition (VMD) **3**:1302–3, 1320
vacuuming methods, evidence **2**:1026; **5**:2294
vaginal injuries **1**:238; **2**:534
vaginal secretions **1**:17
vaginal tract DNA **2**:821–2
validation/validity testing **4**:2221–5
  alcohol analysis **1**:90–2
  assays (toxicology) **4**:2134
  confirmation testing **2**:600
  diatom tests **2**:751–2

validation/validity testing *(cont)*
expert opinion evidence **2**:695, 1002
fire modeling **3**:1185–6
Hare Psychopathy Checklists **4**:2200
mathematical modeling **5**:2575
measurements **3**:1628–30, 1631
novel scientific methods **4**:2009–11
speech analysis **5**:2391–2
time of death determination **5**:2467
urine drug tests **2**:863
Validity Indicator Profile (VIP) **4**:1661
'value for money' CAI **2**:485
values **4**:2187, **5**:2297
valves, diatoms **2**:749
Van der Waals force **1**:362
vanishing points **3**:1626–7
vapor pressures, explosives **2**:1033
variable number tandem repeats (VNTRs) **1**:140; **2**:795–6, 800; **4**:1749; **5**:2595
see also minisatellites
vascular pattern recognition **1**:327–8, 336–7
vasovagal inhibition **1**:232
VCA (vacuum cyanoacrylate) fuming **3**:1301
vector engraving **3**:1273
vector template matching **2**:893
vehicles
air bag systems **1**:51–7; **5**:2543, 2572
air filter pollen evidence **4**:1965
crime scene evidence **2**:617–9
digital data **5**:2572
documents (crime scenes) **2**:612
explosion debris **2**:1031
forensic examination **3**:1636–7
interior environment **5**:2571
light globe/filament examination **3**:1632–8
luminol tests **3**:1652
paints **4**:1932, 1939, 1943–7
speed (reconstructions) **5**:2250–4, 2255–6
see also road traffic accidents
veins
individualization markers **1**:168
pattern recognition **1**:327–8, 336–7
velocity of blood **1**:365, 369–75, 376–7
venoms **1**:301
'venting' ground **1**:204
verbal abuse of elders **2**:904, 905
verbal insults **1**:25
verbal learning tests **3**:1462
verbal probability, handwriting **3**:1440
verbal scale, statements **2**:974
verbatim memory traces **2**:1067, 1068
verdicts, juries **3**:1603–6
verification
biometric devices **1**:322, 324
facts **3**:1601
fingerprint comparison **3**:1285
peer reviews **4**:2010
toolmark examination **5**:2485
vertex posterior **3**:1428
vertical lighting **3**:1595
Vesalius, Andreas **1**:256
vessel elements (hardwoods) **4**:1979; **5**:2642
vessel injuries (burns) **3**:1533
Veterans Administration (VA) **2**:447, 453–4
vibices **2**:703
Vicea graminea **1**:344
victimization rates, child sexual abuse **2**:529–30
victimization risk factors **1**:265; **2**:915
victims
battered child risk factors **1**:265
contact with CSI **2**:615
credibility rehabilitation **2**:541
elder abuse risk factors **2**:915
models (reconstructions) **1**:390–5
perpetrators as **5**:2432–5
shooting scene investigation **3**:1223
of violence **1**:26
witness assistance programs **2**:909
see also crime victims; identification
video game violence **1**:30
video spectral analysis (VSA) **3**:1258–9
video spectral comparator (VSC) **4**:2047
video/videotape
child interviews **2**:551
crime scene photography **2**:641–3
evidence **2**:584, 587, 603, 746
examination of children **2**:657, 658
interrogations **2**:593; **3**:1588
surveillance systems **5**:2611
textile images **2**:995–6
see also surveillance images; video game violence
viewpoint (photography) **2**:627, 632, 633
vigilance **4**:1866
vigilantism **2**:646
Vinella's solution **5**:2327
violence **1**:24–32
assessment **1**:30
behavior **1**:26–7; **2**:572, 744; **4**:1878
causes **1**:28–30
concept of evil **2**:982
consequences **1**:27
crime investigation **2**:934
definitions **1**:25–6
fantasies of **5**:2315, 2321
genetics relationship **3**:1335, 1336
homosexual panic defense **3**:1481
impulsive behavior **2**:572
mental health practitioner threat **5**:2597–8, 2599
perpetrator responses **1**:27
prevalence data **1**:28
prevention **1**:31; **5**:2464–5
risk assessment **2**:671–4; **5**:2272–7, 2461–2, 2599
school threat assessment **5**:2454–8
seizure disorder **5**:2304–5
sex differences/similarities **1**:38–42
stalking **5**:2398, 2399
substance abuse **5**:2415
treatment **1**:30
workplace threat assessment **5**:2462, 2464–5
see also dangerousness; duty to warn
Violence Risk Appraisal Guide (VRAG) **5**:2275
VIP (Validity Indicator Profile) **4**:1661
viral genotyping **3**:1342–7
viral pneumonia **4**:1844–5
Virchow, Rudolf **1**:257–8
Virginia model threat assessment **5**:2456–8

Virginia Tech shooting **3**:1475; **5**:2455, 2456
VIRTOPSY project **5**:2264–6
Virtual Reality Modeling Language (VRML) **5**:2263–4
viruses **1**:301; **2**:501; **3**:1342–7, 1470
  see also viral...
viscosity of blood **1**:362
visible fingermark detection **3**:1292
visible light **3**:1293; **4**:1750
visitation evaluation **5**:2602–11
visual approach, speech analysis **5**:2390
visual examination
  bomb mechanisms **2**:1035
  luminol tests **3**:1653
  serial numbers **5**:2325
  shooting-distance estimation **5**:2351–3
  see also examination...
visual hallucinations **3**:1432, 1433, 1434
visual-motor tasks **3**:1462
visualization
  3D reconstructions **5**:2257–67
  short tandem repeats **5**:2356–7
visuospatial functioning **4**:1867
vitality
  diagnosis **3**:1535–6
  signs of **1**:400
Vitek v. Jones, 445 U.S. 480 (1980) **5**:2583
vitreous humor (VH) **1**:71; **4**:2078, 2103, 2122; **5**:2472–3, 2498–9
VMD *see* vacuum metal deposition
VNTRs *see* variable number tandem repeats
vocational rehabilitation agencies **4**:1729
VOCs (volatile compounds) **1**:92
voice recordings **2**:840
  see also speech analysis
voice stress analysis **2**:723
void bloodstain patterns **1**:362
Voir Dire **2**:759, 1005
  see also expert witnesses
volar pads **3**:1328
volatile components, ink **3**:1544
volatile compounds (VOCs) **1**:92
volatile substance abuse (VSA) **4**:2116, 2133–4
volatile substances **4**:1929–30, 2102, 2105, 2116, 2133–4
volume crime scenes **2**:615–6, 619–25
  see also crime scenes
volume/mass magnetic susceptibility methods **5**:2385
volume percentages, alcohol **1**:59–60
voluntary false confessions **2**:589; **3**:1590
volunteers
  disaster mental health **2**:762
  guardians **3**:1372
vomit/vomiting **1**:78, 427; **2**:874; **4**:1899
voyeurism **5**:2317
VRML (Virtual Reality Modeling Language) **5**:2263–4
VSA *see* video spectral analysis; volatile substance abuse
VSC (video spectral comparator) **4**:2047
vulnerability
  adolescent peer pressure **3**:1610–1
  assessment (workplace) **5**:2461
  biometric devices **1**:329–37
  elder abuse definition **2**:903
  sense of (crime victims) **2**:645

Wainwright v. Ford see Ford v. Wainwright
waivers
  insanity defense **2**:442
  legal counsel **2**:442, 462
  Miranda rights **2**:463–7
  patient **2**:439–40
  privilege **2**:440
  of rights **2**:463–7, 999; **4**:1986
Wales **2**:557
  see also United Kingdom
walk throughs, fire scenes **3**:1124
Walker, Lenore **5**:2432
walking feet impressions **3**:1245
Wallner lines **4**:1687
WAPs (wireless access points) **1**:142
war crimes **3**:1516; **4**:1676, 1677
Ward v. R (1993), 1 WLR 619 **2**:782; **5**:2270
washing machine human factors example **3**:1492
Washington v. Harper, 494 U.S. 210, 110 S.Ct. 1028 (1990) **5**:2586
water content, alcohol **1**:58, 60–1, 68, 71, 82
water-covered impressions **2**:965
water imbalances **4**:2084–5
water-immersed bodies **5**:2474–8
water loading, urine tests **2**:863
water samples, diatoms **2**:753, 754
water sports injuries **1**:102–3
Waterhouse–Friderichsen syndrome **4**:1855
watermarks, documents **3**:1265, 1275
wave cast-off bloodstain patterns **1**:365
wave interference (microscopy) **4**:1778
wave retardation (microscopy) **4**:1778, 1779
wave vibrations **4**:1779, 1780
wavelength dispersive X-ray spectroscopy (WDX) **4**:1800
wavelengths, light **3**:1293
  see also wave...
wax transfer printing **5**:2672, 2673
WDX (wavelength dispersive X-ray spectroscopy) **4**:1800
weak impression fault, printers **5**:2667
weapon focus **2**:1077
weapons **3**:1380–6
  biological agents **1**:301–2
  eyewitness testimony **2**:1077
  packaging **4**:1928–9
  types **3**:1380
  wildlife forensics **5**:2637–8
  see also chemical warfare agents; firearms; *individual weapons*
weapons-grade anthrax **1**:301
weapons of mass destruction (WMD) **2**:500, 501, 1020; **4**:1883
  see also chemical, biological, radiological and nuclear agents
wear and tear
  printers **5**:2667
  shoemark examination **3**:1233–4
  technical surfaces **5**:2487

weather data **2**:939, 943
weave patterns, paper **4**:1973–4
web of blood phenomenon **1**:389
web resources **5**:2619–26
  see also Internet
web servers **1**:145
Wechsler Memory Scale (WMS) **4**:1866
wedge spectrograms **4**:2049
Wehnelt, SEM **4**:1795
weight (bodies) *see* bodyweights
weight of evidence **2**:1016–7
  see also likelihood ratios
weighting fibers **4**:1975
Weir, B.S. **2**:834
Weisgram v. Marley, 528 U.S. (2000) **5**:2627–8
welding (serial number obliteration) **5**:2325
Wenden v. Trikha, 116 A.R. 81 (Q.B.1991) **2**:670
Weschsler, David **4**:2177
Westroads Mall shooting **3**:1475
wet and dry swabbing technique **1**:320, 321
wet items
  packaging **4**:1928, 1929
  storage **4**:1929
  transport **4**:1929
wettability of surfaces **1**:369
Wexler, David **5**:2449, 2450, 2452, 2581
WGA *see* whole genome amplification
Wheat v. State, 527 A.2d 269275 n.5 **2**:541
Whitaker effect **4**:1838
white powder fingerprint enhancement **3**:1318
Whitman, Charles **3**:1475–6
whole genome amplification (WGA) **2**:808–9; **5**:2628–33
whole item removal, trace evidence **5**:2294
whole-sample soil screening **5**:2384–5
Widmark equation **1**:64–6, 67–8
Wiener restoration, CCTV images **3**:1521
Wiener's theory, blood groups **1**:346
wife battering *see* battered woman syndrome
Wiggins v. Smith, 539 U.S. 510 (2003) **4**:1818
'wild beast' test of insanity **5**:2444
wildlife forensics **5**:2635–9
Wildman, A.B. **3**:1403
wills, testator capacity **2**:438–9
wind-pollinated plants **4**:1955, 1956
window glass **3**:1348
window thread, documents **3**:1267
Winnicott, D. W. **4**:1995
wipe approach, sweat **5**:2425, 2426, 2428
wire marks, paper **4**:1973–4
wireless access points (WAPs) **1**:142
wiring, fire investigation **2**:925
Wisconsin **5**:2578, 2579
Wisconsin Card Sorting Test **4**:1867
witchcraft hysteria **2**:978–9
withdrawal from drugs **1**:292–3; **4**:1899
withholding information **2**:774–5
witnesses
  assistance programs **2**:909
  chain of possession **2**:499
  credibility **2**:918–9
  examination of **3**:1505
  materials (explosion scenes) **2**:1026
  memory impairment **2**:918
  older adults **2**:916–9
  reports (cross-contamination) **2**:1067
  testimony (mass graves) **4**:1678
  see also child witnesses; cross-examination; direct examination; expert witnesses; eyewitness. . .; lay witnesses
WMD *see* weapons of mass destruction
WMH-CIDI (World Mental Health–Composite International Diagnostic Interview) **4**:1880
WMS (Wechsler Memory Scale) **4**:1866
women
  legal syndromes **5**:2439
  sentencing **5**:2308–9
  see also battered woman syndrome; female. . .; gender
women's movement **2**:733
wood **5**:2640–6
  botany **1**:426–7
  charring rate **3**:1112
  combustion **3**:1109
  ecological factors **5**:2641–2
  fibers (papermaking) **4**:1976
  firearm discharge residue patterns **5**:2352
  forensic value **5**:2641
  genetic factors **5**:2641–2
  identification **1**:426–7
  microscopic structure **5**:2642–4, 2645
  particle forms **4**:2005
  serial number restoration **5**:2326
  see also hardwoods; softwoods
wood alcohol *see* methanol
wood glue fingerprint spoofing **1**:330, 332
word proportions, handwriting **3**:1442
workflow, 3D reconstruction **5**:2258–60
working surface, tools **5**:2486–7
workplace
  delusions **2**:743–4, 744
  drug testing **2**:860, 866; **3**:1428; **4**:1916–7; **5**:2507
  threat assessment **5**:2460–5
World Mental Health –Composite International Diagnostic Interview (WMH-CIDI) **4**:1880
World Trade Center (WTC) attacks **2**:1031; **3**:1576; **4**:1806
World War II mass graves **4**:1675, 1678
worried child parental alienation **4**:1982
wound track **3**:1383–4
wounding capacity, ammunition **3**:1382–3
wounding channels **5**:2261, 2264–6
wounds
  ballistics **3**:1202–3, 1382
  firearm discharge residue patterns **5**:2353
  position/trajectory **3**:1223
  sharp force injuries **5**:2646–60
  see also gunshot wounds; injuries
woven bone **5**:2558
wrist maturation **1**:188
writing inks **3**:1541
  see also ink. . .
writing instruments **3**:1546–7, 1548, 1550; **5**:2660–3
writing systems **3**:1437
  see also handwriting

wrong treatment category (malpractice) **4**:1702, 1703
wrongdoing, sense of (victims) **2**:645, 647
wrongful execution **4**:1731
wrongfulness, defendant's knowledge **3**:1555
WTC *see* World Trade Center
Wundt, Wilhelm **4**:2177
Wuornos, Aileen **5**:2314
Wyatt v. Stickney, 344 F.Supp. 387 (M.D. Ala. 1972) **5**:2581

X-chromosomal STRs **4**:1806; **5**:2362
X-ray diffraction (XRD) **5**:2384, 2385
X-ray fluorescence (XRF) spectroscopy **3**:1195; **4**:1937; **5**:2353
X-rays **5**:2233, 2234
  aging the living **1**:188
  authentication **2**:840
  best evidence rule **1**:298
  diagnostic use **5**:2235–6, 2236
  explosion debris **2**:1036
  facial reconstruction **3**:1087
  fiber analysis **2**:989–90
  fire investigation **2**:929
  identification use **5**:2237, 2238
  SEM functions **4**:1800
  stab wounds **5**:2647–8
xenobiotics **3**:1427
  see also drug...
xenon lamps **4**:1785
xerography *see* electrophotographic printers
XRD *see* X-ray diffraction
XRF *see* X-ray fluorescence spectroscopy
xylem of wood **5**:2642, 2644
xylenes **3**:1149–50
XYY karyotype **3**:1336

Y-chromosomal short tandem repeats (Y-STRs) **2**:817, 822; **4**:1806; **5**:2362, 2373, 2677–80
Y-chromosomes **2**:821; **4**:1806, 2023
yew tree **4**:2065
Yield scores, suggestibility measurement **2**:1068; **3**:1592
yield strength, materials **4**:1682, 1683
young adults
  age determination of remains **1**:161
  capacity to waive Miranda rights **2**:466
  traffic fatalities **5**:2543
  see also adolescents; children; juveniles; young offenders
Young and Helmholtz theory **4**:2043
young offenders
  death penalty **3**:1608
  Northwestern Juvenile Project **4**:1877–82
  sentencing **5**:2307–8
  separate system policies **3**:1611
  transfer to adult justice **3**:1612–5
  see also juvenile...
Young v State of Maryland, Appellate Court (2005) **2**:832
Youngberg v. Romero, 457 U.S. 307 (1982) **5**:2583
Young's modulus **4**:1682
youth psychopharmacology **4**:2210–6

Z direction (ZD), paper **4**:1972
Zajac, R. **2**:657–8, 659, 660
ZD (Z direction), paper **4**:1972
Zecevic v. DPP (1987), 162 Commonwealth Law Reports 645 **1**:272
zero tolerance approach **5**:2455
  see also per se legislation
zinc chloride **3**:1299
Zinermon v. Burch, 494 U.S. 113 (1990) **5**:2586
Ziskin, Jay **4**:2182
ZivilprozeBordnung (ZPO) **2**:1008, 1009
zolpidem **1**:295; **4**:1915, 1916
zonal approach, textile traces **2**:994
zone models (fire) **3**:1181–3, 1185
zone search strategy **2**:623
zoology **4**:2089
zoom function, video **2**:642
zoom lenses **4**:2038
ZPO (ZivilprozeBordnung) **2**:1008, 1009